PROTEINS: Structure, Function, and Genetics 25:315–334 (1996)

The Application of Different Solvation and Electrostatic Models in Molecular Dynamics Simulations of Ubiquitin: How Well Is the X-Ray Structure "Maintained"?

Thomas Fox and Peter A. Kollman
Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

ABSTRACT **We present molecular dynamics simulations on ubiquitin with explicit solvent molecules and investigate the influence of different force fields [Weiner et al. (J. Am. Chem. Soc. 106:765–784, 1984; J. Comput. Chem. 7:230–252, 1986) vs. Cornell et al. (J. Am. Chem. Soc. 117:5179–5197, 1995)], different treatments of the long-range electrostatic interaction (8 Å cutoff vs. particle mesh Ewald), and different solvation models (periodic box vs. small shell of water molecules) on the structure and the dynamics of the protein. Structural data are monitored by atomic root mean square deviations (RMSDs) from the crystal structure, the radius of gyration, the solvent-accessible surface area, and the pattern of the backbone hydrogen bonds. The dynamic behavior is assessed by the atomic fluctuations and the order parameters of the N-H backbone vectors.**

With the Cornell et al. force field and a periodic box model, the simulated structures stay much closer to the experimental X-ray structure than with the older Weiner et al. force field. A further improvement of the simulation is found when the electrostatic interaction is evaluated with the particle mesh Ewald method; after 1.2 ns of simulation the backbone RMSD amounts to only 1.13 Å. The analysis of the dynamic parameters shows that this good structural agreement is not due to a damping of internal motion in the protein.

For a given length of simulation time, the shell models achieve an agreement between simulated and experimental structures that is comparable to the best models that employ a periodic box of solvent models. However, compared with the box models, the fluctuations of the protein atoms in the shell models are smaller, and only with simulation times as long as 2 ns do they become of comparable size to the experimental ones.

Key words: particle mesh Ewald, cutoff, periodic box, shell, root mean square deviation, atomic fluctuation, order parameters, force field, aqueous solution

INTRODUCTION

Molecular dynamics (MD) simulations have become a powerful tool for studying the structure and dynamics of biologically important molecules.[1–3] With increasing computer power, larger and larger systems can be simulated for increasingly longer periods of time. In the beginning, simulations hardly exceeded a few picoseconds; nowadays, trajectories of hundreds of picoseconds or in the nanosecond region are reported.[4–9] From these studies it has become apparent that water- (and if present) ion-specific interactions are very important in the structural and functional aspects of biopolymer chemistry. Therefore, although simulations that use vacuum conditions[10] or implicit solvent models[4,11,12] have been reported, most current simulation studies use a bath of explicit solvent molecules to maintain a more accurate balance between hydrophobic and hydrophilic interactions.

This can lead to the simulation of a *much larger system;* if one considers explicit water molecules and periodic boundary conditions, the increased number of nonbonded interactions and a concomitant increase of computer time by at least 1–2 orders of magnitude over a simulation in vacuo or with implicit water are consequences. This restricts the sampling in current simulations to the nanosecond timescale or below, whereas many biological processes take at least milliseconds. A limitation of short simulation times, ignoring the problem of sufficient sampling, is that they may not allow one to detect inappropriate approximations or deficiencies in the molecular force field employed, which may become obvious only with longer simulations.

However, even simulations in the 100–500 ps range can provide an important test for currently

Received September 15, 1995; revision accepted January 11, 1996.
Address reprint requests to Peter A. Kollman, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446.

used force fields; a necessary (albeit not sufficient) condition is the production of a stable trajectory that does not deviate too far from the experimentally derived protein structure and that maintains the secondary structure features. This structural stability is particularly important for free energy calculations,[13] in which one would like to identify the effects of small changes in the protein or ligand and not have global, unrelated changes in the overall geometry interfering with this goal. Since often the only way to increase the accuracy of such a calculation is running a longer trajectory, an unrealistic behavior of the MD simulation at longer time scales becomes a crucial problem.

A serious obstacle to creating accurate simulations of macromolecules is the treatment of long-range electrostatic forces.[14–16] To reduce the computational burden from a problem that is proportional to the square of the number of particles to one that scales less steeply with the particle number, the nonbonded interactions are most often neglected beyond a chosen cutoff radius.[17] Experimental justification for this practice stems from mutagenesis studies on T4 lysozyme, which found that the protein stability was influenced only slightly by electrostatic interactions of groups that are separated by more than typical hydrogen-bonding or salt-bridge distances.[18] However, more recent studies found that the truncation of the long-range Coulomb interaction at a finite distance significantly affects the results of MD simulations.[15,19–22]

Alternative ways to treat the calculation of the long-range interactions are now being investigated by a number of research groups. These investigations often take the approach of separating the force on each charge into a short- and a long-range contribution, and they employ special methods for the calculation of the long-range part. Examples of this kind are reaction field methods, which mimic the dielectric response of the bulk solvent[23]; Ewald summation, which assumes a strict periodicity and long-range order[21,24–27]; and various flavors of particle-particle/particle-mesh approaches[28,29] as well as fast multipole methods.[30,31]

The Ewald technique has long been recognized as the correct way to calculate the electrostatic interactions in periodic systems like crystals[32]; however, it has been used less frequently due to its high computational cost. Moreover, as it assumes perfect periodicity, it is still a matter of debate if it can be applied to less ordered solution systems, although simulations of molten salts or concentrated salt solutions have been performed successfully.[33,34]

Recently, an efficient implementation of the Ewald method, the particle mesh Ewald (PME) method, has been developed and incorporated into the AMBER4.1 code.[35] It achieves the full Ewald summation at only a moderate computational burden by interpolating the reciprocal space Ewald sums by B-splines on a grid and by evaluating the convolutions necessary to compute the sums via fast Fourier transform techniques.[7,36] Application to crystal simulations[21,7,37] and to solution simulations of DNA, RNA, and ubiquitin[25] showed the advantage of the Ewald method over conventional cutoff schemes for both crystal and solution simulations of biomolecular systems as well as for simple ions.[38]

In this article we extend and elaborate our previous work on MD simulations of ubiquitin in a periodic box of water molecules.[25,39] We investigate the differences in MD trajectories of up to 1.2 ns length arising from different techniques for handling the long-range electrostatic interaction, calculating it either with a simple cutoff scheme or the PME method. We also address the behavior of the protein calculations with two different all-atom force fields, the Weiner et al.[40,41] and the Cornell et al. force field.[39] The first had been parameterized for vacuum calculations and implicit solvation with a distance-dependent dielectric constant; however, it has been shown that it can also be used with explicit solvent and $\epsilon = 1$.[40,42] In contrast to this, the new Cornell et al. force field has been designed for the use with explicit water molecules.

In this work, we evaluate the stability of the simulated trajectories by monitoring the root mean square deviation (RMSD) of the calculated structures from the experimental X-ray structure, the radius of gyration, the total surface area, and the backbone hydrogen bonds. In addition, we try to assess the dynamic behavior of the proteins by means of the atomic fluctuations from the average structure and the order parameters of the angular correlation function of the backbone N-H vectors.

In addition to the periodic box models, we explored an alternative model to treat protein solvation, whereby the protein is surrounded by only a small shell of water molecules. To damp the effects of the vacuum-solvent boundary on the protein, the electrostatic interaction is calculated with a distance-dependent dielectric constant $\epsilon = \epsilon(R) = R_{ij}$. This model has been suggested as a computationally efficient way to incorporate the effects of explicit solvent into MD calculations and has been applied with considerable success to the trp repressor and BPTI,[10] (Bovine Pancreatic Trypin Inhibitor) resulting in calculated structures that stayed close to the experimental ones. For BPTI, this simulation[10] has been extended to 20 ns[5] with a backbone atom RMSD of the average structure on the order of 1 Å. However, due to the imbalance of the short-range interaction at the water-vacuum surface, the shell exerts an additional pressure on the protein, which might lead to a reduced mobility. Additionally, as TIP3P [53] water was parametrized for $\epsilon = 1$, with a distance-dependent dielectric constant the water-water interaction becomes more attractive, which also might impair the protein's ability to move.

While it has been shown that reducing the water-water interaction energy to its normal TIP3P value by scaling back the charges on the water atoms has almost no effect on the results,[10] the surface tension remains a problem.

Ubiquitin consists of a single polypeptide chain of 76 amino acid residues. It is probably present in all eukaryotic cells and has been isolated from a variety of sources. It appears to be one of the most conserved of all eukaryotic proteins,[43] its primary role being in the intracellular ATP-dependent protein turnover.[44] A crystal structure of human erythrocytic ubiquitin at 1.8 Å resolution has been reported,[45] which shows that the overall structure of ubiquitin is extremely compact and tightly hydrogen bound, but without any disulfide bridges. The most prominent secondary structure features are an α-helix with three and a half turns involving residues 23–34, a short 3_{10}-helix at residues 56–59, and a five-stranded β-sheet employing residues 1–7, 10–17, 40–45, 48–50, and 64–72. Both the α-helix and the β-sheet form a highly hydrophobic core, which accounts for the remarkable stability of ubiquitin.[46] The four C-terminal residues 73–76 are protruding from the protein into the surrounding solvent. They are only poorly refined in the crystal structure, indicating a high mobility. Nuclear magnetic resonance (NMR) studies on the protein in solution confirm these findings on the secondary structure based on NOE (Nuclear Overhauser Effect) connectivity patterns; however, no detailed geometry information was given.[47,48] These studies also report the C-terminus to be highly mobile; in addition, they suggest a high mobility for the Lys side chains of residues 6, 11, and 63. In contrast to the findings from the crystal structure, Lys33 does not seem to be freely rotating in solution.

Given the significant biological importance of ubiquitin, the wealth of experimental structural information, and its rather small size, it is somewhat surprising that only few theoretical studies of ubiquitin have been published so far. However, with the exception of a 3 ns simulation with largely implicit solvent[4] and a series of articles that used high-temperature MD to investigate the unfolding and renaturing process,[8,49,50] no other attempt to investigate the structural characteristics of ubiquitin by means of unrestricted MD has come to our attention. Recently, a 1 ns simulation of ubiquitin in explicit water using Ewald summation to treat the electrostatic interaction was published.[6] However, this work focused on the calculation of radial and reorientational correlations functions to analyze NMR cross-relaxation data, and apart from a net RMSD from the crystal structure, no other data on the simulated structures have been given.

MATERIALS AND METHODS

All simulations were carried out with the SANDER module of the AMBER4.1 suite of programs,[51] using the all-atom force fields by Weiner et al.[40,41] or Cornell et al..[39] The starting point was the reported X-ray structure of human erythrocytic ubiquitin, refined at 1.8 Å.[45] It was taken from the Brookhaven Protein Data Bank (entry 1ubq)[52] and included the 58 crystal waters used in the refinement. The hydrogen atoms were added using standard bond lengths and angles. The protonation states for the charged residues were decided based on their ionization states at pH 7. Thus, Glu and Asp were negatively charged, and Lys and Arg were positively charged. The hydrogen of His68 was arbitrarily added at the δ-position, as neither of the alternatives (δ- or ε-position) leads to the formation of a prominent hydrogen bond. Depending on the solvation model, the program EDIT was used to construct either a rectangular box or a shell of pre-equilibrated TIP3P water molecules[53] around the protein. For the simulations with the Weiner et al. force field this resulted in a system containing 4,428 solvent molecules, whereas in the simulations with the Cornell et al. force field 4,385 water molecules were employed. The different number of water molecules results from the use of lone pairs at sulfur in the Weiner et al. force field that are not present in the Cornell et al. force field, as well as round-off errors in the water placing routine of the EDIT module. The shell model as described by Guenot and Kollman[10] consisted of 500 water molecules (506 for the Cornell et al. force field). Additionally, with the Cornell et al. force field, a larger shell version of 1,439 water molecules was tested. Although several charged side chains are present, the net charge of ubiquitin is zero, and therefore no counterions were added. After initial minimization of the solvent down to an RMS gradient of 0.5 kcal(mol Å)$^{-1}$, the water molecules were subjected to 12 ps of molecular dynamics, holding the protein rigid. This was followed by six minimizations of the whole protein/water system with the protein backbone atoms restraint to the crystal positions by force constants of 1,000, 100, 50, 15, 2, and 0 kcal/mol. Then two 3 ps MD simulations at 100 K and 200 K were performed, followed by the production runs.

We used SHAKE on bonds involving hydrogen atoms, a timestep of 1.5 fs, and an 8 Å residue based nonbonded cutoff. In the CUTSS model (Table I), all nonbonded interactions within the solute were evaluated. The nonbonded pairlist was updated every 10 steps. Temperature was held at 300 K with the Berendsen method,[54] the solute and the solvent being coupled separately to the temperature bath to avoid the hot solvent-cold solute problem.[10] For the box models, constant pressure periodic boundary conditions at 1 bar and a constant dielectric constant of $\epsilon = 1$ were used. This resulted, after equilibration, in a box size of approximately 63 × 50 × 46 Å^3. With the shell models, a distance-dependent dielectric constant of $\epsilon = R_{ij}$ was used. In the Weiner_shell

TABLE I. Models Used in This Study

Model	Force field*	Solvent model	$\epsilon^{\dagger}$	No. of water molecules	Nonbonded cutoff	CPU time (min/ps)‡
Weiner_CUT	Weiner	Box	1.0	4,428	8Å	54.1
Weiner_PME	Weiner	Box	1.0	4,428	8Å§	85.4
Weiner_shell	Weiner	Shell	R_{ij}	500	8Å	13.5
Cornell_CUT	Cornell	Box	1.0	4,385	8Å	67.9
Cornell_CUTSS	Cornell	Box	1.0	4,385	8Å¶	82.7
Cornell_PME	Cornell	Box	1.0	4,385	8Å§	95.8
Cornell_shell	Cornell	Shell	R_{ij}	506	8Å	8.6
Cornell_xshell	Cornell	Shell	R_{ij}	1,439	8Å	22.5

*From Weiner et al.[40] and Cornell et al.[39]
†Dielectric constant.
‡On a HP 9000/735 computer.
§For van der Waals interaction only.
¶All solute–solute interactions evaluated.

simulation, the translational and rotational degrees of freedom were removed every 10 steps. Coordinates were sampled every 0.75 ps in the simulations with the Cornell et al. force field; the corresponding value for the Weiner et al. trajectories was 1.5 ps.

In the PME model a charge grid of 64 × 54 × 48 gridpoints was used, giving a spacing of the grid points of approximately 1 Å. The charge of the unit cell was interpolated with cubic splines on this grid, and the tolerance for evaluating the direct sum was set to 10^{-5}, resulting in estimated RMSD force errors of $5 \cdot 10^{-4}$. The van der Waals contribution to the nonbonded energy was evaluated conventionally employing an 8 Å group-based cutoff radius. In Table I we summarize the pertinent features of the different models used in this study.

The program CARNAL[55] was used for the calculation of the RMSDs and the hydrogen bond analysis of the obtained trajectories. In analogy to previous work,[4] hydrogen bond was considered formed if the distance between donor and acceptor was less than 3.5 Å and the angle between the donor-hydrogen and the donor-acceptor bond was less than 20°. The average structure of a simulation and the fluctuations of the protein atoms around their average positions were calculated with the MDANAL module of AMBER4.1. The atomic fluctuation $\langle \Delta \mathbf{r}^2 \rangle^{1/2}$ can be connected to the crystallographic B factors via

$$B = \frac{8\pi^2}{3} \langle \Delta r^2 \rangle$$

The two-dimensional RMS plots were generated with the program MOIL-VIEW.[56] To obtain the rotational order parameters, we calculated the rotational correlation function

$$C_i(t) = \langle P_2(\mu_i(0) \cdot \mu_i(t)) \rangle$$

with $\mu_i(t)$ the unit vector in the direction of the N-H bond of residue i, and $P_2(x)$ the second Legendre polynomial. The sliding time window for the calculation of the correlation function was 150 ps, with coordinates sampled every 0.75 ps. For the simulations with the Weiner et al. force field, we used a 90 ps window, and a sampling rate of 1.5 ps. $C_i(t)$ was then fitted to a target function[57]

$$g_i(t) = [S^2 - (1 - S^2)e^{-t/\tau}]e^{-t/\tau_{snr}}$$

with order parameter S^2 and rotational correlation time τ, the overall motion of the protein is accounted for by τ_{snr}.[6] The quality of the fit was monitored by

$$\chi^2 = \frac{1}{N}\sum_N (g(t) - C(t))^2$$

which was on the order of 10^{-4} or below. In contrast to the results of Abseher et al.,[6] using a multiexponential target function with more than one order parameter did not improve the fit; thus, we used the simple form of $g(t)$ as given above.

RESULTS

The most direct way to assess the stability of a MD simulation over the course of time is the evaluation of the difference between the initial experimental coordinates and the generated structures, measured by the RMSD.

One could argue about the significance of crystal structures for the evaluation of solution trajectories. However, a recent study of the structures of globular proteins for which both single crystal X-ray data and solution NMR structures are available showed a close coincidence of both methods despite the different environment in which the data had been obtained.[58] Discrepancies can be attributed to real differences due to the different protein environments (packing forces, pH, ionic strength, solvent), but may also be due to subjective assumptions and interpretations made during the analysis of the raw NMR or X-ray data. The RMSDs between the NMR and the X-ray-derived structures amounted to 1–2 Å for all backbone atoms and 1.5–2.5 Å for all heavy

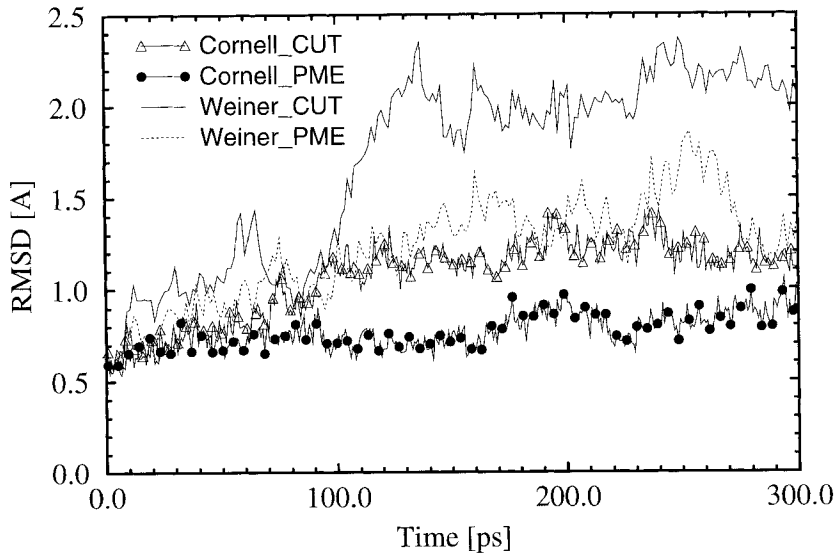

Fig. 1. Backbone atom RMSD from the experimental crystal structure (in Å) as a function of simulation time for the models with periodic box solvent. Only amino acids 1–72 are taken into account. For the names of the models see Table I.

atoms. If the atoms with extremely high deviations were discarded, the deviations reduced to 0.6–1.3 Å for the backbone and to 0.8–1.7 Å for the nonhydrogen atoms.[58] This is also in line with findings on BPTI, where the RMSD values for the three available crystal structures to the mean solution structure are 0.76–0.85 Å for the backbone atoms and 1.23–1.33 Å for all heavy atoms of residues 2–45.[59]

In Figure 1, we present the backbone RMSDs for four of the periodic box models used, Weiner_CUT, Weiner_PME, Cornell_CUT, and Cornell_PME. As discussed above, residues 73–76 at the C-terminal end represent an extended tail sticking out into the solvent and exhibit a high mobility. Therefore this region was excluded from the calculation of the RMSDs. Here, t = 0.0 ps corresponds to the start of the MD simulation at 100 K, and the initial RMSD of 0.6 Å from the crystal structure is due to the minimization of the experimental structure.

With the Weiner et al. force field (model Weiner_CUT; see also Table I), the RMSD of the backbone atoms stays at about 1.1 Å for the first 100 ps of the simulation; then it increases within the next 50 ps to about 2 Å and stays at this value for the rest of the simulation. A second trajectory (not shown), obtained with identical starting conditions on a different computer, does not show this first plateau region, but rather exhibits a steady increase of the RMSD over the 180 ps of the simulation, also ending at an RMS value of 2 Å. With the new Cornell et al. force field (model Cornell_CUT), the backbone RMSD at 300 ps amounts to 1.2 Å, considerably lower than the corresponding value for the Weiner_CUT model. Here again, we observe an initial plateau of the RMSD that changes between 70 and 100 ps to its final value. If we use a modified cutoff scheme whereby all solute-solute interactions are evaluated regardless of the distance of the atomic centers (Cornell_CUTSS), the results hardly change. The initial equilibration to the final RMS value seems to take place somewhat faster, but after 100 ps the RMSDs are essentially the same. However, considering the results for the two trajectories for the Weiner_CUT model, this different behavior at the beginning of the simulation is probably not a significant difference between the two electrostatic models.

Using the Weiner et al. force field with the PME method leads to an almost monotonic increase of the RMSD up to about 1.8 Å over the first 250 ps of the simulation. The RMSD then decreases to about 1.4 Å at the end of the 300 ps trajectory—a clear improvement over the simple cutoff scheme. The least deviation from the experimental structure is found with the Cornell_PME model. Up to 170 ps the RMSD stays below 0.8 Å, rising somewhat toward the end of the simulation. However, with a final value of 1.0 Å for the backbone atoms, it is still significantly lower than those for the pure cutoff models.

Given the success of the Cornell et al. force field in producing trajectories that stay close to the experimental structure, we continued the Cornell_CUT and Cornell_PME models up to 1.2 ns. As can be seen from Figure 2, the RMSD for the backbone atoms fluctuates around 1.3 Å for the Cornell_CUT trajectory. At two points of the simulation, between 500 and 600 ps and between 900 and 1,000 ps, the RMSD is higher and briefly exceeds 1.8 Å. However, at the end of the 1.2 ns trajectory, the RMSD reduces

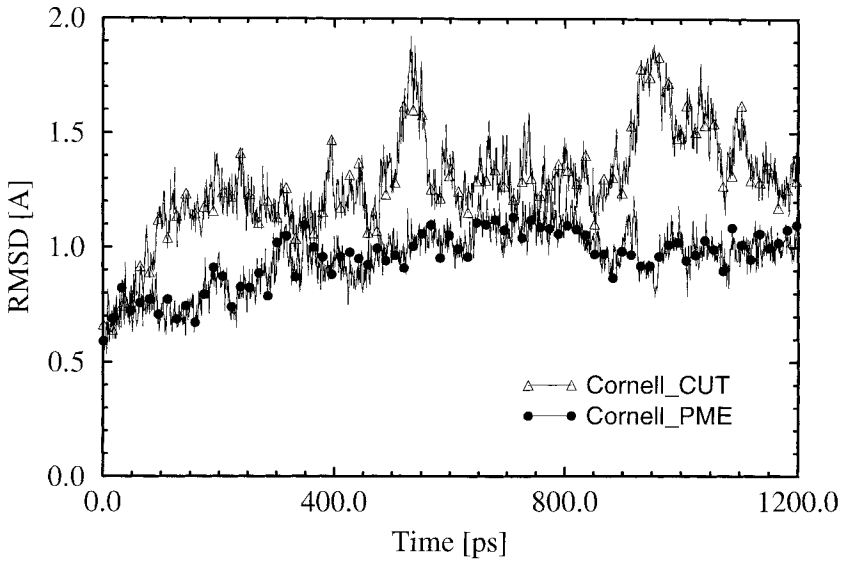

Fig. 2. Backbone atom RMSD from the experimental crystal structure (in Å) for the models Cornell_CUT and Cornell_PME up to a simulation time of 1.2 ns.

to about 1.4 Å. The backbone RMSD of the Cornell_PME simulation increases slowly up to a maximum of 1.2 Å at around 750 ps and then comes back down to an average value of 1 Å. Additionally, the fluctuations of the RMSD are much smaller than those obtained with the Cornell_CUT model.

The RMSD values for both the α carbon atoms and for all heavy atoms follow the trends outlined for the backbone atoms very closely. Therefore we only give their final values at the end of each simulation in Table II, as well as their average values over the whole simulation.

A more informative measure than the gross RMSD from the original crystal structure is a two-dimensional RMSD plot that allows visualization of the mutual RMSDs between the different structures generated during the simulation. This is a tool to assess if a constant RMSD from some reference structure really corresponds to a stable conformation, and it shows (if present) the correlation between different substructures of a trajectory.

The two-dimensional RMSD plot for Cornell_CUT is shown in Figure 3a. Clearly visible are the regions of higher RMSD at around 550 and 950 ps. Evidently the protein returns after these peak regions to the previous conformation, as indicated by the lower mutual RMSD; these short periods do not lead to a new overall conformation of the protein.

Compared with Cornell_CUT, the two-dimensional RMSD plot for Cornell_PME is relatively smooth (Fig. 3b). With the exception of a few frames at about 950 ps, the mutual backbone RMSD during the last 900 ps of the simulation differs by less than 1 Å, in good agreement with the net RMSD from the crystal structure.

With the Weiner et al. force field (model Weiner_CUT, Fig. 3c), a somewhat different behavior is observed. During the course of the 300 ps trajectory, three distinct substructures with transitions at about 100 and 190 ps can be distinguished. Within each of the structures, the backbone RMSD is less than 1.5 Å, the RMS between frames from different structures is at least 1.5 Å, and frames from the first third of the simulation differ from frames taken from the last third by at least 2 Å. Over the whole course of the simulation, the RMSD between any two structures increases with their distance in time, indicating that (although the overall backbone RMSD to the crystal structure reaches a maximum of approximately 2 Å) the protein still undergoes considerable conformational changes.

This behavior is changed when the electrostatic interaction is evaluated with the PME approach (Weiner_PME, Fig. 3d). After an initial equilibration period of about 75 ps, the simulation converges to an ensemble of configurations that have a mutual RMSD of less than 1.5 Å.

The variation of the radius of gyration (R_{gyr}) as a function of time is plotted in Figure 4 for the first 300 ps of the simulations. Starting from the value of 11.63 Å for the crystal structure, all box models evolve to a larger R_{gyr}, indicating a slight expansion of the protein during the simulation. Over the first 100 ps of the simulation, the four box protocols lead to comparable results. Parallel to the evolution of the RMSD, the Weiner_CUT simulation shows a pronounced increase of R_{gyr}, with a maximum of 12.25 Å, 5.4% higher than the starting radius (Fig. 4b). The R_{gyr} for Weiner_PME is also enlarged; however, especially in the last two-thirds of the simula-

TABLE II. RMSD (Amino Acid Residues 1–72) From the Crystal Structure of the Final and the Averaged Simulated Structures for the Different Models in Å

		Final			Average		
Model	Length*	C_α	Backbone	Heavy atoms	C_α	Backbone	Heavy atoms
Weiner_CUT	300	2.004	2.041	2.902	1.311	1.346	1.888
Weiner_PME	300	1.424	1.441	1.943	0.957	0.985	1.315
Cornell_CUT	300	1.149	1.131	1.834	0.992	1.030	1.467
Cornell_CUT	1,200	1.377	1.379	2.133	1.479	1.460	1.941
Cornell_CUTSS	300	1.302	1.304	2.007	0.944	0.944	1.557
Cornell_PME	300	1.011	1.020	1.445	0.554	0.572	0.878
Cornell_PME	1,200	1.075	1.125	1.535	1.028	1.060	1.310
Weiner_shell	300	0.823	0.904	1.155	0.685	0.747	0.951
Weiner_shell	2,142	1.142	1.204	1.689	0.827	0.873	1.208
Cornell_shell	300	0.822	0.825	1.100	0.567	0.576	0.809
Cornell_xshell	300	0.623	0.678	0.999	0.424	0.452	0.673

*Length of trajectory in ps.

tion it remains considerably below the corresponding values of Weiner_CUT. The Cornell_CUT trajectory also shows a significant increase of R_{gyr}, although not as much as the simulations with the Weiner et al. force field. The values for Cornell_CUTSS (not included in Fig. 4) stay at about 11.8 Å up to 180 ps, and then drop to approximately the crystal value. The reason for this behavior is a change in the conformation of the C-terminal tail that folds into a position closer to the rest of the protein. Consequently, when R_{gyr} is calculated only for the first 72 residues, this drop is not observed. The R_{gyr} for the Cornell_PME simulations shows the least variation, oscillating around an average value of 11.76 Å. This feature remains unaltered when the whole 1.2 ns trajectory for Cornell_CUT and Cornell_PME is considered. R_{gyr} for the Cornell_CUT simulation oscillates around an average value of 11.90 Å; for Cornell_PME, the average R_{gyr} is 11.73 Å.

The findings for R_{gyr} are mirrored by the behavior of the solvent-accessible surface area, whose evolution over the first 300 ps is shown in Figure 5. With the Weiner_CUT model, the surface area stays at about 5,000 $Å^2$ for the first 100 ps, and then increases to about 5,500 $Å^2$ for the rest of the simulation. Compared with the value for the crystal structure, 4,785 $Å^2$, this is an increase of about 13%. In contrast, the surface area for Weiner_PME stays over the whole 300 ps simulation around the value reached after the initial equilibration period. The values for the new force field stay rather close to the experimental value; during the entire 1.2 ns simulations, Cornell_CUT oscillates around 5,100 $Å^2$, and Cornell_PME oscillates around 4,900 $Å^2$.

A more detailed picture of the intactness of the initial structure during the simulations is obtained by looking at the hydrogen bonds between the amide hydrogens and the carbon oxygens of the protein backbone. From the X-ray structure, 40 backbone-backbone hydrogen bonds were inferred[45] (Table III). After minimization in the solvent box with the Cornell et al. force field, 32 of these 40 hydrogen bonds are also found. The most prominent differences occur in the large loop formed by residues 51–59. Of the three hydrogen bonds within the 3_{10} helix, only one is found with the Cornell force field, and the hydrogen bonds 56–21 and 57–19 are also missing. These findings also hold for the Weiner force field (results not shown); here two additional hydrogen bonds in the α-helix are misplaced, and the hydrogen bonds 41–38 and 42–70 at the beginning of the β-strand are missing.

With the exception of the hydrogen bond 56–21, all these missing hydrogen bonds are also only rarely populated during the course of the simulations. With all models used, the hydrogen bonds in the experimentally observed 3_{10} helix are formed in less than 40% of the time; of all box models, only with Cornell_PME is the $i \rightarrow i + 3$ pattern found more than 10% of the time. However, the general shape of the protein backbone in this solvent-exposed region remains helix-like. A detailed analysis shows that the hydrogen bonds in this region are continually formed and broken; therefore, these findings are not due to the "locking" of the protein in some particular structure for a longer period of time. The hydrogen bond pattern of the β-sheet and the α-helix is generally retained. The one exception was in the Cornell_CUTSS simulation where the α-helix, starting from its N-terminal end next to a section without a defined secondary structure (residue 34), starts to lose its specific hydrogen-bond pattern and its helix structure. In the periodic box simulation with the Weiner force field, additionally the ends of the individual β-strands are losing their hydrogen bonds; in strands 1–7 and 10–17 all hydrogen bonds are populated only for a small percentage of all saved structures. In most cases this is due to twists in the protein backbone that prohibit the nec-

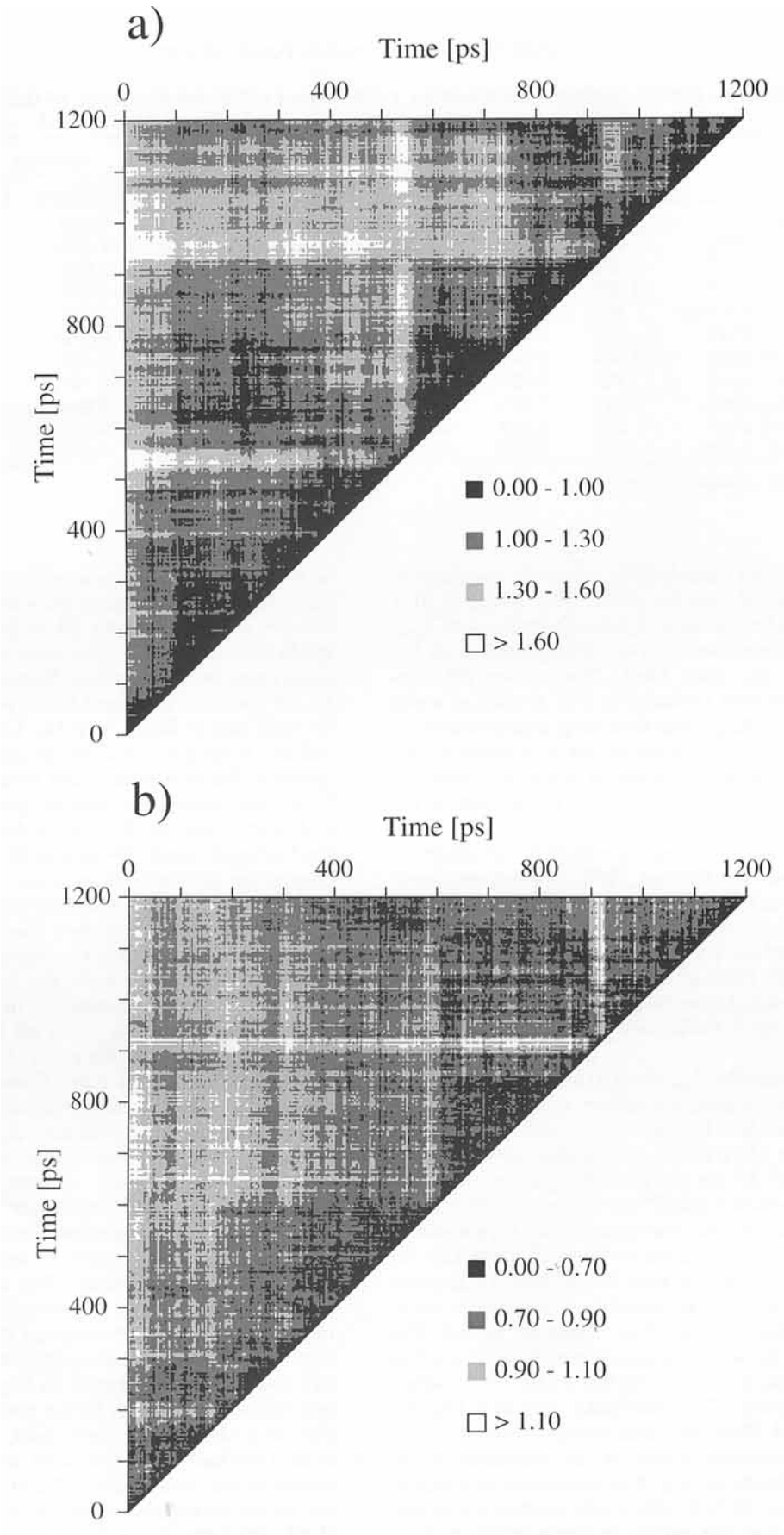

Fig. 3. Backbone atom 2D-RMS plots for (**a**) Cornell_CUT (1.2 ns trajectory); (**b**) Cornell_PME (1.2 ns trajectory); (**c**) Weiner_CUT (300 ps simulation); (**d**) Weiner_PME (300 ps simulation). Only residues 1–72 were taken into account.

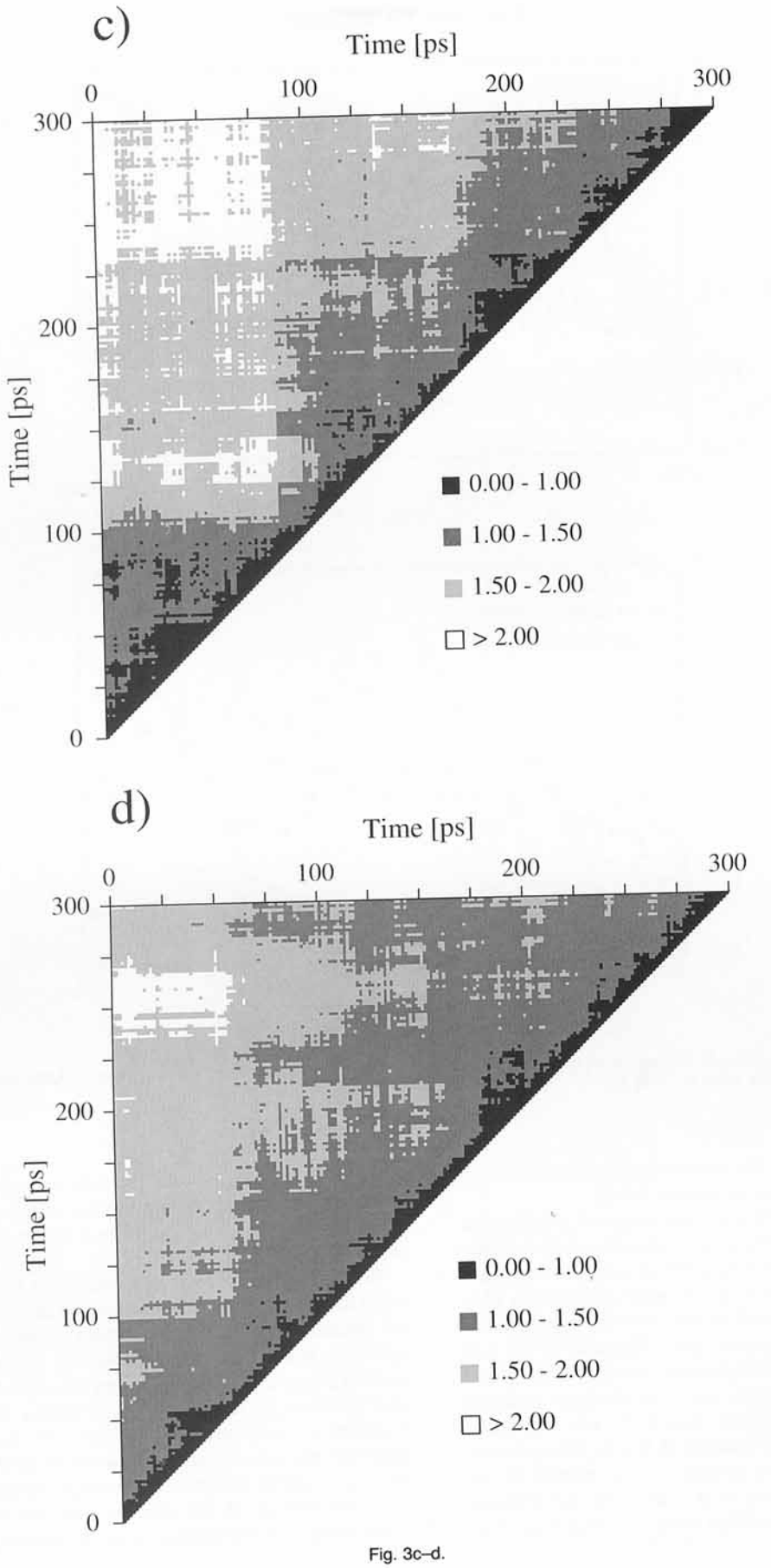

Fig. 3c–d.

Fig. 3c–d.

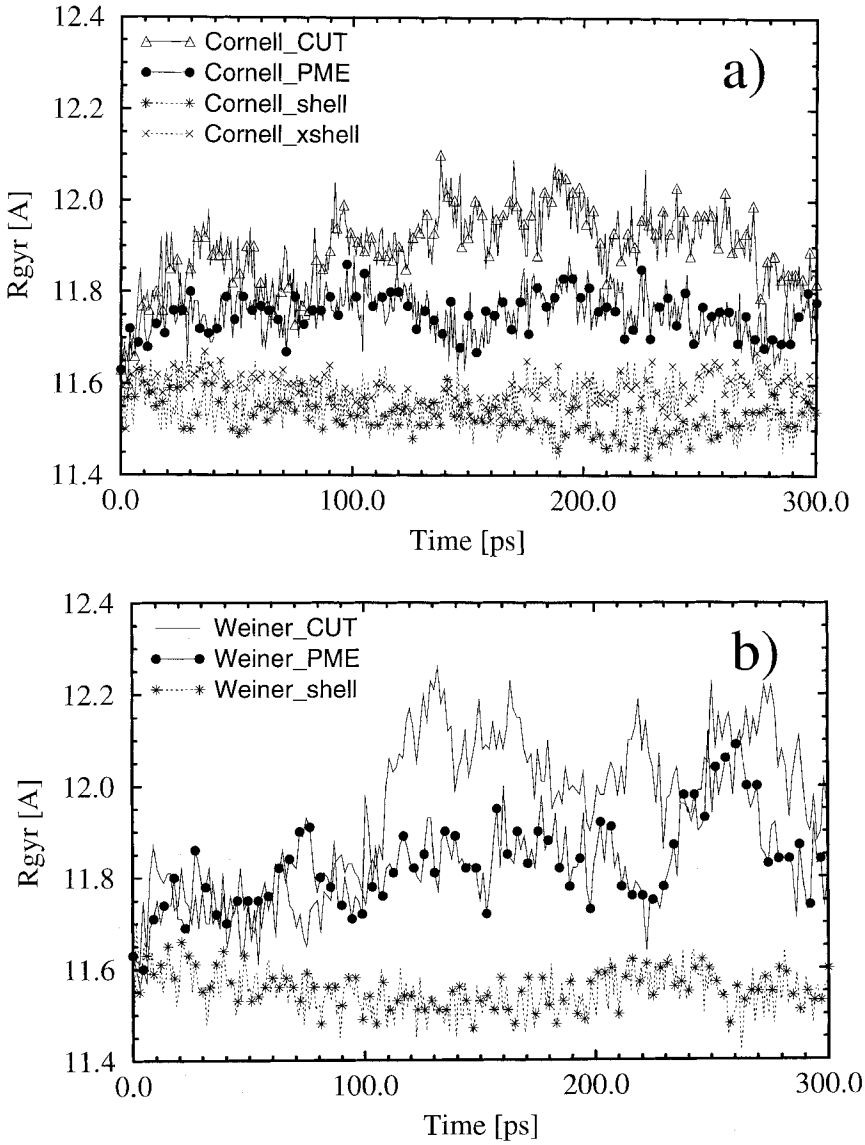

Fig. 4. Radius of gyration (in Å) of ubiquitin as a function of simulation time. **a:** Simulations with the Cornell et al. force field. **b:** Weiner et al. force field. The X-ray value is 11.63 Å.

essary alignment of the carbonyl oxygens and amide nitrogens to form the hydrogen bond.

Whereas the NMR analysis suggests a hydrogen bonding scheme is consistent with the crystal structure, it differs in some significant points. For example, it finds in the β-bulge of residues 7–11 an 11-7 hydrogen bond instead of the 10-7 hydrogen bond inferred from the crystal data. Whereas in all our simulations the 10-7 hydrogen bond is only populated at most 40% of the time, no evidence is found for a 11-7 hydrogen bond. Similarly, the suggested 65-2 hydrogen bond, instead of the 64-2 bond from the crystal structure, cannot be reproduced in our calculations. However, as a 3 ns ubiquitin simulation with largely implicit solvent conditions has shown,[4] it may take considerable time to break a specific hydrogen bond and form another one, so these discrepancies may still be a result of our limited simulation times.

An important concern in performing MD simulations is not only the reproduction of the experimental structure, but also the correct modeling of the dynamic behavior—close agreement to the experimental structure by damping important intramolecular motions or freezing out molecular degrees of freedom is certainly undesirable. To assess this problem, we calculated the fluctuations of all nonhydrogen atoms around the average structure over the first 300 ps of the simulations and compared them with the crystallographic B factors. The re-

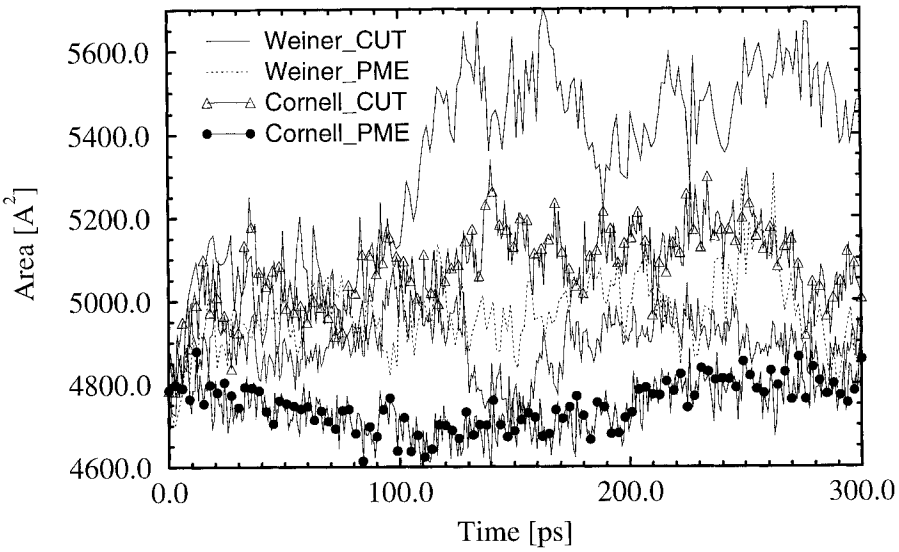

Fig. 5. Solvent accessible surface area (in Å^2) for some of the models studied as a function of simulation time.

sults for the box models, averaged over individual amino acid residues, are displayed in Figure 6.

The fluctuations in the simulations with the Cornell et al. force field are of the same order of magnitude as the experimental values. Most significantly, the PME calculation results in atomic fluctuations that match the experimental B factor pattern closely. Here, the two major discrepancies occur at residues 8/9 and at residues 45–48, where the calculated fluctuations significantly exceed the the experimental ones. Both regions, a β-bulge and a short strand of the β sheet, lie at the surface of the protein and exhibit strong interactions with the solvent. Therefore, a higher mobility than in the crystal is not unreasonable.

The fluctuations obtained with Cornell_CUT are generally larger than those for Cornell_PME. This is especially pronounced for residues 9, 11, 18, 27, 33, 38–39, 49, and 52. In addition to residues 8, 9 and 45–48, these regions constitute the parts of the protein where the calculated fluctuations considerably exceed the experimental ones. With the exception of Lys27, all these residues are located at the surface of the protein. Whereas the enhanced mobility of Lys11 in the Cornell_CUT models concurs with the ^{1}H-NMR analysis, the large fluctuations of Lys33 are not supported by these data. Cornell_PME, in contrast, finds only small fluctuations for either residue. In addition, both models find large fluctuations for Lys48, whereas the solvent NMR data show nondegenerate hydrogen resonances that indicate a rigid side chain. Lys63, which according to the NMR data is mobile, does not show unusually high fluctuations in any of the models.

The Cornell_CUTSS model generally follows the results for Cornell_CUT closely, with significantly increased fluctuations at residues 32, 58, 60, and 68, while residues 11, 27, and 33 are less mobile. Therefore, we did not include it in Figure 6.

The fluctuations obtained with the Weiner_CUT model are significantly larger than those calculated with the Cornell et al. force field or measured experimentally. This is particularly the case for the residues Thr12, Lys33, Arg42, Lys48, Asp52, and Lys63. Interestingly, these are all polar or charged residues at the protein surface capable of forming hydrogen bonds with the explicit solvent molecules. These huge fluctuations are substantially decreased when the PME method is used; they are now generally on the same order of magnitude as those derived experimentally or obtained from the calculations with the Cornell et al. force field. Exceptions are residues 8–11, 18–19, and 45–48; however, these residues also show enhanced fluctuations in the 1.2 ns simulations with the Cornell et al. force field presented below.

In Figure 7 we present the fluctuations of the heavy atoms for the 1.2 ns simulation of Cornell_CUT and Cornell_PME. Compared with the results for the first 300 ps of the trajectory, Cornell_CUT shows significantly enhanced fluctuations for almost all residues. In contrast, the fluctuations for Cornell_PME exhibit only a moderate increase from the 300 ps simulation, with the exception of residues 11, 47, and 54. The fluctuation for residue 11 even exceeds that for the Cornell_CUT simulation. However, in general the calculated fluctuations for the PME model agree very well with the atomic fluctuations derived from the crystallographic B factors. As already seen in the 300 ps simulation, the solvent-exposed residues 8, 11, 45–48,

TABLE III. Percent Occupancies for Backbone Hydrogen Bonds of the Protein (Donor: Amide Nitrogen; Acceptor: Carbonyl Oxygen) Present Either in the Experimental X-Ray or NMR Structure or Found in More Than 10% of the Saved Structures During the 300 ps Trajectories* Were Used

Donor	Acceptor	Exp.	2	Min	WCUT	WPME	CCUT	CCUTS	CPME	Wsh	Csh	Cxsh
Met1	Val17	X N	b	x	68.00	74.50	57.75	40.50	75.94	72.39	90.25	84.50
Ile3	Leu15	X N	b	x	70.50	60.00	86.44	83.75	74.81	81.35	91.00	81.00
Phe4	Glu64				31.00					72.56		
Phe4	Ser65	X N	b	x	28.50	79.50	80.94	85.00	86.81		85.25	86.50
Val5	Ile13	X N	b	x	46.00	59.50	78.19	77.50	76.81	58.46	83.75	88.25
Lys6	Leu67	X N	b	x	36.50	85.50	83.94	87.75	85.75	83.39	81.75	83.75
Thr7	Gly10	X N	b	x		12.00						
Thr7	Lys11	X N	b	x	25.50	31.00	68.81	77.75	75.13	43.98	83.50	89.00
Gly10	Thr7	X	b/t	x		24.00	38.94	17.25	34.31	22.76	26.50	38.00
Lys11	Thr7	N	t									
Ile13	Val5	X N	b	x	22.00	59.00	69.31	41.50	58.06	64.84	91.00	60.50
Leu15	Ile3	X N	b	x	39.50	60.50	62.44	69.00	62.38	35.06	71.25	75.75
Val17	Met1	X N	b	x	33.00	56.50	32.19	17.75	90.81	86.53	77.75	91.25
Asp21	Glu18	X N	t	x	48.50	39.00	77.38	81.25	87.38	89.61	91.75	93.50
Ile23	Arg54	X		x	19.00	78.00	34.00	59.50	85.81	80.01	85.00	89.75
Glu24	Asp52	X				14.00		35.00	22.25	41.95	42.75	44.75
Asn25	Thr22					18.00	19.63					
Val26	Thr22			x	61.00	31.00	78.31	83.25	75.63	81.95	71.75	75.50
Lys27	Ile23	X N	a	x	75.00	88.50	84.69	78.25	87.13	91.34	90.25	95.00
Ala28	Glu24	X N	a	x	73.50	64.00	32.56	54.50	48.19	58.59	60.25	50.50
Lys29	Asn25	X N	a	x	77.00	60.50	81.50	79.00	79.00	63.30	81.00	80.50
Ile30	Val26	X N	a	x	70.00	75.50	64.75	40.75	58.06	56.78	52.50	62.00
Gln31	Lys27	X N	a	x	72.00	74.00	78.63	10.75	69.31	86.43	84.25	90.50
Gln31	Ala28			x				19.00				
Asp32	Ala28	X N	a	x	55.50	68.00	60.31		56.88	63.64	48.25	51.00
Lys33	Lys29	X N	a		21.00	82.50		16.50	10.50	17.65		
Lys33	Lys30						42.56	11.75	22.38	15.11	49.75	44.75
Glu34	Ile30	X N	a	x	13.50	29.00	57.31	36.00	81.00	71.56	86.25	91.75
Gly35	Gln30				20.50	19.00						
Gly35	Gln31				18.50			10.50	10.06	33.02	10.00	
Ile36	Glu34				10.50		11.06			16.11		
Gln40	Pro37	X N	t		21.50	12.50	60.38	69.00	35.31	17.05	40.50	10.00
Gln41	Pro37				37.50	21.00	17.69		14.50	81.68		30.00
Gln41	Pro38			x	11.00	19.50	20.25	36.00	15.75		35.25	
Arg42	Val70	X N	b	x	31.50	49.00	23.69	39.50	49.00	49.30	66.75	45.00
Ile44	His68	X N	b	x	71.50	54.00	80.13	61.75	84.50	76.77	65.00	81.50
Phe45	Lys48	X N	b	x	32.00	73.00	67.94	57.25	81.75	81.78	88.00	67.75
Lys48	Phe45	X N	b/t	x			47.25	60.00	68.69	11.16	57.75	67.25
Lys48	Ala46					12.00				16.44		
Leu50	Leu43	X N	b	x	77.50	66.50	64.38	32.25	81.25	91.21	80.25	86.75
Arg54	Glu51	X	t	x		19.50	26.06	45.00	57.50	60.59	72.50	75.50
Leu56	Asp21	X			46.50	57.00	74.50	72.50	43.06	60.93	66.50	58.00
Ser57	Pro19					19.00	17.63	33.25	20.31	12.83	28.75	59.50
Asp58	Thr55	X N	3		29.50	28.00	20.75		12.81		22.50	12.00
Tyr59	Thr55				33.00		32.56	25.75		68.18		38.00
Tyr59	Leu56	X N	3						15.75		10.50	
Asn60	Leu56				28.50	17.00						
Asn60	Ser57	X N	3	x			11.56	15.75	40.50	21.42	33.50	23.00
Ile61	Leu56			x	34.00	58.00	52.00	65.75	69.25	82.79	68.00	84.25
Glu64	Gln2	X	b	x	38.50	65.50	62.31	81.50	80.69	48.70	92.00	90.25
Ser65	Gln2	N										
Ser65	Gln62	X	t				13.38	17.25			11.25	
Ser65	Gln63				10.50					17.48		
Leu67	Phe4	X N	b	x	33.00	70.50	82.75	76.25	80.63	38.60	89.50	66.75
His68	Ile44	X N	b	x	59.50	20.00	81.56	73.00	73.38	69.28	85.25	73.75
Leu69	Lys6	X N	b	x	78.00	74.00	73.63	71.25	52.38	76.94	65.25	76.25
Val70	Arg42	X N	b	x	16.00	59.00	87.31	90.00	89.94	85.43	91.75	76.75
Arg72	Gln40	X	b	x	61.00	66.50	81.50	76.00	54.69	20.79	60.25	58.50
Leu73	Leu71										31.75	16.25

*Weiner shell, 2.142 ns trajectory; Cornell_CUT and -_PME, 1.2 ns trajectories. Exp., found in experimental X-ray[45] or NMR structure[47,48]; 2, secondary structure (a, alpha helix; b, beta sheet; 3, 3_{10}-helix; t, turn); Min, found in minimization with Cornell force field; WCUT, Weiner_CUT; WPME, Weiner_PME; CCUT, Cornell_CUT (1.2 ns); CCUTS, Cornell_CUTSS; CPME, Cornell_PME (1.2 ns); Wsh, Weiner_shell (2.142 ns); Csh, Cornell_shell; Cxsh, Cornell_xshell.

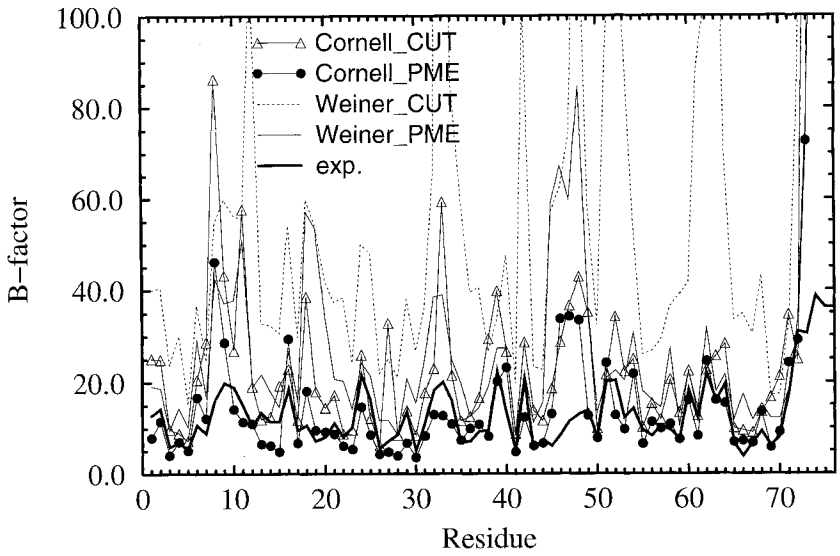

Fig. 6. Calculated B factors for all heavy atoms, averaged over individual residues, for the periodic box models studied. The experimental value is taken from Vijay-Kumar et al.[45]

and 54 show a significantly enhanced mobility compared with the crystal data.

Another possibility for linking simulated protein dynamics to experimental data is order parameters, which are accessible through NMR measurements and can also be calculated from rotational correlation functions. We probed the main-chain dynamics of ubiquitin by calculating the order parameters for the backbone N-H vectors; comparison of both the Cornell_CUT and the Cornell_PME model with experimentally deduced values[60] is shown in Figure 8. With the exception of residue Val17, the calculated order parameters for both simulation protocols are very similar, the values for the Cornell_PME model tending to be somewhat larger, indicating less movement in the protein. The experimental order parameters are consistently lower than the calculated ones; the average order parameter for residues 1–72 is 0.767 compared with 0.865 for Cornell_PME and 0.837 for Cornell_CUT. This discrepancy might again be due to the rather short MD trajectory. However, calculating the correlation function over shorter portions of the 1.2 ns trajectory did not change the order parameters significantly, thus suggesting that a longer simulation would not necessarily result in lower order parameters.

The order parameters for the simulations with the Weiner et al. force field reflect the higher RMSD values and atomic fluctuations. The average values for Weiner_CUT (0.769) and Weiner_PME (0.826) are distinctly lower than the corresponding values for Cornell_CUT and Cornell_PME. The main reason for this behavior is that a number of residues show very small order parameters of 0.3–0.5 (residues 15, 21, 36, 48, and 61 for Weiner_CUT and residues 10, 11, and 54 for Weiner_PME). Considering the fact that a short trajectory of 300 ps introduces a higher statistical error into the calculation of the correlation function and the subsequent fitting procedure, and noting that apart from the few residues with low S^2 the overall features are very similar to the calculations with the Cornell et al. force field, we did not include these data in Figure 8.

We now turn to the analysis of the models employing a small shell of explicit water molecules around the protein in connection with a distance-dependent dielectric constant. As Figure 9 shows, with both force fields the backbone RMSD over the 300 ps simulations stays below 1 Å. With the Weiner_shell model, the value of 0.9 Å is reached within the first 20 ps of simulation and then remains constant. With the Cornell force field the RMSD rises slightly over the first 180 ps to remain constant for the last 120 ps of the simulation. With the extended shell model (Cornell_xshell) the RMSD is even smaller, about 0.7 Å after 300 ps. The computational efficiency of the shell model allowed us to extend the simulation of Weiner_shell model up to 2.142 ns. As Figure 10 shows, the RMSD slowly increases during the first 1.5 ns of the simulation to reach a maximum of 1.2 Å for the backbone atoms and 1.7 Å for all heavy atoms. For the rest of the simulation the RMSD remains approximately constant or even decreases somewhat.

These small values for the RMSD are also reflected in the R_{gyr}. Compared with the box models, the calculated R_{gyr} deviates only slightly from the value of the AMBER-minimized crystal structure of 11.63 Å. However, unlike the box models, R_{gyr} gets smaller (average for Weiner_shell: 11.54 Å; average for Cornell_shell: 11.52 Å), indicating a compaction of the protein. With the extended shell this effect is

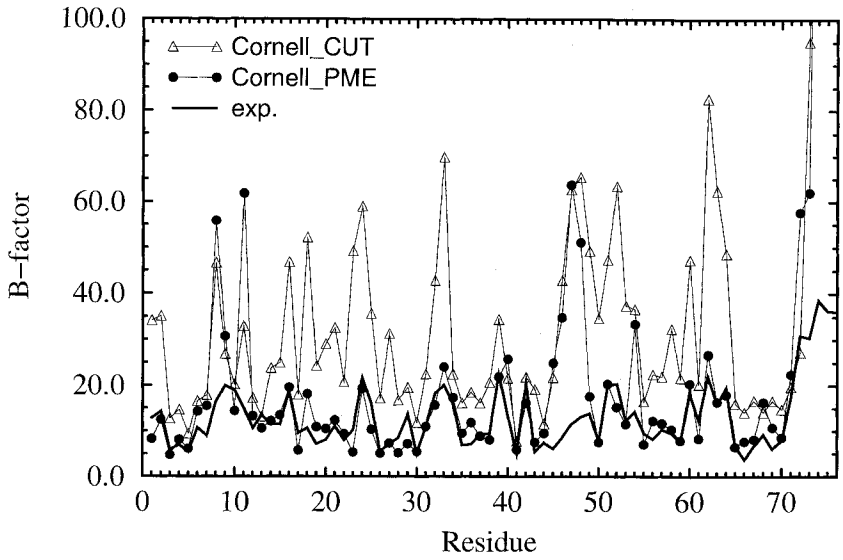

Fig. 7. B factors for all heavy atoms, averaged over individual residues, for the 1.2 ns trajectories of Cornell_CUT and Cornell_PME, together with the experimental data.

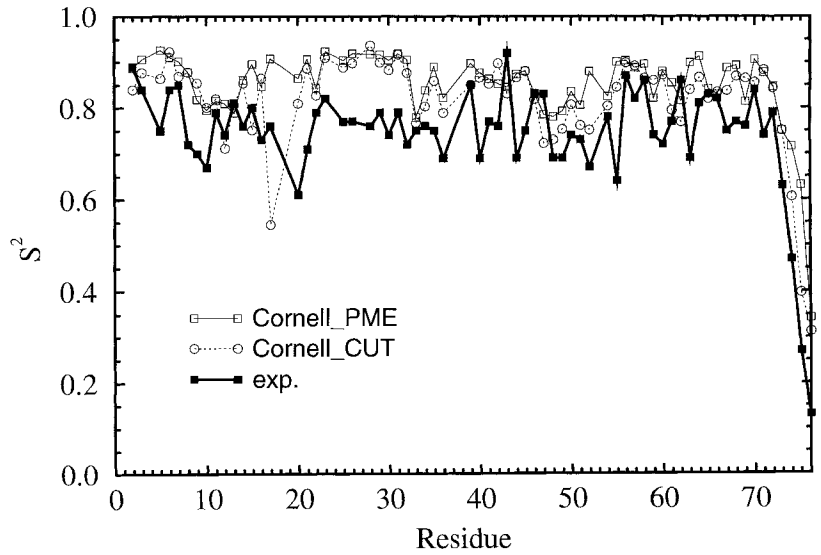

Fig. 8. Order parameters for the backbone N-H vectors. For residues Phe4, Glu18, Pro19, Glu24, Lys27, Pro37, Pro38, and Gly53 no experimental data are available; therefore, we omitted the calculated data for these residues in the graph. Experimental values taken from Schneider et al.[60]

less pronounced (average for Cornell_xshell: 11.58 Å), and R_{gyr} stays at about the crystal value. This is consistent with the notion of the water shell exerting a pressure on the protein—an effect that should be reduced with a larger shell. The exposed surface area of the protein shows a similar trend (Fig. 5); compared with the experimental value of 4,785 Å^3, the values for the shell models either decrease (average for Weiner_shell: 4,748 Å^3; for Cornell_shell: 4,704 Å^3) or hardly change at all (trajectory average for Cornell_xshell: 4,788 Å^3).

As one could expect from the smaller RMSDs from the initial structure, in the shell models with the Cornell et al. force field, the hydrogen bonds found in the initial structure are conserved to a larger extent (Table III). Here, as already found for the periodic box models, however, for residues 58–60 the hydrogen bond pattern typical for a 3_{10} helix as found experimentally cannot be reproduced; the $i \rightarrow i + 3$ hydrogen bonds are formed at most during a third of the simulation.

An interesting artifact of the Weiner et al. force

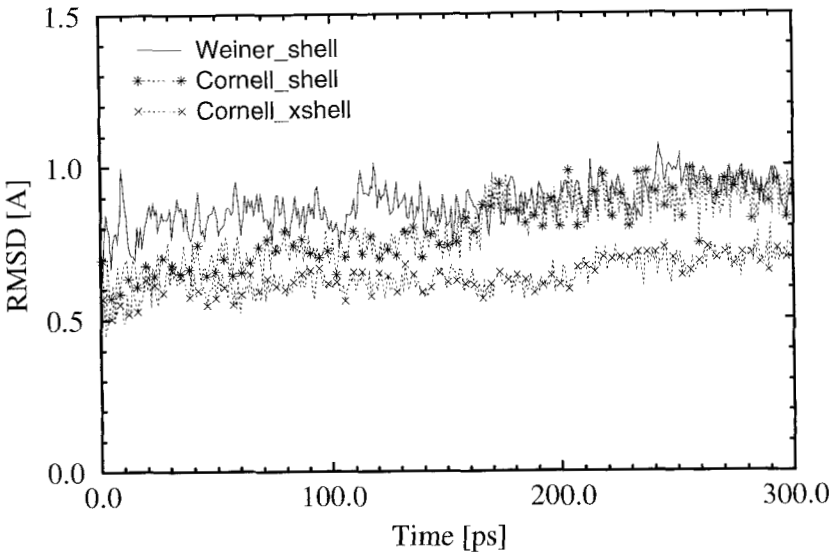

Fig. 9. Backbone atom RMSD (in Å) from the crystal structure for the shell models investigated in this study.

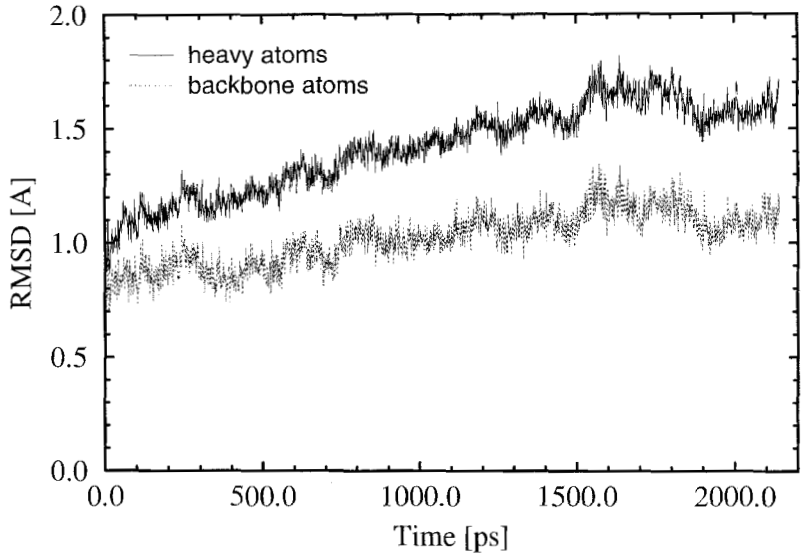

Fig. 10. Time evolution of the RMSD from the crystal structure (in Å) for Weiner_shell from 0 to 2142 ps, for both the backbone atoms and all heavy atoms of the protein.

field is the tendency to form $i \rightarrow i + 2$ γ-turns, as can be observed for the hydrogen bonds Ile36 → Glu34, Lys48 → Ala46, and Ser65 → Gln63 (Table III). This is a consequence of the fact that in this force field the C_7 conformation is overstabilized compared with the extended conformation; in the Cornell et al. force field this has been corrected by the inclusion of specific torsion terms.[39]

A major criticism of the shell model has been the suggestion that the small deviation from the experimental structures is a consequence of damping out important motions due to a rather rigid solvent environment. Indeed, when calculating the atomic fluctuations over the first 300 ps of the simulation, they are significantly reduced compared with the box models, and for many residues they are also much smaller than the experimental B factors. This is especially true for the new force field. However, this picture changes somewhat when using longer intervals of the Weiner_shell simulation to calculate the fluctuations. As shown in Table IV, the average experimental fluctuation for all heavy atoms of residues 1–72 amounts to 0.682 Å. While the calculated value for the first 300 ps of the simulation is only

TABLE IV. Average Atomic Fluctuations and Standard Deviation (in Å) for Weiner_Shell for Different Lengths of Simulation for Residues 1–72 Compared With the Value Obtained From the Experimental Temperature Factors

Simulation length (ps)	Fluctuation
306	0.552 ± 0.222
1,020	0.604 ± 0.273
2,040	0.680 ± 0.336
Experiment	0.682 ± 0.212

0.552 Å, if the fluctuation is calculated over the whole 2.04 ns trajectory, the calculated and experimental fluctuations become comparable.

DISCUSSION

The results presented above clearly demonstrate the superiority of the new Cornell et al. force field over the Weiner et al. force field for MD simulations of proteins in a box of explicit solvent molecules. The Weiner et al. force field shows a larger deviation from the crystal structure than that found using the Cornell et al. force field; this is also evident in the R_{gyr} and especially the solvent-accessible surface of the protein as well as the breaking of backbone hydrogen bonds, which define secondary structure features.

The reason for this is most likely that the Weiner force field was developed for use with *both* vacuum MD simulations (implicit solvent treatment) and explicit solvation. The new Cornell force field, however, has been parametrized exclusively with explicit solvation in mind: the van der Waals parameters were derived from liquid simulations, and the RESP (Restrained Electrostatic Potential)-derived atomic charges provide an improved treatment for the long-range Coulomb interactions. This should result in a better balance among the solute–solute, solute–solvent, and solvent–solvent interactions and subsequently in more stable MD trajectories.

A crucial prerequisite for MD simulations over a long period of time is the accurate evaluation of the long-range electrostatic interaction. For ubiquitin, using an Ewald technique to avoid problems associated with the truncation of the Coulomb interaction at a certain cutoff yields an MD trajectory that stays closer to the experimental structure than the standard 8 Å cutoff simulation for both force fields investigated. This technique, however, is especially successful with the new Cornell et al. force field (RMSD after 1.2 ns: 1.1 Å [Ewald] and 1.4 Å [cutoff]). This should be compared with the results of Alonso and Daggett,[8,49] who use a modified truncation scheme to avoid spurious forces near the cutoff radius and a different force field. They report α-carbon RMSDs of 1.3 Å[49] to 1.8 Å[8] for their 1.4 ns simulations of native ubiquitin.

Compared with RNA or DNA simulations[25] the RMSD differences between trajectories obtained with a pure cutoff model or Ewald summation are rather small. That ubiquitin is less susceptible to errors in long-range electrostatics than DNA or RNA[25] is probably due to the smaller charge density in the solute, but also to its hydrophobic core to which much of its stability is attributed. Another support for the lesser importance of electrostatic interactions in ubiquitin comes from the small difference between the CUT (cutoff of all interactions) and CUTSS (all solute–solute interactions included, and cutoff only for solute–solvent and solvent–solvent) simulations.

However, despite the importance of the accurate treatment of the long-range interactions, the influence of an accurate force field should not be underestimated: Recently Schreiber et al.[6] presented a 1 ns simulation of ubiquitin in explicit water. Although they used Ewald summation to calculate the electrostatic interaction, the backbone RMSD from the crystal structure rose to more than 2 Å during their simulation, significantly higher than both the RMSD for the cutoff or the PME simulations presented in this work with the Cornell et al. force field. This is further corroborated by our investigation of the influence of the force field on the results: starting from the Weiner_CUT model, either using the improved Cornell force field or switching to the PME method considerably improved the stability of the trajectories, and the least deviation from the crystal structure was obtained by the combination of both.

There is important concern about the use of Ewald summation in the context of solution simulations. Since it has been derived for periodic crystallinic systems, it imposes a long-range correlation on the system that is not present in a "true" solution, which might result in unrealistically damped motions in the simulations. While this is in principle true for any periodic boundary simulation, the usual truncation of the long-range interactions prohibits this possible correlation. However, as recent work has shown, using the PME method to evaluate the electrostatic interactions yielded a self-diffusion coefficient of a periodic box of TIP3P or SPC/E [61] water molecules that was actually higher than that for the simple truncation method.[36] Similarly, model simulations of pure water (SPC or TIP4P) or solutions of Cl^- or Fe^{2+} in water[38] showed that the Ewald approach with vacuum boundary conditions (i.e., without the "surface term") results in dissociation constants and Kirkwood *g*-factors in good agreement with spherical cutoff calculations. These findings on model systems indicate that the use of Ewald summation is not necessarily linked to damped motion in the simulation cell.

This hypothesis is supported by our analysis of the atomic fluctuations in ubiquitin. With the Cor-

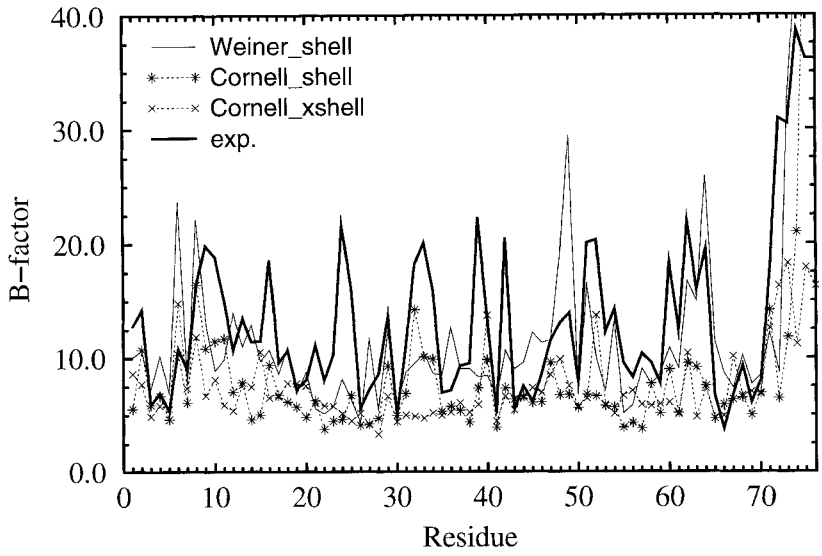

Fig. 11. Comparison of experimental and calculated B factors for all heavy atoms averaged over individual residues over 300 ps trajectories for the shell models.

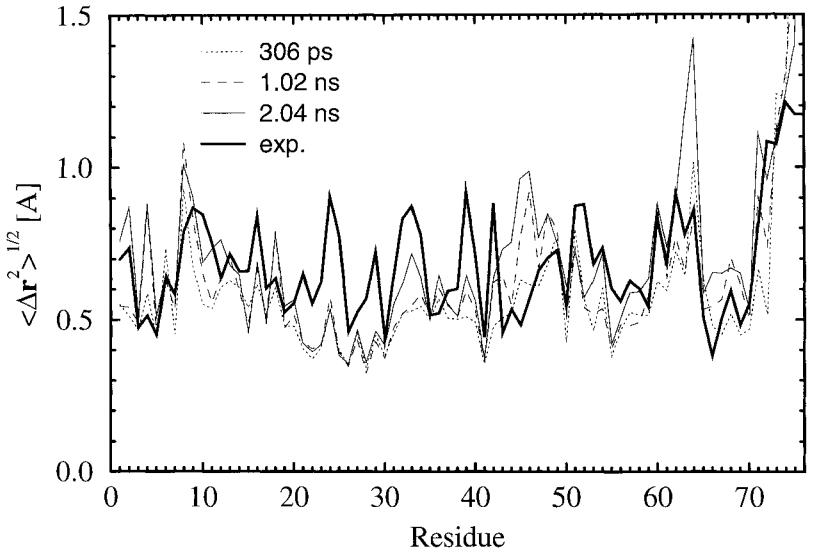

Fig. 12. Dependence of the calculated atomic fluctuations $(\Delta r^2)^{1/2}$ (averaged over amino acid residues) on the length of the trajectory for Weiner-shell. Shown are the fluctuations after 306 ps, 1.02 ns, and 2.04 ns as well as the fluctuations derived from the experimental temperature factors.

nell_PME model, the fluctuations calculated from the 1.2 ns trajectory—with the exception of a few residues at the protein surface—follow the experimental values very closely, indicating that the simulation is able to generate a reasonable amount of internal motion in the protein. We are aware that the experimental values for the fluctuation stem from a crystal study and not from solution data. NMR data on the mobility of individual amino acid residues are rare; however, the 1.2 ns Cornell_PME trajectory finds (in agreement with NMR data) a high mobility for Lys11 and Lys63 as well as less mobility for Lys33. One should keep in mind, however, that even a 1.2 ns trajectory is far too short to pick up all processes occuring on the NMR time scale.

The fluctuations obtained for the cutoff simulations significantly exceed those for the Ewald calculations and those deduced from experiment. One reason for this finding is certainly the higher RMSD for the Cornell_CUT simulation, which automatically leads to higher fluctuation values. Another factor that might contribute to the higher values is unrealistic forces that are a consequence of the sim-

ple truncation method. Therefore, it is not clear if the higher fluctuations obtained with the cutoff methods are really a desirable result. This doubt is underlined by the behavior of residue Lys33, which shows one of the largest fluctuations of all amino acids in the cutoff simulation, but is rather rigid according to solution NMR data.

The solvation models comprised of a shell of water molecules around the protein and employing a distance-dependent dielectric constant showed the least deviation from the experimental crystal structure. This is connected with significantly reduced atomic fluctuations, which confirm concerns that this model inhibits the internal motion of the protein atoms. However, with larger simulation times the calculated fluctuations reach the level of the experimental ones without forfeiting the favorable low RMSDs. Given the computational advantage (the Weiner_shell model is about an order of magnitude faster than Cornell_PME, which should enable routine runs in the nanosecond range), the shell model still might be able to provide valuable information on the structural and the dynamic behavior of proteins. Its use, however, is more appropriate with the Weiner et al. force field, since the new Cornell et al. force field was designed for explicit water molecules and a dielectric constant equal to 1. As a consequence, the shell models with the Cornell force field show even lower atomic fluctuations that will be less well compensated by longer simulation times.

The periodic nature of the Ewald method requires a neutral unit cell. The simulations of ubiquitin as a test system offered the additional advantage that its net charge is zero, so there is no need to modify the setup used for a simple cutoff calculation. In the case of a protein with a net positive or negative charge, we recommend a procedure that adds counterions in a sensible way—possibly guided by the electrostatic potential of the protein—to ensure that the net charge of the unit cell is zero. However, it seems to us that the problem of how many counterions to use and how to best place them, or how these issues affect the actual simulations, is still an open one that warrants further research.

CONCLUSIONS

We have investigated the dependence of the simulated structures of ubiquitin on different force fields, solvation models, and the treatment of the long-range electrostatic interactions. We have shown that with the Cornell et al. force field and a periodic box of water molecules, employing a simple truncation beyond a cutoff radius for the nonbonded interaction, a trajectory up to 1.2 ns could be obtained that is in good agreement with the X-ray and NMR structure of the protein. Whereas the same protocol with the Weiner et al. force field led to a backbone RMSD of 2 Å after 300 ps of simulation, the corresponding value for the simulation with the Cornell force field after 1.2 ns was 1.3 Å. Using Ewald summation instead of simple truncation for the Coulomb interaction further improves these values, giving a backbone RMSD at the end of the 1.2 ns simulation of 1.1 Å. Other observables investigated, such as R_{gyr}, solvent-accessible surface area, and the backbone hydrogen bonds, also show improvement of the MD simulations with the Cornell et al. over the Weiner et al. force field, especially when the PME protocol is used. The atomic fluctuations obtained with the Cornell et al. force field are in good agreement with available experimental data, whereas the order parameters obtained from these simulations were somewhat larger than those deduced from experiment.

The shell models, which consist of a small shell of water molecules and employ a distance-dependent dielectric constant, show the least deviation from the initial protein structure; however, they suffer from a significant decrease in the atomic fluctuations. Due to their high computational efficiency, they allow much longer trajectories and thus may be useful to extend the time range of MD simulations for more extensive conformational sampling.

ACKNOWLEDGMENTS

T.F. is thankful for a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft. P.A.K. was supported by the NIH through grant GM-29072. We thank the Pittsburgh Supercomputer Center (PSC) for allocation of computer time and the Computer Graphics Laboratory at UCSF (RR-1081, director T. Ferrin). We are grateful to Thomas Cheatham for inspiring discussions and valuable help with the Cray T3D version of SANDER. We acknowledge Michael F. Crowley of PSC for parallelizing the PME code for the Cray T3D and supplying us with preliminary versions of this code. We thank Jeanmarie Guenot for her help in the initial stages of this study.

REFERENCES

1. McCammon, J.A., Harvey, S.C. "Dynamics of Proteins and Nucleic Acids." Cambridge: Cambridge University Press, 1987.
2. van Gunsteren, W.F., Weiner, P.K. (eds.) "Computer Simulation of Biomolecular Systems." Leiden: Escom, 1989.
3. 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.
4. Braatz, J.A., Paulsen, M.D., Ornstein, R.L. 3 nsec molecular dynamics simulation of the protein ubiquitin and comparison with X-ray crystal and solution NMR structures. J. Biomol. Struct. Dyn. 9:935–949, 1992.
5. Guenot, J., Kollman, P.A. unpublished results.
6. Abseher, R., Luedemann, S., Schreiber, H., Steinhauser, O. NMR cross relaxation investigated by molecular dynamics simulation: A case study of ubiquitin in solution. J. Mol. Biol. 249:604–624, 1995.
7. York, D.M., Wlodawer, A., Pedersen, L.G., Darden, T.A. Atomic level accuracy in simulations of large protein crystals. Proc. Natl. Acad. Sci. USA 91:8715–8718, 1994.
8. Alonso, D., Daggett, V. Partial refolding of ubiquitin; simulations of the hydrophobic collapse. (submitted)

9. Weerasinghe, S., Smith, P.E., Nohan, V., Cheng, Y.-K., Pettitt, B.M. Nanosecond dynamics and structure of a model DNA triple helix in saltwater solution. J. Am. Chem. Soc. 117:2147–2158, 1995.
10. Guenot, J., Kollman, P.A. Molecular dynamics studies of a DNA binding protein: 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the *Escherichia coli* trp repressor. Protein Sci. 1:1185–1205, 1992.
11. Schiffer, C.A., Caldwell, J.W., Kollman, P.A., Stroud, R.M. Protein structure prediction with a combined solvation free energy-molecular mechanics force field. Mol. Simul. 10:121–149, 1993.
12. Wesson, L., Eisenberg, D. Atomic solvation parameters applied to MD of proteins in solution. Protein Sci. 1:227–235, 1992.
13. Kollman, P. Free energy calculations—applications to chemical and biochemical phenomena. Chem. Rev. 93: 2395–2417, 1993.
14. Steinbach, P.J., Brooks, B.R. New spherical-cutoff methods for long-range forces in macromolecular simulation. J. Comput. Chem. 15:667–683, 1994.
15. Saito, M. Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation. J. Chem. Phys. 101:4055–4061, 1994.
16. Brooks III, C.L. Methodological advances in molecular dynamics simulations of biological systems. Curr. Opin. Struct. Biol. 5:211–215, 1995.
17. Allen, M.P., Tildesley, D.J. "Computer Simulation of Liquids." Oxford: Oxford University Press, 1987.
18. Dao-pin, S., Soderlind, E., Baase, W.A., Wozniak, J.A., Sauer, U., Matthews, B.W. Cumulative site-directed charge-change replacements in bacteriophage-T4 lysozyme suggest that long-range electrostatic interactions contribute little to protein stability. J. Mol. Biol. 221:873–887, 1991.
19. Schreiber, H., Steinhauser, O. Cutoff size does strongly influence molecular dynamics results on solvated polypeptides. Biochemistry 31:5856–5860, 1992.
20. Bader, J.S., Chandler, D. Computer simulation study of the mean forces between ferrous and ferric ions in water. J. Phys. Chem. 96:6423–6427, 1992.
21. York, D.M., Darden, T.A., Pedersen, L.G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. J. Chem. Phys. 99:8345–8348, 1993.
22. Smith, P.E., Pettitt, B.M. Peptides in ionic solutions: A comparison of the Ewald and switching function techniques. J. Chem. Phys. 95:8430–8441, 1991.
23. Schreiber, H., Steinhauser, O. Taming cut-off induced artifacts in molecular dynamics studies of solvated polypeptides. The reaction field method. J. Mol. Biol. 228:909–923, 1992.
24. Forester, T., Smith, W. On multiple time-step algorithms and the Ewald sum. Mol. Simul. 13:195–204, 1994.
25. Cheatham III, T.E., Miller, J.L., Fox, T., Darden, T.A., Kollman, P.A. Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins. J. Am. Chem. Soc. 117:4193–4194, 1995.
26. Zichi, D.A. Molecular dynamics of RNA with the OPLS force field. Aqueous simulation of a hairpin containing a tetranucleotide loop. J. Am. Chem. Soc. 117:2957–2969, 1995.
27. York, D.M., Yang, W., Lee, H., Darden, T.A., Pedersen, L.G. Toward the accurate modeling of DNA: The importance of long-range electrostatics. J. Am. Chem. Soc. 117: 5001–5002, 1995.
28. Hockney, R.W., Eastwood, J.W. "Computer Simulation Using Particles." New York: McGraw-Hill, 1981.
29. Shimada, J., Kaneko, H., Takada, T. Performance of fast multipole methods for calculating electrostatic interactions in biomacromolecular simulations. J. Comput. Chem. 15:28–43, 1994.
30. Greengard, L. "The Rapid Evaluation of Potential Fields in Particle Systems." Cambridge: MIT, 1987.
31. White, C.A., Head-Gordon, M. Derivation and efficient implementation of the fast multipole method. J. Chem. Phys. 101:6593–6605, 1994.
32. Ewald, P.P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann. Phys. (Leipzig) 64:253–287, 1921.
33. Lantelme, F., Turq, P., Quentrec, B., Lewis, J.W.E. Application of the molecular dynamics method to a liquid system with long range forces (Molten NaCl). Mol. Phys. 28: 1537–1549, 1974.
34. Impey, R.W., Madden, P.A., McDonald, I.R. Hydration and mobility of ions in solution. J. Phys. Chem. 87:5071–5083, 1983.
35. Darden, T.A., York, D.M., Pedersen, L.G. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. J. Chem. Phys. 98:10089–10092, 1993.
36. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T.A., Lee, H., Pedersen, L.G. A smooth particle mesh Ewald method. J. Chem. Phys. 103:8577–8593, 1995.
37. Kitson, D.H., Avbelj, F., Moult, J., Nguyen, D.T., Mertz, J.E., Hadzi, D., Hagler, A.T. On achieving better than 1-A accuracy in a simulation of a large protein—*Streptomyces griseus* protease-A. Proc. Natl. Acad. Sci. USA 90:8920–8924, 1993.
38. Roberts, J.E., Schnitker, J. Boundary conditions in simulations of aqueous ionic solutions: A systematic study. J. Phys. Chem. 99:1322–1331, 1995.
39. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz Jr., K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A. A second generation force field for the simulation of proteins and nucleic acids. J. Am. Chem. Soc. 117:5179–5197, 1995.
40. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
41. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A. An all atom force field for simulations of proteins and nucleic acids. J. Comput. Chem. 7:230–252, 1986.
42. Merz Jr., K.M., Kollman, P.A. Free energy perturbation simulations of the inhibition of thermolysin: Prediction of the free energy of binding of a new inhibitor. J. Am. Chem. Soc. 111:5649–5658, 1989.
43. Monia, B.P., Ecker, D.J., Crooke, S.T. New perspectives on the structure and function of ubiquitin. Biotechnology 8:209–215, 1990.
44. Hershko, A., Chiechanover, A. Mechanisms of intracellular protein breakdown. Annu. Rev. Biochem. 51:335–364, 1982.
45. Vijay-Kumar, S., Bugg, C.E., Cook, W.J. Structure of ubiquitin refined at 1.8 Å resolution. J. Mol. Biol. 194:531–544, 1987.
46. Wintrode, P.L., Makhatadze, G.I., Privalov, P.O. Thermodynamics of ubiquitin unfolding. Proteins 18:246–253, 1994.
47. Di Stefano, D.L., Wand, A.J. Two-dimensional ^{1}H NMR study of human ubiquitin: A main chain directed assignment and structure analysis. Biochemistry 26:7272–7281, 1987.
48. Weber, P.L., Brown, S.C., Mueller, L. Sequential ^{1}H NMR assignments and secondary structure identification of human ubiquitin. Biochemistry 26:7282–7290, 1987.
49. Alonso, D.O.V., Daggett, V. Molecular dynamics simulations of protein unfolding and limited refolding: Characterization of partially unfolded states of ubiquitin in 60% methanol and in water. J. Mol. Biol. 247:501–520, 1995.
50. Alonso, D., Daggett, V. Molecular dynamics studies of partially unfolded conformations of ubiquitin in methanol and their refolding in water. (submitted)
51. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham III, T.E., Ferguson, D.M., Seibel, G.L., Singh, U.C., Weiner, P.K., Kollman, P.A. "AMBER 4.1." San Francisco: University of California, 1995.
52. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Tasumi, M. The protein data bank: A computer based archival file for macromolecular structures. J. Mol. Biol. 12:535–542, 1977.
53. Jorgensen, W.L., Chandreskhar, J., Madura, J.D., Imprey, R.W., Klein, M.L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79:926–935, 1983.
54. Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F.,

Di Nola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.
55. Ross, W.S. "CARNAL." San Francisco: University of California, 1995.
56. Simmerling, C., Elber, R., Zhang, J. MOIL-View—a program for visualization of structure and dynamics of biomolecules and STO—a program for computing stochastic paths. In: "Modelling of Biomolecular Structures and Mechanisms." Pullman, A. (ed.) Kluver Acad. Press, The Netherlands, 1995, 241–265.
57. Press, W.H., Flannery, B.P., Teukolksy, S.A., Vetterling, W.T. "Numerical Recipes." Cambridge: Cambridge University Press, 1989.
58. Billeter, M. Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals. Q. Rev. Biophys. 25:325–377, 1992.
59. Berndt, K.D., Guentert, P., Orbons, L.P.M., Wuthrich, K. Determination of high-quality nuclear magnetic resonance solution structure of the BPTI and comparison with three crystal structures. J. Mol. Biol. 227:757–775, 1992.
60. Schneider, D.M., Dellwo, M.J., Wand, A.J. Fast internal main-chain dynamics of human ubiquitin. Biochemistry 31:3645–3652, 1992.
61. Berendsen, H.J.C., Grigesa, J.R., Slsaatsma, J.P. The missing term in effective pair potentials. J. Phys. Chem. 91:6269–6271, 1987.