Biochemistry

 Volume 26, Number 19 September 22, 1987

Accelerated Publications

Main-Chain-Directed Strategy for the Assignment of ^{1}H NMR Spectra of Proteins†

S. Walter Englander
Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104
A. Joshua Wand*
Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111
Received June 9, 1987; Revised Manuscript Received July 21, 1987

ABSTRACT: A strategy for assigning the resonances in two-dimensional (2D) NMR spectra of proteins is described. The method emphasizes the analysis of through-space relationships between protons by use of the two-dimensional nuclear Overhauser effect (NOE) experiment. NOE patterns used in the algorithm were derived from a statistical analysis of the combinations of short proton–proton distances observed in the high-resolution crystal structures of 21 proteins. One starts with a search for authentic main-chain NH–C_αH–C_βH J-coupled units, which can be found with high reliability. The many main-chain units of a protein are then placed in their proper juxtaposition by recognition of predefined NOE connectivity patterns. To discover these connectivities, the 2D NOE spectrum is examined, in a prescribed order, for the distinct NOE patterns characteristic of helices, sheets, turns, and extended chain. Finally, the recognition of a few amino acid side-chain types places the discovered secondary structure elements within the polypeptide sequence. Unlike the sequential assignment approach, the main-chain-directed strategy does not rely on the difficult task of recognizing many side-chain spin systems in J-correlated spectra, the assignment process is not in general sequential with the polypeptide chain, and the prescribed connectivity patterns are cyclic rather than linear. The latter characteristic avoids ambiguous branch points in the analysis and imposes an internally confirmatory property on each forward step.

The development of two-dimensional NMR[1] techniques now makes it possible to resolve essentially all the proton resonances in small to moderate sized proteins and thus unlocks, in principle, the inherent analytic power of NMR methods for structural and functional studies of proteins in solution at a proton resolved level. Before these methods can be applied to any given protein, however, it is first necessary to assign the various resonances to their parent protons. The method of sequential assignment, described by Wüthrich and co-workers (Billeter et al., 1982; Wüthrich et al., 1982, 1984; Wüthrich, 1983), has been used to assign the proton resonances of a number of proteins ranging in size up to about 80 residues. This strategy first attempts to identify the complete spin systems of most amino acid residues by their proton–proton J couplings, which are revealed in two-dimensional ^{1}H NMR spectra that invoke the various properties of the through-bond interaction. The identified amino acid spin systems are then placed within the polypeptide sequence by use of through-space nuclear Overhauser effect connectivities.

The assignment problem can be difficult and becomes rapidly more severe as the size of the protein increases. Here

† This work was supported by NIH Research Grants AM 31847 (S.W.E.) and GM 35940 (A.J.W.), by NIH Grants CA 06927 and RR 05539, by a grant from the Pew Memorial Trust, by an award from Marie Z. Cole Montrose, by an appropriation from the Commonwealth of Pennsylvania awarded to the Institute for Cancer Research, and by Instrumentation Grants RR-02497 (NIH) and DMB 84-13986 (NSF).
* Address correspondence to this author at the Institute for Cancer Research, Fox Chase Cancer Center.

[1] Abbreviations: COSY, J-correlated spectroscopy; DQF, double quantum filter; MCD, main chain directed; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE-correlated spectroscopy; RCT, relayed coherence transfer; TOCSY, total correlation spectroscopy; 2D, two dimensional.

0006-2960/87/0426-5953$01.50/0

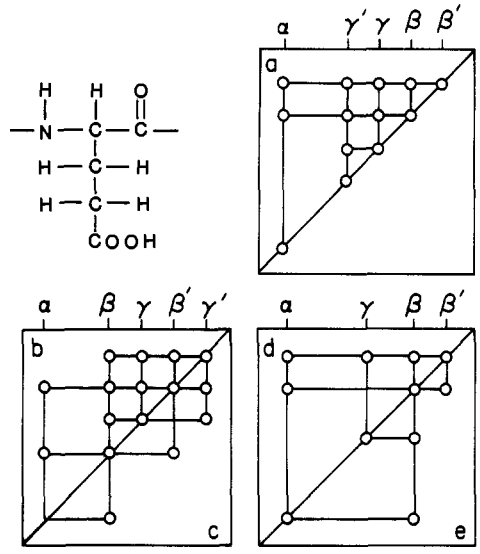

FIGURE 1: Schematic diagrams of COSY cross peak patterns observed for glutamic acid residues in small peptides and several proteins. Only one side of each symmetric spectrum is shown. Panel a represents the ideal cross peak pattern expected for glutamic acid. Panels b–e represent patterns that have been seen in several proteins. Chemical shift differences are caused by differences in local environment. The absence of cross peaks is potentially due to the effects of insufficient *J*-coupling, line width (transverse relaxation), and local side-chain dynamics. Similar variability for other types of amino acid side-chain *J*-coupled spin systems is expected and observed in structured proteins.

we describe a different approach that avoids some of the major difficulties of the sequential assignment method and prescribes a methodical series of steps that seem able to produce a more expeditious solution.

Assignment Strategies

Sequential Assignment Strategy. The sequential assignment procedure involves three ordered steps. These are (Wüthrich, 1983) "... (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 first step one attempts to recognize resonances that belong to the same amino acid side chain, categorize each spin system as to amino acid type, and connect each to its main-chain protons. For the smaller proteins, side-chain recognition is the rate-limiting step.

J-correlated (COSY-type) spectra can in principle display unique spin–spin coupling patterns for 9 of the 20 amino acids and can separate the remaining amino acids into two classes, one class of 9 amino acids for which connectivities are limited to the NH–C_αH–$C_\beta H_2$ fragment (including the aromatics) and a second class composed of Glu and Gln. Ideally, some of the unique amino acid spin systems may be identified by matching them with the chemical shift patterns found in free amino acids (McDonald & Philips, 1969) and small peptides (Bundi & Wüthrich, 1979). Relayed coherence transfer (RCT COSY) and total correlation (TOCSY) experiments can be used to help resolve ambiguities often found in COSY spectra of even small proteins. In our experience with cytochrome *c* (104 residues), we found this procedure to be impractical. The essence of the problem is illustrated in Figures 1 and 2.

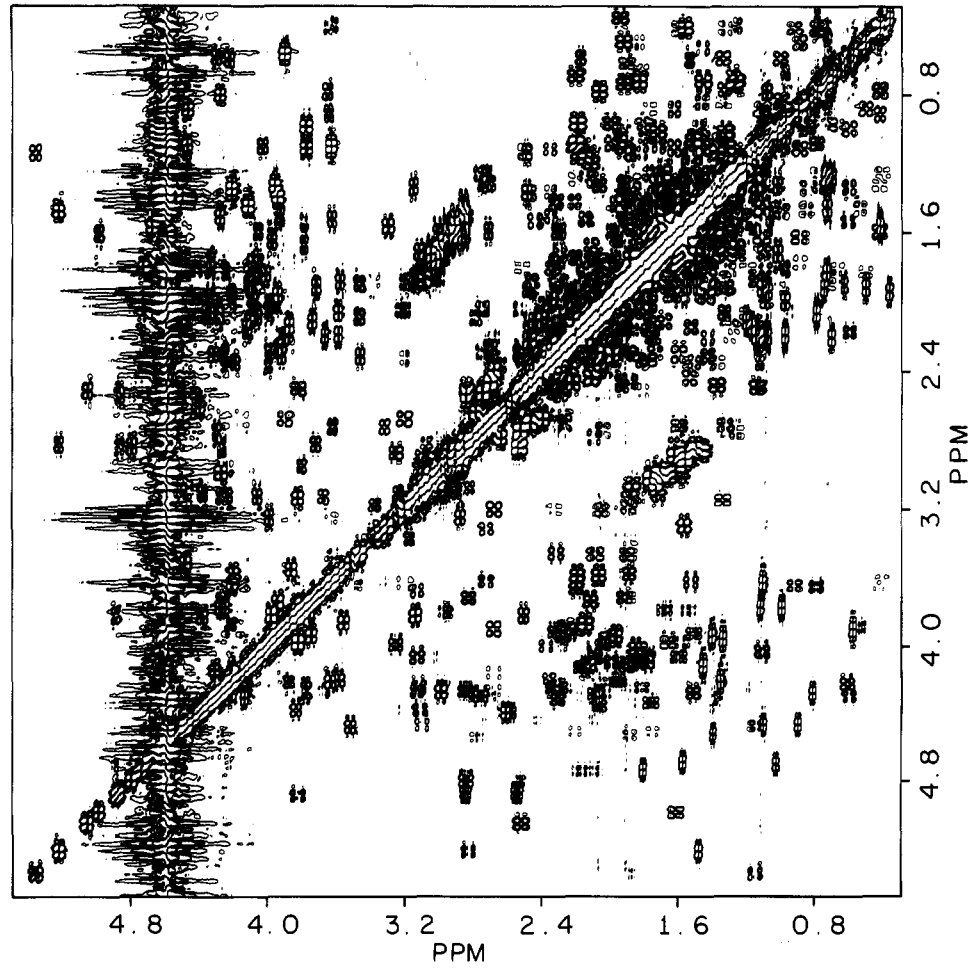

FIGURE 2: Expansion of a DQF COSY spectrum of horse ferrocytochrome *c* obtained at 500 MHz, presented in the phase-sensitive mode. The high cross peak density and the large number of amino acid side-chain *J*-coupling patterns of variable character produce considerable spectral complexity.

Figure 1a shows the ideal COSY pattern for glutamic acid in a structureless oligopeptide. This can be contrasted with Figure 1b–e, which show patterns actually found in several proteins. In a structured protein the chemical shift pattern may be distorted by the local protein environment in an unpredictable way. In addition, cross peak intensity is variable due to a number of factors. For example, most types of *J*-correlated experiments result in antiphase cross peak fine structure. This makes the intensity of cross peaks exceedingly sensitive to the inherent resonance line width (transverse relaxation time), which varies with side-chain rigidity, and the magnitude of ^3J coupling constants, which depend on the dihedral angle between the interacting protons. Since different residues experience different geometries and dynamics, one cannot expect idealized *J*-coupled connectivity patterns to be consistently observed.

Figure 2 is a segment of a COSY spectrum of horse ferrocytochrome *c*. In a sequential assignment effort, one is challenged to distinguish within this dense region the cross peak patterns that represent most of the individual amino acid side-chain spin systems. The high degree of chemical shift degeneracy found for a protein the size of cytochrome *c* (Figure 2) together with the plasticity of the patterns sought (Figure 1) effectively blocked the comprehensive sequential assignment of this protein at the outset.

Main-Chain-Directed Assignment Strategy. Three main criteria were used in designing an alternative assignment strategy: (i) the method must (initially) have only a limited dependence on the analysis of *J*-correlated spectra; (ii) attempts at pattern recognition should involve patterns that will appear with a high degree of frequency and integrity; (ii) the patterns should be resistant to disruption due to chemical shift degeneracy.

In the main-chain-directed (MCD) approach, one puts aside the first step of sequential assignment and adopts a supplementary strategy that emphasizes main-chain protons (NH, $C_\alpha H$, $C_\beta H$) rather than side-chain subspin systems. One uses of course the very same 2D ^1H NMR spectra as before and takes advantage of whatever side-chain information may be available. However, the focus is primarily on the main-chain region of the 2D spectrum where there is a much lower degree of ambiguity than in the side-chain region This, in conjunction with the simple and linear nature of the NH–C_αH–C_βH subspin system, makes the definition of main-chain units in *J*-correlated spectra comparatively straightforward. For example, 60% of the NH–C_αH–C_βH sets of human ubiquitin (76 amino acids) could be unambiguously identified on the basis of a single DQF COSY obtained in H_2O (Di Stefano & Wand, 1987). A single RCT COSY spectrum, used to resolve ambiguities due to chemical shift degeneracies among the amide NH and alpha CH resonances, then allowed 90% of the NH-C_αH-C_βH sets to be unequivocally defined. The remaining main-chain sets had defined amide NH and α-CH resonances but more than one possible β-proton resonance, a situation that does not hamper the analysis described here. Similar success was achieved with the analysis of DQF COSY and RCT COSY spectra of horse ferrocytochrome *c* (104 residues) (unpublished results).

The second and third criteria, aimed at ensuring a robust pattern recognition algorithm, are met by taking advantage of the structural constraints inherent in the secondary structure elements of proteins. The regular elements of protein secondary structure, composed of the main-chain NH–C_αH–C_βH units, reliably produce certain unambiguous NOE patterns in the least dense region of the 2D NOESY spectrum. Considerations discussed below dictate the choice of the patterns to be used in the assignment algorithm and the order in which they should be searched for.

Table I: Calculated Fidelity of Various Combinations of NOEs Applied to Proteins[a]

MCD pattern	fidelity (%) at cutoff distance (Å)						
	3.0	3.2	3.4	3.6	3.8	4.0	4.2
helix							
single	94	91	90	83	83	73	59
double	>99	>99	>99	98	97	93	87
triple	100	100	>99	99	97	95	89
antiparallel							
inner	NO[b]	100	93	98	95	89	87
full	NO	100	100	100	100	100	99
hybrid	100	93	96	95	94	89	81
reduced hybrid	100	>99	98	94	90	81	71
parallel							
single	100	97	86	31	15	15	14
double	100	100	100	95	85	75	65

[a] The fidelity parameter indicates the percent of MCD patterns found within each NOE cutoff distance that actually represent residues in the appropriate sequential and spatial orientation. [b] NO, none observed.

Structural Elements and Their MCD Connectivity Patterns

Many of the NOE interactions expected for elements of the standard protein secondary structures have been detailed by Wüthrich and co-workers (Billeter et al., 1982; Wüthrich, 1983; Wüthrich et al., 1984). Those studies concentrated on the short distances between protons of sequential residues and provide the foundation for the sequential assignment method. Here we augment these analyses with a study of close distances among the amide NH, C_αH, and C_βH protons, with the goal of discovering NOE combinations that satisfy criteria ii and iii above.

Analysis of Known Proteins. The analysis was carried out on 21 proteins whose crystal structures have been determined to 2 Å or less. The structures used were actinidine, alcohol dehydrogenase, azurin, bovine pancreatic trypsin inhibitor, carbonic anhydrase, carboxypeptidase A, concanavalin A, ferricytochrome b_5, ferricytochrome *c*, deoxyhemoglobin, hen egg white lysozyme, high-potential iron protein, lactate dehydrogenase, parvalbumin, plastocyanin, prealbumin, human lysozyme, carboxymyoglobin, melittin, ovomucoid, and T_4 lysozyme. Hydrogen atoms were added to the crystal structures obtained from the Brookhaven Protein Data Bank using standard bond lengths and geometry with a revision of a program kindly provided by Dr. P. E. Wright, Scripps Institute. A computer program was used to extract the distances defined in the text.

Qualitative examination of the commonly observed types of protein secondary structure (helix, sheet, turn) reveals a number of cyclic patterns formed by the *J*-coupled NH–C_αH–C_βH main-chain units and their interproton NOEs (Figure 3). The analysis of known protein structures was designed to test the usefulness of these patterns by assessing their frequency and fidelity. The results of this analysis are summarized in Tables I and II and are discussed below.

Table I lists the fidelity found for each pattern as a function of the spatial reach of the NOE observed. The fidelity parameter gives the percent of all the MCD patterns found that are correct [100(correct hits/correct + incorrect hits)]. As the NOE distance increases, fidelity decreases, since more inappropriate residue orientations manage to produce the MCD pattern in question. Table II indicates the frequency for finding each pattern of proton connectivities as a function

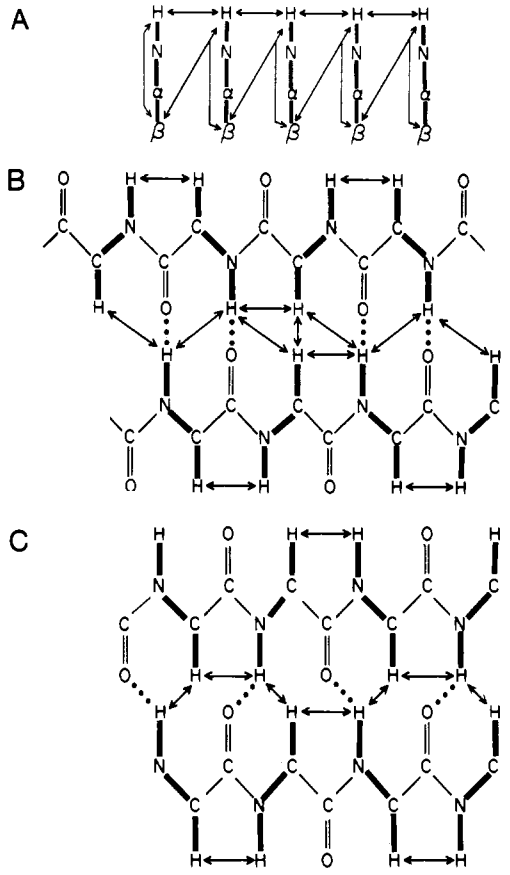

FIGURE 3: Standard secondary structures and their MCD patterns: (A) helix; (B) antiparallel sheet; (C) parallel sheet. Protons are joined by through-bond J-coupled connectivities (heavy lines) and short-range through-space NOEs (arrows), which form a variety of cyclic paths that can provide confirmatory assignment information in 2D NMR spectra.

Table II: Calculated Frequency of Various Combinations of NOEs Applied to Proteins[a]

MCD pattern	frequency (%) at cutoff distance (Å)						
	3.0	3.2	3.4	3.6	3.8	4.0	4.2 (total)
helix							
single	77	80	86	89	92	94	100 (1390)
double	41	57	68	76	83	90	100 (1093)
triple	36	55	65	74	82	90	100 (842)
antiparallel							
inner	0	2	7	22	46	73	100 (175)
full	0	1	4	18	42	73	100 (229)
hybrid	1	4	17	35	60	86	100 (316)
reduced hybrid	33	44	57	69	85	88	100 (426)
parallel							
single	29	52	72	85	92	93	100 (177)
double	32	59	81	91	91	95	100 (78)

[a] The frequency parameter indicates the percent of all the true positives present in the proteins that will be found within any given NOE cutoff distance. The number of correct patterns found when the NOE distance is 4.2 Å is taken to define the 100% value.

of increasing reach of the NOE. Frequency is defined as the fraction of correct patterns that will be found at a given NOE distance, expressed as a percentage of the total number of correct patterns found by using an NOE cutoff distance of 4.2 Å [100(correct hits/total correct patterns present)]. The 4.2-Å distance is considered here to be at the limit of reliable detection by the NOE. At this distance, all units of crystallographically defined secondary structure in the 21 proteins examined display their appropriate MCD pattern at the largest interproton distance examined. The term "correct" here includes not only residues identified by crystallographers as being in the appropriate secondary structural elements but also residues in any analogous chain conformation that places them in the proper sequential and spatial relationship. There is some difficulty in identifying the boundaries of secondary structure on the basis of descriptions given in the data sets deposited in the Protein Data Bank [cf. Wüthrich et al. (1984)]. Thus for assessment of the frequency and fidelity of any MCD pattern at a given NOE distance, "helix" and "sheet" structures are defined here as those residues in helix-like and sheet-like orientations that correctly produce the pertinent MCD pattern when evaluated at a 4.2-Å cutoff NOE distance.

Helix. The commonly occurring α and 3_{10} helices produce short distances between sequential amide NH (Billeter et al., 1982). Short NH–NH distances (d_{NN}), leading to appreciable NOEs in NOESY spectra, are not limited to helix but also occur between stands in the antiparallel sheet and type I reverse turns. In helices, each amide NH also has characteristically strong NOEs to its own $C_\beta H$ (Wand & Englander, 1986) and the $C_\beta H$ of its N-terminal neighbor ($d_{\beta N}$) (Billeter et al., 1982). Analysis of the degree of correlation of the distance between amide and β protons in the 21 proteins studied reveals that, if the interresidue distances NH_i–NH_{i+1} and $C_\beta H_i$–NH_{i+1} meet a certain distance cutoff, then in virtually all cases the intraresidue distance $C_\beta H_i$–NH_i also meets the distance cutoff. Furthermore, when two β-methylene protons are present, the one that experiences the larger J coupling to its α proton, the one most likely to be observed in J-correlated spectra, has been found in all cases to be the one closest to its own amide NH and that of the neighboring $(i + 1)$ residue. This calculation was based on the relationship ${}^3J_{\alpha\beta} = 1.6 - 1.3 \cos \theta_{\alpha\beta} + 9.5 \cos^2 \theta_{\alpha\beta}$ (Jardetsky & Roberts, 1981).

These observations recommend the first MCD pattern, which involves a closed loop formed by interresidue d_{NN} and $d_{\beta N}$ NOEs and an intraresidue $d_{\beta N}$ NOE (Figure 3). We name this combination of connectivities the helical MCD pattern because of its usual, although not unique, association with helical secondary structure. As for the other MCD patterns described below, the J-coupled NH–C_αH–C_βH pathway can be used instead of the intraresidue NOE.

A *crucial* feature of this and other patterns presented here is the cyclic path of the connectivities. This provides a means for dealing with apparent branches in connectivity strings (criterion iii above). Given a number of possible helical neighbors suggested by d_{NN} NOE connectivities, one can check for matching $d_{\beta N}$ NOEs within and between the established NH–C_αH–C_βH sets. This feature selects authentic MCD-defined helical connectivities with high reliability (Table I), rejects incorrect ones, and provides information on chain folding.

As illustrated schematically in Figure 3 and for a real case in Figure 4, the residues in an MCD-defined helix display a series of interlocking helical MCD patterns. Three sequential residues produce the interlocking double helix pattern, and four residues produce the triple pattern. Especially for the interlocking patterns, the fidelity (Table I) of the helical MCD pattern is maintained over the range of distances relevant to the NOE, and these are expected to occur with high frequency (Table II). The MCD-defined helices were often found, at cutoff distances less than 4.2 Å, to be one or two residues shorter than helices defined by the crystallographers. This is due to deviations from standard helical geometry at helix termini.

While helices display a series of interlocking helical MCD patterns, such as that illustrated in Figure 4, type I turns

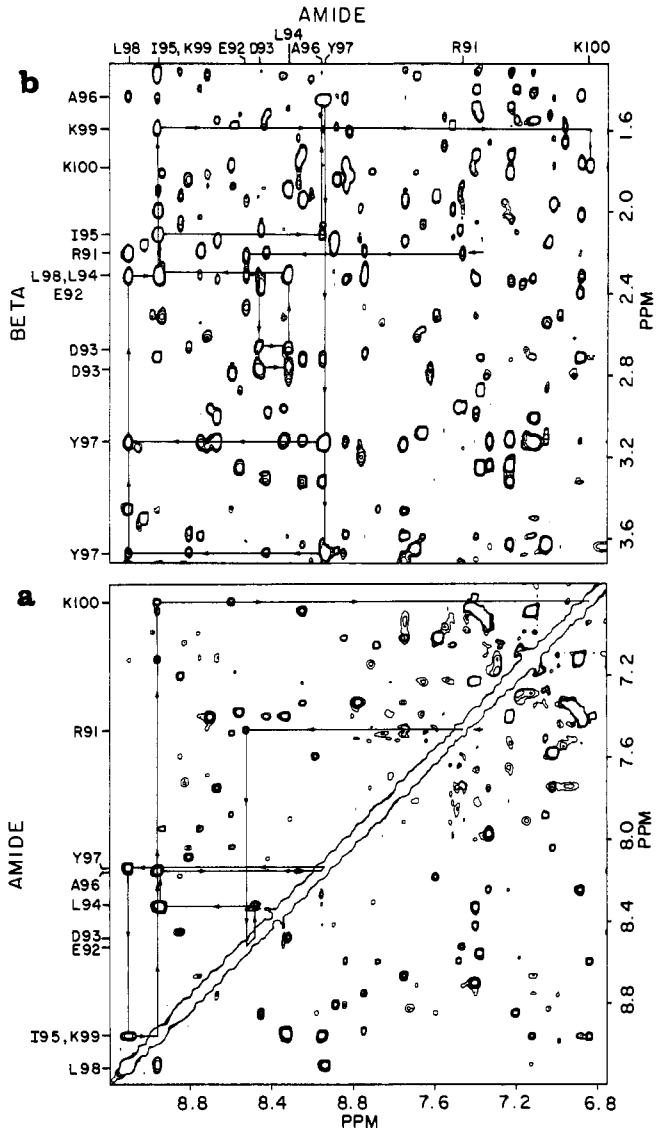

FIGURE 4: Elements of the interlocking helical MCD patterns observed for the C-terminus of horse ferrocytochrome *c*. Panel a shows the amide NH–amide NH NOEs between neighboring residues. Panel b contains the intra- and interresidue β-CH–amide NH NOE cross peaks. The combined use of these data to construct the helical MCD patterns leads to the resolution of many apparent ambiguities. For example, the d_{NN} NOE between Tyr-97 and Ala-96 is lost due to their near degeneracy (panel a) but is overcome by the appearance of inter- and intraresidue $d_{\beta N}$ NOEs (panel b). Situations where all three NOEs of the helical MCD pattern are ambiguous due to degeneracy are also resolved.

display helical MCD patterns involving just three residues.

Antiparallel β-Sheet. The standard antiparallel β-sheet (Figure 3) suggests several closed loop patterns. Four patterns were qantitatively examined. The first is composed of the interstrand NOE interactions arising from the proton involved in the short $C_\alpha H_i$–$C_\alpha H_j$ distance and includes NOEs to the $i + 1$ and $j + 1$ amide NHs (trapezoid shape in Figure 3). This set of interactions, involving four protons and five NOEs, is named the "inner loop" of the antiparallel sheet MCD pattern. An "outer loop" is defined, for example, as the closed loop of $(C_\alpha H\text{–}NH)_i$–$(C_\alpha H\text{–}NH)_{i-1}$–$NH_{j+1}$–$(C_\alpha H\text{–}NH)_i$ NOEs (pentagon shape in Figure 3). Note that this particular outer loop is related to a symmetric counterpart by a C_2 rotation axis. A fusion of the appropriate outer loop with the inner loop, leading to a common interstrand $C_\alpha H_i$ to NH_{j+1} NOE is defined as the "full loop". It too has a symmetric (C_2 rotation) counterpart. Finally, a "hybrid" pattern is defined as the closed loop, $C_\alpha H_i$–NH_{i+1}–$(NH\text{–}C_\alpha H)_{j-1}$–$(NH\text{–}C_\alpha H)_j$–$C_\alpha H_i$, plus the $C_\alpha H_j$–NH_{i+1} NOE. The connectivities enclosed in parentheses, e.g., $(NH\text{–}C_\alpha H)_{j-1}$, are due to J coupling. A "reduced hybrid" pattern is the same loop without the latter NOE.

Although the inner loop and, in particular, the full loop antiparallel MCD patterns have very high fidelity (i.e., when observed give the correct residue orientation), they display low frequency (Tables I and II). This is due to the fact that the interstrand $C_\alpha H$–NH distance is relatively long and is most sensitive to deviations of the two strands away from standard antiparallel β-sheet structure. The hybrid patterns, while still maintaining a high degree of fidelity, provide a much higher frequency index. As for the helical MCD pattern, the major reason for unidentified antiparallel MCD patterns is structural distortion at the ends of crystallographically defined antiparallel strands.

Examination of the NOEs indicated by Wüthrich and coworkers (Wüthrich et al., 1984) for type II turns would suggest a loop composed of $(C_\alpha H\text{–}NH)_i$–NH_{i+1}–$(NH)_{i+2}$–$(C_\alpha H\text{–}NH)_i$ NOEs.

Parallel Sheet. Closed loop patterns involving inter- and intrastrand NOEs expected for standard parallel β-sheet structure can be constructed (Figure 3). The apparent single loop MCD pattern is composed of four $C_\alpha H$–NH NOEs and two $C_\alpha H$–NH scalar couplings [i.e., $(NH\text{–}C_\alpha H)_i$–$(NH\text{–}C_\alpha H)_{i+1}$–$NH_{j+1}$–$C_\alpha H_j$–$(NH\text{–}C_\alpha H)_i$]. The interresidue $C_\alpha H$–NH NOE occurs also in other types of secondary structure (Billeter et al., 1982, Wüthrich et al., 1984), and as indicated in Table I, the single parallel sheet MCD pattern by itself has poor fidelity. When the 21 proteins are examined for two consecutive, parallel β-sheet MCD loop patterns, involving a common interstrand $C_\alpha H$–NH NOE (the double pattern in Table I), the fidelity markedly increases (Table I).

Many of the single parallel sheet patterns found at distances greater than 3.4 Å are false positives; i.e., they involve residues not oriented as in the parallel sheet structure. However, many of the false positive patterns involve residues that participate in other, correct, MCD patterns, especially in the helical pattern at i, $i + 1$, $i + 2$, $i + 3$ residue positions. Thus, if patterns found for helix, antiparallel β-sheets, and turns are removed from consideration, the fidelity of apparent double parallel sheet patterns found then increases significantly at distances leading to very high frequency of observance.

This suggests that the search for MCD patterns should be performed in an ordered manner. A search for the helical MCD pattern should be carried out first as it involves unique interactions, the amide NH to $C_\beta H$ NOEs. Then NH to $C_\alpha H$ NOEs due to the identified helical residues (these NOEs occur often in helices; Wüthrich et al., 1984) should be found and eliminated. This will reduce ambiguities due to degeneracies in amide NH and α-CH resonances in subsequent searches for the antiparallel and parallel sheet patterns, which depend heavily upon NH–$C_\alpha H$ connectivities.

In applying the statistical analysis presented, one wants to set the NOESY mixing time to produce the highest frequency of MCD patterns with a high degree of fidelity. It can be noted that some MCD patterns display the conflicting dependence of frequency and fidelity upon interproton distance. For example, parallel sheet MCD patterns show decreased fidelity at the long NOE distances that lead to high frequency for the antiparallel MCD pattern. The results in Tables I and II indicate that one should fine tune the NOESY mixing time (distance) to allow the high fidelity patterns of the helical and parallel β-sheet patterns to be observed at short mixing times while the antiparallel β-sheet MCD patterns can be defined

in a NOESY spectrum obtained at a longer mixing time. The interdependence of β-sheet MCD patterns will also be useful. β-Sheets on average consist of 6.45 residues per strand and 4.7 strands per sheet (Sternberg & Thornton, 1977). Thus the loss of fidelity in the parallel sheet patterns obtained at long mixing times will be compensated by the appearance of other MCD patterns in the sheet structure formed by one (or both) of the strands with other strands in the sheet. This highly cooperative and confirmatory behavior was observed in the MCD analysis of human ubiquitin based upon a single NOESY spectrum (Di Stefano & Wand, 1987).

MCD Algorithm

These observations lead us to propose the following algorithm for applying the main-chain-directed assignment strategy to the analysis of 2D ^{1}H NMR spectra of proteins.

(1) NH–C_αH–C_βH units: define all apparent NH–C_αH–C_βH J-coupled sets through analysis of J-correlated spectra.

(2) Helix: (a) search for all groups of two consecutive interlocking helical MCD patterns; (b) extend each of these groups in both directions; (c) remove from further consideration all NOEs exclusively involving the main-chain protons of these residues.

(3) Antiparallel sheet: (a) search for the full antiparallel sheet pattern; (b) extend each full pattern by application of the hybrid and, when present, additional full antiparallel sheet patterns; (c) remove from further consideration all NOEs exclusively involving the main-chain protons of these residues.

(4) Parallel sheet: (a) search for the double parallel sheet pattern; (b) extend each double parallel sheet pattern by application of the single parallel sheet pattern; (c) remove from further consideration all NOEs exclusively involving the main-chain protons of these residues.

(5) Reconcile the results of (3) and (4). Construct the MCD-defined sheets found and incorporate turns revealed by (2) and (3).

(6) Extended chain: using only the remaining NOEs and main-chain units, search for unambiguous (i.e., no degeneracy) main-chain NOEs to define random coil regions.

(7) Structure alignment: identify and align the elements of secondary structure in the primary sequence by identification of amino acid side-chain spin systems that can be easily and reliably defined in J-correlated spectra.

(8) If desired, use the constraints imposed by (7) to help define the more complex amino acid side-chain J-correlated spin systems.

One appreciates that certain aspects of this approach have been used by other workers. However, we expect that when the entire approach is used as a general strategy, in conjunction with the tactical application of the closed loop patterns, a considerable gain in efficiency can be achieved. In addition, the MCD approach lends itself to computer-assisted analysis. This will be reported in a future communication.

References

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285–298.

Di Stefano, D. L., & Wand, A. J. (1987) *Biochemistry* (in press).

Jardetsky, O., & Roberts, G. C. K. (1981) in *NMR in Molecular Biology*, p 148, Academic, New York.

McDonald, C. C., & Phillips, W. D. (1969) *J. Am. Chem. Soc. 91*, 1513–1521.

Sternberg, M. J. E., & Thornton, J. M. (1977) *J. Mol. Biol. 110*, 285–296.

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

Wüthrich, K. (1983) *Biopolymers 22*, 131–138.

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.

Purification and Reconstitution of a 75-Kilodalton Protein Identified as a Component of the Renal Na^+/Glucose Symporter†

Jin-Shyun Ruth Wu and Julia E. Lever*

Department of Biochemistry and Molecular Biology, The University of Texas Medical School, Houston, Texas 77225

Received June 22, 1987; Revised Manuscript Received July 16, 1987

ABSTRACT: A 75-kilodalton (kDa) protein was purified from solubilized renal brush border membranes by using high-pressure liquid chromatography (HPLC) and preparative sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Functional and immunological properties identified the 75-kDa protein as a component of the Na^+/glucose symport system. The purified protein was specifically recognized by a monoclonal antibody that functionally interacts with the Na^+/glucose symporter. Na^+-dependent phlorizin binding activity was associated with fractions containing the 75-kDa protein during HPLC fractionation on the anion exchanger Mono-Q and was greatly increased after reconstitution into egg yolk phosphatidylcholine vesicles. The final purified preparation contained glucosamine and a blocked N-terminus.

Transporters localized in the apical membrane of renal proximal tubule and intestinal epithelial cells catalyze the coupled translocation of glucose and Na^+ in a symport (cotransport) mechanism (Crane, 1977). Na^+-dependent binding of phlorizin, a specific high-affinity inhibitor of this transport mechanism, has been established as an indication of the transport activity (Frasch et al., 1970; Lever, 1984). Attempts

†This work was supported by U.S. Public Health Service Grant DK27400.

0006-2960/87/0426-5958$01.50/0