International Tables for Crystallography (2006). Vol. F, Section 20.1.2, pp. 481–482.

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:

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

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, 1981, 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 *GROMOS*96 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

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 *GROMOS*96 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 kJ mol^{-1} 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 kJ mol^{-1} 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 Cα 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 the Cα 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 Cα 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 Cα 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.

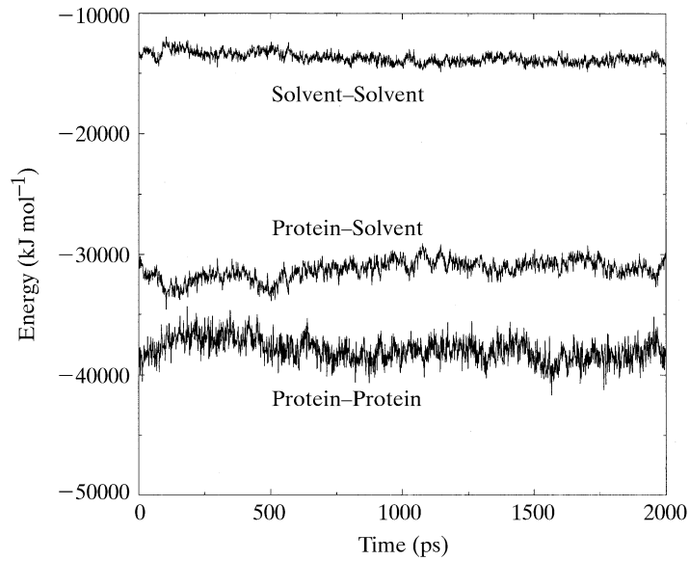

Fig. 20.1.3.1. Non-bonded energies (in kJ mol^{-1}) of the simulated system as a function of time.

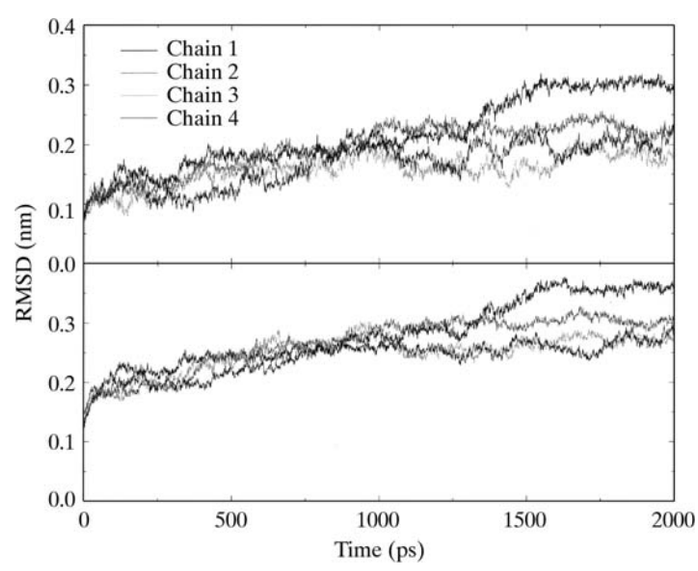

Fig. 20.1.3.2. Root-mean-square atom-positional deviations (RMSD) in nm from the X-ray structure of the four different protein molecules in the unit cell as a function of time. Rotational and translational fitting was applied using the Cα atoms of residues 1–72. The upper and lower graphs show the deviations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

references