Volume 285, number 2, 237–247 FEBS 10004 July 1991

*ADONIS* 001457939100704H

*Research Letters*

# Protein secondary structure determination by NMR

## Application with recombinant human cyclophilin

Kurt Wüthrich[1], Claus Spitzfaden[1], Klaus Memmert[2], Hans Widmer[2] and Gerhard Wider[1]

[1]*Institut für Molekularbiologie und Biophysik, ETH Hönggerberg, CH-8093 Zürich, Switzerland* and [2]*Präklinische Forschung, Sandoz Pharma AG, CH-4002 Basel, Switzerland*

Received 27 May 1991

It is a unique trait of the NMR method for protein structure determination that a description of the polypeptide secondary structure can be obtained at an early stage and quite independently of the complete structure calculation. In this paper the procedures used for secondary structure determination are reviewed and placed in perspective relative to the other steps in a complete three-dimensional structure determination. As an illustration the identification of the regular secondary structure elements in human cyclophilin is described.

Protein structure; Secondary structure; Nuclear magnetic resonance; NMR structure; Cyclophilin

## 1. INTRODUCTION

Protein structure determination based on the use of nuclear magnetic resonance (NMR) spectroscopy in solution for the collection of experimental conformational constraints [1–3] has by now been widely accepted as a second method for studies at atomic resolution, in addition to diffraction techniques with protein crystals. For both methods the typical final result consists of a complete description of the three-dimensional structure (Fig. 1). It is a special feature of the NMR method, however, that it can provide a quite precise characterization of the polypeptide secondary structure at an early stage of a structure determination [1,5–7], i.e. before an exhaustive collection of conformational constraints as input for the structure calculation. During the early years of protein structure determination by NMR, reports on the complete structure determination were therefore always preceded by a description of the secondary structure (e.g. [5,8–11]). In recent studies of larger proteins, determination of the secondary structure or the complete three-dimensional structure have again become two clearly different steps, indicating that perhaps the time has come to recall the fundamentals of secondary polypeptide structure identification by NMR. To place this special aspect in proper perspective, the review starts with a brief survey of protein structure determination by NMR. This is followed by a survey of strategy and practical aspects of secondary structure determination. An illustration is provided by the description of the so far unpublished secondary structure determination of recombinant human cyclophilin.

*Correspondence address:* K. Wüthrich, Institut für Molekularbiologie und Biophysik, ETH Hönggerberg, CH-8093 Zürich, Switzerland.

## 2. PROTEIN STRUCTURE DETERMINATION BY NMR

The left part of Fig. 2 outlines the course of a complete three-dimensional protein structure determination by NMR [1–3]. Sucessive steps involved in this procedure are the preparation of the protein solution, the NMR measurements, obtaining sequence-specific assignments of the NMR lines [12], and the collection of conformational constraints to be used as input for the structure calculations with distance geometry or molecular dynamics techniques. The heavily framed box on the right indicates that at a quite early stage along this avenue, one has the option of applying an empirical pattern recognition approach with a limited set of conformational constraints for identifying the secondary polypeptide structure in the protein [6,7]. As indicated by the dashed arrows in Fig. 2, the secondary structure identification may also provide support for the resonance assingments and the data collection needed for the determination of the complete three-dimensional structure. The scheme of Fig. 2 is largely self-explanatory, and the additional comments in this and the following section should primarily clarify why



*Published by Elsevier Science Publishers B.V.*

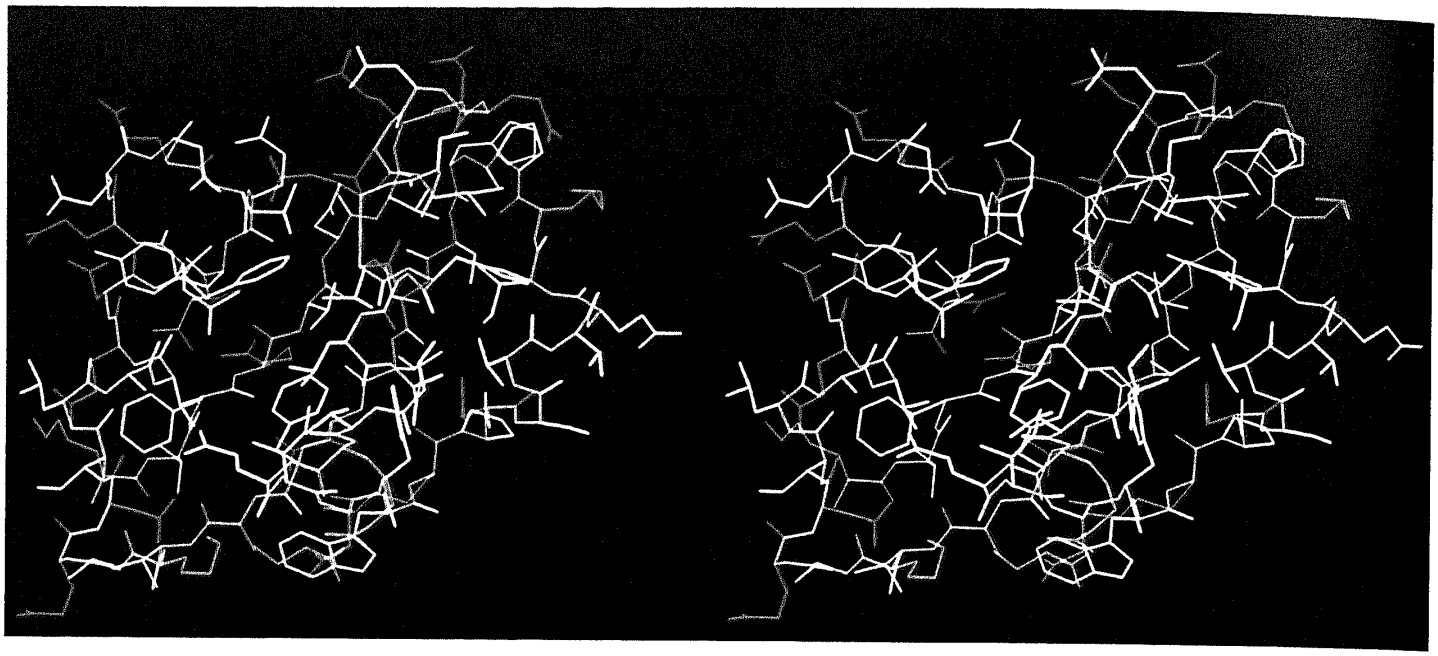

Fig. 1. Stereo view of the NMR solution structure of the activation domain from porcine procarboxypeptidase B. All heavy atoms are shown. Color code: backbone magenta; side chains: acidic red, basic blue, hydrophobic yellow, polar white. (The cover illustration shows the same molecule after a 180° rotation about a vertical axis. The structure determination is described in [4]).

the identification of the polypeptide secondary structure in a protein can be an attractive use of NMR techniques in its own right.

### 2.1. *Sample preparation*

The protein is usually dissolved in 0.5 ml of water, and the ionic strength, the pH and the temperature may be adjusted so as to ensure near-physiological conditions (it is advantageous to be able to work in the slightly acidic pH range from 3 to 5 [1]). The protein concentration should be at least 1 mM, ideally 3–6 mM, so that 15–30 mg of a protein with molecular weight 10 000 should be available for a structure determination. The selection of these experimental parameters is guided by exploratory experiments using circular dichroism and one-dimensional (1D) NMR experiments to determine the solution conditions under which the native form of the protein is stable. So far, structure determinations by NMR have been reported for proteins and protein–DNA complexes with molecular weights up to approximately 18 000. In the future this upper size limit may perhaps be raised to about 30 000. For molecular sizes above approximately 12 000 the NMR study has to include the preparation of protein enriched with $^{15}N$ and/or $^{13}C$, which is best achieved with biosynthetic techniques. As is discussed in more detail in the following, labeling with $^{15}N$ is often sufficient for the identification of the secondary structure, whereas the more costly $^{13}C$ labeling may be indispensable for the determination of the complete three-dimensional structure of the same protein (e.g. [13,14]).

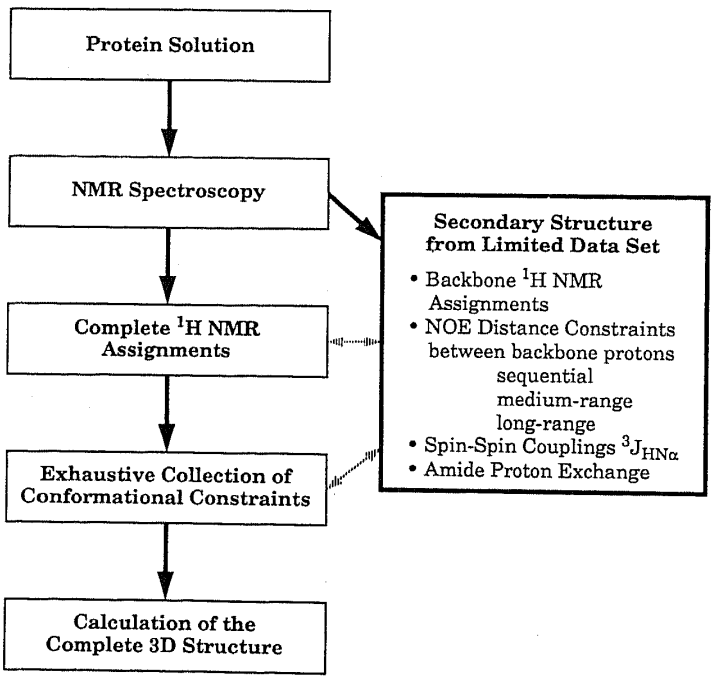


Fig. 2. Diagrammatic presentation of structural studies of proteins by NMR. In the left part of the diagram the determination of the complete three-dimensional structure is presented as a succession of five steps. The box on the right describes the polypeptide secondary structure determination. The position of this box and the arrows connecting it to the left part indicate the relations between the two procedures (see text for further details).

### 2.2. *NMR measurements*

Because of the large number of hydrogen atoms in a protein, a one-dimensional $^1H$ NMR spectrum is crowded with mutually overlapping lines. Therefore, two-dimensional (2D) and higher-dimensional NMR experiments are used [15]. The key experiment is homonuclear 2D $^1H$–$^1H$ nuclear Overhauser effect (NOE) spectroscopy (NOESY) [16]. Suitably executed

NOE measurements, which allow for the fact that NOEs are built up over a time range of milliseconds to seconds [17–19] and that effects of spin diffusion have to be taken into consideration [20,21], can provide information on through-space proton–proton distances in the range from about 2.0 to 5.0 Å [1–3]. Additional homonuclear and heteronuclear 2D NMR experiments provide information on scalar, through-bond relations between nuclear spins located near each other in the covalent polypeptide structure. Examples are correlation spectroscopy (COSY) and total correlation spectroscopy [1,15]. Both homonuclear [22,23] and heteronuclear [24,25] three-dimensional (3D) NMR experiments with further improved spectral resolution can be obtained through suitable combination of two 2D NMR experimental schemes, and this principle can also be extended to higher dimensions [26]. The same types of NMR experiments are used in all structural investigations of proteins, but the final selection will in each case be based on the molecule studied, in particular on its size. Invariably, a less extensive array of experiments is needed for a secondary structure determination, and for the individual measurements the selection of the experimental parameters may be less stringent than for work on the complete three-dimensional structure.

### 2.3. *Resonance assignments*

As proteins contain multiple units of the individual amino acids, spectral assignments are non-trivial. The problem was solved with the *sequential assignment strategy* [1,27–30]. This strategy can rely entirely on the use of homonuclear $^1$H NMR experiments and knowledge of the amino acid sequence, but for molecular sizes above approximately 12 000 Da it usually needs to be supported with heteronuclear NMR experiments using isotope-labeled protein. Sequence-specific $^1$H NMR assignments are the basis of both, secondary structure identification and determination of the complete three-dimensional protein structure [12]. However, although a secondary structure identification can be based on assignments for the polypeptide backbone protons and possibly some $C^\beta$ protons, complete assignments including the amino acid side chains are needed for a complete structure determination [1]. Obtaining complete assignments is laborious and, as was mentioned earlier, for bigger proteins different, more costly isotope-labeling strategies may be needed than for the backbone assignments alone [13,14].

### 2.4. *Protein structures from NMR data*

For the determination of the complete three-dimensional protein structure the maximum possible number of conformational constraints must be collected from the NMR spectra. This is achieved by addition of stereospecific assignments [31,32] to the complete resonance assignments for the polypeptide backbone and the amino acid side chains, complete analysis of the [$^1$H,$^1$H]-NOESY spectra for obtaining a large set of distance constraints, and collection of supplementary conformational constraints that can be obtained, for example, from measurements of spin-spin coupling constants [1]. These data then represent the input for the structure calculations [1,33] leading to a result such as that shown in Fig. 1 [4]. Nowadays, such a structure determination relies typically on an extensive infrastructure ensuring computer-supported spectral analysis and book-keeping in addition to the actual structure calculations (e.g. [35]). In contrast, although a secondary structure identification may benefit from suitable computer support, it can be achieved by visual inspection of the NMR spectra even for relatively large proteins, since only a limited amount of data needs to be surveyed [1,6,7].

## 3. BENEFITS FROM A PROTEIN SECONDARY STRUCTURE DETERMINATION

In an assessment of the significance to be attributed to a secondary structure determination by NMR it must be emphasized again that compared to a complete structure determination, the amount of work needed is small (see section 2). Furthermore, NMR is unique in providing such information early in the course of a structure determination.

With regard to a subsequent determination of the three-dimensional structure by NMR, the identification of regular secondary structure elements can be used for obtaining supplementary conformational constraints which may, for example, enforce the formation of hydrogen bonds [1,36]. Because different ways are used for the two levels of structure determination (Fig. 2), a comparison of the results obtained represents an important check of the complete structure calculated from the NMR data.

Knowledge of the polypeptide secondary structure may be helpful in the initial chain tracing of an X-ray crystal structure determination. This was the case, for example, with human cyclophilin (see section 5; unpublished results).

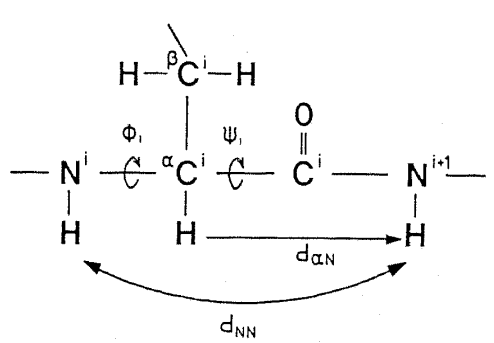


Fig. 3. Polypeptide segment consisting of the amino acid residue i and the NH group of the residue (i + 1). The dihedral angles $\phi_i$ and $\psi_i$ and the sequential distances $d_{\alpha N}$ and $d_{NN}$ are indicated.

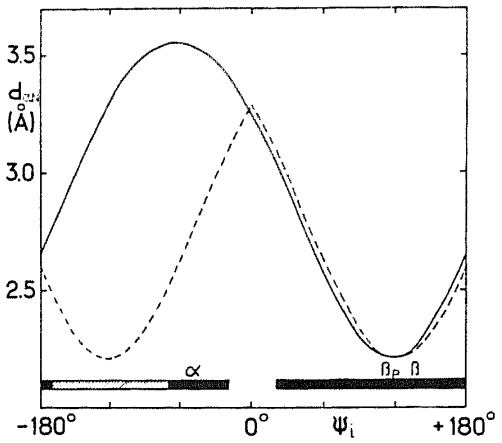


Fig. 4. Plot of the sequential distance $d_{\alpha N}$ between $C^{\alpha}H_i$ and $NH_{i+1}$ (see Fig. 3) vs the dihedral angle $\psi_i$. (—) common L-amino acid in position i. (----) Gly in position i, where for each value of $\psi_i$ the smaller one of the two distances between $NH_{i+1}$ and the two $\alpha$-protons of Gly is displayed. The range of sterically allowed $\psi_i$ values for the common L-amino acids is indicated by the black bar at the bottom of the figure. For glycine, the $\psi_i$ values indicated by the hatched bar are also allowed. The letters $\alpha$, $\beta_p$ and $\beta$ identify, respectively, the $\psi_i$ values for the $\alpha$-helical residues, parallel $\beta$-sheets, and antiparallel $\beta$-sheets. (reproduced from [1])

If some additional experimental constraints are available, the secondary structure may be used as a starting point for interactive model building of the tertiary structure [37,38]. It would come as no big surprise if the combined use of secondary structure determination by NMR and molecular modeling were to gain added interest in the future.

## 4. POLYPEPTIDE SECONDARY STRUCTURE IDENTIFICATION BY NMR

The most important data for secondary structure identification are NOE-derived distance constraints between different polypeptide backbone protons (Fig. 2). A statistical analysis of a representative sample of high resolution crystal structures of globular proteins showed that only a very limited set of such proton-proton distances can adopt short values, which then produce patterns that are characteristic of different secondary structures [1,6]. Supplementary data can be obtained from the amide proton–$C^{\alpha}$ proton coupling constants, $^3J_{HN\alpha}$ [7], and from measurements of amide proton exchange rates [1].

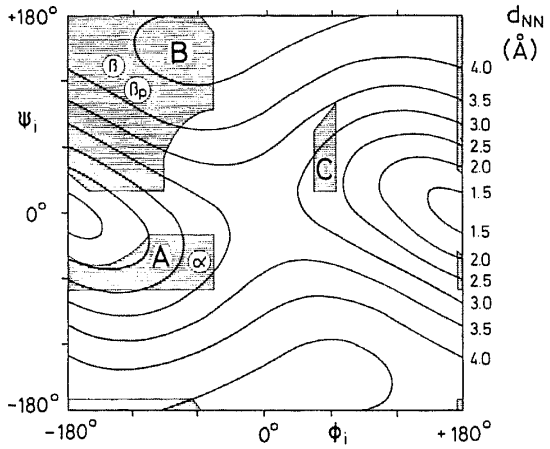


Fig. 5. Contour lines for fixed values of the sequential distance $d_{NN}$ between $NH_i$ and $NH_{i+1}$ (see Fig. 3) in the $\phi_i$–$\psi_i$ plane. The $d_{NN}$ values for the individual contour lines are given on the right. The sterically allowed areas in the $\phi_i$–$\psi_i$ plane (calculated for Ala) are hatched and labelled A,B and C. The $\phi_i$–$\psi_i$ combinations for $\alpha$-helical residues, parallel $\beta$-sheets, and antiparallel $\beta$-sheets are indicated by $\alpha$,$\beta_p$ and $\beta$, respectively. (reproduced from [1])

### 4.1. *Sequential NOEs*

The sequential distances $d_{\alpha N} \equiv d_{\alpha N}(i,i+1)$ and $d_{NN} \equiv d_{NN}(i,i+1)$ (see [1] for the notation used) are used in the early phases of a protein structure determination for obtaining sequential resonance assignments. Figs 3-5 show that these distances are related to the dihedral angles $\phi_i$ and $\psi_i$, which define the local conformation of the polypeptide chain. Specifically, $d_{\alpha N}$ and $d_{NN}$ can be used to discriminate between $\alpha$- and $\beta$-type local conformations, with $d_{\alpha N}$ = 2.2 Å in $\beta$ sheets and 3.4 Å in $\alpha$-helices (Fig. 4), and $d_{NN}$ = 4.3 Å in $\beta$ sheets and 2.8 Å in $\alpha$-helices (Fig. 5). However, because the sterically allowable $\phi$–$\psi$ combinations for nearly all residues in a protein are in either of the two regions A or B of the $\phi$–$\psi$ plane (Fig. 5), the local conformations of *all* residues are either near that characteristic of a $\beta$-sheet residue or near an $\alpha$-helix residue. Therefore, the suitability of a given distance parameter for secondary structure identification must be judged with two criteria, the *extent* and the *uniqueness* with which helical residues or $\beta$-sheet residues are recognized (Table I). Taking $d_{NN}$ as an illustration, Table I shows that 98% of all residues located in a helix have $d_{NN} \leq$ 3.6, but that only 51% of all the residues with $d_{NN} \leq$ 3.6 are located in a helix, whereas 49% must come from turns or 'random coil' polypeptide segments. Improved uniqueness is obtained if trains of several successive short distances $d_{NN}$ are used as a probe, but, as one would expect, the extent of secondary structure iden-

Table I

Extent and uniqueness of the identification of helical secondary structure by one or several subsequent sequential distance constraints $d_{NN} \leq$ 3.6 Å (data obtained from a statistical analysis of a group of globular protein structures, see [6])

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

+ Indicates the number of subsequent short distances $d_{NN} \leq$ 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.

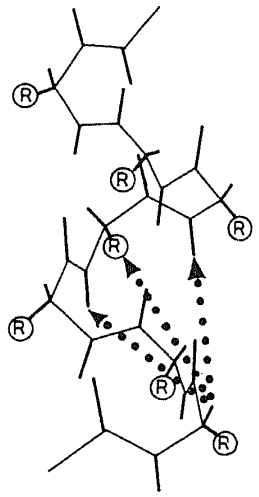

Fig. 6. α-Helix with the medium range $^1$H-$^1$H distances $d_{\alpha N}(i,i+3)$, $d_{\alpha\beta}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ indicated with dotted arrows. The bonds between all backbone atoms are drawn, and the side chains are indicated by ®.

tification is then somewhat decreased (Table I). A reasonable compromise for satisfactory extent and uniqueness of the identification of helical structures is obtained with segments of three to five successive distance constraints $d_{NN} \leq 3.6$ Å. For β-structures, segments of three to five distances $d_{\alpha N} \leq 2.6$ Å give the best results [6].

### 4.2. *Spin–spin coupling constants* $^3J_{HN\alpha}$

The situation is similar to that with the sequential distances $d_{\alpha N}$ and $d_{NN}$. Helical residues have $^3J_{HN\alpha} \approx 4$ Hz, and for β-sheet residues $^3J_{HN\alpha} \approx 9$ Hz, but a satisfactory uniqueness of the identification of helical and β-sheet secondary structure is obtained only with the use of segments of several successive small or large coupling constants, respectively. Segments of 3 to 5 successive $^3J_{HN\alpha}$ values ≤ 6 Hz for helices and 3 to 5 values ≥ 7 Hz for β-sheets give the best results [7].

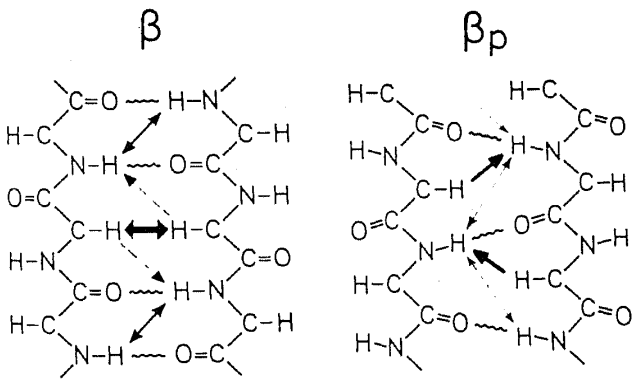


Fig. 7. Short interstrand, long-range $^1$H-$^1$H distances in β-sheets. Wavy lines indicate interstrand hydrogen bonds. All backbone atoms are drawn, the side chains have been omitted. For antiparallel β-structures, short interstrand distances are indicated by thick horizontal arrows for $d_{\alpha\alpha}(i,j)$, thin solid arrows for $d_{NN}(i,j)$, and broken arrows for $d_{\alpha N}(i,j)$. In parallel β-structures, solid arrows indicate $d_{\alpha N}(i,j)$ and broken arrows represent $d_{NN}(i,j)$.

### 4.3. *Medium-range backbone NOEs*

These are NOEs corresponding to distances between polypeptide backbone protons that are located in residues separated by 2–4 positions in the sequence [1]. The distance $d_{\alpha N}(i,i+3)$ (with some limitations also $d_{NN}(i,i+2)$ and $d_{\alpha\beta}(i,i+3)$) is particularly useful, since short values occur nearly exclusively in helices (Fig. 6), i.e. the uniqueness of individual medium-range NOEs for identification of helical structures is significantly higher than for individual sequential NOEs (see Table I), and the uniqueness is near 100% for a succession of 2 to 4 of these medium-range NOEs. Furthermore, additional short values for $d_{\alpha N}(i,i+4)$ are found only in α-helices, whereas in $3_{10}$-helices additional NOEs corresponding to $d_{\alpha N}(i,i+2)$ are expected [6].

### 4.4. *Long-range backbone NOEs*

These are observed between polypeptide backbone protons located in neighboring strands in β-sheets (Fig. 7) and usually relate protons in residues that are at least 5 positions apart in the sequence. Their observation is indispensable for the identification of β-sheets, since the aforementioned analysis of sequential NOEs and coupling constants $^3J_{HN\alpha}$ can only result in the identification of individual extended polypeptide segments [6]. Since in principle only two interstrand NOEs are needed for proper relative positioning of two extended polypeptide segments, this structural feature is usually well determined by the experimental data. As a result, irregularities in β-sheets, such as β-bulges, can usually also be identified (e.g. [9,10]). Of particular interest is the identification of short distances $d_{\alpha\alpha}(i,j)$, since these are highly unique for antiparallel β-sheets (Fig. 7).

### 4.5. *Amide proton exchange rates*

As a rule, for amide protons involved in hydrogen bonds with backbone carbonyl groups the exchange with the bulk solvent is several orders of magnitude slower than for solvent-accessible amide protons. Although protection from solvent contact by other features of the three-dimensional structure than the hydrogen bonds may also occur [39], experience has shown that hydrogen bonds in regular secondary structures are reliably manifested by slow amide proton exchange rates [8,10,11,40]. As a consequence, different patterns of slow and rapid exchange rates along the sequence are found to be typical for helices, peripheral β-strands and interior strands in β-sheets [1].

Each individual one of the aforementioned criteria for identification of secondary structures usually provides a somewhat ambiguous result, mainly because spectral overlap and other technical limitations may prevent one from obtaining a complete data set. The combined information from sequential NOEs, spin–spin couplings $^3J_{HN\alpha}$, medium-range and long-range backbone NOEs, and amide proton exchange rates, however, typically results in an overdetermina-

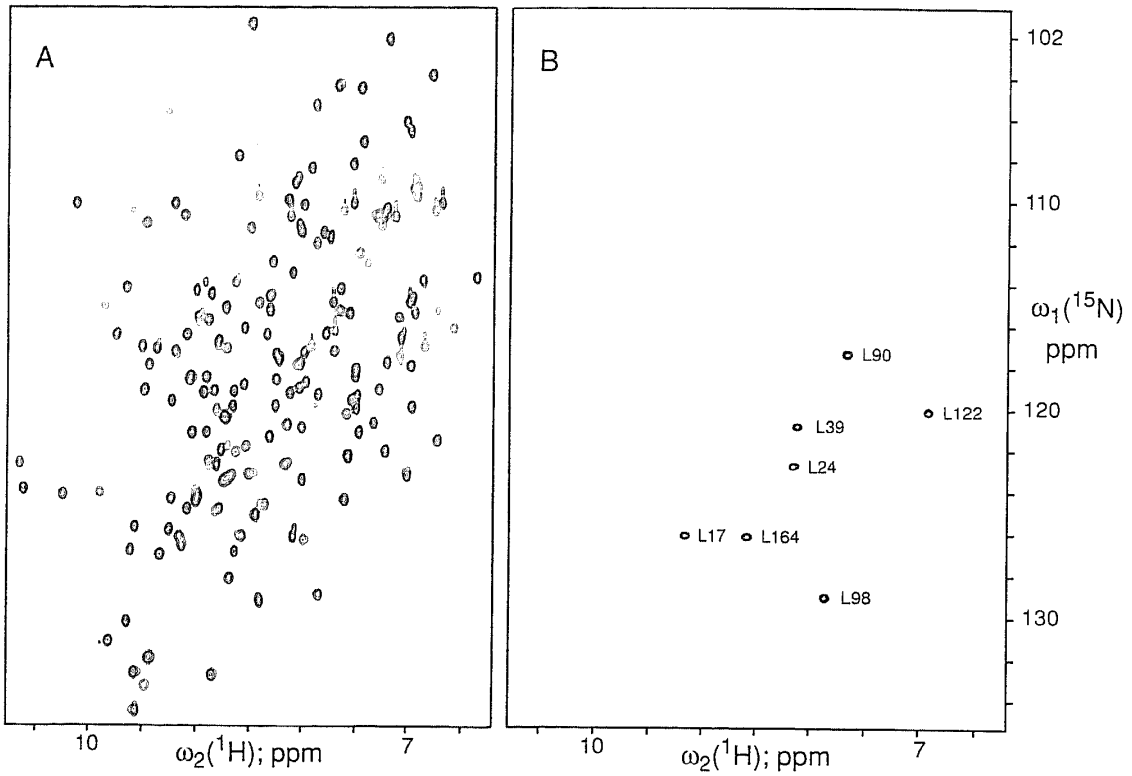


Fig. 8. 2D [$^{15}$N,$^1$H]-COSY spectra (proton frequency 600 MHz, protein concentration 4 mM, solvent 90% $H_2O$ and 10% $D_2O$, pH = 6.0, 26°C). (A) Uniformly $^{15}$N-labeled cyclophilin. (B) Cyclophilin with residue-specific $^{15}$N-labeling of the seven leucyl residues. The peaks are identified by the one-letter amino acid symbol and the sequence position.

tion of the regular secondary structures, and enables one to determine additional features, such as tight turns and β-bulges. As a rule, regular secondary structures extending over 6 or more residues are most reliably characterized (for example, a single short distance $d_{\alpha N}(i,i+3)$ occurs mainly in helices but may also occur in tight turns, whereas a succession of three such short distances is possible exclusively in helical structures). In the following section 5 this is illustrated with the protein cyclophilin.

## 5. THE SECONDARY STRUCTURE OF RECOMBINANT HUMAN CYCLOPHILIN

Cyclophilin is a water-soluble cytosolic protein believed to function as the major cellular receptor of the immunosuppressant cyclosporin A [41,42], and in addition has been found to be active as a proline *cis-trans* isomerase [43,44]. It consists of a polypeptide chain with 165 amino acid residues. Although reports on partial resonance assignments by homonuclear $^1$H NMR experiments have appeared [45], our experience with this protein indicated that isotope-labeling was needed for obtaining sequence-specific resonance assignments. The use of $^{15}$N-labeling described here resulted on the one hand in the identification of several amino acid types to obtain a sufficient number of reference points relative to the amino acid sequence [1], and on the other hand in improved spectral resolution enabling the observation of the sequential $^1$H–$^1$H NOEs.

### 5.1. *Identification of amino acid types*

Because of spectral overlap and peak cancellations only a small number of amino acid spin systems could be unambiguously identified from the analysis of homonuclear 2D $^1$H NMR experiments, and a few of these spin system identifications were at variance with those presented by others [45]. Therefore we decided to base the attribution of individual backbone proton and $^{15}$N resonances to specific ones of the 20 proteinogenic amino acid types mainly on residue-specific $^{15}$N-labeling [46]. Four protein samples were prepared from *E. coli* bacteria grown on amino acid media containing one of the amino acids in the $^{15}$N-labeled form, i.e. Gly (23 residues in human cyclophilin), Leu (7), Lys (14) and Val (9). Thanks to selective cross-labeling effects

Table II

$^{15}$N-labeling of recombinant human cyclophilin by amino acid type

| $^{15}$N-labeled amino acid in the nutrient medium* | $^{15}$N-labels in the protein*† | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Gly | Cys | Ser | Val | Leu | Lys | Ala | Ile |
| Gly | ++ | ++ | ++ | | | | | |
| Val | | | | ++ | ++ | | + | + |
| Leu | | | | | ++ | | | |
| Lys | | | | ++ | | ++ | + | |

*Four experiments with different $^{15}$N-labeled amino acids in the nutrient medium were performed. Selective cross-labeling to other amino acids is due to transamination and other metabolic processes. *†(++) = level of enrichment about 30%; (+) = lower level of enrichment.

[47] in some of these experiments (see Table II), backbone proton and $^{15}N$ assignments were obtained for Ala (9), Cys (4), Ile (10) and Ser (8), in addition to the four labeled amino acid types supplied in the nutrient media. Fig. 8 illustrates how the Leu residues were identified using heteronuclear [$^{15}N$,$^{1}H$]-COSY with [1-$^{15}N$-Leu]-cyclophilin. From four residue-selective labeling experiments (Table II), $^{15}N$ and amide proton chemical shifts were identified for 84 residues. The spin system identifications were extended to $C^{\alpha}H$ (and in many cases to the amino acid side chain) using 2D [$^{1}H$,$^{1}H$]-TOCSY, and 3D $^{15}N$-correlated [$^{1}H$,$^{1}H$]-TOCSY with uniformly $^{15}N$-labeled cyclophilin.

5.2. *Sequential $^{1}H$ NMR assignments*

The standard sequential NOE connectivities $d_{\alpha N}$ and $d_{NN}$ (Fig. 3) were used to identify pairs of neighboring residues in the sequence [1,28-30]. For the measurements of these quantities, uniformly $^{15}N$-labeled cyclophilin was used to record a 3D $^{15}N$-correlated [$^{1}H$,$^{1}H$]-NOESY spectrum, such as the example shown in Fig. 9. The spectral analysis was performed in $\omega_1$-$\omega_3$ cross planes (see Fig. 9 for the orientation of these planes). A compact presentation of the results was obtained in the form of a composite strip plot (Fig. 10) [13]: strips extending over the entire spectral range along $\omega_1$ were selected so that they extend 0.075 ppm to the left and right from each amide proton position along $\omega_3$. These strips were searched for dipeptides or longer polypeptide segments in which the individual residues are linked together by sequential connectivities. Based on these data and the amino acid identifications from residue-specific labeling (Table II), sequence-specific assignments for the polypeptide backbone protons were obtained. The $\omega_1$-$\omega_3$ strips were then reshuffled from the initial coincidential arrangement to the sequential order shown in Fig. 10. A survey of all sequential connectivities observed is afforded by Fig. 11, which shows that a nearly complete set of assignments was obtained. The only remaining gaps in the sequential assignments by $d_{\alpha N}$ and $d_{NN}$ are at the positions 1,81,115,148 and before the six prolyl residues 4,16,30,58,95 and 105, where the gap between residues 115 and 116 was closed on the basis of the two $d_{\beta N}$ NOE connectivities [1].

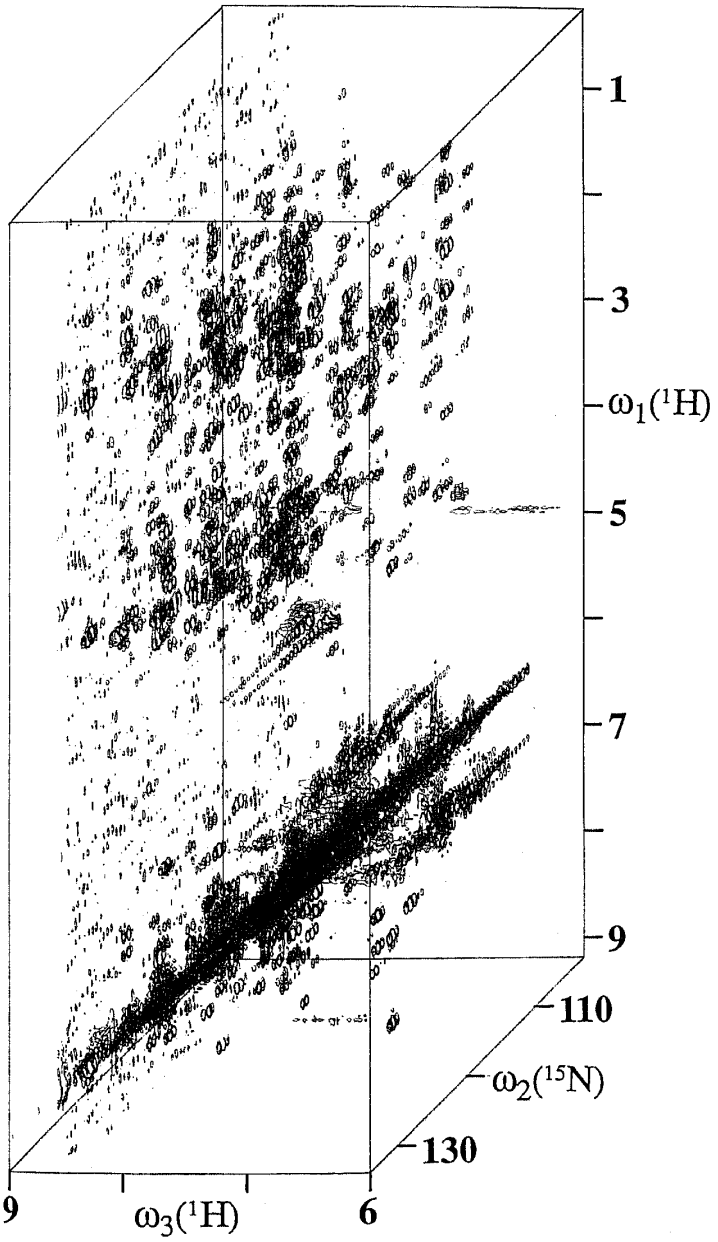


Fig. 9. Spectrtal region ($\omega_1$ = 0.3–9.1 ppm, $\omega_2$ = 103–134 ppm, $\omega_3$ = 6.0–9.0 ppm) of a 3D $^{15}N$-correlated [$^{1}H$,$^{1}H$]-NOESY spectrum of a protein (the DNA-binding domain of P22 c2 repressor).

The procedure adopted for the present study of cyclophilin [13,46,48,49] enables one to obtain sequence-specific assignments without detailed analysis of the amino acid side chain spin systems. As an alternative, main chain-directed resonance assignments using homonuclear $^{1}H$ NMR or heteronuclear experiments with uniformly labeled protein have been proposed for use in situations where the complete amino acid spin systems are difficult to identify [50,51]. When comparing the two approaches, the use of residue-selective $^{15}N$-labeling has the decisive advantage that it enables frequent checks of the sequential assignments against the independently known amino acid sequence.

5.3. *Secondary structure identification*

Initial indications of the presence of regular helical and extended, $\beta$-sheet-type polypeptide segments came from the observation of trains of successive strong sequential $d_{NN}$ or $d_{\alpha N}$ NOEs along the amino acid sequence of cyclophilin (Fig. 11; see also Figs. 4 and 5 and Table I). These data were supplemented by medium-range NOE distance constraints (Fig. 11) obtained from further analysis of the same 3D $^{15}N$-correlated [$^{1}H$,$^{1}H$]-NOESY spectrum (Fig. 10) that was used for the sequential assignments. Due to chemical shift degeneracies of $C^{\alpha}$ protons in the helical regions, few medium-range NOEs were unambiguously assigned. These pro-

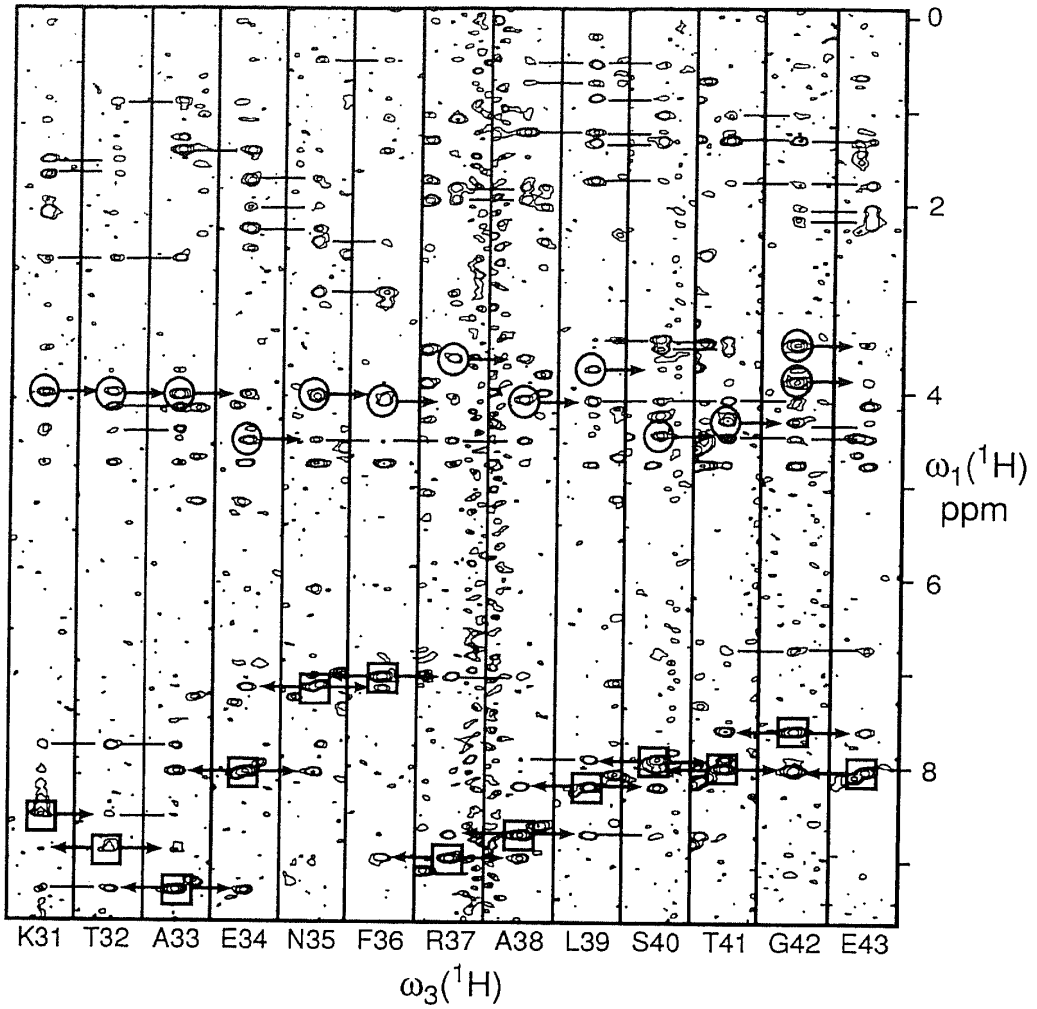


Fig. 10. Presentation of selected spectral regions of a 3D $^{15}$N-correlated [$^1$H,$^1$H]-NOESY spectrum of cyclophilin as a composite plot of strips (same experimental conditions as in Fig. 8, mixing time = 80 ms). The individual strips are taken from different $\omega_1$–$\omega_3$ planes along $\omega_2(^{15}N)$ (see Fig. 9). The individual strips are 0.15 ppm wide and are centered about the amide proton position along $\omega_3$ of the residue indicated at the bottom. The plot was obtained by arranging the amide proton strips following the sequence from Lys-31–Glu-43 (this corresponds to the first helix; see Fig. 11). Squares identify the direct $^{15}$N-$^1$H correlation peaks. (Note that the chemical shifts along $\omega_1$ and $\omega_3$ are identical; the direct peaks for Lys-31 and Thr-32 were folded along $\omega_1$ from their actual positions at 10.67 ppm and 10.31 ppm, respectively, into the spectral region shown here.) Circles are drawn around the intraresidual NH-C$^\alpha$H NOEs. Arrows represent sequential $d_{\alpha N}$ and $d_{NN}$ connectivities. Horizontal lines connect additional NOEs that are common to pairs of sequentially neighboring residues, and thus support the sequential assignment.

vide nonetheless clearcut support for the presence of two helices and a structure that could be either a short helix or an irregular turn (Fig. 11). It should be added that the occurrence of numerous additional medium-range NOEs cannot be excluded, because the locations of these NOESY cross-peaks coincide with other peaks, in particular those corresponding to intramolecular distances $d_{N\alpha}(i,i)$.

An antiparallel $\beta$-sheet with eight strands was identified, using primarily interstrand, long-range NOE connectivities $d_{\alpha N}(i,j)$ and $d_{NN}(i,j)$ [6] observed in the 3D $^{15}$N-correlated [$^1$H,$^1$H]-NOESY spectrum of Fig. 10. Eleven additional interstrand $d_{\alpha\alpha}(i,j)$ connectivities (Fig. 12) were observed in a 2D [$^1$H,$^1$H]-NOESY spectrum recorded in $D_2O$. Independent support for the $\beta$-sheet of Fig. 12 was obtained from the fact that nearly all amide protons that are expected to be hydrogen-bonded in this structure have slowed exchange rates with the bulk solvent [1,40].

### 5.4. *Description of the regular secondary structure in cyclophilin*

The cyclophilin secondary structure includes two helices of about 3.5 turns extending from residues 31–43 and 135–147. For each of these, NOEs corresponding to connectivities $d_{\alpha N}(i,i+4)$ were observed (Fig. 11), indicating standard $\alpha$-helical structure [1,6]. A third structure element including residues 120–124 may either be a short $\alpha$-helix or a helix-like succession of tight turns. The $\beta$-sheet of Fig. 12 consists of the eight strands 5–12, 15–24, 55–57, 61–64, 97–101, 112–116, 130–133, and 155–164 and contains exclusively antiparallel arrangements of neighboring strands. Two pairs of strands, 5–12 and 15–24 and 55–57 and 61–63 form hairpin structures. The regular antiparallel $\beta$-sheet arrangement is interrupted by $\beta$-bulges at positions 17 and 159. In the polypeptide segment 52–57 the residues 55–57 are clearly part of the regular $\beta$-sheet, but the available evidence for the residues 52–54 would

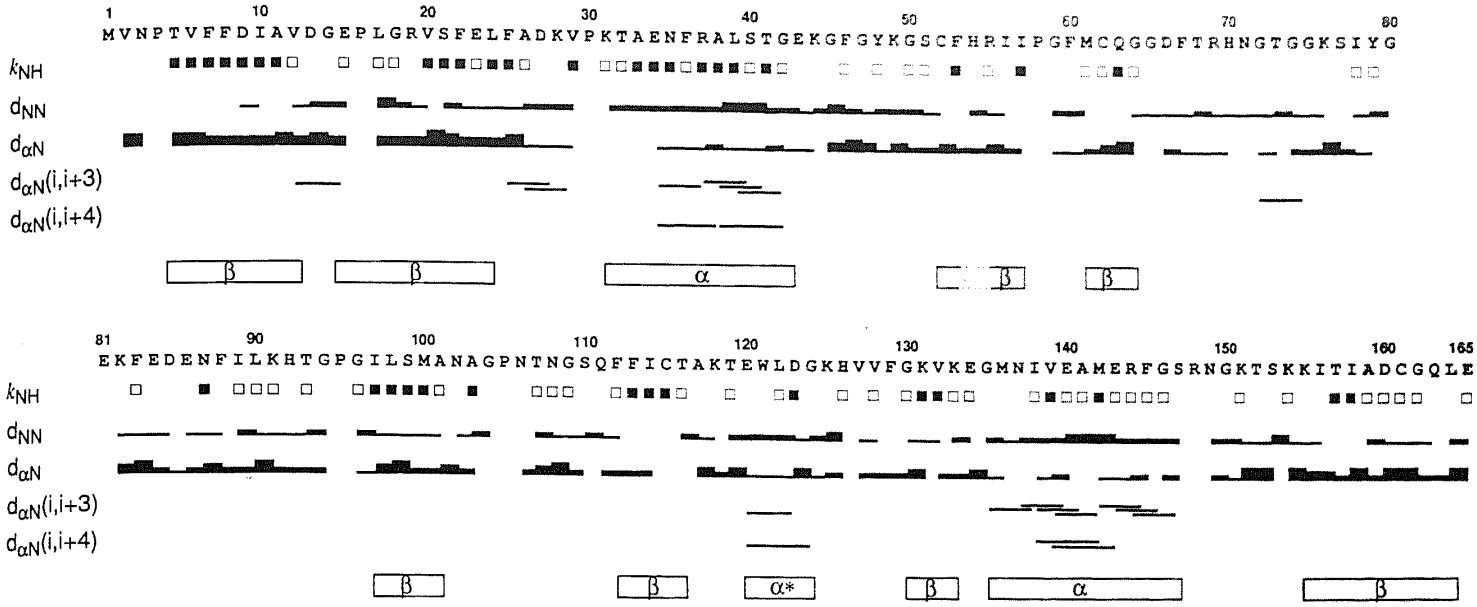


Fig. 11. Amino acid sequence of human cyclophilin, sequential and medium-range NOE connectivities, and amide proton exchange data used to identify elements of the secondary structure. Below the sequence filled and empty squares identify residues with slowed exchange of the amide proton, measured in $D_2O$ at pD 6.0 and 26°C. Open squares: [$^{15}N$,$^{1}H$]-correlation peaks were still visible after 2 h. Filled squares: peaks still visible after 48 h. The next two rows present the sequential $d_{\alpha N}$ and $d_{NN}$ connectivities, where the thickness of the bar is related to the NOE intensities. The medium-range NOE connectivities $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ are represented by lines that start and end at the positions of the two interacting residues. At the bottom, the locations of $\alpha$-helical and $\beta$-sheet segments in the polypeptide chain are identified, where $\alpha^*$ indicates that this structure element could be either a short $\alpha$-helix or a turn-like structure (see text).

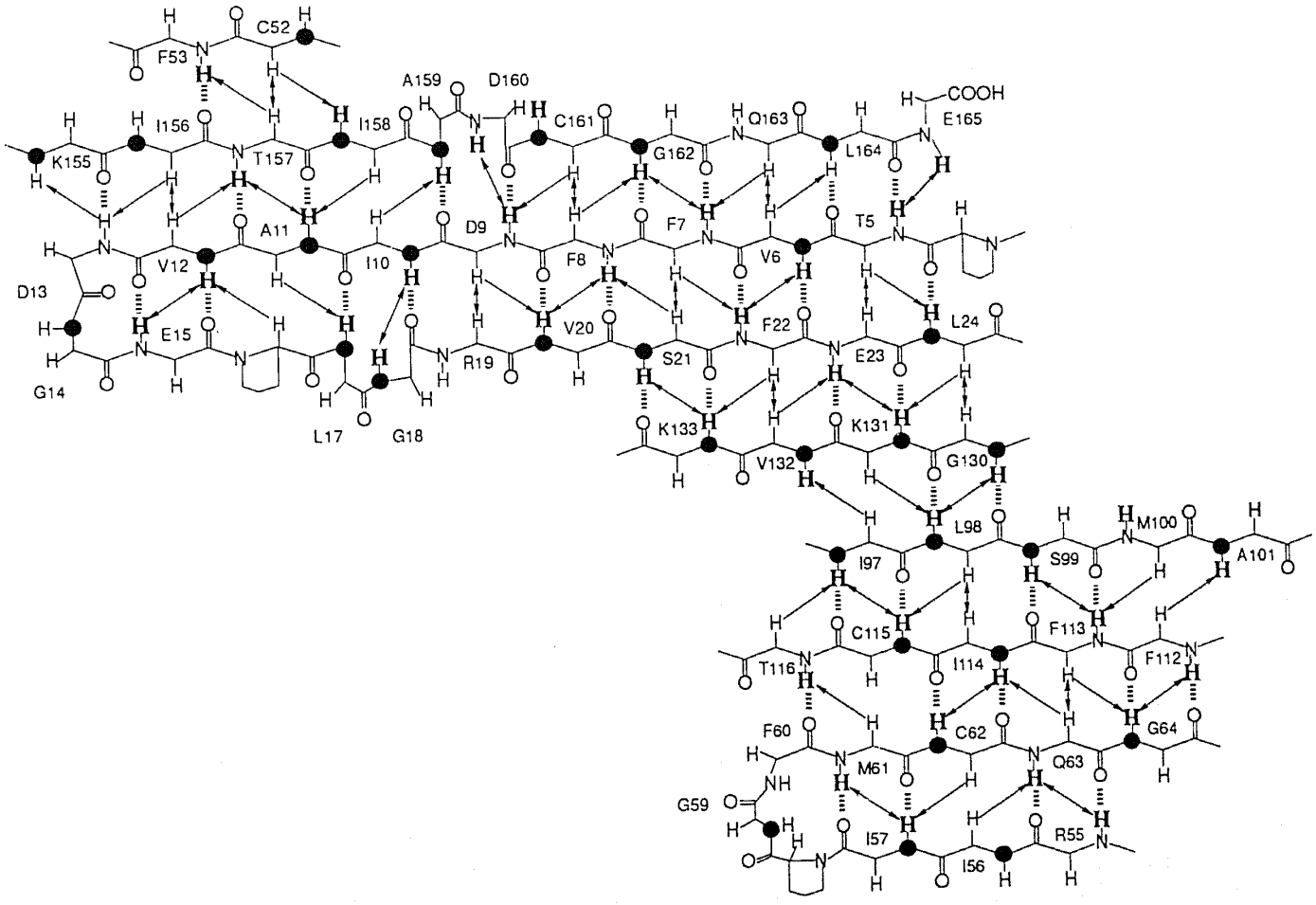


Fig. 12. Schematic representation of the $\beta$-sheet identified in recombinant human cyclophilin. All backbone atoms are shown, and except for Pro the amino acid side chains have been omitted. The individual residues are identified near the $\alpha$-carbon position by the one-letter amino acid symbol and the sequence number. Black circles at the amide nitrogen position identify those residues which were labeled with $^{15}N$ in one of the residue-specific labeling experiments (see text and Table II). Arrows represent the experimentally observed interstrand $^{1}H$-$^{1}H$ NOEs. Slowly exchanging amide protons are printed in bold-face. Hydrogen bonds that were unambiguously identified on the basis of interstrand NOEs and slowed exchange of the amide proton are identified with dashed lines.

be compatible either with β-sheet-type structure or with an irregular spatial arrangement. The possibility remains open that this strand might tie the two apparently peripheral strands in the upper left and the lower right of Fig. 12 together to form a tilted β-barrel.

In Figs 11 and 12 the individual secondary structure elements are drawn as far as they were well defined by resolved NOE cross-peaks and supported by the observation of selectively $^{15}$N-labeled residues (Fig. 12). Possibly, the individual helices and β-strands might eventually be found to extend somewhat further along the sequence. In particular, the observation of some slowly exchanging amide protons at the present periphery of the β-sheet (Figs. 11 and 12) indicates that additional regular hydrogen bonds might be present, although the corresponding interstrand NOE connectivities could not be unambiguously identified with the present analysis of the NMR data.

### 5.5. *Experimental*

Uniformly $^{15}$N-labeled recombinant human cyclophilin was prepared by growing an *E. coli* overexpression system on a glucose minimal medium containing [$^{15}$N]ammonium sulfate as the sole nitrogen source. Four different additional samples were prepared using an amino acid medium containing one amino acid type labeled with $^{15}$N [46]. The 98% $^{15}$N-labeled amino acids Gly, Lys and Val were purchased from Cambridge Isotope Laboratories, and $^{15}$N-labeled Leu was obtained from Sigma. For the protein purification and the NMR sample preparation we used a previously described procedure [52].

The conditions for the NMR measurements were: protein concentration 4 mM in a mixture of 90% $H_2O$/10% $D_2O$ containing 10 mM potassium deuteroacetate and 10 mM potassium phosphate, pH 6.0, 26°C, proton frequency 600 MHz on a Bruker AM600 spectrometer. The following experiments were used: homonuclear 2D clean-TOCSY [53] with a mixing time of 60 ms, homonuclear 2D [$^1$H,$^1$H]-NOESY [16] in $D_2O$ with a mixing time of 120 ms, heteronuclear 2D [$^{15}$N,$^1$H]-COSY [54], 3D $^{15}$N-correlated [$^1$H,$^1$H]-NOESY [24,25,55] with a mixing time of 80 ms, and 3D $^{15}$N-correlated [$^1$H,$^1$H]-TOCSY [24,25] with a mixing time of 40 ms. Water suppression was achieved with a spin-lock purge pulse [55]. The data were processed on a X32 computer using the Bruker UXNMR program and home-written software for the 3D Fourier transformation. The 3D spectra were analysed on SUN computers using the software package EASY [56]. Amide proton exchange data were obtained by equilibrating the uniformly $^{15}$N-labeled sample at pH 6.0 and 4°C against 99.8% $D_2O$ in an Amicon CP 10 ultrafiltration unit. Subsequently a series of [$^{15}$N,$^1$H]-COSY spectra was recorded at 26°C, and the time course of the intensity of the individual $^{15}$N–$^1$H cross peaks was followed.

## 6. CONCLUDING REMARKS

In as far as we can presently oversee the project, human cyclophilin will end up as a typical illustration of the situation alluded to in the *Introduction*, that the secondary structure determination and obtaining the complete three-dimensional molecular model (Fig. 1) are two clearly separated steps. In addition to the $^{15}$N-labeling described here (Table II), which enabled the determination of the regular secondary structure (Figs 11 and 12), we are currently working on the preparation of protein samples labeled with $^{13}$C. On the way toward the complete three-dimensional structure we will then have to record new NMR spectra, complete the sequential assignments of Fig. 11 with the identification of the amino acid side chain spin systems, and further follow the procedures outlined in Fig. 1.

The different aspects mentioned in section 3 that attract interest to polypeptide secondary structure determinations by NMR, all apply for cyclophilin. In addition, data corresponding to those in Figs. 11 and 12 were also obtained for a 1:1 complex of cyclophilin with a water-soluble derivative of cyclosporin A. As will be discussed elsewhere, the comparative analysis of the $^1$H and $^{15}$N chemical shifts of corresponding residues in the free and liganded form of the protein provide the basis for an outline of the intermolecular contacts in the cyclophilin-cyclosporin complex in solution.

Circular dichroism (CD) spectroscopy has long been used for studies of the secondary structure in proteins (e.g. [57]). The results obtained for cyclophilin (Figs. 11 and 12) clearly illustrate the decisive advantage of the NMR approach, which allows not only to estimate the total content of helical and β-sheet structures in a protein, but also to identify the sequence locations of the secondary structure elements.

*Acknowledgements:* The program of NMR studies with proteins at the ETH Zürich receives support from the Schweizerischer Nationalfonds (Project 31.25174.88) and from special grants of the ETH. We thank Dr. M. Billeter for help in preparing Fig. 1, Dr. B. Messerle for recording some of the $^1$H NMR Spectra, and Mr. R. Marani for the careful processing of the manuscript.

## REFERENCES

[1] Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York.
[2] Wüthrich, K. (1989) Science 243, 45–50.
[3] Wüthrich, K. (1990) J. Biol. Chem. 265, 22059–22062.
[4] Vendrell, J., Billeter, M., Wider, G., Avilés, F.X. and Wüthrich, K. (1991) EMBO J. 10, 11–15.
[5] Wagner, G., Anil Kumar and Wüthrich, K. (1981) Eur. J. Biochem. 114, 375–384.
[6] Wüthrich, K., Billeter, M. and Braun, W. (1984) J. Mol. Biol. 180, 715–740.
[7] Pardi, A., Billeter, M. and Wüthrich, K. (1984) J. Mol. Biol. 180, 741–751.
[8] Zuiderweg, E.R.P., Kaptein, R. and Wüthrich, K. (1983) Proc. Natl. Acad. Sci. USA 80, 5837–5841.

[9] Wemmer, D. and Kallenbach, N.R. (1983) Biochemistry 22, 1901–1906.
[10] Williamson, M.P., Marion, D. and Wüthrich, K. (1984) J. Mol. Biol. 173, 341–359.
[11] Kline, A.D. and Wüthrich, K. (1985) J. Mol. Biol. 183, 503–507.
[12] Wüthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) J. Mol. Biol. 155, 311–319.
[13] Driscoll, P.C., Gronenborn, A.M., Wingfield, P.T. and Clore, G.M. (1990) Biochemistry 29, 4668–4682.
[14] Clore, G.M., Kay, L.E., Bax, A. and Gronenborn, A.M. (1991) Biochemistry 30, 12–18.
[15] Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, Oxford.
[16] Anil Kumar, Ernst, R.R. and Wüthrich, K. (1980) Biochem. Biophys. Res. Commun. 95, 1–6.
[17] Solomon, I. (1955) Phys. Rev. 99, 559–565.
[18] Gordon, S.L. and Wüthrich, K. (1978) J. Am. Chem. Soc. 100, 7094–7096.
[19] Anil Kumar, Wagner, G., Ernst, R.R. and Wüthrich, K. (1981) J. Am. Chem. Soc. 103, 3654–3658.
[20] Kalk, A. and Berendsen, H.J.C. (1976) J. Magn. Reson. 24, 343–366.
[21] Hull, W.E. and Sykes, B.D. (1975) J. Chem. Phys. 63, 867–880.
[22] Griesinger, C., Sørensen, O.W. and Ernst, R.R. (1987) J. Am. Chem. Soc. 109, 7227–7228.
[23] Vuister, G.W., Boelens, R. and Kaptein, R. (1988) J. Magn. Reson. 80, 176–185.
[24] Fesik, S.W. and Zuiderweg, E.R.P. (1988) J. Magn. Reson. 78, 588–593.
[25] Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) J. Am. Chem. Soc. 111, 1515–1517.
[26] Kay, L.E., Clore, G.M., Bax, A. and Gronenborn, A.M. (1990) Science 249, 411–414.
[27] Dubs, A., Wagner, G. and Wüthrich, K. (1979) Biochim. Biophys. Acta 577, 177–194.
[28] Billeter, M., Braun, W. and Wüthrich, K. (1982) J. Mol. Biol. 155, 321–346.
[29] Wagner, G. and Wüthrich, K. (1982) J. Mol. Biol. 155, 347–366.
[30] Wider, G., Lee, K.H. and Wüthrich, K. (1982) J. Mol. Biol. 155, 367–388.
[31] Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) J. Am. Chem. Soc. 111, 3997–4004.
[32] Nilges, M., Clore, G.M. and Gronenborn, A.M. (1990) Biopolymers 29, 813–822.
[33] Braun, W. (1987) Quart. Rev. Biophys. 19, 115–157.
[34] Kuntz, I.D., Thomason, J.F. and Oshiro, C.M. (1989) Methods Enzymol. 177, 159–204.
[35] Güntert, P., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K. (1991) J. Mol. Biol. 217, 531–540.
[36] Williamson, M.P., Havel, T.F. and Wüthrich, K. (1985) J. Mol. Biol. 182, 295–315.
[37] Zuiderweg, E.R.P., Billeter, M., Boelens, R., Scheek, R.M., Wüthrich, K. and Kaptein, R. (1984) FEBS Lett. 174, 243–247.
[38] Billeter, M., Engeli, M. and Wüthrich, K. (1985) J. Mol. Graphics 3, 79–83, 97–98.
[39] Englander, S.W. and Kallenbach, N.R. (1983) Quart. Rev. Biophys. 16, 1–87.
[40] Wagner, G. and Wüthrich, K. (1982) J. Mol. Biol. 160, 343–361.
[41] Handschumacher, R.E., Harding, M.W., Rice, J., Drugge, R.J. and Speicher, D.W. (1984) Science 226, 544–547.
[42] Schreiber, S.L. (1991) Science 251, 283–287.
[43] Takahashi, N., Hayano, T. and Suzuki, M. (1989) Nature 337, 473–475.
[44] Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. and Schmid, F.X. (1989) Nature 337, 476–478.
[45] Heald, S.L., Harding, R.E., Handschumacher, R.E. and Armitage, I.M. (1990) Biochemistry 29, 4466–4478.
[46] Senn, H., Eugster, A., Otting, G., Suter, F. and Wüthrich, K. (1987) Eur. Biophys. J. 14, 301–306.
[47] Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W. (1989) Methods Enzymol. 177, 44–73.
[48] Torchia, D.A., Sparks, S.W. and Bax, A. (1989) Biochemistry 28, 5509–5524.
[49] McIntosh, L.P., Wand, A.J., Lowry, D.F., Redfield, A.G. and Dahlquist, F.W. (1990) Biochemistry 29, 6341–6362.
[50] Englander, S.W. and Wand, A.J. (1987) Biochemistry 26, 5953–5958.
[51] Di Stefano, D.L. and Wand, A.J. (1987) Biochemistry 26, 7272–7281.
[52] Weber, C., Wider, G., von Freyberg, B., Traber, R., Braun, W., Widmer, H. and Wüthrich, K. (1991) Biochemistry, in press.
[53] Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) J. Am. Chem. Soc. 110, 7870–7872.
[54] Otting, G. and Wüthrich, K. (1988) J. Magn. Reson. 76, 569–574.
[55] Messerle, B.A., Wider, G., Otting, G., Weber, C. and Wüthrich, K. (1989) J. Magn. Res. 85, 608–613.
[56] Eccles, C., Güntert, P., Billeter, M. and Wüthrich, K. (1991) J. Biomol. NMR, in press.
[57] Greenfield, N. and Fasman, G.D. (1969) Biochemistry 8, 4108–4116.