Sequential ^{1}H NMR Assignments and Secondary Structure Identification of Human Ubiquitin

Paul L. Weber,* Stephen C. Brown, and Luciano Mueller

Department of Physical and Structural Chemistry, Smith Kline and French Laboratories, P.O. Box 1539, King of Prussia, Pennsylvania 19406-0939

Received May 27, 1987; Revised Manuscript Received July 9, 1987

ABSTRACT: ^{1}H NMR assignments of human ubiquitin (76 amino acids, M_r 8565) have been made by a combination of DQF-COSY, DQF-RELAY, NOESY, DQ, and isotropic mixing experiments. Complete NH, $C^\alpha H$, and $C^\beta H$ assignments were obtained; resonances not yet assigned are the side-chain amides of Q-40, Q-41, Q-49, N-60, and Q-62 and the peripheral protons ($C^\gamma H$ and outward) of M-1 and K-27. A total of 558 out of 579 (96%) potentially observable protons were assigned. Particular attention was directed toward obtaining complete assignments of the aliphatic residues (seven Ile, nine Leu, four Val) since these residues form an extensive hydrophobic core and NOEs from these residues are invaluable for structure calculations. The secondary structure elements were also identified from the sequential NOE data and differ slightly in description from the published 2.8 Å resolution crystal structure [Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., & Cook, W. J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3582–3585]; the NMR data suggest that residues 48–50 form a short fifth strand in the β-sheet and that residues 56–61 form a helical turn. The sequential assignment results presented here are in agreement with the main chain directed assignments presented in the preceding paper [Di Stephano, D., & Wand, A. J. (1987) *Biochemistry* (preceding paper in this issue)].

Ubiquitin is a small protein (76 amino acids, M_r 8565) that derives its name from its occurrence throughout the plant and animal kingdoms [for reviews, see Hershko and Ciechanover (1982) and Finley and Varshavsky (1985)]. Ubiquitin has unparalleled amino acid sequence conservation (yeast and human forms differ by only three conservative substitutions) that may be related to the diversity of its biological roles. Ubiquitin was originally believed to be a hormone or autocrine growth factor, inducing T-cell and B-cell differentiation (Audhya & Goldstein, 1985). Ubiquitin has also been found attached to histone H2A and may play a role in nucleosome structure (Busch, 1984). Its most studied role, however, is that of signaling selective cytoplasmic ATP-dependent proteolytic degradation by covalent attachment to target proteins (Hershko & Ciechanover, 1982).

By itself ubiquitin has no known function or physiologically significant enzymatic activity, suggesting that it acts perhaps as a proteinaceous "cofactor". To better understand the contribution of ubiquitin in its varied roles, we and our colleagues have begun investigations of the structure, stability, and folding of ubiquitin using a combination of mutagenesis and NMR[1] methods (Ecker et al., 1987b). As an essential first step to the NMR studies, we have assigned the ubiquitin proton NMR spectrum using the sequential assignment method (Wüthrich et al., 1982). This is a two-step process: first, the *J*-coupled networks of spins for each amino acid are connected and identified by amino acid type when possible; second, NOEs corresponding to short distances (which statistically occur primarily between nearest neighbors) between backbone protons are identified. Comparison of the sequential connectivities to the amino acid sequence allows one to assign resonances to specific protons in the protein (Billeter et al., 1982). The first step, namely, the identification of amino acid spin systems, becomes increasingly difficult in larger proteins (such as ubiquitin) due to spectral crowding, broader resonance line widths, and the frequent appearance of spin systems with missing cross-peaks and/or deviations from canonical chemical shift patterns.

We have attempted to avoid these problems by using a battery of 2D NMR experiments that, taken together, yield redundant information leading to unambiguous identification of spin system networks (Klevit & Drobny, 1986; Bach et al., 1987; Weber et al., 1987). This approach has resulted in the assignment of nearly all resonances in the ubiquitin ^{1}H NMR spectrum. In addition, the sequential NOEs were used to determine the secondary structure of ubiquitin in solution, which differs slightly from that described in the crystal structure reported at 2.8-Å resolution (Vijay-Kumar et al., 1985).

Materials and Methods

The ubiquitin and ubiquitin mutants used in these studies were obtained from *Escherichia coli* cells transformed with a plasmid-encoded synthetic ubiquitin gene and purified as described elsewhere (Ecker et al., 1987a,b). NMR samples of ubiquitin were prepared in 25 mM acetic acid-d_4, pH 4.7 (direct meter reading at room temperature), and were 1–5 mM in protein. Samples were prepared either in 10% D_2O/90% H_2O or in "100%" D_2O buffers. Isotopically labeled solvents were obtained from Stohler.

All NMR experiments were performed at 500 MHz on a JEOL GX500, and data were transferred to a microVAX II for processing with FTNMR software (Hare Research, Inc.). DQF-COSY (Piantini et al., 1982; Shaka & Freeman, 1983),

[1] Abbreviations: NMR, nuclear magnetic resonance; 2D, two dimensional; DQF-COSY, 2D double-quantum filtered correlated spectroscopy; DQ, double quantum; DQF-RELAY, 2D double-quantum filtered relayed coherence transfer spectroscopy; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; PE-COSY, 2D primitive exclusive correlated spectroscopy; MLEV, 2D isotropic mixing experiments employing MLEV-16 or modified MLEV-16 mixing pulses; DSS, 4,4-dimethyl-4-silapentane-1-sulfonate.

0006-2960/87/0426-7282$01.50/0 © 1987 American Chemical Society

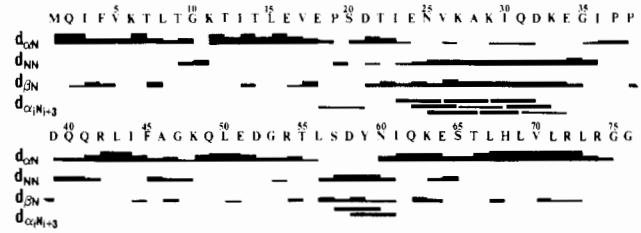

FIGURE 1: Summary of sequential NOEs. Both the type of sequential NOE identified and the intensity observed are summarized here. Thick bars indicate strong NOEs; thin bars indicate weak NOEs. Shown is a combination of the data acquired at 30 and 50 °C (see text).

DQ spectroscopy (Wokuan & Ernst, 1977; Wagner & Zuiderweg, 1983; Rance et al., 1985), PE-COSY (Mueller, 1987), DQF-RELAY (Sørensen, 1984; Weber & Mueller, 1987), NOESY (Macura & Ernst, 1980; States et al., 1982), and MLEV spin-locked experiments (Levitt et al., 1982; Bax & Davis, 1985) were performed as previously described except that the "trim pulses" in the MLEV-17 sequence were deleted since phase errors that these modifications correct are not apparent when a strong (≥10 kHz) radio-frequency field is used (Weber et al., 1987). All experiments used a spectral width of 6250 Hz, and data sets averaged 32–64 transients of 2048 complex points. Quadrature detection in ω_1 was obtained by the hypercomplex method (Mueller & Ernst, 1979; States et al., 1982). Spectra were collected either at 30 or at 50 °C, and solvent suppression was performed by continuous irradiation of the water signal during the relaxation delay (1.5–2.0 s). Data sets were 2K × 2K zero-filled matrices; for spectra in D_2O, data were zero-filled to 4K real points in t_2, but only 2K points (containing all nonaromatic side-chain protons) were saved for subsequent t_1 transformation. An analagous transformation/reduction was carried out for the t_1 transform.

Sequential resonance assignments were performed as originally described (Billeter et al., 1982; Wüthrich et al., 1982), with additional emphasis placed on finding corroborating NOEs that arise from secondary structure contacts. Valine methyl stereospecific assignments were made by interpreting intraresidue and sequential NOEs in light of the possible local structure of the residue in question (Zuiderweg et al., 1985).

RESULTS

A summary of the sequential assignments and the amino acid sequence of ubiquitin is shown in Figure 1.

The four aromatic residues (F-4, F-45, Y-59, H-68) were identified by the routine combination of DQF-COSY and NOESY experiments. First, the ring systems were identified from the DQF-COSY data. The ring systems of H-68 and one phenylalanine were clear, but the $C^{\epsilon}H$ and $C^{\delta}H$ resonances of the second phenylalanine and one of the Y-59 resonances overlapped at 7.28 ppm. The degeneracy was overcome when the ring systems were connected to the $C^{\alpha}H$ and $C^{\beta}H$ resonances with the NOESY data. No NOEs were observed between the H-68 ring and $C^{\beta}H$ or $C^{\alpha}H$ resonances. It was also noted that in D_2O the H-68 $C^{\epsilon_1}H$ resonance was completely exchanged within a few hours at 50 °C and pH 4.7; furthermore, the ring resonances were shifted upfield compared to a free histidine at the same pH.

Four of the six glycine residues were identified by the distinctive pairs of NH–$C^{\alpha}H$ cross-peaks in the DQF-COSY spectrum and by the corresponding $C^{\alpha}H$–$C^{\alpha}H'$ cross-peak. The remaining two glycines could still be identified in the H_2O spectrum, but the α-protons had degenerate or near-degenerate chemical shifts. Confirming evidence for all glycine residues was obtained from DQ-spectroscopy in H_2O. The glycine NH–$C^{\alpha}H$ cross-peaks were also distinctive in the PE-COSY, where they appeared as twin quartets offset by the passive $J_{NH-C^{\alpha}H'}$ coupling.

The seven threonine and two alanine residues were identified with the DQF-COSY and DQF-RELAY experiments; of use was a DQF-RELAY acquired in H_2O, where cross-peaks between the alanine amide and alanine methyl resonances were identified. Two of the threonine $C^{\alpha}H$–$C^{\beta}H$ cross-peaks were very weak (T-22, T-55) in the DQF-COSY spectrum, but still could be identified in spectra plotted at levels with higher noise content. This is a seldom-mentioned advantage of DQF-COSY and DQF-RELAY over the analogous magnitude experiments: weak cross-peaks can be distinguished against a random noise background due to their distinct coupling patterns.

The three serine and seven asparagine or aspartate residues were identified with moderate certainty on the basis of the chemical shifts and coupling patterns of the DQF-COSY cross-peaks. While this is not rigorous and can be misleading, in this case all were correctly identified. The three serine residues were identified by the low-field chemical shifts (ca. 4 ppm) and DQF-COSY multiplet patterns of the $C^{\alpha}H$–$C^{\beta}H$ cross-peaks. The aspartate and asparagine spin systems were identified by the intermediate chemical shift of the β-protons (3.5–2.5 ppm) and again by their antiphase coupling patterns. Some of the Asx $C^{\alpha}H$–$C^{\beta}H$ cross-peaks overlapped, but all identifications were confirmed by $C^{\beta}H^{a}$–$C^{\beta}H^{b}$ cross-peaks. The backbone amide protons were connected with the DQF-COSY and DQF-RELAY spectra, but no attempts to connect the asparagine side chain amide protons were made at this point.

Approximately one-third of the amino acid residues in ubiquitin are aliphatic, each contributing to the high thermal stability of the protein through formation of the hydrophobic core. Since side chain–side chain contacts in these residues will determine to a large extent the tertiary folding of the molecule, care was taken to obtain clear and complete resonance identifications of these residues. Three of the four valine spin systems were identified with DQF-COSY and optimized DQF-RELAY experiments (Weber et al., 1985a). The remaining valine spin system (V-17) had a very weak $J_{C^{\alpha}H-C^{\beta}H}$ coupling; thus, no $C^{\alpha}H$–$C^{\gamma}H_3$ cross-peaks were observed in the DQF-RELAY spectrum. These cross-peaks were observed in the isotropic mixing experiment but were weaker than the other valine $C^{\alpha}H$–$C^{\beta}H$ relays. This led to the ambiguity of whether the spin system in question was actually a valine or a leucine. This spin system was likely to be a valine due to the observed $C^{\alpha}H$–$C^{\beta}H$ cross-peak multiplet pattern, the lack of an observable second $C^{\alpha}H$–$C^{\beta}H$ cross peak, and the apparent connectivity pattern (valine). The same pattern could be obtained from a leucine residue with appropriate weak couplings and resonance degeneracies. Final identification in this case was left until sequential assignments were made. After sequential assignments were completed, it also was possible to obtain stereospecific assignments for the geminal methyl pairs by analyzing the NOE patterns involving the stereorelated protons or methyls in light of the local structure of the residue in question (Zuiderweg et al., 1985).

The seven isoleucine residues were completely identified by comparing the DQF-COSY, DQF-RELAY, and MLEV spectra. From the first two experiments, all $C^{\alpha}H$–$C^{\beta}H$, $C^{\beta}H$–$C^{\gamma_2}H_3$, and $C^{\gamma_1}H$–$C^{\delta}H_3$ cross-peaks could be identified; the DQF-RELAY optimized with a 26-ms mixing time was particularly useful for identifying the latter. The connections

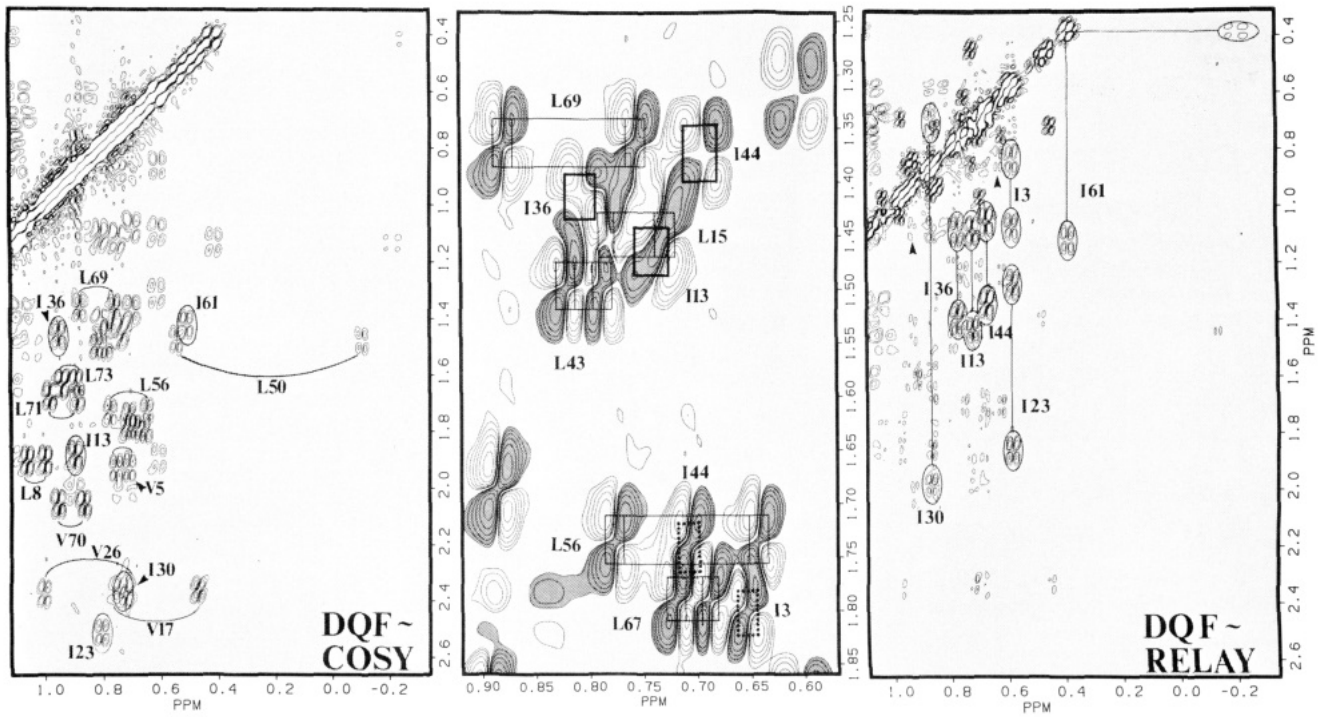

FIGURE 2: Identification of the aliphatic methylene and methyl resonances with DQF-COSY and DQF-RELAY experiments. (Left) Valine $C^{\beta}H$–$C^{\gamma}H_3$ cross-peak pairs, leucine $C^{\gamma}H$–$C^{\delta}H_3$ cross-peak pairs, and isoleucine $C^{\beta}H$–$C^{\gamma 2}H_3$ cross-peaks (in ovals) are identified. (Center) An expansion of the same region showing identification of five leucine $C^{\gamma 1}H$–$C^{\delta}H_3$ cross-peak pairs (connected by thin lines), two isoleucine $C^{\beta}H$–$C^{\gamma 2}H_3$ cross-peaks (dotted lines), and three isoleucine $C^{\gamma 1}H$–$C^{\delta}H_3$ cross-peaks (thick lines). The negative multiplet components are shaded gray in this figure. Note that the three isoleucine $C^{\gamma 1}H$–$C^{\delta}H_3$ cross-peaks have a larger apparent $J_{C^{\gamma 1}H\text{–}C^{\delta}H}$ splitting (along ω_2) due to the passive $J_{C^{\gamma 1}H^{a}\text{–}C^{\gamma 1}H^{b}}$ coupling; the end result is that these peaks are distinguishable from all other methyl cross-peaks in this region. (Right) A DQF-RELAY experiment acquired with a 28-ms mixing time, where modulation of COSY cross-peak intensities has resulted in fortuitous spectral editing, making identification of all isoleucine [$C^{\gamma 1}H$, $C^{\delta}H_3$] subspin systems straightforward (Weber & Mueller 1987). Arrows point to $C^{\gamma 2}H_3$–$C^{\gamma 1}H$ relays (see text). Several methyl–methyl relays from the valine and leucine residues are also apparent.

between the [$C^{\alpha}H$, $C^{\beta}H$, $C^{\gamma 2}H_3$] and [$C^{\gamma 1}H$, $C^{\delta}H_3$] subspin systems were made in at least one of three ways: (1) weak but unambiguous $C^{\beta}H$–$C^{\gamma 1}H$ cross-peaks in the DQF-COSY spectrum (I-13, I-44, I-61); (2) $C^{\alpha}H$–$C^{\delta}H_3$ and/or $C^{\beta}H$–$C^{\delta}H_3$ cross-peaks in the MLEV spectrum (I-3, I-23, I-30, I-61); (3) weak but unambiguous $C^{\gamma 2}H_3$–$C^{\gamma 2}H$ cross-peaks in the DQF-RELAY spectrum (I-3, I-36).

The nine leucine spin systems were identified by a combination of DQF-COSY, DQF-RELAY, and MLEV experiments. The aliphatic region of the DQF-COSY experiment had 13 methyl/methylene pairs, of which 3 (or 4, see above) were already identified as originating from valine spin systems. Often, only one of the two possible $C^{\alpha}H$–$C^{\beta}H$ cross-peaks was observed in the DQF-COSY. The second $C^{\alpha}H$–$C^{\beta}H$ cross-peak was observable in the DQF-RELAY experiment, and the identity of the $C^{\alpha}H$–$C^{\beta}H$ pairs for each leucine was confirmed by the identification of the $C^{\beta}H^{a}$–$C^{\beta}H^{b}$ cross-peaks in the DQF-COSY. Connectivities between $C^{\beta}H$ and $C^{\gamma}H$ resonances were usually not found, but $C^{\alpha}H$–$C^{\delta}H_3$ cross-peaks were always observed in the MLEV spectra. In most cases $C^{\beta}H$–$C^{\delta}H_3$ cross-peaks were also identified; thus the identification of all nine leucine residues was completed with redundant information. Figure 2 show the identification of all aliphatic methyl resonances.

Some simplification of the leucine identifications was also obtained from a mutant ubiquitin with the L-73 residue deleted. This residue lies on the surface of the protein (given the crystal structure); no difference in protein structure was observed as judged by the lack of change in the chemical shifts of all resonances (except for minor shifts in the R-74 $C^{\alpha}H$–$C^{\beta}H$ cross-peaks). The ΔL73 mutant was also more soluble than wild-type ubiquitin and, therefore, was used often instead of the wild-type sample for many of the spin system identifications. Also, as the L-73 $C^{\alpha}H$–$C^{\beta}H$ cross-peaks were located in a crowded portion of the DQF-COSY spectrum, this mutant was also helpful for identifying other spin systems by relieving some spectral crowding.

The remaining residues (Lys, Pro, Arg, Met, Glx) were identified when possible from the DQF-COSY, DQF-RELAY, and MLEV spectra. In most cases, the complete identifications were obtained after the sequential assignments were made, as it is simpler to identify one of these complex spin systems once one knows the identities of the $C^{\alpha}H$–$C^{\beta}H$ cross-peaks (Weber et al., 1985b; Di Stephano & Wand, 1987). Six of the seven lysines could be completely identified with all three spectra. Distinction between the lysine side chain resonances were made by comparing all three spectra: in the DQF-COSY spectrum, the $C^{\alpha}H$–$C^{\beta}H$ and $C^{\delta}H$–$C^{\epsilon}H$ cross-peaks were identified; the DQF-RELAY spectrum contained $C^{\gamma}H$–$C^{\epsilon}H$ cross-peaks (at τ_m = 26–45 ms); in the MLEV spectrum, complete side-chain coherence transfers were observed, bringing together all of the subspin system identifications. Three of the four arginine spin systems were also completely identified with the analogous procedure. The lysine and arginine residues incompletely assigned had $C^{\alpha}H$ resonances under the solvent peak at 50 °C.

Partial identification of the proline resonances could be made in the DQF-COSY spectrum: the $C^{\delta}H^{a}$–$C^{\delta}H^{b}$ cross-peaks were identified by the chemical shifts and cross-peak patterns. In addition, all three proline spin systems had distinctive (i.e., weak) $C^{\gamma}H$–$C^{\delta}H$ cross-peaks though only two or three of the four possible cross-peaks (since the $C^{\delta 2}H^{a}$,H^{b} and $C^{\gamma 2}H^{a}$,H^{b} protons have nondegenerate chemical shifts) were usually found in the DQF-COSY or DQF-RELAY spectrum. The assignments of the proline $C^{\delta}H$ resonances were confirmed by the identification of pseudo-$d_{\alpha N}$ and -d_{NN} sequential connectivities (where the proline $C^{\delta}H$ protons substitute for the am-

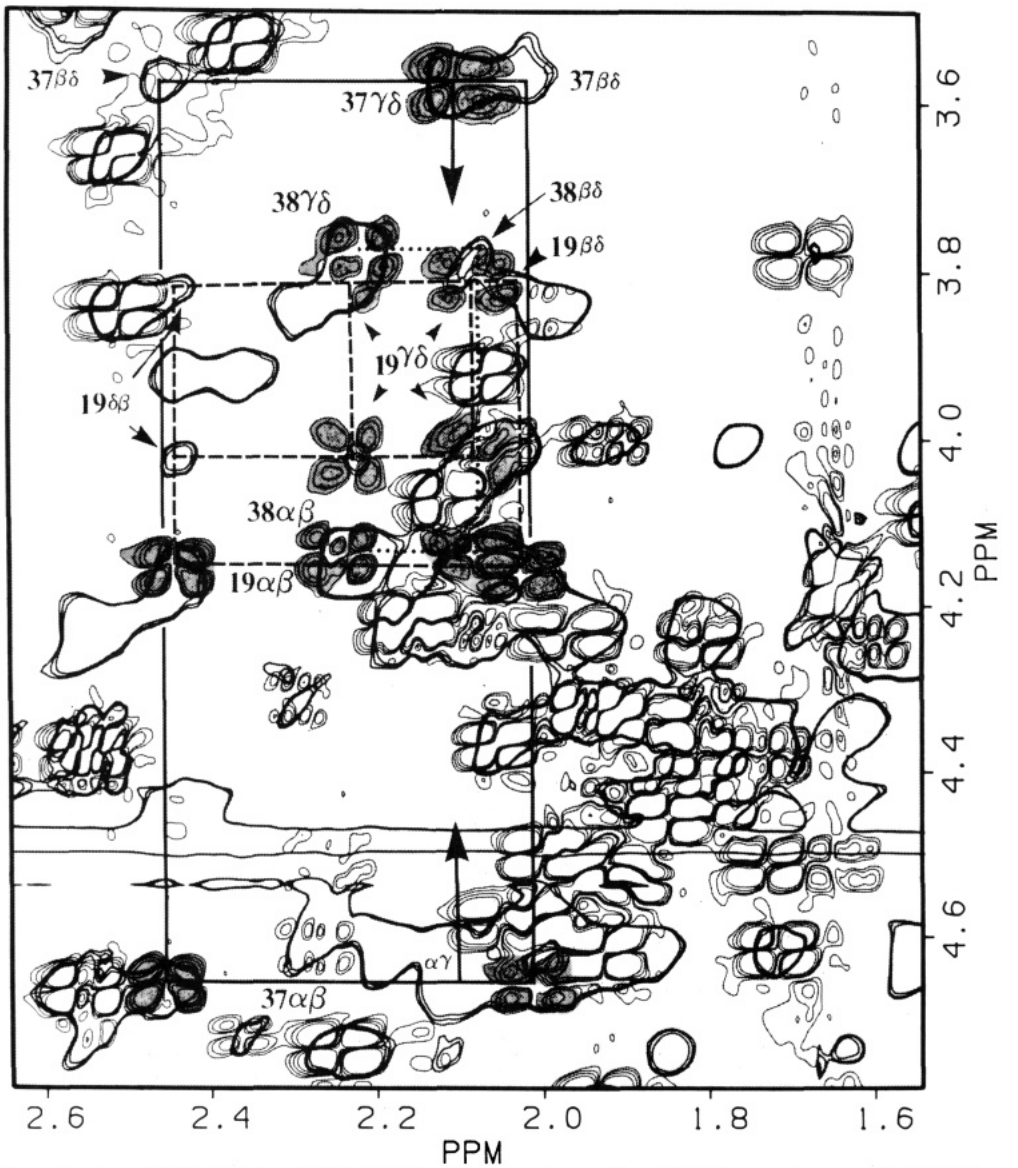

FIGURE 3: Connection of proline [C$^\alpha$H, C$^\beta$H] and [C$^\gamma$H, C$^\delta$H] subspin systems with an MLEV experiment. The MLEV spectrum (in-phase cross-peaks, three closely spaced contours) is superimposed on a DQF-COSY spectrum (antiphase cross-peaks, four more widely spaced contours); the proline COSY cross-peaks are shaded. The P-19 subspin systems are connected (dashed lines) by three C$^\beta$H–C$^\delta$H MLEV cross-peaks. The P-37 subspin systems are connected (solid lines) by two C$^\beta$H–C$^\delta$H MLEV cross-peaks and by one C$^\alpha$H–C$^\gamma$H cross-peak (arrows connect it to the corresponding C$^\gamma$H–C$^\delta$H COSY cross-peak). The P-38 subspin systems are connected (dotted lines) by a single clear C$^\beta$H–C$^\delta$H MLEV cross-peak.

ide). Three sets of C$^\alpha$H–C$^\beta$H cross-peaks were identified as proline subspin systems as they did not connect to amide cross-peaks. The [C$^\alpha$H, C$^\beta$H] subspin systems were connected to the [C$^\delta$H, C$^\gamma$H] subspin systems with the MLEV data (Figure 3). Two unambiguous C$^\beta$H–C$^\delta$H cross-peaks in the MLEV spectrum were used to connect the subspin systems for Pro-37 and Pro-19; the Pro-38 subspin systems were "connected" by elimination at this point, although a single C$^\beta$H–C$^\delta$H cross-peak was observed in the MLEV data. The Pro-38 assignments were also confirmed by the identification of $d_{\beta N}$ NOEs to Asp-39 and a pseudo-$d_{\alpha N}$ NOE between the Pro-37 C$^\alpha$H and a Pro-38 C$^\delta$H resonance.

By far, the most difficult residues to identify completely were the Glx and Met residues; such difficulty was also seen in the only other reported case of virtually complete resonance assignments of a protein near the size of ubiquitin (Kline & Wüthrich, 1986). In the case of ubiquitin, many of the incomplete side-chain assignments are due to the spectral overlap in the relevant spectral regions (six glutamate, six glutamine, and one methionine giving rise to a possible 26 C$^\alpha$H–C$^\beta$H

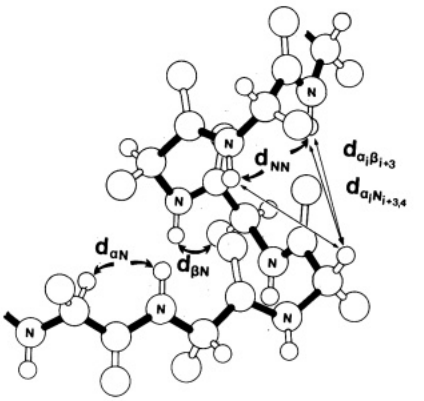

FIGURE 4: Definition of sequential NOEs. The three "major" sequential NOEs are identified with thick arrows, while additional NOEs used in defining helical structure are indicated with thin arrows.

cross-peaks in the same spectral region as the other 31 other long-chain residues). Clear identifications were typically possible for those residues in the β-sheet region of the protein,

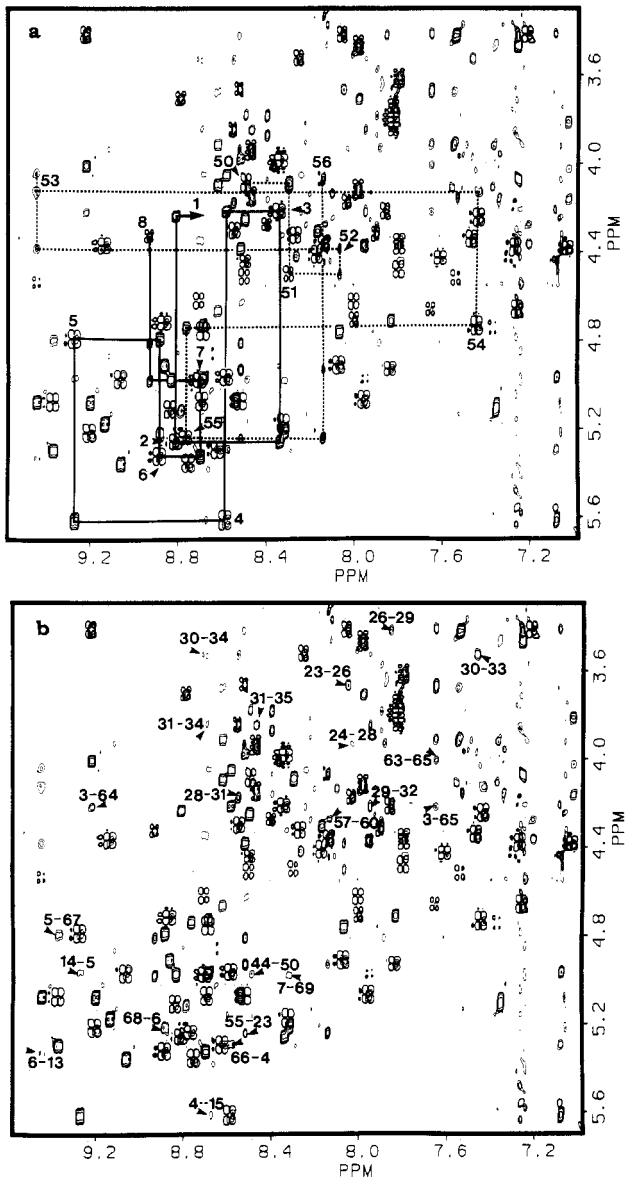

FIGURE 5: $d_{\alpha N}$ NOEs in ubiquitin at 50 °C. NOESY and DQF-COSY spectra are superimposed, but cross-peaks originating from the two experiments can be distinguished by their in-phase (NOESY) or antiphase (DQF-COSY) nature. (a) Connectivities from M-1 to L-8 (solid line) and from L-50 to L-56 (dotted line) are followed. *Arrowheads* identify COSY cross-peaks in crowded regions. (b) Nonsequential $d_{\alpha N}$ NOEs arising from cross-strand proximities in the β-sheet, or to helical repeat. The numbers identify the α and amide resonances, respectively.

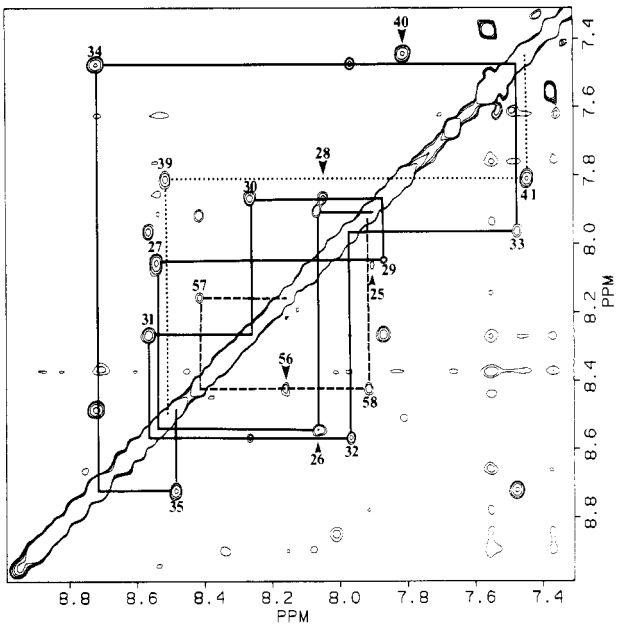

FIGURE 6: d_{NN} NOEs in ubiquitin at 50 °C. Solid, dashed, and dotted lines follow the connectivities through different parts of the molecule. For clarity, the d_{NN} NOEs connecting the helix (N-25–G-35) to residues I-23, E-24, and I-36 are not included in this expansion, nor are the NOEs for the last residues (Y-59, N-60, I-61) in the helical turn. The amide resonances are labeled either directly above or directly below their position in the spectrum.

where the backbone carbonyls give rise to large dispersions in chemical shifts of the amide and $C^{\alpha}H$ resonances.

These residues were identified by comparing the DQF-COSY and DQF-RELAY data to differentiate between the $C^{\beta}H$ and $C^{\gamma}H$ resonances and subsequently confirmed by the MLEV spectrum. In each case the $C^{\alpha}H$–$C^{\beta}H$ cross-peaks were identified in the DQF-COSY spectrum, with identification of corresponding $C^{\beta}H^{a}$–$C^{\beta}H^{b}$ cross-peaks whenever possible. When one of the $C^{\alpha}H$–$C^{\beta}H$ cross-peaks was too weak to observe in the DQF-COSY spectrum, it would be identified in the DQF-RELAY spectrum where its intensity was increased due to a $C^{\alpha}H$–$C^{\beta}H^{a}$–$C^{\beta}H^{b}$ relay (through the geminal coupling). To distinguish this from a $C^{\alpha}H$–$C^{\gamma}H$ cross-peak, a $C^{\beta}H^{a}$–$C^{\beta}H^{b}$ cross-peak that clearly arose from a geminal coupling had to be identified. It may also be possible to sort out the identities of cross-peaks with PE-COSY, where passive couplings can be measured accurately (Mueller, 1987) and can yield information on the identity of a particular cross-peak (e.g., a negative $J_{C^{\beta}H^{a}-C^{\beta}H^{b}}$ geminal coupling could be distinguished from a weaker positive $J_{C^{\beta}H-C^{\gamma}H}$ vicinal coupling). We found this to be important, as the chemical shift values of the Q-31 and E-64 $C^{\gamma}H$ resonances were not downfield of the $C^{\beta}H$ resonances, as normal. Identification using the MLEV spectra and chemical shift arguments would result in incorrect assignments. For further redundancy, an MLEV was obtained with a long mixing time (62 ms, 12.8-kHz applied radio-frequency field); many long-range coherence transfers were observed, and the resolution provided by the spread of NH chemical shifts resulted in additional assignments.

Sequential NOEs (defined in Figure 4) were identified in NOESY spectra acquired in H_2O with relatively long mixing times (175 ms) to yield strong NOE cross-peaks, although the NOE intensity buildup is no longer linear and some spin-diffusion is expected to be present (Weber, 1985; Wagner et al., 1986). Figures 5 and 6 show $d_{\alpha N}$ and d_{NN} sequential connectivities, respectively. Sequential connectivities for those resonances affected by solvent irradiation were obtained by collecting data at 30 °C. No deviations in the NOEs were observed between 30 and 50 °C, consistant with no structural changes occurring between these two temperatures. An exception was that the $d_{\alpha N}$ NOEs for residues 72–75 were weaker in the 50 °C data sets, presumably due to increased motion in the C-terminal tail, which is extremely flexible in solution (as judged by line widths and $J_{NH-C^{\alpha}H}$ values). No NOE was observed between G-75 and G-76 in any data set acquired, but the identity of G-76 was confirmed by obtaining a DQF-COSY spectrum of ubiquitin with a Gly → Ala mutation at position 76. The practice of obtaining additional data sets at different temperatures may be time consuming, but it is helpful in more ways than for just observing NH–$C^{\alpha}H$ cross-peaks. Differential chemical shift changes between two temperatures help corroborate many assignments as correspondence between migrating COSY, NOESY, and MLEV peaks can be made to resolve ambiguity; this type of analysis will be most helpful for assigning larger molecules.

Table I: Ubiquitin ^{1}H NMR Resonance Assignments at 50 °C in 25 mM Acetic Acid, pH 4.7

residue	NH	$C^\alpha H$	$C^\beta H$	others		
M-1	n.o.[a]	4.22	2.17, (2.07)[b]			
Q-2	8.82	5.27	1.67, 1.88	$C^\gamma H$, 2.23, 2.23	$N^\epsilon H$, 6.71, 7.52	
I-3	8.33	4.21	1.80	$C^{\gamma 2}H_3$, 0.66	$C^{\gamma 1}H$, 1.10, 0.88	$C^\delta H_3$, 0.61
F-4	8.57	5.61	2.88, 3.08	$C^\delta H$, 7.11	$C^\epsilon H$, 7.28	$C^\zeta H$, (7.28)
V-5	9.26	4.79	1.93	$C^{\gamma 1}H$, 0.71	$C^{\gamma 2}H$, 0.76	
K-6	8.87	5.34	1.71, 1.40	$C^\gamma H$, 1.31, 1.49	$C^\delta H$, 1.60, 1.60	$C^\epsilon H$, 2.92, 2.92
T-7	8.69	5.00	4.80	$C^{\gamma 2}H_3$, 1.21		
L-8	8.91	4.33	1.95, 1.80	$C^\gamma H$, 1.90	$C^\delta H_3$, 1.06, 1.00	
T-9	7.59	4.41	4.60	$C^{\gamma 2}H_3$, 1.28		
G-10	7.78	3.62, 4.36				
K-11	7.25	4.37	1.73, 1.81	$C^\gamma H$, 1.43, 1.28	$C^\delta H$, 1.66, 1.66	$C^\epsilon H$, 2.95, 2.95
T-12	8.50	5.08	3.98	$C^{\gamma 2}H_3$, 1.11		
I-13	9.44	4.53	1.89	$C^{\gamma 2}H_3$, 0.90	$C^{\gamma 1}H$, 1.12, 1.46	$C^\delta H_3$, 0.75
T-14	8.57	4.97	4.05	$C^{\gamma 2}H_3$, 1.14		
L-15	8.67	4.75	1.26, 1.38	$C^\gamma H$, 1.45	$C^\delta H_3$, 0.74, 0.79	
E-16	8.05	4.92	1.87, 1.94	$C^\gamma H$, 2.25, 2.11		
V-17	8.85	4.71	2.36	$C^{\gamma 1}H_3$, 0.47	$C^{\gamma 2}H_3$, 0.74	
E-18	8.68	5.08	1.67, (1.65)	$C^\gamma H$, 2.27, 2.38		
P-19		4.16	2.44, 2.02	$C^\gamma H$, 2.09, 2.23	$C^\delta H$, 3.82, 4.02	
S-20	7.01	4.38	3.81, 4.17			
D-21	7.98	4.70	2.54, 2.97			
T-22	7.82	4.95	4.80	$C^{\gamma 2}H_3$, 1.28		
I-23	8.51	3.66	2.51	$C^{\gamma 2}H_3$, 0.81	$C^{\gamma 2}H$, 1.90, 1.32	$C^\delta H_3$, 0.61
E-24	9.82	3.92	2.07, 2.07	$C^\gamma H$, 2.37, 2.44		
N-25	7.87	4.56	2.88, 3.21	$N^\delta H$, 6.83, 7.80		
V-26	8.04	3.41	2.38	$C^{\gamma 1}H_3$, 0.72	$C^{\gamma 2}H_3$, 1.00	
K-27	8.52	4.57	2.01, 1.92			
A-28	8.02	4.17	1.65			
K-29	7.84	4.22	2.16, 2.07	$C^\gamma H$, 1.61, 1.69	$C^\delta H$, 1.80, 1.51	$C^\epsilon H$, 3.03, 3.03
I-30	8.24	3.52	2.37	$C^{\gamma 2}H_3$, 0.70	$C^{\gamma 2}H$, 2.02, 0.72	$C^\delta H_3$, 0.90
Q-31	8.54	3.84	2.50, 2.00	$C^\gamma H$, 2.35, 2.35	$N^\epsilon H$, 6.23, 6.50	
D-32	7.94	4.36	2.78, 2.88			
K-33	7.45	4.34	2.02, 1.88	$C^\gamma H$, 2.12, 2.25	$C^\delta H$, 1.74, 1.64	$C^\epsilon H$, 3.14, 3.18
E-34	8.69	4.61	1.72, 2.28	2.11, 2.18		
G-35	8.45	4.14, 3.95				
I-36	6.17	4.45	1.46	$C^{\gamma 2}H_3$, 0.96	$C^{\gamma 2}H$, 1.12, 1.43	$C^\delta H_3$, 0.80
P-37		4.66	2.46, 2.01	$C^\gamma H$, 2.10, 2.05	$C^\delta H$, 3.58, 4.19	
P-38		4.14	2.08, 2.24	$C^\gamma H$, 1.68, 2.21	$C^\delta H$, 3.78, 3.93	
D-39	8.49	4.44	2.70, 2.79			
Q-40	7.78	4.46	1.85, 1.85	$C^\gamma H$, 2.40, 2.40		
Q-41	7.41	4.32	1.97, 1.88	$C^\gamma H$, 2.54, 2.54		
R-42	8.49	4.50	1.73, 1.65	$C^\gamma H$, 1.58, 1.44	$C^\delta H$, 3.16, (3.16)	$N^\epsilon H$, 7.10
L-43	8.74	5.35	1.59, 1.19	$C^\gamma H$, 1.50	$C^\delta H_3$, 0.79, 0.82	
I-44	9.05	4.96	1.74	$C^{\gamma 2}H_3$, 0.71	$C^{\gamma 2}H$, 1.08, 1.38	$C^\delta H_3$, 0.70
F-45	8.82	5.11	2.83, 3.04	$C^\delta H$, 7.39	$C^\epsilon H$, 7.56	$C^\zeta H$, 7.50
A-46	8.77	3.70	0.90			
G-47	7.96	4.10, 3.45				
K-48	7.99	4.61	1.93, (2.01)	$C^\gamma H$, 1.82, (1.82)	$C^\delta H$, 1.56, (1.74)	$C^\epsilon H$, 3.19, 3.19
Q-49	8.54	4.58	2.04, 1.45	$C^\gamma H$, 2.68, 1.69		
L-50	8.48	4.08	1.52, 1.05	$C^\gamma H$, 1.48	$C^\delta H_3$, −0.11, 0.54	
E-51	8.28	4.49	2.26, 2.00	$C^\gamma H$, 2.37, 2.46		
D-52	8.06	4.36	3.32, 2.79			
G-53	9.44	4.03, 4.11				
R-54	7.43	4.73	2.13, 2.23	$C^\gamma H$, 1.65, 1.85	$C^\delta H$, 3.11, 3.17	$N^\epsilon H$, 7.08
T-55	8.75	5.22	4.54	$C^{\gamma 2}H_3$, 1.15		
L-56	8.13	4.06	2.12, 1.26	$C^\gamma H$, 1.74	$C^\delta H_3$, 0.65, 0.78	
S-57	8.38	4.27	3.78, 3.87			
D-58	7.88	4.31	2.30, 2.99			
Y-59	7.24	4.64	2.56, 3.46	$C^\delta H$, 7.28	$C^\epsilon H$, 6.92	
N-60	8.12	4.37	2.65, 2.51			
I-61	7.19	3.41	1.43	$C^{\gamma 2}H_3$, 0.50	$C^{\gamma 1}H$, 1.14, −0.22	$C^\delta H_3$, 0.41
Q-62	7.52	4.51	1.92, 2.27	$C^\gamma H$, 2.33, 2.40		
K-63	8.36	4.00	1.92, 2.06	$C^\gamma H$, 1.52, 1.52	$C^\delta H$, 1.76, 1.76	$C^\epsilon H$, 3.06, 3.06
E-64	9.20	3.40	2.45, 2.56	$C^\gamma H$, 2.31, 2.31		
S-65	7.64	4.65	3.68, 3.90			
T-66	8.61	5.32	4.08	$C^{\gamma 2}H_3$, 0.97		
L-67	9.36	5.08	1.67, (1.49)	$C^\gamma H$, 1.79	$C^\delta H_3$, 0.69, 0.72	
H-68	9.19	5.21	2.96, 3.12	$C^{\delta 2}H$, 7.15	$C^{\epsilon 1}H$, 8.34	
L-69	8.30	5.19	1.12, 1.64	$C^\gamma H$, 1.37	$C^\delta H_3$, 0.76, 0.88	
V-70	9.13	4.36	2.05	$C^{\gamma 1}H_3$, 0.87	$C^{\gamma 2}H_3$, 0.95	
L-71	7.94	5.07	1.59, 2.18	$C^\gamma H$, 1.69	$C^\delta H_3$, 0.88, 0.98	
R-72	8.53	4.29	1.56, 1.78	$C^\gamma H$, 1.53, 1.86	$C^\delta H$, 3.19, 3.19	$N^\epsilon H$, 7.08
L-73	8.15	4.41	1.65, (1.65)	$C^\gamma H$, 1.63	$C^\delta H_3$, 0.89, 0.94	
R-74	8.25	4.32	1.80, 1.89	$C^\gamma H$, 1.65, 1.65	$C^\delta H$, 3.22, 3.22	$N^\epsilon H$, 7.22
G-75	8.32	(3.97), (3.97)				
G-76	7.80	(3.73), (3.82)				

[a] n.o., not observed. [b] Chemical shifts are reported as ±0.01 ppm relative to DSS; values enclosed in parentheses are uncertain assignments.

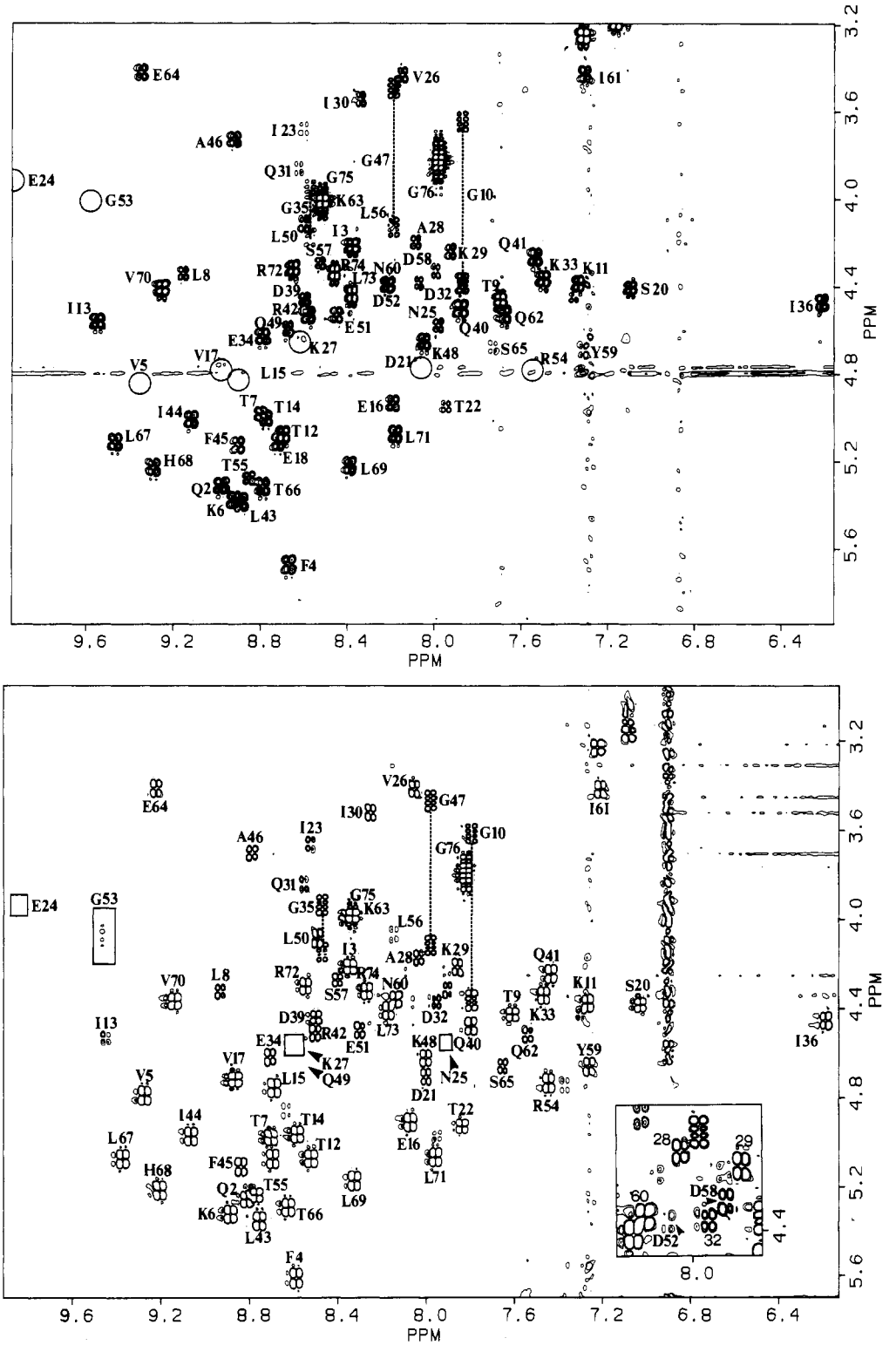

FIGURE 7: DQF-COSY fingerprint regions of ubiquitin at 30 °C (above) and 50 °C (below). Locations of cross-peaks too weak to observe at the lowest contour level plotted or those saturated by the decoupler are indicated by circles in the top spectrum and by boxes in the bottom spectrum. The inset at bottom lower right is an expanded region plotted at a lower level to show the D-52 and D-58 cross-peaks, which occur in a crowded spectral region. Dotted lines in both data sets connect the glycine NH-C$^{\alpha}$H cross-peaks for those with nondegenerate C$^{\alpha}$H chemical shifts.

Table I contains assignments of ubiquitin in 25 mM acetic acid-d_4, pH 4.7 at 50 °C. The fingerprint region is shown in Figure 7 for data collected at 30 and 50 °C. The resolution in this region is typical of a protein this size with extensive β-sheet character. Large variance in the chemical shifts of the amide protons is also seen at both ends of the helices, presumably due to the net helix dipole. Such effects have also been seen in other helix-containing proteins (Zuiderweg et al., 1983; Weber et al., 1985b; Wand & Englander, 1986; Klevit et al., 1986).

Discussion

Resonance assignments are a tedious but essential first step in the NMR characterization of a molecule before the wealth of NMR information can be fully interpreted. We have shown here that such assignments are possible to a very thorough level

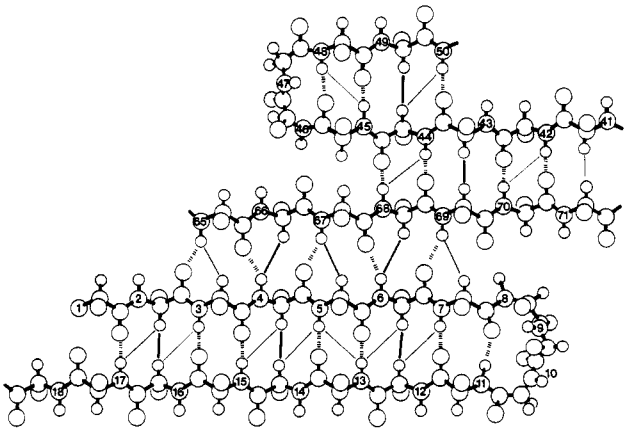

FIGURE 8: β-Structure of ubiquitin. Cross-strand NOEs are indicated by solid lines, where thick lines represent strong NOEs and thin lines weak NOEs. The dashed lines indicate hydrogen bonds inferred from slowly exchanging amide resonances (data not shown); the positions of the hydrogen bonds at the ends of the sheet are speculative. The two reverse turns shown were drawn to roughly correspond to observed NOE patterns.

for a protein with ca. 600 protons, and are likely to be possible for larger proteins. Of course, ubiquitin represents a best case scenario since the protein (i) is highly soluble (5 mM protein concentrations can be attained without noticeable aggregation), (ii) is heat stable, allowing spectra to be recorded at higher temperatures where line widths are narrower and thus solvent saturation is facilitated by increased T_1 of H_2O, and (iii) has a well-resolved spectrum for a protein with 76 amino acids. These qualities aided in the resonance assignments and make ubiquitin an attractive molecule for NMR investigations to study its functions, structure, stability, and folding.

Ubiquitin has been previously studied by NMR (Cary et al., 1980), and some of the structural conclusions from this work were in disagreement with the later crystallographic work (Vijay-Kumar et al., 1985). A discrepancy was noted that the NMR data suggested that H-68, Y-59, and one of the phenylalanine residues were buried, yet the crysal data had all aromatics located on the surface of the protein. Preliminary examination of the NOE data indicate that the H-68 ring is in a very similar if not identical position with that seen in the crystal structure. Thus the position of H-68 in the crystal structure is not the result of an artifact due to mercury binding, but represents the ring's normal position. The position of the other aromatic rings, as evidenced by several NOEs, also appear to be in similar positions to those found in the crystal structure.

The secondary structure was identified from the sequential NOEs and from additional NOEs between cross-strand contacts, providing a high level of redundancy and therefore certainty in the assignments as well as leading to more detailed structural information (Wüthrich et al., 1984). Here we find two very minor discrepancies in the interpretation of the crystal structure. First, an additional short β-strand involving residues 48–50 forms part of a five-stranded β-sheet (see Figure 8). Second, the turn region involving residues 56–61 is described as a "reverse turn" in the crystal work, but we find evidence for a short (possibly 3_{10}) helical turn. Otherwise, the NMR data presented here appear in agreement with the crystal data: the above-mentioned β-sheet; a single helix involving residues 23–34; reverse turns at positons 7–11 (with a β-bulge at position 10), 18–21 (I), and 45–48 (II); other turns at positions 37–41, 56–61, and 62–65 (with a possible β-bulge at residue 64 or 65). Further details concerning the NMR structure as determined by distance geometry and other computational methods will be published elsewhere.

In obtaining the extensive side-chain assignments, we would like to stress the importance of performing the DQF-COSY and DQF-RELAY experiments in addition to the isotropic mixing experiments. It may seem reasonable to make spin system identifications from the MLEV results alone, given (i) the extensive J network information obtained and (ii) that the resultant peaks are (net) absorptive, which avoids loss in sensitivity in the COSY-type experiments due to unresolved antiphase components. But there is a serious pitfall in this approach, since the evolution of cross-peaks in isotropic mixing is quite complex and not simply dependent upon the number of chemical bonds involved: cross-peak intensity oscillates in a manner dependent upon all J couplings in the network. The problem is that it is not possible to unambiguously distinguish direct coherences, single relays, and multiple replays in an isotropic mixing experiment; time dependence of cross-peak appearance may be helpful, but is not unambiguous.

Despite the limited incompleteness of the reported assignments, they represent an "overkill" of the problem at hand. For structural studies, the side chains of many surface residues are not likely to be well-determined in solution, except in those cases where polar groups form hydrogen bonds to the backbone or to other polar side chains. Complete assignments of aliphatic residues, however, are important since contacts between these residues help define the overall folding of the molecule. Careful identification and interproton distance measurements of contacts in the hydrophobic core are desirable for accurate structure determination. Of course, which residues will lie on the surface of the molecule is not strictly known after only performing the sequential assignments, but one can presume that most hydrophobic residues will be buried while hydrophilic residues lie at the surface.

For example, of the seven lysine side chains three (K-6, K-11, K-63) appear to be freely mobile in solution as indicated by degenerate $C^{\epsilon}H^{a,b}$ and $C^{\delta}H^{a,b}$ resonances. Three of the remaining identified lysine side chains (K-29, K 48, K-33) had nondegenerate $C^{\epsilon}H^{a,b}$ and $C^{\delta}H^{a,b}$ resonances, indicating that these side chains are involved in hydrogen bonding to other side chains. For side chains conformationally averaged, no NOEs will be observed and will not contribute to a distance geometry calculation of the structure. Hence, these side chains could be eliminated in distance geometry calculations to reduce conputational time. Side-chain assignments are important when protein–protein or protein–DNA interactions are of vital interest—then it becomes vital to identify longer side-chain resonances, though it will be difficult without resorting to multiple-quantum filtered or multiple-quantum experiments (Müller et al., 1986; Rance & Wright, 1986; Bach et al., 1987), as the spectral region containing many of the side-chain cross-peaks becomes extremely crowded in a DQF-COSY spectrum.

It is clear that the sequential resonance assignment method has no major difficulties for proteins of the size of ubiquitin and is likely to be useful for much larger proteins (perhaps up to M_r 20000). The method is, however, labor intensive and limited in that the process first calls for amino acid identifications, which are increasingly difficult in larger proteins. In the preceding paper, Di Stephano and Wand demonstrate a new assignment method that is more likely to be useful for larger proteins, as well for computer-assisted resonance assignments. Their assignments of the ubiquitin ^{1}H NMR spectrum are in agreement with our independent assignments by the sequential method.

After preparation of this paper, the crystal structure of ubiquitin refined at 1.8-Å resolution was published (Vijay-Kumar et al., 1987). In this structure the same 3_{10} helical turn as described above was seen, as are the fifth strand in the β-sheet (residues 48–51) and the two β-bulges. The second bulge involving E-64 and S-65 results in a left-handed conformation of E-64, which could be causing the high-field shift in the E-64 $C^{\alpha}H$ resonance. In light of the data presented here, the low occupancy in the X-ray data of the C-terminal tail is due to thermal motion and not sample heterogeneity, as our ubiquitin preparations did not suffer from the loss of the terminal Gly-Gly residues (Ecker et al., 1987a,b). Also, the nondegeneracy of the K-33 $C^{\delta}H^{a,b}$ and $C^{\epsilon}H^{a,b}$ resonances suggests that the side chain of this residue is not free in solution; likewise, the degeneracy of the same proton resonances for K-11 suggests that this residue is freely mobile in solution. The reverse was found in the crystal data (Vijay-Kumar et al., 1987).

Acknowledgments

We thank Drs. David Ecker, Tauseef Butt, Stanley Crooke, and Jon Marsh and their colleagues for the wild-type and mutant ubiquitin samples and Drs. Josh Wand and Deena Di Stephano for making available a preprint of their ubiquitin assignments.

Supplementary Material Available

Thirteen figures depicting the following: DQ-COSY and DQF-COSY spectra in H_2O showing the identification of the glycine residues; NOESY and COSY contour plots showing the identification of the aromatic resonances; DQF-COSY spectra showing all $C^{\alpha}H$–$C^{\beta}H$ assignments (three panels); DQF-COSY and DQF-RELAY spectral comparisons that identify $C^{\alpha}H$–$C^{\gamma}H_3$ cross-peaks of valine, isoleucine, and threonine residues; three MLEV contour plots (from a spectrum obtained in D_2O solution) showing the extended coherence transfers in (1) lysine, arginine, glutamate, and glutamine residues, (2) valine, isoleucine, and leucine residues, and (3) isoleucine and leucine $C^{\beta}H$–$C^{\delta}H_3$ cross-peaks; MLEV contour plot (spectrum acquired in H_2O solution) showing extended coherence transfers between amide and side-chain protons; combined NOESY/DQF-COSY spectra showing $d_{\alpha N}$ connectivities not shown in this paper (two panels); and a NOESY spectrum showing cross-strand $C^{\alpha}H$–$C^{\alpha}H$ NOEs and proline pseudo-$d_{\alpha N}$ NOEs (15 pages). Ordering information is given on any current masthead page.

Registry No. Ubiquitin, 60267-61-0.

References

Audhya, T., & Goldstein, G. (1985) *Methods Enzymol. 116*, 279–291.

Bach, A. C., II, Selsted, M. E., & Pardi, A. (1987) *Biochemistry 26*, 4389–4397.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 312–346.

Busch, H. (1984) *Methods Enzymol. 106*, 238–262.

Cary, P. D., King, D. S., Crane-Robinson, C., Bradbury, E. M., Rabbani, A., Goodwin, G. H., & Johns, E. W. (1980) *Eur. J. Biochem. 112*, 577–580.

Di Stephano, D. L., & Wand, A. J. (1987) *Biochemistry* (preceding paper in this issue).

Ecker, D. J., Khan, M. I., Marsh, J., Butt, T., & Crooke, S. T. (1987a) *J. Biol. Chem. 262*, 3524–3527.

Ecker, D. J., Butt, T. R., Marsh, J., Sternberg, E. J., Margolis, N., Monia, B. P., Weber, P. L., Mueller, L., & Crooke, S. T. (1987b) *J. Biol. Chem.* (in press).

Finley, D., & Varshavsky, A. (1985) *Trends Biochem. Sci. (Ref. Ed.) 10*, 343–347.

Hershko, A., & Ciechanover, A. (1982) *Annu. Rev. Biochem. 51*, 335–364.

Klevit, R. E., & Drobny, G. P. (1986) *Biochemistry 25*, 7770–7773.

Klevit, R. E., Drobny, G. P., & Waygood, E. B. (1986) *Biochemistry 25*, 7760–7769.

Kline, A. D., & Wüthrich, K. (1986) *J. Mol. Biol. 192*, 869–890.

Levitt, M. H., Freeman, R., & Frenkiel, T. (1982) *J. Magn. Reson. 47*, 328.

Mueller, L. (1987) *J. Magn. Reson. 72*, 191–196.

Mueller, L., & Ernst, R. R. (1979) *Mol. Phys. 38*, 963–992.

Müller, N., Ernst, R. R., & Wüthrich, K. (1986) *J. Am. Chem. Soc. 108*, 6482–6492.

Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.

Rance, M., & Wright, P. E. (1986) *J. Magn. Reson. 66*, 372–378.

Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169–173.

Sørensen, O. W. (1984) D.N.S. Dissertation, ETH-Zürich, ADAG Administration & Druck AG.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.

Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., & Cook, W. J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3582–3585.

Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., & Cook, W. J. (1987) *J. Mol. Biol. 194*, 531–544.

Wagner, G., & Zuiderweg, E. R. P. (1983) *Biochem. Biophys. Res. Commun. 113*, 854–860.

Wagner, G., Neuhaus, D., Wörgötter, E., Vašák, M., Kägi, J. H. R., & Wüthrich, K. (1986) *Eur. J. Biochem. 157*, 275–289.

Wand, A. J., & Englander, S. W. (1986) *Biochemistry 25*, 1100–1106.

Weber, P. L. (1985) Ph.D. Dissertation, University of Washington, Seattle, WA.

Weber, P. L., & Mueller, L. M. (1987) *J. Magn. Reson. 73*, 184–190.

Weber, P. L., Drobny, G. P., & Reid, B. R. (1985a) *Biochemistry 24*, 4549–4552.

Weber, P. L., Wemmer, D. E., & Reid, B. R. (1985b) *Biochemistry 24*, 4553–4562.

Weber, P. L., Sieker, L., Samy, T. S. A., Reid, B. R., & Drobny, G. P. (1987) *J. Am. Chem. Soc. 109*, 5842–5844.

Wokuan, A., & Ernst, R. R. (1977) *Chem. Phys. Lett. 52*, 407–412.

Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311–319.

Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.

Zuiderweg, E. R. P., Kaptein, R., & Wüthrich, K. (1983) *Eur. J. Biochem. 137*, 279–292.

Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985) *Biopolymers 24*, 601–611.