

20. ENERGY CALCULATIONS AND MOLECULAR DYNAMICS

20.1. Molecular-dynamics simulation of protein crystals: convergence of molecular properties of ubiquitin

BY U. STOCKER AND W. F. VAN GUNSTEREN

20.1.1. Introduction

Molecules in crystals are often believed to have a very rigid structure due to their ordered packing, and the investigation of the molecular motion of such systems is often considered to be of little interest. In contrast to small-molecule crystals, however, the solvent concentration in protein crystals is high, usually with about half of the crystal consisting of water. Thus, in this respect, one can compare protein crystals with very concentrated solutions and expect non-negligible atomic motion. The atomic mobility in proteins can be investigated by experiment (X-ray diffraction, NMR) or by molecular simulation.

Today's experimental techniques are very advanced. They are, however, only able to examine time- and ensemble-averaged structures and properties. In contrast, with simulations one can go beyond averaged properties and examine the motions of a single molecule in the pico- and nanosecond time regime. Such simulations have become possible with the availability of high-resolution structural data, which provide adequate starting structures for biologically relevant systems. Depending on the kind of property in which one is interested, different methods of simulation may be used. Equilibrium properties can be obtained using either Monte Carlo (MC) or molecular-dynamics (MD) simulation techniques, but motions can only be observed with the latter. Current interest in the simulation community mainly focuses on dissolved proteins as they would be in their natural environment. Force fields are parameterized to mimic the behaviour and function of proteins in a solution, and few crystal simulations have been performed. Consequently, a crystal environment provides an excellent opportunity to test a force field on a task for which it should be appropriate, but for which it has not been directly parameterized.

Apart from the analysis of the dynamic properties of a system, MD simulations are also used in structure refinement. In refinement, be it X-ray crystallographic or NMR, a special term is added to the standard physical force field to reflect the presence of experimental data:

$$V(\mathbf{r}) = V^{\text{phys}}(\mathbf{r}) + V^{\text{special}}(\mathbf{r}). \quad (20.1.1.1)$$

In NMR, a variety of properties can be measured, and each of these can be used in the definition of an additional term that restrains the generated structures to reproduce given experimental values. Refinement procedures exist that use nuclear-Overhauser-effect (van Gunsteren *et al.*, 1984; Kaptein *et al.*, 1985), *J*-value (Torda *et al.*, 1993) and chemical-shift (Harvey & van Gunsteren, 1993) restraints. In crystallography, X-ray intensities are used to generate the restraining energy contribution (Brünger *et al.*, 1987; Fujinaga *et al.*, 1989). Combined NMR/X-ray refinement uses both solution and crystal data (Schiffer *et al.*, 1994).

As in an experiment, averages over time and molecules are measured, and instantaneous restraints can lead to artificial rigidity in the molecular system (Torda *et al.*, 1990). This can be circumvented by restraining time or ensemble averages, instead of instantaneous values, to the value of the measured quantity. Time averaging has been applied to nuclear Overhauser effects (Torda *et al.*, 1990) and *J* values (Torda *et al.*, 1993) in NMR structure

determination and to X-ray intensities in crystallography (Gros *et al.*, 1990; Gros & van Gunsteren, 1993; Schiffer *et al.*, 1995). Ensemble averaging has been applied in NMR refinement (Scheek *et al.*, 1991; Fennen *et al.*, 1995). For a more detailed discussion of restrained MD simulations, we refer to the literature (van Gunsteren *et al.*, 1994, 1997).

The first unrestrained MD simulations of a protein in a crystal were carried out in the early 1980s (van Gunsteren & Karplus, 1982). The protein concerned was bovine pancreatic trypsin inhibitor (BPTI), a small (58-residue) protein for which high-resolution X-ray diffraction data were available. The initial level of simulation was to neglect solvent, using vacuum boundary conditions. This was improved gradually by the inclusion of Lennard-Jones particles at the density of water as a solvent (van Gunsteren & Karplus, 1982) to let the protein feel random forces and friction from the outside as well as feel a slightly attractive external field. The next step was to use a simple (simple point charge, SPC) water model (van Gunsteren *et al.*, 1983). Further improvement was achieved by incorporating counter ions into the modelled systems to obtain overall charge neutrality (Berendsen *et al.*, 1986).

Despite these early attempts, few unrestrained crystal simulations have been reported in the literature, and, to our knowledge, these involve one to four protein molecules, simulating one unit cell (Shi *et al.*, 1988; Heiner *et al.*, 1992). The maximum time range covered has been less than 100 ps.

In the work described in this chapter, the current state of MD simulation of protein crystals is illustrated. A full unit cell of ubiquitin, containing four ubiquitin and 692 water molecules, has been simulated for a period of two nanoseconds. Since this simulation is an order of magnitude longer than crystal simulations in the literature, it offers the possibility of analysing the convergence of different properties as a function of time and as a function of the number of protein molecules. Converged properties can also be compared with experimental values as a test of the GROMOS96 force field (van Gunsteren *et al.*, 1996). Finally, the motions obtained can be analysed to obtain a picture of the molecular behaviour of ubiquitin in a crystalline environment.

20.1.2. Methods

Ubiquitin consists of 76 amino acids with 602 non-hydrogen atoms. Hydrogen atoms attached to aliphatic carbon atoms are incorporated into these (the united-atom approach), and the remaining 159 hydrogen atoms are treated explicitly. Ubiquitin crystallizes in the orthorhombic space group $P2_12_12_1$, with $a = 5.084$, $b = 4.277$ and $c = 2.895$ nm. There is one molecule in the asymmetric unit. The protein was crystallized at pH 5.6. The amino acids Glu and Asp were taken to be deprotonated, and Lys, Arg and His residues were protonated, leading to a charge of +1 electron charge per chain. Because this is a small value compared with the size of the system, no counter ions were added. Four chains of ubiquitin, making up a full unit cell of the crystal, were simulated together with 692 water molecules modelled using the SPC water model (Berendsen *et al.*, 1981). 232 water molecules were placed at crystallographically observed water sites, and the remaining 460 were added to obtain

20. ENERGY CALCULATIONS AND MOLECULAR DYNAMICS

the experimental density of 1.35 g cm^{-3} , leading to a system size of 3044 protein atoms and 5120 atoms total.

The crystal structure of ubiquitin [Protein Data Bank (Bernstein *et al.*, 1977) code 1UBQ] solved at 1.8 Å resolution (Vijay-Kumar *et al.*, 1987) was used as a starting point. To achieve the appropriate total density, noncrystallographic water molecules were added, using a minimum distance of 0.220605 nm between non-hydrogen protein atoms or crystallographic water oxygen atoms and the oxygen atoms of the added water molecules, which were taken from an equilibrated water configuration (van Gunsteren *et al.*, 1996). Initial velocities were assigned from a Maxwell-Boltzmann distribution at 300 K. The protein and solvent were coupled separately to temperature baths of 300 K with a coupling time of 0.1 ps (Berendsen *et al.*, 1984). No pressure coupling was applied. Another simulation (results not shown) including pressure coupling showed no significant change in the box volume. Bonds were kept rigid using the *SHAKE* method (Ryckaert *et al.*, 1977), with a relative geometric tolerance of 10^{-4} . Long-range forces were treated using twin range cutoff radii of 0.8 and 1.4 nm (van Gunsteren & Berendsen, 1990). The pair list for non-bonded interactions was updated every 10 fs. No reaction field correction was applied. All simulations were performed using the *GROMOS96* package and force field (van Gunsteren *et al.*, 1996).

The system was initially minimized for 20 cycles using the steepest-descent method. The protein atoms were harmonically restrained (van Gunsteren *et al.*, 1996) to their initial positions with a force constant of $25000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. This minimized structure was then pre-equilibrated in several short MD runs of 500 steps of 0.002 ps each, gradually lowering the restraining force constant from $25000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to zero. The time origin was then set to zero, and the entire unit cell was simulated for 2 ns. The time step was 0.002 ps, and every 500th configuration was stored for evaluation. The first 400 ps of the run were treated as equilibration time, the remaining 1.6 ns were used for analysis.

20.1.3. Results

20.1.3.1. Energetic properties

In Fig. 20.1.3.1, the non-bonded contributions to the total potential energy are shown. The non-bonded interactions comprise Lennard-Jones and electrostatic interactions. Solvent-solvent, solute-solute and solute-solvent interaction energies are shown separately. All of these appear converged after approximately 100 ps. The solvent-solvent energy remains close to its initial value during the whole simulation, the water molecules having relaxed during the pre-equilibration period, while the protein was restrained. The protein internal energy increases during the first few hundred picoseconds, but this is compensated by a decrease in the protein-solvent energy as the protein adapts to the force field and the pre-relaxed solvent environment. This effect becomes negligible after about 200 ps, from which time point the system can be viewed as equilibrated with respect to the energies. The distribution of kinetic *versus* potential energy and the total (bonded and non-bonded) energy of the system relaxes even faster (results not shown).

20.1.3.2. Structural properties

Not all properties converge as fast as the energies. Fig. 20.1.3.2 shows the root-mean-square atom-position deviation (RMSD) from the X-ray structure for each of the four individual chains for both $\text{C}\alpha$ atoms and all atoms. Here, several relaxation periods can be distinguished. After the initial increase, which occurs during the first 50 ps of the simulation, a plateau is reached, and the system is apparently stable until 300 ps. The values reached are 0.12 nm for

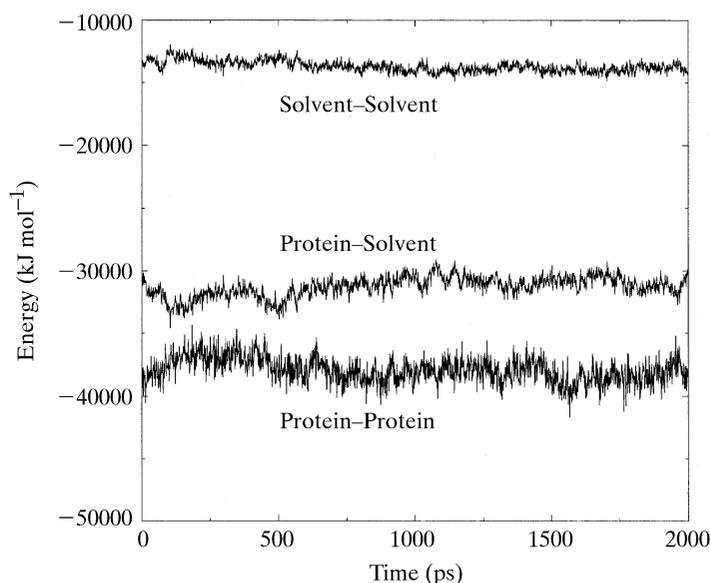


Fig. 20.1.3.1. Non-bonded energies (in kJ mol^{-1}) of the simulated system as a function of time.

the $\text{C}\alpha$ atoms and 0.20 nm if all atoms are considered. These numbers are comparable with results obtained in crystal simulations of other proteins of equivalent length reported in the literature (van Gunsteren *et al.*, 1983; Berendsen *et al.*, 1986; Shi *et al.*, 1988; Heiner *et al.*, 1992; Levitt *et al.*, 1995). After 300 ps, however, the values increase slowly again. For the $\text{C}\alpha$ atoms, there is apparently a second plateau from 300 to 850 ps, but during this period the RMSD for all atoms continues to increase monotonically. After 850 ps, a final plateau is reached. During the second nanosecond of the simulation (1000–2000 ps), the RMSDs are 0.21 nm for the $\text{C}\alpha$ atoms and 0.29 nm for all atoms. The RMSD of chain 1 is an exception. There is a strong increase after 1200 ps due to a movement of a particular part of the chain which will be addressed later. To ensure that the RMSD values have converged, longer runs would be required.

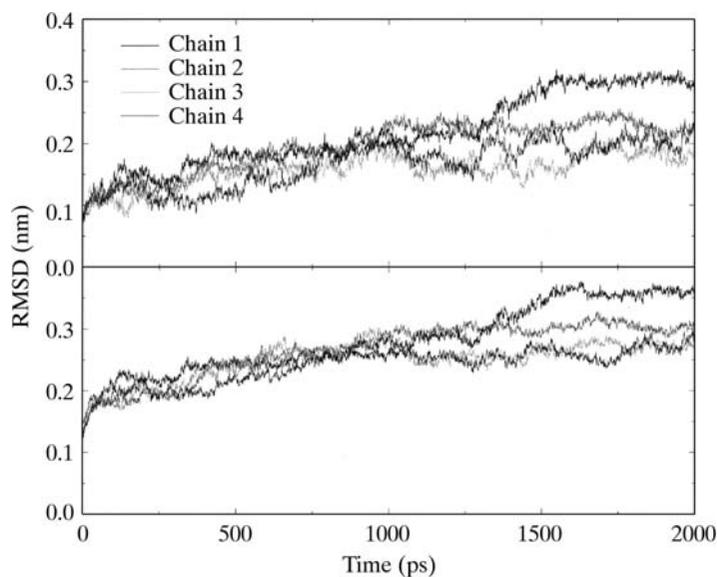


Fig. 20.1.3.2. Root-mean-square atom-positional deviations (RMSD) in nm from the X-ray structure of the four different protein molecules in the unit cell as a function of time. Rotational and translational fitting was applied using the $\text{C}\alpha$ atoms of residues 1–72. The upper and lower graphs show the deviations for the $\text{C}\alpha$ atoms and for all atoms, respectively.

20.1. MOLECULAR DYNAMICS: CONVERGENCE OF UBIQUITIN

Although the RMSD values shown in Fig. 20.1.3.2 grow larger than those usually observed in the course of short simulations, the hydrogen-bonding pattern and thus the secondary-structure elements observed in the X-ray structure are reproduced well (Table 20.1.3.1). Most of the hydrogen bonds reported (Vijay-Kumar *et al.*, 1987) show high occupancies during the whole simulation, especially the ones within secondary-structure elements. Only six out of the 44 hydrogen bonds present in the X-ray structure are disrupted during the simulation. Hydrogen bonds in the α -helix (residues 23–34) show very high occupancies, ranging from 75% at the N-terminus to well over 90% inside the helix. Only its C-terminus shows signs of instability, with the α -helix deforming towards a 3_{10} -helix. The β -sheet pattern is, apart from chain 1 in the region of residues 49–64, as stable as the α -helix. Occupancies range from 55% up to 95%. The six hydrogen bonds not reproduced (five 3_{10} and one α -helical) can be rationalized as follows. The bridges 10–7 and 65–62 are part of the most mobile regions of the protein. These regions involve residues 7–10, 51–54 and 62–65 (Vijay-Kumar *et al.*, 1987). The hydrogen bond at the C-terminal end of the α -helix (35–31) is lost, and the preceding hydrogen bond (34–30) is partly changed, indicating that the end of the α -helix is deformed towards a 3_{10} -helix. The donor of the bond 40–37 is replaced by residue 41, and the four-residue 3_{10} -helix that was stabilized by hydrogen bonds 58–55 and 59–56 is replaced by an α -helical hydrogen bond 59–55. The high occupancy of this particular bond and the complete absence of the two observed experimentally indicate an early rearrangement in this part of the structure (before the analysis period) which is stable during the rest of the simulation.

Backbone–side-chain hydrogen bonds are less well reproduced than backbone–backbone interactions. While some are present 80–90% of the time, others are present less than 50% of the time. Two out of the seven hydrogen bonds in which a backbone atom is the donor are not observed in the simulation; both involve the OG1 atom of Thr7 as an acceptor. The hydrogen-donor atoms are the backbone nitrogen atoms of residues Thr9 and Lys11, both of which have high experimental B factors (18.32 \AA^2 for Thr9 and 13.56 \AA^2 for Lys11). The mean of the experimental B factors is 10.73 \AA^2 for the backbone atoms and 13.41 \AA^2 for all protein atoms. Where a side-chain atom is the donor, three out of the five hydrogen bonds present in the X-ray structure are not found in the simulation. All of these involve the side-chain nitrogen atom of a lysine residue as the donor, the experimental B factors of which range from 23.92 \AA^2 for the NZ atom of Lys48 up to 30.06 \AA^2 for the NZ atom of Lys33. Of the four side-chain–side-chain hydrogen bonds, not one is observed as in the crystal. The 54–58 hydrogen bond is, however, replaced by a 55–58 hydrogen bond. All the others involve very mobile atoms with large experimental B factors as donors and acceptors.

There is one intermolecular hydrogen bond (Table 20.1.3.2) in the starting structure which is not seen in the simulation. The donor is the side-chain nitrogen atom of Lys6, which has an experimental B factor of 20.55 \AA^2 , and the acceptors are the side-chain oxygen atoms of Glu51, with B factors of 32.13 and 33.44 \AA^2 . Most of the hydrogen bonds not reproduced in the simulation contain at least one mobile atom. Although these atoms do not remain fixed at their equilibrium positions, they may still stabilize the structure on average.

In Fig. 20.1.3.3, the deviation of the $C\alpha$ atoms of the different chains from the X-ray structure and from the mean MD structure are presented together with the deviation of the mean MD structure from the X-ray structure. Overall, the individual protein molecules remain close to the experimental structure; however, parts of the structure do deviate substantially. The region involving residues 7–11, which experimentally has high B factors (implying high mobility), has, in three out of the four cases, an RMSD for $C\alpha$ atoms of 0.3 nm or greater. Chain 3, in contrast, is close to the X-ray

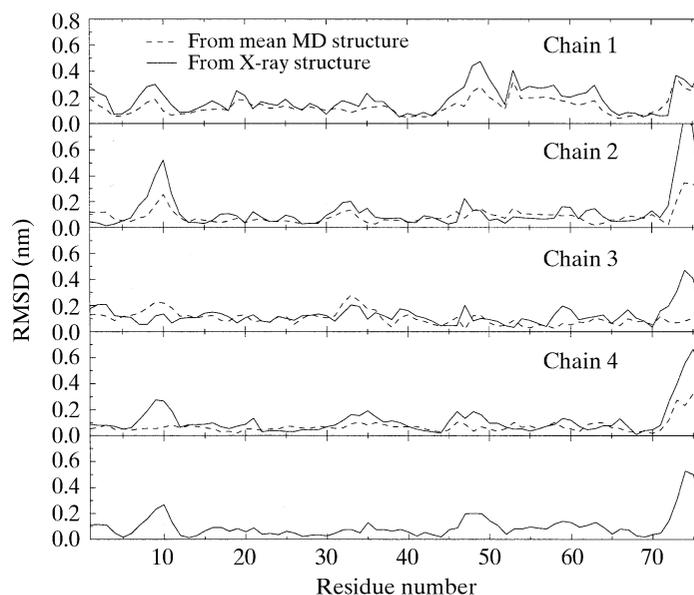


Fig. 20.1.3.3. Root-mean-square $C\alpha$ -atom-position deviation (RMSD) in nm from a reference structure as a function of the residue number using the final 1.6 ns of the simulation. RMSDs of the four protein molecules in the unit cell from the mean molecular-dynamics structure (dashed line) and from the X-ray structure (solid line) are shown in the first four graphs. The bottom graph shows the RMSD of the mean (over the four molecules) MD structure from the X-ray structure.

structure, and the mean MD structure is closer to the X-ray structure than are any of the individual chains. This suggests that the simulation does not deviate systematically from the X-ray structure, but rather that different regions of conformational space are sampled by the different molecules. The same holds for the very mobile region between residues 47 and 64, where chain 1 deviates dramatically from both the mean MD structure and the X-ray structure. In the other chains, this part of the protein remains close to the X-ray structure. Overall, the deviation from the X-ray structure is largest in the C-terminal region. This part is also ill-defined in the experiment, with occupancies of 0.45 for residues 73 and 74, and 0.25 for the terminal two glycines. Other parts of the protein, especially the stable secondary-structure elements, stay close to the X-ray structure. In the average structure, the α -helix, including its C-terminal part which was deformed to a 3_{10} -helix, deviates by a maximum of 0.08 nm from the X-ray structure, although, as seen before, the individual chains may deviate more. The β -sheet regions also stay close to the X-ray structure. As with the helix, residues 1–7, 40–45 and 64–72 stay within 0.1 nm RMSD from the X-ray structure. The β -strands formed by residues 10–17 and 48–50 are not as similar to the experiment, since they lie close to mobile regions and are thus influenced by neighbouring mobile residues. For the strand formed by residues 10–17, from residue 12 onwards the same structural similarity is reached as for all other secondary-structure elements, and residues 48–50 are, again, strongly influenced by the moving part of chain 1.

20.1.3.3. Effect of the translational and rotational fitting procedure

In Fig. 20.1.3.4, the impact of different fitting protocols on atomic mean-square position fluctuations (RMSFs) is examined. B factors are related to mean-square position fluctuations according to

$$B_i = (8\pi^2/3) \langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle, \quad (20.1.3.1)$$

20. ENERGY CALCULATIONS AND MOLECULAR DYNAMICS

Table 20.1.3.1. *Occurrence of intramolecular hydrogen bonds (%) during the final 1.6 ns of the simulation*

The criteria for a hydrogen bond to be present are angle donor–hydrogen–acceptor $\geq 135^\circ$, distance hydrogen–acceptor ≤ 0.25 nm. Hydrogen bonds are shown if they are either present in the X-ray structure or if at least one of the four protein molecules in the unit cell shows the hydrogen bond of interest for at least 50% of the simulation time. The letter h appended to an amino-acid code indicates that the residue is protonated.

Hydrogen bonds		X-ray structure	Molecular dynamics			
Backbone	Backbone		Molecule 1	Molecule 2	Molecule 3	Molecule 4
3Ile H	15Leu O	100	94	94	95	98
4Phe H	65Ser O	100	85	69	87	77
5Val H	13Ile O	100	80	90	87	93
6Lysh H	67Leu O	100	85	82	88	94
7Thr H	11Lysh O	100	65	49	54	62
8Leu H	69Leu O	0	5	52	19	55
10Gly H	7Thr O	100	0	0	0	0
13Ile H	5Val O	100	86	76	70	87
15Leu H	3Ile O	100	87	92	72	82
17Val H	1Met O	100	68	39	79	51
21Asp H	18Glu O	100	68	84	84	90
23Ile H	54Arg O	100	0	74	89	92
24Glu H	52Asp O	100	58	69	63	84
26Val H	22Thr O	100	92	69	78	61
27Lysh H	23Ile O	100	94	97	98	99
28Ala H	24Glu O	100	71	71	84	89
29Lysh H	25Asn O	100	91	79	94	88
30Ile H	26Val O	100	92	76	94	91
31Gln H	27Lysh O	100	85	53	66	93
32Asp H	28Ala O	100	82	27	87	77
33Lysh H	29Lysh O	100	23	13	81	51
33Lysh H	30Ile O	0	59	23	7	19
34Glu H	30Ile O	100	95	54	64	86
35Gly H	31Gln O	100	0	0	0	0
36Ile H	34Glu O	0	62	50	28	35
40Gln H	37Pro O	100	0	0	0	0
41Gln H	37Pro O	0	68	56	72	20
41Gln H	38Pro O	100	14	25	14	50
42Arg H	70Val O	100	82	82	83	88
44Ile H	68Hish O	100	84	96	93	95
45Phe H	48Lysh O	100	20	74	77	91
48Lysh H	45Phe O	100	24	62	59	44
50Leu H	43Leu O	100	29	88	92	85
54Arg H	51Glu O	100	20	60	19	69
56Leu H	21Asp O	100	0	90	81	81
57Ser H	19Pro O	100	3	78	86	83
58Asp H	55Thr O	100	0	0	0	0
59Tyr H	55Thr O	100	58	86	92	85
59Tyr H	56Leu O	100	0	0	0	0
60Asn H	57Ser O	100	38	34	60	58
61Ile H	56Leu O	100	67	7	63	56
64Glu H	2Gln O	100	0	42	6	95
65Ser H	62Gln O	100	0	0	0	0
67Leu H	4Phe O	100	69	74	87	70
68Hish H	44Ile O	100	62	68	83	89
69Leu H	6Lysh O	100	79	72	92	90
70Val H	42Arg O	100	91	89	90	91
72Arg H	40Gln O	100	79	59	85	78

20.1. MOLECULAR DYNAMICS: CONVERGENCE OF UBIQUITIN

Table 20.1.3.1. Occurrence of intramolecular hydrogen bonds (%) during the final 1.6 ns of the simulation (cont.)

Hydrogen bond		X-ray structure	Molecular dynamics			
Backbone	Side chain		Molecule 1	Molecule 2	Molecule 3	Molecule 4
2Gln H	64Glu OE2	0	63	7	84	0
9Thr H	7Thr OG1	100	0	0	0	0
11Lys H	7Thr OG1	100	0	0	0	0
18Glu H	21Asp OD2	100	80	3	0	0
20Ser H	18Glu OE2	0	0	0	55	0
25Asn H	22Thr OG1	100	31	13	61	38
51Glu H	59Tyr OH	100	46	87	56	76
55Thr H	58Asp OD1	100	29	62	22	75
58Asp H	55Thr OG1	100	53	76	72	86
64Glu H	64Glu OE2	0	55	6	16	0

Hydrogen bond		X-ray structure	Molecular dynamics			
Side chain	Backbone		Molecule 1	Molecule 2	Molecule 3	Molecule 4
29Lys HZ2	16Glu O	100	0	0	0	0
33Lys HZ2	14Thr O	100	0	0	0	0
41Gln HE21	27Lys O	100	81	91	47	71
41Gln HE22	36Ile O	100	90	89	60	83
48Lys HZ3	46Ala O	100	0	0	0	0

Hydrogen bond		X-ray structure	Molecular dynamics			
Side chain	Side chain		Molecule 1	Molecule 2	Molecule 3	Molecule 4
11Lys HZ2	34Glu OE2	100	0	0	0	0
20Ser HG	18Glu OE2	0	0	0	60	0
27Lys HZ2	52Asp OD2	100	0	0	0	0
49Gln HE21	16Glu OE1	100	0	0	0	0
54Arg HH12	58Asp OD1	100	0	0	0	0
55Thr HG1	58Asp OD1	0	44	83	29	86

where the angle brackets indicate a time or a combined time and ensemble average. Molecule 4 was selected because it is the most stable. RMSFs of the $C\alpha$ atoms were calculated directly from the simulation trajectory (Fig. 20.1.3.4a) after applying a translational fit using the $C\alpha$ atoms of residues 1–72, which are well defined in the X-ray structure (Fig. 20.1.3.4b), and after applying both a rotational and a translational fit on residues 1–72 (Fig. 20.1.3.4c). In Fig. 20.1.3.4(d), a translational and rotational fit was applied to all $C\alpha$ atoms (residues 1–76). Removal of the overall translational component of motion reduces the positional fluctuations by 0.04 nm on average. Only the RMSFs in the proximity of the end of the large α -helix formed by residues 23–34 are not affected by the

introduction of translational fitting. In contrast, it is exactly this region where the fluctuations are substantially lowered by introducing an additional rotational fit. The regions before residue 27 and after residue 42 are only slightly affected by the removal of overall rotation. These findings suggest that the entire protein translates by about 0.04 nm, while the α -helix region remains close to its initial position, thus rotating relative to the rest of the protein. Inclusion of the four C-terminal residues in the fitting procedure only affects the RMSFs of these residues and residues in the rotating part of the molecule, indicating that these four residues move together with the rest of the molecule. The atom-position fluctuations obtained by applying a full (rotational and transla-

Table 20.1.3.2. Occurrence of intermolecular hydrogen bonds (%) during the final 1.6 ns of the simulation

The criteria for a hydrogen bond to be present are: angle donor–hydrogen–acceptor $\geq 135^\circ$, distance hydrogen–acceptor ≤ 0.25 nm. Hydrogen bonds are shown if they are either present in the X-ray structure or if at least one of the four protein molecules in the unit cell shows the hydrogen bond of interest for at least 50% of the simulation time.

Hydrogen bond		X-ray structure	Molecular dynamics			
			Molecules 1–4	Molecules 2–3	Molecules 3–2	Molecules 4–1
6 Lys HZ3	51 Glu OE1	100	0	0	0	0
12 Thr HG1	18 Glu OE1	0	56	57	75	34
49 Gln H	8 Leu O	0	10	34	0	67
68 Hish HE2	32 Asp OD2	0	0	0	53 (3–1)	13 (4–2)
71 Leu H	58 Asp O	0	65	0	0	0

20. ENERGY CALCULATIONS AND MOLECULAR DYNAMICS

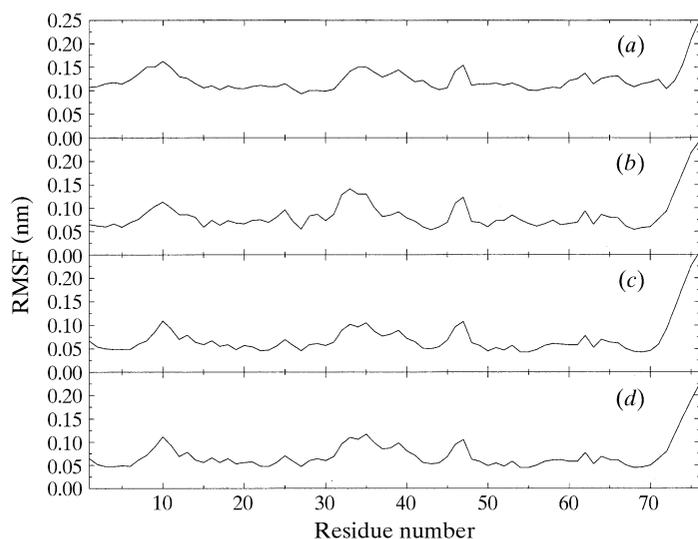


Fig. 20.1.3.4. Root-mean-square $C\alpha$ -atom-position fluctuations (RMSFs) in nm are shown for molecule 4 as a function of residue number. RMSFs are averaged over the final 1.6 ns. In (a), no fitting was applied; in (b), translational fitting was applied using the $C\alpha$ atoms of residues 1–72; and in (c), the rotational component of motion was also removed. In (d), translational and rotational fitting was applied over all $C\alpha$ atoms (1–76).

tional) fit are determined by internal motion only. The largest RMSFs for residues 1–72 are 0.12 nm. RMSFs of the two C-terminal glycines are 0.26 nm if the last four residues are excluded from the fitting and 0.22 nm otherwise.

If the same properties are examined but averaged over all the chains, similar trends can be observed (Fig. 20.1.3.5). If no fitting is applied, the RMSFs of 0.24 nm, on average, indicate that the different molecules show relative translation and rotation. After translational fitting is applied, the mean RMSFs drop to 0.18 nm. Thus, the molecules translate within the unit cell. If, in addition, the rotational component of overall motion is removed, the whole helix region is much less mobile than before, and the mean RMSFs drop to 0.14 nm. The same holds for region 47–64, dominated by the rotation of part of chain 1. Fluctuations are generally much larger than before when only chain 4 was observed, again indicating that the distinct chains behave in an uncorrelated way. The size of the

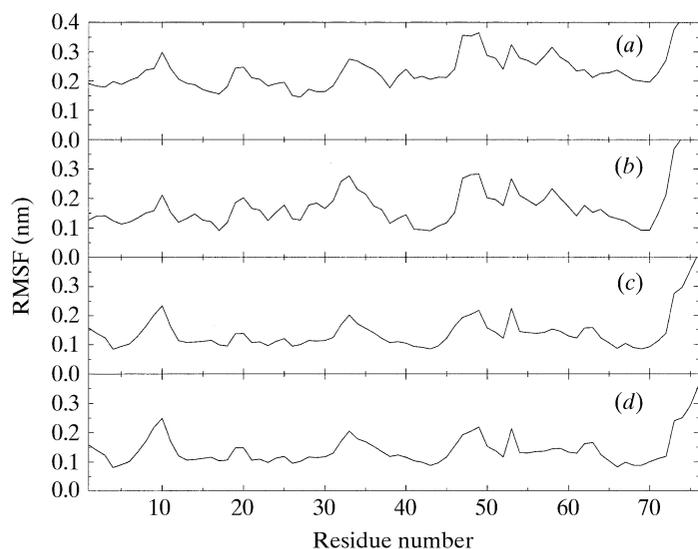


Fig. 20.1.3.5. Root-mean-square $C\alpha$ -atom-position fluctuations (RMSFs) in nm are shown using the same fitting protocols as in Fig. 20.1.3.4, but averaged over all four protein molecules in the unit cell.

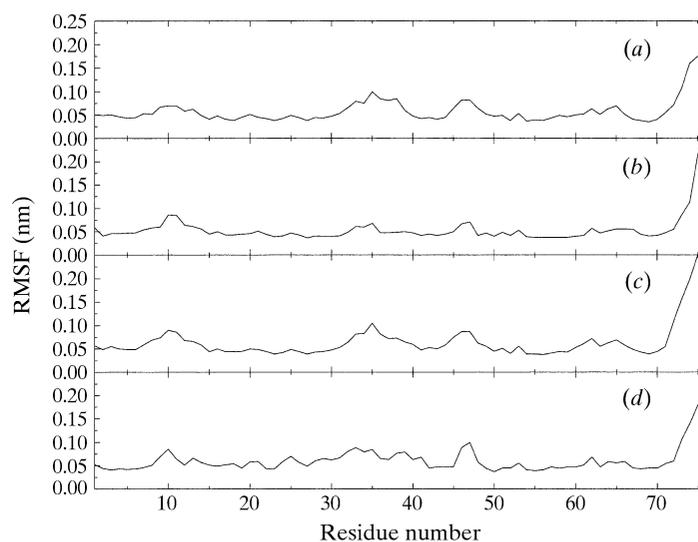


Fig. 20.1.3.6. Root-mean-square $C\alpha$ -atom-position fluctuations (RMSFs) in nm are shown for molecule 4, with full translational and rotational fitting over the $C\alpha$ atoms of residues 1–72. Different averaging periods are compared. (a) shows the RMSFs for the period 400–800 ps, (b) shows those for the period 800–1200 ps. In (c), the results for the previous two periods are averaged (400–1200 ps), and in (d), the results for the period 1200–2000 ps are shown.

peaks of the RMSFs averaged over all chains is around 0.22 nm, compared with 0.27 nm when overall rotation is still present. Thus, in addition to internal rotations, relative rotations of the different molecules occur. If the fit is not applied only to the well defined $C\alpha$ atoms of residues 1–72, the RMSF value becomes slightly higher – apart from the C-terminal region – but this effect is small, with the mean RMSF staying at 0.14 nm. However, the relative heights of the peaks differ, which shows that it is crucial to define a standard fitting protocol that must not be changed during the course of the analysis.

20.1.3.4. Effect of the averaging period

In Fig. 20.1.3.6, we concentrate again on molecule 4. Comparing different averaging periods of 400 and 800 ps with different starting points, it can be seen that, in general, the later the simulation, the less motion observed. During the period 800–1200 ps, only the small region between residues 9 and 12 shows more mobility than that between 400 and 800 ps. In the stable region, the rest of the molecule shows the same mobility as in the earlier time period. In the parts that are most mobile between 400 and 800 ps, the motions decrease significantly after the latter time point, indicating that equilibrium is reached. Focusing on the longer averaging periods, 400–1200 ps versus 1200–2000 ps, we see that over the whole chain mobility is comparable, indicating clear equilibrium as far as internal motions are concerned. The fluctuations during the 400–1200 ps period are of the same size as those of the shorter subperiods, 400–800 ps and 800–1200 ps. They are thus determined by movements on a timescale shorter than 400 ps.

In Fig. 20.1.3.7, the effect of different averaging periods is again considered, but taking the whole unit cell into account. Comparing RMSFs between 400 and 800 ps with those of the following 400 ps period, no significant difference is seen. In fact, contrary to what was observed when chain 4 was examined, somewhat more mobility is evident in the later period compared with the earlier one. This difference shows that although the configurations of the single chains converge rapidly, the different chains visit different parts of phase space. The fact that the fluctuations in the period 400–1200 ps are between those of the two shorter analysis periods is

20.1. MOLECULAR DYNAMICS: CONVERGENCE OF UBIQUITIN

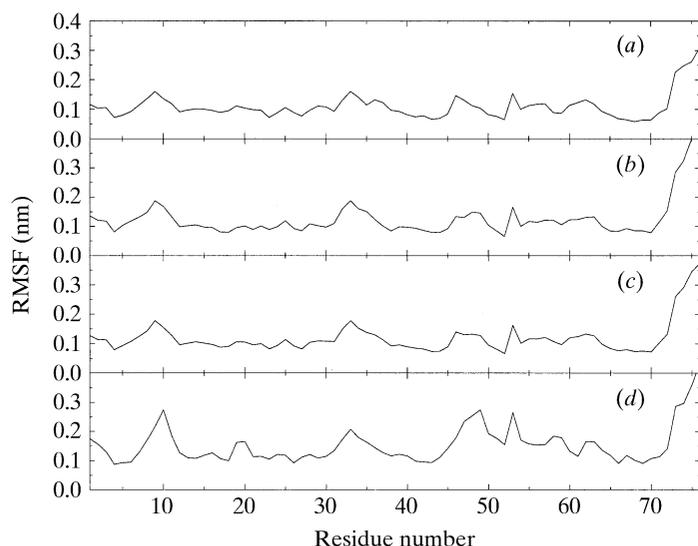


Fig. 20.1.3.7. Root-mean-square $C\alpha$ -atom-position fluctuations (RMSFs) in nm are shown using the same averaging periods as in Fig. 20.1.3.6, but averaged over all four protein molecules in the unit cell.

another indication that the single-chain movements are converged on these short timescales. In the last 800 ps of the simulation, the RMSFs are substantially higher than in the 800 ps window before that. All of the peaks can be traced back to one of the single chains. If only one of the four molecules differs strongly from the other three, this one determines the magnitude of the fluctuations of the average. The peak at residue 10 comes from chain 2, the ones around residue 20 and the whole region 47–64 are determined by chain 1. The peak at residue 33 originates in chain 3, which at this point differs substantially from the mean MD structure (Fig. 20.1.3.3).

20.1.3.5. Internal motions of the proteins

Fig. 20.1.3.8 displays the atomic root-mean-square position fluctuations for the $C\alpha$ atoms of the four protein molecules during the whole analysis period, together with corresponding values obtained using equation (20.1.3.1) and the crystallographic B factors. Rotational and translational fitting was applied using the $C\alpha$ atoms of residues 1–72, and the fluctuations were averaged over the final 1.6 ns. The mobility of the stable secondary-structure elements in the simulation is comparable with that inferred from the experiment. There is a correlation between the more mobile parts of the proteins in the simulation and large B factors in the X-ray structure, but the magnitude of the fluctuations is overestimated in the simulation. The movements of the single chains can be rationalized as follows. In chain 1, the whole region from Gly47 onwards rotates around a stable axis formed by residues 41–46. This part lies, as do all the flexible regions, on the exterior of the protein. Residues 19 and 20, which are stable in all but this single chain, are in contact with this moving part. This rotation, which tends to compact the protein, occurs during the 200 ps period between 1350 and 1550 ps after the start of the simulation, in which the atom-position RMSD from the X-ray structure increases significantly (Fig. 20.1.3.2). Overall, chain 2 is more stable than chain 1. Nevertheless, the end of the unwinding helix shows large fluctuations. In the course of this deformation, the side-chain nitrogen atom of Lys11 moves from close to the OE atom of Glu34 towards the backbone oxygen atom of Lys33, which is associated with a change in the position of Gly10. A similar but smaller motion occurs in chain 4. Both lysines, Lys33 and Lys63, are fully exposed to the solvent and have no intramolecular contacts. In chain 3, the

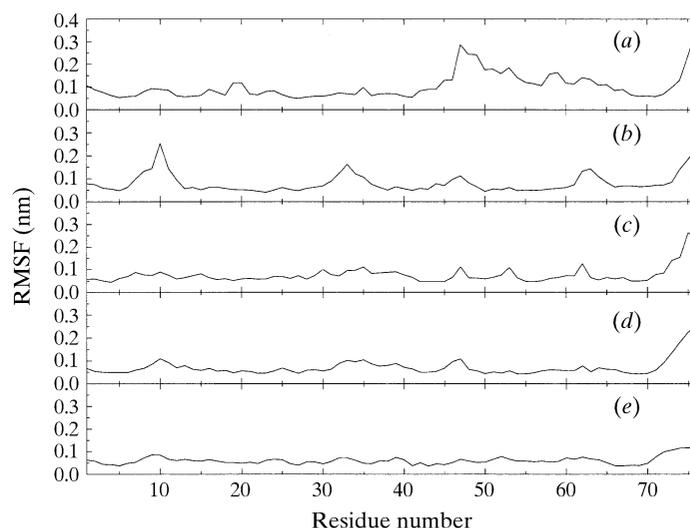


Fig. 20.1.3.8. Root-mean-square $C\alpha$ -atom-position fluctuations (RMSFs) in nm for the four protein molecules in the unit cell as a function of the residue number. Full translational and rotational fitting was applied to the $C\alpha$ atoms of residues 1–72 using the final 1.6 ns of the simulation [(a)–(d)]. (e) shows the corresponding values defined by equation (20.1.3.1), obtained from experimental B factors.

flexible residues are also not part of secondary-structure elements and are located on the outside of the protein. The backbone oxygen atom of Gln62 that moves in all the four chains has, in addition, the closest contact to another heavy atom: the OG1 atom of Ser65 is only 2.51 Å away, and the van der Waals repulsion of these atoms causes them to move further away from one another. The mobile residues in chain 4 are again in contact with the solvent, Gly35, Gly47, Gln62, the end of the helix and Gly10. The terminal residues of all the protein molecules are very mobile, as observed experimentally in the crystal.

20.1.3.6. Dihedral-angle fluctuations and transitions

Backbone dihedral-angle fluctuations and transitions are examined in Tables 20.1.3.3 and 20.1.3.4 using different analysis periods. After the first 400 ps of analysis, the φ/ψ dihedral-angle fluctuations differ only slightly, but if longer averaging times are chosen, the different protein molecules show larger differences from one another. These fluctuations also increase for longer analysis times, indicating that they are not yet converged after 2 ns. In the period from 800 to 1200 ps, chain 3 shows a large increase in mean-square dihedral-angle fluctuations, whereas the $C\alpha$ -atom-position RMSDs with respect to the X-ray structure during the same time fluctuate around a plateau. Thus, there is a lot of flexibility without the simulation structure diverging from the experimental one. Protein molecule 3, for example, shows the largest φ/ψ fluctuations of all the four molecules, and it shows the lowest atom-position RMSDs of $C\alpha$ atoms from the X-ray structure at the end of the simulation (Fig. 20.1.3.2), indicating that it explores phase space around the equilibrium structure. If, in contrast, the $C\alpha$ -atom-position RMSDs, with respect to the X-ray structure, increase significantly, larger dihedral-angle fluctuations are also observed, for example, in molecule 1 after 1200 ps.

Concerning relaxation, observations similar to those made before can be made when analysing dihedral-angle transitions (Table 20.1.3.4). The number of transitions for the different chains differs by about 20%. Within a single chain, however, the number of transitions increases in proportion to the observation time. Again, the protein molecules showing the most transitions do not have the largest $C\alpha$ -atom-position RMSDs from the X-ray structure. Thus,

20. ENERGY CALCULATIONS AND MOLECULAR DYNAMICS

Table 20.1.3.3. Root-mean-square fluctuations of polypeptide backbone and ψ dihedral angles ($^{\circ}$) for the different molecules using different time-averaging periods

The bottom row shows the averages over all four protein molecules.

Molecule	400–800 ps φ/ψ	400–1200 ps φ/ψ	400–2000 ps φ/ψ
1	18.4/22.9	19.5/23.7	31.0/33.6
2	17.2/17.0	18.6/18.7	23.6/26.8
3	18.5/20.4	25.6/26.3	35.3/37.5
4	19.7/18.8	19.4/20.3	21.6/28.8
All	26.1/26.2	28.0/28.6	35.2/38.1

only certain dihedral-angle flips induce large movements that determine the RMSD value.

20.1.3.7. Water diffusion

In Fig. 20.1.3.9, the number of water oxygen atoms with a given atomic root-mean-square position fluctuation (RMSF) are plotted. The time evolution and the shapes of these curves are similar to those obtained for bulk water, a Gaussian distribution with the maximum at larger RMSF values and larger standard deviations when using longer averaging times. Using a diffusion constant of bulk SPC water at 300 K of $3.9 \times 10^{-3} \text{ nm}^2 \text{ ps}^{-1}$ (Smith & van Gunsteren, 1995), the root-mean-square position fluctuation for an average water molecule would be 1.25 nm for a 400 ps period, 1.77 nm for a 800 ps period, and 2.5 nm for a 1600 ps period. Comparison of these values with the distributions in Fig. 20.1.3.9 indicates that the motion of most of the crystal water molecules is restricted by the crystalline environment. A number of water molecules adopt well defined positions for periods of the order of some 100 ps. Several water molecules were also observed to visit the same location repeatedly. This indicates that, although not remaining fixed in space, they stay close to a crystallographically determined site which they revisit periodically, alternating with other water molecules.

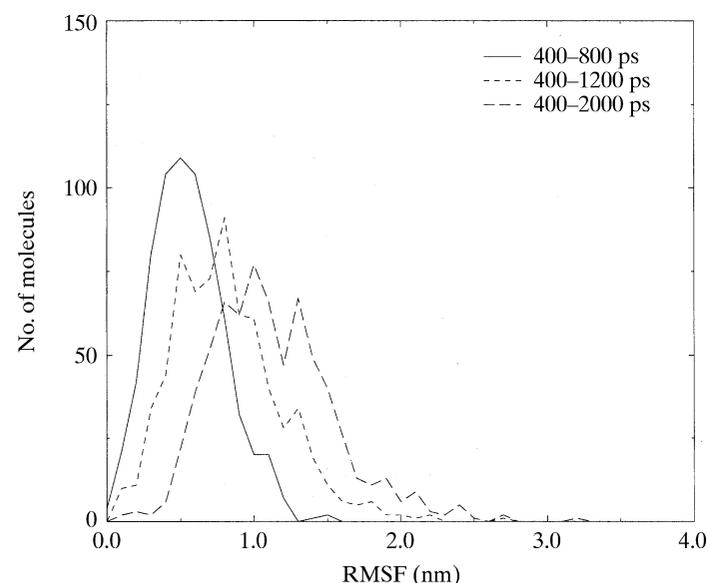


Fig. 20.1.3.9. The number of water molecules with a given root-mean-square oxygen-position fluctuation (RMSF) in nm are shown for different averaging periods: 400–800 ps (solid line), 400–1200 ps (short-dashed line), 400–2000 ps (long-dashed line).

Table 20.1.3.4. Number of protein-backbone dihedral-angle transitions per 100 ps for the different molecules using different time periods

Dihedral angles with threefold and sixfold potential-energy wells are distinguished. The bottom rows show the averages over all protein molecules.

(a) 120° transitions

Molecule	400–800 ps	400–1200 ps	400–2000 ps
1	46.5	45.4	47.7
2	40.5	41.5	47.3
3	50.5	57.1	51.3
4	44.8	46.4	46.4
All	45.6	47.6	48.2

(b) 60° transitions

Molecule	400–800 ps	400–1200 ps	400–2000 ps
1	245.5	246.6	289.3
2	271.5	272.1	261.3
3	381.5	381.0	348.3
4	356.8	325.4	325.4
All	313.8	306.3	306.1

(c) All transitions

Molecule	400–800 ps	400–1200 ps	400–2000 ps
1	292.0	292.0	336.9
2	312.0	313.7	308.6
3	432.0	438.1	399.6
4	401.5	371.8	371.8
All	359.4	353.9	354.2

20.1.4. Conclusions

In the present molecular-dynamics simulation, fast convergence in energy, within about 100 ps, was observed. Other properties, such as dihedral-angle fluctuations and backbone atom-position fluctuations, converged on an intermediate timescale of hundreds of picoseconds. Root-mean-square deviations of the simulated protein molecules from the starting X-ray structure required of the order of 1 ns to reach a plateau. Longer simulations would be necessary to obtain convergence for all molecular properties. The convergence of quickly relaxing properties of the system, such as the energies, is not a reliable indicator of the degree of global convergence in such a molecular-dynamics simulation.

The *GROMOS96* force field used in this simulation largely reproduces the secondary structure and the relative internal mobility of ubiquitin. The simulation does, however, overestimate the magnitude of the fluctuations in the most mobile regions of the protein. The different protein molecules were observed to translate and rotate relative to one another during this simulation. This indicates that the force field would not be able to reproduce the experimental melting temperature of this crystal under the conditions simulated.

Acknowledgements

The authors wish to thank Dr Thomas Huber for fruitful discussions and Dr Alan Mark for critical reading of the manuscript. Financial support was obtained from the Schweizerischer Nationalfonds (Project 21-41875.94), which is gratefully acknowledged.

20. ENERGY CALCULATIONS AND MOLECULAR DYNAMICS

r.m.s.d.'s between MD average structures and the starting X-ray structures correlate with atomic resolution.

The r.m.s.d.'s between main-chain atoms in the starting X-ray structures and simulation snapshots as a function of time are presented in Fig. 20.2.8.2. The 1.1 Å resolution structure has the most stable trajectory during the 500 ps trajectory, with an average r.m.s. value of 1.01 (9) Å. The 1.7 Å resolution structure has an r.m.s. value of 0.98 (22) Å. In this simulation, the r.m.s.d.'s fluctuate more widely from the average value, with small differences in the first 200 ps, larger ones between 200 and 400 ps, and again smaller ones in the last 100 ps. For the 2.7 Å resolution structure, the average over the simulation is 1.28 (21) Å. From the results presented here, it is concluded that the higher-resolution structures are more stable during MD simulations and have a shorter equilibration period (50 ps for 1.1 Å resolution and

over 300 ps for 2.7 Å resolution). This conclusion is consistent with larger errors in the atomic coordinates of X-ray structures determined from lower-resolution data.

Acknowledgements

This work was supported by grants to CBP from the NIH (R01-GM39478, AI39639). CBP was supported by a Research Career Development Award from the NIH (K04-GM00661) and VMD is a DOE/SLOAN Postdoctoral Fellow in computational biology. The computing facilities shared by the Structural Biology group were supported by grants from the Lucille P. Markey Foundation and the Purdue University Academic Reinvestment Program.

References

20.1

- Berendsen, H. J. C., van Gunsteren, W. F., Zwinderman, H. R. J. & Geurtsen, R. (1986). *Simulations of proteins in water*. *Ann. N. Y. Acad. Sci.* **482**, 269–285.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). *Molecular dynamics with coupling to an external bath*. *J. Chem. Phys.* **81**, 3684–3690.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. (1981). *Interaction models for water in relation to protein hydration*. In *Intermolecular forces*, edited by B. Pullman, pp. 331–342. Dordrecht: Reidel.
- 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. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures*. *J. Mol. Biol.* **112**, 535–542.
- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R-factor refinement by molecular dynamics*. *Science*, **235**, 458–460.
- Fennen, J., Torda, A. E. & van Gunsteren, W. F. (1995). *Structure refinement with molecular dynamics and a Boltzmann-weighted ensemble*. *J. Biomol. NMR*, **6**, 163–170.
- Fujinaga, M., Gros, P. & van Gunsteren, W. F. (1989). *Testing the method of crystallographic refinement using molecular dynamics*. *J. Appl. Cryst.* **22**, 1–8.
- Gros, P. & van Gunsteren, W. F. (1993). *Crystallographic refinement and structure-factor time-averaging by molecular dynamics in the absence of a physical force field*. *Mol. Simul.* **10**, 377–395.
- Gros, P., van Gunsteren, W. F. & Hol, W. G. J. (1990). *Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics*. *Science*, **249**, 1149–1152.
- Gunsteren, W. F. van & Berendsen, H. J. C. (1990). *Computer simulations of molecular dynamics: methodology, applications and perspectives in chemistry*. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023.
- Gunsteren, W. F. van, Berendsen, H. J. C., Hermans, J., Hol, W. G. J. & Postma, J. P. M. (1983). *Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data*. *Proc. Natl Acad. Sci. USA*, **80**, 4315–4319.
- Gunsteren, W. F. van, Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P. & Tironi, I. G. (1996). *Biomolecular simulation: the GROMOS96 manual and user guide*. Vdf Hochschulverlag, Zürich, Switzerland.
- Gunsteren, W. F. van, Bonvin, A. M. J. J., Daura, X. & Smith, L. J. (1997). *Aspects of modelling biomolecular structures on the basis of spectroscopic or diffraction data*. In *Modern techniques in protein NMR*, edited by Krishna & Berliner. Plenum.
- Gunsteren, W. F. van, Brunne, R. M., Gros, P., van Schaik, R. C., Schiffer, C. A. & Torda, A. E. (1994). *Accounting for molecular mobility in structure determination based on nuclear magnetic resonance spectroscopic and X-ray diffraction data*. *Methods Enzymol.* **239**, 619–654.
- Gunsteren, W. F. van, Kaptein, R. & Zuiderweg, E. R. P. (1984). *Use of molecular dynamics computer simulations when determining protein structure by 2D-NMR*. In *Proceedings of the NATO/CECAM workshop on nucleic acid conformation and dynamics*, edited by W. K. Olson, pp. 79–97. France: CECAM.
- Gunsteren, W. F. van & Karplus, M. (1981). *Effect of constraints, solvent and crystal environment on protein dynamics*. *Nature (London)*, **293**, 677–678.
- Gunsteren, W. F. van & Karplus, M. (1982). *Protein dynamics in solution and in crystalline environment: a molecular dynamics study*. *Biochemistry*, **21**, 2259–2274.
- Harvey, T. S. & van Gunsteren, W. F. (1993). *The application of chemical shift calculation to protein structure determination by NMR*. In *Techniques in protein chemistry IV*, pp. 615–622. New York: Academic Press.
- Heiner, A. P., Berendsen, H. J. C. & van Gunsteren, W. F. (1992). *MD simulation of subtilisin BPN' in a crystal environment*. *Proteins*, **14**, 451–464.
- Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). *A protein structure from nuclear magnetic resonance data, lac repressor headpiece*. *J. Mol. Biol.* **182**, 179–182.
- Levitt, M., Hirshberg, M., Sharon, R. & Daggett, V. (1995). *Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution*. *Comput. Phys. Commun.* **91**, 215–231.
- Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. *J. Comput. Phys.* **23**, 327–341.
- Scheek, R. M., Torda, A. E., Kemmink, J. & van Gunsteren, W. F. (1991). *Structure determination by NMR: the modelling of NMR parameters as ensemble averages*. In *Computational aspects of the study of biological macromolecules by nuclear magnetic resonance spectroscopy*, edited by J. C. Hoch, F. M. Poulsen & C. Redfield, NATO ASI Series, Vol. A225, pp. 209–217. New York: Plenum Press.
- Schiffer, C. A., Gros, P. & van Gunsteren, W. F. (1995). *Time-averaging crystallographic refinement: possibilities and limitations using α -cyclodextrin as a test system*. *Acta Cryst.* **D51**, 85–92.
- Schiffer, C. A., Huber, R., Wüthrich, K. & van Gunsteren, W. F. (1994). *Simultaneous refinement of the structure of BPTI against NMR data measured in solution and X-ray diffraction data measured in single crystals*. *J. Mol. Biol.* **241**, 588–599.
- Shi, Y.-Y., Yun, R.-H. & van Gunsteren, W. F. (1988). *Molecular dynamics simulation of despentapeptide insulin in a crystalline environment*. *J. Mol. Biol.* **200**, 571–577.

REFERENCES

20.1 (cont.)

- Smith, P. E. & van Gunsteren, W. F. (1995). *Reaction field effects on the simulated properties of liquid water*. *Mol. Simul.* **15**, 233–245.
- Torda, A. E., Brunne, R. M., Huber, T., Kessler, H. & van Gunsteren, W. F. (1993). *Structure refinement using time-averaged J-coupling constant restraints*. *J. Biomol. NMR*, **3**, 55–66.
- Torda, A. E., Scheek, R. M. & van Gunsteren, W. F. (1990). *Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat*. *J. Mol. Biol.* **214**, 223–235.
- Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). *Structure of ubiquitin refined at 1.8 Å resolution*. *J. Mol. Biol.* **194**, 531–544.

20.2

- Abseher, R., Ludemann, S., Schreiber, H. & Steinhauser, O. (1995). *NMR cross-relaxation investigated by molecular dynamics simulation: a case study of ubiquitin in solution*. *J. Mol. Biol.* **249**, 604–624.
- Brooks, C. L. III, Karplus, M. & Pettitt, B. M. (1988). *Proteins: a theoretical perspective of dynamics, structure and thermodynamics*. *Adv. Chem. Phys.* **71**, 1–259.
- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R factor refinement by molecular dynamics*. *Science*, **235**, 458–460.
- Caves, L. S. D., Evanseck, J. D. & Karplus, M. (1998). *Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin*. *Protein Sci.* **7**, 649–666.
- Chatfield, D. C., Szabo, A. & Brooks, B. R. (1998). *Molecular dynamics of staphylococcal nuclease: comparison of simulation with 15N and 13C NMR relaxation data*. *J. Am. Chem. Soc.* **120**, 5301–5311.
- Christiansen, D., Peram, J. W. & Petersen, H. G. (1993). *On the fast multipole method for computing the energy of periodic assemblies of charged and dipolar particles*. *J. Comput. Phys.* **107**, 403–405.
- Cooper, A. (1976). *Thermodynamic fluctuations in protein molecules*. *Proc. Natl Acad. Sci. USA*, **92**, 2740–2741.
- Darden, T., York, D. & Pedersen, L. (1993). *Particle mesh Ewald (PME): a N log(N) method for Ewald sums in large systems*. *J. Chem. Phys.* **98**, 10089–10092.
- Duan, Y. & Kollman, P. A. (1998). *Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution*. *Science*, **282**, 740–744.
- Essmann, U., Perrera, L., Berkovitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). *A smooth particle mesh Ewald method*. *J. Chem. Phys.* **103**, 8577–8593.
- Goodfellow, J. M. & Levy, R. M. (1998). *Theory and simulation*. *Curr. Opin. Struct. Biol.* **8**, 209–210.
- Gunsteren, W. F. van, Billeter, S. R., Eising, A. A., Hünenberger, P. H., Mark, A. E., Scott, W. R. P. & Tironi, I. G. (1996). *Biomolecular simulation: the GROMOS96 manual and user guide*. Vdf Hochschulverlag, Zurich, Switzerland.
- Hamiaux, C., Prangé, T., Riès-Kautt, M., Ducruix, A., Lafont, S., Astier, J. P. & Veessler, S. (1999). *The dcameric structure of bovine pancreatic trypsin inhibitor (BPTI) crystallized from thiocyanate at 2.7 Å resolution*. *Acta Cryst.* **D55**, 103–113.
- Hoover, W. G. (1985). *Canonical dynamics: equilibrium phase-space distributions*. *Phys. Rev. A*, **31**, 1695–1697.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). *Comparison of simple potential functions for simulating liquid water*. *J. Chem. Phys.* **79**, 926–935.
- Jorgensen, W. L. & Tirado-Rives, J. (1988). *The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin*. *J. Am. Chem. Soc.* **110**, 1657–1666.
- Kuriyan, J., Petsko, G. A., Levy, R. M. & Karplus, M. (1986). *Effect of anisotropy and anharmonicity on protein crystallographic refinement: an evaluation by molecular dynamics*. *J. Mol. Biol.* **190**, 455–479.
- Leeuw, S. W. de, Perram, J. W. & Smith, E. R. (1980). *Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants*. *Proc. R. Soc. London Ser. A*, **373**, 27–56.
- Levy, R. M., Karplus, M. & McCammon, J. A. (1981). *Increase of ¹³C NMR relaxation times in proteins due to picosecond motional averaging*. *J. Am. Chem. Soc.* **103**, 994–996.
- Levy, R. M., Karplus, M. & Wolynes, P. G. (1981). *NMR relaxation parameters in molecules with internal motion: exact Langevin trajectory results compared with simplified relaxation models*. *J. Am. Chem. Soc.* **103**, 5998–6011.
- Lipari, G. & Szabo, A. (1982). *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. I. Theory and range of validity*. *J. Am. Chem. Soc.* **104**, 4546–4559.
- McCammon, J. A., Gelin, B. R. & Karplus, M. (1977). *Dynamics of folded proteins*. *Nature (London)*, **267**, 585–590.
- McCammon, J. A. & Harvey, S. C. (1987). *Dynamics of proteins and nucleic acids*. Cambridge University Press.
- MacKerell, A., Wiorkiewicz-Kuczera, J. & Karplus, M. (1995). *An all atom empirical energy function for the simulation of nucleic acids*. *J. Am. Chem. Soc.* **117**, 11946–11975.
- MacKerell, A. D. Jr, Bashford, D., Bellott, M., Dunbrack, R. L. Jr, Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E. III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J. & Karplus, M. (1998). *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. *J. Phys. Chem. B*, **102**, 3586–3616.
- Nose, S. (1984). *A unified formulation of the constant temperature molecular dynamics methods*. *J. Chem. Phys.* **81**, 511–519.
- Parkin, S., Rupp, B. & Hope, H. (1999). *The structure of bovine pancreatic trypsin inhibitor at 125 K: definition of carboxyl-terminal residues glycine-57 and alanine 58*. In preparation.
- Pike, A. C., Brew, K. & Acharya, K. R. (1996). *Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase*. *Structure*, **4**, 691–703.
- Post, C. B. (1992). *Internal motional averaging and three-dimensional structure determination by NMR*. *J. Mol. Biol.* **224**, 1087–1101.
- Post, C. B., Brooks, B. R., Karplus, M., Dobson, C. M., Artymiuk, P. J., Cheetham, J. C. & Phillips, D. C. (1986). *Molecular dynamics simulations of native and substrate-bound lysozyme*. *J. Mol. Biol.* **190**, 455–479.
- Post, C. B., Dobson, C. M. & Karplus, M. (1989). *A molecular dynamics analysis of protein structural elements*. *Proteins*, **5**, 337–354.
- Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. *J. Comput. Phys.* **23**, 327–341.
- Schiffer, C. A., Gros, P. & van Gunsteren, W. F. (1995). *Time-averaging crystallographic refinement: possibilities and limitations using α-cyclodextrin as a test system*. *Acta Cryst.* **D51**, 85–92.
- Smith, P. E., van Schaik, R. C., Szyperski, T., Wuthrich, K. & van Gunsteren, W. F. (1995). *Internal mobility of the basic pancreatic trypsin inhibitor in solution: a comparison of NMR spin relaxation measurements and molecular dynamics simulation*. *J. Mol. Biol.* **246**, 356–365.
- Verlet, L. (1967). *Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules*. *Phys. Rev.* **159**, 98–103.
- Wlodawer, A., Nachman, J., Gilliland, G. L., Gallagher, W. & Woodward, C. (1987). *Structure of form III crystals of bovine pancreatic trypsin inhibitor*. *J. Mol. Biol.* **198**, 469–480.