

Comparison of MD Simulations and NMR Experiments for Hen Lysozyme. Analysis of Local Fluctuations, Cooperative Motions, and Global Changes[†]

Lorna J. Smith,[‡] Alan E. Mark,[§] Christopher M. Dobson,^{*,‡} and Wilfred F. van Gunsteren^{*,§}

Oxford Centre for Molecular Sciences and New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, England, and Department of Physical Chemistry, Swiss Federal Institute of Technology, Zürich, ETH-Zentrum, 8092 Zürich, Switzerland

Received March 7, 1995; Revised Manuscript Received June 12, 1995[®]

ABSTRACT: Three 1000 ps molecular dynamics simulations of hen lysozyme have been compared with a range of experimental NMR parameters in order to gain insight into the dynamical properties of the protein and to assess the significance of the motional events observed in the simulations. The simulations, one in vacuum and two in water, were used to estimate interproton distances (for comparison with NOE data), $^3J_{\text{HN}\alpha}$ and $^3J_{\alpha\beta}$ coupling constants and ^1H – ^{15}N order parameters. Comparison of these values with experimental data, particularly NOEs, enabled force field-induced changes to the structure during the simulations to be recognized. It has been shown, however, that these changes can be largely eliminated by slight modifications to the force field. Using a simulation performed in water with this modified force field, it has been found that ^1H – ^{15}N order parameters calculated for side chain groups in particular correlate well with experimental values and reflect the substantial dependence of these motional properties on the environment, particularly surface exposure, in which the side chain is found. In this case, the simulation then provides models for the motional processes giving rise to the observed experimental data. The results indicate that the order parameter values reflect primarily the number of torsion angles about which rotameric interchange occurs. In addition to local motions, the two different domains of lysozyme have been found to behave differently in the simulations. Possible implications of these differences for the interpretation of unfolding simulations and experimental observations of folding intermediates for lysozyme are discussed.

Proteins in solution undergo a wide variety of motions ranging from local atomic fluctuations and bond oscillations to hinge bending motions, helix-coil transitions, and local and global unfolding processes (Karplus & MacCammon, 1981; Brooks et al., 1988; Dobson, 1993). Both experimental and theoretical techniques have been used extensively in attempts to characterize these dynamic fluctuations and to understand their role in protein function, specificity, and folding (Smith & Dobson, 1995; van Gunsteren et al., 1995a). Nuclear magnetic resonance (NMR)¹ spectroscopy in particular has been able to provide experimental evidence for a wide range of dynamical events in proteins. For example, ^{15}N and ^{13}C relaxation studies enable the extent and time scales of the mobility of interatomic vectors to be probed (Wagner, 1993), while the observation of exchange effects or averaging of NMR parameters such as chemical shifts and coupling constants indicates that there is interconversion between different conformational states (Nagayama & Wüthrich, 1981; Campbell et al., 1985). Although such mobility can be readily recognized using NMR, this technique cannot provide directly a description at an atomic level

for the exact molecular events that give rise to the observed experimental phenomena. This is, first, because of the averaging inherent in the measurements and, second, because of the complex and often highly coupled nature of protein motions (Dobson & Karplus, 1986).

Theoretical molecular dynamics (MD) simulations, by contrast, readily provide a detailed atomistic model for the dynamical processes in proteins (Brooks et al., 1988; van Gunsteren et al., 1995b). How realistic these models are, and how they relate to experimental data, are, however, questions of considerable debate. One approach to assess the reliability of the models used is to compare experimental NMR data with NMR parameters predicted from molecular dynamics simulations. For example, NMR parameters calculated from early simulations of BPTI and lysozyme in vacuum have been compared with experimental NMR coupling constants, ring current shifts, nuclear Overhauser enhancements (NOEs), and ^1H and ^{13}C relaxation data (Hoch et al., 1982; Lipari et al., 1982; Olejniczak et al., 1984; Dobson & Karplus, 1986). Although the lengths of these simulations were relatively short, the comparisons indicated the considerable insight that could be obtained from combining experimental and theoretical techniques in this way. More recently, simulations of interleukin- 1β (Chandrasekhar et al., 1992), glucocorticoid receptor DNA binding domain (Eriksson et al., 1993), and calbindin $\text{D}_{9\text{k}}$ (Kördel & Teleman, 1992) have been compared with ^1H – ^{15}N order parameters, a simulation of the zinc finger peptide Xfin 31 has been compared with ^{13}C relaxation data (Palmer & Case, 1992) and simulations of BPTI have been compared with ^{15}N and

[†] The Oxford Centre for Molecular Sciences is supported by the U.K. Engineering and Physical Sciences Research Council, the Biotechnology and Biological Sciences Research Council and the Medical Research Council.

* To whom correspondence should be addressed.

[‡] University of Oxford.

[§] Swiss Federal Institute of Technology.

[®] Abstract published in *Advance ACS Abstracts*, August 1, 1995.

¹ Abbreviations: NMR, nuclear magnetic resonance; MD, molecular dynamics; NOE, nuclear Overhauser enhancement; S^2 , order parameter; T_1 , spin-lattice relaxation time; T_2 , spin-spin relaxation time; RMSD, root-mean-square difference.

^{13}C T_1 and T_2 relaxation times and $^3J_{\text{HN}\alpha}$ coupling constants (Balasubramanian et al., 1994; Smith et al., 1995). In general, the comparisons with the relaxation data have shown that the fast time-scale fluctuations that result in high-order parameter values can be modeled by the simulations, but the length of current simulations is too short to sample correctly longer time-scale motions such as those that affect the T_1 and T_2 relaxation rates.

The present work extends the scope of previous studies by the analysis of three 1000 ps MD simulations of hen lysozyme. One of these was performed in vacuum and the other two in water, but with different force fields. Predictions from the simulations have been compared with a variety of experimental NMR parameters for hen lysozyme which probe different features of the simulation. The major aim of the study was to gain insight into the nature of the motions occurring in solution that influence the NMR data. A second important objective was to assess the extent to which various NMR parameters can be predicted adequately by an 1000 ps simulation, which is the current practical time limit for simulations of solvated protein systems (van Gunsteren et al., 1995b). It was also hoped that, through detailed comparisons between the calculated and experimental NMR parameters, structural changes and dynamical events occurring as a result of limitations in the force field used in the simulations could be identified and, if possible, rectified. Hen lysozyme is particularly appropriate for a comparison of this type, being a small globular protein that contains both helices and regions of β sheet, together with loop regions, turns, and disulfide bridges. Its structure, dynamics, and folding have already been studied extensively by a wide range of experimental and theoretical techniques (Blake et al., 1965; Pedersen et al., 1991; Mark & van Gunsteren, 1992; Radford et al., 1992; Smith et al., 1993; Dobson et al., 1994; Buck et al., 1995). Its structure can be divided into two domains, one of which (the α domain; residues 1–35 and 85–129) contains four α helices and a short 3_{10} helix. The other domain (the β domain; residues 36–84) comprises a triple-stranded antiparallel β sheet, a long loop, and a 3_{10} helix.

Three different sets of experimental NMR data were compared with parameters calculated from the simulations discussed in this work, ^1H NOEs (Smith et al., 1993), spin-spin coupling constants ($^3J_{\text{HN}\alpha}$ and $^3J_{\alpha\beta}$) (Smith et al., 1991), and ^{15}N relaxation data (Buck et al., 1995), each of which is affected differently by the motional processes occurring in proteins. ^1H NOE effects are used extensively to give interproton distance restraints for protein structure determinations by NMR (Clare & Gronenborn, 1991). The initial buildup rate of a NOE between two protons i and j is proportional to the cross-relaxation rate, σ_{ij} , between these protons. For a rigid isotropically tumbling molecule, assuming the two-spin approximation, σ_{ij} is inversely related to the sixth power of the interproton distance, r_{ij} , and directly related to the correlation time, τ_c . This dependence on interproton distance provides the basis for converting NOE intensities into distance restraints (Kumar et al., 1981; Wüthrich, 1986; Wagner, 1990). However, mobility will affect the NOE data both because it can give different effective correlation times for different interproton vectors in the protein and because it leads to fluctuations in the interproton distances (Dobson & Karplus, 1986; Kaptein et al., 1988; Post, 1992). In this work, the interproton distance restraints derived from the experimental NOEs were compared with the interproton distances in the trajectory. Only

the influence of dynamics on the interproton distances was therefore considered and not the effects of internal motions on the correlation times.

The $^3J_{\text{HN}\alpha}$ and $^3J_{\alpha\beta}$ coupling constants probe the ϕ and χ_1 torsion angles respectively. If conformations with different torsion angles are being populated, and if these give rise to different coupling constant values J_a and J_b , a population-weighted average coupling constant will be observed as long as the rate of interconversion between the conformers is significantly greater than $|J_a - J_b|$ (Jardetzky & Roberts, 1981). This means in general that an averaged coupling constant is observed for conformers that are interconverting at a rate that is greater than approximately 20 Hz (Smith et al., 1991).

^{15}N relaxation methods probe the motions of ^1H – ^{15}N bonds in proteins, i.e., the backbone amide NH group, and the side chain NH groups of Asn, Gln, Trp, and Arg (Wagner, 1993). The experimental data, ^{15}N T_1 and T_2 values, and ^1H – ^{15}N NOE intensities, are generally analyzed using the model-free approach of Lipari and Szabo (1982). This is based on the assumption that the time scale of the internal motions of the ^1H – ^{15}N vector is very different from that of the overall rotation of the molecule and means that the time correlation function describing the orientation of the ^1H – ^{15}N vector, $C(t)$, can be factorized into contributions from the internal motions and from the overall tumbling of the molecule. With this assumption the spectral density function, and hence the expressions for T_1 , T_2 , and NOE, can be written as a function of an order parameter, S^2 , and a correlation time, τ_e , for internal motions. Values for S^2 and τ_e can then be obtained from a fit of the experimental data to these expressions. For certain residues more complex models are required, incorporating, for example, two motions faster than the overall rotation correlation time τ_R or an exchange term to take into account motions on a time scale much slower than τ_R , or considering anisotropic motions (Clare et al., 1990). The order parameters are usually interpreted as giving a measure of the degree of spatial restriction of the ^1H – ^{15}N vector on a time scale much faster than the rotational tumbling time of the molecule, i.e., in general on a picosecond time scale (τ_R for hen lysozyme in H_2O is about 5.7 ns at 308 K). Very restricted motion gives an order parameter close to unity, while completely free motion, where all orientations of the vector are equally probable, gives a value of zero (Lipari & Szabo, 1982).

METHODS

(a) *Simulations.* Three independent simulations of hen lysozyme, one performed in vacuum (V1) and two carried out using explicit solvent water molecules and periodic boundary conditions (W1 and W2) were analyzed in this work. Simulations and analysis were performed using the GROMOS package of programs (van Gunsteren & Berendsen, 1987). The parameters used in each of the simulations are summarized in Table 1. The starting coordinates in each case were taken from the crystal structure of triclinic hen egg white lysozyme (Ramanadham et al., 1987), entry 2LZT of the protein data bank (Bernstein et al., 1977). No crystallographic water molecules or counterions were included in the simulations. In each simulation the temperature was maintained at 300 K by weak coupling to an external bath (Berendsen et al., 1984). The first water simulation, W1, was performed at constant volume, and the second

Table 1: Summary of Simulation Parameters

	vacuum (V1)	water 1 (W1)	water 2 (W2)
starting structure	2LZT	2LZT	2LZT
simulation length (ps)	1000	1000	1100
solute atoms	1264	1264	1321
solvent water molecules	0	5345	4463
periodic boundary conditions		rectangular	truncated octahedral
box dimensions (nm)		$x = 4.88$ $y = 5.31$ $z = 6.81$	$x = y = z = 6.71$
short-range cutoff (nm)	0.8	0.8	0.8
long-range electrostatic cutoff (nm)	1.4	1.4	1.4
pair list update (fs)	10	10 (20 for first 300 ps)	10
time step (fs)	2	2	2
temperature coupling relaxation time (ps)	0.1	0.1	0.1
pressure coupling relaxation time (ps)			0.5
force field	37D4	37C4	37C4 + aromatic H+ modified water – oxygen to carbon interaction

simulation, W2, at constant pressure (1 atm) again by weak coupling to an external bath. The dimensions of the periodic box (Table 1) were chosen such that no protein atom in the starting structure lies within 0.95 nm of the box wall. Treating all nonpolar hydrogens by using a united atom approach results in a total of 1264 protein atoms, a number increasing to 1321 atoms when aromatic hydrogens are also treated explicitly. The two simulations in water included 5345 and 4463 simple point charge (SPC) water molecules (Berendsen et al., 1981) which led to a total system size of 17 299 and 14 710 atoms, respectively. The smaller system size in the second simulation in water was achieved by using the same distance criteria to the box wall, but choosing truncated octahedral rather than rectangular periodic boundary conditions. Bond lengths were constrained to equilibrium values during the simulations using the SHAKE procedure (Ryckaert et al., 1977). Nonbonded interactions were treated using a twin range method (van Gunsteren & Berendsen, 1990). Within the short-range cutoff of 0.8 nm, all interactions were determined at every time step. Longer range electrostatic interactions within a cutoff range of 1.4 nm were updated at the same time that the pair list was generated. During the first 300 ps of the first simulation in water, the pair list was updated every 20 fs (10 steps). The time between updating the pair list and long-range electrostatic forces was reduced to 10 fs after 300 ps of simulation time in order to improve the conservation of energy in the system. In the vacuum and the second water simulation, an update of the pair list of 10 fs was used throughout. A time step of 2 fs for integrating the equations of motion was used throughout the simulations. To allow initial relaxation of the water around the protein, the coordinates of the protein were positionally restrained for 5 and 50 ps in the case of the first and second simulations in water, respectively. Analysis was performed using trajectory coordinates and energies written to disk every 0.05 ps.

The vacuum simulation (V1) was performed using the GROMOS87 vacuum force field, parameter set 37D4. The first water simulation (W1) was performed using the GROMOS87 force field for solvent simulations, parameter set 37C4. The appropriate dielectric permittivity of 1.0 was used in both cases. The vacuum force field is a modified version of the solvent (37C4) parameter set in which the partial charges of atoms in groups bearing a net charge are changed such that the net charge becomes zero, but the hydrogen-bonding capacity is maintained. This effectively mimics the effect of dielectric screening by solvent and reduces the long-

range distorting effects of charged amino acids in vacuum simulations.

Analysis of the first water simulation (W1) indicated that significant changes had occurred in the packing of side chains in the hydrophobic core of lysozyme during the 1 ns simulation time. The GROMOS87 force field uses a united atom representation for nonpolar hydrogens including hydrogens attached to aromatic rings. Although computationally efficient and adequate for the simulation of the properties of many liquid systems, a united atom model does not correctly model a number of features of intermolecular interactions. For example, the minimum energy configuration of benzene dimers or the perpendicular stacking arrangement found in benzene crystals is not reproduced. This can be corrected, however, by the inclusion of explicit hydrogens bearing small partial charges. As a first-order correction to the GROMOS87 force field, explicit hydrogens were added to aromatic carbons on the side chains of phenylalanine, tyrosine, and tryptophan. The partial charge on a hydrogen was taken to be $q_H = -q_C = 0.14e$ in each case and was chosen on the basis of the quadrupole moment of benzene; this partial charge has been shown to reproduce correctly the stacked structure of crystalline benzene (P. M. King, personal communication). The van der Waals parameters for the aromatic hydrogens were $C6^{1/2}(i,i) = 4.5$ (kcal mol⁻¹ Å⁶)^{1/2} and $C12^{1/2}(i,i) = 60.1$ (kcal mol⁻¹ Å¹²)^{1/2} (Hermans et al., 1984). The first water simulation (W1) also showed that the GROMOS87 force field overestimated the interaction between carbon and SPC water [Åqvist personal communication; Mark et al. (1994)]. This had the effect that some nonpolar residues became partially solvated during the simulations, leading to an inappropriate increase in hydrophobic surface area. In the GROMOS87 force field the Lennard-Jones potential, V_{LJ} , is expressed in terms of C6 and C12 parameters where $V_{LJ} = C12/r^{12} - C6/r^6$. The C6 and C12 parameters for the pairwise interaction between two atoms i and j are derived using the following combination rule, i.e., $C12(i,j) = C12^{1/2}(i,i)C12^{1/2}(j,j)$, where different values for $C12^{1/2}(i,i)$ and $C12^{1/2}(j,j)$ can be chosen for given atom types i or j depending on the nature of the interaction. For the interaction between a carbon and SPC water oxygen a value for the repulsive term in the Lennard-Jones potential of $C12^{1/2}(OW,OW) = 421.0$ (kcal mol⁻¹ Å¹²)^{1/2} is used in the standard GROMOS87 force field. Calculation of the free energy of solvation of CH₄ suggests, however, that this results in a free energy of hydration that is too favorable and that use of the value of $C12^{1/2}(OW,OW) = 793.3$ (kcal

Table 2: Comparison of the Three Molecular Dynamics Trajectories with the Experimental NOE Distance Restraints

	no. of violations			av violation (Å) $R_E - R_O$
	>0.5 Å	>1.0 Å	>3.0 Å	
crystal structure	42	27	2	0.06
vacuum simulation (V1, 200–1000 ps)	122	81	20	0.20
first water simulation (W1, 200–1000 ps)	112	81	17	0.19
second water simulation (W2, 300–1100 ps)	94	54	7	0.14

Examples of Large NOE Violations^a

NMR data NOE	R_O	vacuum simulation			first water simulation			second water simulation		
		R	R_E	$R_E - R_O$	R	R_E	$R_E - R_O$	R	R_E	$R_E - R_O$
13HN – 129Hy	6.7	8.4	8.1	1.4	11.4	10.8	4.1	6.1	6.0	0
28He3 – 32Hβ	5.5	5.8	5.7	0.2	11.7	11.4	5.9	6.9	6.8	1.3
28Hζ3 – 56Hγ	3.0	4.7	4.1	1.1	9.6	9.0	6.0	5.8	5.7	2.7
30HN – 120Hβ	5.2	10.1	10.0	4.8	9.5	9.1	3.9	6.3	6.1	0.9
34Hζ – 123Hζ2	6.6	5.5	5.2	0	12.0	11.3	4.7	3.9	3.7	0
43Hα – 51Hy2	4.0	8.6	8.5	4.5	4.9	4.4	0.4	4.1	3.9	0
56Hβ – 108He1	5.4	11.1	11.0	5.6	6.7	5.6	0.2	4.1	3.9	0
95Hβ – 108Hζ2	4.0	8.7	8.3	4.3	5.6	5.3	1.3	8.7	8.5	4.5
99Hβ – 108Hζ3	4.5	6.4	6.1	1.6	9.3	9.0	4.5	8.1	7.4	2.9
112Hα – 116Hβ	3.9	3.4	3.3	0	9.6	8.0	4.1	5.0	4.6	0.7

^a R_O is the upper limit of the experimental NOE distance restraint; R is the average distance from the simulation trajectory; R_E is the distance from the trajectory using $1/r^3$ averaging. 0 indicates no violation. Values in angstroms.

$\text{mol}^{-1} \text{Å}^{12})^{1/2}$ is more appropriate [Åqvist, personal communication; Mark et al. (1994)]. The force field used to perform the second water simulation (W2) incorporated the explicit inclusion of aromatic hydrogens, and this modified interaction between water oxygen and the carbon atoms of the protein.

(b) *NMR Parameters.* The comparisons used experimental data sets of 1158 NOE distance restraints, $^3J_{\text{HN}\alpha}$ and $^3J_{\alpha\beta}$ coupling constants for 106 and 57 residues, respectively, and $^1\text{H}-^{15}\text{N}$ order parameters for 121 backbone amides and the side chain NHs of 28 residues. All these data have been reported previously (Smith et al., 1991, 1993; Buck et al., 1995).

In calculating the NMR parameters from the simulations, 800 ps of the trajectory was used in each case. This excludes the initial 200 ps for the vacuum (V1) and first water (W1) simulations, where these systems may still have been equilibrating. The first 300 ps was excluded in the case of the second water (W2) simulation as this system was relaxed more slowly; in this case the simulation was continued to 1100 ps.

For comparison with the NOE distance restraints, the mean interproton distances, $\langle r \rangle$, and the distances calculated using $1/r^3$ averaging, $\langle (r^{-3})^{-1/3} \rangle$, were calculated from the simulation trajectories. Note that on the short time scale of an MD simulation, where the influence of angular fluctuations should be neglected (Tropp, 1980), it is appropriate to compare the NOE data with $1/r^3$ ($\langle (r^{-3})^{-1/3} \rangle$) averaged distances from a trajectory, rather than $1/r^6$ ($\langle (r^{-6})^{-1/6} \rangle$) averaged distances.

Coupling constants were calculated using the relationship (Karplus, 1959)

$$^3J = A \cos^2 \theta + B \cos \theta + C$$

where for $^3J_{\text{HN}\alpha}$, $\theta = \phi - 60^\circ$ and for $^3J_{\alpha\beta}$, $\theta = \chi_1 - 120^\circ$ for $\text{H}\beta_2$ and $\theta = \chi_1$ for $\text{H}\beta_3$. The values of A , B , and C used in the calculation were 6.4, -1.4, and 1.9 Hz for $^3J_{\text{HN}\alpha}$ (Pardi et al., 1984) and 9.5, -1.6, and 1.8 Hz for $^3J_{\alpha\beta}$ (de Marco et al., 1978).

The internal correlation function $C_1(t)$ for the internal motion of the $^1\text{H}-^{15}\text{N}$ vector was calculated using the relationship

$$C_1(t) = \langle P_2[\cos \theta(t)] \rangle = 1/2 \langle 3 \cos^2 \theta(t) - 1 \rangle$$

where $\theta(t)$ is the angle between the interatomic vector at time t and at time $t = 0$ and the angle brackets denote an average over all time origins (Lipari & Szabo 1982). At long times, $C_1(t)$ should decay to a plateau, the value of which will give the $^1\text{H}-^{15}\text{N}$ order parameter, S^2 . However, estimation of order parameters from the long-time behavior of $C_1(t)$ from a simulation involves large statistical errors. For this reason in this work order parameters were estimated from the trajectories using

$$S^2 = \frac{1}{2} \left[3 \sum_{\alpha=1}^3 \sum_{\beta=1}^3 \langle \mu_\alpha \mu_\beta \rangle^2 - 1 \right]$$

where μ_α ($\alpha = 1, 2, 3$) are the x , y , and z components of the normalized interatomic vector in the molecular frame, which is statistically more reliable (Chandrasekhar et al., 1992). The correlation functions and order parameters were calculated after a least-squares fit on all $\text{C}\alpha$ atom positions in order to remove overall rotational motion.

RESULTS

(a) *Comparison with NOE Data.* A set of 1158 NOEs has been identified for lysozyme. Previous comparisons of these data with crystal structures of the protein, and calculation of NMR solution structures using these data, have revealed the very close similarity of the structure of lysozyme in solution and in crystals (Smith et al., 1993). Although the crystal structures exhibit more NOE violations than the NMR structures do (e.g., the NMR structures have no violations of $>1 \text{Å}$ but the triclinic crystal structure has 31 violations of $>1 \text{Å}$), the majority of these violations result from only small differences in side chain orientations. In this work, the interproton distance restraints derived from these experimental NOEs were compared with the three simulations and the starting crystal structure. Table 2 summarizes the results of the comparison of the mean interproton distances from the trajectory, $\langle r \rangle$, and the distances calculated using $1/r^3$ averaging, $\langle (r^{-3})^{-1/3} \rangle$, with the upper distance limits from the NOE restraints. All three

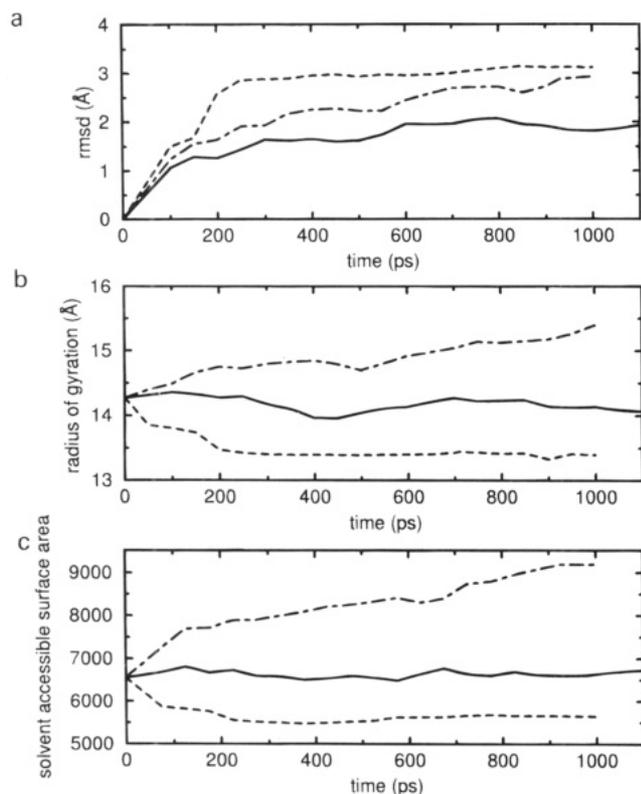


FIGURE 1: Root mean square difference (RMSD) from the starting crystal structure (a, in Å), the radius of gyration (b, in Å), and the total solvent accessible surface area (c, in Å²) through the three simulations as a function of time. The values plotted are for structures averaged over 50 ps windows. Dashed line, vacuum simulation (V1); dot-dashed line, first water simulation (W1); solid line, second water simulation (W2). The solvent accessibilities were calculated using the program NACCESS (Hubbard & Thornton, 1993).

simulations show significantly more NOE violations (i.e., distances averaged over the trajectory being greater than the upper distance bound from NOE) than the starting crystal structure (20, 17, and 7 violations of >3 Å for V1, W1, and W2 compared with 2 of >3 Å for the crystal). Analysis of the simulations shows, however, that these violations result from changes in the overall structure of the protein during the simulation (Figure 1a), as described below, rather than from any dynamical effects. Indeed, the average distances (R) and the $1/r^3$ average distances (R_E) are very similar (e.g., the mean $R - R_E$ for the second water (W2) simulation is 0.02 Å; see also examples in Table 2) confirming that dynamical fluctuations have little effect on the interproton distances over the 800 ps time period considered.

In the vacuum simulation (V1), there is a clear overall contraction of the structure. There is an initial rapid reduction in the radius of gyration and total solvent-accessible surface area followed by a slow drift continuing through the simulation (Figure 1b and c). Various structural changes are involved in this contraction. In particular, there is a large distortion of the triple-stranded antiparallel β sheet which folds over toward the C and D helices. The second and third strands of the sheet become almost buried by the end of the simulation (Figure 2b). These changes result in some of the very large NOE violations (Table 2), e.g., 43H α - 51H γ 2 (4.5 Å violation) and 56H β - 108H ϵ 1 (5.6 Å violation).

Significant structural changes also occur in the first water simulation (W1). The nature of the changes, which ac-

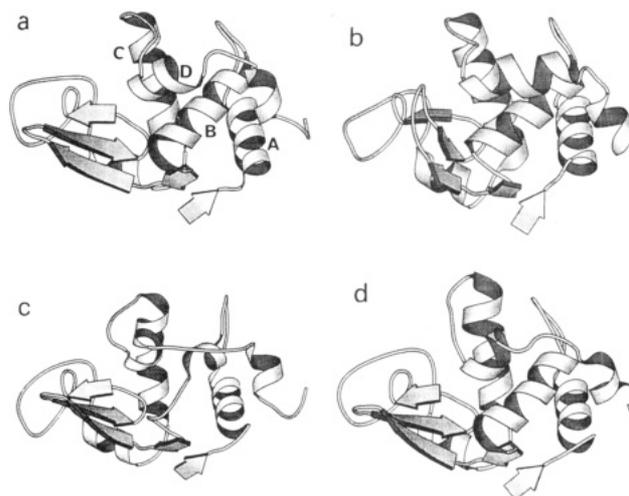


FIGURE 2: Schematic diagrams showing the starting crystal structure (a), and average structures from the last 50 ps of the vacuum, V1 (b), first water, W1 (c), and second water, W2 (d), simulations. The secondary structure shown is that identified in each of the structures using the method of Kabsch and Sander (1983). For the starting crystal structure, the four α helices are labeled A (residues 4–15), B (residues 24–37), C (residues 88–101), and D (residues 108–115). The diagram was generated using the program MOLSCRIPT (Kraulis, 1991).

cumulate throughout the whole simulation, is in marked contrast to those seen in the vacuum simulation (V1). There is a loss of close packing, particularly of hydrophobic residues, in the α domain with water molecules penetrating into the protein core. The hydrophobic side chains of Phe 34 and Met 105 extend out into solution by the end of the simulation. This leads to a significant increase in the overall radius of gyration of the molecule and in the solvent accessible surface area of the structure (see Figure 1b and c). A striking example of the loss of close packing is demonstrated in Figure 3, which shows a set of aromatic (Tyr 20, 23; Trp 28, 108, 111) and other hydrophobic (Leu 17, Ile 98, Met 105) residues, which cluster together in the core of the α domain in the native protein forming the so-called "hydrophobic box" (Blake et al., 1965). The native packing of these residues is completely lost by the end of this simulation. A number of the observed large NOE violations involve residues in this hydrophobic box region e.g., 28H ζ 3 - 56H γ (6.0 Å violation) and 99H β - 108H ζ 3 (4.5 Å violation).

Concurrent with the loss of close packing, examination of the structures indicated that helix B, and to a lesser extent helix D, unwind during the simulation (Figure 2c). Within 100 ps the NH(i)-CO($i-4$) hydrogen bonds typical of an α helix between residues 26–30, 27–31, 28–32, 29–33, 30–34, and 109–113 are almost entirely lost. Side chains of residues from both of these helices are involved in the hydrophobic box; the loss of the native hydrophobic interactions in this region appears, in part, to be responsible for the destabilization of these helices in the simulation. It is interesting to note the differing behavior of the α and β domains of lysozyme in this simulation. While the α domain unfolds substantially the β domain remains virtually intact. This behavior strongly contrasts with the vacuum simulation (V1), where the most significant changes are observed in the β domain and the α domain remains largely folded. The differing behavior of the domains in the simulations reflects an important difference between the interactions within them that stabilize the native fold. In the α domain, interactions

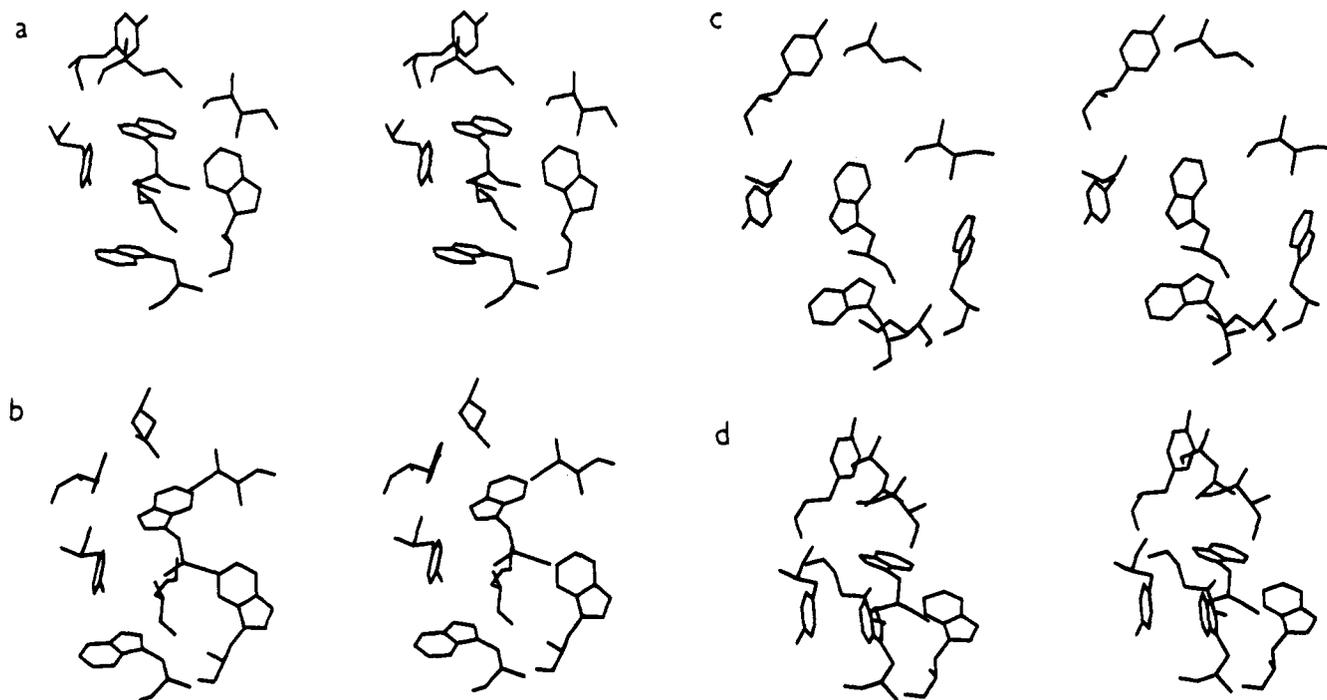


FIGURE 3: Stereodiagrams showing the packing of residues in the hydrophobic box region of lysozyme (Leu 17; Tyr 20, 23; Trp 28, 108, 111; Ile 98; Met 105) for the crystal structure (a) and average structures from last 50 ps of the vacuum, V1 (b), first water, W1 (c), and second water, W2 (d), simulations.

between hydrophobic side chains in the core of the protein are important for holding the helices together while in the β domain interstrand hydrogen bonds are crucial in maintaining its integrity. Interestingly, different characteristics of the two domains of lysozyme are features of the experimentally observed folding behavior of the protein (Dobson et al., 1994); these again are likely to reflect the differing interactions in the two domains.

In the light of the structural rearrangements observed in the first water simulation (W1), a second water simulation was performed using a force field incorporating aromatic hydrogens explicitly and with a modified interaction between the water oxygen atom and the carbon atoms of the protein as described in the Methods section. In this simulation (W2), the overall conformation of the starting structure is retained with only slight variations in the radius of gyration and solvent-accessible surface area (Figures 1 and 2d). The only significant structural changes occur in the region between the C and D helices (CD loop, residues 100–107). These lead to partial unwinding of the end of the C helix and some changes of the packing of Met 105, Trp 108, and Trp 111 in the hydrophobic box by the end of the simulation. Interestingly, this is the region where the backbone conformation differs most between the triclinic (used in this simulation as the starting structure) and tetragonal crystal structures of lysozyme ($C\alpha$ deviation between the two crystal structures of 2.9 Å for Asn 103). The number of NOE violations is significantly reduced for this simulation compared with the vacuum (V1) and first water (W1) simulations, those remaining in general involving residues in the vicinity of the CD loop, e.g., 95H β – 108 H ζ 2 (4.5 Å violation).

(b) *Comparison with Coupling Constants.* Using the Karplus relationship (Karplus, 1959), mean $^3J_{\text{HN}\alpha}$ and $^3J_{\alpha\beta}$ coupling constants have been calculated from the three trajectories. Figure 4 shows comparisons of the experimental $^3J_{\text{HN}\alpha}$ coupling constants with the mean values calculated from the simulation trajectories and also from the starting

triclinic crystal structure. There is a close agreement between the experimental coupling constants and those calculated from the crystal structure (RMSD 0.89 Hz). In all three cases, the values calculated from the trajectories agree less well with the experimental data (RMSD 2.08 Hz for V1, 1.81 Hz for W1, and 1.83 Hz for W2) than the starting structure does. Although this is not surprising for the vacuum (V1) and first water (W1) simulations where large structural changes have been observed in the course of the simulation, it is at first unexpected that there should be no improvement in the correlation of the experimental values with those calculated for the second water simulation (W2).

This result partly reflects the high sensitivity of the $^3J_{\text{HN}\alpha}$ coupling constants to small changes in ϕ , especially for a helical conformation. In this region the Karplus curve is very steep and a 10° change in ϕ can alter $^3J_{\text{HN}\alpha}$ by as much as 1.2 Hz. Regions where only slight conformational changes are induced by the force field can therefore exhibit large deviations of the calculated coupling constants from the experimental values. Examples from the second water simulation (W2) include Cys 94 in the center of helix C and Ile 124 in the C-terminal 3_{10} helix. Helix C becomes more regular in the simulation, the mean $^3J_{\text{HN}\alpha}$ coupling constant for Cys 94 from the trajectory being 4.7 Hz compared to a calculated value of 6.2 Hz for the crystal structure and an experimental value of 6.3 Hz. For Ile 124 in the second water simulation, the effective potential energy well for the ϕ angle is shifted in the simulation away from the crystal structure value of -119.7° , and ϕ undergoes small fluctuations around an average value of -91.0° (calculated $^3J_{\text{HN}\alpha}$ values from the crystal structure and the simulation are 9.7 and 7.8 Hz, respectively; the experimental $^3J_{\text{HN}\alpha}$ value is 10.6 Hz).

Dynamical fluctuations as well as structural transitions affect the coupling constants calculated from the second water (W2) trajectory. As has been shown previously, librations within a ϕ potential energy well have a relatively

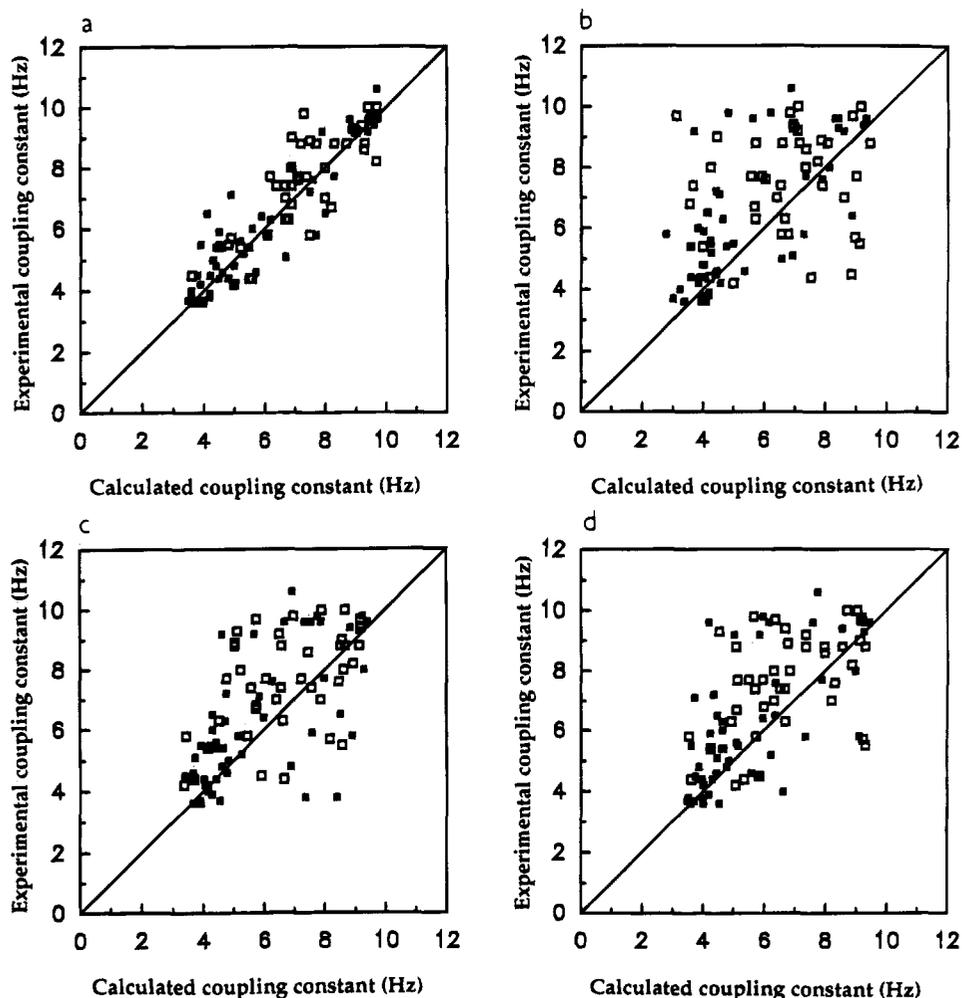


FIGURE 4: Comparison of the experimental and calculated $^3J_{\text{HN}\alpha}$ coupling constants for hen lysozyme. In (a) the coupling constants were calculated from the triclinic crystal structure. In (b–d) the coupling constants were calculated from the trajectories of the vacuum (b), first water (c), and second water (d) simulations. In each case filled boxes indicate residues in secondary structure and open boxes residues in nonsecondary structure regions.

limited effect on mean calculated coupling constants (Hoch et al., 1985). In contrast, jumps of the ϕ torsion angle between potential energy wells can lead to significant averaging of the mean calculated coupling constant values. For a number of residues in this simulation, infrequent jumps of the ϕ torsion angle between potential energy wells are observed (see, for example, Ile 78 in Figure 5b). Experimentally, coupling constants are averaged over a millisecond time scale. Consequently, for a residue where only a small number of transitions is observed during the 800 ps of the simulation, the calculated coupling constant will not be completely averaged and the calculated value will depend very strongly on the time points of the angle transitions between potential energy wells in the simulation. In a few cases, however, many ϕ transitions are observed in the 800 ps trajectory and the calculated $^3J_{\text{HN}\alpha}$ values are adequately averaged. One example of this is Ser 85, where there are 87 transitions between the potential energy wells of $\phi = -60^\circ$ and -120° in the second water simulation (Figure 5d). Here the $^3J_{\text{HN}\alpha}$ value calculated from the trajectory (5.7 Hz) agrees well with the experimental value (5.8 Hz), and significantly better than with that for the starting structure (calculated $^3J_{\text{HN}\alpha}$ of 7.5 Hz). Note, however, that the ϕ angle in the triclinic crystal structure used in this simulation (-86.4°) differs from that in the tetragonal type 2 crystal structure of lysozyme (-72.4°); the $^3J_{\text{HN}\alpha}$ value calculated

from this tetragonal type 2 crystal structure also agrees well with the experimental value (experimental and calculated $^3J_{\text{HN}\alpha}$ values both 5.8 Hz) (Smith et al., 1991).

In the case of $^3J_{\alpha\beta}$ coupling constants, the behavior of the different residues in the protein can be divided into two classes on the basis of the experimental data: those where a single staggered χ_1 conformation is populated and those where multiple χ_1 conformations are adopted. Residues with two β -protons occupying a single staggered χ_1 conformation have either two small coupling constants or one large and one small $^3J_{\alpha\beta}$ coupling constant, while population of multiple χ_1 conformations gives rise to two $^3J_{\alpha\beta}$ values in the range 6–8 Hz. For hen lysozyme, out of the 57 residues whose $^3J_{\alpha\beta}$ coupling constants could be experimentally measured, 16 were found to be occupying multiple χ_1 conformations. All of these residues, with one exception, were in exposed positions on the surface of the protein (Smith et al., 1991).

Many of the exposed residues show no χ_1 transitions in the simulations despite their experimentally averaged $^3J_{\alpha\beta}$ values. However, the absence of transitions in 800 ps is not necessarily incompatible with experimentally averaged coupling constants as the experimental averaging occurs where there is interconversion between conformers at a rate of at least 20 Hz (i.e., with a millisecond time scale or faster). In some cases, such as Arg 45 in the second water simulation

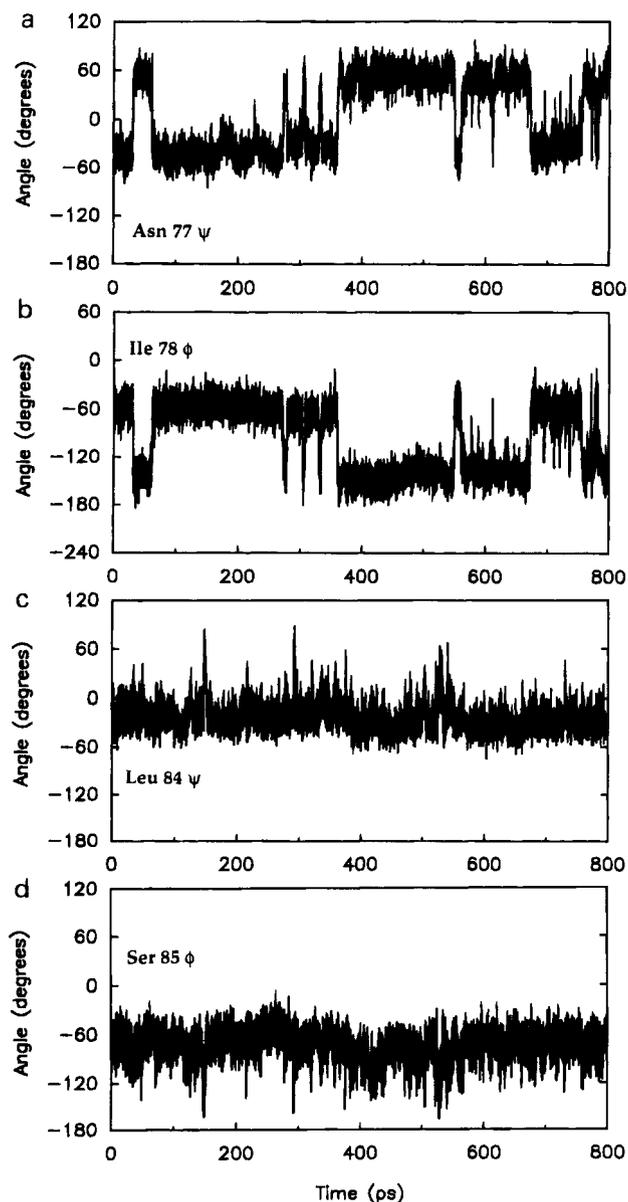


FIGURE 5: Time series showing the fluctuations in torsion angle through the final 800 ps of the second water simulation for (a) Asn 77 ψ , (b) Ile 78 ϕ , (c) Leu 84 ψ , and (d) Ser 85 ϕ .

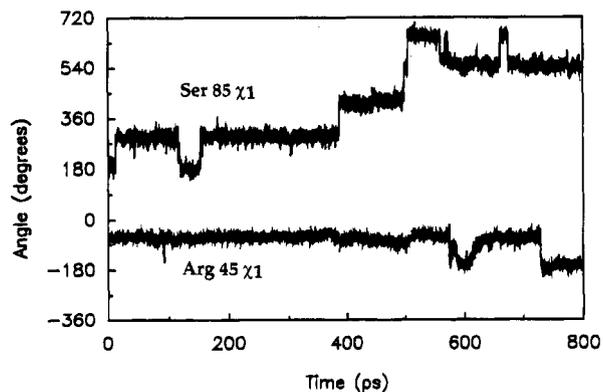


FIGURE 6: Time series showing the fluctuations in Arg 45 χ_1 (lower trace) and Ser 85 χ_1 (upper trace) torsion angles through the final 800 ps of the second water simulation.

(Figure 6), a small number of χ_1 transitions are observed and the calculated ${}^3J_{\alpha\beta}$ values are not significantly averaged (${}^3J_{\alpha\beta}$ experimental 6.9, 6.7 Hz; calculated 11.0, 4.1 Hz). Where more χ_1 transitions are seen in 800 ps, e.g., Ser 85 in

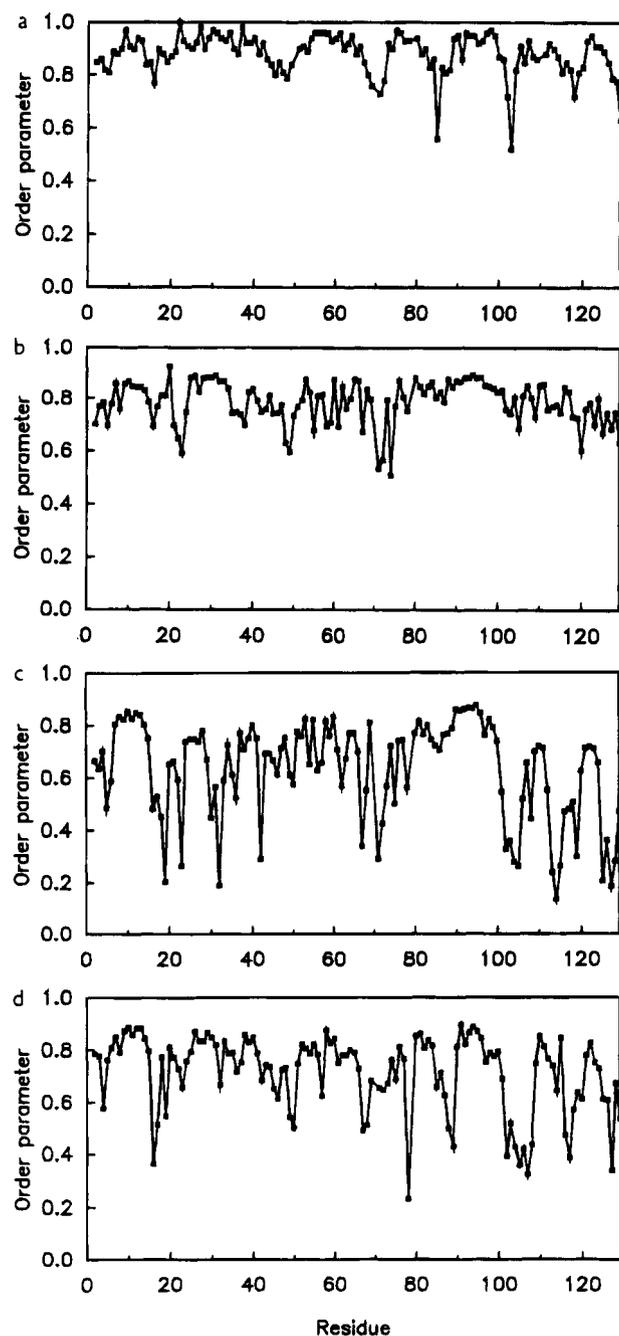


FIGURE 7: ${}^1\text{H}$ - ${}^{15}\text{N}$ order parameters (S^2) for backbone amide groups: (a) Experimental data (Buck et al., 1995); (b) Calculated from the vacuum trajectory (200–1000 ps); (c) Calculated from the first water trajectory (200–1000 ps); (d) Calculated from the second water trajectory (300–1100 ps).

Figure 6, the calculated ${}^3J_{\alpha\beta}$ values are better averaged and agree well the experimental values (${}^3J_{\alpha\beta}$ experimental 7.4, 5.7 Hz; calculated 7.8, 6.3 Hz).

(c) Comparison with ${}^{15}\text{N}$ Relaxation Data. Main Chain. The experimental ${}^1\text{H}$ - ${}^{15}\text{N}$ order parameters for hen lysozyme (Buck et al., 1995) are shown in Figure 7 together with values calculated from the three simulations using the procedure described in the Methods section. Backbone mobility is very restricted in the vacuum (V1) simulation, leading to large order parameters for the main chain amide groups of almost all the residues in the protein. The lowest values are seen for residues in regions of the structure that become distorted in the simulation such as Gly 49 (S^2 0.60), in the turn connecting strands 1 and 2 in the β sheet, and Asn 74 (S^2

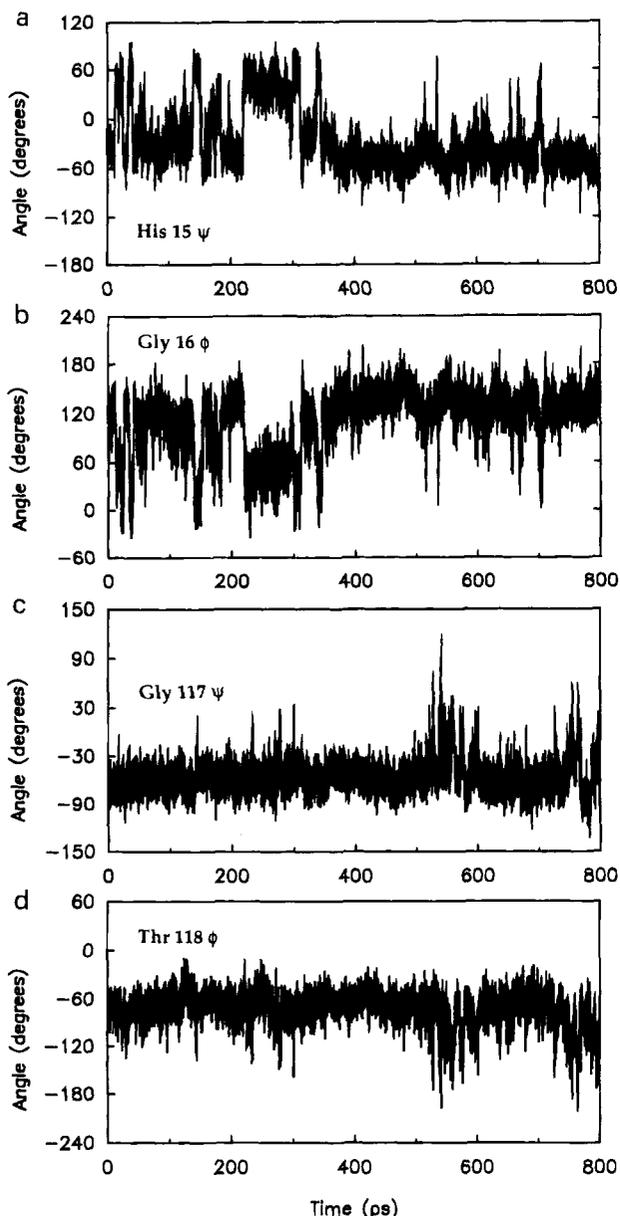


FIGURE 8: Time series showing the fluctuations in torsion angle through the final 800 ps of the second water simulation for (a) His 15 ψ , (b) Gly 16 ϕ , (c) Gly 117 ψ , and (d) Thr 118 ϕ .

0.51) in the long loop region. Low order parameters are not seen, however, for the only two residues that experimentally have order parameters less than 0.6 (Ser 85, Asn 103).

The order parameters calculated from the two water trajectories are shown in Figure 7c and d. Although the correlation with experimental data for the second water (W2) simulation is improved compared with that for the first water simulation (W1), both curves have significantly higher frequency variations with residue number than is found experimentally, a number of isolated residues (or pairs of residues) having low order parameters relative to their near neighbors in the sequence. These low values have been found to arise from concerted ϕ and ψ (of the preceding residue) transitions (i.e., flips of the peptide bond plane). Examples from the second water simulation (W2) are shown in Figures 5 and 8 for Ile 78 (calculated S^2 0.23), Ser 85 (S^2 0.66), Gly 16 (S^2 0.37) and Thr 118 (S^2 0.57). These transitions cover approximately 60° , i.e., between neighboring potential energy wells, and they lead to a rapid drop in the correlation function and a low order parameter (see

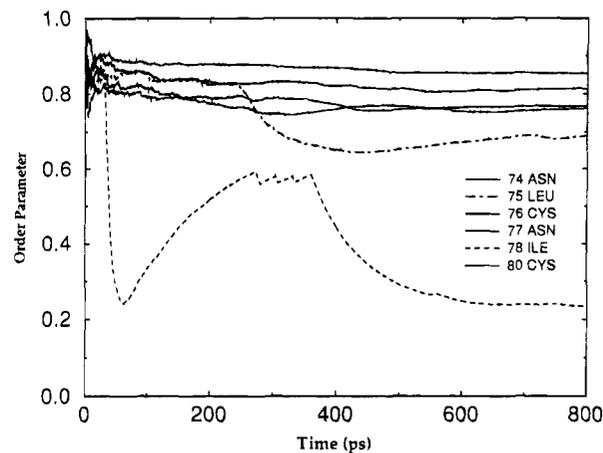


FIGURE 9: Variation in the ^1H - ^{15}N order parameter values with time calculated from the second water trajectory for the backbone amides of residues 74–80.

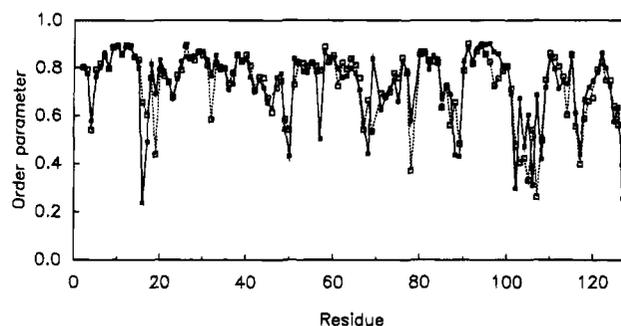


FIGURE 10: ^1H - ^{15}N order parameters for backbone amide groups calculated using 300–700 (filled boxes and solid line) and 700–1100 ps (open boxes and dotted line) of the second water trajectory.

Figure 9, which shows the change in S^2 with time for Ile 78). This results from the dependence of the correlation function for the internal motions on $[3 \cos^2 \theta(t) - 1]$ (see Methods), and we note that the value of the function drops to zero for $\theta = 54^\circ 44'$, and hence 60° jumps result in substantial reductions in the calculated S^2 values.

In general, these ϕ , ψ transitions occur too rarely in the 800 ps block analyzed for the order parameter to be accurately calculated; indeed, the calculated value depends on the exact positions of the transitions in the simulation. This is illustrated in Figure 10, which compares the order parameters calculated, using the second water simulation trajectory, from 300 to 700 ps with those calculated from 700 to 1100 ps. For example, for Gly 16 where the ϕ , ψ transitions are concentrated in the first part of the simulation (see Figure 8), the S^2 values calculated using the 300–700 and 700–1100 ps parts of the simulation are 0.24 and 0.66, respectively, while for Thr 118 the S^2 values are 0.67 and 0.59, respectively (ϕ , ψ transitions in the second part of simulation; see Figures 8 and 10).

Ser 85 is one case where there are sufficient ϕ transitions in the second water simulation for the order parameters to be reliably calculated. This residue has one of the lowest experimental order parameters (0.56), but the motional process giving rise to this is not known. Although the value calculated from the simulation (0.66) is higher than that observed experimentally, frequent peptide bond transitions of the type observed in the simulation could provide an explanation for the low experimental order parameter. In this regard it is interesting to note that the $^3J_{\text{HN}\alpha}$ coupling constant calculated from the simulation for Ser 85 agrees

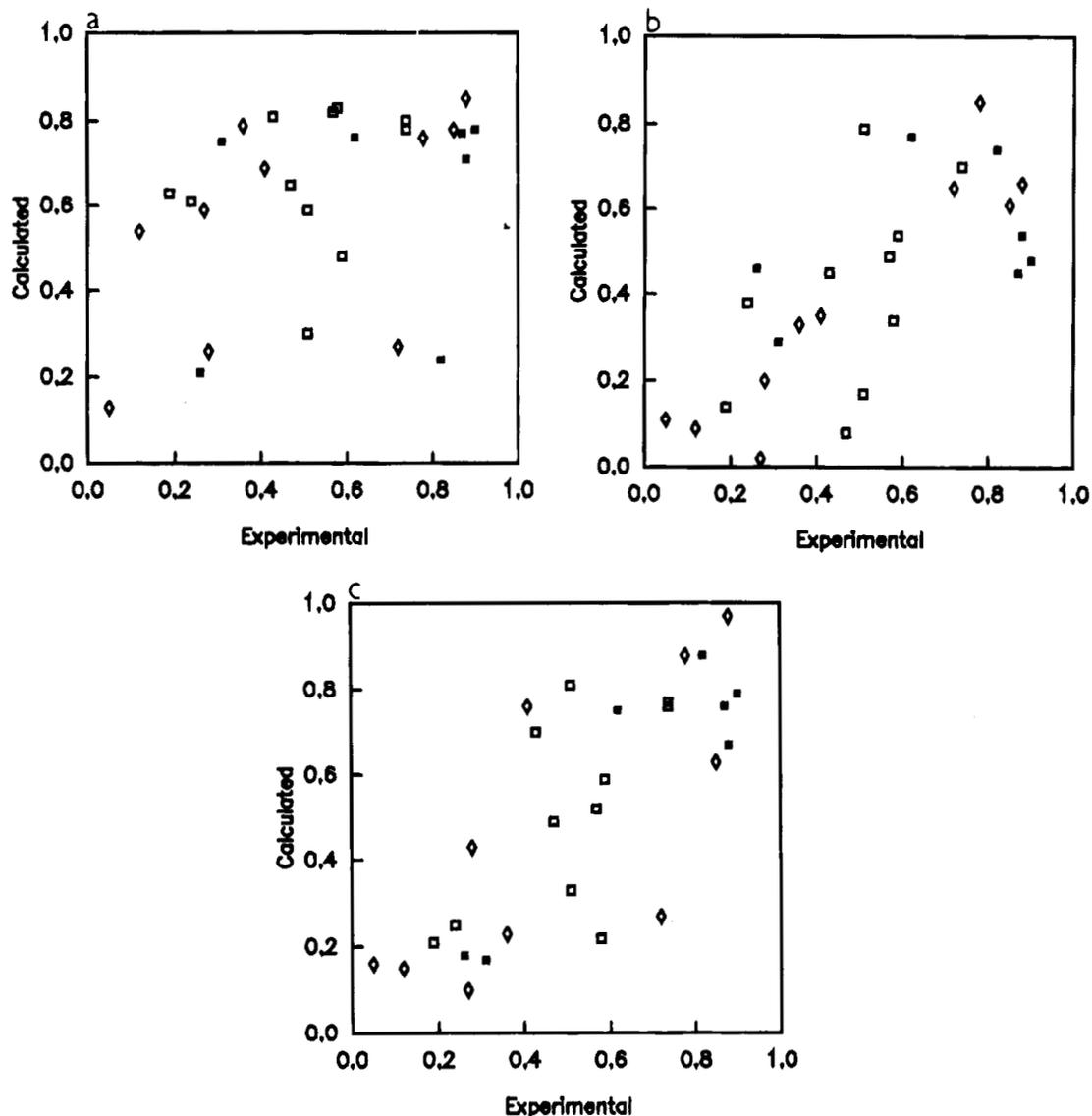


FIGURE 11: Comparison of the experimental ^1H - ^{15}N order parameters of side chain NH groups with those calculated from the vacuum simulation, V1 (a) and from the first, W1 (b), and second, W2 (c), water simulation. For NH_2 groups (Asn, Gln) the mean of the values for the two NH vectors is used. A filled box indicates a buried side chain (solvent accessibility for N atom less than 30%), an open diamond residues with side chains with limited accessibility (percentage solvent accessibility for N atom in range 30–80%), and an open box exposed side chains (solvent accessibility for N atom greater than 80%).

well with the experimental value as discussed earlier. Ser 85 is situated at the end of the first 3_{10} helix in lysozyme, and during the course of the simulation, its amide proton forms hydrogen bonds to the carbonyl oxygens of Ser 81, Ala 82, and Leu 83 for 28%, 40%, and 3% of the time, respectively. This variety of hydrogen bond partners may help to stabilize the conformational changes this residue undergoes.

For the residues where few ϕ transitions are observed in the 800 ps block considered, these angle transitions may be occurring experimentally on too long a time scale for their effects to appear in the experimental order parameter data. However, comparison here of the simulations with other experimental data, particularly the coupling constants, has been particularly valuable as it has suggested that many of these concerted ϕ , ψ transitions are a consequence of limitations in the force field used, the barrier preventing these transitions from being too low. Particularly clear evidence to support this arises when there is a large experimental $^3J_{\text{HN}\alpha}$ coupling constant (>9 Hz). In these cases, the populations of any ϕ angle values except those around $\phi = -120^\circ$ must

be very low or else $^3J_{\text{HN}\alpha}$ would be averaged to a significantly smaller value. An example of this is Thr 118, which undergoes concerted ϕ , ψ transitions in the second half of the simulation. These transitions result in averaging of the calculated $^3J_{\text{HN}\alpha}$ value to a value of 5.7 Hz. This is significantly lower than the experimental value of 9.8 Hz, demonstrating that for this residue at least the ϕ , ψ transitions observed must be an artifact due to the force field used.

Side Chains. As well as probing mobility of the main chain, ^1H - ^{15}N relaxation studies can examine fluctuations of the side chains. Experimental ^1H - ^{15}N order parameters are available for the side chain NH groups of 28 residues in lysozyme (NH_2 of Asn and Gln, NH of Arg, and NH of Trp) (Buck et al., 1995). Figure 11a shows a comparison of these experimental data with order parameters calculated from the vacuum simulation (V1). The correlation here is poor, at least in part, reflecting the large structural changes that are occurring in this simulation. For example, the side chain of Gln 57 in the β sheet changes conformation as this region distorts in the simulation, resulting in a low calculated side chain order parameter (mean calculated 0.24, experi-

mental 0.82). In addition, a number of residues have higher calculated order parameters than have been observed experimentally, suggesting that the overall mobility of the protein in the vacuum simulation is lower than that actually occurring in solution. An example of this is Gln 121 (mean calculated 0.79, experimental 0.30).

Comparisons of the experimental ^1H – ^{15}N side chain order parameters with those calculated from the water simulations are shown in Figure 11b and c. For the first water simulation (W1), a number of the tryptophan residues have much lower calculated order parameters than are observed experimentally (for Trp 28, 108, and 111, the calculated S^2 values are 0.48, 0.45, and 0.54 compared to the experimental S^2 values of 0.90, 0.87, and 0.88, respectively). However, this can be attributed to be a consequence of the disruption of the close packing in the hydrophobic core of the α domain during the trajectory. For the second water (W2) simulation, there is a good agreement of the calculated order parameters with the experiment data (correlation coefficient 0.77). This can be associated with the larger number of torsion angle transitions occurring within 800 ps in this simulation for exposed side chains, compared to the main chain, resulting in better statistics for the order parameter calculations. As was found for the experimental data (Buck et al., 1995), there is a strong correlation of the values of the order parameters with the exposure of the NH groups, the majority of the buried groups having order parameter values comparable with those found experimentally for the backbone (0.7–0.95) and the more exposed groups having order parameters in the range 0.1–0.6.

The large variation in the experimental order parameter values for the side chain NH groups and the good agreement with the values predicted from the second water (W2) simulation give us confidence that we can use this simulation to propose models for the motional processes giving rise to the experimental data. It has been found that the majority of the side chains fall into one of three classes of behavior (see examples in Table 3 and Figure 12), which differ in the frequency of transitions of their χ torsion angles between the staggered rotamers (g^+ , g^- , and t). Population of conformations other than these staggered rotameric states by the side chains in the simulations is very limited, in agreement with results from analyses of high-resolution protein structures (Janin et al., 1978; McGregor et al., 1987). For class I, the order parameter values are high (>0.7) and the fluctuations about χ torsion angles in the simulation are limited in general to librations within single potential energy wells (e.g., Gln 57 in Figure 12a). There are no, or very few (in general less than 10 per residue within the 800 ps analyzed), transitions to neighboring potential energy wells and when these transitions do occur the time spent in a neighboring well is very limited. The side chains of residues where only one χ torsion angle is conformationally disordered fall into class II with S^2 values in the range 0.45–0.65. Here there are many transitions between the potential energy wells of the different rotameric states of the disordered χ torsion angle (100–200 within the 800 ps analyzed) and a significant population of at least two of these rotameric states within the trajectory. In some of these cases, very different values for the order parameters of the two NHs of Asn or Gln are calculated from the trajectory. This arises because of the differing angular dependencies of the two N–H vectors to rotate about, in the case of Asn, the χ_2 torsion angle. However, the average of the two values in

Table 3: Examples of Side Chain ^1H – ^{15}N Order Parameters (Experimental and Calculated from the Second Water Trajectory) for Residues Exhibiting the Three Different Classes of Behavior^a

residue	mean exptl S^2	mean calcd S^2
Class I: Residues Exhibiting No (or Very Few) χ Transitions		
Gln 57	0.82	0.88
Asn 59	0.78	0.88
Trp 108	0.87	0.76
mean (6 residues)	0.79	0.78
Class II: Residues with One Disordered χ Torsion Angle		
Asn 65	0.57	0.52
Asn 93	0.59	0.59
Asn 113	0.47	0.49
mean (6 residues)	0.57	0.59
Class III: Residues with Two or More Disordered χ Torsion Angles		
Gln 41	0.19	0.21
Arg 73	0.12	0.15
Arg 125	0.05	0.16
mean (16 residues)	0.18 ^b	0.28

^a For each class the mean experimental and calculated order parameters are also listed together with the number of residues involved. Note that the behavior of five of the side chains studied does not fall into any of the three categories. ^b This mean value was calculated using experimental data for only 10 residues; for the other 6 (all Arg residues), the experimental S^2 values are not available although the order parameters for all Arg side chains in lysozyme are known to be less than 0.4 (Buck et al., 1995).

general agrees well with the average experimental value. Examples include Asn 65 (calculated S^2 values 0.74, 0.30, mean 0.52; experimental mean 0.57) and Asn 113 (calculated S^2 values 0.77, 0.21, mean 0.49; experimental mean 0.47); the fluctuations occurring about the χ_1 (small fluctuations) and χ_2 (many transitions) torsion angles of Asn 65 during the 800 ps of the simulation are shown in Figures 12c and 13a. The third class of residues, class III, contains those with an order parameter of 0.3 or less. It has been found that these low values involve conformational disorder about more than one side chain torsion angle. An example is shown in Figures 12e and 13b for the side chain of Arg 73, which has a calculated order parameter of 0.15 (experimental value 0.12) and undergoes χ angle transitions about both χ_3 and χ_4 in the 800 ps of the simulation.

DISCUSSION

(a) *NMR Parameter Prediction from MD Simulations.* The NMR parameters considered in this study, NOE distance restraints, 3J coupling constants, and ^1H – ^{15}N order parameters, contain structural information as well as information on the dynamics of the system corresponding to a wide variety of different time scales. To reproduce correctly the experimental observations, the simulations must accurately reproduce the structural as well as the dynamical properties of the system and, for both of these, adequately sample conformational space over an appropriate time scale. Reproduction of the structural and dynamical properties of the system will depend on the underlying model or force field used in the simulations. More specifically, this will require simulating the correct location and relative depths of minima in the (free) energy surface as well as the height of the intervening barriers. Therefore, when comparing the results from the simulations to the experimentally derived parameters, we are primarily interested in two aspects of this behavior. First, whether the simulations violate the structural and dynamical bounds implied by the experimental observa-

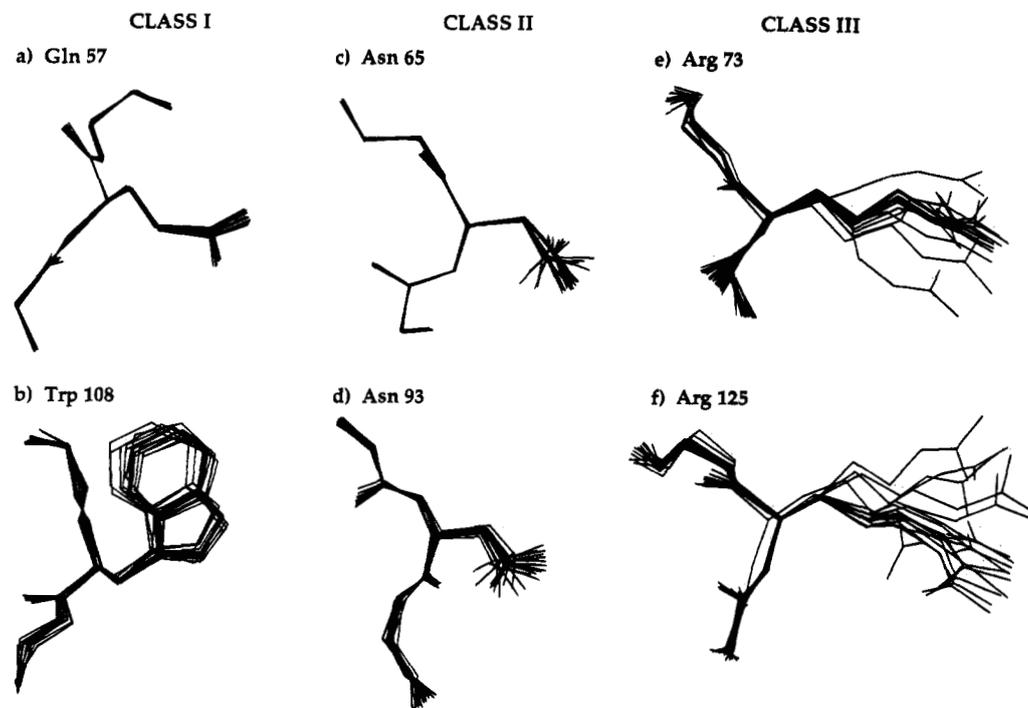


FIGURE 12: Examples of the three classes of side chain behavior. Average structures from 50 ps windows through the second water simulation (300–1100 ps) are shown for the residues: (a) Gln 57, class I; (b) Trp 108, class I; (c) Asn 65, class II; (d) Asn 93, class II; (e) Arg 73, class III; (f) Arg 125, class III. In each case the main chain of the residue concerned and the two residues adjacent to it in the sequence is used in the superposition of the structures. This main chain trace is shown on the left of each diagram and the side chain is shown on the right.

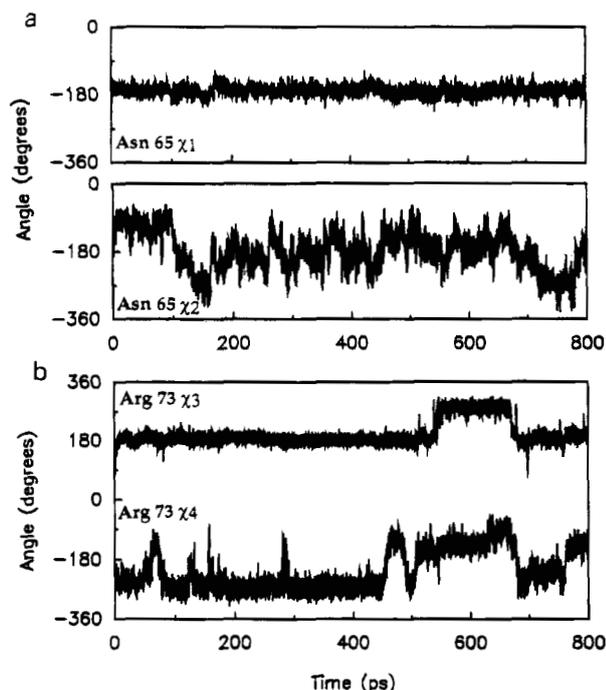


FIGURE 13: Time series showing the fluctuations in the torsion angles of Asn 65 and Arg 73 through the final 800 ps of the second water simulation.

tion, and second, given the ensemble and time averaging inherent in the experimental data, whether the sampling in the simulation is sufficient to extract statistically meaningful results.

Comparison of the three simulations of lysozyme with the experimental NMR data, in particular the NOE distance restraints, has enabled unrealistic changes to the structure of the protein during the course of the simulations to be identified. These changes are particularly significant in the

vacuum (V1) and first water (W1) simulations and clearly arise from the vacuum boundary condition and defects in the force field used. Many of the observed structural changes in the first water (W1) simulation were successfully eliminated in the second water (W2) simulation by use of a force field that incorporates explicit treatment of aromatic hydrogens and has a modified water oxygen–protein carbon interaction. It is important to note, however, that most of the structural changes observed only became apparent as the simulations were continued beyond approximately 500 ps. They therefore would not have been readily recognized in previous shorter simulations of lysozyme or other protein systems. This underlies the fact that as longer simulation times become accessible the results will be more sensitive to slight differences in force field parameterization. The choice of test case is also important. The large hydrophobic core in the α domain appears to make the lysozyme structure particularly sensitive to these force field differences.

The structural changes in the simulations were responsible for virtually all the violations of the experimental NOE distance restraints. Indeed, during the 800 ps time course analyzed, dynamic fluctuations have little effect on the interproton distances. This is in accord with previous studies which have suggested that, although the presence of picosecond motions will affect the NOEs, in general leading to a decrease in intensity, the changes will be too small to have a significant effect on distances derived from the NOE data (Dobson & Karplus, 1986; Post, 1992). The highly nonlinear dependence of NOE intensity on interatomic distance means in addition that rarely sampled configurations can make a dominant contribution to the observed intensity. For this reason, dynamics simulations on a nanosecond or even a microsecond time scale are unlikely to model correctly the atomic motions that give rise to conflicts in NOE distances

without the inclusion of additional (nonphysical) driving forces (Torda et al., 1990).

For coupling constants, the overall agreement between the experimental and calculated values is poor for all three simulations. The observed discrepancies arise in part from the sensitivity of the coupling constants to small changes in torsion angle values, but the effects of partial motional averaging are, however, significant. The 800 ps time course analyzed here was in general found to be insufficient for coupling constants to be accurately calculated if the torsion angle concerned was averaged over multiple conformations. This was because only a few transitions between neighboring potential energy wells were observed in the 800 ps simulation time. As coupling constants are averaged experimentally over a millisecond time scale, the observation of a few or even no torsion angle transitions in 800 ps is not incompatible with experimental data showing averaging between different conformational states. Far longer simulation times or other sampling methods would therefore be needed to predict fully the conformational distribution potentially probed by coupling constant measurements. This said, however, it is also clear that in a number of cases observed in the present study the underlying atomic processes that give rise to the averaged coupling constants occur in the simulations on a time scale of tens to hundreds of picoseconds and may be adequately sampled. A good illustration of such averaging is the $^3J_{\text{HN}\alpha}$ coupling constant of Ser 85; frequent peptide bond transitions are observed in the simulation for this residue and the averaged $^3J_{\text{HN}\alpha}$ coupling constant agrees well with experiment.

$^1\text{H}-^{15}\text{N}$ order parameters are in general interpreted as reflecting motions that occur on a time scale much faster than the overall rotational correlation time of the molecule (Wagner, 1993). For lysozyme, the rotational correlation time in H_2O at 308 K is of the order of 6 ns (Buck et al., 1995). Thus, the time scale of motions that order parameters will be sensitive to is of the same order as the length of the current simulations. It is expected, therefore, that for cases where low order parameters are observed experimentally, indicating significant motion on a subnanosecond time scale, the results from a realistic simulation should correlate well with the experimental data. This is the case for order parameters stemming from mobile side chains in the second water simulation (W2) examined in this work, as illustrated in Figure 11c. However, experimentally observed high order parameters (>0.7) place strict upper limits on the degree of motion in the simulation that is physically reasonable within the 800 ps analyzed. In the case of lysozyme, multiple peptide bond plane flips in the simulation lead to poor predictions for the amide order parameters of certain isolated residues in the second water simulation, (W2). These transitions are indications of limitations in the force field and suggest that the barriers preventing concerted ϕ and ψ torsion angle transitions leading to flips in the peptide bond plane are too low. To eliminate ϕ , ψ transitions for residues like Thr 118, where several transitions are observed in 800 ps, while retaining some transitions for residues such as Ser 85 discussed above, which undergoes many transitions in the current simulation, it would be sufficient to reduce the probability of such transitions by a factor of 2–10 corresponding to an increase in barrier height of 2–5 kJ mol^{-1} .

In cases where no concerted ϕ , ψ transitions occur in the simulation, or where only a small number of transitions are observed, it is impossible to come to any definitive conclu-

sions about the force field from the high or low order parameter values calculated. This is clearly illustrated in Figure 10, which shows that the calculated order parameters depend strongly on the region of the trajectory analyzed. In general, order parameters calculated from the truncated trajectories are higher, and for the backbone $^1\text{H}-^{15}\text{N}$ vectors agree better with experiment, due to more limited sampling. Similar results have been found for comparisons of calculated and experimental order parameters of other proteins (Chandrasekhar et al., 1992; Eriksson et al., 1993). In addition, it should not be concluded from the good agreement on average between the calculated and experimental S^2 values shown in Figure 7b for the simulation of lysozyme in vacuum (V1) that this simulation correctly reproduces the dynamics of the protein. This simulation fails to reproduce the low order parameters observed for mobile side chains (Figure 11a), indicating that the protein is essentially rigid on the relevant time scale, and thus many motions of interest are not revealed in the simulation.

Comparisons of the lysozyme simulations with the experimental NMR data have enabled an assessment to be carried out of how realistic both the structural and the dynamic characteristics of the protein are in simulation force fields. The NOE data in particular enabled global changes to the lysozyme structure arising from inaccuracies of the force field to be detected and then subsequently rectified, while the $^1\text{H}-^{15}\text{N}$ order parameters, in conjunction with the $^3J_{\text{HN}\alpha}$ coupling constants, have led to the identification of too frequent concerted ϕ , ψ transitions. This considerable insight gained from the comparisons carried out for lysozyme stresses the importance for further comparisons of this type to be performed using as long simulation times and as wide a range of experimental parameters as is possible. Moreover, careful analysis of simulations of proteins in the native state is an important step before the results of simulations carried out under more extreme conditions to visualize the unfolding process can be interpreted fully (van Gunsteren et al., 1995a).

(b) Insight into the Structural and Dynamic Properties of Lysozyme. The second water simulation (W2), where the structural features of native folded lysozyme are retained through the trajectory, has been able to contribute significantly to our understanding of the dynamic properties of lysozyme by providing plausible models for the types of motion that might be giving rise to the observed experimental data. In the case of the main chain, one specific example of this is Ser 85 where a low experimental $^1\text{H}-^{15}\text{N}$ amide order parameter has been observed experimentally but a motional process that could give rise to this low value was not apparent previously. This analysis, despite the limitations of the force fields, suggests that frequent flips of the peptide bond plane involving concerted motions of the ψ torsion angle of Leu 84 and the ϕ torsion angle of Ser 85 may be resulting in the low $^1\text{H}-^{15}\text{N}$ order parameter and also in an averaged $^3J_{\text{HN}\alpha}$ coupling constant. In the simulation, the conformational changes involved in this motion are stabilized by the presence of a number of different hydrogen-bonding partners available for the Ser 85 amide proton.

In the case of the side chains, the good correlation of calculated $^1\text{H}-^{15}\text{N}$ order parameters and experimental NMR data has enabled different classes of side chain behavior to be recognized, ranging from limited fluctuations in χ potential energy wells (class I with S^2 values greater than 0.70) to conformational disorder arising from transitions between potential energy wells for two or more χ torsion

angles (class III with S^2 values less than 0.3). The mobility involving multiple χ torsion angles that is indicated by the analysis of the second water (W2) simulation to result in the lowest values of side chain order parameters must clearly be considered when analyzing protein structures determined by NMR or indeed crystallographic techniques. This is of particular importance when interactions with substrates, receptors, or antibodies are being modeled, for example, as these interactions will involve surface residues which the results for lysozyme suggest are likely to have the smallest side chain order parameter values. Here, therefore, the simulations have been able to provide a basis for understanding the experimentally observed relationship between the order parameter values and solvent accessibility of the side chains.

ACKNOWLEDGMENT

We thank Matthias Buck, Paul Smith, and Rafael Brüs-chweiler for helpful discussions. The research of C.M.D. is supported in part by an International Research Scholars award from the Howard Hughes Medical Institute. L.J.S. is a Royal Society University Research Fellow.

REFERENCES

- Balasubramanian, S., Nirmala, R., Beveridge, D. L., & Bolton, P. H. (1994) *J. Magn. Reson. Ser. B* 104, 240.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., & Hermans, J. (1981) in *Intermolecular Forces* (Pullman, B., Ed.) p 331, Reidel, Dordrecht.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984) *J. Chem. Phys.* 81, 3684.
- Bernstein, F. C., Koetzle, T. F., Williams, G. B., Meyer, G. F., Price, M. D., Rodgers, J. R., Kennard, O., Shimanouchi T., & Tasumi, M. (1977) *J. Mol. Biol.* 122, 535.
- Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1965) *Nature* 206, 757.
- Brooks, C. L., III, Karplus, M., & Pettitt, B. M. (1988) *Adv. Chem. Phys.* 71, 1.
- Buck, M., Boyd, J., Redfield, C., MacKenzie, D. A., Jeenes, D. J., Archer, D. B., & Dobson, C. M. (1995) *Biochemistry* 34, 4041.
- Campbell, I. D., Dobson, C. M., & Williams, R. J. P. (1985) *Biochem. J.* 231, 1.
- Chandrasekhar, I., Clore, G. M., Szabo, A., Gronenborn, A. M., & Brooks, B. R. (1992) *J. Mol. Biol.* 226, 239.
- Clore, G. M., & Gronenborn, A. M. (1991) *Annu. Rev. Biophys. Biophys. Chem.* 20, 29.
- Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., & Gronenborn, A. M. (1990) *J. Am. Chem. Soc.* 112, 4989.
- de Marco, A., Llinas, M., & Wüthrich, K. (1978) *Biopolymers* 17, 617.
- Dobson, C. M. (1993) *Curr. Biol.* 3, 530.
- Dobson, C. M., & Karplus, M. (1986) *Methods Enzymol.* 131, 362.
- Dobson, C. M., Evans, P. A., & Radford, S. E. (1994) *Trends Biol. Sci.* 19, 31.
- Eriksson, M. A. L., Berglund, H., Härd, T., & Nilsson, L. (1993) *Proteins* 17, 375.
- Hermans, J., Berendsen, H. J. C., van Gunsteren, W. F., & Postma, J. P. M. (1984) *Biopolymers* 23, 1513.
- Hoch, J. C., Dobson, C. M. & Karplus, M. (1982) *Biochemistry* 21, 1118.
- Hoch, J. C., Dobson, C. M., & Karplus, M. (1985) *Biochemistry* 24, 3831.
- Hubbard, S. J., & Thornton, J. M. (1993) NACCESS, Computer program, Department of Biochemistry and Molecular Biology, University College, London.
- Janin, J., Wodak, S., Lentl, M., & Maignet, B. (1978) *J. Mol. Biol.* 125, 357.
- Jardetzky, O., & Roberts, G. C. K. (1981) *NMR in Molecular Biology*, Academic Press, New York.
- Kabsch, W., & Sander, C. (1983) *Biopolymers* 22, 2577.
- Kaptein, R., Boelens, R., Scheek, R. M., & van Gunsteren, W. F. (1988) *Biochemistry* 27, 5389.
- Karplus, M. (1959) *J. Chem. Phys.* 30, 11.
- Karplus, M., & McCammon, J. A. (1981) *CRC Crit. Rev. Biochem.* 9, 293.
- Kördel, J., & Teleman, O. (1992) *J. Am. Chem. Soc.* 114, 4934.
- Kraulis, P. (1991) *J. Appl. Crystallogr.* 24, 946.
- Kumar, A., Wagner, G., Ernst, R. R., & Wüthrich, K. (1981) *J. Am. Chem. Soc.* 103, 3654.
- Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc.* 104, 4546.
- Lipari, G., Szabo, A., & Levy, R. M. (1982) *Nature* 300, 197.
- Mark, A. E., & van Gunsteren, W. F. (1992) *Biochemistry* 31, 7745.
- Mark, A. E., van Helden, S. P., Janssen, L. H. M., Smith, P. E., & van Gunsteren, W. F. (1994) *J. Am. Chem. Soc.* 116, 6293.
- McGregor, M., Islam, S., & Sternberg, M. J. E. (1987) *J. Mol. Biol.* 198, 295.
- Nagayama, K., & Wüthrich, K. (1981) *Eur. J. Biochem.* 115, 653.
- Olejniczak, E. T., Dobson, C. M., Karplus, M., & Levy, R. M. (1984) *J. Am. Chem. Soc.* 106, 1923.
- Palmer, A. G., & Case, D. A. (1992) *J. Am. Chem. Soc.* 114, 9059.
- Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol.* 180, 741.
- Pedersen, T. G., Sigurskjold, B. W., Andersen, K. V., Kjaer, M., Poulsen, F. M., Dobson, C. M., & Redfield, C. (1991) *J. Mol. Biol.* 218, 413.
- Post, C. B. (1992) *J. Mol. Biol.* 224, 1087.
- Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature* 358, 302.
- Ramanadham, M., Sieker, L. C., & Jensen, L. H. (1987) *Acta Crystallogr. A* 43, (Suppl.), 13.
- Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys.* 23, 327.
- Smith, L. J., & Dobson, C. M. (1995) *Int. J. Quantum Chem.*, in press.
- Smith, L. J., Sutcliffe, M. J., Redfield, C., & Dobson, C. M. (1991) *Biochemistry* 30, 986.
- Smith, L. J., Sutcliffe, M. J., Redfield, C., & Dobson, C. M. (1993) *J. Mol. Biol.* 229, 930.
- Smith, P. E., van Schaik, R. C., Szyperski, T., Wüthrich, K., & van Gunsteren, W. F. (1995) *J. Mol. Biol.* 246, 356.
- Torda, A. E., Scheek, R. M., & van Gunsteren, W. F. (1990) *J. Mol. Biol.* 214, 223.
- Tropp, J. (1980) *J. Chem. Phys.* 72, 6035.
- van Gunsteren, W. F., & Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Nijenborgh 14, 9747 AG Groningen, The Netherlands.
- van Gunsteren, W. F., & Berendsen, H. J. C. (1990) *Angew. Chem., Int. Ed. Engl.* 29, 992.
- van Gunsteren, W. F., Hünenberger, P. H., Kovacs, H., Mark, A. E., & Schiffer, C. A. (1995a) *Proc. R. Soc. London B* 348, 49.
- van Gunsteren, W. F., Hünenberger, P. H., Mark, A. E., Smith, P. E., & Tironi, I. G. (1995b) *Comput. Phys. Commun.*, in press.
- Wagner, G. (1990) *Prog. NMR Spectrosc.* 22, 101.
- Wagner, G. (1993) *Curr. Opin. Struct. Biol.* 3, 748.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic acids*, Wiley, New York.