*Journal of Biomolecular NMR*, **18:** 1–12, 2000.
*KLUWER/ESCOM*



# On the similarity of properties in solution or in the crystalline state: A molecular dynamics study of hen lysozyme

U. Stocker (stocker@igc.phys.chem.ethz.ch), K. Spiegel & W.F. van Gunsteren (wfvgn@igc.phys.chem.ethz.ch)
*Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland*

Received 11 February 2000; Accepted 29 May 2000

*Key words:* lysozyme, molecular dynamics simulation, NMR solution structure, protein dynamics, X-ray crystal structure

**Abstract**

As protein crystals generally possess a high water content, it is assumed that the behaviour of a protein in solution and in crystal environment is very similar. This assumption can be investigated by molecular dynamics (MD) simulation of proteins in the different environments. Two 2*ns* simulations of hen egg white lysozyme (HEWL) in crystal and solution environment are compared to one another and to experimental data derived from both X-ray and NMR experiments, such as crystallographic B-factors, NOE atom–atom distance bounds, $^{3}J_{H\,N\alpha}$-coupling constants, and $^{1}$H-$^{15}$N bond vector order parameters. Both MD simulations give very similar results. The crystal simulation reproduces X-ray and NMR data slightly better than the solution simulation.

## Introduction

Molecular dynamics (MD) simulations of biomolecular systems. such as proteins, DNA, or membranes, are generally performed in their natural environment, in aqueous solution, or, in the case of membranes, as a bilayer with water on both sides or as micelles. Most protein structures, however, are determined by X-ray crystallography, where the molecule of interest is in a different environment. In contrast to small molecule crystals, in protein crystals, the water content is high (generally around 50%). In addition, in protein crystals, only few water molecules occupy well-defined sites, most of the solvent is disordered and thus comparable to the protein environment in solution. Third, if proteins are crystallized in different space groups, the structures are usually not very different from one another. Fourth, for a number of small proteins the structure has been determined in the crystalline state by X-ray diffraction as well as in aqueous solution by NMR spectroscopy (Billeter et al., 1989, 1992; Baldwin et al., 1991; Kallen et al., 1991; Moore et al., 1991; Berndt et al., 1992; Braun et al., 1992; Neri et al., 1992; Fede et al., 1993; Dornberger et al., 1998; Lu et al., 1999; Yang et al., 1999) and structural differences appear to be minor, e.g. involving loop or side-chain conformations. So, it is generally assumed that proteins in crystals and in solution have, apart from polar side chains that take part in crystal packing contacts, a very similar structure in the two different environments. The motions in the crystal, especially side-chain motions, are expected to be of lesser amplitude due to crystal packing contacts.

Molecular dynamics simulation provides a good tool to investigate structural and dynamical differences on a sub-nanosecond time-scale. An early study (van Gunsteren and Berendsen, 1984) compared the structural properties of bovine pancreatic trypsin inhibitor (BPTI) on a picosecond time-scale by comparing the results of two 25 ps MD simulations of BPTI, one in crystalline form and the other in aqueous solution. To our knowledge, however, since then no systematic investigation of simulations of a protein in crystal and solution, with a detailed analysis of the differences observed, has been published. In the present work, two 2 ns MD simulations of hen egg white lysozyme (HEWL) are compared. An orthorhombic unit cell containing four protein molecules was simulated, and

the results are compared to those of a simulation of the protein in solution.

Data computed from the two simulations are compared with data derived from X-ray crystallography (Carter et al., 1997), such as atomic positions, B-factors, and hydrogen bonding patterns, or with data derived from NMR experiments (Smith et al., 1991, 1993; Buck et al., 1995), such as NOE atom–atom distance bounds, $^3J_{H\,N\alpha}$-coupling constants, and $^1$H-$^{15}$N bond vector order parameters. In addition, the data from the two simulations are compared to each other.

## Methods

### *Solution simulation*

Lysozyme consists of 129 amino acids with 1001 non-hydrogen atoms. Hydrogen atoms attached to aliphatic carbon atoms are incorporated into these (the united atom approach), and the remaining 321 hydrogen atoms are treated explicitly. The protein was simulated at pH 6. The amino acids Glu and Asp were taken to be deprotonated; Lys, Arg, and His residues were protonated, leading to a charge of $+9$ electron charges per protein molecule. The crystal structure of lysozyme (entry 1AKI (Carter et al., 1997) of the Brookhaven Protein Database (Bernstein et al., 1977) determined at 1.5 Å resolution (Artymiuk et al., 1982) was used as a starting structure. Truncated octahedron periodic boundary conditions were used with a box length of 7.7392 nm between the quadratic surfaces. 7122 SPC water molecules (Berendsen et al., 1981) were added from an equilibrated cubic box containing 216 water molecules (van Gunsteren et al., 1996). The added water molecules were selected such that no water oxygen atom is closer than 0.23 nm to a non-hydrogen atom of the protein or another water oxygen atom. The system, protein and water was initially energy minimised for 100 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 nm$^2$). The minimised structure was then pre-equilibrated in a short MD run of 100 steps of 0.002 ps, still restraining protein atom positions. Initial velocities were assigned from a Maxwell–Boltzmann distribution at 300 K. 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. A follow-up 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 cut-off radii $R_{cp} = 0.8$ nm for the charge-group (van Gunsteren et al., 1996) pair-list and $R_{cl} = 1.4$ nm for the longer-range non-bonded (van Gunsteren and Berendsen, 1990) interactions. The pair-list for the (short-range) non-bonded interactions and the longer-range forces was updated every 10 fs. Reaction-field forces were included (Tironi et al., 1995) originating from a dielectric continuum beyond a radius of $R_{rf} = 1.4$ nm using a self-consistent relative dielectric permittivity $\epsilon_{rf} = 54$ for SPC water (Smith and van Gunsteren, 1994). Nine counterions were added by replacing water molecules in the following way. The water molecule having the highest electrostatic field at the water oxygen site, after the successive placement of previous ions, was replaced by a chloride anion. After having introduced the ions, the energy was again minimised using 100 steps of steepest descent and protein atom position constraining. The total size of the system was 1322 protein atoms, 9 chloride anions and 7113 water molecules, leading to a total system size of 22670 atoms. The simulation was carried out over 2 ns with a time step of 0.002 ps. Every 500th step, the configuration was saved. The first 300 ps of the simulation were treated as equilibration period, the remaining 1.7 ns were used for analysis. All simulations were performed using the GROMOS96 force field (version 43A1) and software (van Gunsteren et al., 1996; Scott et al., 1999). The simulation parameters have been summarized in Table 1.

### *Crystal simulation*

For the crystal simulation, basically the same set-up was used as for the solution simulation. The differences are the geometry of the periodic box containing the simulated system and the number of protein molecules, ions, and water molecules. Four protein molecules related by the crystallographic symmetry $P2_12_12_1$ were placed in the orthorhombic unit cell with a = 5.9062 nm, b = 6.8451 nm, and c = 3.0517 nm. Seventy-eight crystallographically observed water molecules were added together with their symmetry-related equivalents. After that, 1439 water molecules were added as described in the previous subsection leading to a water content of 42.5% (v/v). Energy minimisation, placing of the ions, and pre-

*Table 1.* Simulation parameters used in the two simulations

| | Simulation type | |
|---|---|---|
| | crystal | solution |
| Protein molecules | 4 | 1 |
| Counterions ($Cl^-$) | 36 | 9 |
| Water molecules | 1715 | 7113 |
| Number of atoms | 10469 | 22670 |
| Periodic boundary conditions | Orthorhombic | Truncated octahedron |
| Box lengths (nm) | 5.9062 | 7.7392 |
| | 6.8451 | 7.7392 |
| | 3.0517 | 7.7392 |
| Box volume ($nm^3$) | 123.38 | 231.77 |
| Mass density ($gcm^{-3}$) | 1.203 | 1.023 |
| Water content (by mass, %) | 34.57 | 89.76 |
| Timestep (fs) | 2 | |
| Relative SHAKE precision | $10^{-4}$ | |
| Temperature coupling at (K) | 300 | |
| Using coupling time (ps) | 0.1 | |
| Pressure coupling | no | |
| Pair-list, cut-off radius $R_{cp}$ (nm) | 0.8 | |
| Pair-list update frequency ($ps^{-1}$) | 100 | |
| Non-bonded interaction cut-off $R_{cl}$ (nm) | 1.4 | |
| Poisson–Boltzmann reaction field (PBRF) | yes | |
| Beyond radius $R_{rf}$ (nm) | 1.4 | |
| Using relative dielectric permittivity $\epsilon_{rf}$ | 54 | |

equilibration simulations were performed as described in the previous subsection. The simulation parameters have been summarized in Table 1.

*Analysis*

The 1158 NOE distances, the 95 $^3J_{H\,N\alpha}$-coupling constants, and the 124 backbone and 28 side-chain $S^2$ order parameters were calculated form the trajectory as in (Smith et al., 1995). The $S^2$ order parameters were calculated using a 200 ps averaging window (moving through the whole 1700 ps analysis period) which approximates the time-scale of the N-H bond vector motions which determine the $S^2$ order parameters as derived from NMR experiments for HEWL (Evenäs et al., 1999). Similarly, 200 ps windows were used to compute the $r^{-3}$ averages for the NOE distances. The experimental values were taken from (Smith et al., 1991, 1993; Buck et al., 1995).

**Results**

Root-mean-square atom positional deviations (RMSD) from the starting (X-ray) structure are shown in Figure 1. In the top graph, values for $C_\alpha$ atoms of the solution simulation and of the four molecules in the crystal simulation are displayed, in the bottom graph values for all atoms. RMSD values for the solution simulation are converged after 800 ps for both $C_\alpha$ and all atoms. In the crystal simulation, the RMSD of $C_\alpha$ atoms are converged earlier, after 500 ps, whereas the side-chain RMSD only converge after 1 ns. The crystal simulation stays – in terms of RMSD – closer to the X-ray structure than the solution simulation. In the second nanosecond of the simulations, the RMSD for $C_\alpha$ atoms is 0.13 nm compared to 0.16 nm, the RMSD for all atoms is 0.22 nm compared to 0.26 nm in the crystal and the solution simulation, respectively. The different molecules in the crystal simulation show, after being converged, very similar behaviour. In the solution simulation, larger structural fluctuations are observed, which should be reflected in the calculated

B-factors (Figure 2). Atomic isotropic B-factors from simulation trajectories were calculated from atom-positional root-mean-square fluctuations for $C_\alpha$ atoms according to

$$B_i = \frac{8\pi}{3}\langle(\vec{r_i} - \langle\vec{r_i}\rangle)^2\rangle \ . \quad (1)$$

When comparing atomic B-factors resulting from structure refinement based on crystallographic X-ray diffraction data with B-factors obtained through Equation 1 from the atomic mean square positional fluctuations calculated from an MD trajectory, it should be kept in mind that these two types of B-factors are not wholly comparable (Hünenberger et al., 1995).

(1) The configuration space sampled in the experiment is incompletely sampled in the simulation since: (i) crystallographic B-factors derived from X-ray data include static disorder due to averaging over a collection of molecules; (ii) the simulation time (nanoseconds) is much shorter than the data aquisition time (>seconds).

(2) Crystallographic B-factors form an incomplete measure of the motion since: (i) they are a measure of the spread of the electron density as a function of position in the crystal, irrespective of which particular atom is contributing to the electron density (i.e., they result from fitting on the electron density map); (ii) they are, in general, restricted to a given maximum value during structure refinement, whereas the simulated B-factors are true atomic positional fluctuations, which may in principle grow infinitely with atomic mobility; (iii) systematic errors in the data, e.g. due to absorption, extinction and thermal diffuse scattering, may not have been corrected.

Although the backbone is more mobile in the solution simulation, the agreement of calculated and experimentally derived B-factors is similar for both simulations (Figure 2). Residues having large experimental B-factors also showed enhanced mobility in the simulation. Secondary structure regions (α-helices 4Gly-14Arg, 25Leu-36Ser, 88Ile-99Val, 108Trp-115Cys, and $3_{10}$-helix 119Asp-124Ile, and β-sheet 42Ala-60Ser, see Table 2) are stable and thus, in both simulations, not much flexibility is observed for these parts of the chain. Exceptions are residue 1Lys in the solution simulation, which in the crystal lies close to some crystallographically observed water molecules, and the considerably rearranged loop region on the outside of the protein between residues 115Cys and 119Asp. The overall trends in mobility are thus correctly sampled, magnitudes, however, are overestimated. Flexible regions show more mobility in the simulations compared to experimentally derived B-factors, stable regions show less mobility, which might be due to the fact that single-molecule B-factors were calculated from the simulations, whereas the experimentally derived B-factors contain contributions from many molecules.

Table 2 shows the backbone hydrogen bonding pattern in the X-ray structure and the percentage of hydrogen bonding as obtained by averaging over the single molecule trajectories from the simulations. A hydrogen bond is counted if the distance between the hydrogen atom and the acceptor atom is maximally 0.25 nm and the angle between donor atom, hydrogen atom and acceptor atom is larger than 135°. Hydrogen bonds are very well reproduced in both simulations. Generally, the solution simulation shows stronger hydrogen bonding than the crystal simulation. Of the 63 hydrogen bonds found in the X-ray structure, 59 show occupancies of over 20% in the solution simulation or over 20% average occupancy in the crystal simulation. The hydrogen bonds not reproduced can be rationalized as follows: 84Leu-81Ser is replaced by 84Leu–80Cys and 85Ser-81Ser. The other three hydrogen bonds lost in both simulations all lie in the very mobile region form residue 115Cys onwards. 118Thr-115Cys is partly replaced by 118Thr-114Arg (not shown in Table 2), 122Ala-119Asp is replaced by 123Trp-119Asp. Also the two following hydrogen bonds, 123Trp-120Val and 124Ile-121Gln, are, although still present over 20% in at least one of the simulations, shifted from a $3_{10}$-helical pattern towards an α-helical structure: 124Ile-120Val and 125Arg-121Gln. The hydrogen bond 127Cys-124Ile at the end of this helix shows slightly less than 20% occupancy in the simulations. The two hydrogen bonds showing the highest occupancies in the simulations without being recognized in the X-ray structure are 97Lys-93Asn and 100Ser-96Lys, where the chosen geometric criterion for counting a hydrogen bond is only slightly missed (donor–acceptor distances of 0.258 nm and 0.262 nm for the two hydrogen bonds). These residues are part of an α-helix running from residue 88Ile to residue 99Val. We note that MD simulation of the protein α-lactalbumin in aqueous solution using the same GROMOS96 force field does show the tendency to prefer α-helical structure over $3_{10}$-helical structure at pH = 2, but does not show this tendency at pH = 8 (Smith et al., 1999).

NOE distance bound violations are summarized in Table 3. The average violation for the analysis period, 300 ps–2000 ps, is lower for all chains in the crystal

*Table 2.* Occurrence of backbone–backbone hydrogen bonds in the X-ray structure, the crystal simulation, and the solution simulation

| Residue | | Percentage hydrogen bonding | | | | | |
|---|---|---|---|---|---|---|---|
| Donor | Acceptor | X-*ray* | *cry*1 | *cry*2 | *cry*3 | *cry*4 | *sol* |
| 3Phe | 38Phe | 100.0 | 91.0 | 91.4 | 90.6 | 86.2 | 70.5 |
| 8Leu | 4Gly | 100.0 | 30.9 | 94.4 | 97.9 | 98.2 | 97.1 |
| 9Ala | 5Arg | 100.0 | 3.4 | 93.8 | 84.4 | 92.3 | 90.8 |
| 10Ala | 6Cys | 100.0 | 91.5 | 96.5 | 96.6 | 97.5 | 98.7 |
| 11Ala | 7Glu | 100.0 | 64.4 | 86.8 | 89.9 | 89.2 | 85.3 |
| 12Met | 8Leu | 100.0 | 99.1 | 99.1 | 99.1 | 99.2 | 99.3 |
| 13Lys | 9Ala | 100.0 | 99.2 | 97.9 | 95.8 | 96.9 | 96.6 |
| 14Arg | 10Ala | 100.0 | 94.9 | 91.2 | 93.8 | 87.7 | 84.3 |
| 15His | 11Ala | 0 | 68.6 | 40.0 | 55.5 | 46.8 | 73.2 |
| 15His | 12Met | 0 | 18.1 | 36.9 | 27.1 | 36.1 | 11.6 |
| 16Gly | 12Met | 0 | 30.9 | 19.8 | 14.1 | 41.6 | 56.4 |
| 16Gly | 13Lys | 100.0 | 35.4 | 39.3 | 51.1 | 28.9 | 32.0 |
| 17Leu | 12Met | 100.0 | 25.9 | 25.1 | 53.3 | 15.1 | 33.4 |
| 20Tyr | 17Leu | 100.0 | 47.1 | 51.2 | 0 | 51.4 | 35.6 |
| 20Tyr | 23Tyr | 0 | 25.1 | 11.6 | 78.5 | 17.7 | 44.6 |
| 22Gly | 19Asn | 100.0 | 40.2 | 27.7 | 0 | 46.1 | 28.8 |
| 22Gly | 41Gln | 0 | 39.5 | 52.8 | 42.7 | 22.7 | 0 |
| 23Tyr | 2OTyr | 100.0 | 64.4 | 68.5 | 70.3 | 62.9 | 69.0 |
| 27Asn | 24Ser | 100.0 | 26.4 | 32.2 | 52.9 | 53.3 | 26.8 |
| 28Trp | 24Ser | 0 | 7.4 | 30.0 | 15.9 | 10.3 | 37.6 |
| 29Val | 25Leu | 100.0 | 95.1 | 89.2 | 89.2 | 84.9 | 94.8 |
| 30Cys | 26Gly | 100.0 | 98.4 | 95.2 | 99.1 | 98.8 | 98.3 |
| 31Ala | 27Asn | 100.0 | 97.2 | 96.9 | 98.0 | 96.7 | 97.5 |
| 32Ala | 28Trp | 100.0 | 93.4 | 97.6 | 88.6 | 92.7 | 96.2 |
| 33Lys | 29Val | 100.0 | 79.0 | 81.9 | 34.5 | 88.5 | 93.0 |
| 34Phe | 30Cys | 100.0 | 66.6 | 69.1 | 28.4 | 52.1 | 69.4 |
| 35Glu | 31Ala | 100.0 | 9.2 | 93.2 | 3.9 | 93.0 | 87.3 |
| 36Ser | 32Ala | 100.0 | 0 | 19.1 | 0 | 22.2 | 25.5 |
| 37Asn | 34Phe | 0 | 40.0 | 2.4 | 5.5 | 27.6 | 22.2 |
| 38Phe | 32Ala | 100.0 | 83.4 | 59.2 | 75.2 | 77.3 | 81.5 |
| 39Asn | 36Ser | 0 | 71.6 | 22.2 | 43.5 | 8.0 | 9.4 |
| 40Thr | 1Lys | 100.0 | 94.0 | 91.3 | 91.1 | 87.1 | 10.1 |
| 42Ala | 39Asn | 0 | 8.5 | 3.4 | 9.3 | 20.7 | 27.2 |
| 42Ala | 40Thr | 0 | 46.2 | 50.9 | 52.8 | 31.9 | 33.7 |
| 44Asn | 52Asp | 100.0 | 93.5 | 83.4 | 89.5 | 51.6 | 89.2 |
| 46Asn | 50Ser | 100.0 | 75.0 | 65.9 | 0 | 54.2 | 69.8 |
| 49Gly | 46Asn | 100.0 | 67.4 | 76.9 | 0 | 72.5 | 80.2 |
| 52Asp | 44Asn | 100.0 | 75.3 | 92.9 | 85.9 | 90.8 | 74.8 |
| 53Tyr | 58Ile | 100.0 | 97.4 | 96.8 | 94.2 | 96.2 | 95.8 |
| 54Gly | 42Ala | 100.0 | 90.7 | 78.2 | 82.1 | 42.2 | 68.4 |
| 57Gln | 54Gly | 100.0 | 78.6 | 0 | 31.6 | 10.1 | 84.0 |
| 58Ile | 53Tyr | 100.0 | 98.7 | 95.4 | 92.6 | 79.4 | 98.4 |
| 60Ser | 61Thr | 100.0 | 87.2 | 81.9 | 89.9 | 86.6 | 92.8 |
| 63Trp | 59Asn | 100.0 | 72.1 | 75.4 | 52.1 | 83.5 | 84.5 |
| 64Cys | 59Asn | 0 | 65.8 | 65.8 | 18.5 | 29.5 | 17.5 |
| 64Cys | 60Ser | 0 | 10.3 | 9.0 | 11.8 | 23.6 | 43.4 |
| 65Asn | 78Ile | 100.0 | 83.2 | 92.9 | 81.2 | 77.0 | 80.0 |
| 73Arg | 61Arg | 100.0 | 82.6 | 77.3 | 76.8 | 79.4 | 90.3 |
| 74Asn | 61Arg | 0 | 51.5 | 52.8 | 0 | 43.4 | 0 |
| 75Leu | 62Trp | 100.0 | 91.7 | 97.4 | 95.9 | 91.3 | 91.8 |

*Table 2.* Continued

| Residue | | Percentage hydrogen bonding | | | | | |
|---|---|---|---|---|---|---|---|
| Donor | Acceptor | X-*ray* | *cry*1 | *cry*2 | *cry*3 | *cry*4 | *sol* |
| 76Cys | 63Trp | 100.0 | 75.2 | 64.4 | 81.7 | 90.1 | 91.5 |
| 77Asn | 74Asn | 100.0 | 56.4 | 58.8 | 53.1 | 41.3 | 72.7 |
| 78Ile | 76Cys | 0 | 29.9 | 83.8 | 74.0 | 73.1 | 6.2 |
| 80Cys | 65Asn | 100.0 | 94.8 | 84.5 | 93.5 | 90.9 | 93.6 |
| 82Ala | 79Pro | 100.0 | 17.4 | 46.0 | 32.8 | 34.9 | 6.5 |
| 83Leu | 79Pro | 0 | 21.4 | 15.7 | 9.1 | 3.8 | 80.8 |
| 83Leu | 80Cys | 100.0 | 21.1 | 40.2 | 27.6 | 35.8 | 2.6 |
| 84Leu | 81Ser | 100.0 | 5.6 | 11.6 | 12.0 | 19.4 | 5.4 |
| 84Leu | 80Cys | 0 | 78.8 | 32.4 | 56.9 | 46.4 | 82.6 |
| 85Ser | 81Ser | 0 | 43.1 | 51.9 | 38.5 | 44.8 | 39.3 |
| 85Ser | 82Ala | 0 | 18.5 | 21.8 | 23.9 | 25.9 | 30.2 |
| 89Thr | 87Asp | 0 | 0 | 0 | 2.8 | 0 | 25.6 |
| 92Val | 88Ile | 100.0 | 93.1 | 60.1 | 95.8 | 88.2 | 3.6 |
| 93Asn | 89Thr | 100.0 | 96.9 | 91.1 | 98.2 | 92.8 | 95.8 |
| 94Cys | 90Ala | 100.0 | 97.8 | 96.9 | 98.7 | 98.4 | 98.3 |
| 95Ala | 91Ser | 100.0 | 98.5 | 95.2 | 98.8 | 96.7 | 98.1 |
| 96Lys | 92Val | 100.0 | 94.9 | 97.1 | 95.1 | 97.6 | 98.1 |
| 97Lys | 93Asn | 0 | 69.8 | 77.6 | 82.2 | 80.0 | 80.6 |
| 98Ile | 94Cys | 100.0 | 93.8 | 90.4 | 93.4 | 93.6 | 95.2 |
| 99Val | 95Ala | 100.0 | 91.1 | 93.7 | 94.7 | 96.5 | 69.2 |
| 100Ser | 96Lys | 0 | 95.6 | 91.2 | 89.8 | 91.2 | 93.8 |
| 101Asp | 98Ile | 100.0 | 24.6 | 42.1 | 25.5 | 29.9 | 5.2 |
| 101Asp | 97Lys | 0 | 66.2 | 41.7 | 52.7 | 58.8 | 68.5 |
| 104Gly | 98Ile | 0 | 0 | 0 | 0 | 0 | 34.9 |
| 106Asn | 103Asn | 100.0 | 69.3 | 17.5 | 0 | 6.6 | 55.1 |
| 107Ala | 104Gly | 0 | 17.6 | 79.4 | 42.7 | 40.1 | 7.9 |
| 108Trp | 105Met | 100.0 | 80.5 | 82.9 | 67.2 | 79.5 | 81.4 |
| 112Arg | 108Trp | 100.0 | 78.8 | 84.0 | 29.9 | 74.5 | 71.6 |
| 113Asn | 109Val | 100.0 | 0 | 92.9 | 13.6 | 93.6 | 82.1 |
| 114Arg | 110Ala | 100.0 | 0 | 26.6 | 40.9 | 72.7 | 65.5 |
| 115Cys | 110Ala | 0 | 60.1 | 89.6 | 18.6 | 89.6 | 87.2 |
| 115Cys | 111Trp | 100.0 | 26.8 | 7.1 | 62.9 | 5.8 | 9.2 |
| 116Lys | 111Trp | 100.0 | 43.7 | 95.6 | 85.3 | 96.5 | 79.7 |
| 118Thr | 115Cys | 100.0 | 13.4 | 19.4 | 13.3 | 16.2 | 14.3 |
| 122Ala | 119Asp | 100.0 | 3.4 | 0 | 2.9 | 7.1 | 4.7 |
| 123Trp | 119Asp | 0 | 79.9 | 26.3 | 3.2 | 69.9 | 39.3 |
| 123Trp | 120Val | 100.0 | 7.6 | 27.0 | 49.5 | 4.4 | 11.1 |
| 124Ile | 120Val | 0 | 92.9 | 40.2 | 3.6 | 63.3 | 67.4 |
| 124Ile | 121Gln | 100.0 | 0 | 19.2 | 50.7 | 10.6 | 14.5 |
| 125Arg | 121Gln | 0 | 0 | 16.4 | 73.7 | 5.6 | 12.1 |
| 125Arg | 122Ala | 0 | 4.1 | 17.7 | 17.5 | 36.4 | 25.1 |
| 125Arg | 123Trp | 0 | 17.4 | 21.3 | 0 | 8.3 | 24.5 |
| 127Cys | 124Ile | 100.0 | 12.5 | 28.2 | 17.2 | 18.1 | 6.0 |

Intramolecular hydrogen bonds are listed if they are present in the X-ray crystal structure (*X-ray*), if they are present for more than 20% of the simulation time in the solution simulation (*sol*), or if they are present for more than 20% simulation time averaged over the four molecules (*cry*1–4) in the crystal simulation. A hydrogen bond is counted if the distance between the hydrogen atom and the acceptor atom is maximally 0.25 nm and the angle between donor atom, hydrogen atom and acceptor atom is larger than 135°.

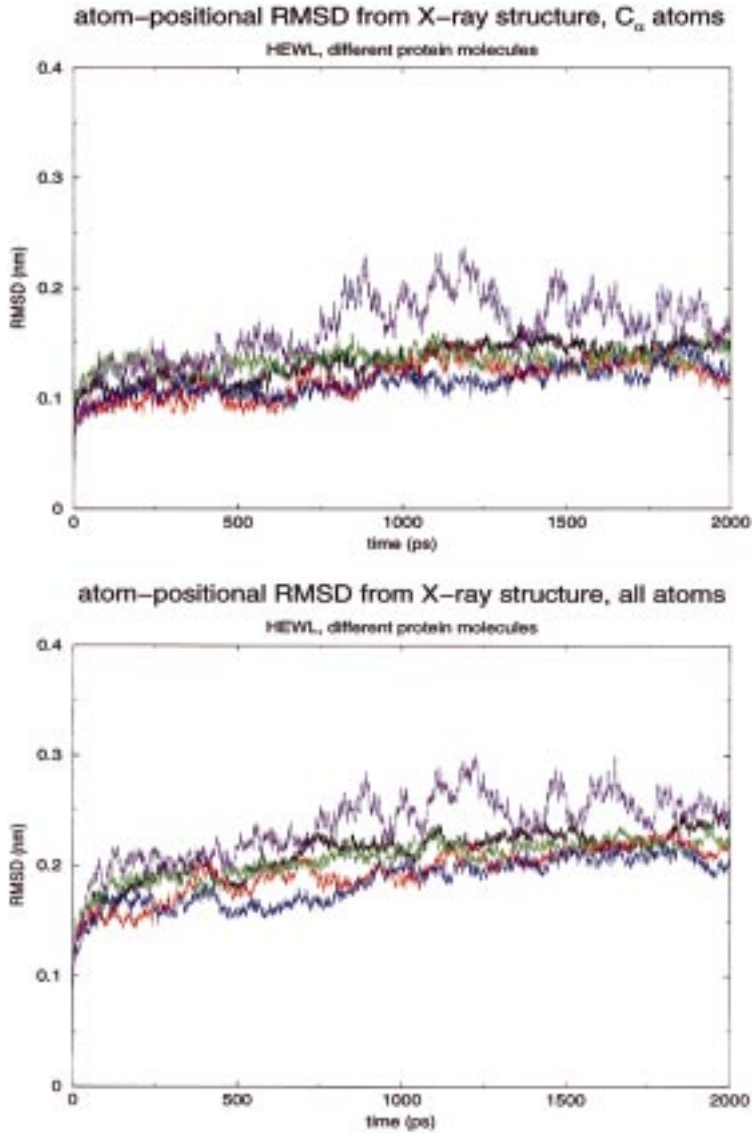


*Figure 1.* Root-mean-square atom positional deviation (RMSD) in nm from the X-ray (crystal) structure (Carter et al., 1997) as a function of time in ps. Rotational and translational fitting was applied using all 129 $C_\alpha$ atoms. RMSD of $C_\alpha$ atoms (top graph) and of all atoms (bottom graph) are shown. RMSD of the solution simulation are shown in magenta, of the four chains in the crystal simulations in black (molecule 1), red (molecule 2), green (molecule 3), and blue (molecule 4).

simulation compared to the solution simulation. Also, the number of large violations is always lower in the crystal simulation, reflecting the larger backbone fluctuations in the solution simulation. The X-ray crystal structure satisfies the NOE bounds even better.

The $95\,^3J_{H\,N\alpha}$-coupling constants (Figure 3) are slightly better reproduced in the solution simulation (RMSD of calculated versus experimentally derived J-coupling constants are 1.70 Hz in the solution simulation and 1.72 Hz, 1.77 Hz, 1.88 Hz, and 1.64 Hz for the four chains in the crystal simulation). The higher mobility in solution allows for a larger part of phase space being sampled and thus, averaging may be better. The X-ray crystal structure reproduces the experimental J-coupling constants well with an RMSD of 0.88 Hz, as is seen in the lower right corner of Figure 3 where J-coupling constants calculated from the X-ray structure are compared with J-coupling constants obtained from NMR experiments. Calculated J-coupling constants are generally slightly lower than experimentally

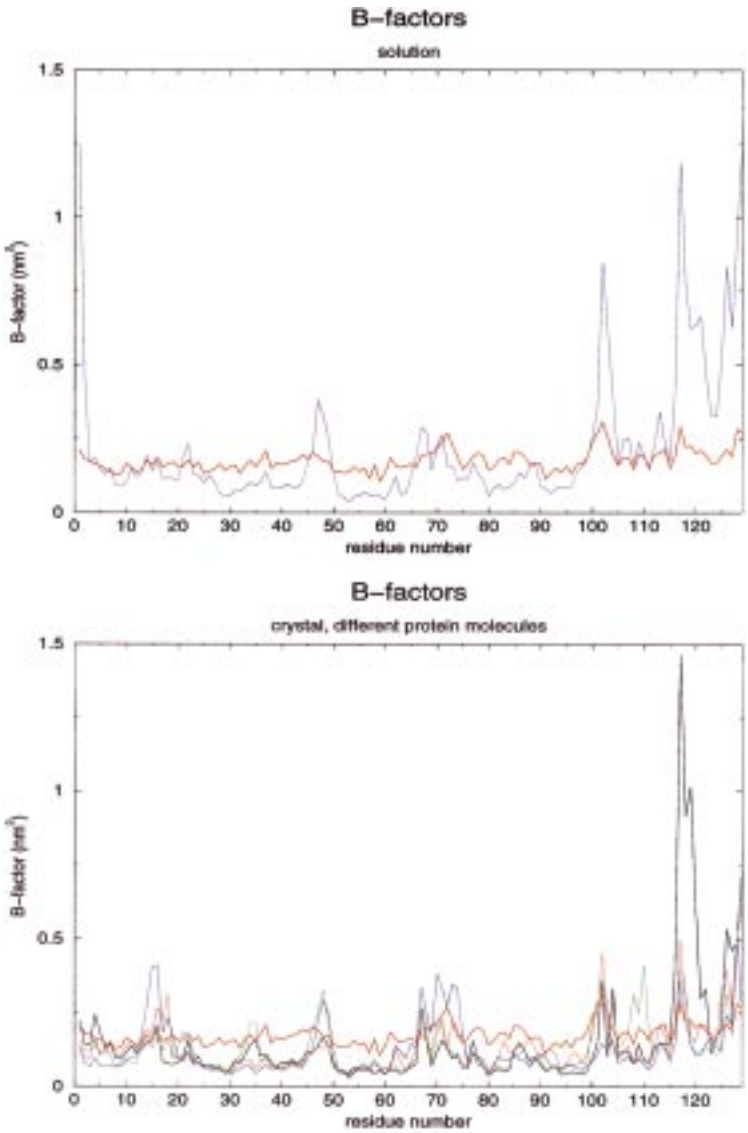


*Figure 2.* Isotropic B-factors for backbone $C_\alpha$-atoms in $mm^2$. In the top graph, B-factors calculated from the solution simulation (magenta) are shown together with experimentally derived values (orange). In the bottom graph, B-factors for the different protein molecules calculated from the crystal simulation are shown in black (molecule 1), red (molecule 2), green (molecule 3), and blue (molecule 4) together with experimentally derived values (orange).

derived ones but longer simulation times could bring these values closer to each other.

Backbone order parameters, displayed in Figure 4, are equally well reproduced in both simulations. The secondary structure elements correspond to large calculated $S^2$ values. Order parameters calculated from the simulation are generally slightly lower compared to those determined using NMR data. In the crystal simulation, different chains generally have very similar order parameters with a handful of exceptions. This happens if one protein molecule has sampled a particular movement, while another has not. This effect is even stronger for side-chain order parameters (Figure 5). Slow motions observed only once in a simulation lead to different order parameters for the same residue in the four protein molecules. The crystal simulation shows significantly larger side-chain order parameters than experimentally derived ones. Its order parameters are also larger than in the solutions simulation, which is expected, since side chains involved in

*Table 3.* Number of NOE distance bound violations larger than a given value computed from the X-ray structure, the crystal simulation, and the solution simulation

| Protein molecule | NOE bound violations | | | |
|---|---|---|---|---|
| | >0.05 nm | >0.1 nm | >0.3 nm | mean $\langle R_E - R_0 \rangle$ (nm) |
| *X-ray* | 45 | 27 | 2 | 0.006 |
| *cry*1 | 65 | 37 | 6 | 0.009 |
| *cry*2 | 57 | 35 | 3 | 0.008 |
| *cry*3 | 67 | 39 | 7 | 0.009 |
| *cry*4 | 56 | 33 | 4 | 0.007 |
| *sol* | 75 | 48 | 9 | 0.010 |

The total of number of experimentally determined NOE bounds $R_0$ is 1158 (Smith et al., 1991, 1993; Buck et al., 1995). $R_E$ is the distance from the trajectory using $r^{-3}$ averaging over 200 ps windows throughout the whole 300-2000 ps analysis period. The symbol $R_E$ -$R_0$ indicates a (mean) violation, calculated as a mean of the larger value of $R_E$-$R_0$ and zero. Average violations are in nm. The X-ray structure is indicated by the symbol *X-ray*. The four molecules in the crystal unit cell are indicated by the symbol *cry*1–4, the one molecule in the solution simulation by the symbol *sol*. NOE distance bound violations of the X-ray crystal structure are denoted by the symbol *X-ray*.

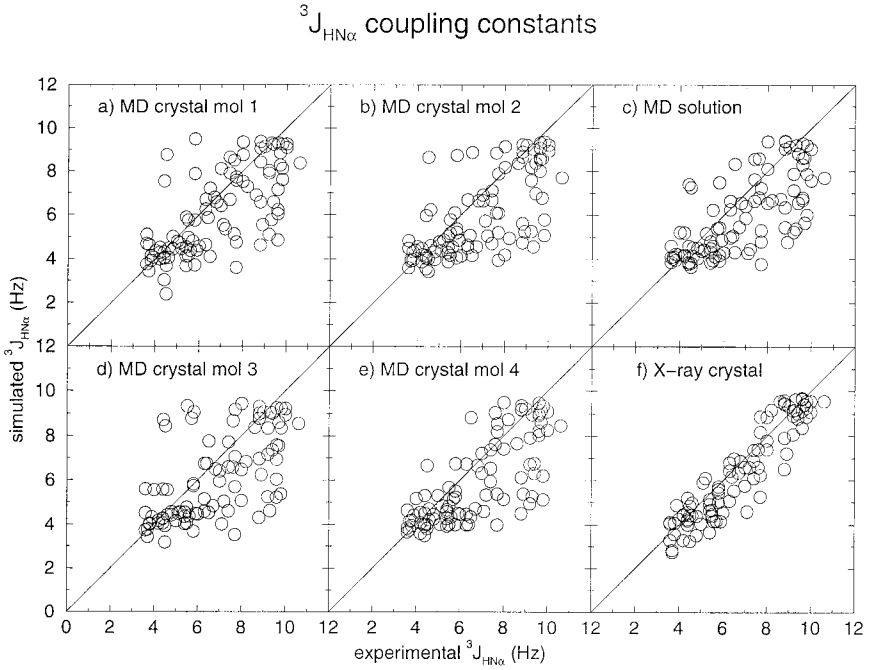


*Figure 3.* Comparison of 95 experimental (Smith et al., 1991) and calculated $^3J_{H\,N\alpha}$ coupling constants (in Hz). In graphs a, b, d, and e, $^3J_{H\,N\alpha}$ coupling constants calculated from the crystal simulation are displayed for the four different protein molecules in the unit cell. In graph c, values calculated from the solution simulation are shown. Graph f shows J-values calculated from the experimentally determined X-ray crystal structure.

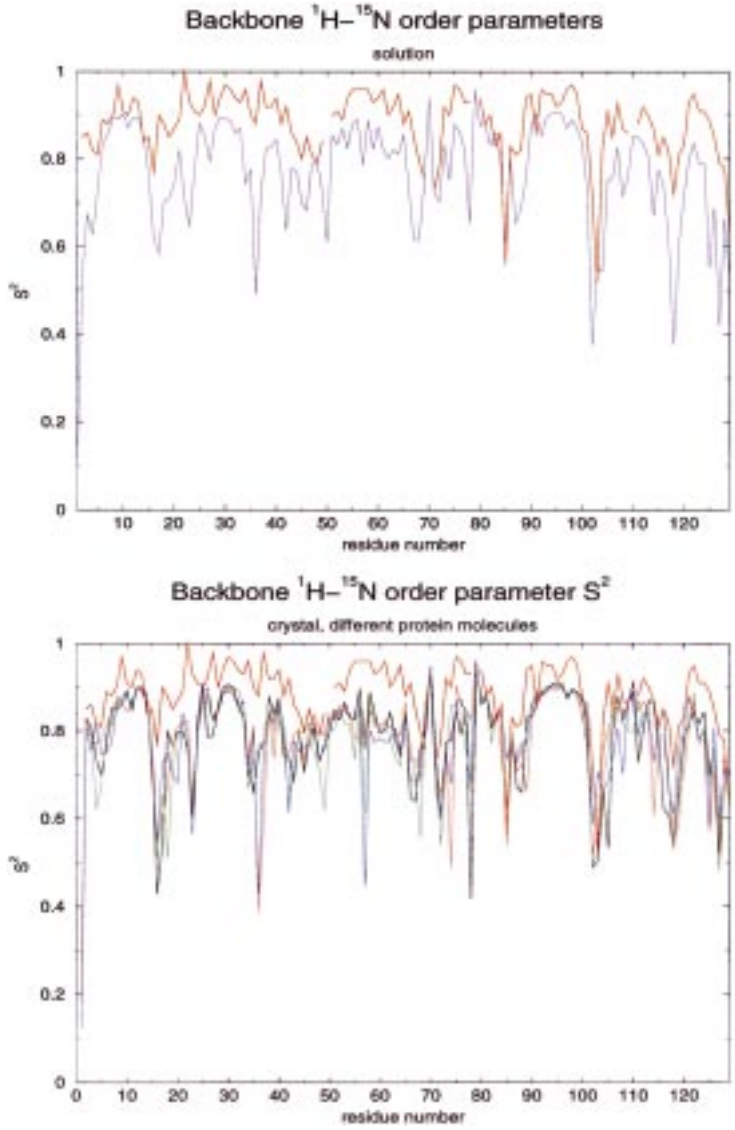


*Figure 4.* Backbone $^{1}$H-$^{15}$N order parameters ($S^2$) as a function of residue number. The order parameters were calculated using a 200 ps averaging window moving through the whole 1700 ps analysis period. In the top graph, order parameters calculated from the solution simulation (magenta) are shown together with experimentally derived values (orange). In the bottom graph, order parameters for the four different protein molecules calculated from the crystal simulation are shown in black (molecule 1), red (molecule 2), green (molecule 3), and blue (molecule 4) together with experimentally derived values (orange). There is no experimental value available for residues Lys1, Ser50, Pro70, Pro79, and Ala110.

crystal packing contacts will be more rigid than in solution. Order parameters calculated from the solution simulation correspond very well to experimentally derived ones.

## Conclusions

The two 2 ns molecular dynamics simulations of hen egg white lysozyme (HEWL), one in solution, the other in crystalline environment, give very similar results, slight differences are, however, present. Not unexpectedly, the four protein molecules in the crystal simulation stay closer to the X-ray structure. They also fulfill atom–atom distance bounds derived from

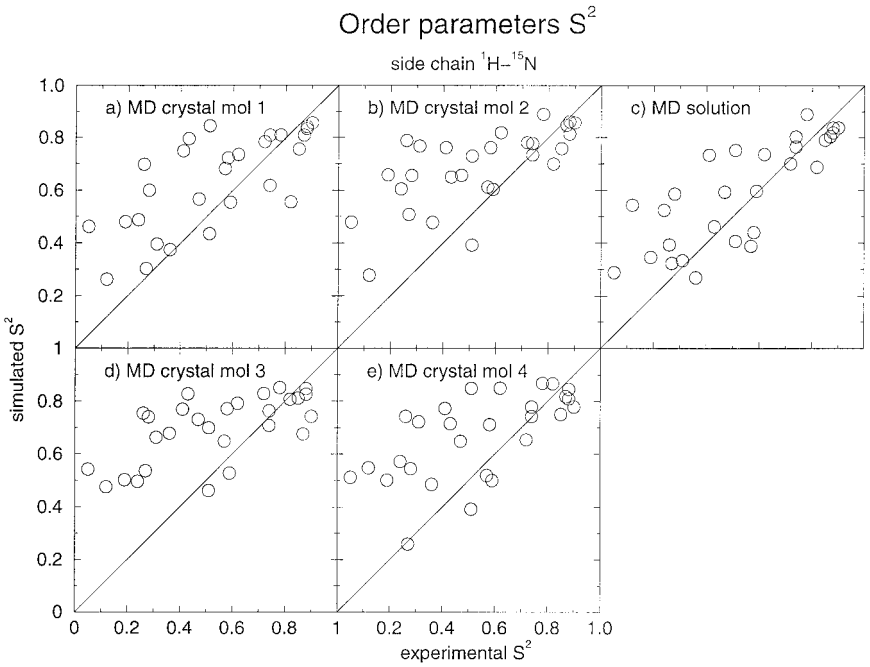


*Figure 5.* Comparison of 28 experimental (Buck et al., 1995) $^{1}$H-$^{15}$N order parameters $S^2$ of side-chain NH groups with calculated values. The order parameters were calculated using a 200 ps averaging window moving through the whole 1700 ps analysis period. In graphs a, b, d, and e, $^{3}J_{H\,N\alpha}$ coupling constants calculated from the crystal simulation are displayed for the four different protein molecules in the unit cell. In graph c, values calculated from the solution simulation are shown. For $NH_2$ groups the average of the order parameters for the two NH vectors is displayed.

NMR solution data significantly better than the solution simulation. B-factors, representing the magnitude of atom positional fluctuations and order parameters $S^2$ which are determined by orientational mobility of N-H bond vectors, show the same degree of agreement with experimentally derived date for both simulations. The root-mean-square deviation from the X-ray structure shows larger fluctuations for the solution simulation, indicating enhanced sampling of configuration space and perhaps better converged J-coupling constants than in the crystal simulation. The backbone hydrogen bonding network is very well maintained in both simulations. Occupancies of intramolecular hydrogen bonds are even higher in the solution simulation, probably because in the crystal, the higher atom density leads to less flexibility: the protein in solution is less restrained than in the crystal. The four different molecules in the crystal show highly similar properties, especially in stable regions of secondary structure. In the more mobile regions of the molecule, a rare event, such as a jump over a rotational barrier of a dihedral angle, can lead to different observations between single molecules. With increasing simulation length, more of these events would be observed, and the same events could also happen in another protein molecule. Longer simulations would thus enhance the similarity between the different molecules in the crystal simulation.

The present results for hen egg white lysozyme confirm that proteins in crystalline environment and in solution show very similar behaviour. Fluctuations are somewhat larger in solution, structural properties are almost identical.

## Acknowledgements

The authors wish to thank Dr. Lorna J. Smith for making data available and useful discussions and Prof. Dr. Alan E. Mark for fruitful discussions. Financial support was obtained from the Schweizerischer Nationalfonds, project number 21-41875.94, which is gratefully acknowledged.

## References

Artymiuk, P.J., Blake, C.C.F., Rice, D.W. and Wilson, K.S. (1982) *Acta Crystallogr.*, B**38**, 778–783.

Baldwin, E.T., Weber, I.T., Charles, R.S., Xuan, J.-C., Appella, E., Yamada, M., Matsushima, K., Edwards, B.F.P., Clore, G.M., Gronenborn, A.M. and Wlodawer, A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 502–506.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J. (1981) In *Intermolecular Forces* (Ed., Pullman, B.), Reidel, Dordrecht, pp. 331–342.

Berndt, K.D., Güntert, P., Orbons, L.P.M. and Wüthrich, K. (1992) *J. Mol. Biol.*, **227**, 757–775.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Billeter, M., Kline, A.D., Braun, W., Huber, R. and Wüthrich, K. (1989) *J. Mol. Biol.*, **206**, 677–687.

Billeter, M., Vendrell, J., Wider, G., Avilès, F.X., Coll, M., Gausch, A., Huber, R., and Wüthrich, K. (1992) *J. Biomol. NMR*, **2**, 1–10.

Braun, W., Vašák, M., Robbins, A.H., Stout, C.D., Wagner, G., Kägi, J.H.R., and Wüthrich, K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10124–10128.

Buck, M., Boyd, J., Redfield, C., MacKenzie, D.A., Jeenes, D.J., Archer, D.B. and Dobson, C.M. (1995) *Biochemistry*, **34**, 4041–4055.

Carter, D., He, J., Ruble, J.R. and Wright, B. (1997) Protein Data Bank, entry 1AKI.

Dornberger, U., Flemming, J. and Fritzsche, H. (1998) *J. Mol. Biol.*, **284**, 1453–1463.

Evenäs, J. Forsén, S. and Akke, M. (1999) *J. Mol. Biol.*, **289**, 603–617.

Fede, A., Billeter, M., Leupin, W. and Wüthrich, K. (1993) *Structure*, **1**, 177–186.

Hünenberger, P.H., Mark, A.E. and van Gunsteren, W.F. (1995) *J. Mol. Biol.*, **252**, 492–503.

Kallen, J., Spitzfaden, C., Zurini, M.G.M., Wider, G., Widmer, H., Wüthrich, K. and Walkinshaw, M.D. (1991) *Nature*, **353**, 276–279.

Lu, J., Lin, C.-L., Tang, C., Ponder, J.W., Kao, J.L.F., Cistola, D.P., and Li, E. (1999) *J. Mol. Biol.*, **286**, 1179–1195.

Moore, J.M., Lepre, C.A., Gibbert, G.P., Chazin, W.J., Case, D.A. and Wright, P.E. (1991) *J. Mol. Biol.*, **221**, 533–555.

Neri, D., Billeter, M. and Wüthrich, K. (1992) *J. Mol. Biol.*, **223**, 743–767.

Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) *J. Comput. Phys.*, **23**, 327–341.

Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. and van Gunsteren, W.F. (1999) *J. Phys. Chem.*, **A103**, 3596–3607.

Smith, L.J., Dobson, C.M. and van Gunsteren, W.F. (1999) *Proteins*, **36**, 77–86.

Smith, L.J., Mark, A.E., Dobson, C.M. and van Gunsteren W.F. (1995) *Biochemistry*, **34**, 10918–10931.

Smith, L.J., Sutcliffe, M.J., Redfield, C. and Dobson, C.M. (1991) *Biochemistry*, **30**, 986–996.

Smith, L.J., Sutcliffe, M.J., Redfield, C. and Dobson, C.M. (1993) *J. Mol. Biol.*, **229**, 930–944.

Smith, P.E. and van Gunsteren, W.F. (1994) *J. Chem. Phys.*, **100**, 3169–3174.

Tironi, I.G., Sperb, R., Smith, P.E. and van Gunsteren, W.F. (1995) *J. Chem, Phys.*, **102**, 5451–5459.

van Gunsteren, W.F. and Berendsen, H.J.C. (1984) *J. Mol. Biol.*, **176**, 559–564.

van Gunsteren, W.F. and Berendsen, H.J.C. (1990) *Angew. Chem. Int. Ed. Engl.*, **29**, 992–1023.

van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G. (1996) Vdf Hochschulverlag, Zürich, Switzerland.

Yang, F., Bewley, C.A., Louis, J.M., Gustafson, K.R., Boyd, M.R., Gronenborn, A.M., Clore, G.M. and Wlodawer, A. (1999) *J. Mol. Biol.*, **288**, 403–412.