Chapter 5

THE SOLUTION STRUCTURE OF Zα

Introduction

Biological macromolecules, such as proteins, are only active if they are folded into their native three-dimensional structure. Three-dimensional structures of biomacromolecules provide a thorough insight into their function and mechanism of action on the atomic level. This structural information allows one to interpret biological data from site-directed mutagenesis, to map binding sites with interaction partners and to design further functional experiments. High resolution structures of biomacromolecules can be obtained from two distinct biophysical techniques, single crystal X-ray diffraction and multi-dimensional NMR spectroscopy.

For X-ray diffraction, biomacromolecules have to be crystallized in single crystals of 0.1 – 1 mm size. The crystallization process often poses a time-consuming bottle-neck or even prevents structure determination in unfavorable cases. The high resolution structure is obtained from the characteristic X-ray diffraction pattern of the macromolecule in a solid crystal lattice. In some cases, crystal packing forces deform the structure of the molecule, an effect often observed for extended flexible substructures of the molecule and for surface residues that are involved in crystal lattice contacts. However, in most cases the core parts of crystal structures of proteins are virtually identical to solution structures determined by NMR spectroscopy because protein crystals can be considered a highly concentrated frozen solution with approximately 50% of their total mass occupied by solvent.

High resolution NMR spectroscopy is a much younger field than X-ray crystallography. The physical principle of NMR was discovered in 1946 when Bloch et al. [180] and Purcell et al. [181] recorded the first NMR spectra on solid compounds. 15 years ago, the first protein structure was determined by NMR spectroscopy [182]. The fundamental work of Ernst and coworkers on the mathematical description of NMR experiments by the product operator formalism [183], and the development of a general strategy for resonance assignment and structure determination of proteins and nucleic acids by Wuthrich and coworkers [182] has triggered a rapid development in biological NMR spectroscopy. In 1990 protein NMR was greatly advanced by new biosynthetic methods for labeling proteins with stable NMR-sensitive ^{15}N and ^{13}C isotopes [184]. Such labeled proteins allowed the application of multi-dimensional heteronuclear double [185] and triple resonance NMR experiments [186,187], which tremendously reduce signal overlap by dispersing the NMR signals into a third or even fourth dimension. By this technology, protein structures with molecular weights up to ~ 35 kD can be solved by NMR spectroscopy [188]. The size limit of protein NMR can be further extended to approximately 60 kD by random fractional deuteration [189], amino acid selective isotope labeling in a deuterated background [190] and segmental isotope labeling by means of protein splicing [216] using bacterial expression systems. Deuteration as well as selective and segmental labeling diminish the number of NMR-sensitive nuclei that contribute to transverse

relaxation causing rapid loss of signal, considerable line broadening and spectral congestion. A recent advancement in the containment of relaxation in heteronuclear NMR experiments, designated transverse relaxation optimized spectroscopy (TROSY) [191,192], promises the structural investigation of proteins with molecular weights of 100 kD and more. In conclusion, high resolution NMR spectroscopy is an alternative technique to X-ray crystallography for the determination of atomic resolution structures of proteins and nucleic acids under semi-physiological conditions in solution. Therefore, solution-state NMR structures are valuable to validate solid-state crystal structures, and NMR is invaluable for those biomacromolecules that are not amenable to crystallization.

Structure determination by NMR spectroscopy requires two steps of data analysis, resonance assignment and NOE assignment. In resonance assignment, the resonance frequencies (synonymous to chemical shifts, see p. 109) of NMR-sensitive nuclei have to be assigned to their atom identifiers (e.g. L176.Hα) in the chemical structure of the molecule under study. This attaches a chemical shift number to each detectable ^{1}H, ^{13}C and ^{15}N atom of the $^{13}C/^{15}N$-isotope labeled Zα protein domain. In the second step, NOE assignment, cross-peaks in NOESY spectra are assigned to pairs of atom identifiers (e.g. L176.Hα - L179.H_N) on the basis of the chemical shift numbers attached during resonance assignment. The intensities of identified NOE cross-peaks are converted into distance restraints, which are used for structure calculation by dynamic simulated annealing or distance geometry methods.

Semi-automated resonance assignment

A resonance assignment to a certain atom identifier can be derived from four sources of information:

1) The chemical shift range known for a particular atom type from chemical shift databases (e.g. protein Hα's = 3 – 5.5 ppm versus H_N's = 6.5 – 10 ppm)
2) Spectral editing and filtering techniques as well as selective pulses make sure that only certain atom types occur in a NMR spectrum (e.g. only H_N's on one axis of the spectrum).
3) Connections to characteristic or known atoms through scalar couplings (chemical bonds) (e.g. H's closely bonded to backbone HN groups are Hα's.)
4) Connections to characteristic or known atoms through dipolar couplings (NOEs) (e.g. aromatic H's closest to Hβ's are aromatic Hδ's provided that there are no long-range NOEs.)

Resonance assignments of ^{1}H atoms based solely on chemical shifts are often unreliable because chemical shifts can be strongly changed by the tertiary fold of proteins. Likewise, resonance assignment relying on NOEs can be misleading in some cases because these NOEs may originate from unexpected long-range ^{1}H-^{1}H interactions. Spectral editing and filtering techniques simplify spectra but require additional information for resonance assignment. Consequently, NMR spectra connecting ^{1}H, ^{13}C and ^{15}N atoms through one and two bond scalar couplings, which are independent of the tertiary fold, provide the most robust information for resonance assignment and are almost exclusively used when $^{13}C/^{15}N$-labeled protein is available.

Since NMR can only correlate atoms separated by a few chemical bonds or short (< 6 Å) distances, resonance assignment must be carried out sequentially. The polymeric architecture of proteins allows one to subdivide resonance assignment into two steps:
(1) Sequential backbone assignment, which relies on the short distances within the backbone.
(2) Side chain assignment, which relies on the short distances along the side chain.

Sequential backbone assignment

In folded proteins the ^{1}H and ^{15}N resonances of the backbone amide moiety (HN) show a large chemical shift dispersion. The chemical shift of amide protons is far separated from that of side chain protons. Furthermore, backbone amide protons are close to vicinal amino acids allowing one to link sequential residues by scalar or dipolar coupling. Therefore, the backbone HN moiety is a well-suited NMR anchor to which most other atoms can be readily connected.

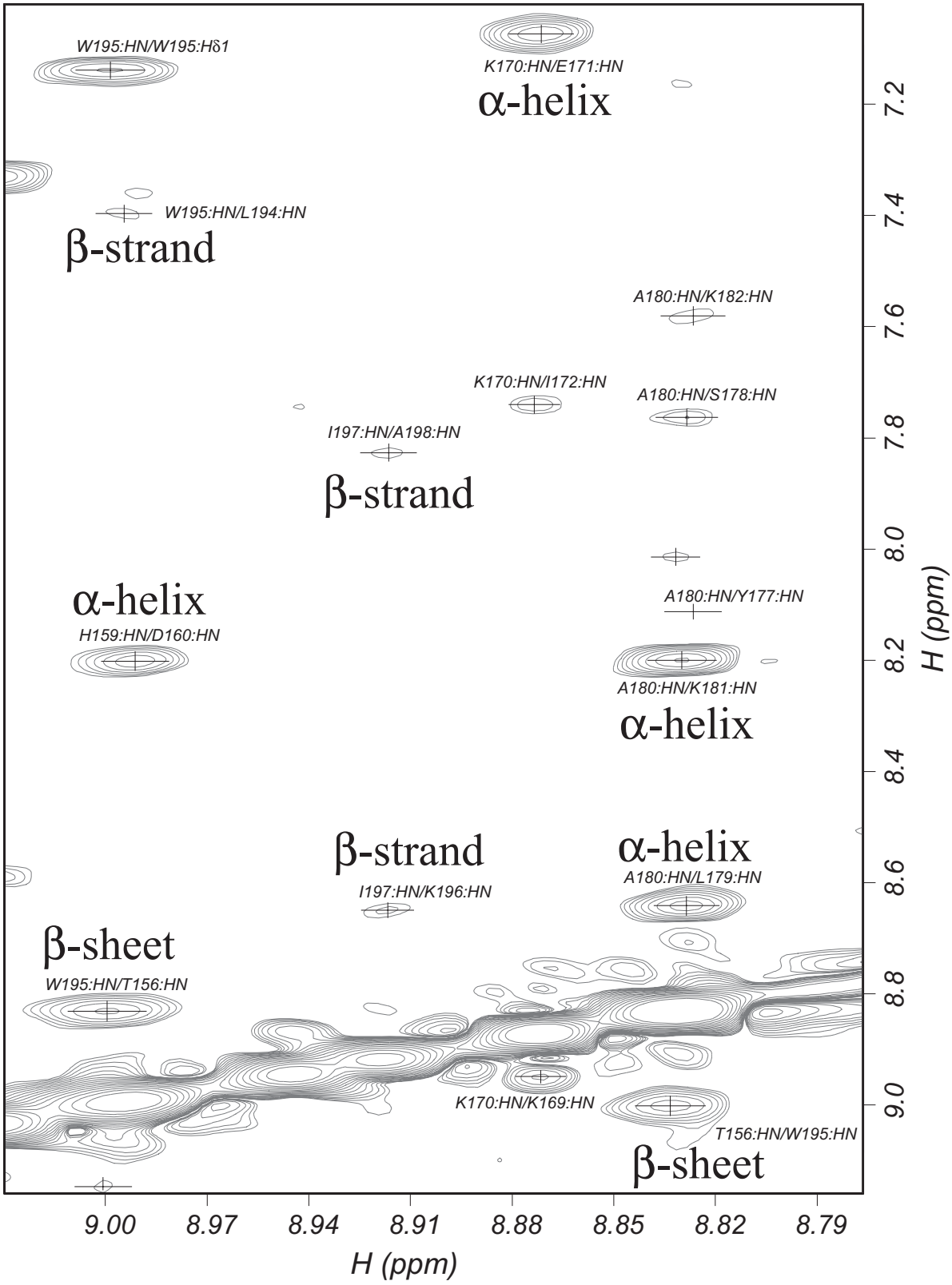
Either of the following three different strategies can be used for sequential backbone assignment dependent on the availability of unlabeled, ^{15}N-labeled or $^{13}C/^{15}N$-double labeled protein:

- homonuclear resonance assignment for unlabeled protein
- ^{15}N-separated resonance assignment for ^{15}N-labeled protein
- triple resonance based resonance assignment for $^{13}C/^{15}N$-labeled protein

Backbone assignment by homonuclear NMR experiments

The resonance assignment of peptides and small proteins of less than ~ 70 residues can often be achieved using homonuclear 2D NMR experiments alone [182,188]. In this homonuclear approach, NOEs between backbone amide protons (H_N), which are strongest for direct neighbors (between residue i and i±1) (table 3), are used to connect sequential residues in 2D NOESY spectra (fig. 28). Additional homonuclear 2D TOCSY and 2D DQF-COSY experiments link the side chain ^{1}H atoms to the backbone H_N, thereby completing the assignment process. This assignment approach fails or delivers only incomplete assignments for poorly dispersed small proteins and for larger proteins because many peaks overlap in homonuclear 2D spectra, especially in the aliphatic region of side chain ^{1}H atoms. Since heavily overlapped cross-peaks provide insufficient distance restraints, protein structures determined purely from homonuclear spectra with considerable spectral congestion are often of poor quality.

fig. 28 **Close-up of the H_N region of a 2D NOESY spectrum (40 ms mixing time) of Zα in H_2O** showing sequential and medium-range H_N-H_N NOEs in α-helices and β-strands which are fundamental for sequential assignment **(see next page)**. The NOE intensity (illustrated by the number of contour lines) of sequential H_N-H_N NOEs is significantly higher in α-helices than in β-strands reflecting the shorter $H_N(i)$-$H_N(i+1)$ distance in α-helices (see table 3). The sequential α-helical NOE between K169:NH and K170:NH is an exception because the signal intensity from the H_N proton of K169, which is the first N-terminal residue of helix α3, is most likely weakened by chemical exchange. α-Helical $H_N(i)$-$H_N(i+2)$ NOEs are of weak intensity due to the long inter-proton distance (4.2 Å). The intense cross-β-strand H_N-H_N NOE between W195 and T156 indicates a short distance between β1 and β3. The NOE between the H_N and Hδ1 of W195 demonstrates that aromatic side chains can be linked to backbone H_N protons by virtue of intraresidual NOEs.


W195:HN/W195:Hδ1
K170:HN/E171:HN
α-helix
W195:HN/L194:HN
β-strand
A180:HN/K182:HN
K170:HN/I172:HN
A180:HN/S178:HN
I197:HN/A198:HN
β-strand
α-helix
H159:HN/D160:HN
A180:HN/Y177:HN
A180:HN/K181:HN
α-helix
β-strand
α-helix
A180:HN/L179:HN
I197:HN/K196:HN
β-sheet
W195:HN/T156:HN
K170:HN/K169:HN
T156:HN/W195:HN
β-sheet
9.00
8.97
8.94
8.91
8.88
8.85
8.82
8.79
H (ppm)
7.2
7.4
7.6
7.8
8.0
8.2
8.4
8.6
8.8
9.0
H (ppm)

Backbone assignment by ^{15}N-separated NMR experiments

Backbone assignment by ^{15}N-separated NMR experiments is principally identical to the homonuclear approach described above. The only difference is that the ^{15}N-chemical shift of the backbone amide is utilized to disperse 2D ^{1}H-^{1}H NOESY and TOCSY spectra into a third dimension. This greatly reduces spectral overlap while the total number of cross-peaks remains unchanged. A schematic 3D spectrum (fig. 29) illustrates the underlying principle of spectral simplification by spreading 2D spectra into a third or even fourth dimension. The third dimension is obtained from the chemical shift of a chemically bonded heteronucleus (in this case the ^{15}N atom of the amide group), which is recorded in the HSQC part of the NMR pulse sequence (see Appendix, fig. 58). The magnetization transfer in backbone ^{1}H-^{15}N and ^{1}H-^{13}C groups is very efficient because of large one-bond coupling constants with $^{1}J_{H\text{-}N}$ = 90 Hz and $^{1}J_{H\text{-}C}$ = 140 Hz, respectively [18].

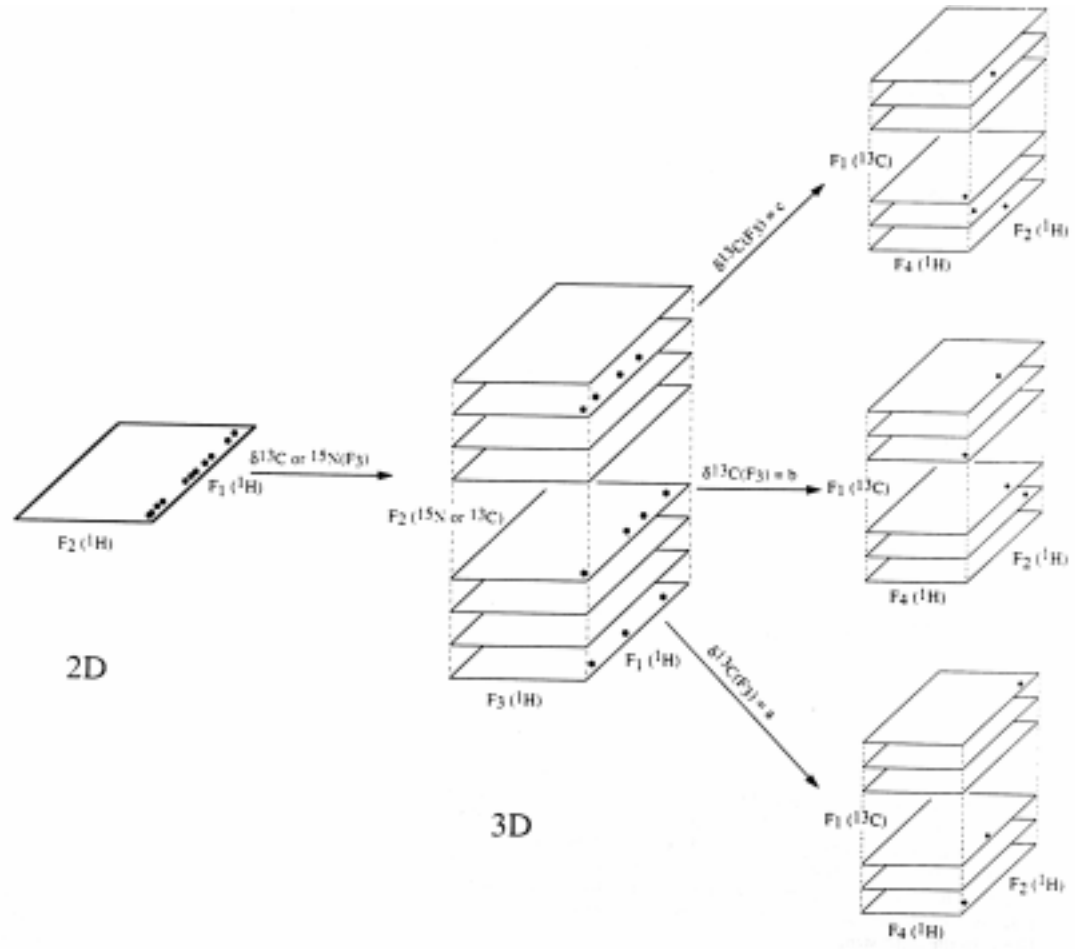


fig. 29 **Schematic illustration of spectral simplification by dispersing a 2D spectrum into a third and fourth dimension** (adapted from [188]). A number cross-peaks, of which several overlap in a 2D spectrum, are distributed on 8 - 64 planes governed by the ^{15}N or ^{13}C chemical shift of the attached heteronucleus, thereby building a 3D spectrum. Cross-peaks from the planes of the 3D spectrum may be further distributed on separated planes to give a 4D spectrum. The allocation in the fourth dimension is given by the chemical shift of a second attached heteronucleus. The gain in information from 4D spectra with respect to 3D spectra is limited because it is compromised by reduced signal-to-noise, poor digital resolution and time-expensive NMR experiments.

The gain in resolution for the Zα protein achieved by 3D ^{15}N-separated spectra becomes clear from the 2D ^{15}N-^{1}H HSQC (Heteronuclear Single-Quantum Coherence) spectrum of Zα (fig. 30). ^{15}N-^{1}H HSQC spectra give a single peak with the chemical shifts of the H_N proton and the backbone ^{15}N atom for each non-proline residue of a protein sample. In addition, side chain N-H moieties of tryptophan and arginine, and N-H_2 moieties of asparagine and glutamine residues are detected in ^{15}N-^{1}H-HSQC spectra (labeled in fig. 30). Of a total of 81 detectable non-proline residues of Zα, 70 are well resolved in the HSQC spectrum. Hence, each of these 70 backbone amide peaks occur as well-separated 2D ^{1}H-^{1}H NOESY strips in the 3D ^{15}N-HSQC-NOESY spectrum. In contrast, all H_N-N cross-peaks in fig. 30 with identical $^{1}H_N$ chemical shifts but different ^{15}N chemical shifts, overlap in a 2D homonuclear NOESY spectrum. This demonstrates that 3D ^{15}N-separated NMR spectroscopy greatly improves the information content of spectra recorded on ^{15}N-labeled protein, or alternatively on ^{15}N/^{13}C-double labeled protein.

Sequential backbone assignment using ^{15}N-separated spectra relies on NOEs between adjacent residues because sequential residues cannot be connected by scalar coupling in ^{15}N-labeled proteins. NOEs between sequential backbone amide protons (denoted $d_{NN}(i, i+1)$) and between the Hα of residue i and the backbone amide proton of residue i+1 (denoted $d_{\alpha N}(i, i+1)$) give the most indicative cross-peaks for sequential assignment because these atoms are close in space (see table 3). Moreover, NH and Hα resonances are separated from the crowded aliphatic region in the spectrum which facilitates their analysis. Sequential NOEs between side chain ^{1}H atoms and the backbone amide protons, such as $d_{\beta N}(i, i+1)$, confirm the sequential assignment and become crucial if sequential H_N and Hα NOEs are ambiguous. Prior to the usage of Hα and side chain NOEs, intraresidual NOEs must be identified because ^{15}N-HSQC-NOESY spectra contain all, intraresidual, sequential, medium-range (defined as $1<|i-j|<5$) and long-range (defined as $|i-j|>4$) NOEs (see fig. 31). Identification and partial assignment of intraresidual ^{1}H atoms is readily achieved by using a 3D ^{15}N-HSQC-TOCSY experiment [18] that correlates the side chain ^{1}H atoms with the backbone amide of the same residue by isotropic mixing. This information allows one to discriminate between intraresidual and sequential NOEs in ^{15}N-HSQC-NOESY spectra. Long-range NOEs can be neglected at this stage because they are rare, as compared with sequential and medium-range NOEs to amide protons. Medium-range NOEs are readily recognized once four or more residues have been sequentially assigned.

fig. 30 **The 2D ^{15}N-^{1}H HSQC spectrum of Zα (see next page),** in which each peak represents the backbone amide of one non-proline residue, shows that 70 of a total of 81 expected residues are resolved in the spectrum. The backbone amide peaks of N173 and K196, and E140 and K145 partially overlap (see labels), and those of L133, A155, D160 and K181 fall onto a single spot. Thus, the side chain resonances of these residues cannot be discriminated by virtue of their backbone amide chemical shifts in 3D ^{15}N-edited spectra. In addition, the side chain N-H moieties of tryptophan (upper left corner) and arginine ('folded' into lower right corner) and the N-H_2 moieties of asparagine (labeled N173.Hδ2∗) and glutamines (boxed and labeled Q.Hε2∗) appear in characteristic regions of this spectrum.

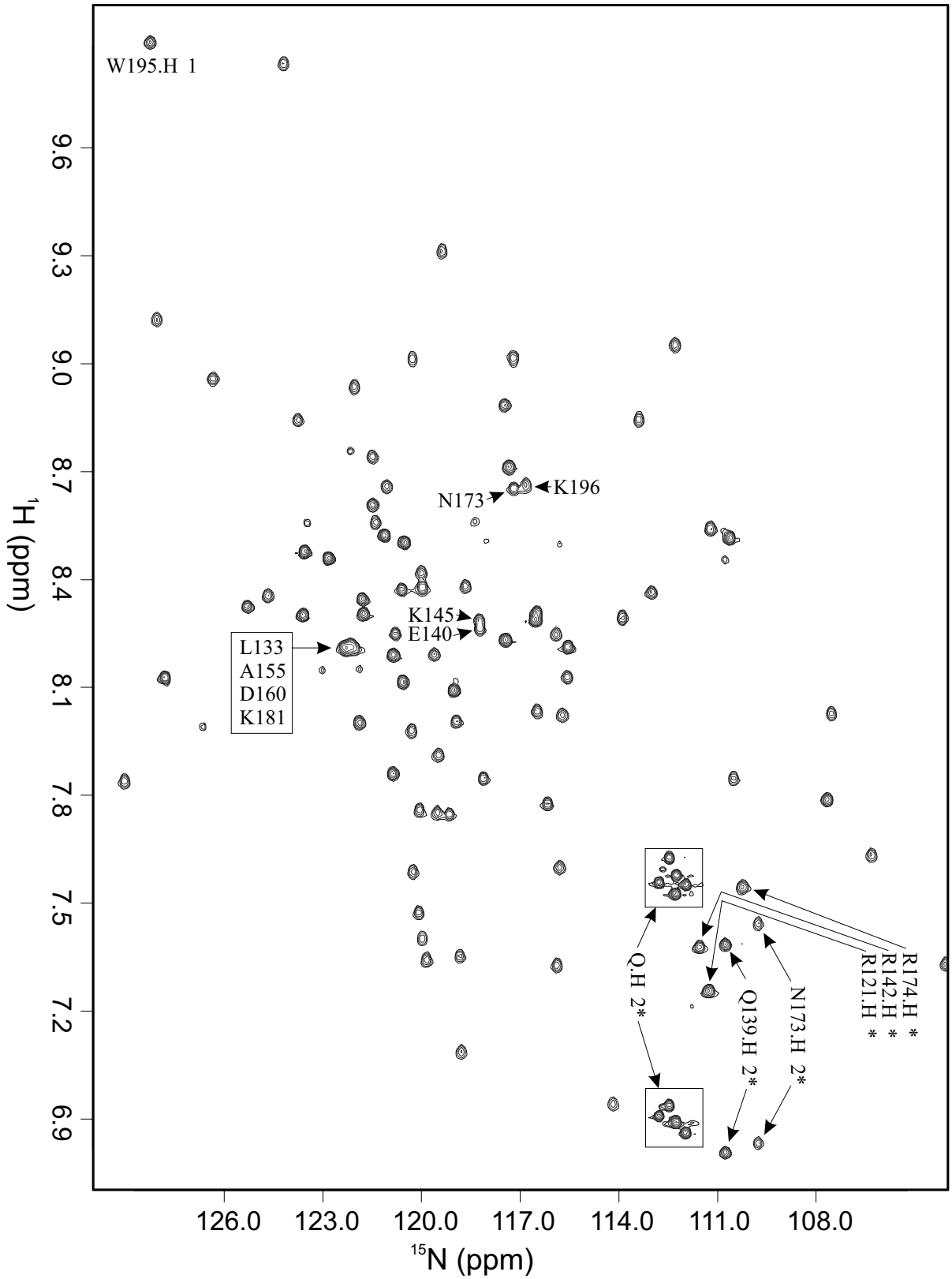

W195.H 1
N173
K196
K145
E140
L133
A155
D160
K181
Q.H 2*
Q139.H 2*
N173.H 2*
R174.H *
R142.H *
R121.H *
^{1}H (ppm)
^{15}N (ppm)
9.6
9.3
9.0
8.7
8.4
8.1
7.8
7.5
7.2
6.9
126.0
123.0
120.0
117.0
114.0
111.0
108.0

table 3 **Short (< 4.5 Å) sequential and medium-range ^{1}H-^{1}H distances in polypeptide secondary structures** [182]. Distances resulting in the most striking NOEs are marked in **bold**.

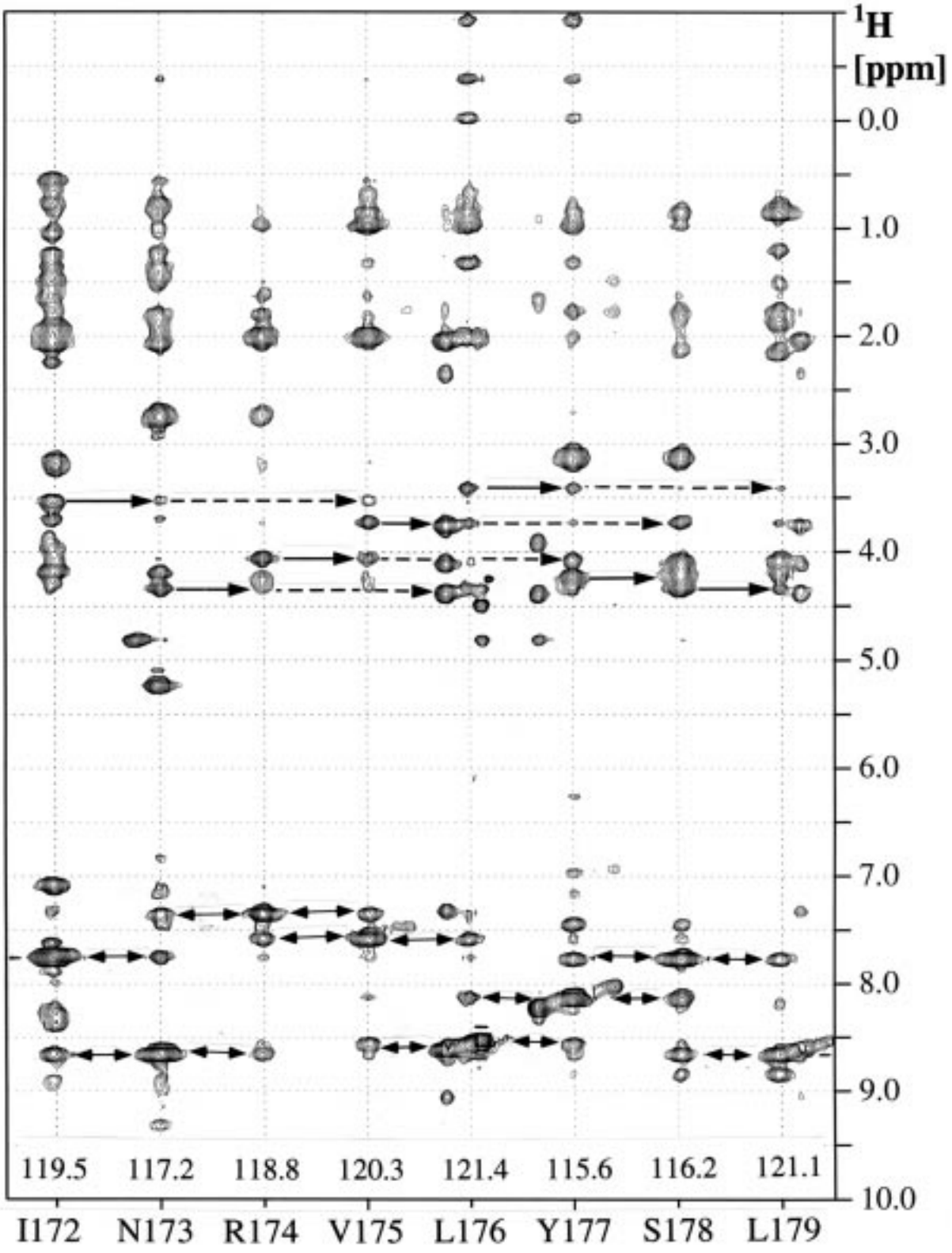
Distance	α-helix	3_{10}-helix	antiparallel β-sheet	parallel β-sheet
$d_{NN}(i, i+1)$	**2.8**	**2.6**	4.3	4.2
$d_{NN}(i, i+2)$	4.2	4.1		
$d_{\alpha N}(i, i+1)$	3.5	3.4	**2.2**	**2.2**
$d_{\alpha N}(i, i+2)$	4.4	3.8		
$d_{\alpha N}(i, i+3)$	**3.4**	**3.3**		
$d_{\alpha N}(i, i+4)$	**4.2**			
$d_{\beta N}(i, i+1)$	2.5 - 4.1	2.9 – 4.4	3.2 – 4.5	3.7 – 4.7
$d_{\alpha\beta}(i, i+3)$	2.5 – 4.4	3.1 – 5.1		

The 3D ^{15}N-HSQC-NOESY spectra of Zα were recorded at three different mixing times, 70, 150 and 250 ms, in order to optimize the mixing time towards a maximal number of H_N-H_N NOEs and a minimal degree of spin diffusion. The 3D spectra with 150 and 250 ms mixing time were selected and dissected into 2D ^1H_N-^{1}H strips using the ^{15}N-^{1}H HSQC spectrum of Zα as a template for the ^{15}N and ^1H_N chemical shifts. Automated 'strip picking' from a clean template spectrum brings about the advantage that strips are only taken from those parts of the 3D spectrum where real signals occur. Noise and other artifacts elsewhere in the 3D space are excluded making time-consuming spectral cleaning after peak-picking unnecessary. This semi-automated spectral analysis strategy, which uses the backbone amide as an anchor group, was successfully applied to all 3D spectra of Zα.

Sequential assignment with 2D strips has the advantage that they can be easily aligned sequentially allowing one to use both line shape and chemical shift information for assignment. It is apparent from fig. 31 that the shape of contour lines varies among H_N cross-peaks. Cross-peaks arising from the same H_N proton are more similar in magnitude and line shape allowing one to confirm sequential assignment.

The search for sequential neighbors is started with a 2D strip i, showing at least two well-resolved and strong H_N-H_N cross-peaks. This strip is compared to all other strips to find a strip j that shows a diagonal peak at one of the two H_N-H_N cross-peak positions of strip i and a cross-peak at the H_N diagonal peak position of strip i. Then both strips are checked for the presence of matching Hα and other side chain NOEs to confirm the sequential assignment. Since NOEs between aliphatic ^{1}H atoms and backbone amide protons are directional, the two strips are sequentially ordered so that the strip containing sequential aliphatic NOEs takes the C-terminal position. This strip is the point of departure for the next search and so on, till no new matching strip can be identified unambiguously. Then the string of strips is extended in the N-terminal direction. If the sequential extension fails on both termini due to spectral overlap,

fig. 31 **Sequential and secondary structure assignment in helix α3 (see next page).** 2D strips of α3 residues extracted from a 3D ^{15}N-edited NOESY-HSQC spectrum of Zα, depict sequential NOEs between amide protons (marked by double arrows in the 7-9.5 ppm region), sequential Hα(i)→NH(i+1) NOEs (solid arrows in the 3.4 - 5.3 ppm Hα region) and medium range Hα(i)→NH(i+3) NOEs (dashed arrows in the 3.4 - 5.3 ppm Hα region) characteristic for α-helices. Backbone ^{15}N chemical shifts are shown above each residue label at the bottom (adapted from [5]).


^{1}H
[ppm]
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
119.5
117.2
118.8
120.3
121.4
115.6
116.2
121.1
I172
N173
R174
V175
L176
Y177
S178
L179

signal loss or a proline residue, a new starting strip is selected from the pool of so far unassigned strips, and the search for sequential neighbors begins again. This repetitive procedure of searching and comparing 2D strips is supported by various NMR analysis software, such as NMRpipe [20], Aurelia (Bruker GmbH) and Felix'97 (MSI Inc.), all of which have been used for this work. Starting with H_N-H_N cross-peaks is most favorable for α-helical regions of the protein, while Hα-H_N cross-peaks are the most reliable point of departure for extended regions (table 3).

In cases with ambiguous H_N-H_N cross-peaks, additional side chain information can be used for sequential assignment. A number of amino acids show characteristic chemical shift patterns in 3D ^{15}N-HSQC-TOCSY spectra. For instance, glycine residues exhibit two Hα peaks, provided that the Hα chemical shifts are not degenerate, but no other aliphatic peaks. In contrast, hydrophobic amino acids often show Hα, Hβ and methyl peaks, all of which occur in different regions of the chemical shift scale. Thus, chemical shift information about the ^{1}H atoms of the side chains can be converted into amino acid type information. This helps in the aforementioned sequential assignment procedure because the primary amino acid sequence of the protein is known from its genotype. For example, if three candidate 2D strips match the H_N-H_N cross-peaks of a strip, whose amino acid type is known from the context of already assigned sequential neighbors, one of the three strips may be unambiguously assigned based on the fact that the primary sequence requires a glycine residue, and two of the three strips have peaks in the Hβ or methyl chemical shift region. Consequently, side chain information confirms sequential assignment and vice versa.

Another source of additional information for sequential assignment is the characteristic NOE pattern for α-helices and β-strands. Sequential H_N-H_N cross-peaks in α-helical regions are much more intense than in β-stranded regions, as seen in fig. 28. In contrast, $d_{\alpha N}(i, i+1)$ NOEs are significantly stronger in β-sheets than in α-helices reflecting distinct Hα-H_N distances in α-helices and β-strands (see table 3). Furthermore, residues in α-helices show $d_{NN}(i, i+2)$, $d_{\alpha N}(i, i+2)$, $d_{\alpha N}(i, i+3)$ and $d_{\alpha N}(i, i+4)$ NOEs (fig. 31), which are absent for β-stranded residues (fig. 28). The main chain directed approach (MCD) for sequential assignment [193], relies on these regular NOE patterns in α-helices and β-strands to overcome problems of ambiguity with $d_{NN}(i, i+1)$ and $d_{\alpha N}(i, i+1)$ NOEs. However, the MCD approach does not work well in segments of random secondary structure, such as long loops. Therefore, it is most effective to start sequential assignment with the $d_{NN}(i, i+1)$, $d_{\alpha N}(i, i+1)$ and $d_{\beta N}(i, i+1)$ NOE-based approach, as described above. Medium-range NOEs and side chain/primary sequence information is then used to overcome cases of ambiguity and to check for internal consistency. By this ^{15}N-separated sequential backbone assignment strategy, all but two residues of the Zα core fold (residues 136-198) could be unambiguously assigned. However, several residues of the unfolded N-terminus (residues 117 – 135) could not be assigned sequentially due to poor dispersion of the H_N chemical shifts.

Backbone assignment by triple resonance experiments

A triple resonance NMR experiment transfers magnetization through scalar (through bond) coupling between three different nuclear species, ^{1}H, ^{15}N and ^{13}C, which requires a ^{15}N/^{13}C-double labeled protein sample. In triple resonance experiments, magnetization can be transferred through a peptide bond onto the adjacent residue. This allows one to connect

sequential residues solely by through-bond coherence rather than by sequential NOEs that may be misleading in some cases.

Several 3D triple resonance experiments have been developed that create coherence between various nuclei in the protein backbone [18]. The 3D CBCA(CO)NH and the 3D CBCANH triple resonance experiments [18] are particularly useful because they deliver both sequential assignment and chemical shift assignment of Cα and Cβ nuclei, which are required for NOE assignment later on. The CBCA(CO)NH experiment transfers magnetization by the INEPT pulse sequence (see Appendix fig. 57) from the Hβ and Hα atoms to the attached Cβ and Cα atoms of residue i (fig. 32). After chemical shift labeling of Cβ and Cα, magnetization is transferred to the backbone amide of residue i-1. The ^{15}N and $^{1}H_N$ chemical shifts of this amide group are labeled during the evolution period t_2 and the acquisition period t_3, respectively. Thus, each non-proline residue gives raise to two cross-peaks in the 3D spectrum at the positions (F1, F2, F3) = ($\Omega_{C\beta(i)}$, $\Omega_{N(i-1)}$, $\Omega_{HN(i-1)}$) and ($\Omega_{C\alpha(i)}$, $\Omega_{N(i-1)}$, $\Omega_{HN(i-1)}$).

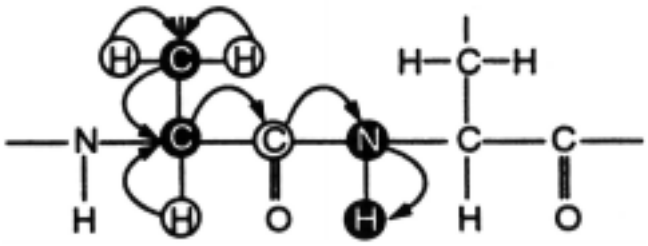


fig. 32 **Schematic magnetization transfer in the CBCA(CO)NH triple resonance experiment**. A dipeptide segment of a protein is shown to illustrate how the CBCA(CO)NH experiment correlates the Cβ and Cα chemical shifts of residue i with the backbone ^{15}N and $^{1}H_N$ chemical shifts of residue i-1.

The related 3D CBCANH experiment also records the chemical shifts of the Cβ and Cα atoms, but then correlates them to the backbone ^{15}N and $^{1}H_N$ chemical shifts of the same residue i (fig. 33). In addition, the Cβ and Cα chemical shifts are correlated to the backbone ^{15}N and $^{1}H_N$ chemical shifts of residue i+1. The cross-peaks with the backbone amide of residue i+1 are less intense than those with the amide of the same residue i, because the two bond coupling constant across the peptide bond, $^{2}J_{NC\alpha}$ = 4 – 9 Hz, is slightly smaller than the one bond coupling constant to the intraresidual ^{15}N atom, $^{1}J_{NC\alpha}$ = 7 – 11 Hz. In the CBCANH experiment, Cβ and Cα cross-peaks have opposite sign, while they have like sign in the CBCA(CO)NH experiment. In the 3D CBCANH spectrum, each non-proline residue shows four cross-peaks at the chemical shift positions (F1, F2, F3) = - ($\Omega_{C\beta(i)}$, $\Omega_{N(i)}$, $\Omega_{HN(i)}$), ($\Omega_{C\alpha(i)}$, $\Omega_{N(i)}$, $\Omega_{HN(i)}$), - ($\Omega_{C\beta(i-1)}$, $\Omega_{N(i)}$, $\Omega_{HN(i)}$) and ($\Omega_{C\alpha(i-1)}$, $\Omega_{N(i)}$, $\Omega_{HN(i)}$).

In practice, both 3D spectra are dissected into 2D ^{13}C-$^{1}H_N$ strips using the ^{15}N-^{1}H

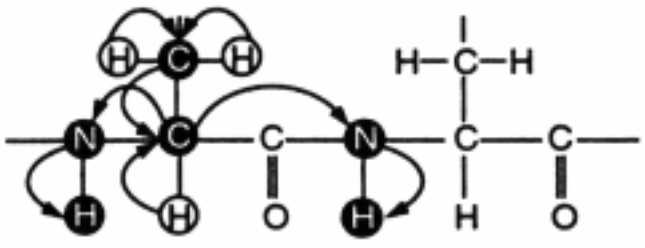


fig. 33 **Schematic magnetization transfer in the CBCANH triple resonance experiment,** which correlates the Cβ and Cα chemical shifts of residue i with the backbone ^{15}N and $^{1}H_N$ chemical shifts of residue i. In addition, a smaller proportion of magnetization is transferred to the subsequent backbone amide yielding additional cross-peaks with the ^{15}N and $^{1}H_N$ chemical shifts of residue i+1.

HSQC spectrum of Zα as a template. Analogous to the 2D strip analysis of 3D ^{15}N-HSQC-NOESY spectra, chemical shift and line shape information of the 2D ^{13}C-$^{1}H_N$ strips are used to achieve sequential assignment. One strip from the CBCA(CO)NH spectrum and one from the CBCANH spectrum belonging to the same backbone amide, are selected as a starting point provided that they exhibit well-resolved peaks. The intraresidual Cβ and Cα cross-peaks of the CBCANH-derived strip are then used to search for a match in the CBCA(CO)NH-derived pool of strips. If the search was successful the CBCANH-derived strip corresponding to the same backbone amide is picked, and a new sequential search for the next C-terminal neighbor begins (see Appendix, fig. 60 for complete sequential alignment of 2D ^{13}C-$^{1}H_N$ strips). It is advantageous to use the intraresidual Cβ and Cα peaks as the bait and to fish in the CBCA(CO)NH pool because the CBCA(CO)NH strips contain less spectral overlap, and more intense and cleaner cross-peaks.

If both the Cβ and the Cα cross-peaks are ambiguous with respect to chemical shift and line shape, additional information is required to overcome ambiguity. If ^{15}N-HSQC-NOESY spectra are available sequential backbone assignment applying the ^{15}N-separated and triple resonance approach in conjunction can resolve almost all problems of ambiguity. By this twin-track assignment strategy, all backbone resonances of Zα, including those of the 20 unfolded N-terminal residues, could be assigned. Furthermore, primary sequence information in combination with the chemical shift information of the Cα and Cβ peaks may help in sequential assignment. A number of amino acids, such as glycine, alanine, threonine and serine, have characteristic Cα and Cβ chemical shifts that allow one to automatically predict the sequential assignment from CBCA(CO)NH and CBCANH spectra with some precision.

Alternatively, further triple resonance experiments can clear ambiguity. Sequential backbone assignment using the HBHA(CBCACO)NH and HBHA(CBCA)NH pair of NMR experiments is identical to the CBCA(CO)NH / CBCANH–based approach but records the Hβ and Hα chemical shifts instead of the Cβ / Cα shifts. Since it is extremely rare that Cβ, Cα, Hα and Hβ chemical shifts are degenerate, the usage of more than one pair of triple resonance experiments allows sequential backbone assignment in almost all cases. The remarkable robustness of the triple resonance-based sequential assignment approach even enables assignment of partially denatured and unfolded proteins, thereby allowing high resolution structural investigation of protein folding.

Moreover, the extremely clean triple resonance spectra exhibiting excellent signal-to-noise ratios, are particularly suited for automated assignment procedures. However, overlap of backbone amide peaks poses a severe bottleneck for both ^{15}N-separated and triple resonance based assignment that often causes automated algorithms to fail. For instance, the backbone amide $^{1}H_N$-^{15}N cross-peaks of L133, A155, D160 and K181 of Zα are degenerate (fig. 30) leading to a crowded pair of CBCA(CO)NH / CBCANH strips, in which not all of the expected Cβ and Cα peaks are resolved. The $^{1}H_N$-^{1}H strip derived from the ^{15}N-HSQC-NOESY shows even more severe spectral congestion. Therefore, both triple resonance and ^{15}N-separated sequential assignment fail at the positions of these four residues. The resulting four gaps can be filled in if the surrounding residues complement the sequential information. In this case, the CBCANH strips of the preceding residues 132, 154, 159 and 180 allow one to assign the i-1 cross-peaks in the degenerate strips, and the CBCA(CO)NH-strips of the succeeding residues 134, 156, 161 and 182 allow one to identify the degenerate intraresidual Cα and Cβ cross-peaks. In conclusion, cases of peak overlap necessitate manual inspection and consultation of

additional spectral information. Therefore, sequential backbone assignment may be carried out in a semi-automated manner, but full automation remains to be elusive in such cases.

Side chain assignment

Side chain assignment of $^{15}N/^{13}C$ double labeled proteins is most efficiently achieved by using 3D NMR experiments that correlate side chain ^{1}H and ^{13}C atoms to the backbone HN anchor group. The side chain atoms of Zα were assigned by a semi-automated procedure with the program Felix'97 (MSI Inc.) analyzing the following 3D NMR experiments in a stepwise fashion. Prior to automated side chain assignment, the 3D spectra have been peak picked, but not manually edited.

In the first step, the Hα(i) resonances were automatically collected from a 3D HNHA spectrum using the 2D ^{15}N-HSQC spectrum of Zα as a template spectrum. In the second step, the Hα(i-1) and Hβ(i-1) resonances were automatically collected from a 3D HBHA(CO)NH spectrum, again using the 2D ^{15}N-HSQC spectrum as a seed spectrum. The analysis of both the 3D HNHA and HBHA(CO)NH spectrum brings about the advantage that Hα and Hβ atoms of serine and threonine can be unambiguously discriminated. This is not possible from a 3D HBHA(CO)NH spectrum alone because the Hα and Hβ chemical shifts of serine and threonine residues are too similar. The Hα chemical shift value is taken from the HBHA(CO)NH spectrum because it contains cleaner cross-peaks due to a higher signal-to-noise ratio. The HBHA(CO)NH spectrum is favored over the HBHANH spectrum because scalar coherence transfer across the carbonyl group is more efficient than transfer to the HN group of the same residue. Finally, the table of resonances is checked for completeness. Hα or Hβ resonances that have been missed by the automated routine, are manually collected from the 3D HNHA and HBHA(CO)NH spectra.

The remaining aliphatic side chain ^{1}H atoms, Hγ(i–1), Hδ(i-1) and Hε(i-1), are automatically extracted from a 3D H(CCO)NH spectrum using the 2D ^{15}N-HSQC spectrum as before. In the H(CCO)NH experiment, the equilibrium ^{1}H magnetization of all side chain protons is chemical shift labeled and then transferred through ^{1}H-^{13}C and ^{13}C-^{13}C scalar coherence to the backbone HN group of the succeeding residue, whose ^{1}H and ^{15}N chemical shifts are also recorded. Again the resonance table is checked for completeness, and missing resonances are manually added. This is feasible because the number of expected resonances is known for each residue from the sequential assignment.

Hγ, Hδ and Hε resonances cannot be discriminated on the basis of their chemical shifts. These aliphatic resonances are assigned by the combined analysis of a 3D HCCH-COSY and a 3D HCCH-TOCSY spectrum. These two 3D NMR experiments are analogous to homonuclear 2D COSY and 2D TOSCY experiments, but show less spectral congestion by dispersing the cross-peaks into a third ^{13}C dimension. In the 3D HCCH-COSY, equilibrium ^{1}H magnetization is transferred to the directly attached ^{13}C atom via the $^{1}J_{CH}$ coupling (~ 140 Hz), then from that ^{13}C atom to all neighboring ^{13}C atoms via the $^{1}J_{CC}$ coupling (32 - 40 Hz), and finally from that ^{13}C atom to the directly attached ^{1}H atoms again via a $^{1}J_{CH}$ coupling. The flow of magnetization from atom to atom gave rise to the name 'HCCH' experiments.

In the 3D HCCH-TOCSY, the magnetization transfer from ^{1}H to ^{13}C atoms is performed in the same way, but then the transverse ^{13}C magnetization is distributed along the carbon chain by isotropic mixing. Finally, the ^{13}C magnetization is transferred back to the directly attached ^{1}H atoms and detected. For proteins of the size of Zα or larger, the three-step

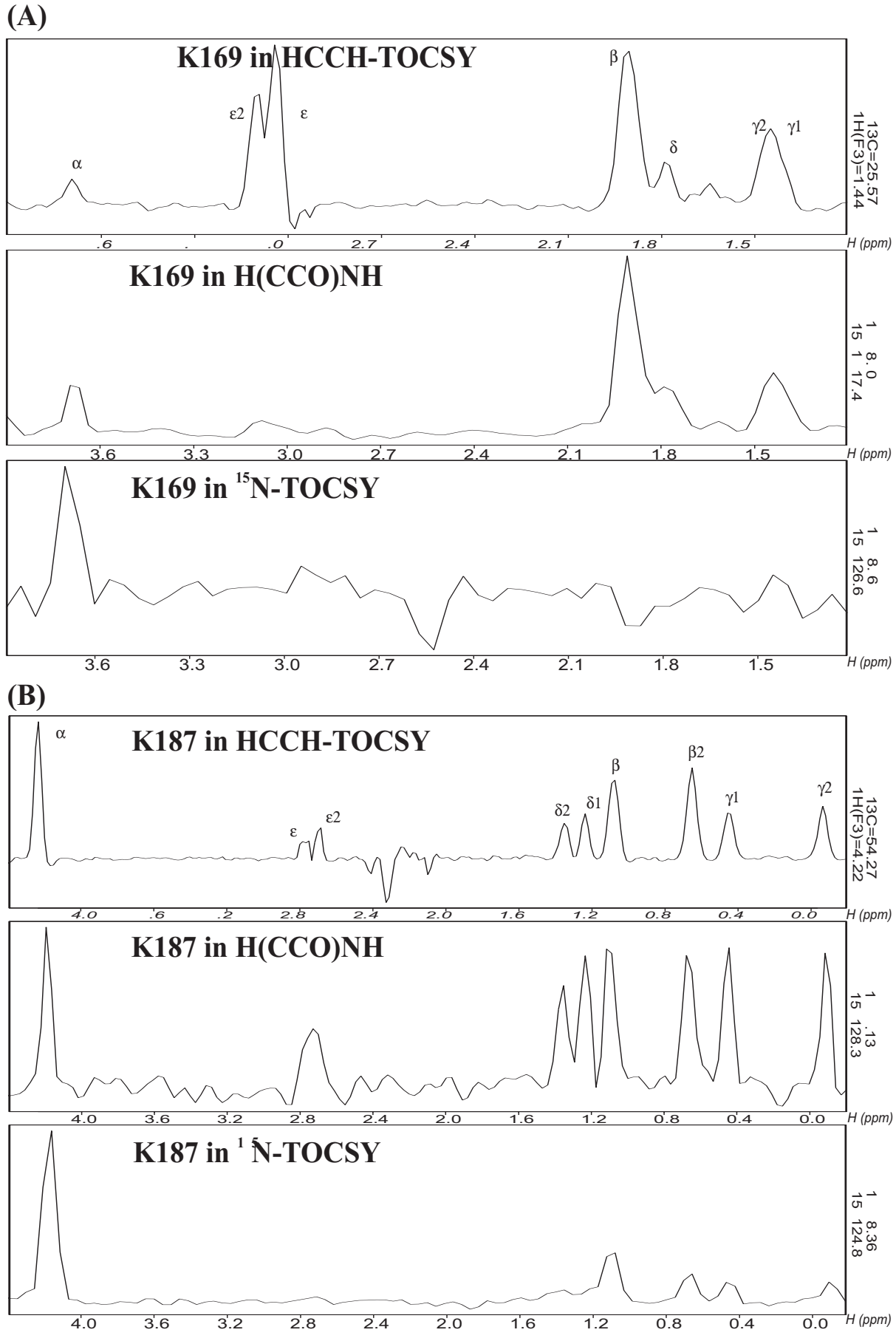
HCCH magnetization transfer is significantly more efficient than the one-step ^{1}H - ^{1}H magnetization transfer via the three-bond $^{3}J_{HH}$ coupling. This is illustrated by the comparison of three different 3D NMR experiments that all utilize isotropic mixing for magnetization transfer through the side chain. The 2D strips of lysine 169 extracted from the three TOCSY experiments 3D HCCH-TOCSY, 3D H(CCO)NH and 3D ^{15}N-HSQC-TOCSY of Zα (fig. 34A) show that only the HCCH-TOCSY spectrum contains a complete set of aliphatic side chain ^{1}H atoms. The 3D H(CCO)NH experiment lacks the terminal Hε atoms of K169, and the ^{15}N-HSQC-TOCSY experiment only detects the Hα atom. The side chain ^{1}H resonances of lysine 187 are completely present in both the HCCH-TOCSY and 3D H(CCO)NH experiments, but signal-to-noise and resolution are better in the HCCH spectrum (fig. 34B). Again, the ^{15}N-HSQC-TOCSY spectrum lacks several side chain resonances. Consequently, both HCCH-TOCSY and 3D H(CCO)NH experiments were useful for the assignment of ^{1}H side chain resonances of Zα, whereas the ^{15}N-HSQC-TOCSY experiment showed inadequate data completeness.

In practice, aliphatic side chain resonances are assigned from a 3D HCCH-COSY experiment in the following way. The (Hβ, Cβ, Hβ) diagonal peak in the HCCH-COSY, whose chemical shift position is already known from the CBCA(CO)NH and HBHA(CO)NH experiments, is the point of departure. In the 2D ^{1}H-^{1}H plane at the ^{13}C position of Cβ, it shows cross peaks to the Hα at position (Hβ, Cβ, Hα), possibly to a second Hβ at (Hβ1, Cβ, Hβ2), and to the Hγ resonances at position (Hβ, Cβ, Hγ). At the position of this Hγ resonance, a perpendicular 2D (^{1}H, ^{13}C) plane is inspected in order to find the Cγ position at (Hγ, Cγ, Hγ). All putative Cγ cross-peaks lying on the (Hγ, Hγ) plane within the expected chemical shift range of Cγ atoms, are further investigated one at a time in 2D (^{1}H, ^{1}H) planes taken at these putative Cγ positions. If the (Hγ, Cγ, Hγ) diagonal peak correlates with a (Hγ, Cγ, Hβ) cross-peak, this putative Cγ position is further verified by inspecting the 3D HCCH-TOCSY spectrum because the (Hγ, Cγ, Hβ) cross-peak may be degenerate or otherwise misleading.

The 3D HCCH-TOCSY shows almost all side chain ^{1}H resonances in the 2D (^{1}H, ^{1}H) at the putative ^{13}C position of Cγ, including the well-resolved Hα resonance. Thus, the assignment is supported by more than one cross-peak, thereby greatly alleviating problems with spectral congestion in the aliphatic region. The chemical shift of the Cγ is most accurately taken from a clean (Hγ, Cγ, Hα) cross-peak because diagonal peaks and HCCH-COSY cross-peaks are often less precise due to signal overlap.

The Hδ/Cδ and Hε/Cε resonances are identified in the same manner as described for the Hγ/Cγ pair using the (Hγ, Cγ, Hγ) and (Hδ, Cδ, Hδ) diagonal peaks, respectively, as the starting points in the 3D HCCH-COSY spectrum. This procedure led to a complete list of Zα assignments which is attached in the Appendix (table 13 and table 14).

fig. 34 **Data completeness in three different heteronuclear 3D TOCSY experiments utilized for side chain assignment (see next page). (A)** In the 3D HCCH-TOCSY, all aliphatic side chain ^{1}H atoms of K169 are clearly visible. The 3D H(CCO)NH experiments shows all resonances of K169 but the terminal Hε ones. The 3D ^{15}N-HSQC-TOCSY shows only the Hα. The backbone H_N of K169 is most likely weakened by chemical exchange which diminishes the sensitivity of H_N-based experiments. **(B)** For K187, the 3D HCCH-TOCSY and H(CCO)NH are approximately equally sensitive. Again, the ^{15}N-HSQC-TOCSY shows poor signal-to-noise lacking the Hδ and Hε resonances of K187. Consequently, 3D H(CCO)NH and HCCH-TOCSY are well-suited NMR experiments for aliphatic side chain assignment whereas the ^{15}N-edited TOCSY is only of limited use.


(A)
K169 in HCCH-TOCSY
α
ε2
ε
β
δ
γ2
γ1
13C=25.57
1H(F3)=1.44
H (ppm)
K169 in H(CCO)NH
H (ppm)
K169 in 15N-TOCSY
H (ppm)
(B)
K187 in HCCH-TOCSY
α
ε
ε2
δ2
δ1
β
β2
γ1
γ2
13C=54.27
1H(F3)=4.22
H (ppm)
K187 in H(CCO)NH
H (ppm)
K187 in 15N-TOCSY
H (ppm)

The simultaneous inspection of both the HCCH-COSY and HCCH-TOCSY is supported by the „frame connection“ option in the Felix'97 program that links zoom and display operations in both spectra. Furthermore, the analysis of these two HCCH spectra may be fully automated. However, this is only of limited use because the aliphatic region of the spectrum is particularly affected by spectral congestion. For instance, most methyl groups of valine and leucine residues are not well separated in Zα. Moreover, Hγ resonances of leucines often show weak intensities preventing the usage of a high peak-pick threshold to alleviate problems with partial spectral overlap. Consequently, spectral analysis of the poorly dispersed aliphatic region is most reliably performed in a primarily manual fashion.

The manual search for the Hγ/Cγ, Hδ/Cδ and Hε/Cε resonances in the HCCH spectra was guided by the Hγ, Hδ and Hε resonances collected previously from the 3D H(CCO)NH spectrum in a semi-automated manner. The 3D H(CCO)NH spectrum lacked terminal ^{1}H resonances only for a few (< 10 %) residues with long side chains. Hence, the 3D H(CCO)NH experiment allows fast and almost complete collection of side chain ^{1}H resonances that can be largely automated. In conclusion, automated assignment procedures work well with 3D spectra utilizing the well-dispersed HN anchor group but do not significantly facilitate the assignment of poorly separated aliphatic resonances.

Assignment of aromatic side chains

The side chains of aromatic residues cannot be assigned from the aforementioned 3D H(CCO)NH or ^{15}N-HSQC-TOCSY spectra since the through-bond coherence transfer between the aromatic ring and the Cβ/Hβ atoms is very inefficient. 3D HCCH spectra lack ^{1}H and ^{13}C resonances of aromatic side chains because aromatic carbons resonate at a much higher frequency (110 – 140 ppm) than aliphatic carbons. This frequency is not sufficiently covered by the radio-frequency pulses on aliphatic carbons in HCCH experiments. Consequently, additional NMR experiments are required to acquire the aromatic ^{1}H and ^{13}C resonances.

The ^{1}H resonances of most aromatic side chains were obtained from homonuclear 2D DQF-COSY and TOCSY experiments in D_2O, which correlate the ^{1}H atoms within an aromatic ring. The isolated aromatic ring systems were assigned sequence specifically by analyzing NOEs between backbone H_N and aromatic Hδ atoms in the 3D ^{15}N-HSQC-NOESY spectra. This assignment was subsequently confirmed on grounds of many NOEs between aromatic side chain ^{1}H atoms and Hβ and Hα atoms in 3D ^{13}C-HSQC-NOESY spectra recorded for both the aliphatic and aromatic range of carbon frequencies. The aromatic 3D ^{13}C-HSQC-NOESY spectrum was also used to obtain all ^{13}C chemical shifts of the aromatic ring systems and to resolved ambiguities in the ^{1}H resonance assignments arising from ^{1}H chemical shift degeneracy in the Zα residues W195, F146 and F130.

Assignment of side chain ^{15}N resonances

The side chain ^{15}N-^{1}H and ^{15}N-$^{1}H_2$ resonances of tryptophan, arginine, glutamine and asparagine residues occurring in the 2D ^{1}H-^{15}N HSQC spectrum (fig. 30), were assigned sequence specifically by analyzing their cross-peaks with the side chain ^{1}H atoms in a 3D ^{15}N-HSQC-TOCSY spectrum (70 ms mixing time). The aromatic imino moiety of W195 was assigned based on a NOE to the backbone H_N of W195 in the 3D ^{15}N-HSQC-NOESY (150 ms

mixing time). Side chain imino groups of arginines can be readily distinguished from backbone H_N groups by virtue of their distinct ^{15}N chemical shift of 80 - 85 ppm. In order to gain spectral resolution in the ^{15}N-dimension, these imino groups are conventionally 'folded' into the H_N region of the spectrum where they are recognized from their opposite sign if the correct number of folding operations is used. The side chain amides of the six glutamines and one asparagine of Zα could be identified by the intense cross-peaks in the 3D ^{15}N-HSQC-TOCSY spectrum reflecting strong geminal coupling between both amide protons. In NOESY spectra, the short distance between those amide protons causes rapid spin diffusion between them, thereby preventing unambiguous stereo-specific assignment of the Hδ21 and Hδ22 protons of asparagine and of the Hε21 and Hε22 protons of glutamine. However, this is not required because the same problem occurs with NOESY spectra used to derive distance information.

Assignment of prolines and of residues preceding prolines

Since proline residues lack a backbone amide proton, they cause a gap in all three sequential assignment strategies basing on the H_N proton. Knowing the primary sequence of the protein under study, this does not pose a problem because the residues surrounding prolines can be assigned sequence specifically first. The Cα, Cβ and ^1H resonances of P168 and P193 of Zα were readily obtained from the CBCA(CO)NH, HBHA(CO)NH and H(CCO)NH experiments utilizing the magnetization transferred to the H_N proton of the succeeding residue. The ^1H and ^{13}C resonances of P192 were retrieved from the HCCH-TOCSY and HCCH-COSY spectra without having prior knowledge of any chemical shift of this residue.

Of the residues preceding prolines, the Cα and Cβ chemical shifts were collected from the CBCANH spectrum, and the ^1H side chain resonances were recorded from the 3D ^{15}N-HSQC-TOCSY spectrum (70 ms mixing time). Missing ^{13}C and ^1H resonances were obtain from the HCCH-COSY and HCCH-TOCSY spectra, as described for the conventional side chain assignment.

The secondary structure of Zα

The secondary structure of a protein can be derived from two classes of NMR observables:

1) Sequential and medium-range NOE in the protein backbone
2) Backbone dihedral angles derived from coupling constants

α-Helices and β-strands have distinct sequential and medium-range backbone inter-proton distances (table 3) resulting in characteristic NOE pattern. α-Helices are identified by strong $d_{NN}(i, i+1)$ NOEs and medium or weak $d_{\alpha N}(i, i+3)$ and $d_{\alpha N}(i, i+4)$ NOEs. In contrast, β-strands show weak $d_{NN}(i, i+1)$ NOEs and lack $d_{\alpha N}(i, i+3)$ and $d_{\alpha N}(i, i+4)$ NOEs, but give rise to strong $d_{\alpha N}(i, i+1)$ NOEs. Regular α-helices can be distinguished from rare 3_{10}-helices if $d_{\alpha N}(i, i+4)$ NOEs are present which are unique to regular α-helices (table 3). Furthermore, β-sheets show strong long-range $d_{NN}(i, j)$, $d_{\alpha N}(i, j)$ and $d_{\alpha\alpha}(i, j)$ NOEs between hydrogen-bridged β-strands arising from short inter-proton distances (table 4).

table 4 **Short cross-strand distances in antiparallel and parallel β-sheets.**

Distance	antiparallel β-sheet	parallel β-sheet
$d_{NN}(i, j)$	3.3	4.0
$d_{\alpha N}(i, j)$	3.2	3.0
$d_{\alpha\alpha}(i, j)$	2.3	4.8

The aforementioned characteristic backbone NOEs were measured from the 3D ^{15}N-HSQC-NOESY spectrum (150 ms mixing time) of Zα (fig. 31) and classified by their intensity into strong, medium and weak (fig. 35). The Zα core domain (residues 136 – 198) shows three almost contiguous stretches of strong $d_{NN}(i, i+1)$ NOEs that are accompanied by many $d_{\alpha N}(i, i+3)$ and $d_{\alpha N}(i, i+4)$ NOEs. The NOE pattern in these three regions are indicative of regular α-helices, which have been sequentially labeled α1, α2 and α3 in fig. 35.

Furthermore, the Zα core domain shows three short continuous stretches of strong $d_{\alpha N}(i, i+1)$ NOEs, while $d_{\alpha N}(i, i+1)$ NOEs in other parts of Zα are only of weak to medium intensity. These three stretches show weak or medium $d_{NN}(i, i+1)$ NOEs and almost no $d_{\alpha N}(i, i+3)$ and $d_{\alpha N}(i, i+4)$ NOEs typical for α-helices. Thus, these three NOE pattern are characteristic for β-strands, which have been sequentially designated β1, β2 and β3 in fig. 35. In addition, the β-stranded structure of β2 and β3 is supported by a number of strong inter-strand NOEs between β2 and β3 consistent with an antiparallel β2β3-sheet (fig. 36). The H_N-H_N NOE between T156 of β1 and W195 of β3 (fig. 28) and the H_N-Hα NOE between T156 and L194 indicate that β1 and β3 are close in space. However, this does not allow one to deduce a parallel or antiparallel β-sheet orientation (fig. 36).

The NOE pattern at the termini of α-helices and β-strands is less indicative making it difficult to determine the very ends of the α-helices and β-strands unambiguously. For instance, the strong $d_{NN}(i, i+1)$ NOEs in the loop between α1 and β1 are due to the conformation and

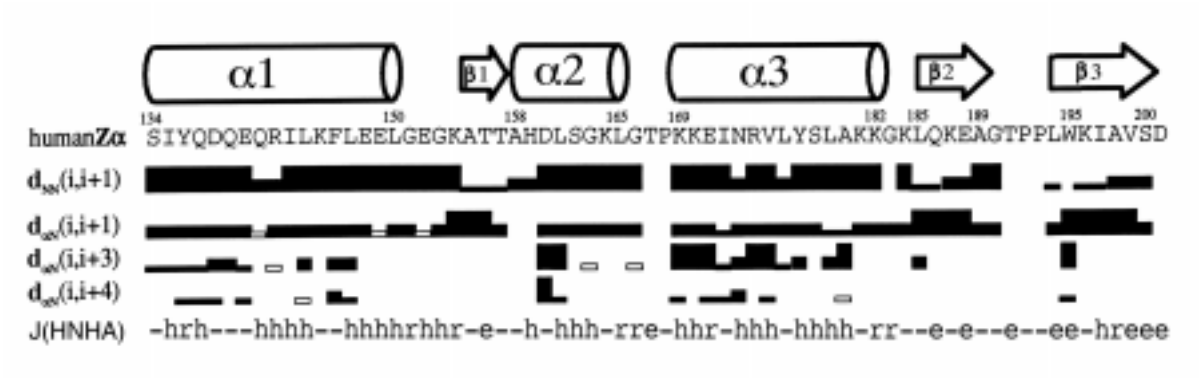


fig. 35 **The preliminary secondary structure of human Zα .** The sequence of Zα from residues 134-200 is shown. d_{NN} and $d_{\alpha N}$ represent backbone NH-NH- and Hα-NH-NOEs and are indicated by strong (thick bar), medium (intermediate bar) and weak (thin bar) intensity. $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ NOEs are characteristic for α -helices, and strong $d_{\alpha N}(i,i+1)$ NOEs are diagnostic of β-strands. The line J(HNHA) shows the quantitative analysis of backbone H_N-Hα coupling constants. Coupling constants less than 6 Hz are characteristic of α-helices (identified by an h) while those greater than 8 Hz suggest an extended chain conformation (marked with an e). Those residues with a coupling constant between 6 and 8 Hz are marked with an r representing random conformation.

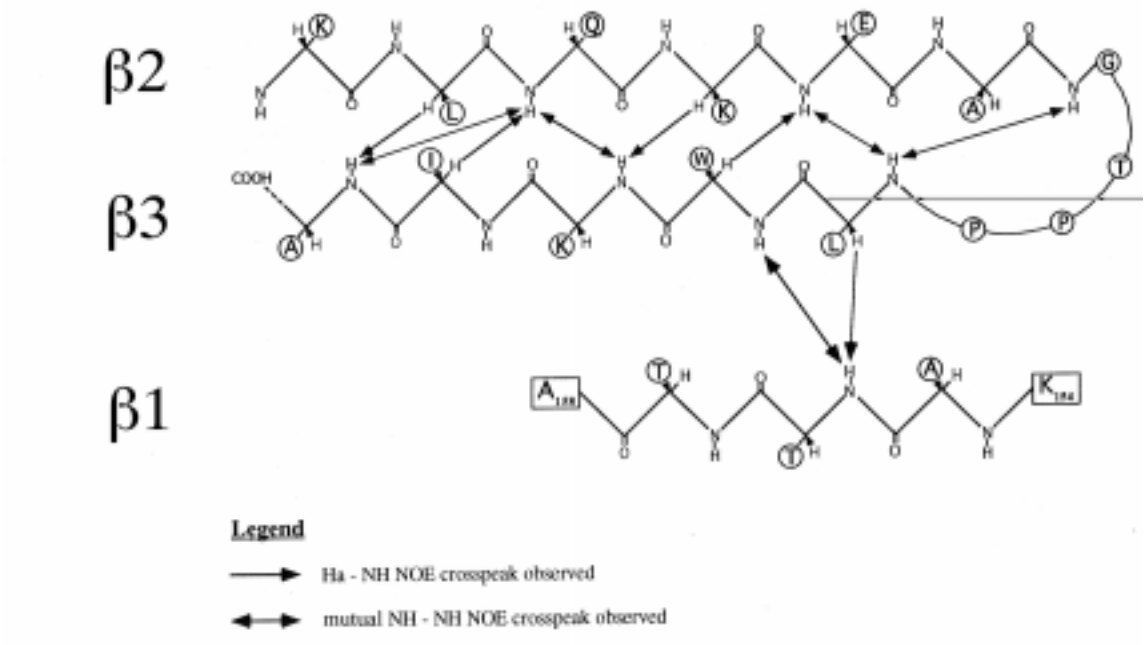


fig. 36 **Schematic representation of β-sheet NOEs in the backbone derived from a 3D ^{15}N-HSQC-NOESY** (150 ms). An almost complete set of $d_{NN}(i, j)$ NOEs (double-headed arrows) and $d_{\alpha N}(i, j)$ NOEs (single-headed arrows) was observed for the β-sheet between β2 and β3. Only two such NOEs reflect the interaction between β1 and β3 making it impossible to deduce the orientation between β1 and β3.

local flexibility of this loop. Local flexibility can lead to reduced line widths and thus produce apparently more intense NOE cross-peaks. On the contrary, the first two residues of α2 show weaker $d_{NN}(i, i+1)$ NOEs than other α-helical residues because these two H_N protons are not hydrogen bonded and exposed to solvent facilitating signal loss through chemical exchange. Furthermore, the pattern of $d_{\alpha N}(i, i+3)$ and $d_{\alpha N}(i, i+4)$ NOEs is incomplete in all α-helices because some of the five α-helical NOEs, $d_{\alpha N}(i, i)$, $d_{\alpha N}(i, i+1)$, $d_{\alpha N}(i, i+2)$, $d_{\alpha N}(i, i+3)$ and $d_{\alpha N}(i, i+4)$, overlap for most residues. Consequently, secondary structure analysis from sequential and medium-range NOEs allows one to determine the greater part of α-helices and β-sheets but leaves a twilight zone at their termini. The comparison with the high resolution NMR structure of Zα (fig. 47) shows that the N-terminus of α1 and the C-terminus of β3 are longer by 2 and 3 residues, respectively, in the preliminary secondare structure.

Determination of dihedral phi angles

Secondary structure determination from NMR coupling constants relies on the fact that α-helices and β-strands differ significantly in their backbone dihedral angles. The dihedral angle phi, which describes the rotation around the N-Cα bond in proteins, adopts values around –57, -60 and –119 to –139 degrees in α-helices, 3_{10}-helices and β-strands, respectively (table 5). Regions of random secondary structure have phi values between those of α-helices and β-sheets. Three-bond coupling constants are sensitive to the angular orientation of the three chemical bonds H-N-Cα-Hα, with a maximum if the three electronic orbitals are all in one plane (torsion angle = 180 degrees), and a minimum if the orbitals are orthogonal (torsion angle = 90 degrees). Applying the Karplus equation (see Appendix A, p. 110) parameterized for

$^3J_{HNH\alpha}$ couplings in proteins [194], this leads to the following relationship between phi angles and $^3J_{HNH\alpha}$ coupling constants in secondary structure elements (table 5):

table 5 **Relationship between phi angles in proteins and $^3J_{HNH\alpha}$ coupling constants** (adapted from [182].

Secondary structure element	phi [degrees]	$^3J_{HNH\alpha}$ [Hz]	I(cross-peak) / I(diagonal peak)
α-helix	- 57	3.9	0.13
3_{10}-helix	- 60	4.2	0.15
antiparallel β-sheet	- 139	8.9	0.99
parallel β-sheet	- 119	9.7	1.31

This table shows that phi angles, $^3J_{HNH\alpha}$ coupling constants and the ratio between cross-peak and diagonal peak intensities differ significantly for α-helices and β-sheets.

Experimentally, $^3J_{HNH\alpha}$ coupling constants of non-proline residues in ^{15}N-labeled proteins are readily accessible from a 3D HNHA spectrum, in which the intensity of a H_N diagonal peak, $I_{diagonal}$, and a Hα cross-peak, I_{cross}, relate to the $^3J_{HNH\alpha}$ coupling constant as given by [19,195]:

$$I_{cross} / I_{diagonal} = - \tan^2 (2\pi \xi \, {}^3J_{HNH\alpha})$$

The parameter ξ represents the de-/rephasing periods in the pulse sequence. The apparent coupling constants were multiplied by a factor of 1.11 to correct for Hα spin flips during the ξ periods [19,195]. Consequently, $^3J_{HNH\alpha}$ coupling constants can be calculated for all non-proline residues whose diagonal and cross-peak intensities are measurable.

In the HNHA pulse sequence, a certain amount of magnetization is transferred from the H_N proton to the Hα atom which depends on the size of the $^3J_{HNH\alpha}$ coupling constant. For residues in β-sheets, which possess a strong $^3J_{HNH\alpha}$ coupling (table 5), the intensity of the Hα cross-peaks becomes approximately as high as that of the H_N diagonal peaks of Zα (fig. 37, A). In contrast, for residues in α-helices, which have a weak $^3J_{HNH\alpha}$ coupling, the intensity of the Hα cross-peaks is considerably smaller than that of the H_N diagonal peaks (fig. 37, B). For some α-helical residues, the intensity of the Hα cross-peaks was barely measurable by peak integration, confining this method to protein samples with good signal-to-noise ratios.

For Zα, the $^3J_{HNH\alpha}$ coupling constants of 40 residues of a total of 85 residues could be determined by this method (see Appendix, table 15 for results). For the remaining 45 residues of Zα (marked with dashes in fig. 35), accurate peak integration was primarily hindered by overlap of either H_N diagonal peaks or Hα cross-peaks. On grounds of table 5, the $^3J_{HNH\alpha}$ coupling constants were categorized into the three groups, α-helical (< 6 Hz), random (6 – 8 Hz), and extended (> 8 Hz) conformation, and used to confirm the NOE-derived secondary structure assignment (fig. 35). $^3J_{HNH\alpha}$ coupling constants diagnostic of α-helices and β-strands cluster in the three α-helices and the three β-strands of Zα, respectively (fig. 35). However, α-helical coupling constants also occur in the loop between α1 and β1, and at I197 in β3. Likewise, coupling constants indicative of β-sheets are observed in the two loops between α2 and α3, and between β2 and β3. These coupling constants are misleading and do not help to overcome the problem of defining the termini of α-helices and β-strands. Thus, correct secondary structure assignment solely on the basis of $^3J_{HNH\alpha}$ coupling constants is impossible for Zα.

In order to use the $^3J_{HNH\alpha}$ coupling constants in structure calculation, they were converted into phi angles by using the Karplus equation parameterized for $^3J_{HNH\alpha}$ couplings in proteins [194]. The coupling constants were measured with an accuracy of approximately ± 1 to 2 Hz which is approximately equivalent with a phi angle accuracy of ± 30 degrees. A lower boundary of 3 Hz was used to correct for cases of very weak Hα cross-peaks in α-helices. The phi angles derived from coupling constants were checked for consistency with the three-dimensional Zα structure determined solely from NOE restraints. All of the 40 $^3J_{HNH\alpha}$ coupling constants are in agreement with the NOE-derived Zα structure when they are used with the above-mentioned phi angle tolerance of ± 30° in the structure calculation protocol.

The low precision with which the coupling constants can be measured makes it possible that misleading coupling constants with values characteristic for α-helices or β-strands occur within loops and segments of random conformation. Consequently, $^3J_{HNH\alpha}$ coupling constants

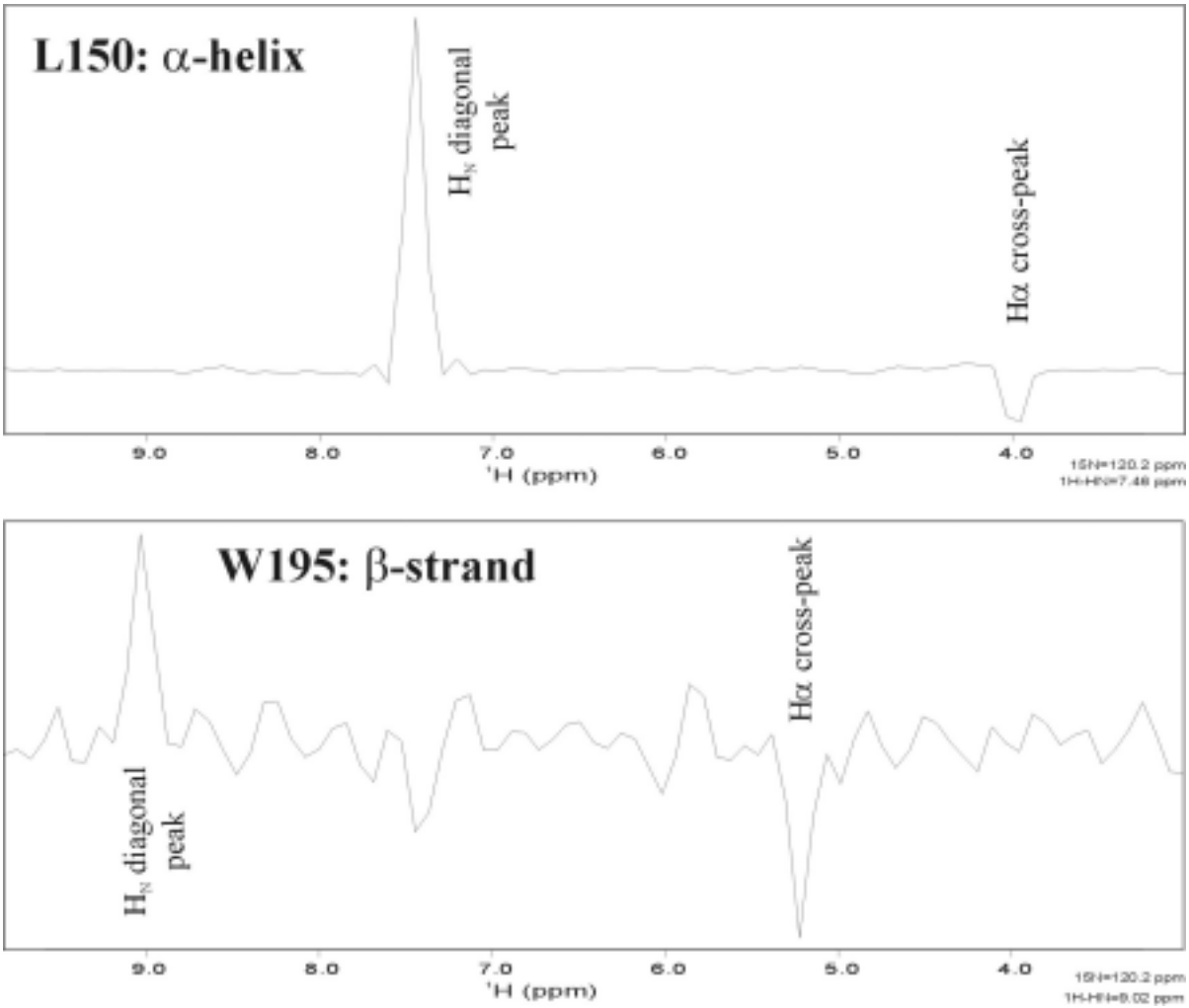


fig. 37 **Determination of $^3J_{HNH\alpha}$ coupling constants from a 3D HNHA spectrum** (ξ = 14 ms) of ^{15}N-labeled Zα. (A) Residues in β-strands, such as W195 of Zα, show almost equal intensities for H_N diagonal and Hα cross-peaks. (B) In contrast, α-helical residues, such as L150 in α1 of Zα, show Hα cross-peaks that are much less intense than the corresponding H_N diagonal peaks.

are useful for the validation of secondary and tertiary structures but do not provide high resolution structural information.

Summary

In high resolution NMR spectroscopy, almost complete resonance assignment is a prerequisite for structure determination. Backbone assignment of the Zα domain was achieved by using ^{15}N-separated HSQC-NOESY and CBCA(CO)NH/CBCANH-based triple resonance assignment strategies in combination. Side chains of Zα were principally assigned from the 3D triple resonance experiments HBHA(CO)NH and H(CCO)NH, and from the double resonance experiments HCCH-COSY/HCCH-TOCSY. These experiments allowed the resonance assignment of all Zα residues except for the N-terminal glycine.

Automated assignment routines have the potential to accelerate resonance assignment considerably. The side chains of Zα were assigned by a semi-automated two-step approach using the program Felix'97. Firstly, the side chain resonances were collected from 3D HBHA(CO)NH, HNHA and H(CCO)NH spectra by a fully automated routine for each spectrum. Secondly, the resulting table of resonances was manually inspected for missing or erroneous resonances and corrected by manual spectral analysis. The semi-automated assignment approach worked well with the well-resolved 3D spectra of Zα.

The secondary structure of Zα was determined from the analysis of characteristic interresidual and medium range NOEs which were obtained from a 3D ^{15}N-HSQC-NOESY spectrum (150 ms mixing time). $^{3}J_{HNH\alpha}$ coupling constants were useful to confirm the NOE-derived secondary structure. The Zα domain contains three α-helices and three β-strands giving rise to a α1β1α2α3β2β3 topology. Several termini of these α-helices and β-strands could not be determined with one-residue precision from these data. In particular, the N-terminus of α1 is poorly defined because the residues preceding tyrosine 136 show an ambiguous pattern of NOE signals in which NOEs diagnostic of α-helical structure are intercepted by NOEs diagnostic of random structure. After structure calculation, the 3D structure of Zα revealed that the Zα core domain comprises residues 136 – 198 and that residues 126 – 133 form a separate α-helix preceding α1 (designated prehelix).

The tertiary structure of Zα

Introduction

The determination of three-dimensional structures by solution state NMR spectroscopy relies primarily on the measurement of NOE distance restraints. Other NMR-accessible structural restraints, such as dihedral angles obtained from coupling constants, and the relative orientation of various chemical bond vectors derived from residual dipolar couplings [196], supplement the body of NOE restraints, but cannot define a three-dimensional structure of their own. Thus, the quality of a three-dimensional structure depends heavily on the number and accuracy of measured NOE restraints. For a high-resolution protein structure, an average of approximately 16 or more NOE restraints per residue are to be collected [188]. During structure calculation, this large number of short distance restraints forces a random polypeptide chain into a certain well-defined three-dimensional conformation.

The accuracy of a NOE-derived distance varies with the quality of a NOE cross-peak in a certain NOESY spectrum and with the type of ^{1}H atoms giving rise to a NOE cross-peak. Although the intensity of the NOE cross-peak is strongly dependent on the inter-proton distance (r^{-6} dependence under extreme narrowing conditions, see p. 112), chemical shift degeneracy, spin diffusion and spectral overlap often necessitate the allowance of broad distance tolerances of ± 1 Å or more. For instance, a pair of Hδ atoms on the aromatic ring of phenylalanine or tyrosine cannot be discriminated in most solution structures due to fast aromatic ring flips on the NMR time scale. Since these atoms are 4.3 Å apart, a NOE-derive distance including such Hδ atoms must take into account an effective pseudo atom correction of approximately + 2.0 Å and a correspondingly large distance tolerance [197]. In this case, the NOE restraint only defines the center of gravity (designated ‘pseudo atom’) of both Hδ atoms, but not their actual position. A similar, however smaller pseudo atom correction (+ 0.4 Å to + 1 Å) must be allowed for NOE restraints involving methyl groups, whose three ^{1}H atoms are magnetically equivalent by virtue of fast rotational averaging.

Spin diffusion lowers the precision of NOE distance measurements in cases where one or more ^{1}H atoms are close to the pair of ^{1}H atoms causing the NOE cross-peak. For example, NOE cross-peaks including one ^{1}H atom of a methylene moiety are often less intense than expected from the actual inter-proton distance because a certain share of magnetization has been moved on to the other methylene ^{1}H atom. This abundant effect can be minimized by choosing appropriately short mixing times, but remains a problem for configurations of ^{1}H atoms in which two or three short-distance NOE transfers compete with one long-distance NOE transfer.

Another abundant obstacle to the determination of accurate distance restraints is spectral overlap which is a more severe problem in NOESY spectra than in those spectra used for resonance assignment because of the larger number of cross-peaks. If NOE cross-peaks overlap their intensities cannot be measured accurately. As a result, such NOE cross-peaks are converted into qualitative rather than quantitative distance restraints with a very broad distance tolerance (1.8 – 6.0 Å) reflecting the physical limits of NOE distances. Taken together, most NOE distance restraints possess a distance precision of less than ± 1 Å.

Despite the comparatively low precision of many individual NOE restraints, they lead in summa to a well-defined three-dimensional structure with a sub-Ångstrom coordinate

precision. With more than 16 restraints per residues, each ^{1}H atom is constrained by more than one NOE restraint. Each NOE restraint defines the surface of a sphere with a thickness given by its distance tolerance (often more than 2 Å), which is also called lower and upper distance boundary. The ^{1}H atom is confined to the volume of intersection of these spheres, which is often of sub-Ångstrom size, resulting in a coordinate precision comparable to crystal structures. This holds in particular true for residues within the hydrophobic core which form NOE interactions with numerous ^{1}H atoms in the 1.8 – 6.0 Å surroundings. For instance, the L176.Hδ1 methyl group of Zα shows interresidual NOE cross-peaks with 41 ^{1}H atoms of which magnetically equivalent ^{1}H atoms (e.g. in methyl groups) were counted as one pseudo ^{1}H atom. In contrast, residues on the surface of proteins only show a few NOE cross-peaks leading to flexible side chains that are distributed over the allowed conformational space in the ensemble of solution structures. In these flexible parts of proteins, solid-state crystal structures and solution-state NMR structures are most distinctive reflecting the fact that a crystal 'freezes' a single out of many conformations during the crystallization process.

Another problem in structure determination using NMR spectroscopy is the consideration of stereo centers in amino acid side chains. Stereo-specific assignment is particularly important for the two methyl groups of valine and leucine residues, whose center-averaged H* positions are 2.8 Å apart [197]. A NOE distance restraint with the apparent L176.Hδ1* methyl group may be off by up to 2.8 Å if this methyl group is the real L176.Hδ2* methyl group. Stereo-specific assignment of ^{1}H atoms can be obtained from measuring coupling constants in the side chain that are distinctive for chiral ^{1}H atoms [18]. For methyl groups, stereo-specific assignment can be obtained from 2D ^{13}C-HSQC experiments on a fractionally ^{13}C-labeled sample [198]. Since these approaches require additional NMR experiments and additional NMR samples, the structure calculation of Zα was carried out with floating stereo-specific assignment [25-28], which allows one to forgo stereo-specific assignment. Floating stereo-specific assignment means that both chiral conformations are repetitively tested during structure calculation by dynamic simulated annealing, and that the lower energy stereo-assignment is kept. In other words, the force field permanently switches between both enantiomers, and the stereo-specific assignment 'floats' between both alternatives. Although the NOE data set lacks the stereochemical information in this approach, structure calculation using floating stereo-specific assignment results in high quality structures, as evidenced by the solution structure of Zα. This is a result of the large number of NOE restraints determining the stereochemistry of well-defined structures unambiguously.

Calculation of a template structure

The chemical shift information on ^{1}H, ^{13}C and ^{15}N resonances is insufficient to assign all NOE cross-peaks in 2D and 3D NOESY spectra because many chemical shifts are ambiguous. Therefore, only a small percentage of NOE cross-peaks can be assigned unambiguously at the beginning of a NOE assignment project. These include interresidual and medium-range NOEs in the protein backbone that were obtained from well-resolved 3D ^{15}N-HSQC-NOESY spectra and allowed the determination of the secondary structure (see p. 69). In addition, a small number of unambiguous long-range NOE restraints, which are essential for the determination of a preliminary model structure, can be obtained from well-resolved regions of NOESY spectra. For the Zα domain, 17 such long-range NOE restraints (table 6) were collected from a 3D ^{15}N-HSQC-NOESY spectrum (150 ms mixing time) and a 2D NOESY

spectrum (40 ms mixing time) in D_2O. Well-resolved 3D ^{13}C-HSQC-NOESY spectra were not available at that time.

table 6 **Long-range NOE in Zα.**

A. β-sheet NOE				**B. Fold restraining NOE**	
β2	**β3**	**β1**	**β3**		
L185.Hα	A198.NH	T156.NH	L194.Hα	W195.H6	Y177.H2/6
Q186.NH	A198.NH	T156.NH	W195.NH	W195.H5/7	L176.Hδ2
"	I197.Hα	W195.Hβ1/2	T157.NH	W195.H2	A158.NH
"	K196.NH			A158.Hβ	L176.Hδ2
K187.Hα	K196.NH			T156.NH	E152.Hα
E188.NH	W195.Hα			L165.NH	F146.H2/6
"	L194.NH			L165.Hδ1/2	L147.NH

The backbone NOE restraints observed in the C-terminal β-sheet and between β-strand β1 and β3 are summarized in column **A**. For instance, the Hα of L185 and the backbone H_N of A198 are connected through a NOE. **B.** Further unambiguous long-range NOEs that determine the tertiary fold, were obtained from 3D-^{15}N-separated NOESY-HSQC and 2D NOESY (D_2O) experiments. For each H_N-H_N NOE two independent cross-peaks were observed.

In conjunction with the secondary structure of Zα, these long-range restraints lead to the topology of Zα (fig. 38). Zα contains a double-stranded antiparallel β-sheet composed of β2 and β3, to which β-strand β1 aligns. The three α-helices are packed against this β-sheet. This arrangement of three α-helices and three β-strands is similar to the (α+β) helix-turn-helix fold [36] occurring in several DNA binding proteins, such as histone H5 and hepatocyte nuclear factor 3γ (HNF-3γ). Furthermore, the length of the α-helices, β-strands and loops is very similar for histone H5 and Zα (compare fig. 27) rendering the crystal structure of histone H5 a suitable template for the structure of Zα.

The homologous template structure was built by changing all residues of histone H5 to the equivalent residues of Zα, while maintaining the backbone coordinates of the crystal structure of histone H5. The loops between α-helices and β-strands were replaced with those of Zα. The stereochemistry and conformation of the remodeled loops was corrected by a few steps of local minimization using the program Sybyl6.4beta (Tripos Inc.). Finally, the whole structure of the template structure was remodeled by 20 runs of 10 ps simulated annealing calculations in a waterbox (force field AMBER 4.1), enforcing the known long-range NOE restraints of table 6, and keeping the backbone coordinates of α-helices and β-strands locked by using regular H-bond and dihedral angle constraints. The ensemble of template structures showed a high convergence, and the lowest energy structure was used to guide the semi-automated NOE assignment of Zα.

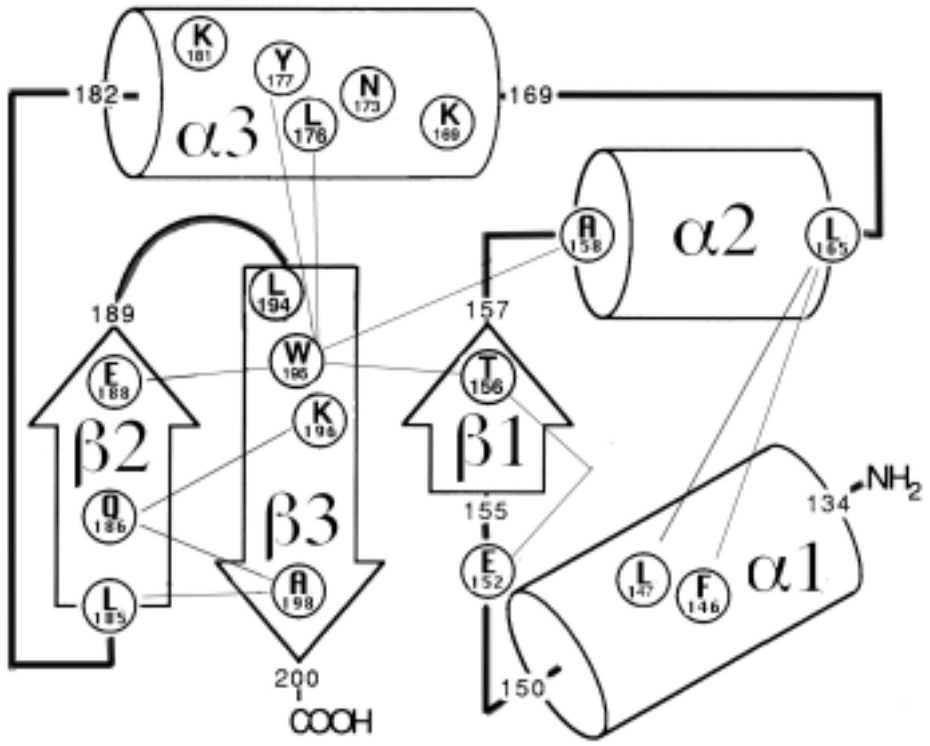


fig. 38 **Topology and candidate Z-DNA contacts of Zα.** The topology and location of candidate DNA contacts of Zα are similar to those of histone H5 and HNF-3γ [36]. α-Helices and β-strands are represented as boxed zylinders and arrows, respectively. Long-range NOEs are indicated with thin lines. By analogy with HNF-3, β1 is shown as antiparallel to β3. The relative orientation between β1 and β3 could not be determined unambiguously with the NOEs of table 6. Residues K169, N173, Y177 and K181 on α3 are candidate Z-DNA contacts, as suggested by mutagenesis.

The relative positioning of the α-helices and β-strands to each other poses the information input derived from the crystal structure of histone H5. The small number of 17 long-range NOE would be insufficient to yield a three-dimensional structure of comparable quality in a structure calculation attempt starting from the random polypeptide chain of Zα. Therefore, the usage of the histone H5 homology model was preferred over the conventional way of gradually collecting a larger number of unambiguous long-range NOE restraints prior to structure calculation.

Semi-automated NOE assignment

Using the chemical shift information and the template structure in conjunction, the majority of NOE cross-peaks can be assigned unambiguously. The role of the template structure in NOE assignment becomes clear if one considers a NOE cross-peak that may be consistent with five potential assignments based on the list of chemical shifts. If four of these assignments include pairs of ^{1}H atoms that are further apart in the template structure than a reasonable threshold (10 Å), these NOE assignments can be excluded, and the cross-peak can be assigned unambiguously. The threshold may adopt any value between the detection limit of NOE cross-peaks (6 Å) and the diameter of the Zα core domain (32.5 Å). A threshold of 10 Å

is a good compromise between the exclusion of as many irrelevant NOE assignment as possible and the allowance of a coordinate error of 4 – 8 Å in the template structure with respect to the real structure of Zα.

The automated NOE assignment routine of the Felix'97 software utilizes both the chemical shift and template structure information to make one out of three possible assignment decisions for each NOE cross-peak that has been peak-picked previously: no assignment, single assignment or multiple assignment. In the 3D ^{15}N- and ^{13}C-HSQC-NOESY spectra of Zα, approximately 50% of all NOE cross-peaks were assigned, and about 10 % received a single assignment. Single assignments were primarily obtained for intense intraresidual NOE cross-peaks, which are not required for structure calculation. These results show that many valuable long-range NOE cross-peaks received multiple assignments or were not assigned at all, despite the usage of a template structure.

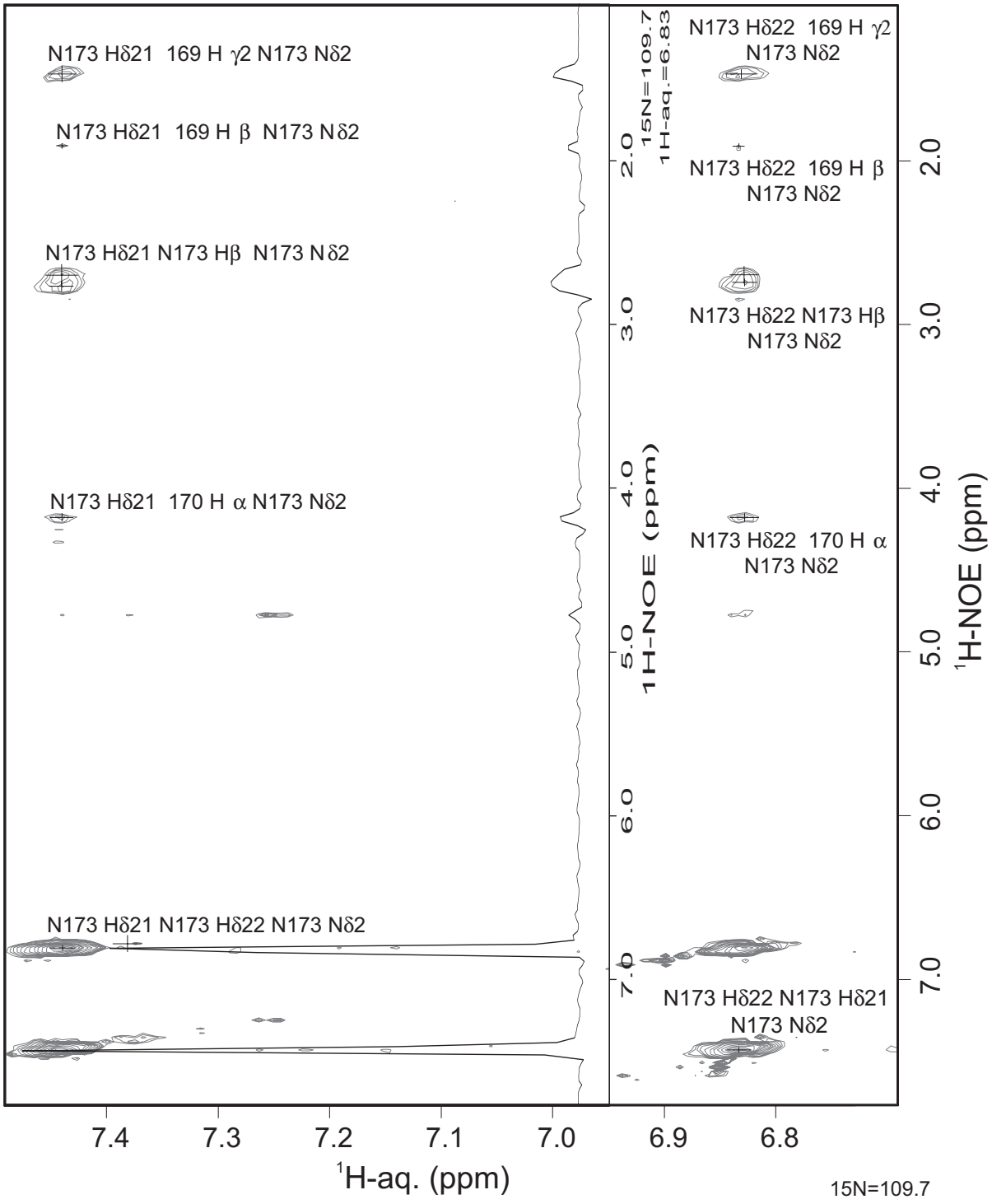
Multiple assignments are useful for focusing manual assignment on a few candidate assignments. In some cases, manual editing of these multiple assignments led to single assignments. In ambiguous cases, multiple assignments were retained or even complemented by further potential assignments, and subsequently used as ambiguous NOE restraints in structure calculation. In order to be less restricted by the template structure, the NOE distance threshold for multiple assignments was set to a more conservative value of 12 Å.

Even though well-separated 3D NOESY spectra were exclusively used for the first rounds of NOE assignment, spectral overlap in the aliphatic region of these spectra caused problems for the automated NOE assignment routine. Approximately 10 – 20 % of the single and multiple assignments were erroneous, and many important long-range NOE cross-peaks in the aliphatic region were missed. Erroneous assignments were often found if the true center of a cross-peak was shifted by overlap with other cross-peaks. Missing assignments occurred principally for overlapped cross-peaks that exceeded the line width cut-off, which was defined on the basis of a small number of well-resolved cross-peaks for each spectrum. In consequence, the majority of NOE cross-peaks had to be manually edited. The automated NOE assignment provided a useful pre-selection of NOE assignments rather than reliable assignments that can be directly used for structure calculation.

3D ^{15}N-HSQC-NOESY spectra

The Nuclear Overhauser Effect (NOE) leads to a linear build-up of cross-peak intensity as a function of mixing time in NOESY spectra. With increasing mixing time, the cross-peak intensity levels off to reach a maximum and decreases thereafter. In order to optimize signal-to-noise and minimize spin diffusion, the mixing time is to be chosen so that the intensity of the majority of NOE cross-peaks is in the upper linear range of the NOE build-up curve. For this reason, ^{15}N-HSQC-NOESY spectra were recorded with three different mixing times, 70, 150 and 250 ms. The 70 ms spectrum contained less than half as many NOE cross-peaks as the 150 ms spectrum. The cross-peak intensity in the 70 ms spectrum was significantly lower than in the 150 ms spectrum indicating that a mixing time of 70 ms is too short for ^{15}N-HSQC-

fig. 39 **NOE cross-peaks in a ^{15}N-HSQC-NOESY spectrum (see next page)** (150 ms mixing time) between N173.Hδ21/Hδ22 and neighboring residues that rigidify the side chain of the surface residue N173. An additional cross-peak with water suggests either chemical exchange or the presence of a H_2O molecule with a long residence time (> 0.5 ns) [199].


N173 Hδ21 169 H γ2 N173 Nδ2
N173 Hδ21 169 H β N173 Nδ2
N173 Hδ21 N173 Hβ N173 Nδ2
N173 Hδ21 170 H α N173 Nδ2
N173 Hδ21 N173 Hδ22 N173 Nδ2
N173 Hδ22 169 H γ2 N173 Nδ2
N173 Hδ22 169 H β N173 Nδ2
N173 Hδ22 N173 Hβ N173 Nδ2
N173 Hδ22 170 H α N173 Nδ2
N173 Hδ22 N173 Hδ21 N173 Nδ2
15N=109.7 1H-aq.=6.83
1H-NOE (ppm)
2.0
3.0
4.0
5.0
6.0
7.0
7.4
7.3
7.2
7.1
7.0
6.9
6.8
1H-aq. (ppm)
15N=109.7

NOESY spectra of Zα. The 250 ms spectrum which contains only a few more NOE cross-peaks than the 150 ms spectrum, showed cross-peak intensities similar to those of the 150 ms spectrum. This comparison suggests that the spectrum with 150 ms mixing time shows the best signal-to-noise ratio in conjunction with the least spin-diffusion of these three ^{15}N-HSQC-NOESY.

H_N-H_N, H_N-Hα and H_N-Hβ NOE cross-peaks are well-resolved in the 150 ms ^{15}N-HSQC-NOESY spectrum of Zα, making it the most important source of NOE distance restraints in the protein backbone, as has been demonstrated in the paragraph on 'The secondary structure of Zα'. In addition, a number of NOE cross peaks were observed for the side chain amides and imides of asparagine, glutamine, arginine and tryptophan residues. For instance, the two side chain amide protons of N173 show NOE cross-peaks to ^{1}H atoms of three other residues (fig. 39) that restrict the flexibility of this amide group. The potential biological relevance of this will be discussed later (see page 96).

Although a large number of H_N–aliphatic side chain (Hγ, Hδ and Hε) cross-peaks are detected in the ^{15}N-HSQC-NOESY spectrum (see fig. 31), many of these cannot be assigned unambiguously due to spectral overlap in the aliphatic region. Therefore, another NOESY experiment is required that better resolves aliphatic resonances.

3D ^{13}C-HSQC-NOESY spectra

The 3D ^{13}C-HSQC-NOESY experiment, which disperses the crowded aliphatic ^{1}H-^{1}H region into a third ^{13}C dimension, is pivotal for collecting NOE restraints between aliphatic side chain ^{1}H atoms. Of the three 3D ^{13}C-HSQC-NOESY spectra recorded at 40, 70 and 100 ms mixing time, the 70 ms spectrum showed the best signal-to-noise ratio and minimal spin-diffusion. A 2D ^{1}H-^{13}C plane from this spectrum demonstrates how NOE cross-peaks involving T191.Hγ2* and A198.Hβ* are separated along the ^{13}C axis, which would overlap in a conventional 2D NOESY spectrum (fig. 40). In this way, numerous NOE distance restraints were obtained from the ^{13}C-HSQC-NOESY spectrum that were obscured in non-^{13}C-separated NOESY spectra.

However, many aliphatic cross-peaks are still overlapped in 3D ^{13}C-HSQC-NOESY spectrum because of the small ^{13}C chemical shift dispersion of aliphatic Cγ, Cδ and Cε resonances. The spectral congestion was particularly troublesome for automated NOE assignment necessitating time-consuming manual editing of the majority of cross-peaks in the aliphatic region.

In order to further simplify the 3D ^{13}C-HSQC-NOESY spectrum, a heteronuclear quadruple-quantum coherence (HQQC) filtered 3D ^{13}C-NOESY [200] spectrum was recorded that selects for NOE cross-peaks involves at least one methyl group. Due to the multiple quantum filter, the HQQC-NOESY experiment suffers from a reduced signal-to-noise ratio, as compared with conventional ^{13}C-HSQC-NOESY spectra. Moreover, overlap of methyl-methyl cross-peaks, which is a central problem for Zα containing 11 leucines and four valines, is not resolved. Therefore, the HQQC-NOESY experiment was only of limited use in the case of Zα.

Aromatic ^{13}C resonances are recorded in a separate ^{13}C-HSQC-NOESY spectrum because aromatic and aliphatic ^{13}C resonances are far apart on the chemical shift scale. Furthermore, the NOE build-up curve for aromatic ^{1}H cross-peaks is shifted to shorter NOESY mixing times. Therefore, aromatic 3D ^{13}C-HSQC-NOESY spectra were recorded at shorter

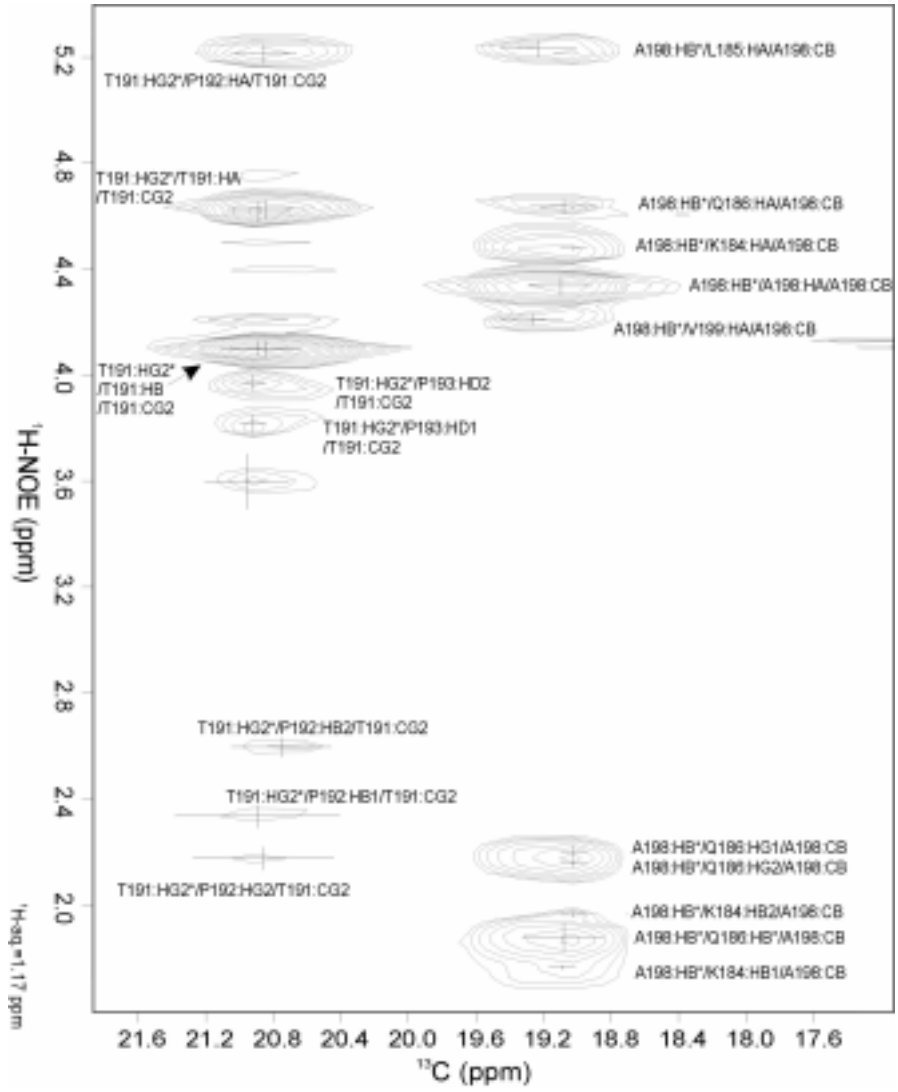


fig. 40 **2D ^{1}H-^{13}C plane from a 3D ^{13}C-HSQC-NOESY spectrum** (70 ms mixing time) shows that the Hγ2* methyl group of T191 and the Hβ* methyl group of A198, which possess the same ^{1}H chemical shift (1.17 ppm), are separated in the third dimension by virtue of their distinct ^{13}C chemical shifts. In a 2D NOESY, both columns of NOE cross-peaks fall on top of each other, thereby obscuring a large number of these NOE restraints.

mixing times of 35 and 70 ms. Intense NOE cross-peaks corresponding to short ^{1}H-^{1}H distances, were collected from the 35 ms spectrum to exclude marked spin-diffusion. NOE cross-peaks of medium and weak intensity corresponding to long ^{1}H-^{1}H distances, were measured in the 70 ms spectrum because many of them were barely detectable or absent in the 35 ms spectrum.

2D NOESY spectra in D_2O and H_2O

3D HSQC-NOESY experiments bring about the disadvantage that inter-proton distances cannot be accurately quantified from cross-peak intensities because of spin diffusion

and differential INEPT magnetization transfer efficiencies [201]. Such quantitative distance restraints can be obtained from well-resolved regions of homonuclear 2D NOESY experiments. These regions include H_N-H_N cross-peaks in the backbone, cross-peaks with aromatic ring ^{1}H atoms (fig. 41) and other well-resolved cross-peaks.

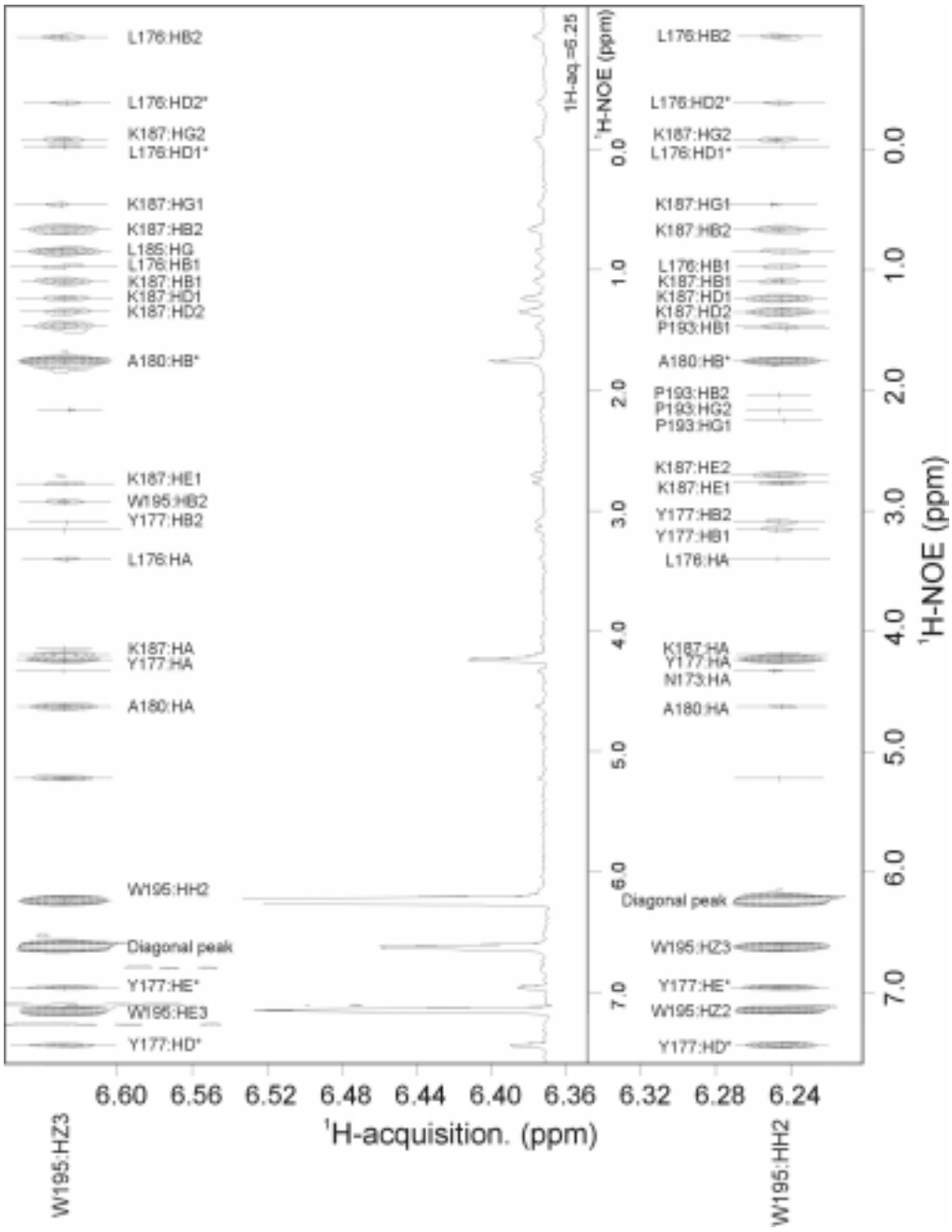


fig. 41 **Close-up of a 2D NOESY in D_2O** (40 ms mixing time) showing NOE cross-peaks to the two well-resolved aromatic ^{1}H atoms Hz3 and Hh2 of W195. The spectrum shows that more than 20 NOE cross-peaks could be assigned for these two aromatic tryptophan ^{1}H atoms within the hydrophobic core of Zα. The very weak cross-peaks which are represented by one contour line or less, are below the detection limit of the 3D NOESY spectra. Thus, the analysis of well-resolved regions in 2D NOESY spectra provides additional NOE distance restraints.

Accurate H_N-H_N distances allow one to set the distance tolerance to less than 0.6 Å for clean cross-peaks, thereby improving the stereochemical quality of α-helices and β-sheets which often contain irregularities in unrefined solution structures. From the well-dispersed 2D NOESY spectra of Zα, a large number of NOE distance restraints could be collected confirming and supplementing those restraints previously derived from 3D ^{15}N- and ^{13}C-HSQC-NOESY spectra.

Calibration of distance restraints

The intensity of a NOE cross-peak is proportional to the inverse sixth power of the inter-proton distance within the linear range of the NOE build-up curve (under extreme narrowing conditions). From comparison with the intensity, I_{ref}, of an appropriate reference cross-peak, whose inter-proton distance, r_{ref}, is known, an unknown inter-proton distance, r, can be calculated from its cross-peak intensity, I, as given by:

$$r = r_{ref} (I_{ref} / I)^{1/6}$$

Since this equation is only a practicable approximation of the complete relaxation matrix description of NOE cross-peak intensities, systematic deviations occur in practice. In general, inter-proton distances longer than the reference distance are underestimated, and inter-proton distances shorter than the reference are overestimated [202]. Underestimated distance restraints are problematic because they over-constraint the protein structure. Therefore, reference distances of intermediate length, such as the $d_{\alpha N}(i, i+3)$ distance (3.4 Å) in α-helices, were chosen for the calibration of NOESY spectra.

The relaxation rates for individual types of ^{1}H atoms (e.g. H_N, $H_{aliphatic}$, $H_{aromatic}$) may differ resulting in varying apparent NOE intensities for similar inter-proton distances. The reference cross-peak should thus be of the same type as the most important NOE cross-peaks in a NOESY spectrum. For example, if H_N-Hα and H_N-Hβ distances are to be measured, the reference cross-peak should also be of the type H_N-Hα or H_N-Hβ provided that reference

Hα of residue:	H_N of residue:	reference distance [Å]
K169	K170	3.5
K170	N173	3.4
E171	I172	3.5
I172	N173	3.5
I172	V175	3.4
N173	L176	3.4
R174	V175	3.5
R174	Y177	3.4
V175	L176	3.5
V175	S178	3.4
L176	L179	3.4
Y177	A180	3.4
S178	L179	3.5
L179	A180	3.5

table 7 **Reference distances for ^{15}N-HSQC-NOESY spectra.** 14 well-resolved $d_{\alpha N}(i, i+1)$ and $d_{\alpha N}(i, i+3)$ NOE cross-peaks in helix α3 of Zα were used to calibrate distance restraints in the 3D ^{15}N-HSQC-NOESY spectrum of Zα (150 ms mixing time). The average of these cross-peak intensities weighted by the reference distance was used to calculate the calibration constant.

distances of this type are available. For the solution structure of Zα, integrated cross-peak intensities from the 3D ^{15}N-HSQC-NOESY spectrum were calibrated with $d_{\alpha N}(i, i+1)$ and $d_{\alpha N}(i, i+3)$ NOE cross-peaks in helix α3 (table 7). For aliphatic 3D ^{13}C-HSQC-NOESY spectra, calibration was achieved by using the Hα-Hβ* cross-peaks of the four alanine residues taking into account a suitable multiplicity correction (apparent reference distance: 2.13 Å). Cross-peaks of the aromatic 3D ^{13}C-HSQC-NOESY spectrum with 35 ms mixing time were calibrated with NOEs between the vicinal aromatic ring ^{1}H atoms of W195 (2.48 Å). The 70 ms ^{13}C-HSQC-NOESY spectrum was referenced with NOE cross-peaks across this tryptophan ring (4.3 Å). The homonuclear 2D NOESY spectra in H_2O and D_2O were calibrated with analogous reference distances as the 3D ^{15}N-HSQC-NOESY and aliphatic ^{13}C-HSQC-NOESY spectra, respectively.

Calculated NOE distance restraints are afflicted with both random and systematic errors that dependent on the magnitude of the cross-peak intensity and thus the inter-proton distance. Therefore, a first-order error allowance was applied to raw NOE distance restraints that subtracts and adds a certain percentage value to the calculated distance value yielding a lower and upper distance boundary. This ensures that short distance restraints derived from a cross-peak with an excellent signal-to-noise ratio and a small systematic error, receive small tolerances, whereas long distance restraints receive broad tolerances.

^{1}H atoms of methyl groups and of chemical equivalent moieties, such as degenerate methylene groups, are commonly represented by a pseudo-atom at the center of gravity of the ^{1}H atoms involved. However, a pseudo-atom does not consider that a subset of its ^{1}H atoms is closer to the interacting ^{1}H atom that causes the NOE. Consequently, the NOE-derived distance between the virtual pseudo-atom and the interacting ^{1}H atom is underestimated and has to be corrected. For Zα, the pseudo-atom corrections listed in table 8 were added to the upper boundary of the distance restraints yielding a NOE data set that was used directly for structure calculation. Multiplicity correction, which corrects for the fact that more than ^{1}H atom contributed to the cross-peak intensity of such an NOE, was carried out automatically by the Felix'97 program.

table 8 **Pseudo-atom corrections for groups of equivalent or non-stereo-assigned ^{1}H atoms** (adapted from [197]).

Equivalent group	Pseudo-atom correction [Å]
-CH_3	1.0
-CH_2	1.0
aromatic Hδ and Hε	2.0
-$CH(CH_3)_2$	2.4
non-regio-assigned -$CONH_2$ (in N, Q)	1.0

Structure calculation

Structure calculation combines the experimental data set of NOE distance restraints and of dihedral angle restraints with the chemical data set containing bond lengths, bond angles, stereochemistry and van der Waals repulsion. For the conversion of this information into a set of three-dimensional structures, two principal computational methods are used, distance

geometry (DG) and restrained molecular dynamics (MD), which is also known as dynamic simulated annealing.

Distance geometry

In distance geometry, the distance restraints are used to set-up a metric matrix of upper and lower distance bounds with one set of bounds for each pair of atoms within the protein. Unknown distance bounds, which include all experimentally inaccessible distances, for example distances longer than 6 Å, are filled with the van der Waals radius for the lower bound and an arbitrarily large distance for the upper bound. These unknown distances are calculated by using the triangle inequality:

$$d_{ij} <= d_{ik} + d_{kj}$$

An unknown distance, d_{ij}, between atom i and j must be shorter than or of equal length as the sum of two known distances between atom i and k, d_{ik}, and between atom j and k, d_{kj}. In this manner, unknown distances are gradually replaced by the sum of two known distances leading to a range of three-dimensional structures in distance space. From this data, trial distances are selected in such a way that the conformational space is well sampled. These distances are converted into cartesian coordinates to give an ensemble of three-dimensional structures.

A recent advancement of distance geometry operates in torsional space rather than distance space, which drastically reduces the number of variables in the calculation. It uses restrained minimization of a variable target function and is computationally less expensive than metric matrix distance geometry and dynamic simulated annealing. Drawbacks of distance geometry include insufficient sampling of conformational space and overall expanded structures with respect to crystal structures of the same molecule [203].

Dynamic simulated annealing

Dynamic simulated annealing computationally mimics protein folding from a denatured conformation at high temperatures (2000 K) to a native conformation at low temperature (50 K). The target function, which is to be minimized during simulated annealing, is given by the potential energy function, which is composite of energy terms for chemical bond lengths (E_{bond}), bond angles (E_{angle}), improper angles ($E_{improper}$), van der Waals repulsion ($E_{vdWaals}$), NOE distance restraints (E_{NOE}) and dihedral angles restraints ($E_{dihedral}$) derived from NMR coupling constants:

$$E_{tot} = E_{bond} + E_{angle} + E_{improper} + E_{vdWaals} + E_{NOE} + E_{dihedral}$$

In order to determine the trajectory of the structure, the force acting on each atom at a given time is calculated by differentiating the potential energy function with respect to the three spatial coordinates. Applying Newton's equation of motion (F = m*a), the velocity of each atom can be calculated by numerical integration over very small time steps. The velocity allows one to calculate the displacement of each atom during these time steps. A large number of such successive displacements allows a denatured protein of random conformation to fold into a loose native conformation guided by the experimentally NOE and dihedral angle restraints. The

high temperature (2000 K) in the initial phase ensures that displacements occur fast due to high velocities and that local energy minima are jumped over. During the subsequent cooling phase, the displacements gradually become smaller, and the loose structure adopts a tight low energy conformation consistent with all restraints. Finally, minimization corrects local minute deviations from optimal chemical geometry.

The structure calculation of Zα was performed by dynamic simulated annealing in the force field PARALLHDG.PRO with the program XPLOR 3.1 [24] using asymptotic soft-square potentials and the experimental restraints summarized in table 9. The simulated annealing protocol begins with 50 steps of Powell energy minimization of the initial randomized coordinates. The protein was thereafter heated to 2000 K and allowed to rearrange during 30000 steps (1 step = 5 fs) of simulated annealing. The cooling phase consisted of two steps: firstly, 3800 steps of simulated annealing to arrive at a temperature of 1000 K, and secondly, 1500 steps of simulated annealing to reach the final temperature of 50 K. The final Powell minimization included 500 steps.

table 9 **Summary of the experimental NOE distance and dihedral angle restraints** used for the calculation of the final ensemble of Zα structures.

Distance restraints from NOE	
Interresidue sequential (|i-j|=1)	695
Interresidue medium range (1<|i-j|<5)	669
Interresidue long range (|i-j|>4)	800
All interresidual NOE restraints	2164
Dihedral angle restraints from coupling constants	47

In order to sample the entire conformational space allowed to the protein by the experimental constraints, a total of 200 structures were computed by dynamic simulated annealing from randomized coordinates. 70% of the 200 structures showed a total energy only marginally higher than the lowest energy, while about 10 % of those structures showed total energies higher by more than an order of magnitude than the lowest energy (fig. 42). This indicates that the initial coordinates were efficiently randomized so that not all of the loose high temperature structures found the absolute minimum of total energy, but were trapped in local minima of elevated energy.

Iterative structure refinement

NOE assignment and structure calculation are iterative processes that are gradually refined in order to finally arrive at the best available fit between the body of NOE data and its structural interpretation. After the first round of structure calculation using exclusively NOE restraints from 3D NOESY spectra, numerous NOE restraints were in disagreement with the corresponding inter-proton distances in the calculated structure. Inspecting these violating NOE restraints in the NOESY spectra showed that they were erroneously assigned or that their intensity was erroneously quantified. This was caused either by wrong automated NOE assignment, incomplete manual editing or inaccurate automated peak integration. After these erroneous restraints had been removed or corrected, another ensemble of simulated annealing structures was calculated. The refined structure was used as the new template structure for automated and manual NOE assignment. Again, heavily violating restraints were verified on

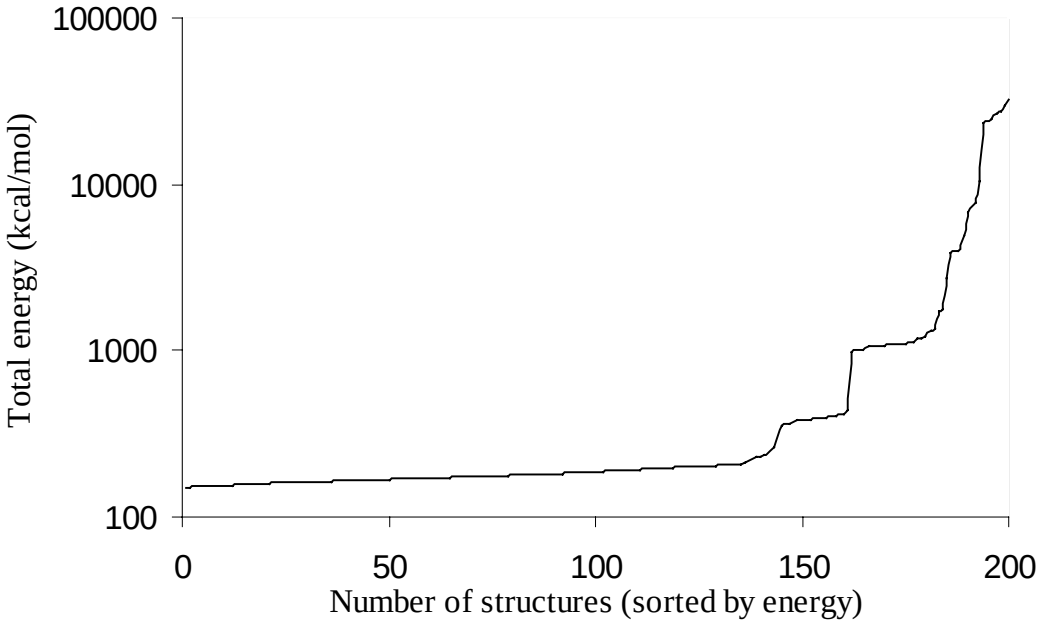


fig. 42 **The total energies of the final 200 structures of Zα**. About 70 % of the 200 Zα structures (residues 117 – 201) calculated by simulated annealing from random coordinates show a total energy close to that of the lowest energy structure. About 10% of all structures possess a total energy higher by more than an order of magnitude than the lowest energy. This indicates efficient sampling of the conformational space during structure calculation.

the spectral level and dismissed or amended. This iterative cycle of structure refinement was continued until no structure of the ensemble of the 15 lowest energy structures showed a NOE distance violation greater than 0.2 Å or a dihedral phi angle violation greater than 2°.

The dihedral phi angle restraints and the restraints from the 2D NOESY spectra, which were obtained by semi-automated NOE assignment using a pre-refined structure, were added to the NOE data set in the late rounds of refinement. In addition, 24 hydrogen bonds in the three α -helices and 4 hydrogen-bonds in the C-terminal β-sheet of Zα were included in the late rounds of refinement in order to improve the geometry of the α-helices and β-sheet. Intraresidual restraints were omitted because they did not improve the precision of the structure.

Precision and quality of the final ensemble of Zα structures

The final ensemble of the 15 lowest energy structures of Zα (residues 117-201) (fig. 43) superimposed on the mean structure, shows a very good agreement in the backbone trace of the Zα core domain (residues 136–198). The backbone rms deviation about the mean structure of the Zα core domain is 0.26 Å omitting one flexible loop (residues 151-153). The rms deviation of all non-hydrogen atoms is 0.65 Å including both rigid side chains within the hydrophobic core and flexible side chains on the surface of Zα. The excellent coordinate precision is due to the large number of NOE restraints with an average of 12.7 long-range NOE restraints per residue of the core domain.

The stereochemistry of the final ensemble of 15 Zα structures (residues 136-198) was checked by Ramachandran analysis using the program PROCHECK [204]. 92.6 % of all

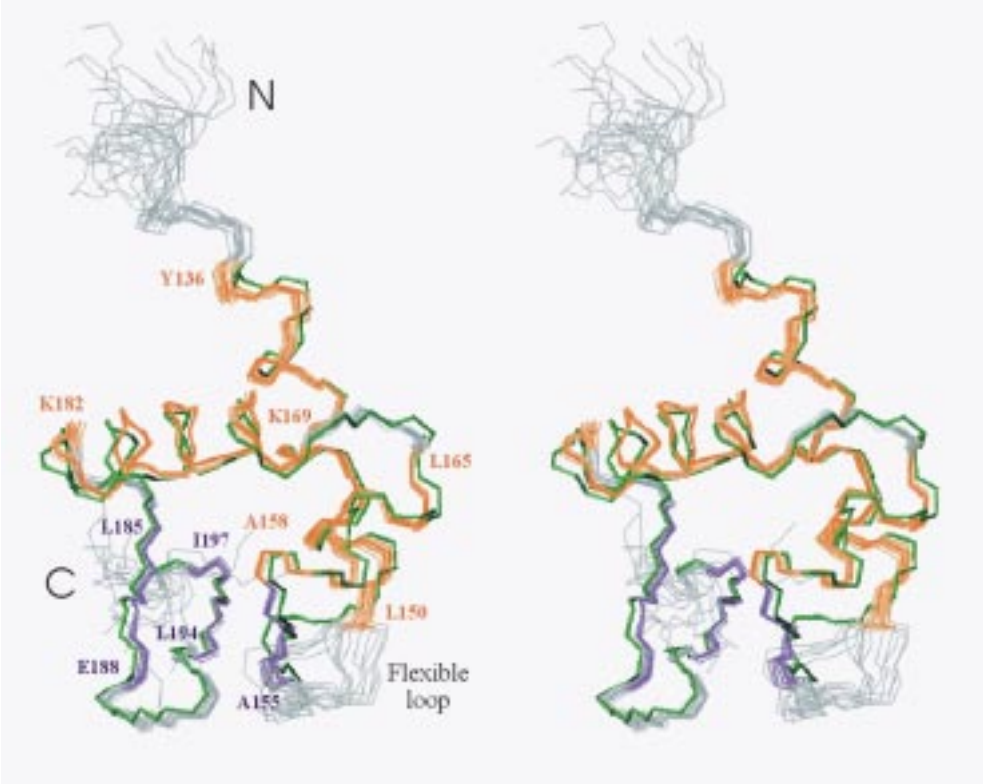


fig. 43 **Ensemble of 15 Zα structures.** The stereo view of the backbone atoms of the 15 lowest energy structures (residues 126-200) shows that Zα possesses a rigid backbone with one flexible loop between α1 and β1. The terminal residues of α-helices (orange) and β-strands (violet) are labeled as well as the N- and C-terminus. For comparison, the crystal structure of bound Zα (green) was superimposed on the lowest energy structure of unbound Zα showing that both structures are almost identical except for one flexible loop between α1 and β1.

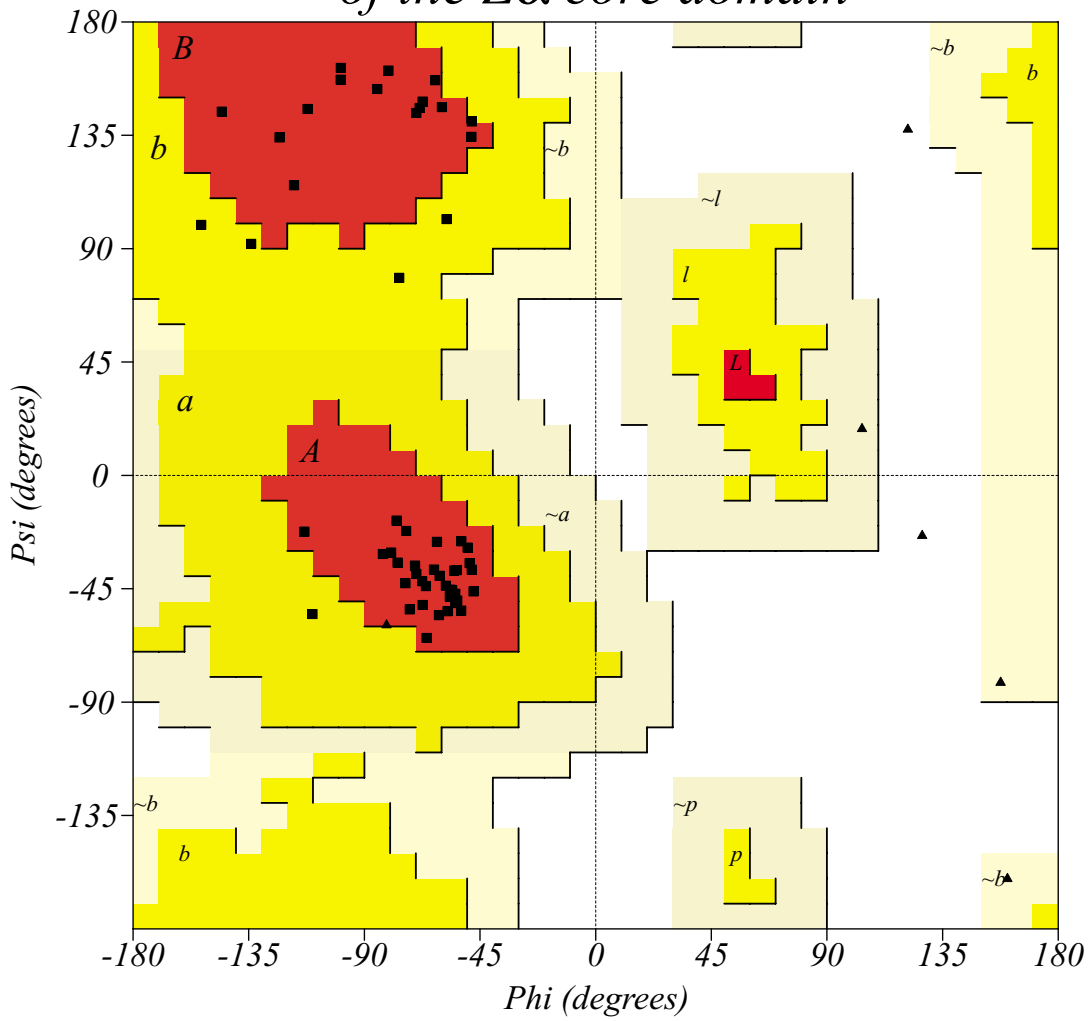
residues of the Zα core domain are found in the most favored regions of the Ramachandran plot, and the remaining 7.4 % of the residues occur in the additional allowed regions. In the lowest energy structure of Zα, 88.5 % and 11.5 % of all residues occupy the most favored and additionally allowed regions of the Ramachandran plot (fig. 44), respectively. These data show that the solution structure of Zα is of a high precision and stereochemical quality.

Another criterion for well-refined high-quality structures are low potential energies. The XPLOR potential energies of the ensemble of 15 Zα structures (residues 136-198) show that the total energy of 143.6 ± 10.1 kcal/mol is within the range expected for high-quality structures (table 10). The potential energies for experimental dihedral angle (E_{cdih}) and NOE

fig. 44 **The Ramachandran plot of the lowest energy structure of the Zα core domain (see next page)** shows a high stereochemical quality of the Zα backbone with 46 residues of a total of 52 non-glycine and non-proline residues occupying the most favored regions of the plot. Only 6 residues lie in the additional allowed regions. The α-helices show particularly high stereochemical quality with only one residue in the additional allowed α-helical region (intermediately gray area labeled with an ‘a’). The most favored regions of α-helices and β-strands are shown in dark gray and marked with an ‘A’ and a ‘B’, respectively. Each residue is represented by a black square except for glycines which are shown as black triangles.

PROCHECK

Ramachandran Plot
of the Zα core domain



Plot statistics

Residues in most favoured regions [A,B,L]	46	88.5%
Residues in additional allowed regions [a,b,l,p]	6	11.5%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	52	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	6	
Number of proline residues	3	
Total number of residues	63	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

(E_{NOE}) restraints are particularly low consistent with a very small average NOE violation of 0.008 Å. The energy of chemical bond angles (E_{angle}) is the highest single energy term suggesting that a number of bond angles adopt non-ideal values. For example, the chemical bonds around the *cis* peptide bond of proline 192 possess uncommon angles. The high bond angle energy may be partially due to non-ideal pairs of $\chi 1$-$\chi 2$ angles which received a sub-optimal score in the coordinate analysis with the programs PROCHECK [204] and WHAT-IF [205]. In addition to the analysis of the XPLOR potential energies, the Lennard-Jones potential energies of the ensemble of 15 structures were calculated with the CHARMM PARAM19/20 force field (Quanta97, MSI Inc.) resulting in a normal van der Waals energy of –69.8 ± 13.2 kcal/mol.

table 10 **XPLOR potential energies and coordinate precision of the final ensemble of 15 Zα structures.** The potential energies for experimental NOE and dihedral angle restraints are particularly low, while the energy for chemical bond angles is less optimal. The Ramachandran analysis and the coordinate precision are indicative of a high quality solution structure.

XPLOR potential energies (kcal mol^{-1})		
E_{total}	143.6	± 10.1
E_{bond}	5.5	± 0.6
E_{angle}	102.5	± 7.2
$E_{improper\ angles}$	11.5	± 1.0
E_{repel}	16.2	± 1.7
E_{NOE} [2]	7.9	± 0.8
E_{cdih} [2]	0.02	± 0.015
PROCHECK Ramachandran analysis [2,3]		
Residues in most favored regions	92.6 %	
Residues in additional allowed regions	7.4 %	
Residues in generously or disallowed regions	0 %	
Lennard-Jones potential energies [3] (kcal mol^{-1}) [4]		
$<SA>_{15\ structures}$ [5]	-69.8	± 13.2
Coordinate precision (Å) [6]		
Backbone (N, Cα, C, O)	0.26	
All non-hydrogen atoms	0.65	

[2]No structure shows NOE distance violations greater than 0.2 Å or phi angle violations greater than 2°. [3]The core domain (residues 136-198) was used. [4]The Lennard-Jones van der Waals energy was calculated with CHARMM PARAM19/20. [5]Solution-state structures derived by simulated annealing. [6]The core domain excluding the flexible loop (residues 151-153) was used.

The solution structure of Zα

The solution structure of the Z-DNA binding domain Zα (residues 119 – 200) of human ADAR1 was determined by heteronuclear multi-dimensional NMR spectroscopy, as described in the previous sections. Since the wild-type construct showed aggregation within days at room temperature and at concentrations required for NMR spectroscopy, cysteine 125 was mutated to

serine. Binding of the C125S-mutant to Z-DNA is indistinguishable from wild-type in surface plasmon resonance, circular dichroism spectroscopy (CD) and analytical ultracentrifugation experiments [4]. In 2D ^{15}N-HSQC spectra, no difference between wild-type and the C125S-mutant was discernible suggesting that the C125S-substitution has no effect on the protein conformation.

The solution structure of the Zα core domain comprises residues 136 – 198. Residues preceding Y136 and residues succeeding A198 do not show long-range NOE and are exposed to solvent. The Zα domain consists of three α-helices (designated α1, α2 and α3) and three β-strands (designated β1, β2 and β3) having an α1β1α2α3β2β3 topology (fig. 47). The three helices are roughly perpendicular to each other enclosing a hydrophobic core which is packed against the C-terminal antiparallel β2β3-sheet (fig. 45). β1 is oriented almost perpendicular to the C-terminal β-sheet, contacting β3 through two backbone hydrogen bonds between T156 and W195. This arrangement of three α-helices and β-strands is classified as the (α+β) helix-turn-helix (α+βHTH) fold, which has been found in numerous eukaryotic and prokaryotic B-DNA binding protein domains (table 11).

Zα has a flexible loop between α1 and β1 and an unexpectedly rigid loop between β2 and β3 in solution (fig. 43). The rigidity of the latter can be partially accounted for by the restricted flexibility of the backbone at P192-P193, where P192 forms an unusual *cis* peptide bond. In addition, long-range NOE restraints between P192 and T157 of β1 define the conformation of this hairpin precisely (fig. 46). In contrast, we found no long-range NOEs for the α1β1-loop residues G151 - G153. The mutations P192A and P193A strongly reduce Z-DNA affinity [5], suggesting that the rigid proline loop is important for binding Z-DNA.

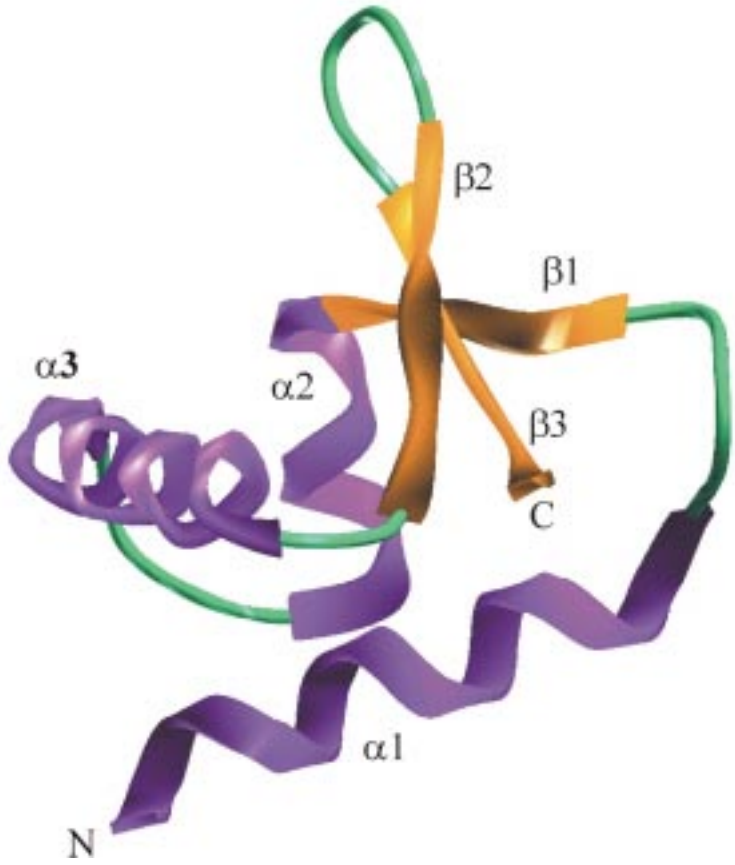


fig. 45 **Shaded ribbon representation of Zα.** Helix α2 and α3 form a HTH motif which is crossed by the N-terminal helix α1. This three helix bundle is packed against a C-terminal β-sheet enclosing a contiguous hydrophobic core. The short β-strand, β1, inserted between α1 and α2, crosses the C-terminal β-sheet almost perpendicular forming two backbone hydrogen bonds with β3.

Comparison of the Zα structure with mutants of the hydrophobic core

Evidence for an important structural role of residues in the hydrophobic core of Zα is obtained from previous mutagenesis studies where changing these residues to alanine enhances the rate of proteolytic degradation within *E.coli* [5]. Such increased degradation suggests either a loss of protein stability or an impairment of correct protein folding. In the solution structure of Zα, α1 is packed against α2 through extensive van der Waals interactions between the buried residues I143 and F146 of α1 and L161 and L165 of α2 (fig. 46). All four residues are conserved as hydrophobic residues in human, mouse, rat, bovine and frog Zα and Zβ domains [3]. The mutations I143A, L161G and L165P resulted in severe proteolytic degradation suggesting that the interactions observed in the NMR structure are important for the packing between α1 and α2 [5].

Helix α3 forms van der Waals contacts to both hydrophobic and polar side chains on α1. The hydrophobic residues I143 and L144 on α1 interdigitate with I172, L176 and L179 on α3 (fig. 46). Mutations in each of these highly conserved hydrophobic residues strongly reduced protein stability [5]. Although the aliphatic moieties of the polar residues E140 and Q139 on α1 interleave with V175 and L179 on α3, the mutations E140A and V175G had no effect on protein stability. Thus, the centrally located van der Waals interactions between α3 and the hydrophobic residues on α1 play a more important role for the stability of Zα than those between α3 and the polar residues.

α+βHTH proteins differ from related HTH proteins because they possess an additional C-terminal β-sheet, which is packed against the α-helical core. In Zα the aromatic ring system of W195 is sandwiched between L176 of α3 and K187 of β2, thereby linking the C-terminal β-

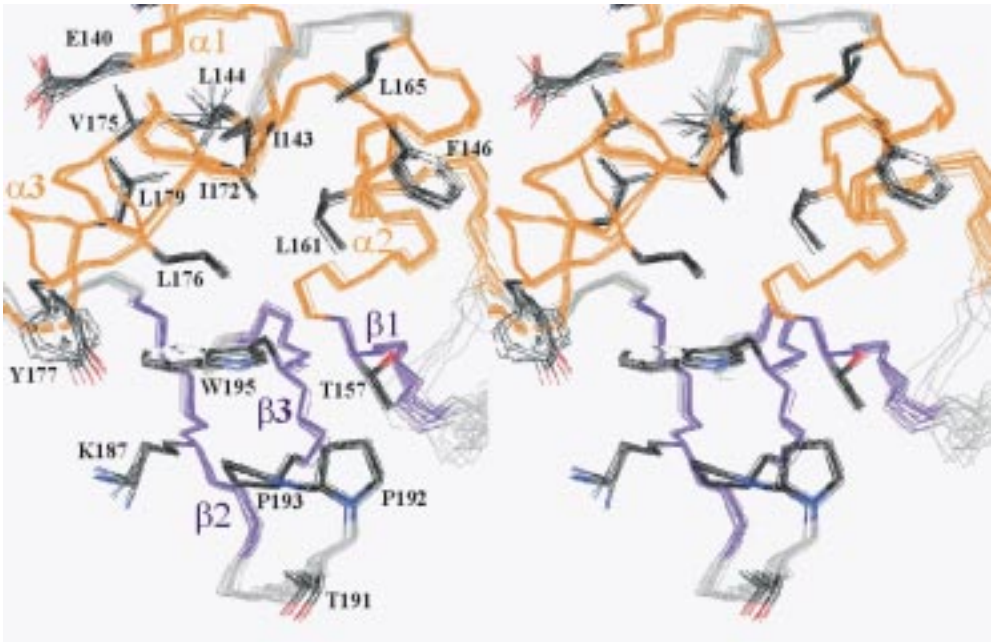


fig. 46 **Close-up stereoview of the hydrophobic core of Z**α shows the pivotal position of W195 in packing the C-terminal β-sheet against the α-helical core. The side chains of the hydrophobic core residues are well-defined in the ensemble of 15 NMR structures.

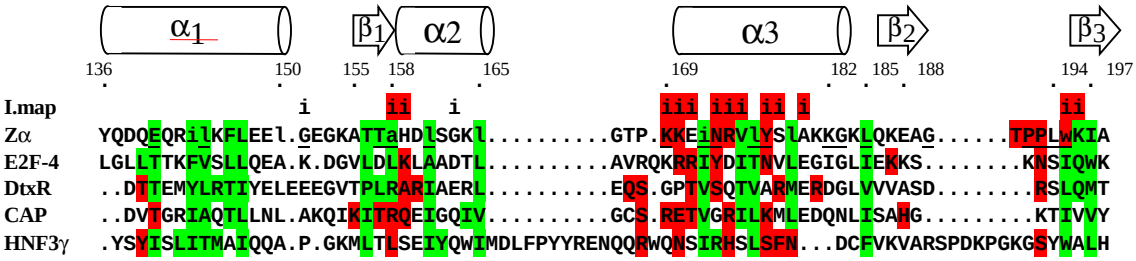


fig. 47 **Sequence alignment of structural homologues.** The human Zα domain (residues 136-198) is shown with the secondary structure derived from this NMR study. Residues absolutely conserved in human, mouse, rat, bovine and *Xenopus* Zα are underlined. Residues of Zα identified as essential for protein stability by scanning mutagenesis are in lower case. Residues identified by interaction mapping are superscripted with an 'i', which is shown on dark grey background for those residues forming a roughly contiguous binding surface. Structural homologues were aligned with Zα based on the DALI [206] structural alignment. Buried residues (light grey) were detected using PROCHECK [204] and DNA contacting residues (dark grey) were taken from the cocrystal structures.

sheet to the α-helical core (fig. 46). The W195A mutation is deleterious to the stability of Zα, and even the conservative W195Y mutation increased proteolytic degradation [5]. Mutant W195A showed the lowest affinity for Z-DNA of all mutants screened. The pivotal role of W195 is underscored by its absolute conservation in Zα and Zβ domains (fig. 17). The solution structure of Zα thus provides an explanation for the effects of different mutations on protein stability and for the conservation pattern seen in the sequence alignment of Zα and Zβ. This structure/function analysis of hydrophobic core interactions may also provide valuable insight into packing forces in homologous α+βHTH proteins because buried residues are markedly conserved throughout this entire class of protein (fig. 47).

Summary

The secondary structure of Zα in conjunction with a small number of unambiguous long-range NOE restraints allowed the identification of a suitable structural homolog, histone H5. After the homologous structure was remodeled to harbor the structural constraints known for Zα, it was used as a template structure to guide automated NOE assignment in 3D ^{15}N- and ^{13}C-HSQC-NOESY spectra. However, severe spectral congestion in the aliphatic region of these spectra necessitated considerable manual editing. Thus, the automated NOE assignment routine was only useful to provide a pre-selection of potential assignments rather than delivering an output that can be used directly for structure calculation.

By using this semi-automated approach of NOESY analysis, an exceptionally large number of NOE distance restraints was collected containing an average of 12.7 long-range NOE restraints per residue of the Zα core domain (residues 136-198). This led to a high coordinate precision with a backbone rmsd of 0.26 Å and a rmsd of 0.65 Å for all non-hydrogen atoms of the Zα core domain excluding one flexible loop (residues 151-153). The solution structure of Zα was computed by dynamic simulated annealing from random coordinates using floating stereo-specific assignment. The three-dimensional structure of Zα

comprises three α-helices and three β-strands consistent with the previously derived α1β1α2α3β2β3 topology. The α-helices are approximately perpendicular to each other. The two β-strands at the C-terminus form an antiparallel β-sheet which is packed against the α-helical bundle. Zα possesses a (α+β) helix-turn-helix fold occurring in several B-DNA binding proteins. The Zα structure is further validated by comparison with mutants of the hydrophobic core showing that residues buried in the structure were found to be essential for protein stability in the mutagenesis study.