Structural Determinants of Protein Dynamics: Analysis of ^{15}N NMR Relaxation Measurements for Main-Chain and Side-Chain Nuclei of Hen Egg White Lysozyme†

Matthias Buck,‡ Jonathan Boyd,‡ Christina Redfield,‡ Donald A. MacKenzie,§ David J. Jeenes,§ David B. Archer,§ and Christopher M. Dobson*,‡

New Chemistry Laboratory and Department of Biochemistry, Oxford Centre for Molecular Sciences, University of Oxford, South Parks Road, Oxford OX1 3QT, England, and Agricultural and Food Research Council, Institute of Food Research, Norwich Research Park, Norwich NR4 7UA, England

Received October 19, 1994; Revised Manuscript Received December 15, 1994⊗

ABSTRACT: ^{15}N-labeled hen lysozyme has been studied by 2D and 3D NMR in order to characterize its dynamic behavior. The resonances of all main-chain amide nitrogen atoms were assigned, as were resonances of nitrogen atoms in 28 side chains. Relaxation measurements for the main-chain and arginine and tryptophan side-chain ^{15}N nuclei used standard methods, and those for the ^{15}N nuclei of asparagine and glutamine side chains used pulse sequences designed to remove unwanted relaxation pathways in the NH_2 groups. The calculated order parameters (S^2) show that the majority of main-chain amides undergo only small amplitude librational motions on a fast time scale ($S^2 \geq 0.8$). Increased main-chain motion ($0.5 < S^2 < 0.8$) is observed for a total of 19 residues located at the C-terminus, in loop and turn regions, and in the first strand of the main β-sheet. Order parameters derived for the side chains range from 0.05 to 0.9; five of the six tryptophan residues have high order parameters ($S^2 \geq 0.8$), consistent with their location in the closely packed core of the protein, whereas the order parameters between 0.05 and 0.3 for arginine residues confirm increased side-chain mobility at the protein surface. Order parameters for the side chains of asparagine and glutamine residues range from 0.2 to 0.8; high values are found for side chains that have low solvent accessible surfaces and well-defined χ_1 values, as measured by $^3J_{\alpha\beta}$ coupling constants. Many of the main-chain and side-chain groups with low order parameters have higher than average temperature factors in X-ray crystal structures and increased positional uncertainty in NMR solution structures. They also tend to lack persistent hydrogen bond interactions and protection against amide hydrogen exchange. The most significant correlations are found between residues with low order parameters and high surface accessibility in both crystal and solution structures. The results suggest that a lack of van der Waals contacts is a major determinant of side-chain and main-chain mobility in proteins.

It is now accepted that protein molecules are not rigid entities, but experience a variety of motions covering a wide range of time scales and amplitudes, from fast local oscillations about bonds to slower cooperative motions of segments of the overall structure (Karplus & McCammon, 1983; Gerstein et al., 1994). The manner in which these motions are related to molecular function is a subject of intense interest and some controversy. Molecular motion has been related to binding and catalytic events and to the mode of interaction of proteins with other macromolecules, such as receptors, antibodies, and nucleic acids. In addition to the relationship of molecular motion to function, it is clear that the dynamic properties of proteins are intimately related to their stability and the manner in which they fold (Richards, 1992; Dobson, 1993). Unfolded and partly folded proteins have very extensive mobility compared to the native state, and only when the specific interactions between side chains are established, and close packing of different sections of the polypeptide has been achieved, does the characteristic order of a globular protein structure appear. Information about the dynamic properties of the native as well as denatured states therefore should provide clues concerning the nature of the interactions within proteins and their influence on structure and stability. Despite the importance of these in fields such as protein modeling and design, the structural determinants for protein dynamics are not well established.

Of particular value in the investigation of molecular dynamics is NMR spectroscopy (Dobson & Karplus, 1986; London, 1989). ^{15}N relaxation measurements recently have become an attractive means of studying the main-chain dynamics of proteins in solution (Kay et al., 1989; Clore et al., 1990; Peng & Wagner, 1992). The analysis of the relaxation data in conjunction with the model-free approach of Lipari and Szabo (1982) yields order parameters (S^2) commonly interpreted as a measure of the geometrical restriction of the $^{15}N-^1H$ bond vectors on a time scale faster than the stochastic rotational tumbling of the molecule in solution. Molecular dynamics simulations have reproduced

† Contribution from the Oxford Centre for Molecular Sciences, which is supported by the U.K. Engineering and Physical Sciences Research Council, the Biotechnology and Biological Sciences Research Council, and the Medical Research Council. The research of C.M.D. is supported in part by the award of an International Scholarship from the Howard Hughes Medical Foundation and a Leverhulme Trust Senior Research Fellowship by the Royal Society.

* Author to whom correspondence should be addressed at the New Chemistry Laboratory, University of Oxford.

‡ University of Oxford.

§ Institute of Food Research.

⊗ Abstract published in *Advance ACS Abstracts,* February 15, 1995.

0006-2960/95/0434-4041$09.00/0

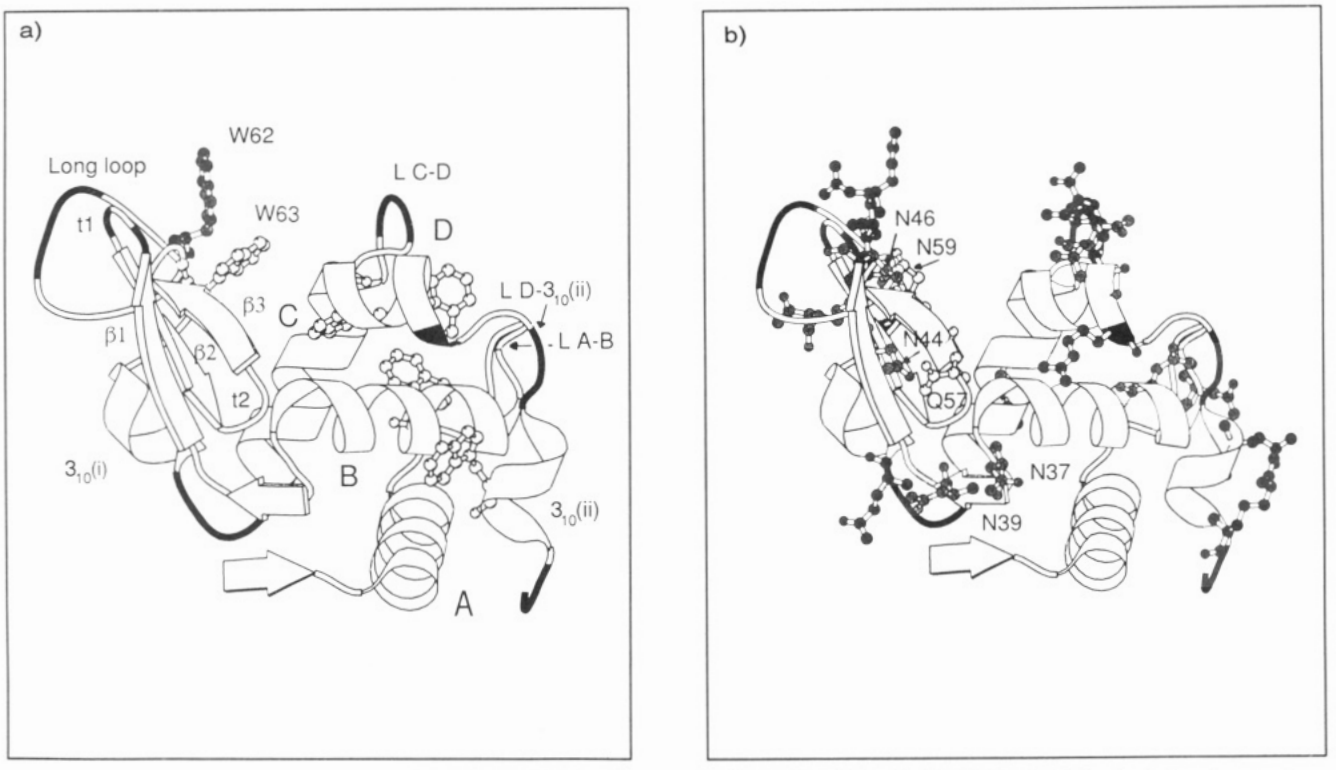

FIGURE 1: Schematic outline of the structure of hen lysozyme drawn using Molscript (Kraulis, 1991). In the helical subdomain, the loops are labeled L. (a) The four α-helices and the N- and C-termini are labeled. The active site is located in the cleft between the α-helical and β-sheet domains. The regions of the main chain and the tryptophan side chain (Trp 62) with order parameters S^2 less than or equal to 0.80 are shaded. (b) The side chains of arginine, asparagine, and glutamine are shaded according to the magnitude of the S^2 for their terminal NH or NH_2 groups: black, $S^2 < 0.4$; gray, $0.4 < S^2 < 0.75$; white, $S^2 > 0.75$. S^2 values for the six asparagine/glutamine side chains located in the active site cleft range from 0.5 to 0.8 (mean S^2 value = 0.65 ± 0.13); S^2 values for the eleven asparagine/glutamine side chains located elsewhere range from 0.2 to 0.7 (mean S^2 value = 0.46 ± 0.19).

experimentally determined order parameters with some success (Olejniczak et al., 1984; Chandrasekhar et al., 1992). The main strategy for the interpretation of order parameters has become focused on their magnitude, their variation along the polypeptide chain, and their relationship with other structural parameters. For example, good correlations between order parameters and *B*-factors of X-ray structures or rmsd values from ensembles of NMR[1] structures have been reported for at least some of the proteins studied (Kördel et al., 1992; Redfield et al., 1992; Grasberger et al., 1993; Powers et al., 1993). Similarly, a tendency of order parameters to be related to secondary structure or hydrogen bonding was suggested in several but not all studies (e.g., Schneider et al., 1992; Kay et al., 1989; Stone et al., 1993).

In order to identify features that may account for dynamic behavior, we have derived order parameters from the ^{15}N relaxation data of a protein for which a wide range of structural data is available. Hen egg white lysozyme was the first enzyme whose three-dimensional structure was determined by X-ray crystallography (Blake et al., 1965). Since then a series of crystal structures showing essentially the same conformation have been refined to high resolution (Handoll, 1985; Ramanadham et al., 1990; Kundrot & Richards, 1991; Wilson et al., 1992; Young et al., 1994). Recently, restraints from 2D ^1H NMR measurements have been used to calculate an ensemble of NMR structures in good agreement with the X-ray structures (Smith et al., 1993). The protein has a globular fold, and its structure, shown schematically in Figure 1, consists of two subdomains, predominantly comprising either α- or β-secondary structure, with the active site cleft at their interface. Although the two subdomains appear to become structured largely independently during kinetic refolding (Radford et al., 1992a), hen lysozyme is known to unfold cooperatively under equilibrium conditions.

The dynamic behavior of hen lysozyme in crystals has been investigated by analysis of *B*-factors. In order to attempt to differentiate between static and dynamic disorder, structures were solved at different temperatures, pressures, and levels of hydration (Kundrot & Richards, 1987; Kodandapani et al., 1990; Young et al., 1994), in different crystal forms (Sternberg et al., 1979; Ramanadham et al., 1989; Young et al., 1994) and of different species (Artymiuk et al., 1979). Mobility in the native state of the protein has also been inferred from changes in the conformation of the enzyme on antibody or ligand binding (Sheriff et al., 1987; Strynadka & James, 1991). However, a comparison of hydrogen exchange behavior in the crystalline state and in solution (Pedersen et al., 1991) has suggested that at least some fluctuations may be significantly dampened by crystal packing contacts.

[1] Abbreviations: 2D and 3D, two- and three-dimensional; COSY, *J*-correlated spectroscopy; CPMG, Carr−Purcell−Meiboom−Gill; CSA, chemical shift anisotropy; HEWL, hen egg white lysozyme (EC 3.2.1.17); HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; R_1, longitudinal relaxation rate ($=1/T_1$); R_2, transverse relaxation rate ($=1/T_2$); rmsd, root-mean-squared deviation; std, standard deviation.

NMR spectroscopy has also been used to investigate the dynamic behavior of hen lysozyme. Structural variation among the members of the ensemble of NMR structures arises as a consequence of insufficient or conflicting experimental restraints, some of which reflect the dynamic behavior of the protein in solution while some reflect limitations in the experimental ^1H NMR methods for a protein of this size (Smith et al., 1993). Direct evidence for structural dynamics was obtained by detection of aromatic ring flips (Campbell et al., 1975), ^1H cross-relaxation dynamics (Olejniczak et al., 1984), analysis of hydrogen exchange (Radford et al., 1992b; Thomson & Poulsen, 1993), and measurement of coupling constants (Smith et al., 1991; Bartik et al., 1993). These NMR studies concluded that the main chain of hen lysozyme is well ordered, but that side chains, particularly at the surface of the protein, are undergoing fluctuations that are significantly greater than is suggested by their definition in crystal structures (Smith et al., 1991).

Parameters calculated from molecular dynamics simulations of hen lysozyme have been compared with experimental data, such as diffuse X-ray scattering (Clarage et al., 1992; Faure et al., 1994) and NMR proton relaxation measurements (Olejniczak et al., 1984). Such simulations indicate that fluctuations of the loop regions are significantly greater than those of the helices and β-sheet (Post et al., 1989). In addition, computational techniques have been used to investigate the possibility of hinge bending (Bruccoleri et al., 1986; Gibrat & Go, 1990). Its existence was suggested by the small reorientation of the subdomains in crystals upon substrate binding (Imoto et al., 1972), but little additional evidence has been found to support such motion in hen lysozyme.

In order to investigate these phenomena in more detail, in this paper we report relaxation measurements and order parameters for the main-chain ^{15}N nuclei of lysozyme and for a comprehensive set of side-chain ^{15}N nuclei. Many of these side chains are located at the protein surface. The relaxation data are compared with the rmsd values derived from the ensemble of NMR structures and *B*-factors from several crystal structures. Hydrogen bonding and surface accessibility are investigated as possible primary determinants of the mobility of NH and NH_2 groups.

MATERIALS AND METHODS

Collection of NMR Data. Hen egg white lysozyme was expressed in *Aspergillus niger* using $^{15}NH_4Cl$ as the sole nitrogen source and purified from filtered culture medium (Roberts et al., 1992) with the following modifications. *A. niger* transformant B1, an improved lysozyme production strain, was used. This strain has been transformed with a glucoamylase−lysozyme fusion vector (Jeenes et al., 1993) and produces about 60 mg of ^{15}N-labeled lysozyme/L. Prior to FPLC purification on Mono S, lysozyme was isolated from filtered culture medium using ion exchange chromatography on ProductiveTMS cartridge columns (BPS Separations Ltd., Durham, UK). The NMR samples were prepared to contain ca. 4 mM protein in 95% H_2O/5% D_2O (v/v), and the pH was adjusted to 3.8. NMR experiments were performed at 35 °C on three homebuilt NMR spectrometers belonging to the Oxford Centre for Molecular Sciences with ^1H operating frequencies of 500.1, 600.2, and 750.2 MHz. The ^{15}N assignments were made using data recorded from a single

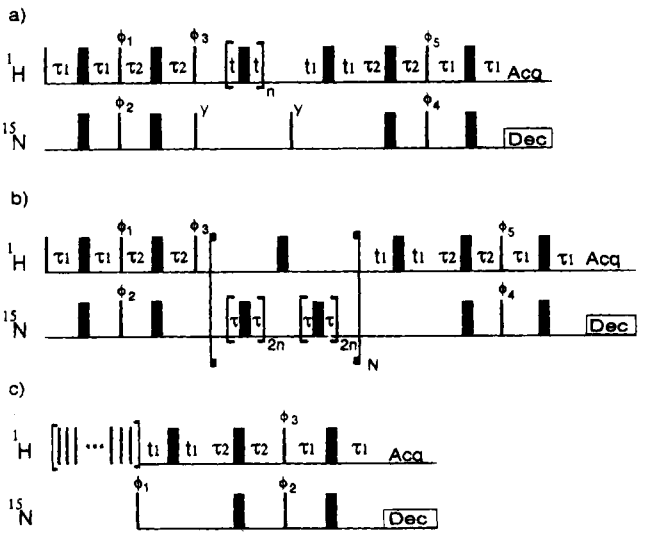

FIGURE 2: Pulse schemes for the measurement of (a) T_1, (b) T_2, and (c) $^{15}N-^1H$ NOE relaxation parameters of ^{15}N nuclei in asparagine and glutamine side chains. All π pulses have phase $+y$. The phase cycling used for sequences a and b was $\varphi_1 = 8(y)$, $8(-y)$, $\varphi_2 = 4(x)$, $4(-x)$, $\varphi_3 = 2(x)$, $2(-x)$, $\varphi_4 = (x)$, $(-x)$, and $\varphi_5 = 16(y)$, $16(-y)$, and the acquisition was $2(x, -x)$, $4(-x, x)$, $2(x, -x)$. The receiver phase was further inverted every 16 scans. Sign discrimination in F_1 was achieved by incrementing the phase of φ_4. The ^1H π pulses applied during the relaxation periods of sequences a and b serve to remove the effects of $^{15}N-^1H$ CSA−dipole cross-correlations (Palmer et al., 1992; Kay et al., 1992). Solvent suppression, where used, was achieved with a combination of RF pulse presaturation and trim pulses (Messerle et al., 1989). For the $^{15}N-^1H$ NOE experiment, no RF presaturation was used. φ_2 was incremented for F_1 sign discrimination, and the phase cycling closely followed that described in Kay et al. (1989). The delays τ_1 and τ_2 were set to 2.45 and 1.38 ms, respectively for all experiments. τ in sequence b was typically set to 300 μs.

3D NOESY-HMQC experiment (Messerle et al., 1989) with a mixing time of 200 ms, in conjunction with the known ^1H assignments (Redfield and Dobson, 1988). The data set comprised (t_1, t_2, t_3) 128, 32, and 512 complex points recorded with digital resolution of 42.7, 37.8, and 12.4 Hz/point, respectively, at 500 MHz and was zero filled to 256, 64, and 1024 real points.

Pulse sequences for the measurement of the longitudinal (T_1) and transverse (T_2) relaxation times and the heteronuclear nuclear Overhauser effects (NOEs) of ^{15}N nuclei in NH groups have been described previously (Kay et al., 1989; Boyd et al., 1990; Palmer et al., 1992). Suppression of the solvent resonance was achieved by low-power presaturation or by exploiting RF inhomogeneity (Messerle et al., 1989). The T_1 and T_2 relaxation measurements used a series of nine experiments with relaxation delays ranging from 20 to 1400 ms and from 8.5 to 340 ms, respectively. A recycle delay of 4.0 s was used for the T_1 relaxation and NOE experiments and of 2.4 s for the T_2 experiments. The data sets were acquired using 128 complex data points in T_1 with resolution of 9.5 Hz/point.

Experiments involving $^{15}NH_2$ resonances used the pulse sequences described in Figure 2. These data were recorded at 600 MHz with 64 complex data points in t_1, giving a resolution of 11.4 Hz/point. The T_1 and T_2 data sets were recorded with a digital resolution in t_2 of 6.2 Hz/pt, and the NOE experiments were recorded using a digital resolution of 7.4 Hz/pt. For the NOE experiments, where the dynamic range is large, 4 times oversampling in F_2 was employed.

The 2D spectra were zero filled once in F_2 and twice in F_1 and were processed with Felix (Hare Research and Biosym, Inc.). The residual water signal was removed during data processing by the application of a low-frequency deconvolution in t_2 (Marion et al., 1989).

Analysis of NMR Relaxation Data. The heteronuclear NOE effect was calculated as the ratio of peak heights in spectra recorded with and without ^{1}H saturation. Rates R_1 ($=1/T_1$) and R_2 ($=1/T_2$) were fitted as single-exponential decays to peak height data; uncertainties were estimated using published procedures (Palmer et al., 1991; Stone et al., 1992). The errors in the primary intensity data were taken as 1−3 times the spectral noise. The ^{15}N relaxation parameters obtained were analyzed using the formalism of Lipari and Szabo (1982) with the extension of Clore et al. (1990). In the case of analysis of the relaxation parameters for ^{15}N in NH_2 groups, the dipolar contribution to spectral density functions was increased by a factor of 2 to account for the presence of two ^{1}H nuclei. The appropriate model for the spectral density function used to calculate T_1, T_2, and NOE was selected according to the criteria of Clore et al. (1990) on the basis of the observed T_1/T_2 ratio and the magnitude of the NOE effect. Parameters calculated using the appropriate spectral density function, S^2, τ_e, and Δex for simple models and S_s^2, S_f^2, and τ_s for the extended model, were optimized by minimization of the rmsd between the calculated and the experimental relaxation data using published procedures (Clore et al., 1990; Palmer et al., 1991; Stone et al., 1992; Constantine et al., 1993). The order parameters and their uncertainties were estimated by Monte Carlo sampling using MODELFREE2.1 (Arthur G. Palmer, Columbia University) and in-house software.

Analysis of Crystal and NMR Structures. Six crystal structures of hen lysozyme were analyzed, five of which were obtained from the Brookhaven Protein Databank. 2LYM (Kundrot & Richards, 1987), and 2LZT (Ramanadham et al., 1990) are tetragonal and triclinic structures both at 2.0 Å resolution; 4LYT and 6LYT (Young et al., 1994) are monoclinic and tetragonal structures at 1.9 Å resolution; and 1HEL (Wilson et al., 1992) is a tetragonal lysozyme structure solved to 1.7 Å resolution. Hentyp2 is a tetragonal structure refined to 2 Å by Handoll (1985) on the basis of the original data of Blake et al. (1967) and has been used frequently for comparisons with NMR data. The *B*-factors of individual nitrogen atoms were used for comparison of main-chain and tryptophan side-chain dynamics in crystals with the ^{15}N relaxation data. The positioning of arginine, asparagine, and glutamine side-chain atoms can be more ambiguous in X-ray structures, and the average of *B*-factors for the groups of $N^{\epsilon}H + (N^{\eta}H_2)_2$ and $O + N^{\delta/\epsilon}H_2$ atoms was used for arginine $N^{\epsilon}H$ and for asparagine and glutamine $N^{\delta/\epsilon}H_2$ side-chain groups, respectively. Global and local (superposition window size of 2−5 residues) rmsd values were calculated for the main-chain superposition of the six crystal structures and for the ensemble of NMR structures (Smith et al., 1993).

Accessible surface area of the heavy atoms in the structures was calculated using the van der Waals radii of Chothia (1976) and a 1.4 Å radius probe in an implementation of the algorithm of Lee and Richards (1971) as part of the program NACCESS (Dr. Simon Hubbard, EMBL, Germany, and Prof. Janet Thornton, University College, London, UK). All hydrogen bond interactions involving a particular atom were determined in the crystal structures, and their energies were calculated explicitly using X-PLOR (Brünger, 1992). When several hydrogen bonds were found to a particular atom, their energies were summed. A stable hydrogen bond is taken to exist if this energy is more negative than −0.5 kcal mol^{-1} (Kabsch & Sander, 1983). Hydrogen bond interactions are considered a barrier to mobility and therefore are plotted as an "activation energy" term, exp(H-bond energy), throughout, giving better correlations with S^2 than the linear relationship. The peptide bond unit (C_{i-1}, O_{i-1}, and N_i) and the groups of side-chain atoms $O + N^{\delta/\epsilon}H_2$ for asparagine and glutamine and $N^{\epsilon}H + (N^{\eta}H_2)_2$ for arginine comprise structural units whose mobility is likely to be determined in a concerted manner by the surrounding environment. For example, hydrogen bonding with one $N^{\eta}H_2$ group of an arginine residue is likely to affect the mobility of the entire side-chain end unit, but the effects on accessibility and hydrogen bond energy are predominantly on one of the three donating nitrogen atoms involved. Therefore, the absolute accessibilities (in Å^2) and hydrogen bond energies are reported as unweighted averages over these atom groups. This grouping yielded better correlations with S^2 than the parameters of the single atoms. The *B*-factors, accessibilities, and hydrogen bond energies were calculated for individual structures and then averaged to yield their mean values over the six crystal structures and over the ensemble of NMR structures.

RESULTS

^{15}N Resonance Assignments. ^{1}H NMR assignments previously reported for hen lysozyme (Redfield & Dobson, 1988) were used together with a 2D ^{1}H COSY experiment as the starting point for ^{15}N resonance assignment. All of the remaining information was provided by a single 3D NOESY-HMQC spectrum (Messerle et al., 1989), which allowed connectivities to be traced along the entire main chain using nearest neighbor NOE connectivities. Assignments for the ^{15}N−^{1}H correlations for the 126 main-chain amides and for the six tryptophan indole ^{15}N−^{1}H correlations in the HSQC spectrum of hen lysozyme are shown in Figure 3. Additional resonances in the HSQC spectrum belong to the side chains newly assigned in this study; these include pairs of resonances in the region around 112 ppm, which arise from side-chain $^{15}NH_2$ groups of the 14 asparagine and 3 glutamine residues in hen lysozyme. Resonances in the region boxed around 108 ppm are folded into the spectrum and arise from arginine $^{15}N^{\epsilon}$−^{1}H side-chain correlations. These residues could be assigned in the 3D spectrum by their NOEs to main-chain and side-chain protons of their own and neighboring residues, and in the case of five arginines also by the use of through-bond information. None of the resonances arising from arginine $(^{15}N^{\eta}-^1H_2)_2^+$ side-chain correlations is resolved, and they are likely to correspond to the broad resonances near 118 ppm that are folded in the HSQC spectrum. [Table 1S, in the supplementary material, presents a full listing of the ^{15}N and additional ^{1}H assignments made in this study.] The ^{1}H chemical shifts of resonances identified in the spectra are in agreement with those reported previously (Redfield & Dobson, 1988) to within ±0.06 ppm. Assignments for main-chain ^{15}NH and asparagine and glutamine side-chain $^{15}NH_2$ resonances have also been reported for human lysozyme by Ohkubo et al. (1991). As has been found for ^{1}H chemical shifts (Redfield & Dobson, 1991), there is good correlation between ^{15}N main-chain and side-

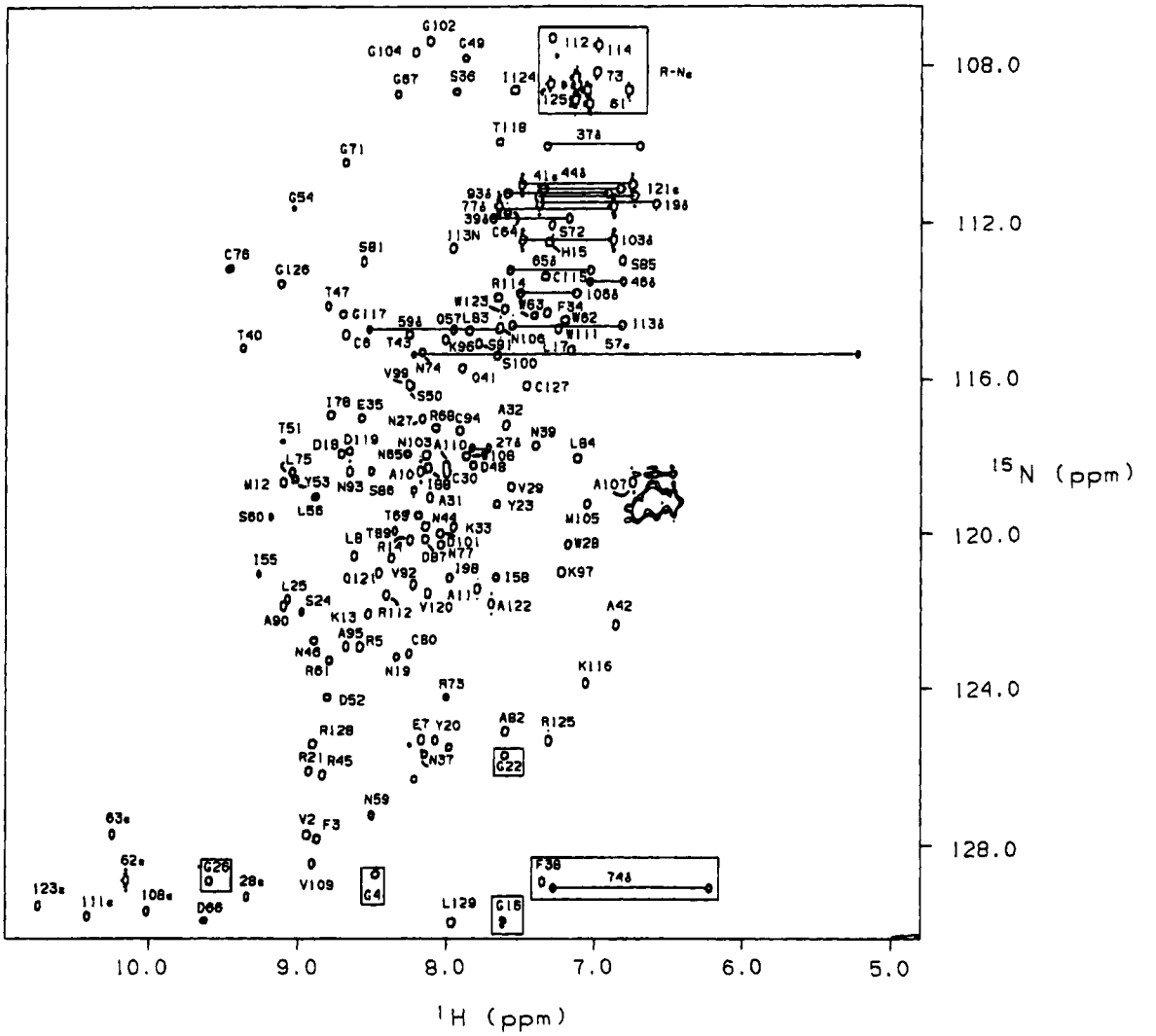

FIGURE 3: 500 MHz $^{15}N-^{1}H$ HSQC spectrum of uniformly ^{15}N-labeled hen lysozyme in H_2O at 35 °C and pH 3.8. All of the expected 126 main-chain $^{1}H-^{15}N$ correlations are labeled, together with the six tryptophan $^{15}N^{\epsilon}-^{1}H$ and pairs of the resonances of the seventeen asparagine/glutamine $^{15}N^{\delta/\epsilon}-^{1}H_2$ side-chain groups. The five arginine $^{15}N^{\epsilon}-^{1}H$ resonances that could be assigned are labeled. Cross-peaks shown in boxes are contoured at negative levels as they are folded in from outside the region of the spectrum displayed. Unlabeled peaks are low-level impurities. Cross-peaks C30/A110, S50/V99, and G16/impurity overlap significantly and were not used for subsequent analysis of ^{15}N relaxation.

chain chemical shifts, particularly for the residues that are conserved in these two homologous proteins.

Relaxation Data and Order Parameters for Main-Chain ^{15}N Nuclei. Relaxation parameters were measured for 121 of the 129 main-chain ^{15}N nuclei of hen lysozyme (Figure 4). A value for the global rotational correlation time, τ_R, was obtained from the average T_1/T_2 ratio (Clore et al., 1990) of 100 main-chain ^{15}N nuclei, which did not show evidence for significant fast motion or exchange broadening. The mean T_1/T_2 ratio was 3.32 ± 0.13 and yielded an optimized value for τ_R of 5.7 ± 0.2 ns. This value is consistent with previous experimental data for hen lysozyme, after correction for temperature and viscosity effects (Olejniczak et al., 1984), and compares well with the τ_R value of 6.0 ns that can be calculated for hen lysozyme under these experimental conditions (Cantor & Schimmel, 1980).

The main-chain order parameters, S^2, the internal correlation times, τ_e or τ_s, and the magnitude of the exchange effects, Δex, are plotted as a function of the protein sequence in Figures 4 and 5 (all experimental and calculated data are given in the supplementary material, Table 2Sa,b). The uncertainties estimated for these parameters are found to be small for S^2 (<4%), larger for Δex (typically 10−30%), and much larger for certain τ_e values (several exceeding 100%). These estimated uncertainties correspond in magnitude to those reported for other proteins (Palmer et al., 1991; Constantine et al., 1993). It is notable, in comparing Figures 4 and 5, that the profile of S^2 in the range 0.75−0.95 is predominantly determined by the variations in the T_1 and T_2 data, the NOE value being uniformly high. By contrast, values of S^2 below 0.75 reflect slower motions of increased amplitude, which result in a reduced magnitude for the NOE. Residues with such behavior are not randomly distributed in the protein (Figure 4). It is noticeable that residues undergoing motions described by similar values of the internal correlation time, τ_e, are clustered in the A-helix, A-helix−loop−B-helix, and B-helix region as well as the C-helix and loop-to-D-helix region, all located in the α-subdomain of the protein (Figure 1). Residues in the latter region are undergoing motions of increased amplitude, and one residue (Asn 103) undergoes motions that require an extended form of the model for the spectral density function

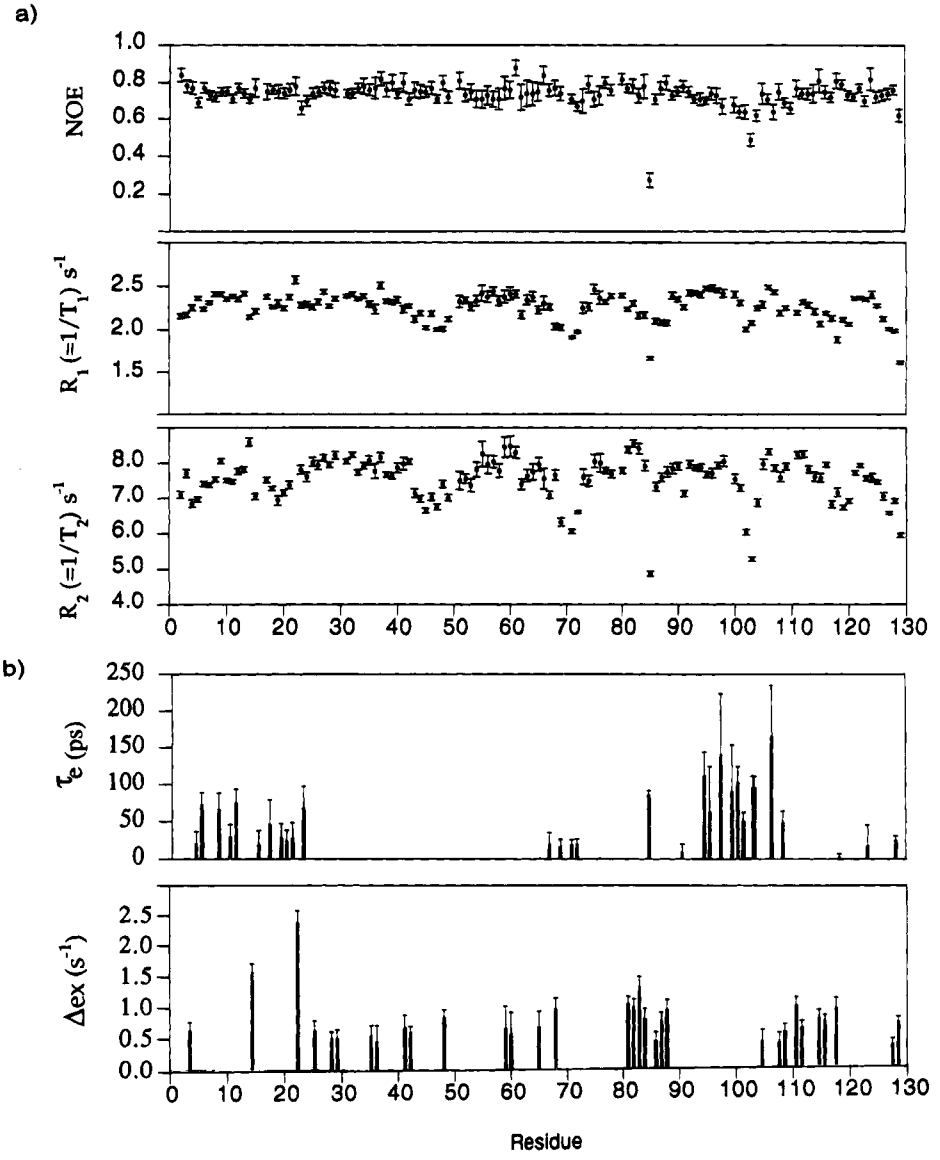

FIGURE 4: (a) NOE, R_1 (=1/T_1), and R_2 (=1/T_2) values for main-chain amide nitrogens and (b) τ_e and Δex values are each plotted versus lysozyme sequence. The R_2 value for residue 22 in plot a is 10.8 s^{-1}. The value of τ_s for Asn 103 is 1708 ± 186 ps. Values of −160 ppm for the CSA constant (Stone et al., 1992; Schneider et al., 1992) and 5.7 ± 0.2 ns for the rotational correlation time (τ_R) were used in the calculations. All order parameters in this study were calculated with the assumption of fixed and equal NH bond lengths of 1.02 Å throughout the protein. A lengthening of the bond by 1% would cause an increase of approximately 6% in the value of S^2; such vibrational lengthening may itself be part of internal motion (Kay et al., 1989; Schneider et al., 1992).

by Clore et al. (1990) involving internal motions on two time scales. Furthermore, motions on a microsecond to millisecond time scale, which result in an exchange contribution to relaxation, are clustered in the region of the first 3_{10}-helix and the D-helix. Such clustering is indicative of the regional nature of the environments of the polypeptide chain and may, at least in part, reflect segmental flexibility. Residues 22 and 85 have order parameters significantly different from those of their neighbors. At present, the structural reasons for their relaxation behavior are not understood. However, it is interesting to consider that both are located near the termini of helices.

The order parameters, S^2, derived from the analysis of the main-chain ^{15}N relaxation data are greater than 0.80 for the majority of residues (Figures 1 and 6). The protein backbone is, therefore, well ordered except for a few main-chain ^{15}N atoms that undergo significant motions associated with S^2 below 0.70 for a few residues in hen lysozyme. These are located in the long loop (residues 65−75) and in a short loop connecting the C- and D-helices (residues 101−104), the C-terminal residue, and residue 85. The lowest order parameters are around 0.5. Therefore, lysozyme does not belong to the class of proteins that includes, for example, staphylococcal nuclease (Kay et al., 1989) and that lacks any appreciable internal mobility; with the exception of the residues in the termini, no value of S^2 for the latter protein is below 0.8. Lysozyme can best be compared to proteins such as ubiquitin, thioredoxin, and interleukin-8 [Schneider et al. (1992), Stone et al. (1993), and Grasberger et al. (1993), respectively], in which several residues, usually located in loop regions, are undergoing motions of moderate amplitude (lowest value of S^2 usually >0.5). By contrast, in several proteins, for example, interleukin-1β, interleukin-4, calbindin, calmodulin, and eglin *c* (Clore et al., 1990; Redfield et al., 1992; Kördel et al., 1992; Barbato et al., 1992; Peng & Wagner, 1992), continuous stretches of residues are found

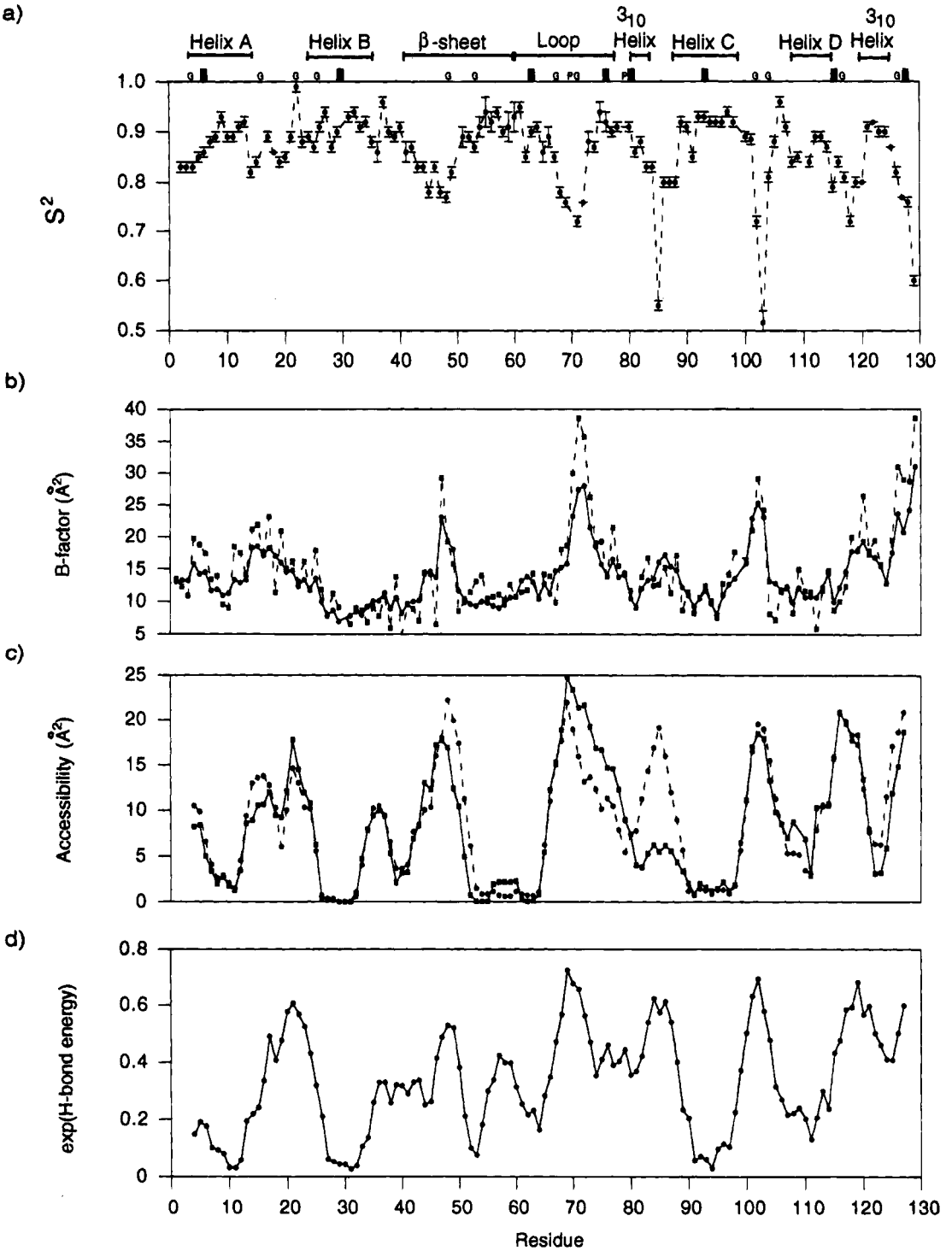

FIGURE 5: (a) Main-chain S^2 values and their uncertainties, calculated as described in the text, as a function of sequence. The units of protein secondary structure are shown, and the positions of the cysteine residues, involved in disulfide bonds, the glycine, and the proline residues are indicated as ■, G, and P, respectively. (b) *B*-factors of main-chain nitrogen atoms in the hentyp2 crystal structure (dashed lines) and as the average of the six hen lysozyme crystal structures, which are listed in the Materials and Methods section. (c) Mean peptide unit accessibilities of the six crystal structures (continuous lines) and the members of the ensemble of NMR structures (Smith et al., 1993) (dashed lines). (d) Mean of hydrogen bond energies for the peptide units in the six crystal structures, plotted in exponential form. The data in (c) and (d) were smoothed (unweighted) by averaging data points over a sliding window of five residues.

with values of S^2 close to or less than 0.5, substantially reduced from those of the remainder of the protein.

Proteins with regions characterized by low order parameter values all possess long, disordered loops. Correlations of S^2 with other structural parameters have been made most convincingly for this class of protein. By contrast, most of the loops and turns in hen lysozyme are short. Turns joining strands of the β-sheet and loops joining the two 3_{10}-helices to other structural elements are four residues long, and the loops that join α-helices A−B and C−D are seven and eight residues long, respectively. There is, however, one long loop comprising residues 61−78 in hen lysozyme. Of these loops, only the C−D loop and the center of the long loop appear to undergo significant fluctuations on a time scale faster than

Table 1: Summary of Order Parameters and Structural Data for Side Chains[a]

residue	S^2	S_f^2, S_s^2	τ_e (τ_s) (ps)	Δex (s^{-1})	*B*-factor ($Å^2$)[f]	accessibility ($Å^2$)[f]	χ_1[g]	exp(H-bond E)[h]
Trp[b]								
28	0.90 ± 0.01		49.73 ± 10.54		11.8	0.0	no	0.67
62	0.41 ± 0.01	0.73, 0.56	(1317.1 ± 31.90)		42.4	19.6		1.00
63	0.88 ± 0.01		46.03 ± 14.82	0.84 ± 0.17	19.4	11.4		1.00
108	0.87 ± 0.01		65.59 ± 10.36		15.8	2.8	no	0.63
111	0.88 ± 0.01			0.83 ± 0.16	10.9	0.0	no	0.68
123	0.85 ± 0.01				13.4	10.6	no	0.67
Arg[c]								
61	0.28 ± 0.02		74.96 ± 8.69		40.2	17.7	no	0.53
73	0.12 ± 0.01		101.16 ± 12.63		78.9	36.2	yes	0.93
112	0.31 ± 0.04		72.51 ± 13.24	1.59 ± 1.40	46.4	23.4	yes	0.33
114[d]	0.27 ± 0.02		34.92 ± 4.16	0.52 ± 0.53	13.2	28.9	no	0.97
125	0.05 ± 0.01		120.53 ± 6.15	2.37 ± 0.97	57.9	19.4	yes	0.89
Asn/Gln[e]								
N19	0.43 ± 0.00		24.74 ± 0.92		43.2	27.2	yes	1.00
N27	0.72 ± 0.08			2.82 ± 1.35	12.2	8.8	no	0.56
N37	0.51 ± 0.04		20.02 ± 1.94		28.3	29.8	yes	0.80
N39[d]	0.74 ± 0.02			1.29 ± 0.37	12.0	20.7	no	0.11
Q41[d]	0.19 ± 0.01	0.51, 0.38	(228.2 ± 12.7)		15.6	41.2	no	0.78
N44[d]	0.51 ± 0.01		25.70 ± 1.86		27.8	31.5	yes	0.31
N46	0.62 ± 0.02			1.88 ± 0.48	28.5	15.1	no	0.26
Q57	0.82 ± 0.04				10.9	1.9	no	0.30
N59	0.78 ± 0.13			4.68 ± 2.25	17.8	4.4	no	0.01
N65	0.57 ± 0.01		11.69 ± 2.87		23.8	30.3	no	0.74
N74	0.74 ± 0.03			1.75 ± 0.57	17.4	9.7	no	0.58
N77	0.24 ± 0.01	0.57, 0.42	(389.2 ± 36.2)		40.2	39.6	yes	0.88
N93	0.59 ± 0.01		26.79 ± 3.13	0.65 ± 0.21	25.7	22.6	no	0.98
N103	0.26 ± 0.02	0.56, 0.47	(699.5 ± 57.8)		50.4	41.6	yes	0.78
N106[d]	0.58 ± 0.02		21.80 ± 4.10	0.82 ± 0.29	13.4	14.2	no	0.61
N113[d]	0.47 ± 0.01	0.71, 0.66	(397.5 ± 31.8)		16.2	24.0	yes	1.00
Q121	0.36 ± 0.00		55.73 ± 1.05		48.5	40.6	yes	1.00

[a] Data at pH 3.8, 35 °C. Hydrogen exchange rates for the side chains are estimated to be between 0.25 and 0.7 s^{-1} (Krishna et al., 1982; Wedin et al., 1982; Bai et al., 1993) and are too slow to cause significant errors in the relaxation measurements. [b] Calculated using a CSA constant of −89 ppm (Opella & Cross, 1983). [c] The CSA constant is not known. A value of −160 ppm was used in the calculation. S^2 values vary by ±20% with values of −90 and −210 ppm, respectively. [d] Side-chain atom groups involved in crystallographic contacts. [e] A CSA constant of −163 ppm was used; this was calibrated for NH_2 in asparagine (Herzfield et al., 1987) and was presumed to be identical for glutamine. The differences in derived parameters, fitted to data for the two components of the $^{15}N-^{1}H$ correlation separately, are small and are shown as averages. [f] Hentyp2 crystal structure. [g] A yes indicates that the side chain is found in a single χ_1 rotamer state, and a no indicates that averaging about χ_1 has been detected by NMR (Smith et al., 1991; Bartik et al., 1993). [h] Mean hydrogen bond energy of the six crystal structures.

that of the overall rotational tumbling of the protein. Hinge bending fluctuations modeled by dynamics simulations (Bruccoleri et al., 1986; Gibrat & Go, 1990) have periods of approximately 20 ps and are within the frequency range contributing to the NMR relaxation parameters studied here. The amplitudes of the fluctuations are small, however, and predicted to be less than 3°. Such an angle would cause a maximal atomic displacement of 1 Å. Theoretical analysis suggests that the order parameters described here would not be sufficiently sensitive to monitor such small hinge bending dynamics, which are obscured by more local fluctuations (X. Li and R. Brüschweiler, personal communication).

Relaxation Data and Order Parameters for Side Chains. Relaxation data for ^{15}N nuclei in tryptophan and arginine side-chain NH groups were extracted from the same spectra using the same procedure that was used for the main-chain NH groups (supplementary material Table 3). The data for the tryptophan $^{15}N^{\epsilon}$ atoms are similar to those obtained for the majority of main-chain atoms, with the exception of Trp 62, which has a lower NOE value of 0.25 ($S^2 = 0.56$). NOE ratios for five arginine $^{15}N^{\epsilon}$ atoms are, in contrast to those of tryptophan side chains, significantly smaller, with all values being less than 0.04 and in some cases negative, such as Arg 125 with the lowest NOE value in this study at −2.8. In addition, T_1 and T_2 relaxation time constants are longer than those for typical main-chain nuclei, suggesting that the arginine side chains are undergoing considerable fluctuations. Order parameters for these side chains are shown in Table 1.

The relaxation of ^{15}N nuclei in NH_2 groups of asparagine and glutamine side chains is more complicated than that of ^{15}N in NH groups, and pulse sequences designed to monitor selectively the ^{15}N longitudinal and transverse relaxations of these side-chain nuclei were used in their measurement (see Materials and Methods, Figure 2, and Table 1). For this purpose, a simplified model was adopted to describe the relaxation properties of the ^{15}N nucleus in the asparagine and glutamine side chains. Justifications for the use of pulse sequences in Figure 2 have recently been presented (Boyd, 1995). The ^{15}N nucleus is considered to be part of an isolated, weakly coupled three-spin system wherein the two heteronuclear one-bond coupling constants and bond lengths are assumed equal. However, the two protons have distinct chemical shifts, giving rise to an I_1I_2S spin system where $J(I_1I_2) \ll \Delta\nu(I_1I_2)$. The relaxation mechanisms that must be considered in addition to those possible in NH spin systems include dipolar auto- and cross-correlation, chemical shift anisotropy of the ^{15}N nucleus (including CSA−dipole cross-correlation), and a chemical exchange term defining the intramolecular bond rotation. The intramolecular chemical exchange is not anticipated to contribute to the relaxation of either the single spin transverse or longitudinal processes. The importance of cross-correlation terms, both ^{15}N CSA−dipole and dipole−dipole, must, however, be considered. The

Table 2: Correlation Coefficients for Comparisons between Structural Parameters

	Crystallographic versus NMR Structures of Hen Lysozyme[a]				
	B-factor	rmsd (NMR)	accessibility (crystal)	accessibility (NMR)	exp(H-bond E) (crystal)
B-factor	0.84	0.46	0.59	0.58	0.57
rmsd (NMR)	0.50	N/A	0.38	0.54	0.45
accessibility (crystal)	0.76	0.50	0.96	0.84	0.57
accessibility (NMR)	0.74	0.69	0.71	0.89	0.58
exp(H-bond E) (crystal)	0.71	0.61	0.74	0.76	0.94

	Crystallographic and NMR Structures versus S^2 [b]		
comparison of S^2 with	individual residues (121)	smoothed window of five residues	average of structural units (17)
B-factor (crystal)[c]	0.53	0.53	0.63
B-factor mean (crystal)	0.65	0.69	0.77
rmsd (NMR)	0.53	0.68	0.62
exp(H-bond E)	0.50	0.67	0.69
accessibility			
single crystal[c]	0.52	0.66	0.85
mean crystal	0.53	0.66	0.85
NMR ensemble	0.62	0.81	0.92

[a] Diagonal: average correlation coefficient for the comparisons of parameters of individual structures with the parameters averaged over the structures. Above diagonal: comparisons of the six crystal structure or NMR ensemble averaged structural parameters. Below diagonal: comparison of parameters smoothed over a window of five residues. [b] Statistical significance (at 5% critical value) requires correlation coefficients of greater than 0.14 for 121 data points (residues) and 0.41 for 17 data points (structural units) (Hoel, 1984). The relationships need not be linear, or in the case of hydrogen bond energy, exponential, and there can be no proof of a causal link. [c] Representative data from the hentyp2 crystal structure.

magnitude of these cross-correlations will dictate the extent of the deviation from an exponential longitudinal or transverse relaxation process for the ^{15}N nucleus.

The experimental protocols developed for the suppression of the ^{15}N CSA−dipole cross-correlation process in two-spin heteronuclear systems were found to be effective here also (Kay et al., 1989; Palmer et al., 1992). The deviations from exponential behavior caused by the dipolar cross-relaxation terms (Overhauser effects) are small and may be neglected. Pulse sequences a and b of Figure 2 create spin operators as a result of the dipole−dipole cross-correlation process during the relaxation delay that would, if significant, lead to observable magnetization and amplitude distortion. However, the magnitude of the dipole−dipole cross-correlation process between the pair of internuclear vectors $^{15}NH_1$ and $^{15}NH_2$ is scaled by the angular term $(3\cos^2\theta_{NH_1NH_2} - 1)/2$, which for $\theta_{NH_1NH_2} = 120°$ is 0.125. Simulations using a rotational correlation time of 5.7 ns, appropriate for hen lysozyme, show that this cross-correlation process does not strongly affect the longitudinal or transverse recovery. Therefore, the very small deviations from exponential behavior introduced by this dipole−dipole cross-correlation process were ignored in both cases, and the relaxation was treated as an exponential decay. The order parameters calculated from the data are shown in Table 1; in contrast to the order parameters for the main chain or the tryptophan and arginine side chains, the order parameters for asparagine and glutamine side chains show considerable variation in their values, ranging from 0.16 to 0.82.

The dynamic behavior of the side chains studied here can be grouped into three classes by their S^2 values. The first class of side chains ($S^2 = 0.8-0.9$) comprises five of the six tryptophan and one of the fourteen asparagine side chains. These S^2 values are similar to those obtained for the majority of main-chain residues. Similar results have been reported for the two tryptophan side chains in thioredoxin, which were studied by ^{15}N relaxation (Stone et al., 1993). A second class of side chains ($S^2 = 0.4-0.7$) comprises Trp 62 and the majority of asparagine and glutamine side chains, which are located at or near the protein surface. A third class of side chains undergoes significantly greater fluctuations, having S^2 values of less than 0.4. It comprises those of the remaining two asparagine and two glutamine residues and all arginines. The location of these side chains is shown in the hentyp2 crystal structure in Figure 1; the classification of side chains by their dynamics is indicated by shading. This classification of side-chain dynamics is not, in general, simply by amino acid type. A report of NOE measurements in eglin *c* suggested that two of the four arginine residues in the protein were significantly restrained in their mobility (Peng & Wagner, 1992). However, the data of the present study indicate that all arginine side chains of hen lysozyme for which measurements were made experience considerable fluctuations.

DISCUSSION

Correlation of Main-Chain S^2 Values with Features of X-ray and NMR Structures. Good correlations of main-chain S^2 values with *B*-factors have been reported in several studies (Clore et al., 1990; Kördel et al., 1992; Grasberger et al., 1993; Powers et al., 1993). Figure 5a,b shows the comparison of the experimental order parameters with *B*-factors of the crystal structure of hen lysozyme frequently used in comparison with NMR data (referred to as hentyp2). It is apparent that all regions that have high *B*-factors in this tetragonal crystal structure also have moderately low values of S^2. However, the converse does not appear to hold, as residues 83−88 have low S^2 values but do not appear to have *B*-factors higher than the average. However, *B*-factors vary considerably between the six crystal structures used in this analysis, especially when comparing structures solved in different crystallographic space groups (Table 2a), and appear to depend, at least in part, on the precise experimental conditions. However, several regions of the main chain, for example, the long loop residues 61−78, have increased *B*-factors in the majority of crystal structures. This suggests that *B*-factors do indeed reflect internal dynamics, except

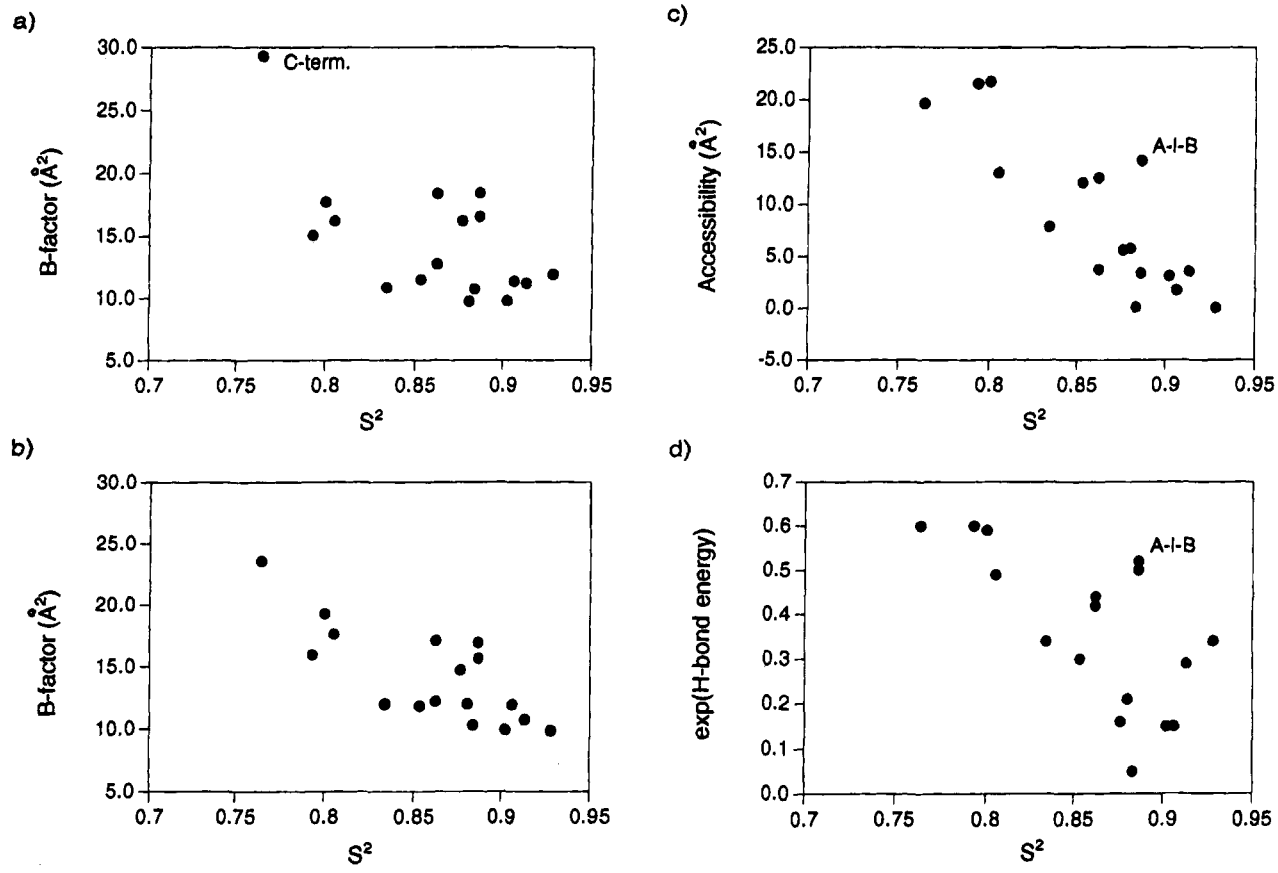

FIGURE 6: Correlations of S^2 values averaged over structure segments with (a) *B*-factors of main-chain nitrogen atoms in the hentyp2 crystal structure; (b) mean of *B*-factors of the six crystal structures; (c) mean peptide unit accessibilities of the six crystal structures; (d) mean hydrogen bond energies of the peptide units in the six crystal structures. The correlation coefficients are given in Table 2.

Table 3: S^2 Values and Structural Parameters Averaged over 17 Different Structural Units of Hen Lysozyme

structural units[a]	residues	S^2	*B*-factor[b] (Å²)	rmsd (NMR)[c] (Å²)	accessibility (crystal)[b] (Å²)	accessibility (NMR)[c] (Å²)	exp(H-bond E)[d]
helix A	4−15	0.88	16.2	1.3	5.7	6.8	0.16
helix B	24−36	0.90	9.8	1.1	3.2	3.7	0.15
helix C	89−99	0.91	11.4	1.0	1.8	1.2	0.15
helix D	108−115	0.85	11.5	1.7	12.1	9.6	0.30
3_{10}(i)	80−84	0.86	12.7	4.4	3.8	11.3	0.42
3_{10}(ii)	120−124	0.89	18.4	2.2	3.4	6.3	0.50
strand 1	41−45	0.83	10.9	1.9	7.9	8.3	0.34
turn 1	46−49	0.80	17.7	5.4	21.8	21.8	0.59
strand 2	50−53	0.88	10.7	2.4	0.1	1.4	0.05
turn 2	54−57	0.93	11.9	1.1	0.1	0.9	0.34
strand 3	58−60	0.91	11.2	1.6	3.6	0.9	0.29
loop A−B	16−23	0.89	16.5	1.3	14.2	11.1	0.52
loop C−D	100−107	0.81	16.2	1.5	13.1	13.7	0.49
long loop[e]	61−78	0.86	18.4	2.3	12.6	10.8	0.44
loop D−3_{10}(ii)	116−119	0.79	15.1	4.1	21.6	21.0	0.60
C-terminus	125−129	0.76	29.3	1.5	19.7	20.9	0.60
small sheet	1−3/37−40	0.88	9.7	2.9	5.8	7.7	0.21

[a] Defined in Handoll (1985). [b] Hentyp2 crystal structure. [c] Ensemble of structures. [d] Mean of the six crystal structures. [e] The average hydrogen bond energy and peptide unit accessibility over the central region of the long loop (residues 68−72) are −0.4 kcal mol^{-1} and 23.4 Å², respectively. The parameters for the terminal residues of the loop (residues 61−67 and 73−78) are −1.0 kcal mol^{-1} and 9.8 Å², respectively. The average residue hydrogen bond energy and the accessibility of peptide units for the entire protein are −1.1 kcal mol^{-1} and 8.4 Å², respectively.

when they are dampened by intramolecular contacts (Artymuik et al., 1979; Handoll, 1985; Smith et al., 1986; Young et al., 1994).

In order to explore this further, a mean *B*-factor was calculated by simply averaging the reported *B*-factors from the six crystal structures, shown in Figure 6b. As shown in Table 2b, such average *B*-factors show a better correspondence with S^2 values than those taken from individual structures. Furthermore, the variations in both S^2 values and *B*-factors are largely regional, suggesting that averaging over the structural segments of hen lysozyme may lead to further improvement in the correlations. The correlation of S^2 with these regional *B*-factors, averaged over 17 structural units of the protein [Handoll (1985) and Table 3], is shown in Figure 6. The rmsd, resulting from a global superposition of the ensemble of NMR structures (Smith et al., 1993), also

correlates well with individual S^2 values. A similar result has been found for other proteins (e.g., Redfield et al., 1992; Peng et al., 1992; Grasberger et al., 1993). In the case of lysozyme the correlation is comparable to that found for *B*-factors. No significant improvement in the correlation with S^2 is seen when local rmsd values, derived from the superposition of structures over a window of 2−5 residues, are used. Correlations between smoothed and regionally averaged NMR data are also compared in Table 2.

Given that these different indicators of molecular dynamics correlate well, it is possible to examine the factors that determine the extent of dynamics in the different regions of the protein. A molecular dynamics study of hen lysozyme has suggested a possible relationship between increased fluctuations and a greater than average surface accessibility of regions of the protein (Post et al., 1989). In contrast to *B*-factors, surface accessibilities of peptide bond units are nearly identical when calculated for single protein molecules taken from different crystal structures having the same space group and very similar when calculated from structures in other space groups and the ensemble of NMR structures. Therefore, the effect of crystal contacts is, at least in part, removed when individual molecules are considered in this manner. Comparisons of individual S^2 values with peptide bond accessibilities (Figure 6c, Table 2) indicate that there is a significant correlation between them. Furthermore, the correlation of S^2 with accessibilities is considerably enhanced when the data are averaged over structural segments of the protein (Figure 6a,b, Tables 2 and 3). Interestingly, comparisons of S^2 values with average accessibilities derived for the ensemble of NMR structures are even better than those with crystal structure accessibilities, especially when smoothed over several residues or averaged over the structural segments of the protein. These results imply that some of the determinants of the dynamic behavior may often be similar for neighboring residues in the protein sequence.

The extent of main-chain fluctuations at particular sites is also likely to be affected by the participation of the latter in hydrogen bonding and in secondary structure. This has been found in some studies (Schneider et al., 1992; Kördel et al., 1992), but not in others (Kay et al., 1989; Stone et al., 1993). Figure 5d suggests that there is significant correspondence between regions of the main chain with strong hydrogen bonding and high order parameters. From Figure 7a, it is apparent that peptide groups that are hydrogen bonded in crystal structures have, on average, higher order parameters than those that are not. However, these populations overlap considerably, as there are sites that, according to their S^2 values, are highly restrained despite the absence of hydrogen bonding (energy > −0.5 kcal mol^{-1}). Thus, while for example 14 of the 31 peptide units that are not persistently hydrogen bonded (as either donor or acceptor) in the crystal structures have values of $S^2 \leq 0.83$, 12 have values of $S^2 > 0.86$. This suggests that structural features other than hydrogen bonding can significantly determine the mobility of main-chain sites.

It is apparent that hydrogen-bonding strength is, at least in part, related to surface accessibility. Thus, the most strongly hydrogen-bonded groups are completely inaccessible in hen lysozyme. However, this relationship is complicated because inaccessibility does not always result in strong hydrogen bonding (McDonald & Thornton, 1994), and in these cases the lack of main-chain mobility must be rationalized solely in terms of the van der Waals packing interactions within the interior of the protein. The effects of hydrogen bonding and accessibility on the dynamics therefore are largely complementary for the NH groups considered here. By comparing the distribution of S^2 values for amides with and without protection against hydrogen exchange [Classes II−IV in Pedersen et al. (1991); Figure 7b], it is clear that the latter class has, on average, lower values of S^2. As both of these factors contribute to hydrogen exchange (Pedersen et al., 1991), it might be expected that there would be good quantitative correlation between S^2 values and amide exchange protection factors. This is not, however, found experimentally, presumably because fluctuations that give rise to hydrogen exchange from protected amides represent rare excursions and are not reflected on the time scale of the dynamic processes examined in this study by relaxation measurements.

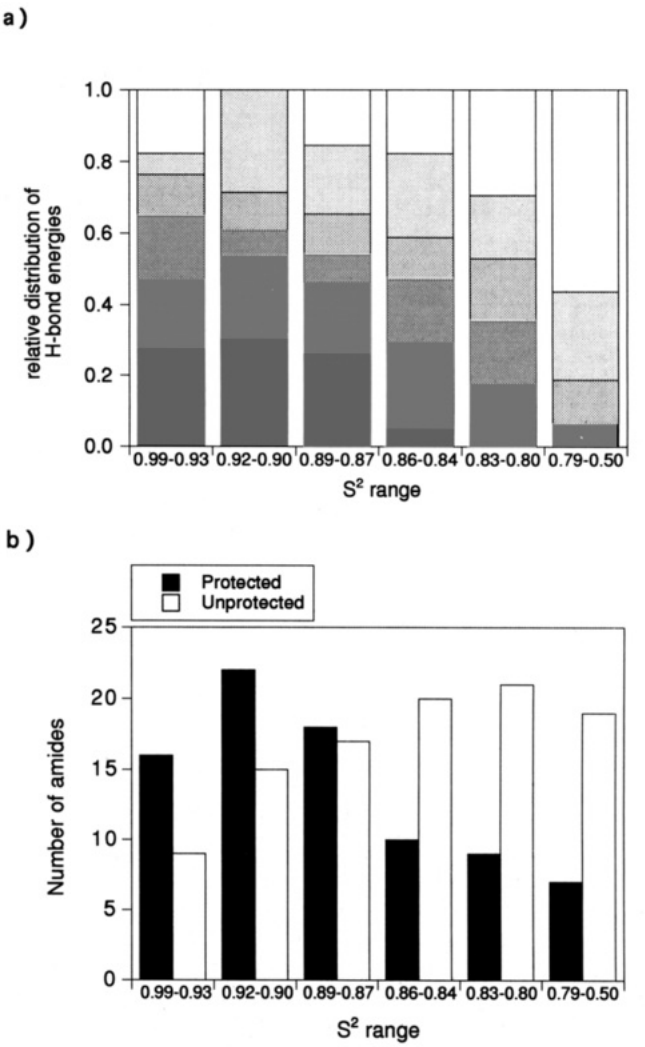

FIGURE 7: (a) Relative distribution of the peptide units in categories of hydrogen bond energy within ranges of S^2 values. Hydrogen bond energy categories are 0.0 to −0.5, −0.5 to −1.0, −1.0 to −1.5, −1.5 to −2.0, −2.0 to −3.0, and less than −3.0 kcal mol^{-1} in order of increased shading. (b) The number of amides having slowed hydrogen exchange [classes II−IV defined as protected in Pedersen et al. (1991)] (black) and fast exchange (white) within ranges of S^2 values.

In addition to the general correlations discussed in the preceding paragraphs, some more specific features of the protein structure do, however, appear to be of considerable importance in determining main-chain dynamics. As an

example, we consider the observation that only the central region, residues 68−72, of the long loop spanning residues 61−78 in hen lysozyme experiences substantial fluctuations (Table 3). Several features of the protein primary sequence in this region are pertinent in this respect. The region contains two proline residues (70 and 79) and several bulky hydrophobic residues (Trp 62, Trp 63, Leu 75, and Ile 78), and the side chains of several residues may form salt bridges (Asp 66, Arg 61, Arg 68, and Arg 73) or hydrogen bonds (Asn 65, Asn 73, and Asn 77). These factors are all likely to contribute to the attenuation of the mobility of this loop. The presence of cysteine residues participating in disulfide cross-links (Cys 64, Cys 76, and Cys 80) is, however, particularly likely to be of importance for the reduction in mobility of this and other regions of the protein (shown in Figure 5). In contrast to these features, which provide conformational stability to the loop terminal residues, glycine residues 67 and 71 are likely to be responsible for the relatively high flexibility in the central region of the loop. Interestingly, a second mobile loop in the protein, residues 101−104, is also bounded by glycine residues 102 and 104. Other main-chain regions with glycine repeats are, however, not as flexible. An example is the region of the main-chain joining the A- and B-helices (16−23), which does not experience substantial motion on the fast time scales considered here, although Gly 22 and Gly 26 are in close proximity, and despite the fact that the main chain in this region has higher than average surface accessibility and lower than average hydrogen-bonding interactions (Figure 7b,c). Three of the loop residues, however, have side chains that participate in the hydrophobic core of the protein. Thus, as in the case of disulfides earlier, it is possible that specific side-chain interactions may dampen and restrict main-chain dynamics and may therefore well be superimposed on the correlation of mobility with accessibility. Similar structural themes have been described as the determinants of loop and turn conformations in antibodies (Tramontano et al., 1989) and have been exploited as design strategies to build in rigidity to *de novo* protein structures (Chemeris et al., 1994).

Correlation of Side-Chain S^2 Values with Features of the Crystal and NMR Structures. S^2 values of side chains have been compared to structural parameters similar to those considered in the analysis of the main-chain values (see Table 1). Again, *B*-factors of side chains correlate well with their order parameters derived in this study. The majority of side chains with low *B*-factors have, as expected, high values for S^2 and many of the sidechains with low S^2 values have high *B*-factors in all crystal structures. Indeed, electron density has often been reported as poor or nonexistent in crystallographic maps for several of these side chains; this is the case, for example, for Arg 73, Arg 125, and Gln 121 in tetragonal crystal structures (Grace, 1980; Handoll, 1985). There are, however, side chains with low S^2 values but with low *B*-factors in all structures, showing exceptions to this trend. As with main-chain atoms, *B*-factors for some of the side-chains can vary considerably between different crystal structures, presumably as a result of crystal packing contacts specific to a certain structure or space group. The side chains of Gln 41, Asn 106, and Asn 113, which deviate from the approximately linear relationship shown in Figure 8a between S^2 and *B*-factors for the asparagine and glutamine side chains in the hentyp2 structure, are part of the group of side-chain atoms known to make intermolecular contacts in tetragonal

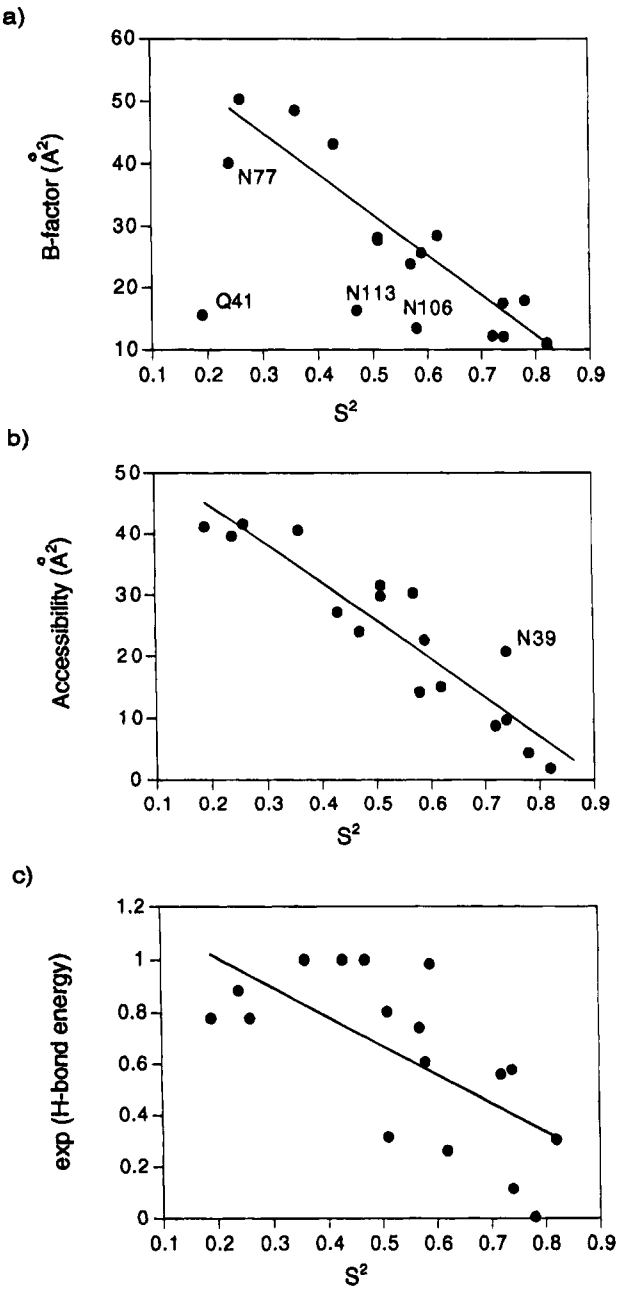

FIGURE 8: S^2 values for asparagine and glutamine side-chain groups as a function of (a) *B*-factors, (b) accessibility (both of the hentyp2 crystal structure), and (c) hydrogen bond energy plotted in exponential form, averaged over the six crystal structures. The correlation coefficients for these relationships are 0.65 for (a), 0.92 (for b), and 0.67 for (c). Direct comparison of results for tryptophan and arginine with those for asparagine and glutamine side chains is difficult given their fewer numbers, the differences in the absolute accessibility scale used, and the likely dependence of the minimum value of S^2 attainable on the number of intervening bonds to the main chain. The data for tryptophan and arginine side chains are summarized in Table 1; the trends are in good agreement with those seen for asparagine and glutamine side chains.

crystal structures (Handoll, 1985). Overall, however, the results indicate that both *B*-factors and S^2 values reflect similar motions.

The correlation of S^2 with accessibility found for the main chain is even more evident for the side chains, where the variation in both of these parameters is greater. Thus, an excellent correlation is obtained when the S^2 values of asparagine and glutamine side chains are compared with their

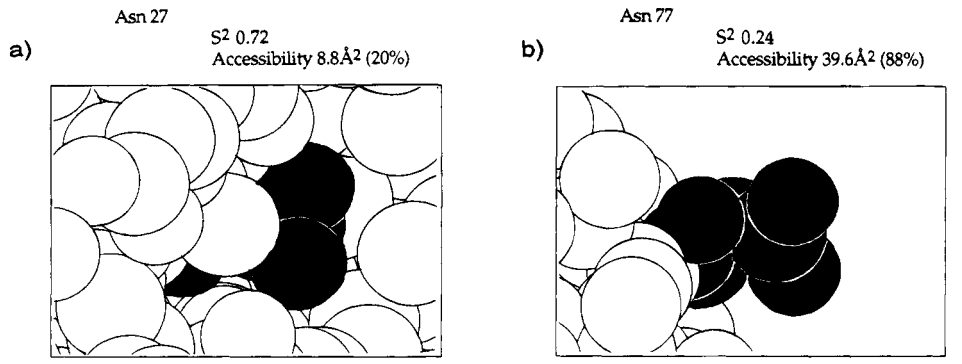

FIGURE 9: Examples of the environments of asparagine side chains: (a) Asn 27 with a well-defined conformation and (b) Asn 77 with a poorly defined conformation in all six crystal structures. Accessibilities of the hentyp2 structure and S^2 values are indicated. van der Waals radii were used on the hentyp2 crystal structure to generate this picture using InsightII (Biosym Inc.).

accessibilities in crystal structures (plotted for hentyp2 in Figure 8b). The locations of the side chains of Asn 27 and Asn 77 in the hentyp2 crystal structure are shown in Figure 9; these represent examples of low and high side-chain accessibility, respectively. Side-chain group accessibilities averaged over individual members of the ensemble of the NMR solution structures also correlate well with S^2 values, although not as well as the crystallographic accessibilities. This may be due to the fact that few NOE restraints were included for these side-chain atoms in the structure calculations, and the restraints on their position are therefore indirect. Low values of S^2 for the arginine side chains and for Trp 62 are, on average, also accompanied by both high *B*-factors and high accessibilities in crystal structures (Table 1). Furthermore, S^2 values, which could be obtained for four of the remaining six unassigned arginine side chains, are also less than 0.3; all arginines have high accessibilities in the crystal structures.

Many of the side chains considered here are found to be hydrogen bonded in one or more of the six crystal structures. Four tryptophan indole hydrogen are on average involved only in weak hydrogen bonding, but these residues are confined to the hydrophobic core of the protein, which is likely to be responsible for their high S^2 values. The indole hydrogens of Trp 62 and Trp 63 are not hydrogen bonded in any of the crystal structures. Two of the five assigned arginine residues are, on average, hydrogen bonded in the crystal structures (bond energy ≤ -0.5 kJ mol^{-1}). The hydrogen bond partners are, however, different in different individual structures, and the low order parameters suggest that hydrogen bonding is unlikely to be persistent in solution. The asparagine and glutamine side chains can be divided roughly into two groups by the hydrogen-bonding energy associated with them in crystal structures. When hydrogen bond energies are averaged over the six crystal structures, eight of these side chains are hydrogen bonded; they have a mean order parameter of 0.69 ± 0.10, well above that for side chains that are, on average, not hydrogen bonded in crystal structures (mean $S^2 = 0.40 \pm 0.15$). S^2 is plotted as a function of hydrogen bond energy in Figure 8c and shows an overall trend similar to that observed for the main-chain atoms. Side chains with high order parameters therefore are more likely to be involved in hydrogen bonding, as judged by the frequency and strength of their hydrogen bonding in different crystal structures, and are invariably buried from the protein surface.

Evidence for side-chain mobility of hen lysozyme in solution has previously come from $^3J_{\alpha\beta}$ coupling constants measured directly (Smith et al., 1991) or from the extent of averaging of the χ_1 angle identified from peak shapes (Bartik & Redfield, 1993). Thus, the extent of conformational averaging about the $C^\alpha-C^\beta$ bond is known for the great majority of the side chains considered here and is summarized together with the order parameters in Table 1. The six asparagine side chains with a highly averaged χ_1 value in solution have a mean S^2 value of 0.40 ± 0.12. By contrast the eight asparagine side chains found to exist in a single conformation about χ_1 have a mean S^2 value of 0.67 ± 0.09. Therefore, there is a correspondence between the lack of rotational averaging about the $C^\alpha-C^\beta$ bond and the degree of restraint of the $^{15}N-^1H$ bond vector as measured by ^{15}N relaxation. This is despite the fact that there is an additional $C^\beta-C^\gamma$ bond (torsion angle χ_2), about which the asparagine side-chain end groups may rotate. For the longer side chains of glutamines, this correspondence may break down, as seen in Gln 41 which has an S^2 value of 0.19. For arginines, which have even longer side chains, there appears to be no correlation with the extent of averaging about the $C^\alpha-C^\beta$ bond. Interestingly, however, for some of the groups of side chains there is significant correspondence between main-chain and side-chain mobility (shown in Figure 1a,b). This may be anticipated if the features that are the determinants of conformational freedom are general and local to particular regions of the protein.

Determinants of Conformational Freedom. Close packing of the majority of the main chain and of a significant fraction of side chains in the interior of proteins is a feature specific to their native folds (Chothia, 1975, 1976; Miller et al., 1987; Richards, 1992; Lee & Richards, 1971, 1994). Equally, it has been observed that the great majority of main-chain and side-chain groups with hydrogen-bonding potential are involved in such interactions within the interior, as well as near and at the surface of folded proteins (Baker & Hubbard, 1984; Jeffrey & Saenger, 1991; McDonald & Thornton, 1994). Thus, burial of the main and side chains and the optimization of their hydrogen bonding are recognized as the main determinants of protein structure and folding (Rose & Wolfenden, 1993). It follows that these features of protein structure must also significantly affect, if not determine, the dynamic behavior of the polypeptide main chain and the side chains of its component residues. Evidence for such relationships has been provided from previous analysis of the crystal and solution structures of hen lysozyme and other

proteins. Examples are the correlation of the apparent thermal motions (a derivative of the *B*-factor) with the square of the distance from the molecular centroid (Sternberg et al., 1979; Artymiuk et al., 1979) and the correlations found by MacArthur and Thornton (1993) between surface accessibility and a lack of side-chain rotamer definition in NMR structures.

In this study, we find statistically significant correlations of S^2 values with *B*-factors of crystal structures and rmsd values of the ensemble of NMR structures. This suggests that these indicators of positional disorder have at least some common determinants. The data derived here from ^{15}N relaxation measurements on the dynamics of the main chain and the collection of 28 side chains of hen lysozyme in solution have allowed us to explore further some of the possible origins of these determinants. A significant correspondence is found between limited mobility, as defined by high S^2 values, and the strength of hydrogen bonding, although the origin of this relationship is complicated by the known dependence of hydrogen bond strength on low solvent accessibility. However, the most striking correlation both for the main chain (when regionally averaged) and particularly for side chains is between S^2 values and surface accessibilities. The quality of these correlations implies that van der Waals contacts or proximity to neighboring side-chain and main-chain groups, here crudely estimated by static solvent accessibility, imposes the most effective constraints on mobility.

It will be interesting as further data become available to assess how similar the details of correlations between accessibility and structure are for different proteins and to what extent differences between them can be attributed to specific features of the structures. Thus, for example, the long loop of lysozyme (residues 61−78), while mobile compared to other regions of this protein, experiences more limited fluctuations than loops of some other proteins such as the four-helix bundle interleukin-4 (Redfield et al., 1992). This may be attributed to the specific nature of the residues in this region of the sequence and to the presence of disulfide bonds. Furthermore, although the location of binding sites for other molecules will of necessity be on or near surfaces of proteins, and hence the residues located in them will frequently be highly mobile in the unbound state, this may not always be the case. In the case of lysozyme, for example, a number of side-chain residues located in the active site are actually more restricted in the extent of their motions than other residues of the same amino acid type located elsewhere in the molecule (Figure 1). This can be attributed to the fact that this region of the protein surface forms a deep cleft, which gives rise to closer packing and reduced solvent accessibility for at least some of the residues within it. The extent to which such variation in behavior is of biological significance or simply reflects the overall architecture of the protein remains to be established.

ACKNOWLEDGMENT

We thank Dr. Simon Hubbard (EMBL) and Professor Janet Thornton (UCL) for making available their program NACCESS for surface accessibility calculations; Dr. Arthur Palmer (Columbia University) for providing his program package to fit and analyze NMR relaxation data; and Drs. Sheena Radford and Lorna Smith for advice and assistance.

SUPPLEMENTARY MATERIAL AVAILABLE

Table of ^{15}N main-chain and tryptophan side-chain chemical shifts and shifts of ^{1}H and ^{15}N resonances for the arginine, asparagine, and glutamine side chains assigned in this study; summary and table of relaxation data and derived order parameters, including further details on the derivation of order parameters (12 pages). Ordering information is given on any current masthead page.

REFERENCES

Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C., & Sternberg, M. J. E. (1979) *Nature 280*, 563−568.

Bai, Y., Milne, J. S., Mayne, L. M., & Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.

Baker, E. N., & Hubbard, R. E., (1984) *Prog. Biophys. Mol. Biol. 44*, 97−179.

Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., & Bax, A. (1992) *Biochemistry 31*, 5269−5278.

Bartik, K., & Redfield, C. (1993) *J. Biol. NMR 3*, 415−428.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1965) *Nature 206*, 757−761.

Boyd, J. (1995) *J. Magn. Reson.* (in press).

Boyd, J., Hommel, U., & Campbell, I. D. (1990) *Chem. Phys. Lett. 175*, 477−482.

Bruccoleri, R. E., Karplus, M., & McCammon, J. A. (1986) *Biopolymers 25*, 1767−1802.

Brünger, A. T. (1992) *X-PLOR version 3.1: A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.

Campbell, I. D., Dobson, C. M., & Williams, R. J. P. (1975) *Proc. R. Soc. London B. 189*, 503.

Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry*, Vol. II, Freeman & Co., San Francisco.

Chandrasekhar, I., Clore, M. G., Gronenborn, A. M., & Brooks, B. R. (1992) *J. Mol. Biol. 226*, 239−250.

Chemeris, V. V., Dolgikh, D. A., Fedrov, A. N., Finkelstein, A. V., Kirpichnikov, M. P., Uversky, V. N., & Ptitsyn, O. B. (1994) *Protein Eng. 7*, 1041−1052.

Chothia, C. (1975) *Nature 254*, 304−308.

Chothia, C. (1976) *J. Mol. Biol. 105*, 1−14.

Clarage, J. B., Clarage, M. S., Phillips, W. C., Sweet, R. M., & Caspar, D. L. D. (1992) *Proteins: Struct., Funct., Genet. 12*, 145−157.

Clore, G. M., Driscoll, P. C. Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 7387−7401.

Constantine, K. L., Friedrichs, M. S., Goldfarb, V., Jeffrey, P. D., Sherriff, S., & Mueller, L. (1993) *Proteins: Struct., Funct., Genet. 15*, 290−311.

Cross, T. A., & Opella, S. J. (1983) *J. Am. Chem. Soc. 105*, 306−308.

Dobson, C. M. (1993) *Curr. Biol. 3*, 530−532.

Dobson, C. M., & Karplus, M. (1986) *Methods Enzymol. 131*, 362−389.

Faure, P., Micu, A., Perahia, D., Doucet, J., Smith, J. C., & Benoit, J. P. (1994) *Nature Struct. Biol. 1*, 124−128.

Gerstein, M., Lesk, A. M., & Chothia, C. (1994) *Biochemistry 33*, 6739−6749.

Gibrat, J.-F. & Go, N. (1990) *Proteins: Struct., Funct., Genet. 8*, 258−279.

Grace, D. E. P. (1980) D.Phil. Thesis, University of Oxford, Oxford, UK.

Grasberger, B. L., Gronenborn, A. M., & Clore, G. M. (1993) *J. Mol. Biol. 230*, 364−372.

Handoll, H. H. G. (1985) D.Phil. Thesis, University of Oxford, Oxford, UK.

Herzfeld, J., Roberts, J. E., & Griffin, R. G. (1987) *J. Chem. Phys. 86*, 597−602.

Hoel, P. G. (1984) *Introduction to Mathematical Statistics*, 5th ed., J. Wiley & Sons, New York.

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C., & Rupley, J. A. (1972) *The Enzymes* (Boyer, P. D., Ed.) Vol. VII, Academic Press, New York.

Jeenes, D. J., Marczinke, B., MacKenzie, D. A., & Archer, D. B. (1993) *FEMS Microbiol. Lett. 107*, 267−272.
Jeffrey, G. A., & Saenger, W. (1991) *Hydrogen Bonding in Biological Structures*, Springer, New York.
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577−2637.
Karplus, M., & McCammon, J. A. (1983) *Annu. Rev. Biochem. 53*, 263−300.
Kay, L. E., Torchia, D. A., & Bax, A. (1989) *Biochemistry 28*, 8972−8979.
Kay, L. E., Nicholson, L. K., Delagio, F., Bax, A., & Torchia, D. A. (1992) *J. Magn. Reson. 97*, 359−375.
Kodandapani, R., Suresh, C., & Vijayan, M. (1990) *J. Biol. Chem. 27*, 16126−16131.
Kördel, J., Skelton, N. J., Akke, M., Palmer, A. G., & Chazin, W. J. (1992) *Biochemistry 31*, 4856−4866.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Krishna, N. R., Sarathy, K. P., Huang, D.-H., Stephens, R. L., Glickson, J. D., Smith, C. K., & Walter, R. (1982) *J. Am. Chem. Soc. 104*, 5051−5053.
Kundrot, C. E., & Richards, F. M. (1987) *J. Mol. Biol. 193*, 157−170.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559−4570.
London, R. E. (1989) *Methods Enzymol. 176*, 358−375.
MacArthur, M. W., & Thornton, J. M. (1993) *Proteins: Struct., Funct., Genet. 17*, 232−251.
Marion, D., Ikura, M., & Bax, A. (1989) *J. Magn. Reson. 84*, 425−430.
McDonald, I. K., & Thornton, J. M. (1994) *J. Mol. Biol. 238*, 777−793.
Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 608−613.
Miller, S., Janin, J., Lesk, A., & Chothia, C. (1987) *J. Mol. Biol. 196*, 641−656.
Ohkubo, T., Taniyama, Y., & Kikuchi, M. (1991) *J. Biochem. 110*, 1022−1029.
Olejniczak, E. T., Poulsen, F. M., & Dobson, C. M. (1984) *J. Am. Chem. Soc. 106*, 1923−1930.
Palmer, A. G., Rance, M., & Wright,P. E. (1991) *J. Am. Chem. Soc. 113*, 4370−4380.
Palmer, A. G., Skelton, N. J., Chazin, W. J., Wright, P. E., & Rance, M. (1992) *Mol. Phys. 75*, 699−711.
Pedersen, T. G., Sigurskjold, B. W., Andersen, K. V., Kjaer, M., Poulsen, F. M., Dobson, C. M., & Redfield, C. (1991) *J. Mol. Biol. 218*, 413−426.
Peng, J. W., & Wagner, G. (1992) *Biochemistry 31*, 8571−8586.
Post, C. B., Dobson, C. M., & Karplus, M. (1989) *Proteins: Struct., Funct., Genet. 5*, 337−354.
Powers, R., Clore, G. M., Garret, D. S., & Gronenborn, A. M. (1993) *J. Magn. Reson. B 101*, 325−327.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992a) *Nature 358*, 302−305.
Radford, S. E., Buck, M., Topping, K. D., Evans, P. A., & Dobson, C. M. (1992b) *Proteins: Struct., Funct., Genet. 14*, 237−248.
Ramanadham, M., Sieker, L. C., & Jensen, L. H. (1990) *Acta Crystallogr. B46*, 63−69.
Redfield, C., & Dobson, C. M. (1988) *Biochemistry 27*, 122−136.
Redfield, C., & Dobson, C. M. (1990) *Biochemistry 29*, 7201−7214.
Redfield, C., Boyd, J., Smith, L. J., Smith, R. A. G., & Dobson, C. M. (1992) *Biochemistry 31*, 10431−10437.
Richards, F. M. (1974) *J. Mol. Biol. 82*, 1−11.
Richards, F. M. (1992) in *Protein Folding* (Creighton, T. E., Ed.) Freeman, New York.
Roberts, I. N., MacKenzie, D. A., Jeenes, D. B., Archer, D. B., Radford, S. E., Robinson, C. V., Aplin, R. T., & Dobson, C. M. (1992) *Biotechnol. Lett. 14*, 897−902.
Rose, G. D., & Wolfenden, R. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 381−415.
Schneider, D. M., Dellwo, M. J., & Wand, J. A. (1992) *Biochemistry 31*, 3645−3652.
Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C., & Davis, D. R. (1987) *Proc. Natl. Acad. Sci U.S.A. 84*, 8075−8079.
Smith, J. L., Hendrickson, W. A., Honzatko, R. B., & Sheriff, S. (1986) *Biochemistry 25*, 5018−5027.
Smith, L. J., Sutcliffe, M. J., Redfield, C., & Dobson, C. M. (1991) *Biochemistry 30*, 986−996.
Smith, L. J., Sutcliffe, M. J., Redfield, C., & Dobson, C. M. (1993) *J. Mol. Biol. 229*, 930−944.
Sternberg, M. J. E., Grace, D. E. P., & Phillips, D. C. (1979) *J. Mol. Biol. 130*, 231−253.
Stone, M. J., Fairbrother, W. J., Palmer, A. G., Reizer, J., Saier, M. H., & Wright, P. E. (1992) *Biochemistry 31*, 4394−4406.
Strynadka, N. C. J., & James, M. N. G. (1991) *J. Mol. Biol. 220*, 401−424.
Thomson, N. K., & Poulsen, F. M. (1993) *J. Mol. Biol. 234*, 234−411.
Tramontano, A., Chothia, C., & Lesk, A. M. (1989) *Proteins: Struct., Funct., Genet. 6*, 382−394.
Wedin, R. E., Delepierre, M., Dobson, C. M., & Poulson, F. M. (1982) *Biochemistry 21*, 1098−1103.
Wilson, K. P., Malcolm, B. A., & Matthews, B. W. (1993) *J. Biol. Chem. 267*, 10842−10849.
Young, A. C. M., Tilton, R. F., & Dewan, J. C. (1994) *J. Mol. Biol. 235*, 302−317.

BI942443J