Two-Dimensional ^{1}H NMR Study of Human Ubiquitin: A Main Chain Directed Assignment and Structure Analysis†

Deena L. Di Stefano and A. Joshua Wand*
Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111
Received May 27, 1987; Revised Manuscript Received July 6, 1987

ABSTRACT: The ^{1}H resonances of human ubiquitin were studied by two-dimensional nuclear magnetic resonance techniques. A recently introduced assignment algorithm termed the main chain directed (MCD) assignment [Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953–5958] was applied. This approach relies on an ordered series of searches for prescribed patterns of connectivities in two-dimensional *J*-correlated and nuclear Overhauser effect spectra and centers on the dipolar interactions involving main-chain amide NH, α-CH, and β-CH. Unlike the sequential assignment procedure, the MCD approach does not rest upon definition of side-chain *J*-coupled networks and is generally not sequential with the primary sequence of the protein. The various MCD patterns and the general algorithm are reiterated and applied to the analysis of human ubiquitin. With this algorithm, the vast majority of amino acid residue amide NH–C_αH–C_βH *J*-coupled subspin systems could be associated with and aligned within units of secondary structure without any knowledge of the identity of the side chains. This greatly simplified recognition of side-chain spin systems by restricting their identity. Essentially complete resonance assignments are presented. The MCD method is compared with the sequential assignment method in some detail. The MCD method is highly amenable to automation. Human ubiquitin is found, at pH* 5.8 and 30 °C, to be composed of an extensive β-sheet structure involving five strands. Three of these strands form an antiparallel set sharing a common strand and have a parallel orientation to two antiparallel strands. Two helical segments were also observed. The largest, spanning 13 residues, shows dipolar interactions consistent with an α-helix while the smaller 4-residue helical segment appears, on the basis of observed nuclear Overhauser effects, to be a 3_{10} helix. Five classical tight turns could be demonstrated.

Two-dimensional nuclear magnetic resonance techniques have now led to the nearly complete resonance assignments of a number of small proteins. The assignment of proton resonances in small proteins has usually proceeded by the application of the direct sequential assignment method introduced and developed by Wüthrich and co-workers (Wüthrich et al., 1982; Billeter et al., 1982; Wüthrich, 1983). This procedure hinges on the ability to define and identify the majority of amino acid *J*-coupled side-chain spin systems and commonly entails the comprehensive application of a wide variety of types of *J*-correlated spectroscopy (Wider et al., 1984; Wüthrich, 1986). In general, identification of amino acid side-chain *J*-coupled networks requires analysis of the most dense and complicated region of *J*-correlated spectra. Successful application of the sequential assignment procedure necessarily requires comprehensive analysis of this region. Such an analysis becomes rapidly more difficult with increasing protein size.

During the course of our work with horse ferrocytochrome *c* (Wand & Englander, 1985; Wand et al., 1986), which is the largest protein for which essentially complete resonance assignments are available,[1] we were forced to develop an assignment strategy that was distinct from the sequential assignment procedure. This method, termed the main chain directed assignment procedure or MCD[2] (Englander & Wand, 1987), is based on the formal pattern recognition analysis of the NOESY spectrum and does not require the difficult extensive analysis of side-chain resonances to provide main-chain proton assignments and alignment of residues within units of secondary structure. Furthermore, the assignment is generally not linear with the primary sequence. The assignment of the ^{1}H resonances of the ubiquitin protein, described here, serves to illustrate the MCD method and its advantages and disadvantages with respect to the sequential assignment method.

Ubiquitin is a relatively small, 76 amino acid protein that has been found to be a central component of the ATP-dependent eucaryotic system for control of protein turnover (Ciechanover et al., 1978; Hershko et al., 1979). In targeting proteins for degradation, ubiquitin is covalently linked, via its C-terminus, to the target protein (Ciechanover et al., 1980a; Hershko et al., 1980). A detailed structural analysis of ubiquitin is important for understanding the physical properties of this interaction. The NMR data analysis yields the general secondary structure of the ubiquitin protein in solution. Some differences with the initially reported crystal structure (Vijay-Kumar et al., 1985) appear and are consistent with the more recently reported analysis (Vijay-Kumar et al., 1987).

Materials and Methods

Human ubiquitin was a gift from Drs. A. Hershko and I. Rose and was isolated as described (Ciechanover et al., 1980b). All solvents and buffers were of reagent grade and were used without further purification. Quoted pH* values are uncor-

† Supported, in part, by NIH Grant GM 35940 awarded to A.J.W. and by NIH Grants CA 06927 and RR 05539, by an award from Marie Z. Cole Montrose toward the purchase of the AM 500, by a grant from the Pew Memorial Trust, and by an appropriation from the Commonwealth of Pennsylvania awarded to the Institute for Cancer Research.

[1] Manuscripts in preparation.

[2] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; MCD, main chain directed; COSY, *J*-correlated spectroscopy; NOESY, NOE-correlated spectroscopy; RCT, relayed coherence transfer; DQF, double-quantum filter; TOCSY, total *J*-correlated spectroscopy; FID, free induction decay; ppm, parts per million; HPLC, high-performance liquid chromatography.

0006-2960/87/0426-7272$01.50/0

rected for the isotope effect. All spectra were obtained at 30 °C and at pH* 5.8 with samples prepared to 8 mM ubiquitin in 50 mM K_3PO_4 buffer in 90% H_2O/10% D_2O or 99.9% D_2O as required. Quoted chemical shifts are referenced to an external standard of sodium 3-(trimethylsilyl)tetradeuteriopropionate in D_2O (coaxial capillary).

NMR Spectroscopy. All NMR spectra were obtained on a Bruker Instruments AM 500 NMR spectrometer operating at 500.13 MHz for protons. Phase-sensitive DQF COSY (Shaka & Freeman, 1983; Rance et al., 1984), NOESY (Macura & Ernst, 1980), and isotropic mixing (TOCSY) (Braunschweiler & Ernst, 1983; Bax & Davis, 1985) spectra were obtained as described and employed time proportional phase incrementation (Redfield & Kuntz, 1975; Marion & Wüthrich, 1983) to provide quadrature detection during the incremented time domain. The TOCSY experiments used a repetitive MLEV-17 spin-locking pulse train (Bax & Davis, 1985) without flanking "trim" pulses to effect isotropic mixing in the rotating frame. Mixing times for NOESY and TOCSY spectra are provided in the figure legends. Relayed coherence transfer experiments (RCT COSY) (Eich et al., 1983; Bax & Drobny, 1985) employed the phase cycling of Wagner (1984) to eliminate axial peaks and NOEs generated during the relay 90–τ–180–τ–90 mixing sequence. RCT COSY experiments used a mixing time (2τ) of 60 ms. For all spectra, each free induction decay (FID) was the result of signal averaging of at least 64 scans. The sweep width used was 5813 Hz for most spectra shown. Four dummy scans were used for each FID, and a recycle delay of 2 s including acquisition was employed. FIDs for phase-sensitive spectra were composed of 2048 complex data points. Phase-sensitive data sets were generally composed of at least 700 FIDs spanning an incremented time domain of 3 μs to at least 60 ms. A total of 512 FIDs of 1024 complex data points were collected for the RCT COSY spectrum. When required, solvent suppression was by the method of direct presaturation at all times except during acquisition. Decoupling efficiency was dramatically increased by the use of an electromagnetic shield kindly provided by R. Dykstra (Dykstra, 1987).

Data Processing. Time domain data were transferred to a Vax 11/785 cluster and processed with the software package FTNMR due to Dr. Dennis Hare, Infinity Systems Design. Phase-sensitive data were Gaussian filtered and processed to give 2048 × 2048 real data points providing 2.8 Hz/point resolution. RCT COSY data were resolution enhanced with unshifted sine bell filters and processed to give 2048 × 2048 real data points and are shown in the magnitude mode.

Results

The MCD approach prescribes a hierarchical sequence of searches for predicted NOE connectivity patterns. One attempts to first define amide NH–C_αH–C_βH subspin systems using *J*-correlated spectroscopy. Searches are then carried out, within the NOESY spectrum, for NOE patterns, connecting the amide NH–C_αH–C_βH *J*-coupled units, characteristic of helices, then antiparallel sheets and turns, then parallel sheets, and finally extended chain (Englander & Wand, 1987).

The only *J*-coupled spin system information that is required for the application of the MCD approach is the definition of the amide NH–C_αH–C_βH *J*-coupled subspin system of each residue. This subspin system is by far the easiest component of amino acid *J*-coupled networks to identify and does not require the exhaustive analysis often necessary to fully define and hence identify each amino acid side chain. The NH–C_αH–C_βH subspin systems are easily determined by combined use of the DQF COSY and RCT COSY experiments obtained in H_2O solvent (Figure 1). Examination of the DQF COSY obtained for human ubiquitin unambiguously identified 45 potential amide NH–C_αH–C_βH sets. The comparison of the COSY and RCT COSY removes many ambiguities due to degeneracies among NH, C_αH, and C_βH resonances (Figure 1) and provided an additional 21 potential amide NH–C_αH–C_βH sets. Overall, this analysis yielded 66 fully defined candidate amide NH–C_αH–C_βH *J*-coupled sets and 15 incompletely defined sets. Some of these correspond to side-chain NH–CH–CH *J*-coupled sets. This ambiguity is resolved later in the analysis. Of the incomplete sets, three were due to saturation of the α-proton by the decoupler. The remaining ill-defined NH–C_αH–C_βH sets had an ambiguity at the β-proton level.

Elimination of Extraneous NOEs. The MCD method defines a series of searches, in the NOESY spectrum, for NOE patterns connecting the previously defined NH–C_αH–C_βH *J*-coupled sets. These searches are restricted to the region of the NOESY spectrum spanned by the chemical shifts of the NH, C_αH, and C_βH resonances. Most of this region is shown in Figure 2a. Figure 2b represents those main-chain NOE cross-peaks that remain after all unambiguous intraresidue NOEs and NOEs involving aromatic and other side-chain protons have been eliminated. Ambiguous NOEs are retained, including those that may arise from aromatic protons degenerate with exchangeable NH.

Search for the Helical Pattern. Helical structure, both α and 3_{10}, is the first type of secondary structure considered by the MCD algorithm (Englander & Wand, 1987). Helices consistently display NOE connectivities due to short interresidue $d_{\beta N}$ and d_{NN} distances (Billeter et al., 1982). For helices, the NH–C_αH–C_βH *J*-coupled sets of neighboring residues are connected, in the NOESY spectrum, by $d_{\beta N}$ and d_{NN} NOES that form chains of closed, self-confirmatory loops (Wand & Englander, 1986; Englander & Wand, 1987). In this case, the closed loop is formed by the specific combination of β–amide and amide–amide NOEs leading to a $NH_i \rightarrow NH_{i+1} \rightarrow C_\beta H_i \rightarrow NH_i$ connectivity string. All MCD patterns have this circular or closed-loop property. The NOESY spectrum is searched for loops like these that connect the previously defined NH–C_αH–C_βH *J*-coupled sets. An example of these helical MCD NOE patterns is shown in Figure 3. Search of the ubiquitin NOESY spectrum (Figure 2) for such loops led to the *unambiguous* definition of three helical stretches. Two of these are joined by a glycine (which fails to display the $d_{\beta N}$ NOE). The total length of this MCD-connected helical stretch is 13 residues. The other, smaller helical stretch involves four residues. The location of these helical residues within the primary sequence cannot be specified at this point in the analysis.

The two helical units were then examined for NOEs arising from all possible interactions between the NH–C_αH–C_βH sets involved that are not considered by the MCD pattern for helix (e.g., $d_{\alpha N}$). There were 61 such NOEs in addition to the 20 NOEs included in the standard helical MCD patterns. Additional NOEs [i.e., $d_{\alpha N}(i, i+2)$] indicate that the small helix is a 3_{10} helix. All of these NOEs were then eliminated from further consideration.

Search for the Antiparallel β-Sheet Pattern. Antiparallel β-sheet structure gives rise to a characteristic pattern of main-chain NOEs. As shown in the inset of Figure 4, suitable combinations of NH–C_αH *J*-coupled pathways and NOE interactions form two basic closed loops with a common edge that explicitly involves both strands. In contrast, the sequential assignment would consider each strand independently and

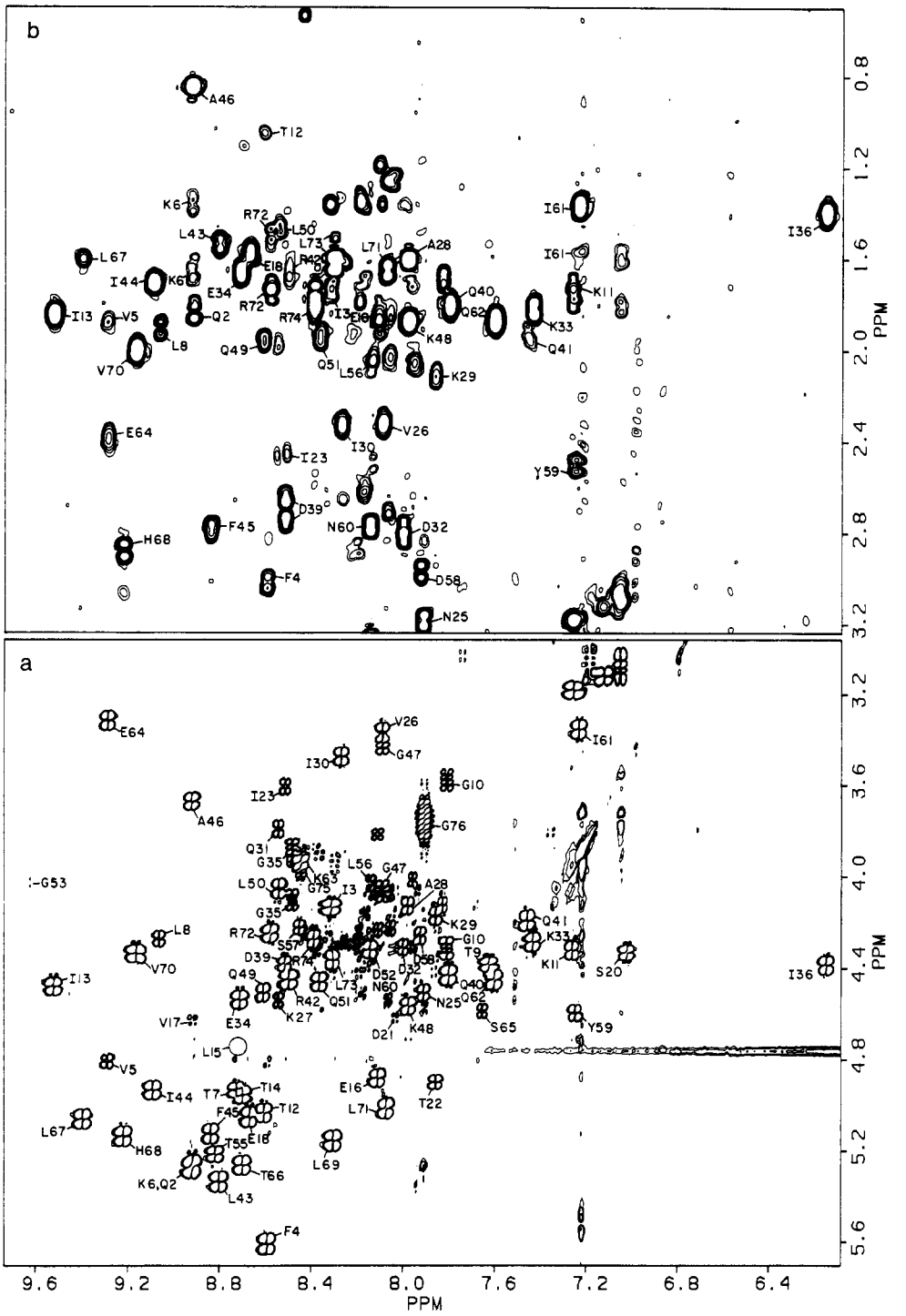

FIGURE 1: (a) Fingerprint region of a DQF COSY of 8 mM human ubiquitin obtained in 90% H_2O/10% D_2O–50 mM K_3PO_4 buffer at pH* 5.8 and 30 °C. Assigned amide–α-proton cross-peaks are indicated. (b) Expansion of a RCT COSY obtained under conditions as described in (a). The region shown predominantly contains amide–$C_\beta H$ relayed coherence transfer cross-peaks and was assigned as indicated. These two spectra were sufficient to unambiguously identify the amide–$C_\alpha H$–$C_\beta H$ J-coupled networks of each amino acid residue.

determine the correct pathway of NOE connectivities by appealing to the identity of the amino acid side chains and the primary structure. Only after each strand has been recognized and defined is the type of secondary structure explicity considered in the sequential assignment. The presence of internally confirmatory closed loops in the connectivity string allows ambiguous branchings of apparent connectivities to be resolved and the true connectivity to be discovered.

Application of the MCD pattern for antiparallel β-sheet to the NOESY spectrum of human ubiquitin led to the unambiguous identification of five β-strands. Three of these strands form one antiparallel set sharing a common strand, and two form another antiparallel set. As before, NOEs arising exclusively from combinations of the NH, $C_\alpha H$, and $C_\beta H$ protons of the residues involved in antiparallel β-sheet structures were then eliminated from further consideration.

Search for Reverse Turns. Tight turns not involving proline were revealed by NOE connectivities that failed to meet the requirements of the MCD patterns for helix and antiparallel β-sheet applied above. Type I turns (Wüthrich et al., 1984) show the d_{NN} and $d_{\beta N}$ NOEs of the helical MCD pattern but terminate after three residues. Type II reverse turns display a subset of the antiparallel β-sheet MCD pattern (Wüthrich et al., 1984; Englander & Wand, 1987). Patterns corre-

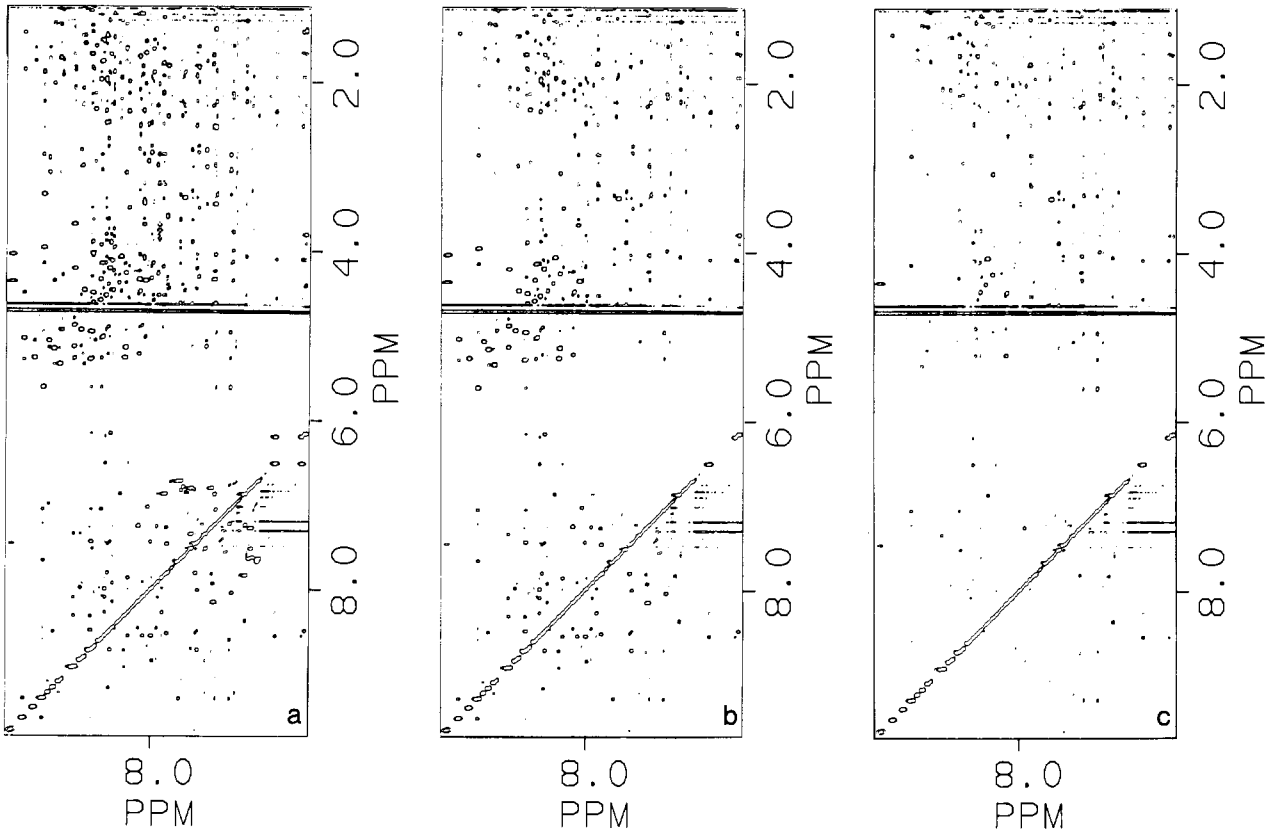

FIGURE 2: Summary of the application of the main chain directed assignment algorithm to the analysis of a phase-sensitive NOESY spectrum of human ubiquitin. A mixing time of 120 ms was used. Other conditions are as described in Figure 1. Panel a is the initial NOESY spectrum. The region shown contains cross-peaks arising from NOEs involving both intra and inter amino acid interactions and includes cross-peaks due to protons other than amide NH, α-CH, and β-CH. Panel b shows the same region of the NOESY spectrum after NOE cross-peaks involving intraresidue interactions and those involving protons other than amide NH, α-CH, and β-CH have been eliminated (in the plot by "white out"). This is possible through complete definition of the NH–C_αH–C_βH J-coupled subspin systems of each amino acid (see Figure 1). Ambiguous cross-peaks (e.g., those possibly due to lysine side chain NH and aromatic protons degenerate with amide NH) are retained. Panel c shows the remaining NOE cross-peaks after exhaustive application of the MCD pattern search for helix, antiparallel, and parallel sheets and type I and II turns not involving proline. The region is considerably simplified and allowed extended-chain NH–C_αH NOE connectivities to be identified with little or no ambiguity.

sponding to two type I turns and three type II turns were found in the NOESY spectrum.

Search for Parallel β-Sheet Search Pattern. The parallel β-sheet structure also leads to a closed-loop connectivity as is indicated in the inset to Figure 5. The exhaustive elimination of $d_{\alpha N}$ NOEs arising from helical and antiparallel β-sheet (see above) makes the search for these patterns much easier and less ambiguous. Application of the MCD pattern indicated in Figure 5 to the NOEs that remain for consideration in the NOESY spectrum led to the identification of a parallel β-sheet orientation of two of the strands involved in the antiparallel β-strand pairs defined above. This immediately leads to the construction of an extensive sheet of antiparallel and parallel strands as indicated in Figure 8. As before, NOEs arising exclusively from combinations of protons due to residues involved in the parallel sheet were then eliminated from further consideration.

Search for Irregular Extended Chain. At this point in the analysis of the NOESY spectrum of human ubiquitin, a total of 58 of the 76 residues could be unequivocally associated with and aligned within a particular unit of secondary structure (helix, sheet, turn). This point was reached without any reference to the character of the side chains of the involved amino acids beyond the distinction of glycine (see the helix above). Following elimination of all NOEs between protons contained within these sets of MCD-defined units of secondary structure, it is now possible to search for the linear and potentially irregular NOE patterns characteristic of extended or "disordered" chain. Unambiguous main-chain NOE connectivities, indicative of extended chain, between nine of the remaining NH–C_αH–C_βH sets were found. Six of these connected previously determined elements of secondary structure. A total of 67 residues were thus unambiguously associated with units of secondary structure. Nine residues, including the three prolines, could not, at this point in the analysis, be unambiguously aligned within or relative to such units of MCD-defined secondary structure. Twelve apparent NH–C_αH–C_βH sets remained, indicating that three of these represent J-coupled side-chain subspin systems and do not involve amide NH.

Sequence Alignment of Secondary Structure Elements. Given incorporation of *all* residues having amide NH by the MCD algorithm, the alignment of these within the primary sequence would be immediately evident. As might be expected, the MCD algorithm will generally fail to do this with proteins of significant size. In the case of human ubiquitin, 67 of 76 residues were aligned in helical, antiparallel, parallel, turn, and extended-chain secondary structure by the MCD algorithm. Thus some side-chain identification was required to complete the assignment of main chain proton resonances. However, it must be emphasized that the task being presented following the MCD analysis is clearly less demanding than the side-chain spin system definition required for the successful application of the sequential assignment. The extensive alignment of residues provided by the MCD algorithm allows one to pick and choose the amino acid side-chain J-coupled

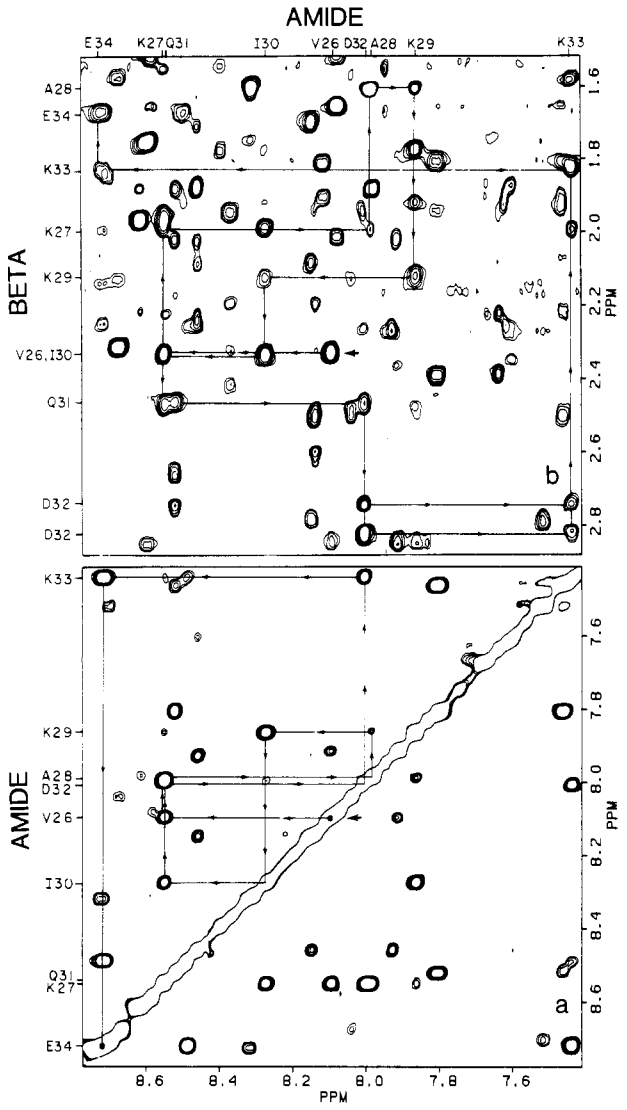

FIGURE 3: Expansion of a phase-sensitive NOESY spectrum of human ubiquitin. A mixing time of 140 ms was used. Other conditions as described in Figure 1. Shown are the basic NOE elements recognized and ordered by the MCD pattern of a helical unit of secondary structure. This particular segment corresponds to the α-helical segment incorporating residues 26–34. Panel a shows the d_{NN} NOE connectivities which run in parallel to the $d_{\beta N}$ NOEs shown in panel b. Many intraamino acid $d_{\beta N}$ NOEs are observed and, in conjunction with the previously determined J-coupled amide NH–C_αH–C_βH sets, lead to efficient recognition of helical secondary structures. Note that the combined use of $d_{\beta N}$ and d_{NN} NOEs, which compose the closed-loop helical MCD pattern, allows many apparent degeneracies to be resolved.

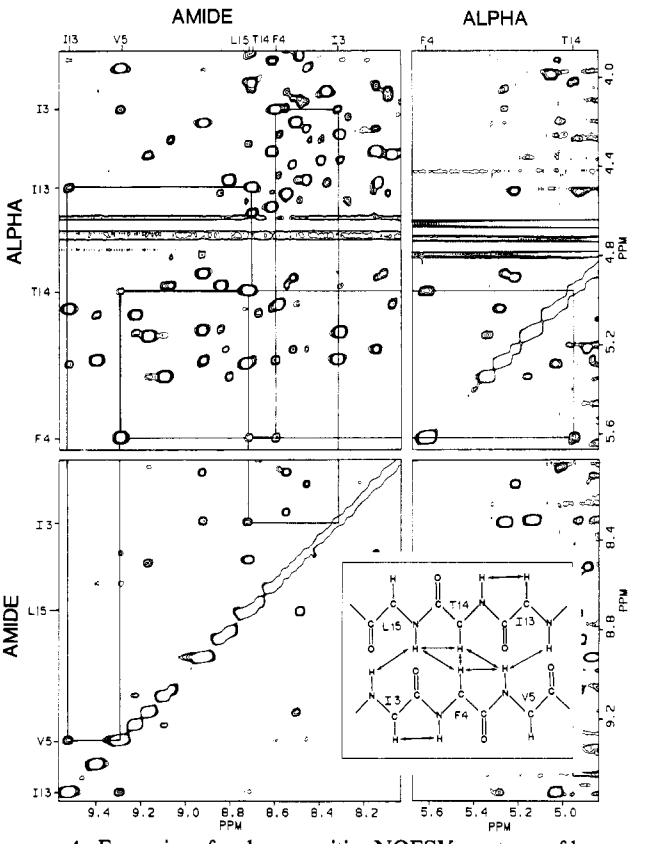

FIGURE 4: Expansion of a phase-sensitive NOESY spectrum of human ubiquitin. A mixing time of 140 ms was used. Other conditions are as described in Figure 1. Shown are the basic NOE elements recognized and ordered by the MCD pattern of an antiparallel sheet. This particular segment involves residues 3–5 on one strand and 13–15 in its antiparallel partner. The inset shows the protons participating in the NOE pattern.

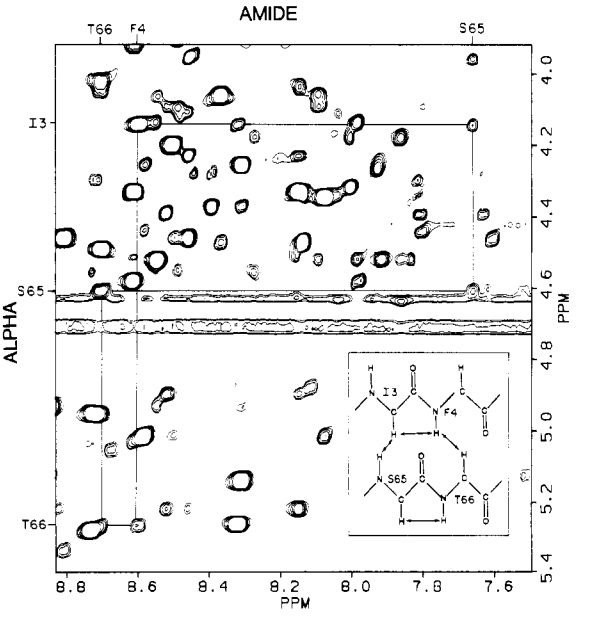

FIGURE 5: Expansion of a phase-sensitive NOESY spectrum of human ubiquitin. A mixing time of 140 ms was used. Other conditions are as described in Figure 1. Shown are the basic NOE elements recognized and ordered by the MCD pattern of a parallel sheet. This particular segment involves residues 3 and 4 on one strand and 65 and 66 in its parallel partner. The inset shows the protons participating in the NOE pattern.

networks to be defined in order to further align the MCD elements.

The advantage of this can be simply illustrated by choosing the easiest side-chain spin systems to identify. These might be considered to be those of glycine, alanine, and threonine. Glycine is easily identified by it unique ability to display *two* NH–C_αH cross-peaks. The two α-protons are in turn coupled. The distinction between glycine and, for example, the lysine side-chain NH–$C_\epsilon H_2$–$C_\delta H_2$ spin system is based on the involvement of the authentic main-chain subspin system in MCD patterns. Alanine and threonine are the simplest amino acids having a methyl group. Due to the magnetic equivalence of the three methyl protons, chemical shift correlation via J-coupling with a methyl group leads to particularly intense and efficient coherence transfer (Sorensen et al., 1983; Bax & Drobny, 1985) and hence to a large captured volume (Levitt et al., 1984). Taken together, these properties lead to a distinctively intense cross-peak as well as a characteristic splitting in the DQF COSY (Figure 6). For the most part, cross-peaks

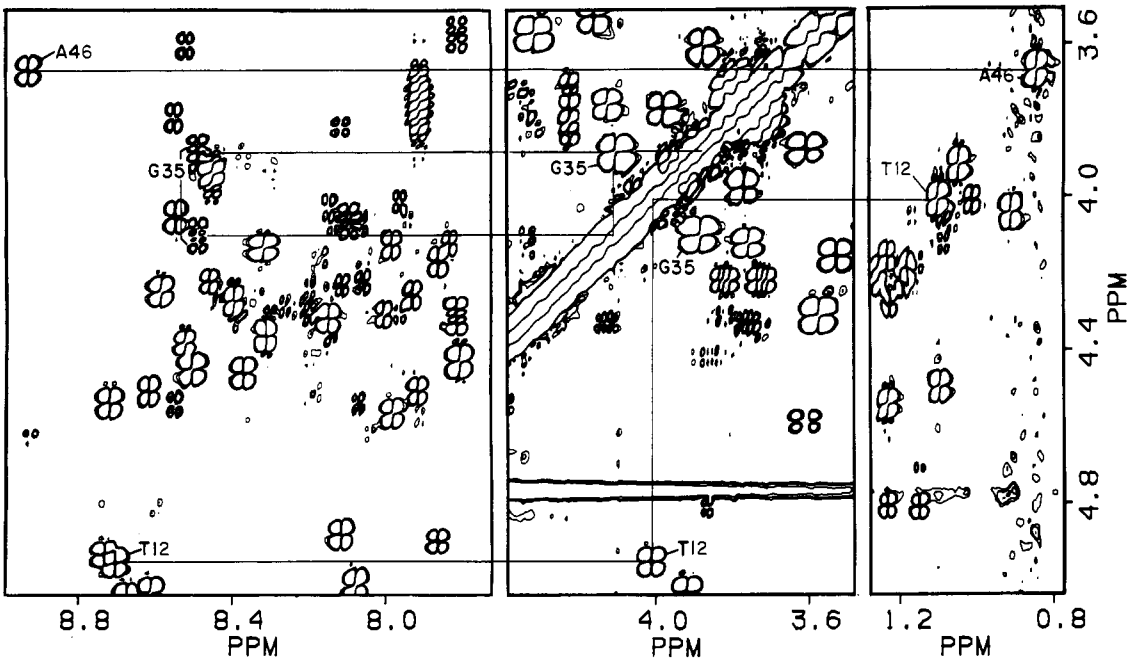

FIGURE 6: Sequence alignment of MCD-defined secondary structure elements by identification of *J*-coupled spin systems of simple amino acids. MCD elements are aligned within the primary sequence by reference to the placement of simple amino acid spin systems such as glycine, alanine, and threonine. Examples of *J*-coupled connectivities of each are shown.

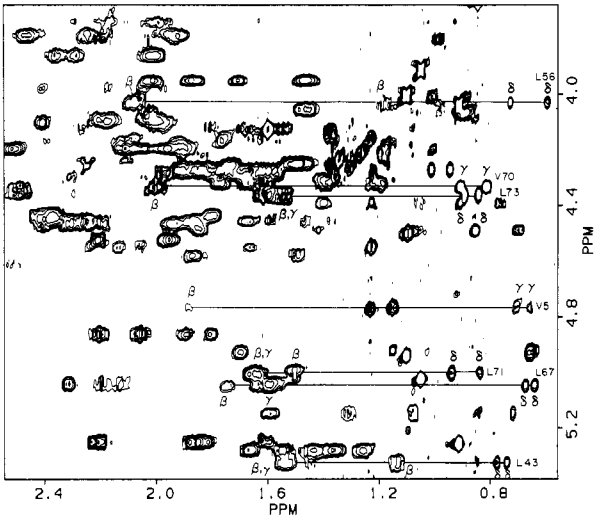

FIGURE 7: Identification of side-chain *J*-correlated spin systems following MCD analysis. Following MCD analysis and sequence alignment, the amino acid type of each NH–C_αH–C_βH subspin system is known, and the remaining portions of the side chain are easily recognized in DQF COSY, RCT COSY, and TOCSY spectra. Shown are the remote correlations of the C_αH proton with the rest of the side-chain protons of several valine and leucine residues as revealed by a TOCSY experiment. A 75-ms spin-locked isotropic mixing period was used.

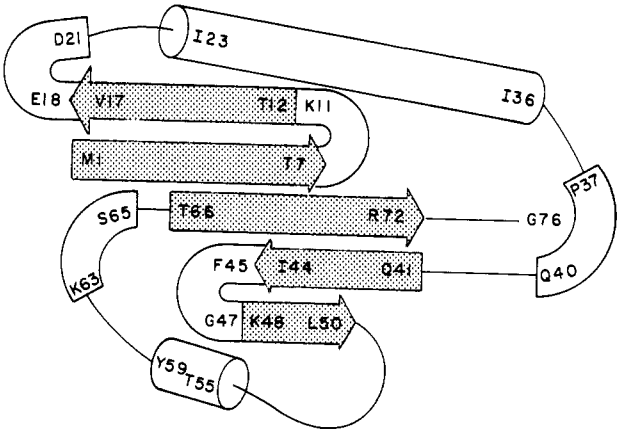

FIGURE 8: Schematic diagram of the secondary structure of human ubiquitin as defined by the MCD analysis. It is presented in a planar manner to emphasize the secondary structural information provided by the MCD analysis; i.e., no attempt is made to suggest the tertiary character of the molecule.

involving a methyl group can be identified by inspection, and the demonstration of alanine (e.g., by an NH to β-CH_3 relay) and threonine is generally efficient and reliable.

Consider the largest helical segment defined by the MCD analysis. This stretch contained a glycine. One of the two alanines defined (Figure 6) is found seven residues from the glycine. This immediately locates and confirms the orientation of the helical segment, previously indicated by the polarity of the connectivity conferred by the $d_{\beta N}$ NOE. The remaining alanine (residue 46) is found in a turn that connects two strands of the smallest antiparallel sheet. One of these strands is part of another MCD-defined antiparallel β-strand that terminates seven residues toward the N-terminus (see Figure 8). This residue is expected to be a proline and is consistent with the termination of the MCD pattern. Once again, the MCD pattern, in this case for an antiparallel β-sheet, determines the polarity of the two strands. Similar arguments led to the placement of all the MCD-defined units of secondary structure and only required the recognition of glycine, alanine, and threonine.

Side-Chain Definition following Sequence Alignment. The final step in the MCD analysis involved the definition of turns incorporating proline. In one case, a turn involving Pro-19, the C_δH protons mimic the NOE interactions of the analogous amide NH in a type I turn. The remaining turn, involving Pro-37 and Pro-38, displays NOEs consistent with a trans configuration (Wüthrich et al., 1984).

The alignment of the units of secondary structure defined by the MCD algorithm within the primary sequence also allows side-chain *J*-coupled spin systems to be efficiently discovered. If complete assignments are not required, e.g., for hydrogen-exchange studies of peptide group NH, then extensive side-chain identification is not necessary and only serves to raise the level of confidence in the MCD results obtained.

Table I: Chemical Shifts of Assigned Proton Resonances of Human Ubiquitin[a]

	chemical shift (ppm)					
amino acid	NH	$C_\alpha H$	$C_\beta H$	$C_\gamma H$	other	
Met-1		4.20				
Gln-2	8.92	5.27	1.87	2.24		
			1.60			
Ile-3	8.31	4.13	1.76	1.07, CH		0.56, δ-CH_3
				0.84, CH		
				0.63, CH_3		
Phe-4	8.59	5.61	3.02		7.21, ortho	
			2.86		7.05, meta	
					6.75, para	
Val-5	9.29	4.81	1.89	0.71		
				0.76		
Lys-6	8.93	5.28	1.69	1.29	1.42, δ-CH	2.91, ϵ-CH
			1.37	1.57		
Thr-7	8.72	4.93	4.80	1.24, CH_3		
Leu-8	9.06	4.27	1.73	1.85	1.02, ϵ-CH_3	
			1.91		0.95, ϵ-CH_3	
Thr-9	7.62	4.38	4.56	1.24, CH_3		
Gly-10	7.81	4.31				
		3.58				
Lys-11	7.26	4.32	1.78	1.36		2.91, ϵ-CH
			1.68	1.20		
Thr-12	8.60	5.04	3.92	1.06, CH_3		
Ile-13	9.52	4.48	1.85	1.08, CH	0.70, δ-CH_3	
				1.41, CH		
				0.85, CH_3		
Thr-14	8.69	4.96	4.01	1.11, CH_3		
Leu-15	8.71	4.74	1.34	1.43	0.79, δ-CH_3	
			1.21		0.74, δ-CH_3	
Glu-16	8.11	4.87	1.81	2.08		
			1.92	2.22		
Val-17	8.91	4.63	2.33	0.44, CH_3		
				0.70, CH_3		
Glu-18	8.67	5.05	1.58			
Pro-19		3.97	2.22	2.03	4.06, δ-CH	
			2.42	1.58	3.78, δ-CH	
Ser-20	7.02	4.33	4.14			
			3.78			
Asp-21	8.03	4.62	2.93			
			2.90			
Thr-22	7.86	4.90	4.78	1.16, CH_3		
Ile-23	8.51	3.61	2.46	0.76, CH_3	0.57, δ-CH_3	
				1.90, CH		
				1.30, CH		
Glu-24	10.01	3.87	2.01	2.36		
Asn-25	7.91	4.51	3.19			
			2.84			
Val-26	8.09	3.38	2.33	0.97, CH_3		
				0.68, CH_3		
Lys-27	8.54	4.54	1.99	1.82		
				1.58		
Ala-28	7.98	4.13	1.61, CH_3			
Lys-29	7.86	4.17	2.12	1.93, β or γ	1.76, δ-CH	2.99, ϵ-CH
				1.64	1.43, δ-CH	
Ile-30	8.27	3.47	2.33	1.97, CH	0.65, δ-CH_3	
				0.84, CH		
				0.67, CH_3		
Gln-31	8.54	3.79	2.47			
Asp-32	8.00	4.31	2.80			
			2.73			
Lys-33	7.43	4.28	1.82		2.01	3.14, ϵ-CH
					1.71	
					1.60	
Glu-34	8.71	4.54	1.66	2.16		
				2.08		
Gly-35	8.48	3.90				
		4.11				
Ile-36	6.15	4.39	1.40	1.37, CH	0.76, δ-CH_3	
				1.07, CH		
				0.93, CH_3		
Pro-37		4.60	2.40	2.08	4.17	
			1.95	2.04	3.54	
Pro-38		4.09	2.19	2.16	3.73	
			2.03	1.62	3.87	
Asp-39	8.51	4.38	2.75			
			2.66			

Table I (Continued)

amino acid	chemical shift (ppm)					
	NH	$C_\alpha H$	$C_\beta H$	$C_\gamma H$	other	
Glu-40	7.80	4.43	1.81	2.41		
			1.83			
Gln-41	7.46	4.19	1.88	2.08 (?)		
			1.96	2.52 (?)		
Arg-42	8.50	4.45	1.67	1.49	3.11, δ-CH	
				1.61		
Leu-43	8.80	5.34	1.54	1.45	0.75, δ-CH_3	
					0.78, δ-CH_3	
Ile-44	9.09	4.94	1.71	0.68, CH_3	0.64, δ-CH_3	
				1.32, CH		
				1.03, CH		
Phe-45	8.84	5.13	2.78		7.33, ortho	
			2.99		7.51, meta	
					7.44, para	
Ala-46	8.92	3.67	0.85, CH_3			
Gly-47	8.09	4.07				
		3.42				
Lys-48	7.97	4.57	1.88	1.49	1.82, δ-CH	3.13, ϵ-CH
Gln-49	8.61	4.51	1.97	2.50		
			2.22	2.04		
Leu-50	8.54	4.06	1.47	1.45	0.50, δ-CH_3	
			0.99		−0.17, δ-CH_3	
Gln-51	8.36	4.46	1.95			
Asp-52	8.14	4.35	2.49			
			2.61			
Gly-53	9.66	4.04				
Arg-54[b]						
Thr-55	8.82	5.22	4.51	1.11, CH_3		
Leu-56	8.14	4.03	2.08	1.70	0.73, δ-CH_3	
			1.19		0.60, δ-CH_3	
Ser-57	8.45	4.22	3.83			
			3.73			
Asp-58	7.92	4.26	2.95			
			2.26			
Tyr-59	7.25	4.60	2.51		7.27, ortho	
			3.43		6.87, meta	
Asn-60	8.14	4.33	3.28			
			2.77			
Ile-61	7.23	3.35	1.37	−0.37, CH	0.39, δ-CH_3	
				1.08, CH		
				0.48, CH_3		
Glu-62	7.60	4.45	1.88	2.29		
				2.35		
Lys-63	8.44	3.94	2.02	1.49	1.71, δ-CH	3.01, ϵ-CH
			1.87			
Glu-64	9.28	3.32	2.39	2.24		
			2.54			
Ser-65	7.65	4.59	3.61			
			3.89			
Thr-66	8.70	5.26	4.05	0.92, CH_3		
Leu-67	9.39	5.06	1.61	1.75	0.68, δ-CH_3	
			1.60		0.65, δ-CH_3	
His-68	9.22	5.14	2.89		8.01, C_4H	
					7.01, C_2H	
Leu-69	8.30	5.16	1.60	1.31	0.84, δ-CH_3	
					0.73, δ-CH_3	
Val-70	9.16	4.34	2.01	0.91, CH_3		
				0.82, CH_3		
Leu-71	8.07	5.02	1.65	1.64	0.96, CH_3	
			1.53		0.86, CH_3	
Arg-72	8.58	4.25	1.51	1.50	3.13, δ-CH	
			1.75			
Leu-73	8.30	4.36	1.54	1.63	0.92, δ-CH_3	
					0.87, δ-CH_3	
Arg-74	8.39	4.27	1.74	1.61	3.19, δ-CH	
			1.85			
Gly-75	8.44	3.92				
		3.97				
Gly-76	7.91	3.71				
		3.79				

[a] Chemical shifts are referenced to external sodium 3-(trimethylsilyl)tetradeuteriopropionate and are quoted to 0.01 ppm. For ubiquitin in 50 mM potassium phosphate buffer, pH* 5.8 at 30 °C. [b] Not assigned, due to α-proton saturation by decoupler.

If complete resonance assignments are required, e.g., for distance–geometry studies, then the MCD approach provides a great advantage over de novo side-chain assignment, since it restricts or even fully defines in advance each spin system type. This is especially useful since overlap and ambiguity in any protein ^{1}H spectrum is most severe in the aliphatic side chain region.

The restrictions imposed by the MCD analysis led to efficient and complete definition of most of the leucine, valine, and isoleucine side chains (Figure 7 and Table I). Amide NH–C_αH–C_βH sets that were incorporated into secondary structure with the MCD algorithm had known side-chain spin system type, which could then be extracted from the difficult aliphatic region of *J*-correlated spectra. The remaining NH–C_αH–C_βH sets that were not incorporated during the application of the MCD algorithm had highly restricted identities. Thus, sequential assignment of these few remaining residues (i.e., side-chain definition followed by sequential linkage) proceeded smoothly. The chemical shifts of the assigned resonances are summarized in Table I.[3]

Discussion

Summary. The assignment of the majority of ^{1}H resonances of human ubiquitin has been presented in some detail. The assignment was initiated by the application of a main chain directed algorithm introduced by Englander and Wand (1987). The MCD method successfully identified the general character of the secondary structure elements of the protein with no foreknowledge of the amino acid side chains. These results then led to the rapid assignment of the amino acid side chain *J*-coupled networks by providing the identity of most side chains and thereby restricting apparent *J*-coupling pathways to be considered. The MCD method is summarized and compared to the sequential assignment method below. The observed secondary structure elements and their alignment are also discussed and compared to the published crystal structure.

Methodology. The MCD method (Englander & Wand, 1987) and the sequential assignment method (Wüthrich et al., 1982; Wüthrich, 1983) both use the very same data and seek the very same goal, but they represent distinct approaches to the assignment problem. As defined (Wüthrich, 1983), the sequential assignment method involves "...(i) identification of amino acid side-chain spin systems, (ii) identification of neighboring residues in the amino acid sequence, and (iii) suitable combination of the results from (i) and (ii) for obtaining individual resonance assignments in the primary structure of the protein". The MCD method, while relying on the same structural observations (Billeter et al., 1982; Wüthrich et al., 1984; Englander & Wand, 1987), seeks to define units of secondary structure that need not be sequential with the primary sequence nor require any reference to the identity of the involved amino acids. The de novo definition of side chains, through the application of *J*-correlated techniques, is the most difficult (and first) step of the sequential assignment procedure. The aim of the MCD method is to avoid, at the outset of the assignment, the need to exhaustively define the identities of each amino acid residue. The MCD algorithm provides a formal procedure that links amino acid residues together by a logical and straightforward dissection of overlapping NOE patterns. The sequential assignment method *discovers* the type of secondary structure *after* sequentially linking previously recognized amino acid spin systems together, via the NOE. In contrast, the MCD method *proposes* a given kind of secondary structure and seeks the NOE interactions required to satisfy the companion pattern involving main-chain protons. Furthermore, when the various MCD patterns are sought in the appropriate order (Englander & Wand, 1987), many ambiguities due to accidental chemical shift degeneracy or to the appearance of NOEs involving the same type of protons in different types of secondary structure are successfully and efficiently accommodated. This is due to the closed-loop, internally confirmatory nature of the MCD patterns. Given a successful MCD analysis, the otherwise most difficult definition of side-chain *J*-coupled networks becomes a relatively straightforward task, as illustrated here with ubiquitin. One anticipates that the relative advantage of the MCD approach will become greater as the size of the protein being studied increases.

While the power and sophistication of *J*-correlated techniques has dramatically increased in the very recent past (e.g., TOCSY), it has been suggested that dipolar contributions for large (i.e., greater than 20 kDa) proteins will lead to unmanageable line widths and relaxation behavior (LeMaster & Richards, 1987). Clearly genetic approaches to the assignment are available but can be anticipated to be both inefficient and expensive. A more comprehensive approach could be the random deuteriation method introduced by LeMaster and Richards (1987) to reduce the dipolar contribution to observed resonance line widths and simplify cross-peak fine structure. Nevertheless, this would effectively prevent the application of long-range *J* correlation via, for example, an isotropic mixing experiment. However, such a system would be completely amenable to analysis by the MCD method.

The MCD approach is not without limitations. It does require the initial definition of the amino acid residue amide NH–C_αH–C_βH *J*-coupled subspin systems. However, this is clearly an easier task than the resolution and assignment of the side-chain *J*-coupled networks. Thus, if the NH–C_αH–C_βH *J*-coupled subspin systems cannot be successfully defined, it is unlikely that any comprehensive application of two-dimensional ^{1}H NMR techniques would be successful. As with the sequential assignment method, the application of MCD to the analysis of NOESY spectra requires some consideration of the mixing time used to generate NOEs. Spin–diffusion (Kumar et al., 1981) should be avoided although limited spin–diffusion, leading to additional secondary cross-peaks while retaining primary NOEs, would be accommodated by the MCD analysis. This is because the MCD analysis is restricted to selected combinations of NOE interactions. All other NOEs arising from the involved protons are initially ignored. Extensive spin–diffusion, leading to loss of primary NOEs, would not, in general, be handled by either the MCD or sequential assignment approaches. Quantitative consideration of observed NOEs is not required. It was claimed before (Englander & Wand, 1987) and has been observed here with ubiquitin that, if the closed-loop MCD patterns are satisfied, the linkage of the involved amino acid residues within the considered unit of secondary structure is virtually guaranteed. This, however, does lead to a potential limitation of the present description of the MCD method. That is, distortion of secondary structure elements (e.g., β-bulges in β-sheets) may result in the inability of the observed NOEs to meet the stringent MCD patterns described here and thus limit the identification of secondary structure elements. This will have to be accommodated by development of further self-confir-

[3] Following application of the assignment procedure described here, several weak unassigned *J*-correlated cross-peaks remained. Subsequent analysis, by HPLC, indicated the presence of five minor species each representing 2% of total protein. The presence of these minor forms did not inhibit the analysis described.

matory closed-loop MCD patterns for the various distortions of standard elements of secondary structure that may be observed.

The MCD patterns themselves require discrete information in the sense that only the presence or absence of a given NOE is considered. Therefore, the method is highly adaptable to automation. The adaptation of the algorithm to computer-assisted analysis is currently under development and can be expected to allow the efficient application of many more MCD patterns designed to accommodate structural irregularities in units of regular secondary structure. We shall report on this elsewhere.

Structure. The MCD analysis of the NOESY spectrum of human ubiquitin suggests the secondary structure schematically illustrated in Figure 8. The NMR analysis shows that ubiquitin in solution has five strands organized in an extensive β-sheet structure. The MCD analysis suggests a hydrogen-bonding scheme that is consistent with the more recently reported crystal structure (Vijay-Kumar et al., 1987). The C-terminus, which is the locus for covalent linkage of ubiquitin to other proteins (Goldknopf & Busch, 1977; Ciechanover et al, 1980a; Hershko et al., 1980), appears to be rotating freely in solution.

ACKNOWLEDGMENTS

We are most grateful to Drs. P. Weber, S. Brown, and L. Mueller for critical discussion, for encouragement, and for providing a preprint of their independent assignment of ubiquitin. We thank Dr. H. Roder for help with implementation of the TOCSY experiment on the AM 500. We also thank R. Dykstra for continued support of the AM 500 and critical development and installation of the electromagnetic shield used for water suppression.

Registry No. Ubiquitin, 60267-61-0.

REFERENCES

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Bax, A., & Drobny, G. (1985) *J. Magn. Reson. 61*, 306–320.
Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
Ciechanover, A., Hod, Y., & Hershko, A. (1978) *Biochem. Biophys. Res. Commun. 81*, 1100–1105.
Ciechanover, A., Heller, H., Elias, S., Haas, A. L., & Hershko, A. (1980a) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1365–1368.
Ciechanover, A., Elias, S., Heller, H., Ferber, S., & Hershko, A. (1980b) *J. Biol. Chem. 255*, 7525–7528.
Dykstra, R. (1987) *J. Magn. Reson. 72*, 162–167.
Eich, G., Bodenhausen, G., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 3731–3732.
Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953–5958.
Goldknopf, I. L., & Busch, H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 864–868.
Hershko, A., Ciechanover, A., & Rose, I. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3107–3110.
Herhsko, A., Ciechanover, A., Heller, H., Haas, A. L., & Rose, I. A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1783–1786.
Kumar, A., Wagner, G., Ernst, R. R., & Wüthrich, K. (1981) *J. Am. Chem. Soc. 103*, 3654–3658.
LeMaster, D. M., & Richards, F. M. (1987) *Biophys. J. 51*, 235.
Levitt, M. H., Bodenhausen, G., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 462–472.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Rance, M., Sorensen, O. W., Bodenhausen, O., Wagner, G., Ernst, R. R., & Wüthrich, K. (1984) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Redfield, A. G., & Kuntz, S. D. (1975) *J. Magn. Reson. 19*, 250–254.
Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169–173.
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., & Cook, W. J. (1987) *J. Mol. Biol. 194*, 531–544.
Wagner, G. (1984) *J. Magn. Reson. 55*, 151–156.
Wand, A. J., & Englander, S. W. (1985) *Biochemistry 24*, 5290–5294.
Wand, A. J., & Englander, S. W. (1986) *Biochemistry 25*, 1100–1106.
Wider, G., Macura, S., Kumar, A., Ernst, R. R., & Wüthrich, K. (1984) *J. Magn. Reson. 56*, 207–234.
Wüthrich, K. (1983) *Biopolymers 22*, 131–138.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Chapter 5, Wiley, New York.
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.