INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONES: 224281/2/3/4/5/6
CABLE: CENTRATOM - TELEX 460392-1

SMR/111 - 3

SECOND SUMMER COLLEGE IN BIOPHYSICS

30 July - 7 September 1984

SEQUENTIAL INDIVIDUAL RESONANCE ASSIGNMENTS IN THE ^{1}H-NMR SPECTRA OF POLYPEPTIDES AND PROTEINS.

K. Wüthrich
Institut für Molekularbiologie und Biophysik
Eidgenössische Technische Hochschule
ETH-Hönggerberg
CH-8093 Zürich
Switzerland

Sequential Individual Resonance Assignments in the ^{1}H-NMR Spectra of Polypeptides and Proteins

KURT WÜTHRICH, *Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule, CH-8093 Zürich, Switzerland*

Synopsis

Recently, a new procedure for the assignment of protein ^{1}H-nmr spectra was introduced that relies on stereochemical considerations of proton–proton distances in polypeptides and on the use of two-dimensional nmr for obtaining ^{1}H-^{1}H through-bond and through-space connectivity maps. In the present paper a particular aspect of this assignment procedure is discussed in more detail, i.e., how to obtain individual resonance assignments from identification of amino acid side-chain spin systems and identification of neighboring residues in the amino acid sequence.

INTRODUCTION

Full use of the potentialities of nmr spectroscopy for structural studies of polypeptides and proteins depends on the identification of the individual resonance lines, and lack of reliable, generally applicable methods for obtaining individual assignments in polypeptide chains has long been a limiting factor. Once individual assignments can be obtained from the nmr data alone, i.e., without reference to three-dimensional structures of the protein obtained with other methods, all the "conventional" nmr studies of protein conformation, e.g., delineation of active centers in enzymes, comparison of crystal and solution conformations from ring-current calculations of the chemical shifts or studies of internal mobility in globular proteins,[1,2] attain a much higher degree of relevancy and reliability and their scope can be greatly widened. Furthermore, experience from work with smaller sets of assigned resonances[3,4] indicates that with nearly complete individual assignments of the resonance lines in a protein ^{1}H-nmr spectrum, the three-dimensional structure could be determined from the known amino acid sequence and a suitable set of nmr data.[5] This could then also, for the first time, provide a meaningful detailed comparison of polypeptide conformations in single crystals and in noncrystalline environments.

A general scheme for obtaining individual resonance assignments in protein ^{1}H-nmr spectra has recently been introduced.[6–8] It relies on stereochemical considerations of ^{1}H-^{1}H distances in polypeptide chains and on two-dimensional (2D) nmr experiments in H_2O solution,[9] which provide

through-bond and through-space ^1H-^1H connectivity maps for the entire protein. For the present discussion, three principal steps may be distinguished in the assignment procedure. These are (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 present paper starts with a brief survey of the data obtained in steps (i) and (ii), which provides the basis for a detailed discussion of the fundamental considerations that lead to individual resonance assignments.

IDENTIFICATION OF THE AMINO ACID SIDE-CHAIN SPIN SYSTEMS

For 15 of the 20 common amino acid residues the spin system of nonlabile hydrogen atoms[2] can in principle be identified entirely from studies of through-bond spin-spin coupling connectivities. These can be elucidated efficiently with 2D spin echo correlated spectroscopy (SECSY)[10,11] or 2D correlated spectroscopy (COSY).[11-13] To simplify the spectra, the labile protons are carefully replaced by deuterium prior to the nmr measurements. Then the SECSY or COSY spectra are screened for the typical connectivity patterns of the different types of amino acid residues.[8,14] The results that have thus far been obtained in practice in studies of small proteins at proton resonance frequencies of 360 and 500 MHz are illustrated in Figs. 1 and 2. For the side chains of Gly, Ala, Val, Ile, and Thr, which have unique and relatively simple systems of nonlabile hydrogen atoms, the complete spin systems can usually be identified. Then there is a group of eight amino acids that give rise to AMX spin systems originating from the $C^\alpha H$—$C^\beta H_2$ fragment (Fig. 2). This group includes the four aromatic amino acids, since the $C^\alpha H$—$C^\beta H_2$ fragments cannot usually be connected by J couplings with

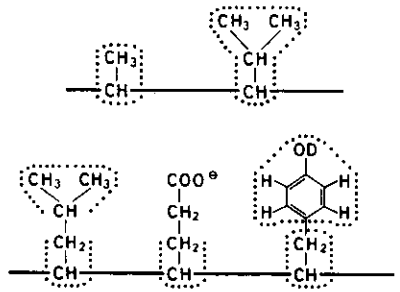

Fig. 1. Illustration of the extent to which the side-chain spin systems of the common amino acid residues can usually be identified in the initial analysis (i.e., before sequential assignments of the polypeptide backbone protons) of the protein ^1H-SECSY spectra recorded in D_2O (see text).

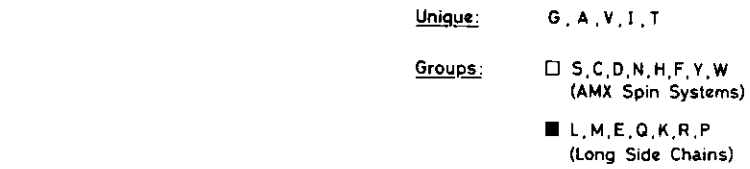

Fig. 2. Survey of the classification of the common amino acid residues obtained from the initial analysis of protein ^1H-SECSY spectra prior to the sequential assignments of the polypeptide backbone protons (see text).

the separately identified spin systems of the aromatic protons[2] (Fig. 1). Finally, there is a group of seven residue types with "long side chains," all of which, with the exception of Met, could be completely identified by J connectivities in principle. In practice, however, it has been difficult to find the connectivities beyond $C^\beta H_2$ in this initial analysis of the SECSY or COSY spectra. Leu presents a special case, since the peripheral fragment $-C^\gamma H(CH_3)_2$ is readily identified, whereas the connectivity between $C^\beta H_2$ and $C^\gamma H$ is usually not found at this stage (Fig. 1). In principle, Pro can be separated from the other residue types in this group when spectra in H_2O are inspected, since Pro $C^\alpha H$ is unique in that it does not have a spin–spin coupling connectivity with an amide proton.

IDENTIFICATION OF NEIGHBORING RESIDUES IN THE AMINO ACID SEQUENCE

Since hydrogen atoms in sequentially adjoining residues are separated by four or more covalent bonds (Fig. 3), the homonuclear spin–spin coupling constants are too small to be used for identification of neighboring residues. Therefore, the through-space distances d_1, d_2, and d_3 as defined in Fig. 3 are the crucial quantities for "sequential assignments." d_1, d_2, and d_3 depend on at least one of the dihedral angles ϕ_i, ψ_i, and χ_i^1, and stereochemical considerations show that for all sterically allowed conformations at least one of these distances is shorter than 3.0 Å.[6,15-17] With suitably executed 2D nuclear Overhauser enhancement (NOESY) experiments[18-21]

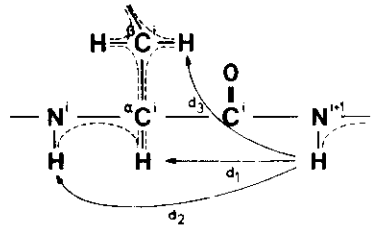

Fig. 3. Polypeptide backbone segment. The through-space distances d_1, d_2, and d_3 used for the sequential resonance assignments are indicated by arrows. The broken lines indicate through-bond spin–spin coupling connectivities within amino acid residue i.

in H_2O, where all the amide protons are present (Fig. 3), semiquantitative measurements of the distances d_1, d_2, and d_3 in the range from ca. 2.0 to 4.0 Å can be obtained. From a statistical analysis of ^1H-^1H distances in refined protein crystal structures, it was estimated that neighboring amino acid residues can be identified with a reliability of ca. 80% when one of the connectivities d_1, d_2, or d_3 is found to be $\lesssim$3.0 Å, and with a reliability of ca. 90% when any two of the three connectivities are $\lesssim$3.0 Å.[6]

When the NOESY experiments are complemented by COSY or SECSY experiments in H_2O, which delineate the $C_i^\beta H_n$—$C_i^\alpha H$—$N_i H$ through-bond spin-spin couplings (Fig. 3), the resonances of the backbone and C^β hydrogen atoms of successive residues in the amino acid sequence can be identified step by step (Fig. 3). In practice, these identifications extend usually over peptide segments of 2 to ca. 8 neighboring amino acid residues.

LOCATING SEQUENTIALLY IDENTIFIED PEPTIDE SEGMENTS IN THE PRIMARY STRUCTURE

Except in particularly favorable cases—for example, when the protein contains only a single residue of Gly, Ala, Val, Ile, Thr, Leu, or one of the aromatic amino acids—no individual resonance assignments result from the experiments described in the two preceding sections. It therefore remains to locate the sequentially connected peptide segments in the primary structure. Here, we shall discuss how this additional information can be extracted from combination of the data on the partial identification of amino acid side-chain spin systems (Fig. 2) and the identification of neighboring residues in the sequence.

Sequential identification of neighboring residues (Fig. 3) and identification of amino acid side-chain spin systems (Fig. 1) both cover the C^α and C^β hydrogen atoms of the amino acids. When the chemical shifts for the $C^\alpha H$—$C^\beta H_n$ fragments in these two sets of data are matched, each residue in the sequentially connected peptide segments can be identified either

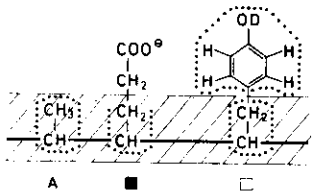

Fig. 4. Illustration of how the data from identification of the side-chain spin systems (indicated by dotted lines; see Fig. 2 for an explanation of the code at the bottom of the figure) and sequential identification of neighboring residues (inside shaded ribbon) are combined. Both data cover the C^α and C^β hydrogen atoms and thus, from matching the $C^\alpha H$ and $C^\beta H_n$ chemical shifts in the two data sets, each sequentially connected residue can be identified either as a unique amino acid or as one of a group of amino acid types (see Fig. 2).

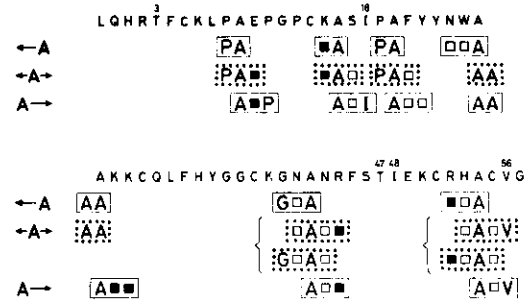

Fig. 5. Illustration of how individual resonance assignments were obtained in the ^{1}H-nmr spectrum of the proteinase inhibitor E from the venom of *Dendroaspis polylepis polylepis*. The procedure is explained in the text. The symbols □ and ■ are explained in Figs. 2 and 4.

as a unique amino acid residue, as a residue with an AMX spin system, or as a residue with a long side chain (Figs. 2 and 4). The resulting fragmentary peptide sequences are then checked against the protein primary structure to locate their positions in the overall amino acid sequence. In the following this is illustrated for a small protein of molecular weight 6500, the proteinase inhibitor E from the venom of *Dendroaspis polylepis polylepis*.[22]

It is advisable to start work on the resonance assignments with the amino acid residues with unique proton spin systems (Fig. 2). In the inhibitor E (Fig. 5), identification of the spin system of Val 56 resulted in the first individual resonance assignment, since this is the only valine in this protein. Further, there are two isoleucines in positions 18 and 48, and two threonines in positions 3 and 47. Once the sequential connectivity between Thr 47 and Ile 48 was found, these four spin systems could also be individually assigned. The sequence of the inhibitor E contains seven alanines in positions 9, 16, 20, 26, 27, 42, and 54, of which the spin systems could readily be identified.[22] Starting from any one of the seven alanines, sequential identification of the neighboring residues can be pursued either in the direction toward the N-terminus or the C-terminus or in both directions (see Fig. 5). When going toward the N-terminus, identification of a dipeptide fragment suffices to locate Ala 16 and Ala 27 in the sequence, whereas tripeptide segments are needed for Ala 26, Ala 42, and Ala 54. The alanines 9 and 20, which both follow proline in the sequence, cannot generally be distinguished by this experiment, since sequential assignments across prolyl residues have so far been obtained only in particularly favorable instances.[7,22] However, when sequential identification towards the C-terminus is used, Ala 26 can be assigned from a dipeptide segment and each of the other six alanines from a tripeptide segment. Further, all but two alanines can be assigned from identification of the two neighboring residues in the sequence (Fig. 5).

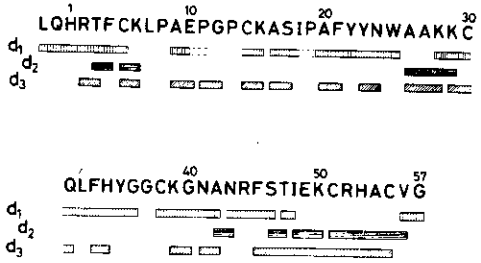

Fig. 6. Amino acid sequence of trypsin inhibitor E from *Dendroaspis polylepis polylepis* and survey of the sequential connectivities bv which the individual resonance assignments were obtained. The numeration of residues starts with Leu −1, Gln 0, His 1, ... *Symbols:* ▣, sequential assignments via d_1 (NOE from NH_{i+1} to $C^{\alpha}H_i$); ▨, sequential assignments via d_2 (NOE from NH_{i+1} to NH_i); ▩, sequential assignments via d_3 (NOE from NH_{i+1} to $C^{\beta}H_i$; ⬚, sequential assignments via NOEs from proline $C^{\delta}H_{i+1}$ to $C^{\alpha}H_i$; Leu 7 and Gly 37 were individually assigned, since there were no other leucine or glycine spin systems left unassigned at the end of the spectral analysis (adapted from Ref. 22).

Once the location of a sequentially connected segment of neighboring amino acid residues in the primary structure has been determined, further extension of the sequential assignments is usually considerably facilitated, since each step can then be checked against the amino acid sequence. The final result obtained for the inhibitor E is shown in Fig. 6. Furthermore, individual assignments result for the AMX-type residues (Fig. 2), and some of the long side chains can usually be completely identified when one starts from the sequentially assigned $C^{\alpha}H$—$C^{\beta}H_2$ fragments.[7,8,22]

DISCUSSION

The main purpose of this paper is to illustrate the fundamental ideas that have resulted in nearly complete individual assignments of the ^{1}H-nmr spectra of several polypeptides and small proteins.[7,8,22] For ease of presentation, it was helpful to divide the overall assignment procedure into the three distinct steps described in the preceding sections and to assume that to the extent indicated in Fig. 2, the identifications of amino acid side-chain spin systems can be obtained without ambiguity. In practice, there may be some limitations. For example, since for many of the long side chains the $C^{\beta}H$—$C^{\gamma}H$ J connectivities cannot usually be readily detected in SECSY or COSY, the distinction from the $C^{\alpha}H$—$C^{\beta}H_2$ three-spin systems relies initially also on measurements of the $C^{\beta}H_2$ chemical shifts. However, even though in all the long side chains the random-coil shifts for $C^{\beta}H_2$ are at least 0.6 ppm to higher field than in any of the AMX spin systems,[2,23] the chemical-shift criterion is somewhat ambiguous in globular proteins because of the conformation-dependent contributions to the observed shifts.[2] From the combined data on side-chain spin systems and sequentially neighboring residues, the assignments are usually sufficiently

overdetermined to clarify these uncertainties, but it may be advisable with certain peptide fragments to pursue all three steps of the assignment procedure in parallel. It is, of course, also possible that in addition to those listed in Fig. 2, unique identifications are obtained for other residues, e.g., for Asp and Glu from pH titrations[14] or for the aromatic residues from NOEs between $C^{\beta}H_2$ and the ring protons.[6-8,22] This can quite obviously provide further help in eliminating ambiguities in the assignments of nearby residues. For the future, it is quite conceivable that the entire assignment procedure will be pursued with ^{1}H-nmr spectra recorded in H_2O, and the three distinct steps described here might then be combined in the analysis of a single set of 2D-nmr data.

The strategy for assignment of protein ^{1}H-nmr spectra outlined in Figs. 1–5 should be quite widely applicable, since in natural polypeptide chains one observes few repetitions of identical peptide segments of four or more residues. Similar principles could be applied for fragmentary (see Fig. 2) polypeptide sequence analysis by nmr,[16] but in view of the well-established chemical procedures, it appears rather unlikely that polypeptide sequencing would be the most useful application of nmr. However, the possibility of checking known primary structures in selected positions may well become of considerable interest.

It can be expected that future developments of nmr instrumentation and nmr methodology will further enhance the usefulness of the approach to resonance assignments in proteins described here. For example, it seems reasonable to expect that a more extensive differentiation of amino acid spin systems (Fig. 2) will be achieved in the future. Depending on further improvements of the sensitivity of 2D-nmr experiments, it should also be feasible to automate many of the assignment procedures. Detailed structural studies in noncrystalline environments should thus become practical for a wide variety of small and medium-size polypeptides and proteins.

I would like to thank Mrs. E. H. Hunziker and Mrs. E. Huber for the careful preparation of the figures and the manuscript and Drs. A. S. Arseniev, W. Braun, G. Wagner, and G. Wider and Mr. M. Billeter for enlightening discussions on the subject of this paper. Financial support by the Schweizerischer Nationalfonds (Project 3.528.79) is gratefully acknowledged.

References

1. Roberts, G. C. K. & Jardetzky, O. (1970) *Adv. Protein Chem.* **24**, 447–545.
2. Wüthrich, K. (1976) *NMR in Biological Research: Peptides and Proteins*, North-Holland, Amsterdam.
3. Braun, W., Bösch, C., Brown, L. R., Gō, N. & Wüthrich, K. (1981) *Biochim. Biophys. Acta* **667**, 377–396.
4. Crippen, G. M., Oppenheimer, N. J. & Connolly, M. L. (1981) *Int. J.Pep. Protein Res.* **17**, 156–169.
5. Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982) *J. Mol. Biol.* **155**, 311–319.

6. Billeter, M., Braun, W. & Wüthrich, K. (1982) *J. Mol. Biol.* **155,** 321–346.
7. Wagner, G. & Wüthrich, K. (1982) *J. Mol. Biol.* **155,** 347–366.
8. Wider, G., Lee, K. H. & Wüthrich, K. (1982) *J. Mol. Biol.* **155,** 367–388.
9. Anil Kumar, Wagner, G., Ernst, R. R. & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **96,** 1156–1163.
10. Nagayama, K., Wüthrich, K. & Ernst, R. R. (1979) *Biochem. Biophys. Res. Commun.* **90,** 305–311.
11. Nagayama, K., Anil Kumar, Wüthrich, K. & Ernst, R. R. (1980) *J. Magn. Reson.* **40,** 321–334.
12. Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976) *J. Chem. Phys.* **64,** 2229–2246.
13. Bax, A. & Freeman, R. (1981) *J. Magn. Reson.* **44,** 542–561.
14. Nagayama, K. & Wüthrich, K. (1981) *Eur. J. Biochem.* **114,** 365–374.
15. Leach, S. J., Némethy, G. & Scheraga, H. A. (1977) *Biochem. Biophys. Res. Commun.* **75,** 207–215.
16. Kuo, M. & Gibbons, W. A. (1979) *Peptides: Structure and Biological Function, Proceedings of the Sixth American Peptide Symposium*, Gross, E. & Meienhofer, J., Eds., Pierce Chem. Co., Rockford, Ill., pp. 229–232.
17. Dubs, A., Wagner, G. & Wüthrich, K. (1979) *Biochim. Biophys. Acta* **577,** 177–194.
18. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979) *J. Chem. Phys.* **71,** 4546–4553.
19. Anil Kumar, Ernst, R. R. & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95,** 1–6.
20. Anil Kumar, Wagner, G., Ernst, R. R. & Wüthrich, K. (1981) *J. Am. Chem. Soc.* **103,** 3654–3658.
21. Macura, S., Wüthrich, K. & Ernst, R. R. (1982) *J. Magn. Reson.* **46,** 269–282.
22. Arseniev, A. S., Wider, G., Joubert, F. J. & Wüthrich, K. (1982) *J. Mol. Biol.* **159,** 323–351.
23. Bundi, A. & Wüthrich, K. (1979) *Biopolymers* **18,** 285–297.

Received June 20, 1982
Accepted August 24, 1982