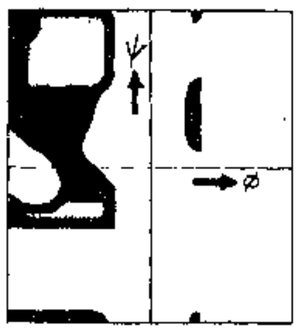


Special issue:
The Ramachandran map and
protein and peptide structures

The Ramachandran plot and the NMR method for protein structure determination

Kurt Wüthrich

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

The Ramachandran plot for visualization of the sterically allowable combinations of polypeptide backbone dihedral angles ϕ_i and ψ_i in the ϕ–ψ plane has had considerable impact on the development of the now widely used NMR method for protein structure determination in solution. This paper first presents a brief survey of the determination of NMR structures of proteins. It then focuses on the role played by the Ramachandran plot in the development of a general strategy for sequential NMR assignments, and the identification of regular secondary polypeptide structures by NMR.

Survey of the NMR method for protein structure determination

THE scheme in Figure 1 presents an outline of a protein structure determination by nuclear magnetic resonance (NMR) spectroscopy[1, 2]. For the data collection the protein is dissolved in an aqueous solvent. The protein concentration must be of the order of 1–5 mM (or possibly higher). The sample volume is 0.5 ml, so that 5–25 mg of a protein with molecular weight 10000 are needed for the preparation of a NMR sample. The pH, ionic strength and temperature can be adjusted so as to mimic the physiological milieu of the protein.

The foundations of the method are: (i) NMR experiments enabling the use of *nuclear Overhauser effects (NOE) for measurements of proton–proton distances* in native protein structures in the presence of spin diffusion[3–5]. This can be achieved by measurements of NOE buildup curves[6] in one-dimensional[6, 7] or multidimensional[8] experiments. Suitably executed NOE measurements can thus provide information on proton–proton distances in the range from about 2.0 to 5.0 Å. In a qualitative interpretation the observation of each ^1H–^1H NOE manifests an upper bound on a ^1H–^1H distance. The information contained in a small region of a two-dimensional NOE (NOESY) spectrum (Figure 2, *a*) can then be represented by the scheme in Figure 3, *a*. (ii) *Sequential resonance assignments* as an efficient technique for obtaining sequence-specific ^1H NMR assignments. By the fact that polypeptide chains in proteins generally contain multiple units of each amino acid type, NMR spectral assignments are non-trivial and no generally applicable assignment procedure was available until 1982, when the sequential assignment strategy was introduced[9–12]. The crucial importance of the resonance assignments as a basis for protein structure determination[9] is illustrated with Figures 2 and 3. The small spectral region from a NOESY

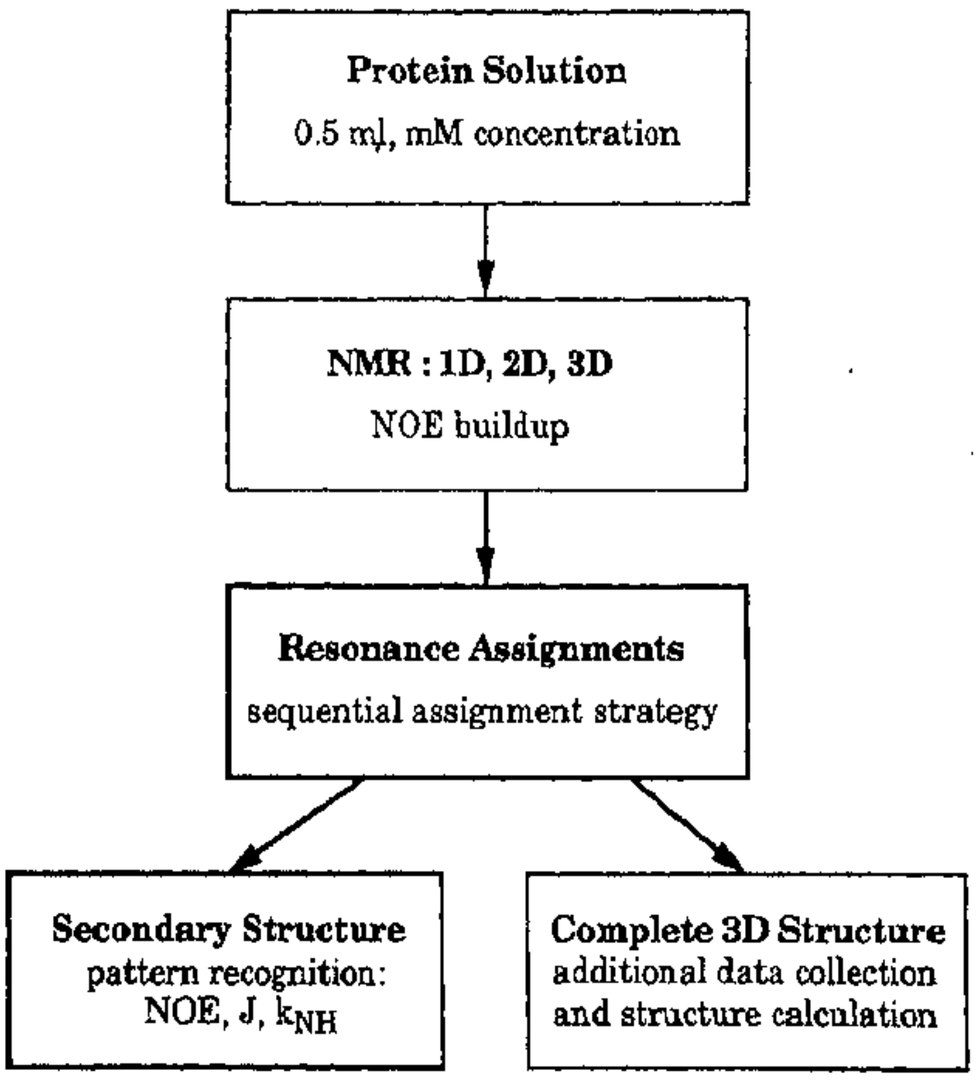


Figure 1. Diagram outlining the course of a protein structure determination by NMR (see text for further explanations).

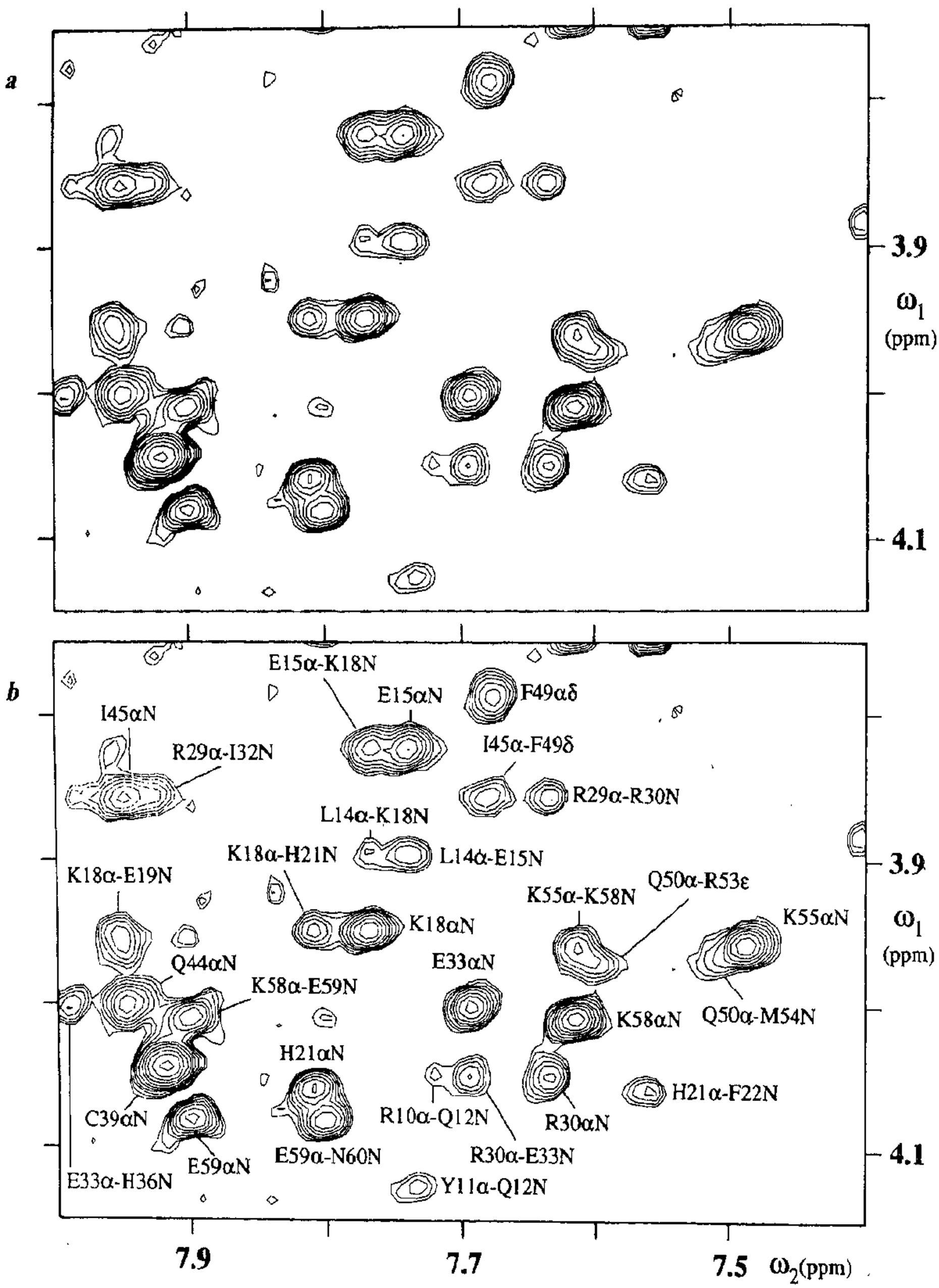
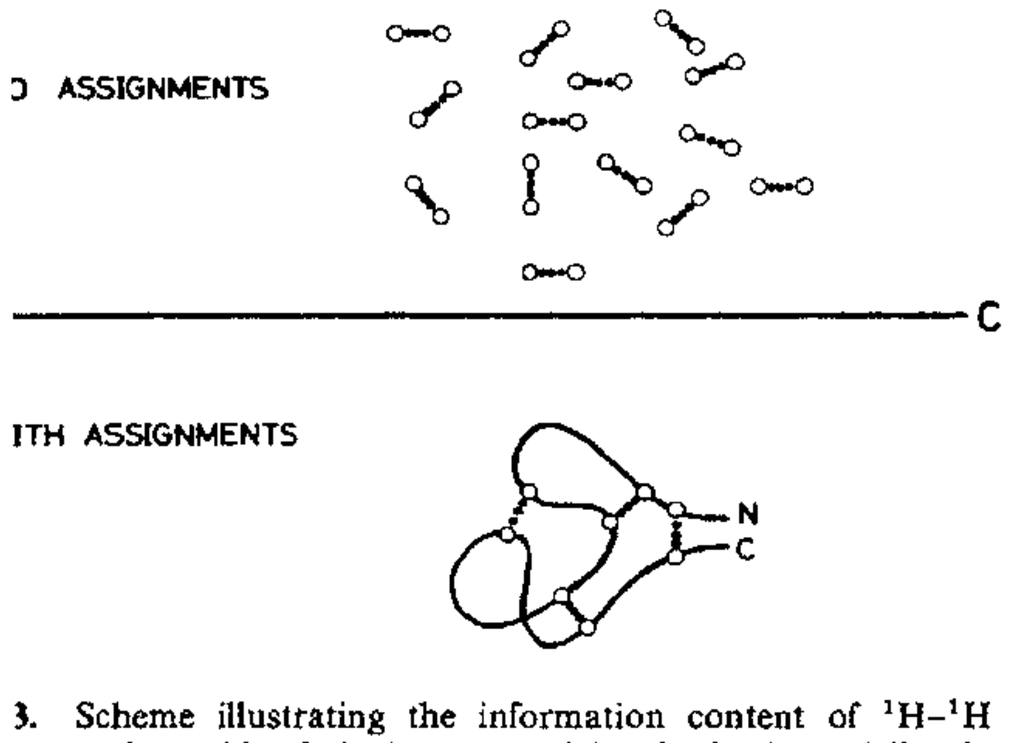


Figure 2. *a*, Contour plot of a small region from a ^{1}H soft-NOESY spectrum of the *Antennapedia* homeodomain. *b*, Same spectral region with sequence-specific resonance assignments. For interresidual NOEs the cross peaks are identified by the one-letter amino acid symbols of the two residues, their sequence positions and the proton types, and for intraresidual NOEs by the identification of one residue and two proton types.

spectrum shown in Figure 2, *a* contains 30 cross peaks. In the absence of sequence-specific resonance assignments each of these peaks merely indicates the presence of two nearby hydrogen atoms, which may be located anywhere in the polypeptide chain (Figure 3, *a*). When resonance assignments have been obtained, each cross



3. Scheme illustrating the information content of ^1H–^1H n a polypeptide chain (represented by the horizontal line in re) with and without sequence-specific resonance assignments. :cles represent hydrogen atoms of the polypeptide, and dotted ort ^1H–^1H distances manifested by the NOEs (see text). uced from ref. 1.)

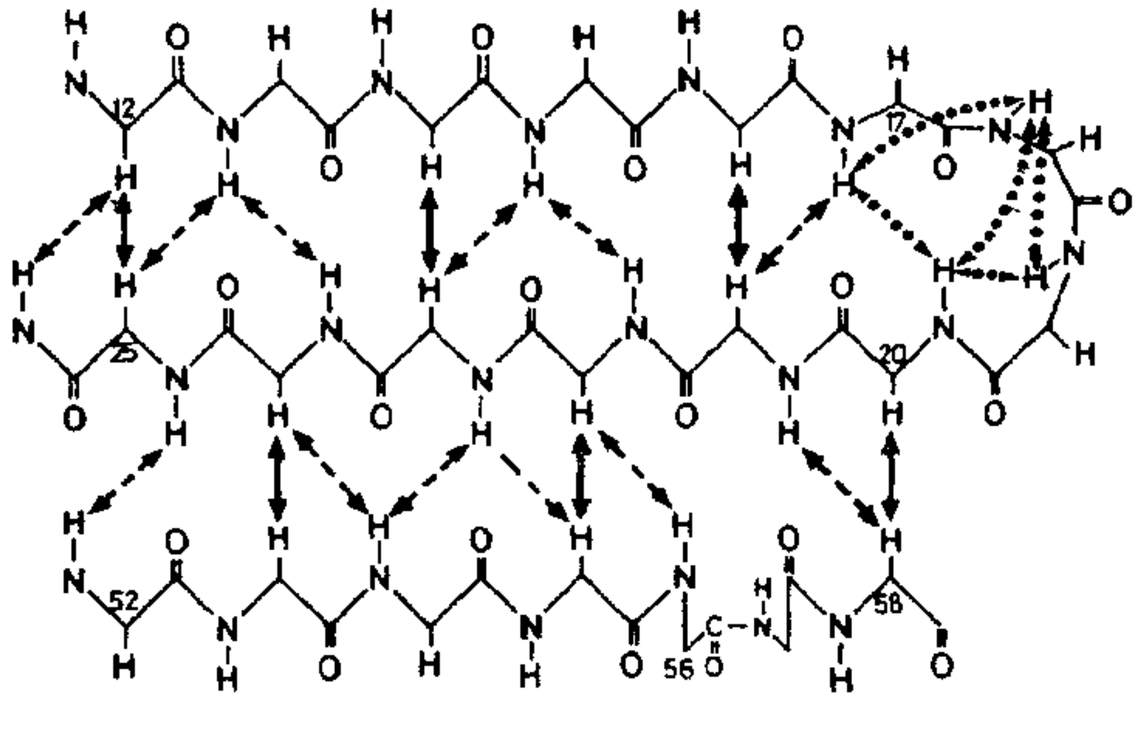


Figure 4. An antiparallel β-sheet in Tendamistat (Figure 5) identified by NMR[24] using the empirical pattern recognition approach[14]. The arrows indicate long-range interstrand NOEs, which provided the experimental basis for docking the individual strands relative to each other (reproduced from ref. 24).

dentifies an upper limit on the distance between istinct locations along the polypeptide chain : 3, *b*), which is the information needed for the ination of the three-dimensional protein structure. l, in its impact on the NMR structure determina- ethod, the sequential assignment strategy can be red to the use of isomorphous heavy atom ives for solving the phase problem in protein lography[13]. (iii) Methods for the *structural :tation of the individually assigned intramolecular e constraints* obtained from NOE experiments. ing the scheme of Figure 1, two different ches can be taken. One is an empirical search for s of NOE distance constraints, dihedral angle .ints from measurements of vicinal spin–spin ıg constants, and slow amide proton exchange dentifying hydrogen-bonded backbone amide s[1, 10, 14], which enable the identification of secondary structures (Figure 4), and in some ıble instances, also tight turns. Figure 4 illustrates e sequence locations and the lengths of regular ary structures can be precisely defined by this ch. Alternatively, the complete three-dimensional structure (Figure 5) can be determined from an :ive analysis of the NMR spectra, which yields ıut for the structure calculation. Mathematical ues used for this purpose include metric matrix e geometry, a variable target function algorithm, trained molecular dynamics calculations[1, 2, 15]. :ions with the Ramachandran plot have been hed for those steps of a NMR structure deter- ın (Figure 1) which are focused on the local nation. Local structure can be characterized by *idual* NOEs and spin–spin coupling constants, *sequential* NOEs between protons attached to s that are neighbours in the amino acid ce. Sequential resonance assignments rely entirely on such short-range interactions. Sequential NOEs and vicinal amide proton–C^α proton scalar couplings contribute also to the identification of secondary structures by pattern recognition. In the following sections these two steps in the scheme of Figure 1 are discussed in some more detail.

Sequential resonance assignments

Besides NOESY, obtaining resonance assignments in a protein makes use of NMR experiments which delineate scalar, through-bond relations between nuclear spins, for example, 2D correlated spectroscopy (COSY) and total correlation spectroscopy (TOCSY)[1, 16]. In proteins these techniques delineate networks of scalar spin–spin couplings between hydrogen atoms that are separated from each other by three or less covalent bonds, and in this way the *spin systems of the individual amino acid residues* can be identified[1]. However, because typical proteins contain multiple copies of the 20 common amino acids, this identification of the groups of scalar-coupled spins belonging to individual amino acid residues is in general not sufficient to define a unique sequence location. The problem can be solved if instead of individual amino acid spin systems, the identity of two or several neigbouring spin systems in the primary structure can be established by NMR experiments, as is illustrated in Figure 6. The dotted lines indicate the ^1H–^1H connectivities within the amino acid residues, which can be established via scalar spin–spin couplings. In the example of Figure 6 these spin systems unambiguously identify the amino acid residues alanine and valine[1]. Relations between protons in sequentially neighbouring amino acid residues are established by sequential NOEs manifesting close approach between $C^\alpha H(i)$ and $NH(i+1)$, or both. In

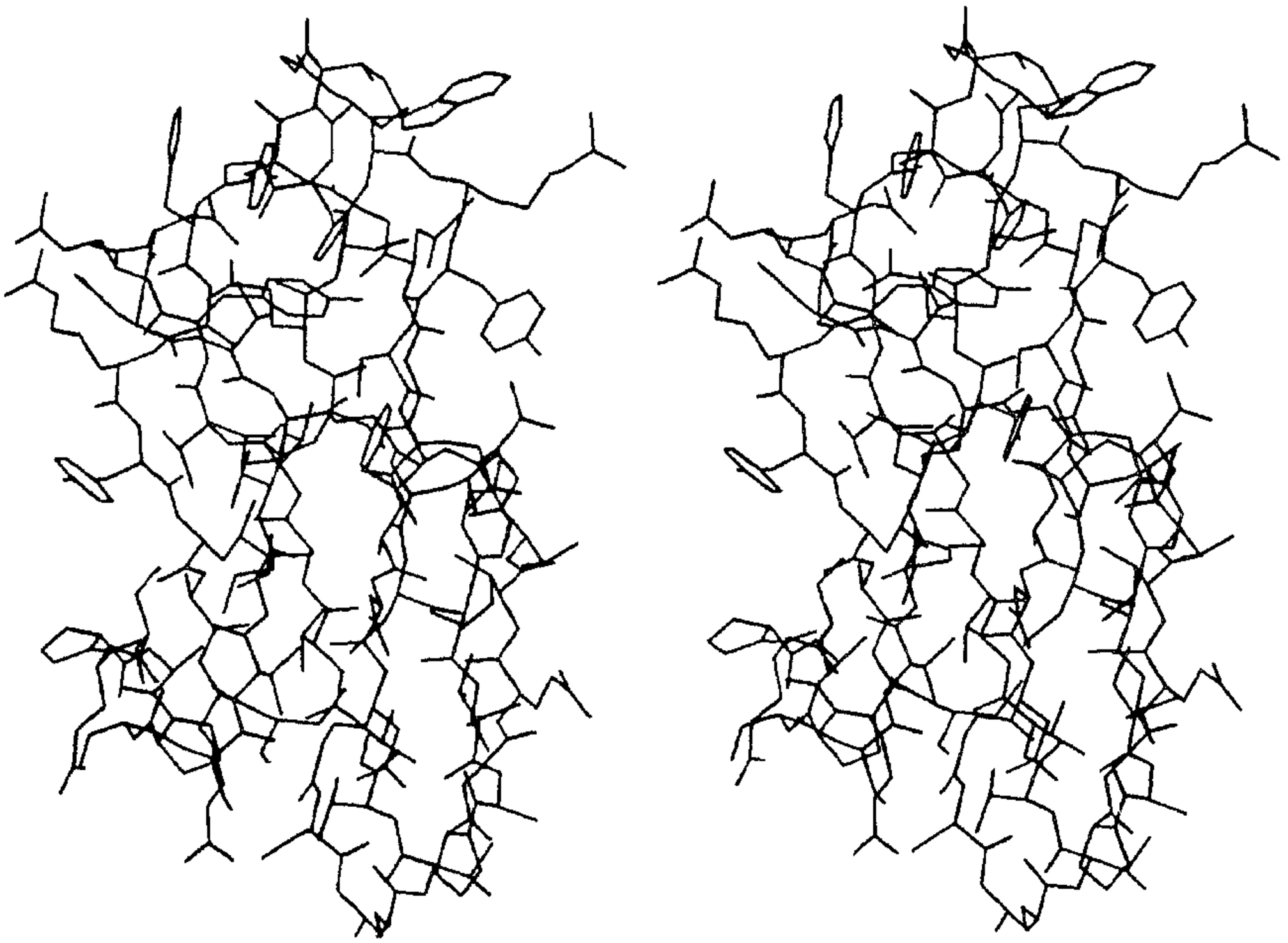


Figure 5. Stereo view of the three-dimensional structure of the protein Tendamistat determined from NMR measurements in aqueous solution. All bonds connecting heavy atoms are drawn for the residues 5–73 (ref. 25). The complete molecule contains 74 amino acid residues and has a molecular weight of 9000.

Figure 6 we obtain the result that the protein studied contains a dipeptide segment Ala–Val. This dipeptide is then matched against the independently known amino acid sequence. If the latter contains Ala–Val only once, the assignment problem is solved. Otherwise, to distinguish between the different Ala–Val sites, tri- or tetrapeptide segments including Ala–Val must be identified by NMR and matched against the amino acid sequence.

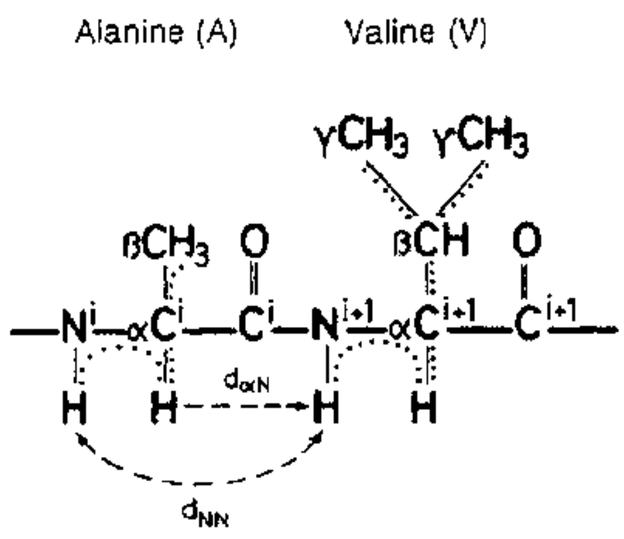


Figure 6. Illustration to the description of sequential resonance assignments. In the dipeptide segment –Ala–Val– the dotted lines indicate ^1H–^1H relations which can be established by the scalar spin–spin couplings observed, for example, in COSY. The broken arrows indicate relations between protons in sequentially neighbouring residues, which can be established by NOESY cross peaks manifesting short sequential distances $d_{\alpha N}$ (between $C^{\alpha}H$ and the amide proton of the following residue) and d_{NN} (between the amide protons of neighbouring residues in the amino acid sequence).

When working on the development of the sequential resonance assignment technique we were primarily guided by the following two observations[10, 17]: (i) A statistical analysis of the protein sequence data bank showed that there is only a very small probability that identical tri- or tetrapeptide segments occur repeatedly in the primary structure of globular proteins with less than 200 amino acid residues[1]. This ensures that the sequential assignment strategy of identifying several sequentially neighbouring amino acid spin systems is a generally applicable avenue to obtaining the desired sequence-specific ^1H NMR assignments. Thereby the sequential connectivities may be established either by ^1H NOESY (Figure 6) or by heteronuclear correlation experiments with isotope-labelled proteins[1]. (ii) Reference to the information contained in the *Ramachandran plots*[18, 19] for the proteinogenic amino acids revealed that in the sterically allowed regions of local conformation space, at least one of the two sequential distances between polypeptide backbone protons, $d_{\alpha N}$ and d_{NN}

(Figure 6), is always shorter than 3.0 Å, independent of the conformation. Such closely spaced protons are related by intense NOEs, which give prominent peaks in one-dimensional or multi-dimensional NOE experiments. This showed that the homonuclear ^{1}H NMR approach to sequential assignments (Figure 6) is generally applicable as long as the NOE cross peaks corresponding to $d_{\alpha N}$ (Figure 2) and d_{NN} can be resolved. Experience has since shown that the spectral resolution in two-dimensional ^{1}H NOESY spectra at 500 or 600 MHz usually becomes a limiting factor when the protein has a molecular weight of the order 10000–15000.

Figure 7 illustrates the relation between the proton–proton distance $d_{\alpha N}$ and the dihedral angle ψ_i, and between d_{NN} and the two angles ϕ_i and ψ_i (the peptide bond is assumed to be in the planar *trans* form). Figures 8 and 9 display curves representing the dependence of the sequential ^{1}H–^{1}H distances on the dihedral angles, which were computed using the average of the standard ECEPP geometry[20] for the common amino acid residues[10]. Figure 8 shows that for a NOE measurement registering ^{1}H–^{1}H distances up to $\sim$3.0 Å (this would be with the use of a very short mixing time in NOESY[1, 16]), $d_{\alpha N}$ is an observable parameter for the entire sterically allowed regions B and C in the ϕ_i–ψ_i plane (Figure 9). A particularly small $d_{\alpha N}$ value of $\sim$ 2.2 Å prevails for β-structures. In the region A, which includes the α-helix, $d_{\alpha N}$ is between 3.3 and 3.5 Å. For glycine in position i, $d_{\alpha N}$ is $\leqslant$3.1 Å in the entire sterically accessible region. With today's NMR instrumentation, NOE experiments can be set up so that all $d_{\alpha N}$ connectivities should produce well observable NOEs. Figure 9 shows a Ramachandran plot[18], with the sterically allowed regions A, B and C calculated for an Ala dipeptide unit[19], and supplemented with contour lines for fixed values of d_{NN} (ϕ_i, ψ_i) between 1.5 and 4.5 Å. It is seen that d_{NN} is $\lesssim$3.0 Å for the entire region A and about half of the region C. Most of the region B with β_p and β lies in the range of d_{NN} values between 3.5 and 4.5 Å, so that only part of the

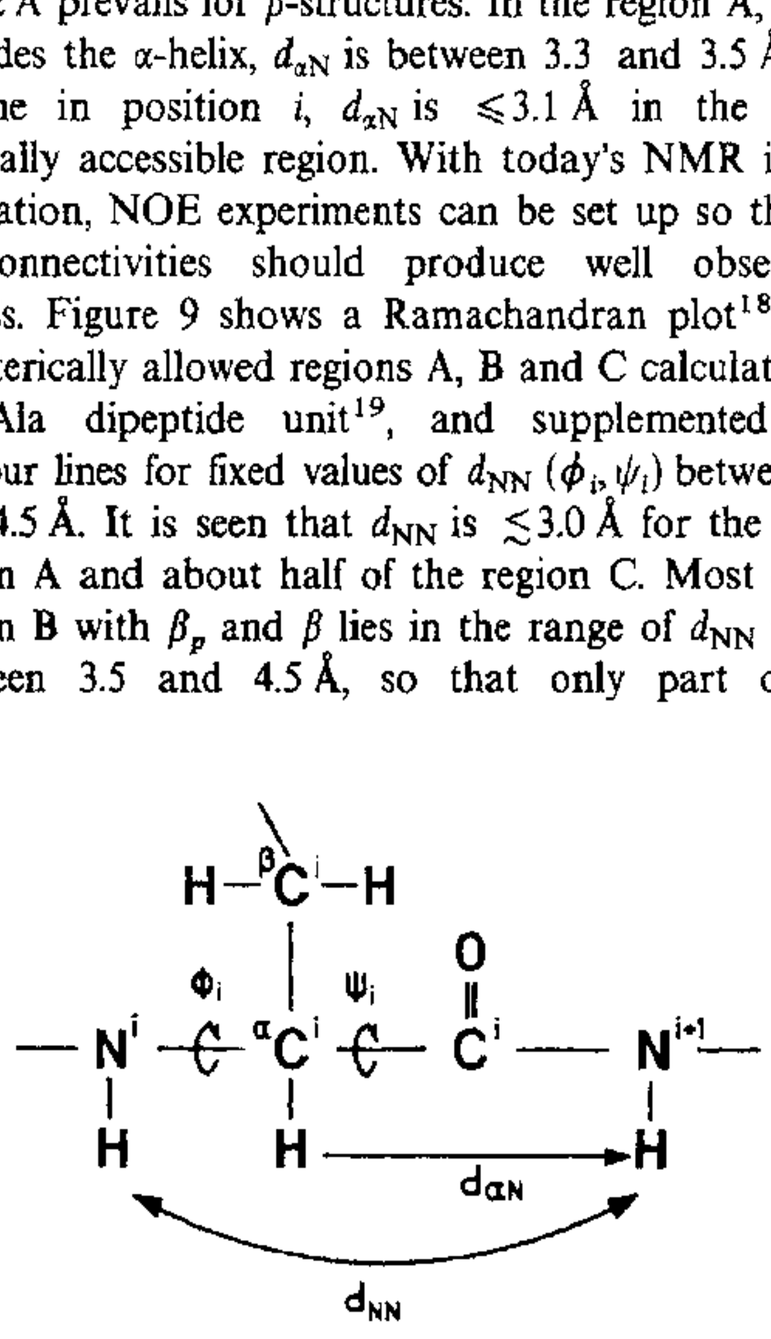


Figure 7. Polypeptide segment consisting of the amino acid residue i and the NH group of the residue $(i+1)$. The dihedral angles ϕ_i and ψ_i and their relations with the distances $d_{\alpha N}$ and d_{NN} are indicated.

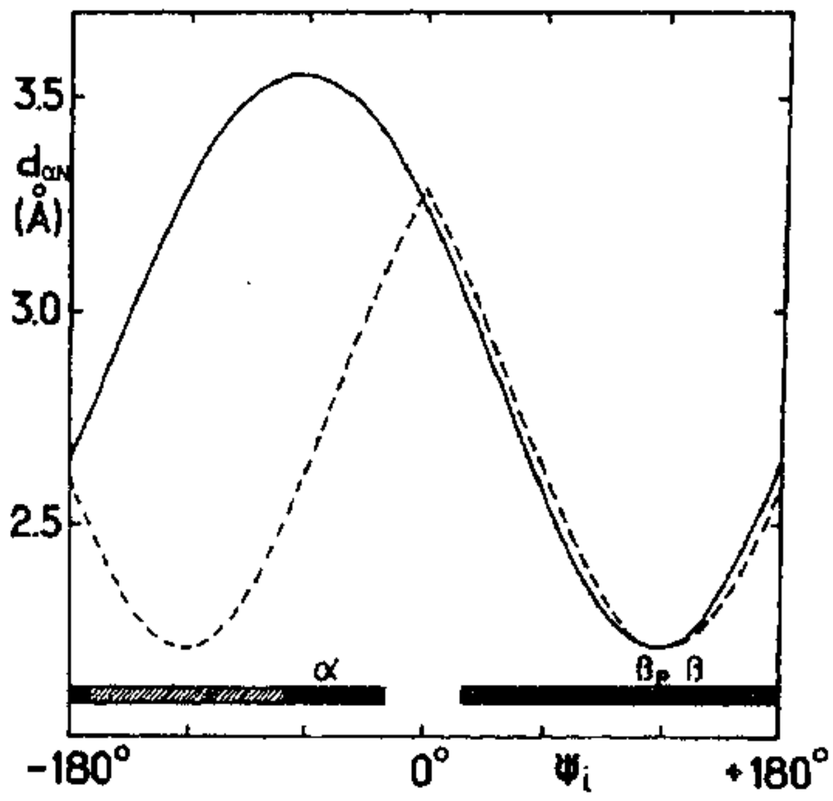


Figure 8. Plot of the distance $d_{\alpha N}$ between $C^{\alpha}H_i$ and NH_{i+1} (see Figure 7) vs. the dihedral angle ψ_i. (———), Common L-amino acid in position i; (— — —) Gly in position i, where for each value of ψ_i the smaller one of the two distances between NH_{i+1} and the two α-protons of Gly is displayed. In the range from 0 to +180°, small differences between the two curves arise as a consequence of small deviations between the ECEPP parameters for Gly and for the other residues[20]. The range of sterically allowed ψ_i values for the common L-amino acids is indicated by the black bar at the bottom of the figure. For glycine, the ψ_i values indicated by the hatched bar are also allowed. The letters α, β_p and β identify the ψ_i values for the α-helix, the parallel β-structure and the antiparallel β-structure. (Reproduced from ref. 10.)

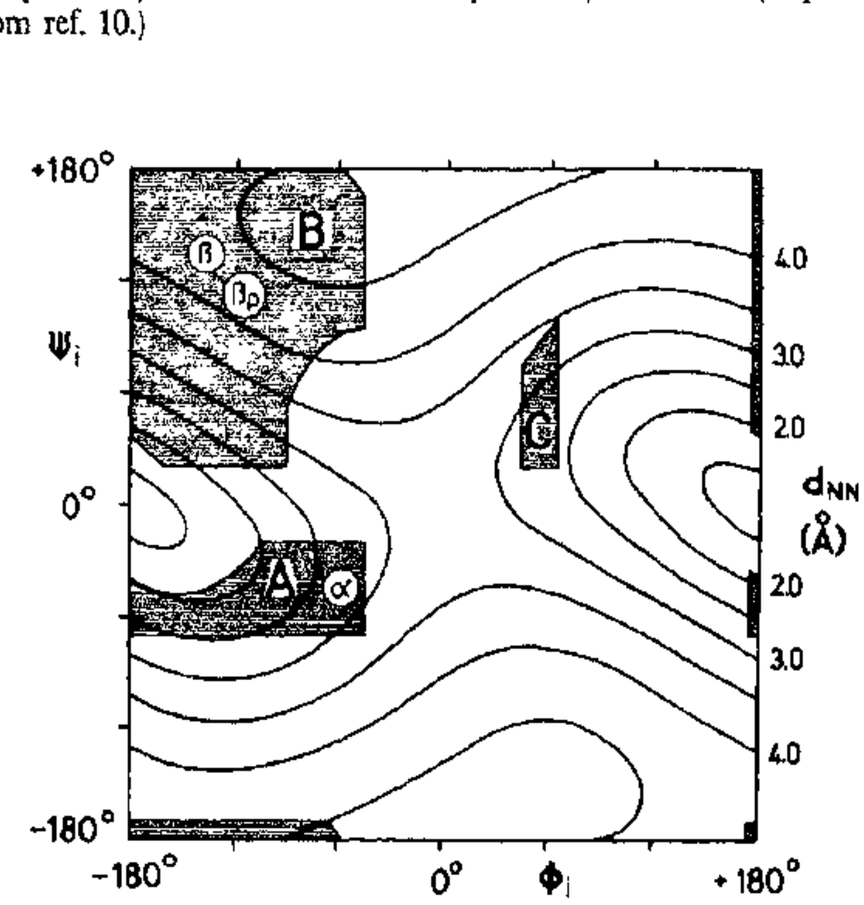


Figure 9. Contour lines for fixed values of the distance d_{NN} between NH_i and NH_{i+1} (see Figure 7) in the ϕ_i–ψ_i plane. The d_{NN} values for the individual contour lines are given on the right. The areas in the ϕ_i–ψ_i plane that are sterically allowed for Ala (ref. 19) are hatched and labelled A, B and C. The ϕ_i–ψ_i combinations for the regular α-helix and the parallel and antiparallel β-structures are indicated by α, β_p and β. (Reproduced from ref. 10.)

d_{NN} connectivities should be observable in NOE measurements with typical globular proteins.

In summary, in the Figures 7–9 the experience of the 1960s regarding the description and nature of local conformations in polypeptide chains, to which Prof. Ramachandran has made so many seminal contribu-

tions, is extended for use in the space spanned by the sequential ^1H–^1H distances. This presentation makes intuitively clear that the sequential assignment procedure outlined in Figure 6 is generally applicable with proteins. The impact of this approach is best illustrated by the fact that since 1982 (refs. 9–12) sequence-specific NMR assignments were published for over 200 polypeptides and proteins, which provided the basis for the determination of over 50 NMR structures of proteins (Figure 5).

Sequential NOEs and the identification of regular secondary polypeptide structures

Because of their pivotal role in the procedures used for obtaining sequence-specific resonance assignments, the sequential distances $d_{\alpha N}$ and d_{NN} are extensively investigated in the early phases of a protein structure determination by NMR. It was therefore of practical interest to explore possible uses of $d_{\alpha N}$ and d_{NN} for polypeptide spatial structure determination. Figures 7–9 show that there are direct relations between the sequential ^1H–^1H distances $d_{\alpha N}$ and d_{NN} and the *local* conformation in the principal regular secondary structures found in globular proteins. Correspondingly, using the search parameters $d_{\alpha N} \leqslant 2.6$ Å or $d_{NN} \leqslant 3.6$ Å for residues with the local conformations typical of a β-sheet or an α-helix, respectively, β-sheets and helical structures can be identified to an extent of nearly 100% (Tables 1 and 2). This result was obtained from a statistical study[14] of a group of 19 protein structures from the Brookhaven protein data bank containing 3227 amino acid residues. These protein structures had all been determined at a resolution of at least 2.0 Å. For an objective and consistent identification of secondary structure elements in these proteins, we used the criteria of Kabsch and Sander[21]. Overall, of the total of 3227 residues 796 were assigned to helical secondary structure, 767 to β-sheets, 818 to turns and 846 to random coil segments.

Table 1. Extent and uniqueness of the identification of helical secondary structure by several subsequent sequential distance constraints $d_{NN} \leqslant 3.6$ Å.

Length of segment*	Extent**	Uniqueness†
1	98	51
3	97	68
5	91	78
7	85	80

*Indicates the number of subsequent short distances $d_{NN} \leqslant 3.6$ Å.
**Indicates the percentage of the total number of helical residues that are recognized by the specified segment.
†Indicates the percentage of the residues recognized by the specified segment that are actually located in helices.

Table 2. Extent and uniqueness of the identification of β-sheet secondary structure by several subsequent sequential distance constraints $d_{\alpha N} \leqslant 2.6$ Å.

Length of segment*	Extent**	Uniqueness†
1	95	46
3	90	55
5	79	63
7	59	65

*Indicates the number of subsequent short distances $d_{\alpha N} \leqslant 2.6$ Å.
**Indicates the percentage of the total number of β-sheet residues that are recognized by the specified segment.
†Indicates the percentage of the residues recognized by the specified segment that are actually located in β-sheets.

In contrast to the high *extent* of β-sheet or helix identification by $d_{\alpha N} \leqslant 2.6$ Å or $d_{NN} \leqslant 3.6$ Å, respectively, the *uniqueness* of identification of these secondary structures is only of the order of 50% or less (*extent* and *uniqueness* are defined in the footnotes to the tables). Inspection of the Ramachandran plot in Figure 9 provided the clue for a qualitative explanation of this observation: Since the sterically allowable ϕ–ψ combinations for nearly all residues in a protein are in either of the two regions A or B of the ϕ–ψ plane, the *local* conformations of *all* residues are either near that characteristic of β-sheet residues or near that for α-helix residues. Hence, with the search criteria used, the majority of the residues in turns or irregular 'random coil' polypeptide segments were also recognized, which led to the low uniqueness for secondary structure identification (Tables 1 and 2). In view of this situation, we further investigated how regular secondary structures could be identified by segments of successive short distances $d_{\alpha N}$ or d_{NN}. The results of this study (Tables 1 and 2) show that, as one would expect, the extent of secondary structure identification decreases and uniqueness increases, with increasing length of such segments. For identification of helical structures, segments of three to five successive distance constraints $d_{NN} \leqslant 3.6$ Å give excellent results. For β-structures, an optimal compromise for satisfactory extent and uniqueness is obtained with segments of three to five distances $d_{\alpha N} \leqslant 2.6$ Å. The uniqueness is lower for β-sheets than for helical secondary structure, because all regular extended polypeptide segments are not part of a double- or multiple-stranded β-sheet.

In today's practice, the pattern recognition approach for secondary structure determination (bottom left of Figure 1; Figure 4) is an attractive complement to the determination of complete NMR structures of proteins (bottom right of Figure 1; Figure 5). Its main advantages are that it can largely be based on the data accumulated for obtaining the NMR assignments (a complete structure determination requires extensive

additional work of both data collection and structure calculations), and that its potential for studies of protein folding intermediates seems to be unequalled by any other presently available method (e.g., refs. 22, 23). When combined with medium-range NOE distance constraints, coupling constants $^3J_{HN\alpha}$ and amide proton exchange rates in helices, or with long-range NOEs, $^3J_{HN\alpha}$ and amide proton exchange rates in β-sheets, segments of successive short sequential constraints d_{NN} (Table 1) or $d_{\alpha N}$ (Table 2), respectively, represent an important part of the experimental parameters that make secondary polypeptide structure determination by NMR reliable, and usually actually largely overdetermined[1, 14].

Epilogue

During the first visit to India in 1974, I also had the great privilege of meeting with Professor G. N. Ramachandran. I had written to Prof. Ramachandran about my interest to visit his laboratory, and in return was offered to present a seminar at his famous Molecular Biophysics Unit at the Indian Institute of Science in Bangalore. Following my diary, the visit was arranged for 7 January 1974. After my seminar on *Non-Planar Peptide Bonds*, Prof. Ramachandran asked me to his office for several hours of discussions on polypeptide conformation (I remember that some noisy workmen had to stop their activities in order not to interfere with our concentration on science), and in the evening invited my wife and me to his house to have dinner with his family. It was a great day for the very junior scientist I was at the time!

It is a great privilege again today to have been asked to contribute to a special issue of *Current Science* honouring Prof. Ramachandran. I do hope that I have succeeded with my article to convey the message that Professor Ramachandran's seminal contributions to molecular biophysics and structural biology continue to have a profound impact on current developments in these fields.

1. Wüthrich, K., *NMR of Proteins and Nucleic Acids*, 1986, Wiley, New York.
2. Wüthrich, K., *Science*, 1989, **243**, 45.
3. Solomon, I., *Phys. Rev.*, 1955, **99**, 559.
4. Hull, W. E. and Sykes, B. D., *J. Chem. Phys.*, 1975, **63**, 867.
5. Kalk, A. and Berendsen, H. J. C., *J. Magn. Reson.*, 1976, **24**, 343.
6. Gordon, S. L. and Wüthrich, K., *J. Am. Chem. Soc.*, 1978, **100**, 7094.
7. Wagner, G. and Wüthrich, K., *J. Magn. Reson.*, 1979, **33**, 675.
8. Anil Kumar, Wagner, G., Ernst, R. R. and Wüthrich, K., *J. Am. Chem. Soc.*, 1981, **103**, 3654.
9. Wüthrich, K., Wider, G., Wagner, G. and Braun, W., *J. Mol. Biol.*, 1982, **155**, 311.
10. Billeter, M., Braun, W. and Wüthrich, K., *J. Mol. Biol.*, 1982, **155**, 321.
11. Wagner, G. and Wüthrich, K., *J. Mol. Biol.*, 1982, **155**, 347.
12. Wider, G., Lee, K. H. and Wüthrich, K., *J. Mol. Biol.*, 1982, **155**, 367.
13. Green, D. W., Ingram, V. M. and Perutz, M. F., *Proc. R. Soc.*, 1954, **A225**, 287.
14. Wüthrich, K., Billeter, M. and Braun, W., *J. Mol. Biol.*, 1984, **180**, 715.
15. Braun, W., *Quart. Rev. Biophys.*, 1987, **19**, 115.
16. Ernst, R. R., Bodenhausen, G. and Wokaun, A., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
17. Wüthrich, K., *Acc. Chem. Res.*, 1989, **22**, 36.
18. Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V., *J. Mol. Biol.*, 1963, **7**, 95.
19. Ramachandran, G. N. and Sasisekharan, V., *Adv. Protein Chem.*, 1968, **23**, 281.
20. Momany, F. A., McGuire, R. F., Burgess, A. W. and Scheraga, H. A., *J. Phys. Chem.*, 1975, **79**, 2361.
21. Kabsch, W. and Sander, C., *Biopolymers*, 1983, **22**, 2577.
22. Oas, T. G. and Kim, P. S., *Nature*, 1988, **336**, 42.
23. Dyson, H. J., Rance, M., Houghten, R. A., Wright, P. E. and Lerner, R. A., *J. Mol. Biol.*, 1988, **201**, 201.
24. Kline, A. D. and Wüthrich, K., *J. Mol. Biol.*, 1985, **183**, 503.
25. Kline, A. D., Braun, W. and Wüthrich, K., *J. Mol. Biol.*, 1988, **204**, 675.

ACKNOWLEDGEMENT. The author's research is supported by the Schweizerischer Nationalfonds (project Nr. 31.25174.88).