

PEPTIDE NMR

1. STRUCTURE DETERMINATION BY NMR

NMR spectroscopy enables the determination of structures of proteins in solution under near-physiological conditions. In contrast to single crystal diffraction it does not require the protein to be crystallized, which still presents a major bottleneck. Unfortunately, spin physics change with increasing molecular weight such that proteins with molecular weights larger than about 50 kDa present a major challenge. NMR and crystallography have therefore often been considered as complementary techniques. Structure determination using NMR heavily utilizes the so-called nuclear Overhauser effect (NOE). The magnitude of the NOE depends on the distance separation of the interacting spins, but also on its motional properties¹:

$$NOE \sim r^{-6} \cdot f(\tau_c) \quad (\text{GL 1})$$

The NOESY spectrum contains all the information about spatial proximities of protons and therefore encoded the three-dimensional arrangement of atoms, the structure.

Structure determination by NMR may be divided into the following steps:

- Production of the peptide/protein (by recombinant methods, isolated from natural sources, or, in the case of smaller peptides, by solid-phase peptide synthesis)
- Establishment of suitable conditions for recording spectra (protein must be non-aggregating, stable over time at the chosen pH and temperature, folded...)
- Measurement of a series of 2D or 3D NMR spectra
- sequential assignment of all non-labile proton frequencies
- assignment of all cross-peaks encountered in the NOESY spectra
- integration of cross peaks and transformation into upper-distance bounds (calibration)
- (determination of scalar couplings for dihedral angle restraints)
- structure calculation
- energy refinement in a full force-field with explicit solvent
- assessment of the quality of the structure

1. τ_c denotes the correlation time, which is the time the molecule needs to rotate by 360° about an arbitrary axis.

2. SAMPLE-PREPARATION

The parameter should be checked for successful structure determination by NMR:

1. the protein should be in its native conformation:
 - check by CD spectroscopy whether the peptide/protein is folded, the temperature stability, the denaturing temperature (reversible?)
 - choose a proper *pH* (close to physiological pH, but protein should still be well-soluble (aggregation*), stable, amide-proton exchange not too fast.))
 - choose a proper *temperature*: Usually, spectral quality (linewidths) improve with increasing temp., but protein stability is often limiting
 - aggregation depends on the ionic strength of the solution (*Hofmeister series*)
 - suitable conditions can often be screened by CD, which requires less material *
 - Solubility should be reasonable at the chosen pH (pI!). Concentrations should usually be higher than 1mM (in 0.3-0.5ml). At half the concentration the measurement time is four times as long to get the same signal-to-noise! Sometimes addition of salt/detergents improves spectra remarkably.
2. purity of the protein should be very high (>95%).
3. Spectral quality usually decreases with increasing molecular weight:
 - The larger the protein the longer is its correlation time τ_c (η : viscosity, r : hydrodynamic radius, T : temperature in Kelvin, k : Boltzmann-constant)

$$\tau_c = \frac{4\pi\eta r^3}{3kT} \quad (\text{GL 2})$$

long correlation times lead to broad signals and less signal intensity, more signal overlap!

- resonance assignment using homonuclear proton spectra is feasible up to approx. 50AA, 50-100 AA requires ^{15}N -labelling, > 100 AA additional ^{13}C (plus possibly ^2H) labelling.
4. aggregation state:
 - dynamic light scattering allows to determine the aggregation state
 5. check stability under the chosen condition of pH, temp., concentration etc. by CD, FPLC.

An often used approach is to express the protein on “normal media” (LB), then record a 1D spectrum and a NOESY to judge about spectral quality (folding, aggregation, signal dispersion), then produce ^{15}N -labelled protein to check the ^{15}N , ^1H correlation. If these spectra look promising the protein may be produced in its fully labelled form.

Distance limits were derived from NOEs. Most of the time spent during an NMR structure determination is actually required to determine the frequencies of all non-labile protons in the sequence in order identify the proton pairs leading to each

particular cross peak. In order to prevent loss of signals of amide protons from exchange with solvent deuterons hydrophilic peptides are usually measured in 90% H₂O/D₂O. Backbone-NH exchange is slowest at approx. pH 3.0, but for reasons of stability (amide bonds may be hydrolyzed in acidic solutions) the chosen pH is mostly between 4 and 5. (Fig. 1) In the case of globular (folded) proteins amide protons are usually part of hydrogen bonds and therefore exchange much more slowly, so that neutral or even basic values of the pH may still be chosen. Similar arguments apply for solvent shielded protons (those in the core).

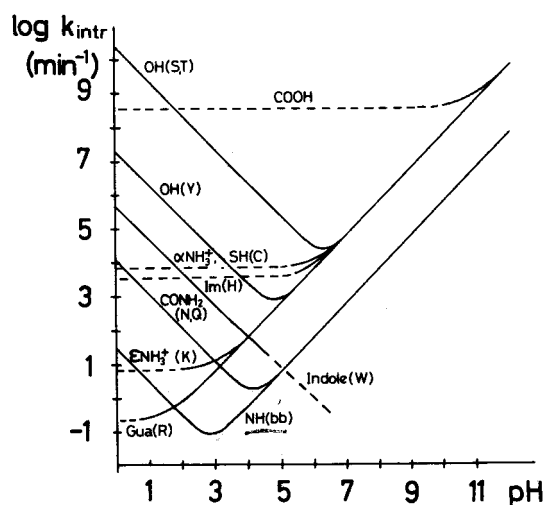


FIGURE 1. Logarithmic presentation of the intrinsic exchange rates vs. pH for solvent accessible, labile protons in aqueous solution at 25°C.

3. MEASUREMENT OF SPECTRA

Structure calculation of proteins is based on the Nuclear Overhauser Effect (NOEs) between protons. In order to assign the NOE to specific resonances all non-labile protons must be assigned to their sequence-specific position (sequence-specific sequential resonance assignment). Because of the large number of protons per residue (3-13) and the resulting resonance overlap even smaller peptides cannot be assigned from 1D spectra but require use of two or three dimensional correlation spectra. The following figure indicates the positions of the protons in the 1D spectrum:

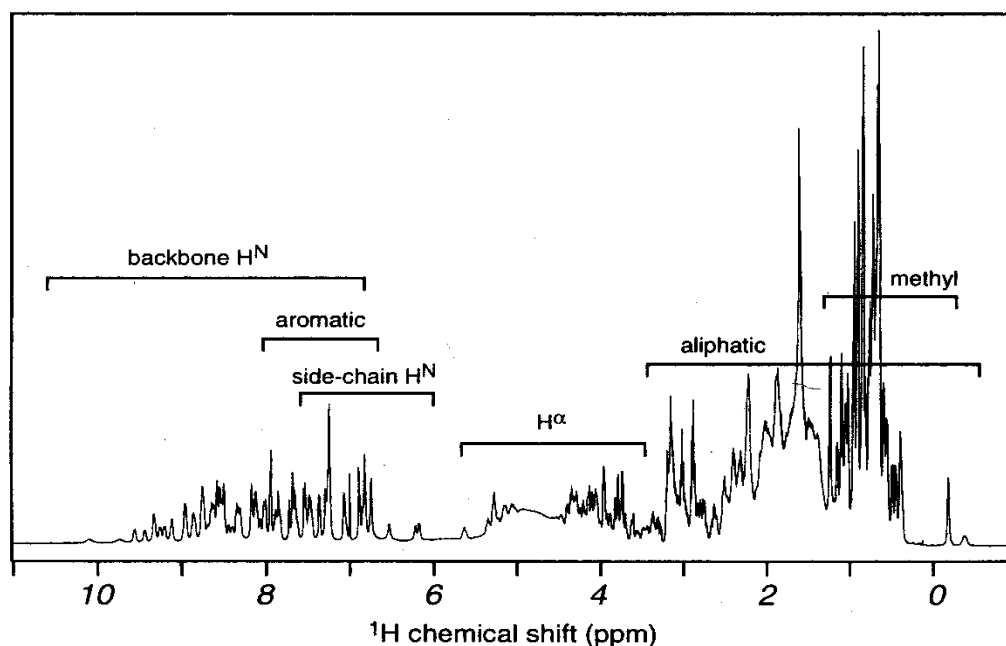


FIGURE 2. ^1H chemical shift positions of chemical groups in ubiquitin¹ (from: Cavanagh *et al.*: Protein NMR Spectroscopy).

In 2D spectra correlations between two frequencies are recorded. The correlating partners are found by forming the orthogonal projections onto the two frequency axis F2 (direct dimension) and F1 (indirect dimension). Peak intensities are encoded as contours like in a roadmap. By which mechanism spins appear to be correlated in the spectrum is determined by the type of experiment: Main mechanisms are scalar couplings (COSY, TOCSY), whose magnitude depends on the number of intervening bonds (usually 3 or smaller), but also on the dihedral angle about the central bond. The second important mechanism is the dipolar (direct, through-space) couplings (NOESY).

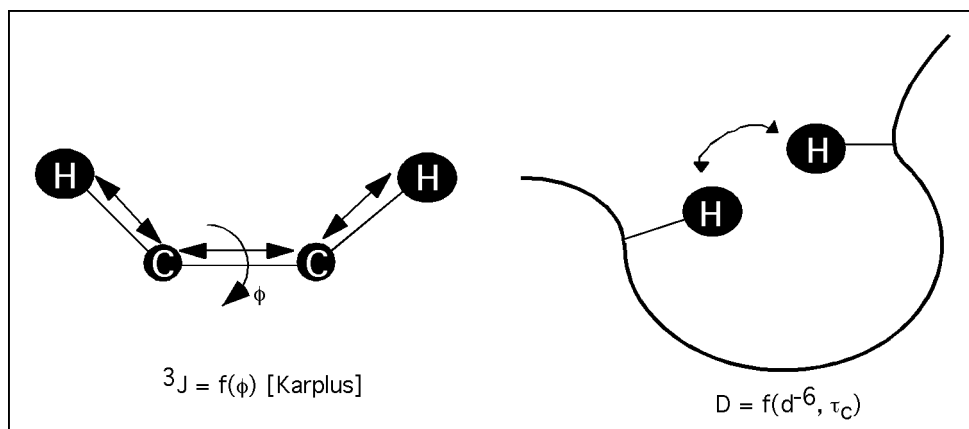


FIGURE 3. Scalar (left) and dipolar (right) couplings.

All homonuclear spectra (e.g. proton frequencies in both dimensions) display also the auto correlation peaks, placed on the diagonal (diagonal peaks), which are mostly very intense. The peaks on the diagonal reflect the 1D spectrum of the protein

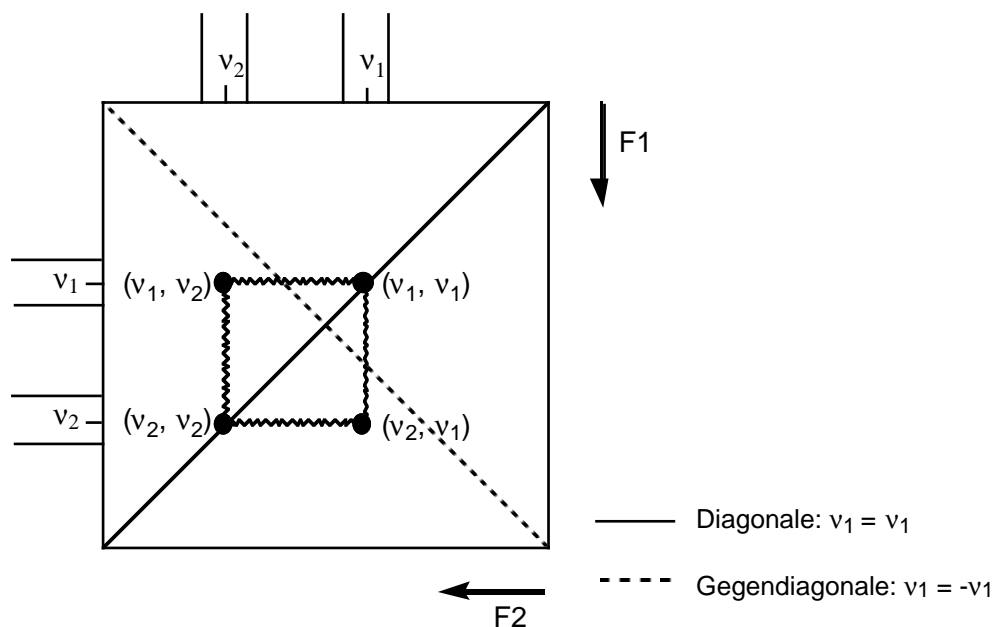


FIGURE 4. Schematic representation of a 2D spectrum, external projections are added.

For resonance assignments of small (non-labelled) peptides usually a set of three different spectra is recorded: a COSY, a TOCSY and the NOESY.

A short remark: Spectra are usually recorded in 90% H₂O, 10% D₂O. The very strong water signal must be experimentally suppressed and the residual water usually appears in the middle of the spectrum. In 2D spectra the residual water signal is manifested as a band of noise in the center of the spectrum. It may obscure peaks close to the water resonance (such as the H α protons).

3.1 [¹H,¹H]-COSY:[¹H,¹H]-Correlated Spectroscopy

The COSY displays [¹H,¹H]-correlations due to scalar (through-bond) couplings.

Efficiency of the coherence transfer largely decreases with increasing linewidth (which is related to the molecular weight). The COSY experiment is therefore almost exclusively used for smaller (non-labelled) peptides.

Positions of cross peaks in the COSY are characteristic for the amino acids and can be classified according to the following rules (see the next figure for their position in the spectra):

- all non-labile, non-aromatic sidechain protons except those from $\beta\text{H} - \gamma\text{CH}_3$ of Thr, $\delta\text{H} - \delta\text{H}$ of Pro and $\beta\text{H} - \beta\text{H}$ of Ser.
- $\alpha\text{H} - \beta\text{CH}_3$ of Ala and $\beta\text{H} - \gamma\text{CH}_3$ of Thr.
- $\alpha\text{H} - \beta\text{H}$ of Val, Ile, Leu, Glu, Gln, Met, Pro, Arg, and Lys.
- $\alpha\text{H} - \beta\text{H}$ of Cys, Asp, Asn, Phe, Tyr, His and Trp.
- $\alpha\text{H} - \alpha\text{H}$ of Gly, $\alpha\text{H} - \beta\text{H}$ of Thr, $\delta\text{H} - \delta\text{H}$ of Pro, $\alpha\text{H} - \beta\text{H}$ and $\beta\text{H} - \beta\text{H}$ of Ser.

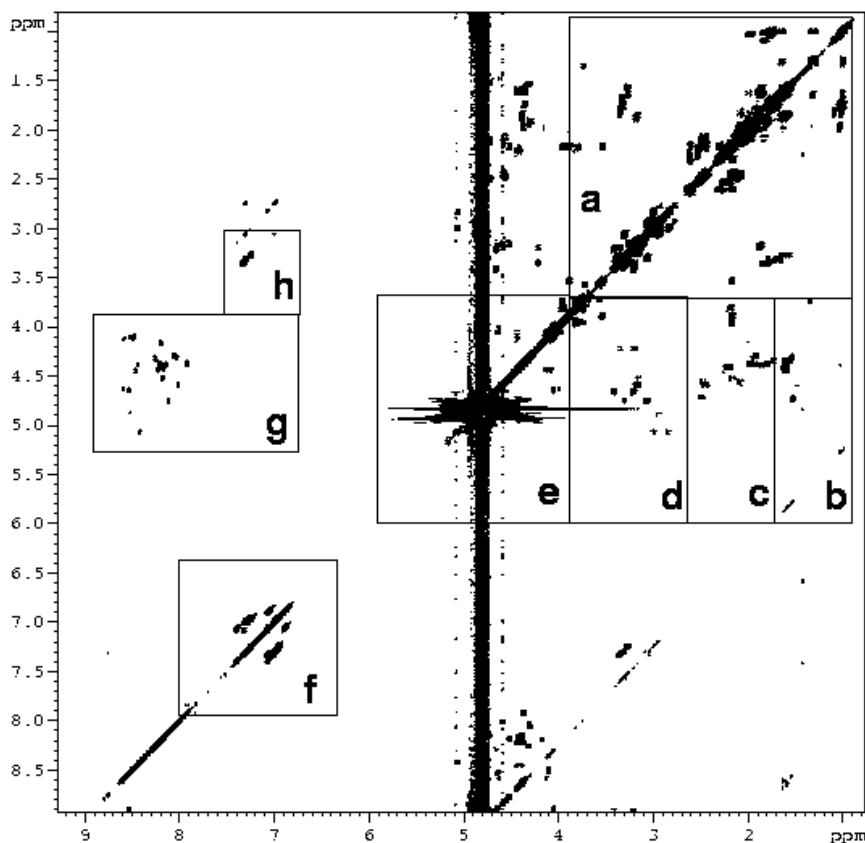


FIGURE 5. [$^1\text{H}, ^1\text{H}$]-COSY of Neuropeptide Y (NPY)

- aromatic ring protons, including 2H-4H of His, as well as sidechain protons from Asn and Gln.
- backbone NH- αH .
- $\delta\text{CH}_2 - \epsilon\text{NH}$ of Arg.

The region comprising the NH- αH -Region (g) (in spectra recorded in 90% $\text{H}_2\text{O}/$ 10% D_2O , in D_2O backbone NH-protons exchange with solvent D) is called the *fingerprint*-region. Spectral resolution and appearance in that part serves as an indicator whether the peptide/protein can be successfully investigated by NMR. For

each amino acid (except Pro) a single peak is observed in that region. The N-terminal amino proton cannot be observed due to very rapid exchange with the solvent.

Scalar couplings can only be observed between protons separated by not more than 3 bonds. Moreover, the protons need to have a different resonance frequency. However, signal overlap, exchange broadening, small values of the $J_{\text{NH}-\alpha\text{H}}$ -coupling constants (which in stable α -Helices is < 4 Hz), resulting from dihedral angles close to 90° , may lead to the fact that less than the expected number of signals are observed.

COSY crosspeaks display a multiplet fine structure, which is caused by the scalar couplings. Active couplings (those that lead to the cross peak) display anti-phase components (components with positive and negative signal intensity) whereas all other couplings (passive couplings) are in-phase:

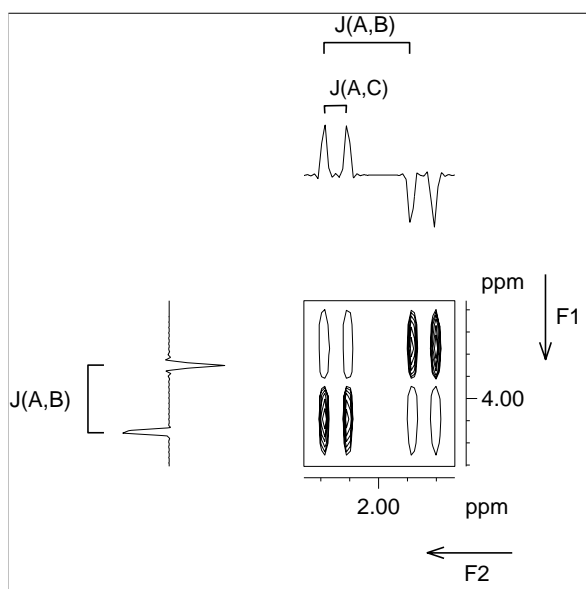
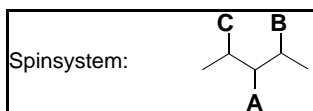


FIGURE 6. COSY cross peak pattern for the following spin system



3.2 [$^1\text{H},^1\text{H}$]-TOCSY: Total Correlation Spectroscopy

A TOCSY experiment contains all cross peaks due to protons of the same spin system. *Protons from different amino acids always belong to different spin systems, because there is no scalar coupling across the amide bond.* Some amino acids consist of a single spinsystem (Ile), some contain two (e.g. Phe) or three (e.g. Trp):

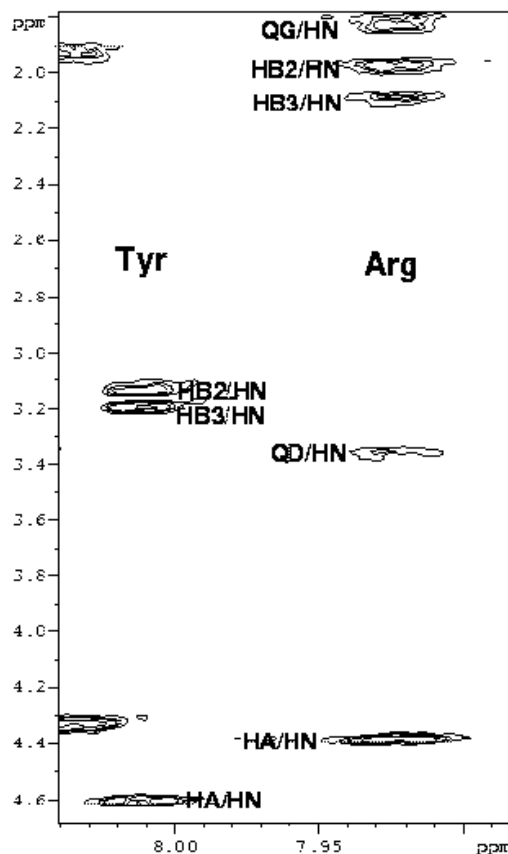


FIGURE 7. Spinsystems of Tyr (J) and Arg(X) in the TOCSY

Analysis of spin systems allows to decide to which type of amino acids the spin system belongs. Possible criteria are: Occurrence or absence of methyl groups, length of the spin system, positions of chemical shifts in the spin system. Although the exact amino acid can rarely be derived many amino acids can be excluded from such an analysis. The mixing time (a parameter that can be varied during setup of the experiment) determines whether neighboring correlations (e.g. for a 12ms TOCSY, which displays information similar to a COSY) or long-range correlations (e.g. for a 80ms TOCSY) are detected. Such an analysis is conducted in a region of the spectrum which displays the smallest overlap. This part is usually the region containing the amide protons. Possible overlap may be removed by recording a second set of spectra at a slightly different temperature.

The following figure shows the difference between a COSY and a 80ms TOCSY. Whereas only a single correlation from each amide proton is found in the COSY, many of them exists in the TOCSY:

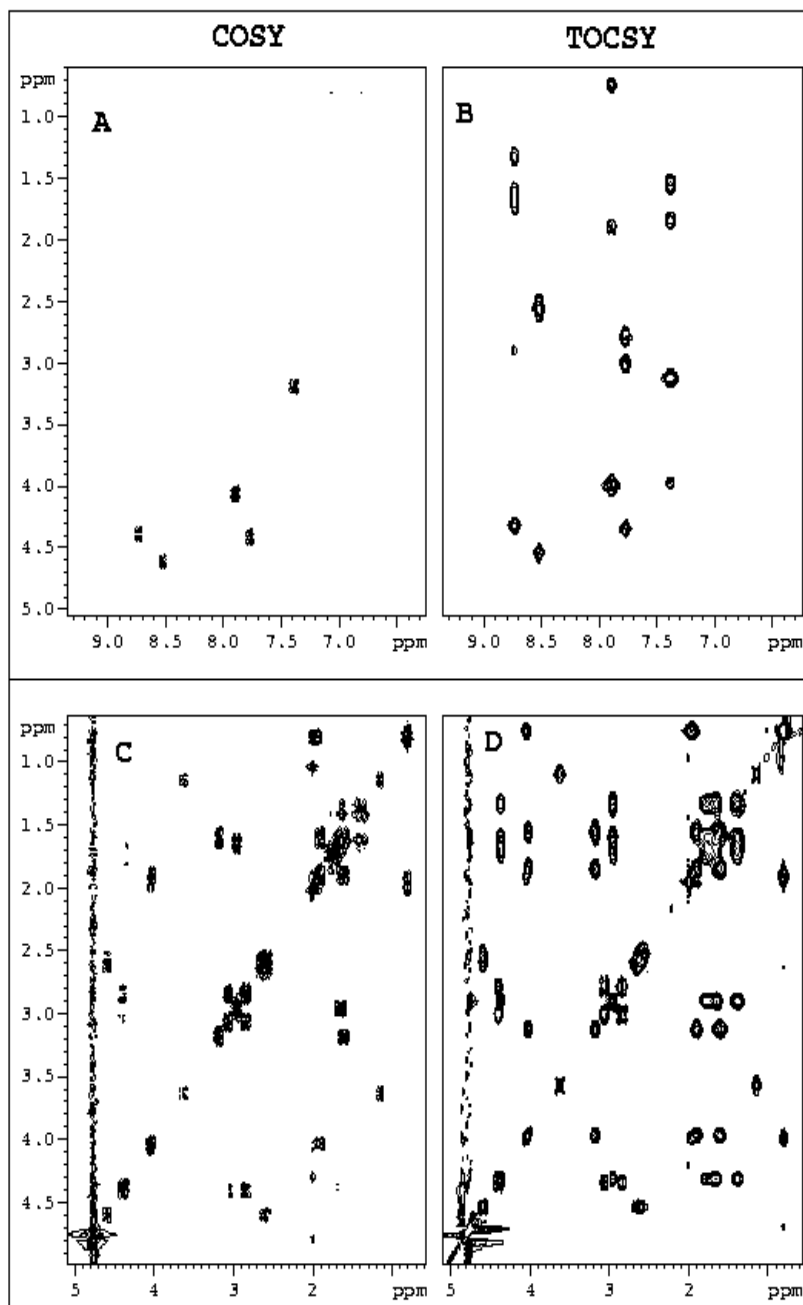


FIGURE 8. Comparison of the region containing the amide protons (F2), aliphatic protons (F1) for a COSY (A) or TOCSY (B) as well as correlations between aliphatic protons in the COSY (C) and TOCSY (D).

3.3 [$^1\text{H}, ^1\text{H}$]-NOESY: Nuclear Overhauser Effect Spectroscopy

Cross peaks in the NOESY are due to dipolar couplings resulting from interactions of spins *via space* and hence only depend on the distance but not on the number of intervening bonds:

$$D \approx r^{-6} \quad (\text{GL 3})$$

Dipolar couplings are averaged to zero in solution but give rise to the very important relaxation phenomena, one of which is the NOE (nuclear Overhauser effect, NOE) - The strong dependence of the cross peak intensity on the distance separation explains why this parameter is the most useful for structure determination.

The sign of the NOE depends on the magnitude of the rotational correlation time and hence on the molecular weight (but also on the viscosity of the solvent)

TABLE 9. Signal phases in NOESY-experiments:

	signal phase of the diagonal peak	signal phase of the cross peak
small molecules	positive	negative
medium-size molecules	positive	very weak signals (positive or negative)
large molecules	positive	positive

The maximum proton-proton NOE in dependence of the correlation time is shown in the figure below:

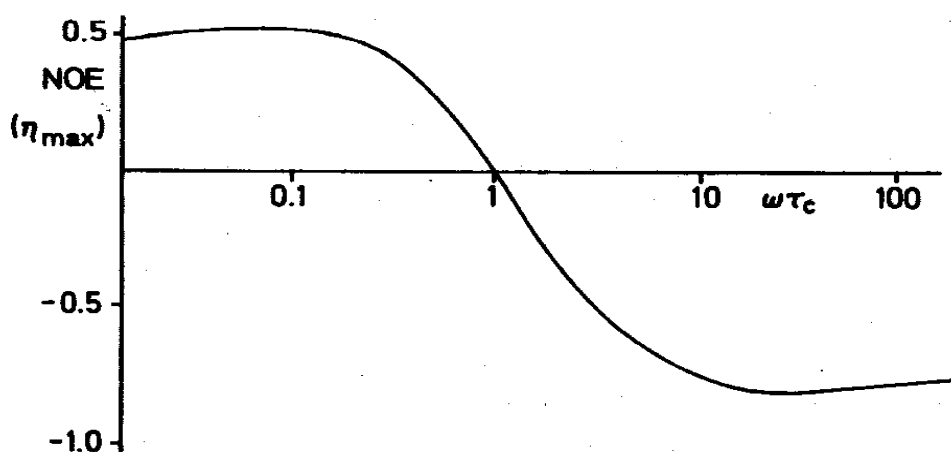


FIGURE 10. Dependence of the $^1\text{H}, ^1\text{H}$ NOE on the product of the correlation time and the resonance frequency.

Similarly to the situation encountered in TOCSY/COSY spectra peaks may be classified according to in which region of the spectrum they are found:

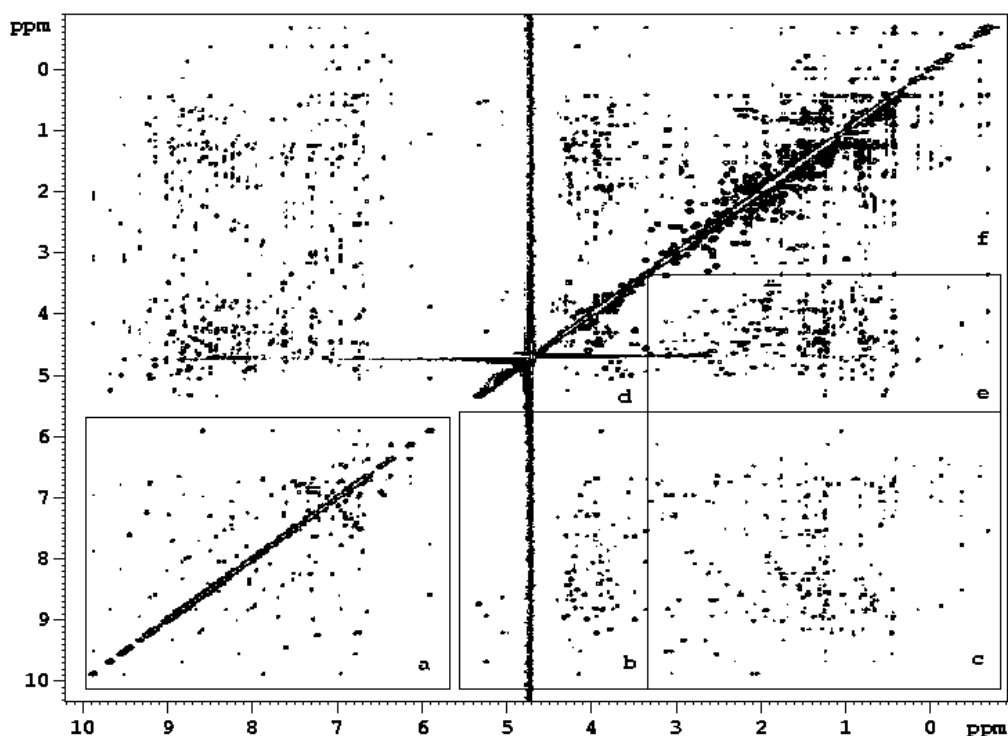


FIGURE 11. (H,H)-NOESY of ProteinD (phage envelope protein)

- a. NH; aromatic - NH; aromatic-aromatic.
- b. NH; aromatic - α H; δ H of Pro; β H of Ser and Thr.
- c. NH; aromatic - aliphatic sidechains.
- d. α H; δ H of Pro; β H of Ser and Thr - α H; δ H of Pro; β H of Ser and Thr.
- e. α H; δ H of Pro; β H of Ser and Thr - aliphatic sidechains.
- f. Aliphatic sidechains - aliphatic sidechains.

The NOESY not only contains peaks from which distances are derived for the structure calculation but is also heavily used during the sequential resonance assignment process. Considering that scalar couplings are restricted to protons within a single amino acids, sequential correlations (correlations between proton of neighboring amino acids) need to be taken from the NOESY.

4. STRATEGIES FOR RESONANCE ASSIGNMENT OF SMALL NON-LABELLED PEPTIDES

4.1 Optimizing spectral quality

The spectral quality can be rated by inspection of the fingerprint region (number of peaks should roughly match the number of non-proline residues). If more than 10% of peaks are missing, the conditions should be varied (temperature, pH, salt etc.) in order to remove peak overlap or aggregation.

In the first step peaks must be referenced correctly. In protein NMR the water signal (usually in the center of the spectrum) is used, whose frequency depends on the temperature according to (T in [K]):

$$\delta(H_2O) = 7,83 - \frac{T}{96,9} ppm \quad (\text{GL 4})$$

Sometimes it is advisable to record a second set of spectra at a slightly different temperature to remove peak overlap.

4.2 Spinsystem-Identification: NH- α H- β H-...

In a first step the spin systems are classified. This is best done in the TOCSY selecting the amide proton region (approx. 6.5-12 ppm) in F2 and the aliphatic region (0-5 ppm) in F1. Spin systems will line up vertically in that region, and the number and position of peaks give valuable information. Useful criteria are

- the length of the spin systems (number of peaks)
there are short spin systems (type J, e.g. Ser), long spinsystems (e.g. Lys)
- pattern of peak positions are characteristic (see appendix)

Some amino acids can be identified based on the chemical shifts rather easily:

Ser: β -protons low-field (> 4ppm), no methyl group

Thr: β -protons low-field (> 4ppm), methyl group around 1.2 ppm

Ala: no β -protons, but methyl group around 1.3 ppm

Gly is the only amino acid that shows up as a triplet for the amide proton (coupling to two α -protons!), Pro has no amide proton, but displays a characteristic pattern in the aliphatic region of the TOCSY. Ala, Thr, Val, Leu and Ile are methyl-group containing amino acids and can be distinguished from each other based on their COSY connectivities. Short spin systems (only NH, α - and β -protons) are from Ser or Cys, Asp, Asn, Phe, His, Trp and Tyr (so called type J-spin systems). Typical

regions in which the protons from the amino acids are found are displayed in the figure below:

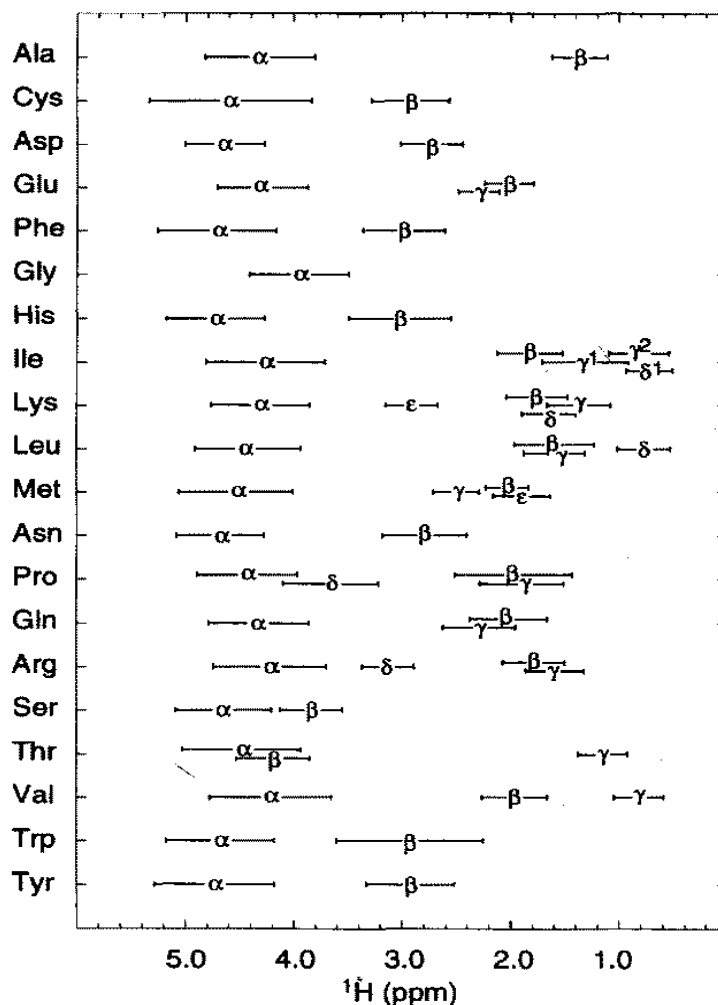


FIGURE 12. ^1H -chemical shift ranges of aliphatic protons (taken from Cavanagh et al.: Protein NMR Spectroscopy)

The following table contains the random-coil chemical shifts, as measured in unstructured small peptides. The values rather serve as an indication for approx. values, in structured proteins deviations can be substantial!

TABLE 13. Random coil chemical shifts for dipeptides (following Ala residues)

residue	NH	H α	H β	other
Ala	8.24	4.32	1.39	
Cys(red)	8.32	4.55	2.93, 2.93	
Cys(ox)	8.43	4.71	3.25, 2.99	
Asp	8.34	4.64	2.72, 2.65	
Glu	8.42	4.35	2.06, 1.96	γCH_2 2.31, 2.31
Phe	8.30	4.62	3.14, 3.04	2,6H 7.28; 3,5H 7.38; 4H 7.32
Gly	8.33	3.96		
His	8.42	4.73	3.29, 3.16	2H 8.58; 4H 7.29

TABLE 13. Random coil chemical shifts for dipeptides (following Ala residues)

residue	NH	H α	H β	other
Ile	8.00	4.17	1.87	γ CH ₂ 1.45, 1.16; γ CH ₃ 0.91; δ CH ₃ 0.86
Lys	8.29	4.32	1.84, 1.75	γ CH ₂ 1.44, 1.44; δ CH ₂ 1.68, 1.68; ϵ CH ₂ 2.99, 2.99; ϵ NH ₃ ⁺ 7.81
Leu	8.16	4.34	1.62, 1.62	γ CH 1.59; δ CH ₃ 0.92, 0.87
Met	8.28	4.48	2.11, 2.01	γ CH ₂ 2.60, 2.54; ϵ CH ₃ 2.10
Asn	8.40	4.74	2.83, 2.75	γ NH ₂ 7.59, 6.91
Pro	-	4.42	2.29, 1.94	γ CH ₂ 2.02, 2.02; δ CH ₂ 3.63, 3.63
Gln	8.32	4.34	2.12, 1.99	γ CH ₂ 2.36, 2.36; δ NH ₂ 7.52, 6.85
Arg	8.23	4.34	1.86, 1.76	γ CH ₂ 1.63, 1.63; δ CH ₂ 3.20, 3.20; ϵ NH 8.07
Ser	8.31	4.47	3.89, 3.87	
Thr	8.15	4.35	4.24	γ CH ₃ 1.21
Val	8.03	4.12	2.08	γ CH ₃ 0.94, 0.93
Trp	8.25	4.66	3.29, 3.27	2H 7.27; 4H 7.65; 5H 7.18; 6H 7.25; 7H 7.50
Tyr	8.12	4.55	3.03, 2.98	2,6H 7.14; 3,5H 6.84

The long spin systems like Lys, Arg, Met, Gln, Glu and Pro all contain two γ -protons, coupled to the β -protons. Usually, the β -protons are observed at frequencies higher than 2.2 ppm and were therefore called type U (upfield) spin systems. Met, Gln and Glu have their γ H-frequencies lower than those of the β -protons whereas they are higher for Arg, Lys, Pro and Leu.

Characteristic patterns for cross peaks as found in COSY or TOCSY spectra are summarized in the appendix.

4.3 Sequence-specific assignment

Sequential resonance assignment means connecting spin systems in their sequential order. Considering the fact that scalar couplings will never occur between protons of different amino acids it is clear that NOEs must be used for that purpose. Since NOEs may be found between all protons close in space the use of them introduces some ambiguity. For sequential correlations usually a set of peaks are used, largely depending on the secondary structure in the corresponding segment. Two principal elements of secondary structure exist: Extended chains (corresponding to β -strands, β -

sheets) or helical regions. In helical regions sequential amide protons are close in space:

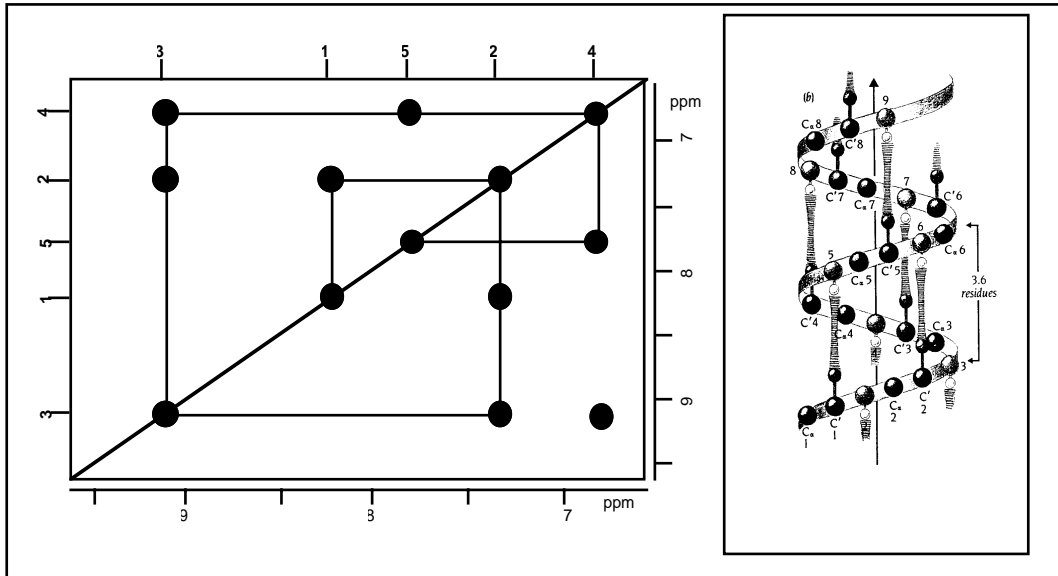


FIGURE 14. NOESY-diagonal- and crosspeaks in the NH (F2)/ NH (F1)- region. Strong (sequential) d_{NN} -crosspeaks are observed in helical segments (NOESY-walk in the NH/NH-region)

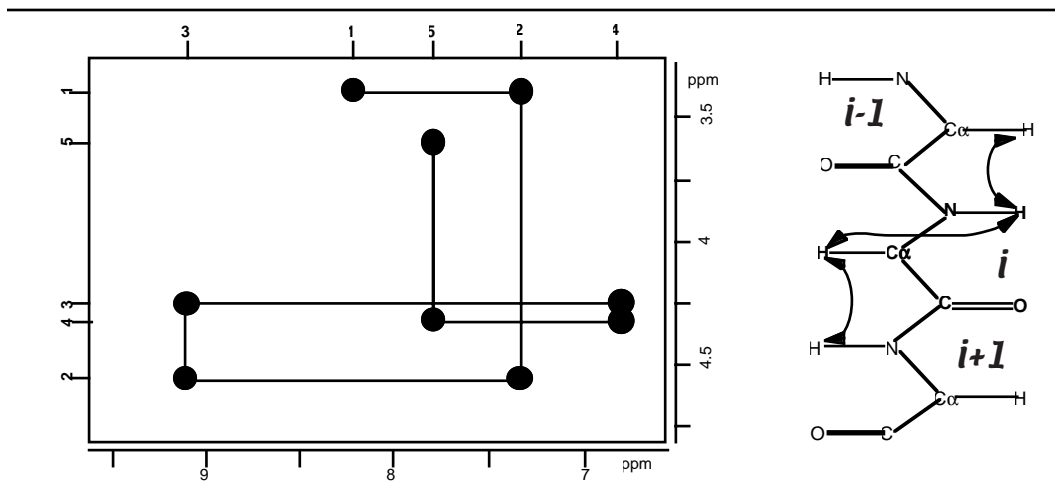


FIGURE 15. NOESY-crosspeaks in the NH(F2)/ α H (F1)-region. Strong (sequential) $d_{\alpha N}$ -crosspeaks occur in extended (β -sheets and random coil) conformations.

In extended chains or random coil regions sequential NH_{i+1} - αH_i peaks are observed. By comparing fingerprint regions of COSY (or TOCSY) and NOESY intraresidual and sequential NH- αH peaks can be distinguished:

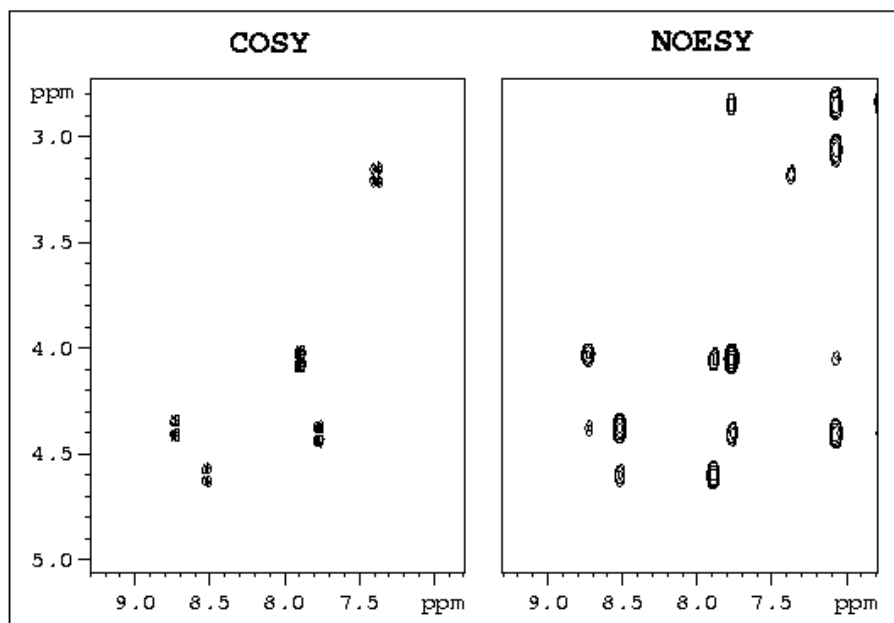


FIGURE 16. : Fingerprint region in COSY (left) and NOESY (right) spectra.

Once it is clear which spin systems are sequential over a segment of 3 to 4 residues the additional information from the spin system identification is usually sufficient to assign the spin systems to their exact position in the amino acid sequence.

5. STRUCTURE CALCULATIONS USING NMR DATA

5.1 Chemical Shifts. The presence of an α -helix or β -sheet has a decisive influence onto the chemical shifts of the α -protons. If the chemical shifts of three consecutive residues are by more than 0.1 ppm higher than the corresponding random coil values the presence of a helix is probable. Are they lower by the same amount the presence of a β -sheet is likely.

5.2 $^3J_{\text{HN-}\alpha\text{H}}$ -Coupling constants. are related to the torsion angle θ via the Karplus relation:

$$J(\theta) = 6,98(\cos\theta)^2 - (1,38(\cos\theta)) + 1,72 \quad (\text{GL 5})$$

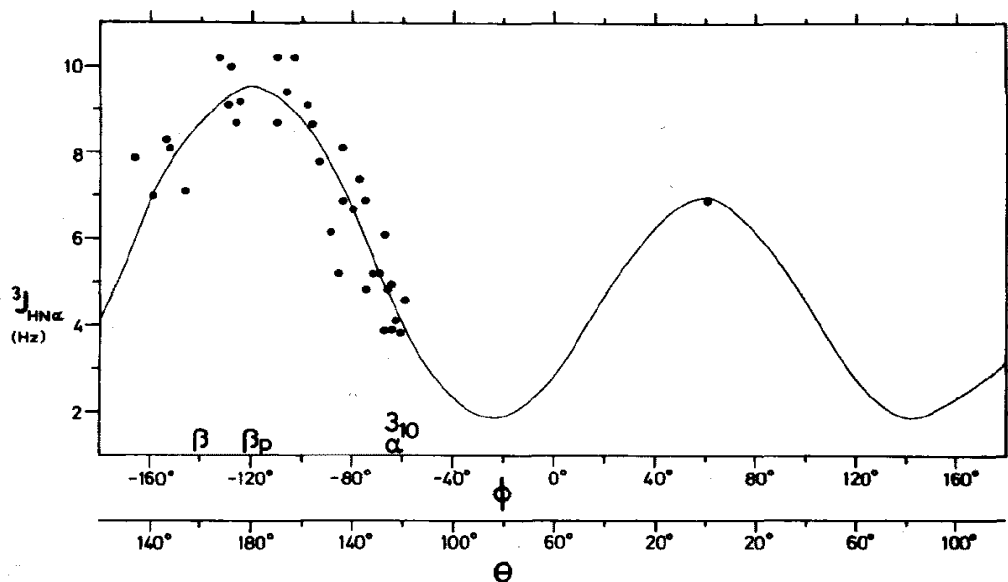


FIGURE 17. Plot of $^3J_{\text{HN-}\alpha\text{H}}$ -coupling constants versus the dihedral angle $\theta=|\phi-60^\circ|$, as derived from a crystal structure, in BPTI. (taken from Wüthrich: NMR of Proteins and Nucleic Acids)

For most values of J more than one solution exists. Scalar couplings may be used to define dihedral angle restraint *ranges*, which sometimes help to improve convergence of the structure calculation:

TABLE 18. Theoretical values of $^3J_{\text{HN-}\alpha\text{H}}$ in typical secondary structural elements:

secondary structure	θ	$^3J_{\text{HN-}\alpha\text{H}}$
α -helix	-57°	3.9 Hz
3_{10} -helix	-60°	4.2 Hz

TABLE 18. Theoretical values of ${}^3J_{\text{HN-}\alpha\text{H}}$ in typical secondary structural elements:

secondary structure	θ	${}^3J_{\text{HN-}\alpha\text{H}}$
antiparallel β -sheet	-139°	8.9 Hz
parallel β -sheet	-119°	9.7 Hz

Internal flexibility leads to rotationally averaged values of the scalar coupling constants, resulting in typical values of about 7 Hz in short, unstructured peptides, and hence values between 6 and 8 Hz are usually not included in the structure calculations.

5.3 NOEs

NOEs are the by far most important source of information for the structure calculation. They are usually observed for protons separated by less than 5 Å. They may be due to protons far in distance in the amino acid sequence (e.g. cross-strand NOEs in β -sheets)!

A severe problem for larger molecules is that NOEs are not only observed between spins close in space but transfer via a relay nucleus may have occurred (spin-diffusion), depending on the mixing time of the NOESY experiment. The two-spin approximation is strictly valid only for smaller peptides and short mixing times. Identification of speaks due to spin diffusion is crucial. In order to reduce artefacts due to spin diffusion but also due to internal mobility *upper-distance bounds* are employed instead of exact distance bounds (the mixing time can also not be chosen too short because signal buildup takes some time). Too large upper distance limits will lead to less-well defined structures whereas too small upper distance bounds will lead to wrong structures, because the correct distance is not contained in the allowed solutions.

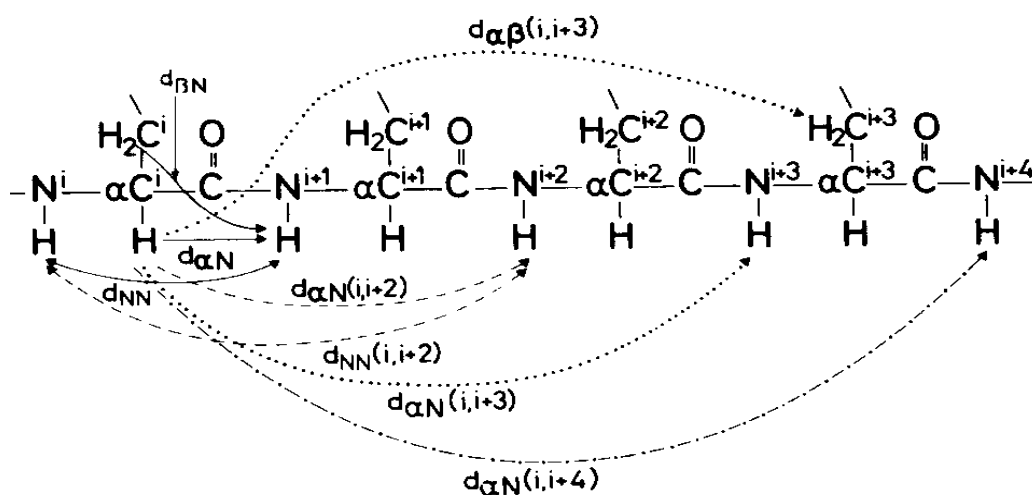
In a first step peak volumes, derived from integration of the peaks, will be translated into distances according to:

$$V = \frac{k}{u^6} \quad (\text{GL } 6)$$

in which the constant k is often chosen such that known NOEs of certain distances will correspond to typically observed values. e.g. sequential distances $d(\text{H}_i\alpha, \text{H}_{i+1}\text{N})$ and $d(\text{H}_i\text{N}, \text{H}_{i+1}\text{N})$ in regular secondary structural elements.

TABLE 19. Short (<4.5Å) sequential and medium-range ^1H - ^1H distances in polypeptide secondary structures

Distance	α -Helix	$^3_{10}$ -Helix	β	βp	turn I	turn II
$d_{\alpha\text{N}}$	3.5	3.4	2.2	2.2	3.4 3.2	2.2 3.2
$d_{\alpha\text{N}}(i,i+2)$	4.4	3.8			3.6	3.3
$d_{\alpha\text{N}}(i,i+3)$	3.4	3.3			3.1-4.2	3.8-4.7
$d_{\alpha\text{N}}(i,i+4)$	4.2					
d_{NN}	2.8	2.6	4.3	4.2	2.6 2.4	4.5 2.4
$d_{\text{NN}}(i,i+2)$	4.2	4.1			3.8	4.3
$d_{\beta\text{N}}$	2.5-4.1	2.9-4.4	3.2-4.5	3.7-4.7	2.9-4.4 3.6-4.6	3.6-4.6 3.6-4.6
$d_{\alpha\beta}(i,i+3)$	2.5-4.4	3.1-5.1				

FIGURE 20. Sequential and medium-range ^1H - ^1H distances in polypeptide chains (taken from Wüthrich: NMR of Proteins and Nucleic Acids).

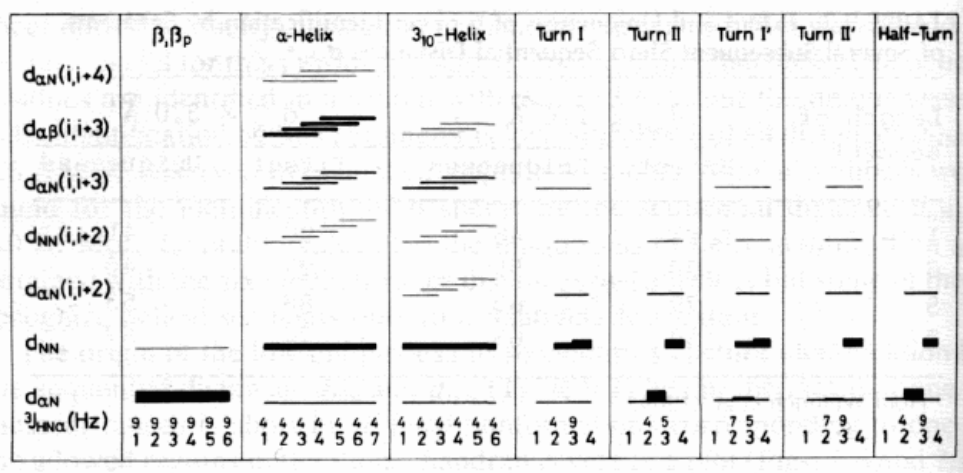


FIGURE 21. Overview over sequential and medium-range ^1H - ^1H NOEs and spin-spin-coupling constants $^3J_{\text{HN}\alpha}$ in various secondary structural elements. Short distances are indicated by horizontal lines connecting the corresponding proton positions and the thickness of the bars is proportional to the intensity of the observed NOE. (taken from Wüthrich: NMR of Proteins and Nucleic Acids)

According to IUPAC rules the backbone conformation of a peptide is defined by four dihedrals: For $\Theta = \Phi - 60^\circ$. Sidechain conformations are usually restricted to the staggered conformations of χ_1 (60° , -180° , -60°).

Structure calculation programs use either distance geometry approaches or force-field/molecular dynamics methods. Therein, additional potentials are introduced representing the upper distance and dihedral angle restraints. The program DYANA for example uses molecular dynamics in torsion angle space.

A force field incorporates potentials for classical terms such as bond-stretching, electrostatics, van der Waals interactions etc.:

$$U_{pot} = U_{Torsionangle} + U_{vdW} + U_{Coulomb} + U_{NMR}$$

For programs working in dihedral angle space only dihedral angles are altered, and hence no bond distances or bond angles need to be corrected.

Violations of the upper distance limits as derived from the NOE data are added in form of a (semi hyperbolic) potential:

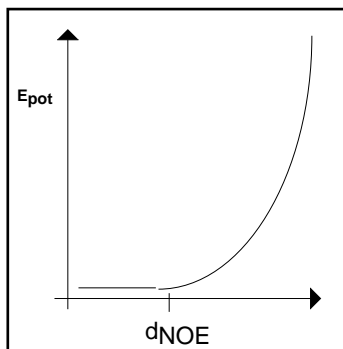


FIGURE 22. d_{NOE} in U_{NMR} is derived via distance-volume calibration from the volume integrals of NOESY cross peaks.

In a typical calculation a set of conformers with arbitrary dihedral angles (random conformers) is produced. Unfortunately, simple energy minimization algorithms are unlikely to lead to the correct structure since the energy landscape is complex and force field methods will only find the nearest local minimum. Instead of a simple force-field calculation (much) kinetic energy is therefore delivered to the system such that it can undergo *major* conformational transitions. Standard procedures usually follow simulated annealing protocols, in which the temperature is raised to high values in order to deliver large kinetic energy to the system:

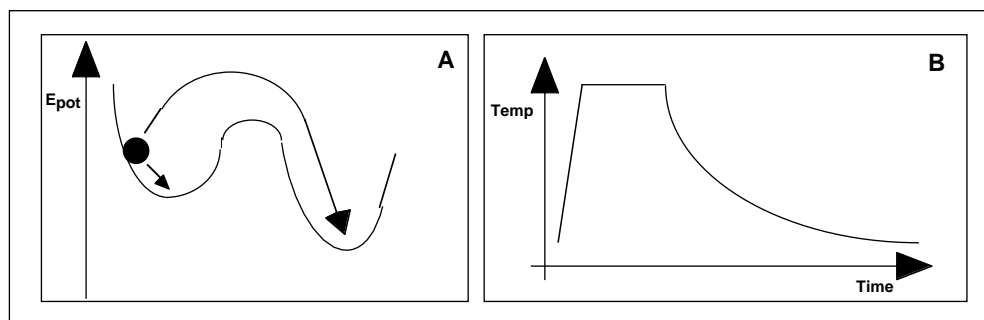


FIGURE 23. Left: Potential energy of a system. The global minimum is separated from the local one by an energy barrier. Right: Temperature vs. time during a simulated annealing protocol.

Provided that the equilibration period was sufficiently long, subsequent cooling should secure that a reasonable number of conformers assume low-energy conformations (recognized by low values of remaining potential energy). These low-energy conformers then represent the NMR ensemble.

In a MD calculation positions of atoms are corrected such that they follow a trajectory determined by the direction of the force acting on them. The force is computed from the classical newton equation of motion:

$$m_i \frac{d^2}{dt^2} r_i(t) = F_i(t) \quad i = 1, \dots, N \quad (\text{GL } 7)$$

in which the force is determined by the (negative) gradient of the potential energy. The latter is derived from a classical force-field to which extra terms representing the NMR-derived restraints have been added:

$$F_i = -\frac{d}{dr_i} U(r_1, r_2, \dots, r_N) \quad (\text{GL } 8)$$

Bond-lengths, bond-angles, chirality and planar bonds etc. are kept fixed in order to enable reasonably short calculation times.

In case peaks have been wrongly assigned in the NOESY or conformational equilibria exist not all distance limits can be fulfilled by the computed conformers simultaneously and consistent violations of the restraints remain. Such can be detected by high values of the remaining potential energy (for which usually a penalty function, called the target function, is introduced).

The NMR ensemble is fitted afterwards so that the backbone atoms of the best-defined region superimpose the best. Residue-specific root mean square deviations (RMSDs) to the mean coordinates indicate how well the bundle of conformers is defined in that segment. Therefore, the bundle of structures is often shown since it easily allows to recognize which part of the protein is likely to be rigid (usually elements of secondary structure) and which parts are more flexible:

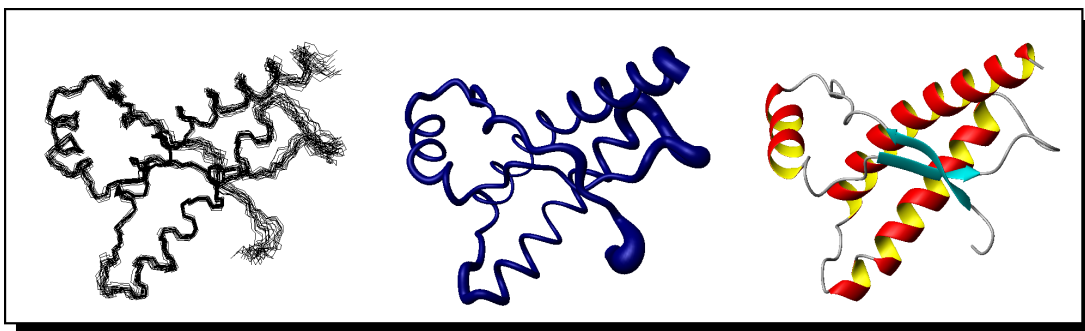


FIGURE 24. Left: Backbone representation of the 25 computed structures of the human prion protein. Middle: Spline representation through the C α trace. Right: Schematic representation highlighting the secondary structure.

In principle, poor definition of the protein conformation could also present a sampling problem often encountered for surface-exposed protons (NOEs can only be measured into a single direction in space).

Methylen protons can mostly not be assigned stereospecifically. Structure calculation programs therefore try to decide from local NOEs etc. if it is possible to decide which of the two possibilities is correct. Otherwise both possibilities are kept by introducing a so-called pseudoatom correction.

6. CRITERIA TO EVALUATE THE QUALITY OF NMR STRUCTURES

- Percentage of assigned resonances
- number of upper distance limits per residue
- value of the target function representing consistent violations of restraints
- number of violations
- distribution of dihedral angles in the Ramachandran-plot
- RMSDs for backbone and heavy atoms
in well defined structures RMSD < 0.5Å for backbone atoms and < 1Å for sidechain atoms

Example: Human prion protein

TABLE 25. Collection of the input for the structure calculation and characterization of the energy-minimized NMR structures of the polypeptide segment 121-230 in different human PrP constructs

Quantity*	hPrP(23-230)	hPrP(90-230)	hPrP(121-230)
NOE upper distance limits	1,732	1,705	1,752
Dihedral angle constraints	429	453	436
Residual target function, value, Å ²	0.25 ± 0.06	0.34 ± 0.08	0.39 ± 0.05
Residual distance constraint violations, Number [≥] 0.1, Å	0.3 ± 0.5	0.7 ± 0.9	0.1 ± 0.3
Maximum, Å	0.10 ± 0.01	0.11 ± 0.01	0.10 ± 0.00
Residual dihedral angle constraint violations Number [≥] 2.0 degrees	1.8 ± 1.0	1.5 ± 1.0	5.6 ± 2.0
Maximum, degrees	2.9 ± 0.8	3.0 ± 1.2	3.7 ± 0.9
AMBER energies, kcal/mol			
Total	[-]4824 ± 85	[-]4533 ± 79	[-]4698 ± 83
Van der Waals	[-]352 ± 16	[-]315 ± 15	[-]325 ± 16
Electrostatic	[-]5398 ± 84	[-]5164 ± 71	[-]5283 ± 67
rms deviation from ideal			

TABLE 25. Collection of the input for the structure calculation and characterization of the energy-minimized NMR structures of the polypeptide segment 121-230 in different human PrP constructs

geometry			
Bond lengths, Å	0.0084 ± 0.0002	0.0089 ± 0.0002	0.0084 ± 0.0003
Bond angles, degrees	2.25 ± 0.04	2.41 ± 0.04	2.29 ± 0.04
rms deviation to the averaged coordinates, Å			
N, C[α], C' (125-228)	0.65 ± 0.10	0.79 ± 0.11	0.81 ± 0.11
All heavy atoms (125-228)	1.06 ± 0.09	1.27 ± 0.10	1.26 ± 0.13
N, C[α], C' of regular secondary structures †	0.51 ± 0.12	0.60 ± 0.12	0.68 ± 0.13

* Except for the top two entries, the data characterize the group of 20 conformers that is used to represent the NMR structure; the mean value and the standard deviation are given.

† Secondary structure elements are formed by residues 128-131 ([β]-strand 1), 144-154 ([α]-helix 1), 161-164 ([β]-strand 2), 173-194 ([α]-helix 2), and 200-228 ([α]-helix 3).

7. LITERATURE

- Kurt Wuethrich: NMR of Proteins and Nucleic Acids, Wiley 1986
- John Cavanagh, Wayne J. Fairbrother, Arthur G. Palmer III, Nicholas J. Skelton: Protein NMR Spectroscopy Principles and Practice, Academic Press 1996
- NMR of Macromolecules - A practical approach, Oxford University Press 1993
- Ad Bax: Two-Dimensional NMR and Protein Structure, Annu. Rev. Biochem. 1989. 58:223-56

8. APPENDIX: GRAPHICAL PRESENTATION OF THE SPIN SYSTEMS FROM AMINO ACIDS

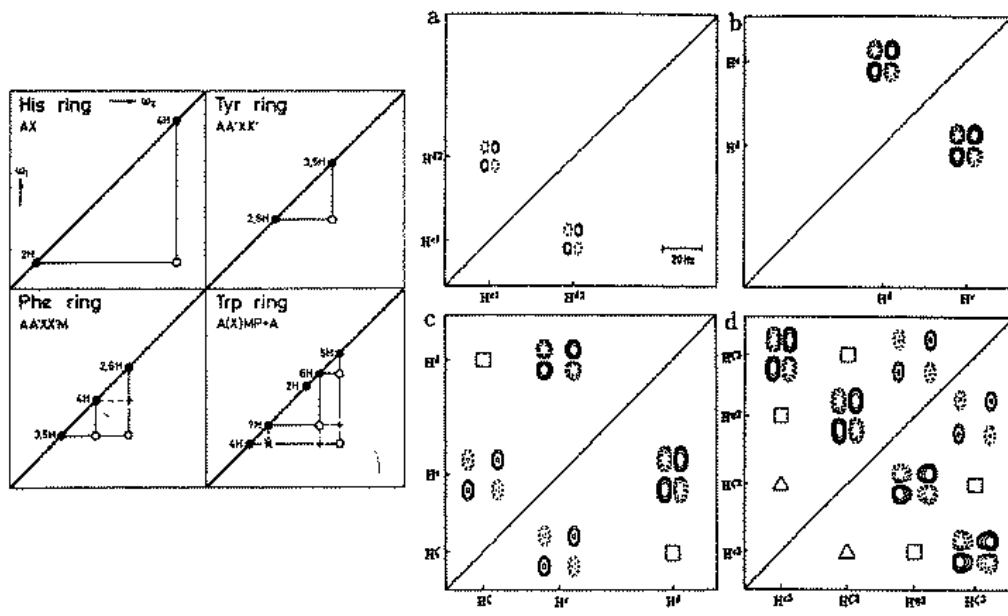


FIGURE 26. Sidechains of aromatic residues.

TABLE 27. Side-chain spin system pattern as observed in COSY or TOCSY spectra

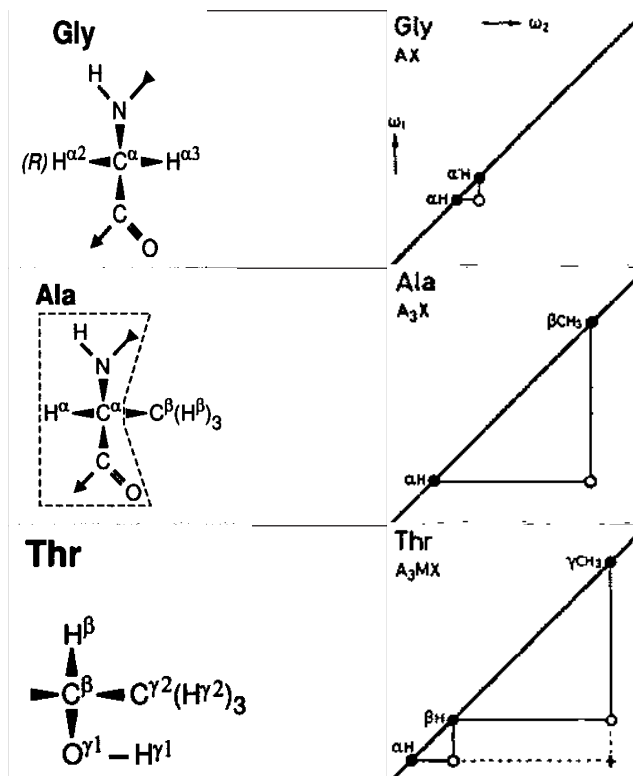


TABLE 27. Side-chain spin system pattern as observed in COSY or TOCSY spectra

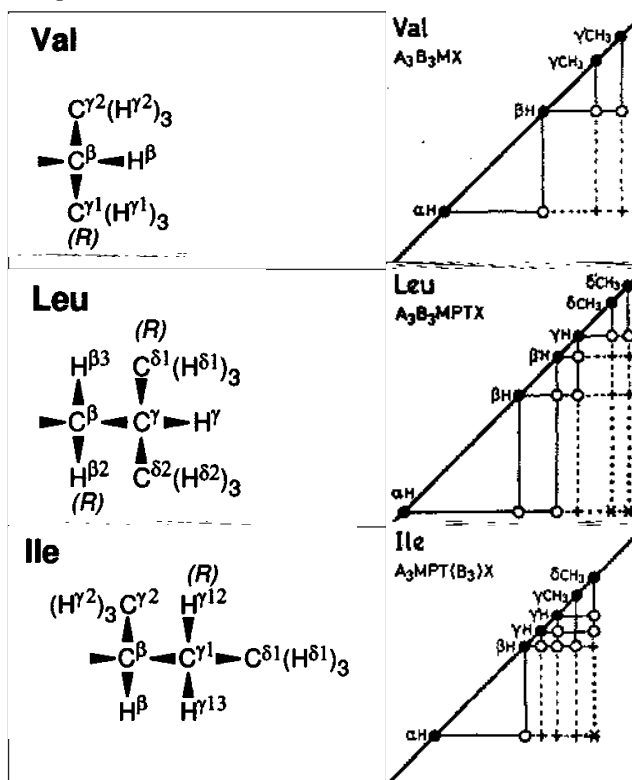


TABLE 28. Type-J spinsystems

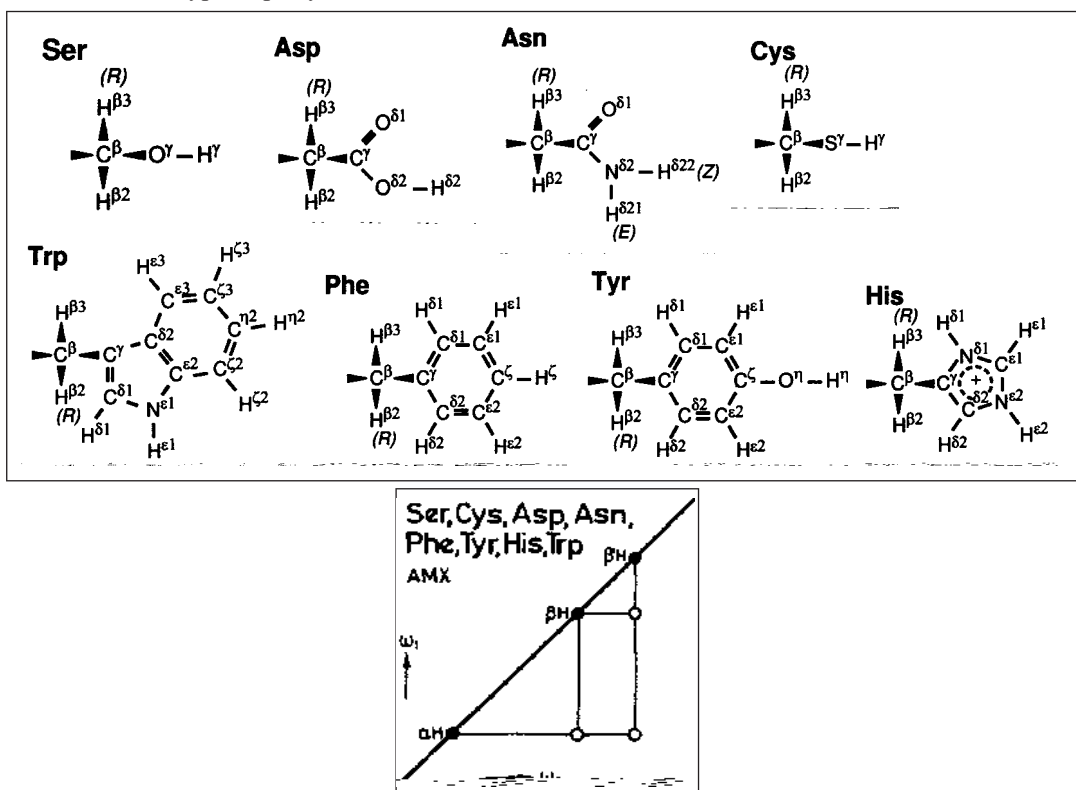


TABLE 29. Type-U spinsystems

