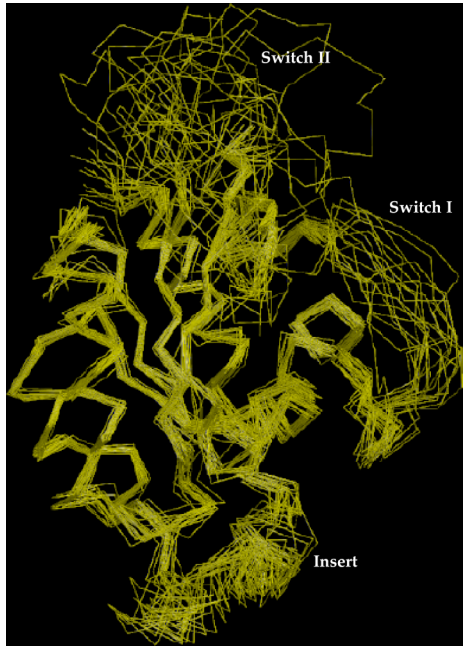
Solving protein structures by NMR



Qinghua Wang
1232B LGRT
545-3473
qwang@biochem.umass.edu

Concepts to be covered

- Why use NMR instead of crystallography?

- The basics of structure determination:
  What is the data, and how is it used?

- How to evaluate the quality of an NMR structure.
  How does it compare to a crystal structure?

We'll focus on solving protein structures using solution-state NMR, but many of the same considerations apply to nucleic acids and/or solid state techniques.

Comparing NMR and crystal structure determination

NMR structures	Crystal structures
Family or ensemble of structures calculated	Single structure calculated (more satisfying??)
Limited to smaller proteins	Applicable to small or large proteins
Protein must be soluble at 0.1-1 mM	Protein must be soluble at high concentration, and must then be crystallized
Can observe protein dynamics	Dynamic information is more difficult to extract

NMR has historically been applied to smaller proteins that do not crystallize, but this is changing.

- NMR is especially useful for studying protein complexes
- 20% of protein structures deposited in the Protein Data Bank in 2005 were determined by NMR

A typical representation of an NMR structure

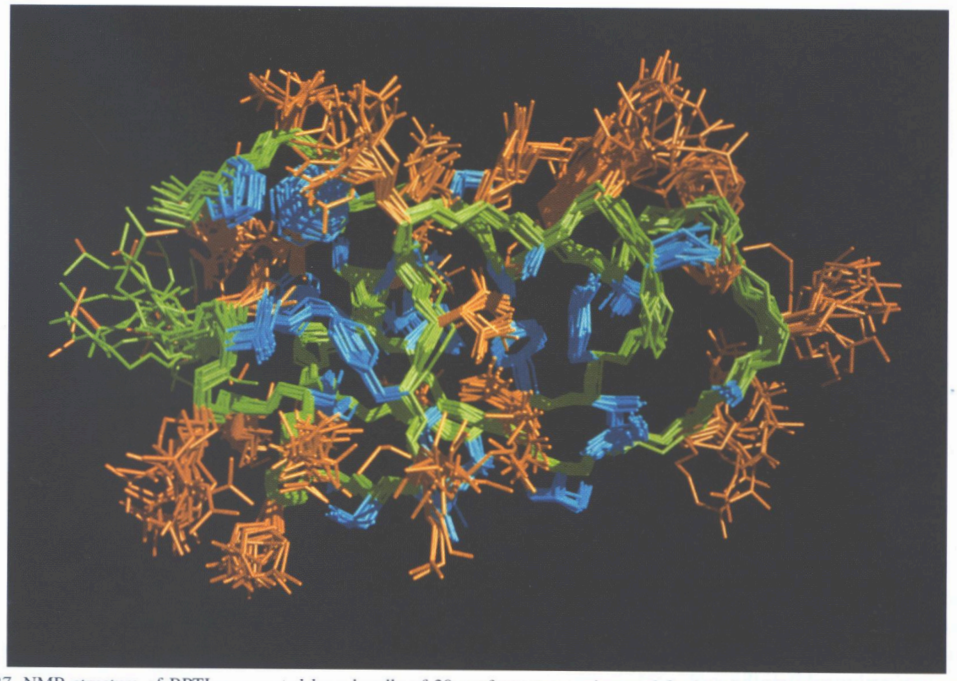


Figure 27. NMR structure of BPTI represented by a bundle of 20 conformers superimposed for best fit of the polypeptide backbone. The polypeptide backbone is green, core side-chains are blue, and solvent-accessible surface side-chains are red.

20 structures superimposed, all consistent with the available data

Wüthrich, *J. Biomol. NMR,* **27:** 13-39, 2003

- Backbone and core side chains usually better defined than the solvent-exposed side chains and the chain termini.
- Ill-defined regions may indicate conformational dynamics in solution or a lack of data-derived constraints in that region.
  - Dynamics can be confirmed by relaxation measurements
- Remember, proteins are not static! Dynamics can be substantial **and** functionally important.

Lecture Overview

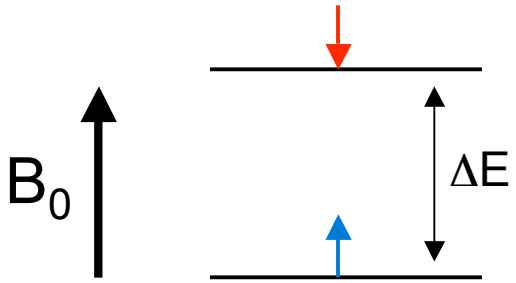
- Acquiring the data that go into the structure calculation
  - what is the NMR signal, and how is it measured?
- Introduction to the NOE and distance restraints
- The problem of sequential assignment

A brief description of nuclear magnetic resonance

Some nuclei have an intrinsic angular momentum called “spin”, which is quantized. A general rule:

Those with odd mass numbers have “spin 1/2”, e.g. ^{1}H, ^{15}N, ^{13}C, ^{31}P, ^{19}F.
Those with even mass numbers but odd atomic numbers are “spin 1”, e.g. ^{2}H, ^{14}N.

Consider the simplest case of a spin 1/2 nucleus when an external magnetic field B_0 is applied. Two energy states are possible.



High energy, spin aligned against magnetic field

Low energy, spin aligned with magnetic field

The energy difference between the two states (ΔE) determines the frequency of transitions between them

$$E = -\mu_z B_0 \qquad \mu_z = \gamma \hbar m_z$$

μ_z = magnetic moment
m_z = spin quantum number (+1/2 or -1/2)
γ= gyromagnetic ratio (constant for each nucleus type)
$\hbar = h/2\pi$ where h is Planck's constant

$$\therefore E = -m_z \hbar \gamma B_0$$

$$\Delta E = \hbar \gamma B_0 (= h\nu)$$

So the frequency of transitions is $\nu_L = \frac{\gamma}{2\pi} B_0$

$$\nu_L = \frac{\gamma}{2\pi} B_0$$

In order to achieve resonance (induce transitions), apply electromagnetic radiation of frequency ν_L.

- ν_L is also called the Larmor frequency of a particular nucleus.
- This frequency is in the radiofrequency (rf) region of the electromagnetic spectrum.

For attainable B_0 fields with even a high γ nucleus (e.g. ^1H), $\Delta E << kT$, so at equilibrium, the two states are nearly equally populated

{e.g. out of a million ^1H nuclei in an 800 MHz spectrometer (18.1 T magnet), only ~60 more spins will be aligned with the field than against it}

Small population differences means low sensitivity, so sample concentrations are usually in the millimolar-submillimolar range.

In the classical view, the interaction of the nuclear magnetic moment with the external magnetic field leads to rotation (or precession) about the axis of the field

•similar to the motion of a gyroscope

•precession frequency corresponds to ν_L

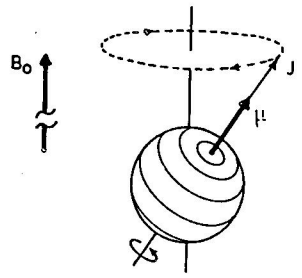


Figure 4.1 A spinning charge with angular momentum J, such as a proton, gives rise to a magnetic moment μ (= γJ). In a magnetic field its axis of rotation precesses around the direction of the field, like a gyroscope.

The surplus of spins aligned with the B_0 field form a cone of vectors, all rotating about the z axis.

In bulk view, they correspond to a net magnetization aligned with the z axis (M_0).

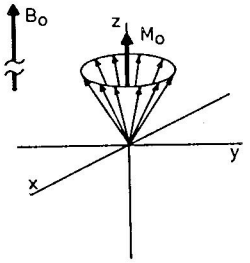


Figure 4.2 Out of a large collection of moments, a surplus have their *z* components aligned with the applied field, so the sample becomes magnetised in that direction.

Continuous wave NMR: In the early days, spectra were obtained by sweeping the magnetic field strength with a constant rf excitation applied (similar to measurement of an absorption spectrum)

- extremely time-intensive

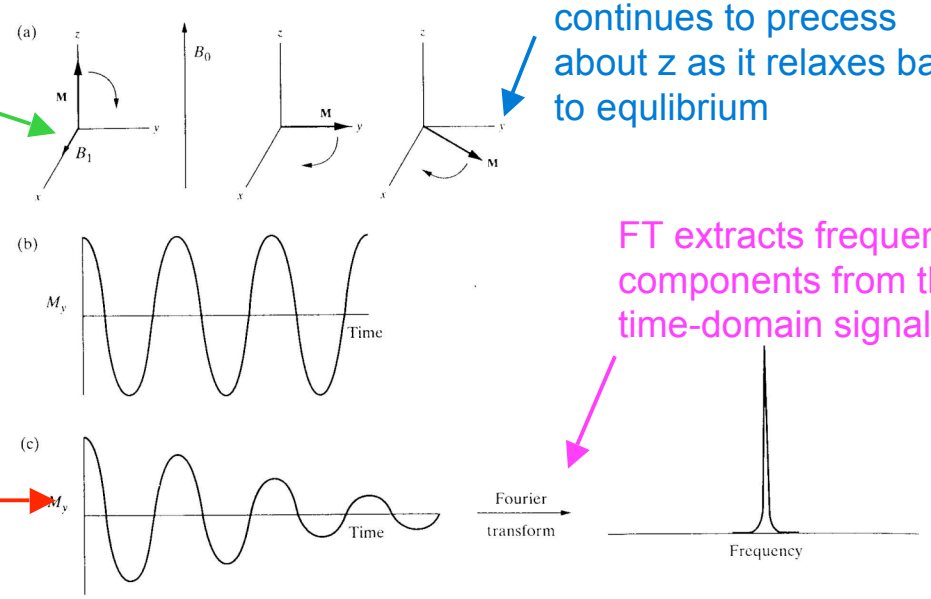
Pulse NMR: A short high power rf pulse is applied that perturbs the system from equilibrium, and then the system is observed as it relaxes.

- the time savings allows collection of multiple similar experiments in order to improve the signal-to-noise ratio.

Apply a 90° rf pulse to flip net magnetization to y axis

Net magnetization continues to precess about z as it relaxes back to equlibrium

FT extracts frequency components from the time-domain signal



the free induction decay or FID
•the rf signal monitored along y axis
•decays as the system returns to alignment with z

▲ FIGURE 10.30
from Modern NMR Techniques in Chemistry by Derome

The key to extracting structural information from the NMR signal:

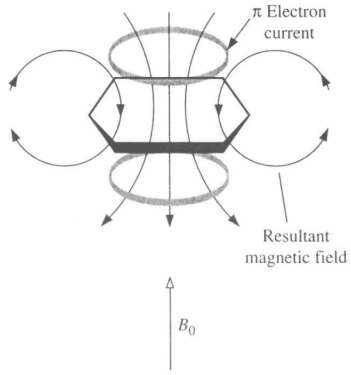
- not all spins precess at the same frequency
- dispersion along frequency axis is due to local electronic structures, which can shield or accentuate the external magnetic field.
- high electron density causes shielding, low electron density deshields.

$$\Delta E = (1-\sigma)\hbar\gamma B_0$$

$$\nu_L = (1-\sigma)\frac{\gamma}{2\pi}B_0$$

where σ is the shielding constant

Ring currents can cause complex effects.



◀ FIGURE 10.32
The ring current effect in NMR. When a planar aromatic molecule is placed in a magnetic field B_0, the mobile π electrons can be considered to flow in a current around the periphery of the planar molecule. Classically, a current loop generates a magnetic field as shown in the figure. The magnitude of the field decreases approximately as the inverse cube from the center of the loop. This model is approximate, but it provides a good qualitative understanding of the effect of neighboring aromatic molecules on chemical shifts.

from Modern NMR Techniques in Chemistry by Derome

Frequency is typically expressed relative to a reference frequency, and normalized to make it independent of magnetic field strength.

$$\delta = \frac{\nu - \nu_0}{\nu_0} \times 10^6$$

δ=chemical shift (ppm)

Each hydrogen atom (proton) in a protein resonates at a characteristic frequency on the chemical shift scale, defined by its local structural environment.

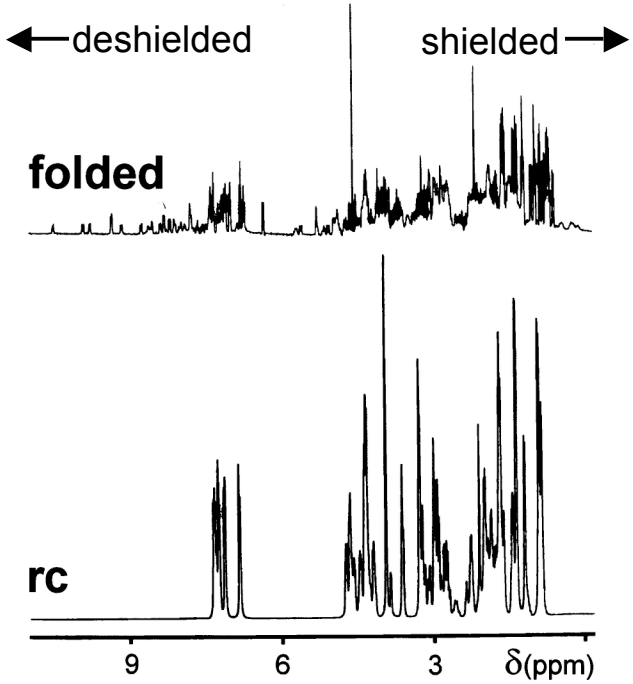


Figure 5. One-dimensional (1D) ^{1}H NMR spectra of the small protein bovine pancreatic trypsin inhibitor (BPTI, M ≈ 6000). Top: experimental spectrum of folded, active BPTI in a freshly prepared $^{2}H_2O$-solution. Bottom: simulated spectrum for the unfolded, random coil form of the BPTI polypeptide chain.

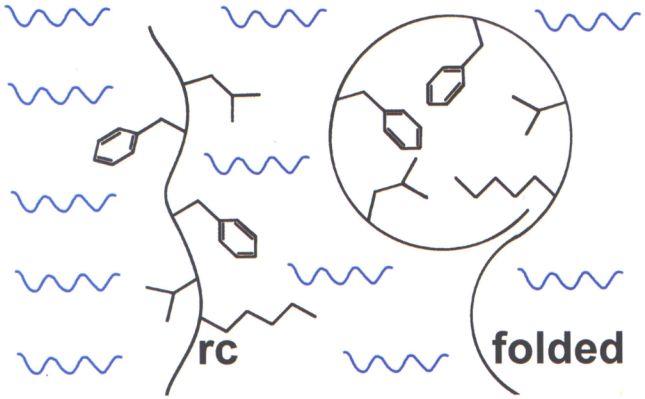


Figure 6. Scheme illustrating that the solvent (blue) has free access to all parts of a random coil polypeptide chain (rc), whereas it is excluded from the core of a folded globular protein.

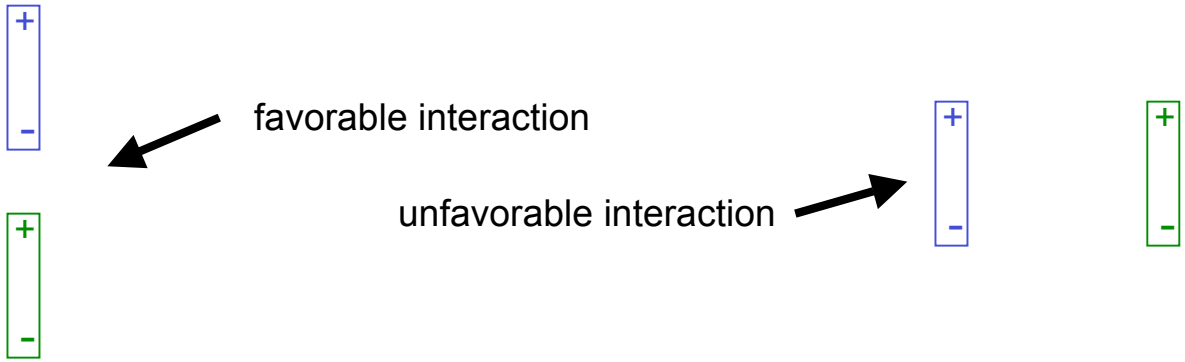
Wüthrich, *J. Biomol. NMR,* **27:** 13-39, 2003

Resonances in folded proteins are much better dispersed than in unfolded ones, due to shielding or enhancement of the externally-applied magnetic field by electrons in the local structure.

Nuclei can interact with one another:

Scalar couplings between nuclei connected by three or fewer bonds
- spin state of one nucleus can oppose or accentuate the external magnetic field at another, via the bond electrons

Dipolar couplings between nuclei that are close together in space (<~5Å)
Picture two dipoles at a fixed distance from one another.



Molecular tumbling causes rotation of the green nucleus about the blue nucleus, which experiences an oscillating electromagnetic field.

This leads to an optimal relaxation pathway in large slowly tumbling proteins in which the blue nucleus relaxes to the ground state by transferring magnetization to the green one.
Dipolar coupling is the basis of the nuclear Overhauser effect, or NOE.

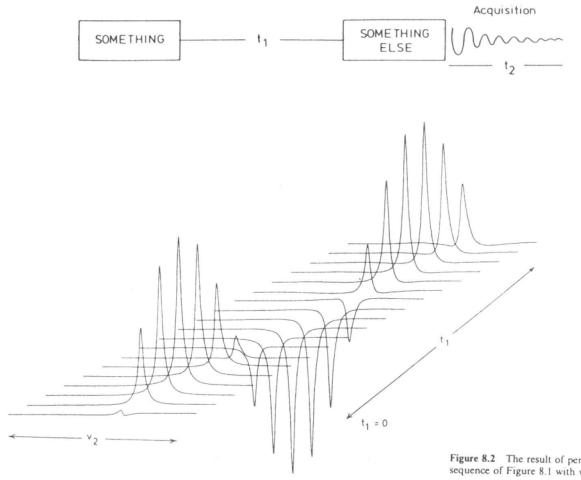


Figure 8.2 The result of performing the sequence of Figure 8.1 with variable t_1.

Figure 8.3 A slice from the data of Figure 8.2, taken parallel with t_1 through the tops of the peaks.

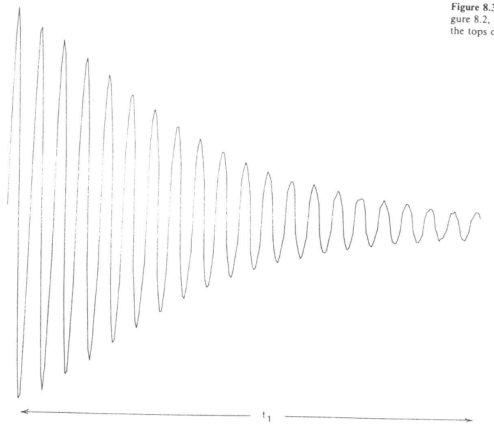
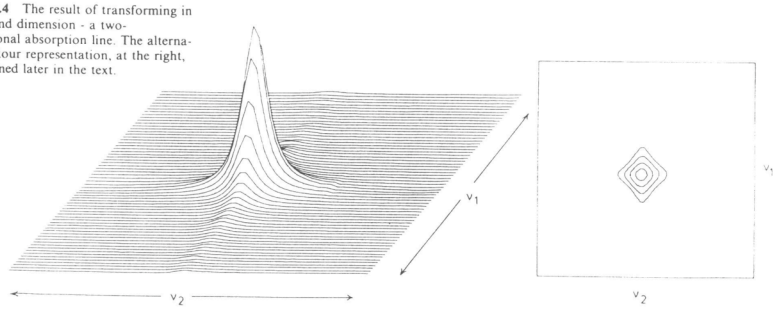


Figure 8.4 The result of transforming in the second dimension - a two-dimensional absorption line. The alternative contour representation, at the right, is explained later in the text.



from Modern NMR Techniques in Chemistry by Derome

1D NMR experiments are typically simple - an rf pulse is applied, followed by measurement of the free induction decay.

In a 2D NMR experiment, a more complicated sequence of pulses and delays is applied.

- 'something' is usually an rf pulse to get things going
- 't1' delay is initially short, but is incremented in successive experiments to create an indirectly-sampled time domain signal that encodes the chemical shifts of the first nucleus
- 'something else' is usually a mixing period to exchange magnetization between nuclei (via scalar or dipolar coupling)
- The directly-sampled FID at the end contains the chemical shift of the second nucleus.

For example: a two-dimensional NOESY correlation spectrum

Off-diagonal peaks correspond to NOEs between two protons that are close to each other in the protein

The intensity of the peak is proportional to r^{-6} (r = distance between protons)

Limited to protons within about 5 Å of each other

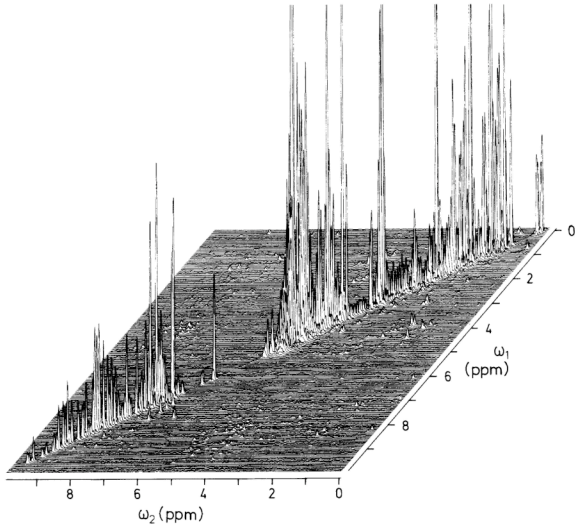


Figure 18 Two-dimensional (2D) [^{1}H,^{1}H]-NOE spectroscopy ([^{1}H,^{1}H]-NOESY). A stacked plot representation of a spectrum of the small protein bull seminal proteinase inhibitor IIA (BUSI IIA, M ≈ 6000) is shown (500 MHz, 45 °C, H_2O-solution).

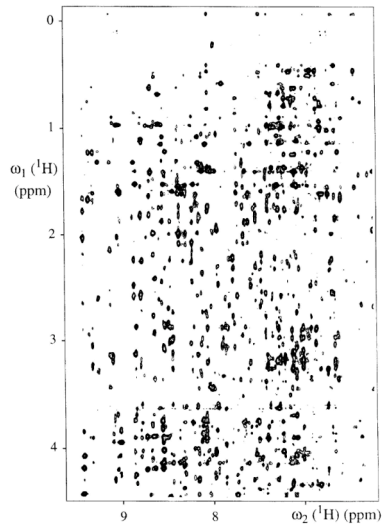


Figure 19 2D [^{1}H,^{1}H]-NOESY spectrum of the plant pathogenesis-related protein P14A (M ≈ 15000). A contour plot of the spectral region [ω_1(^{1}H) = 0–4.3 ppm, ω_2(^{1}H) = 6.3–9.5 ppm] is shown (750 MHz, 30 °C, H_2O-solution).

Wüthrich, *J. Biomol. NMR,* **27:** 13-39, 2003

NMR structure calculation relies primarily on NOEs

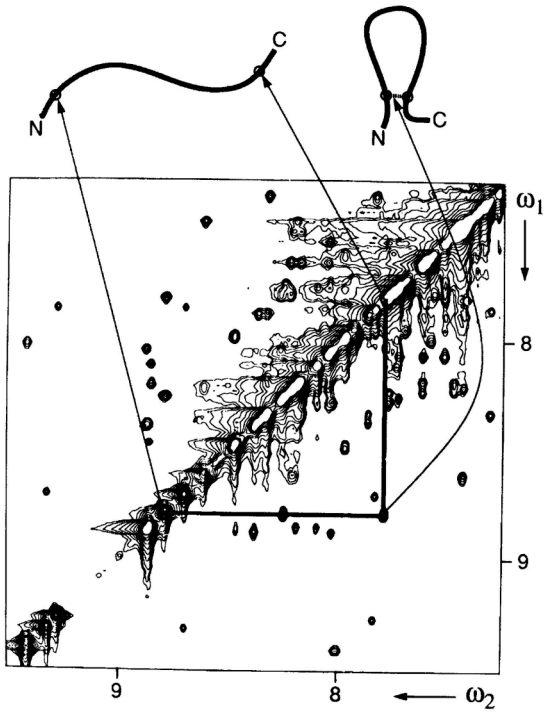


Figure 24 Scheme indicating the relations between an experimental 2D [^{1}H,^{1}H]-NOESY spectrum, a polypeptide with the chain ends indicated by N and C, sequence-specific assignments for two hydrogen atoms in the polypeptide chain indicated by circles, and the NOE upper distance constraint derived from the NOESY cross-peak connecting the chemical shift positions of the two assigned hydrogen atoms (see text).

Wüthrich, *J. Biomol. NMR,* **27:** 13-39, 2003

The Sequential Assignment Problem!

- If the chemical shift of each proton is known, every off-diagonal peak can be assigned to a short-distance interaction between two specific protons within the protein sequence.
- Peak volume or intensity relates to the interproton distance.
- Hundreds or thousands of inter-proton distance restraints are used to calculate three-dimensional structures that are consistent with the NOE data.

The Problem of Sequential Assignment

Solution: Use “through-bond” scalar couplings (as opposed to the “through-space” correlations that underlie the NOE) to discover which resonances in the spectrum are connected through bonds in primary sequence.

Different strategies are utilized for small proteins versus larger proteins, where peak overlap becomes more of a problem.

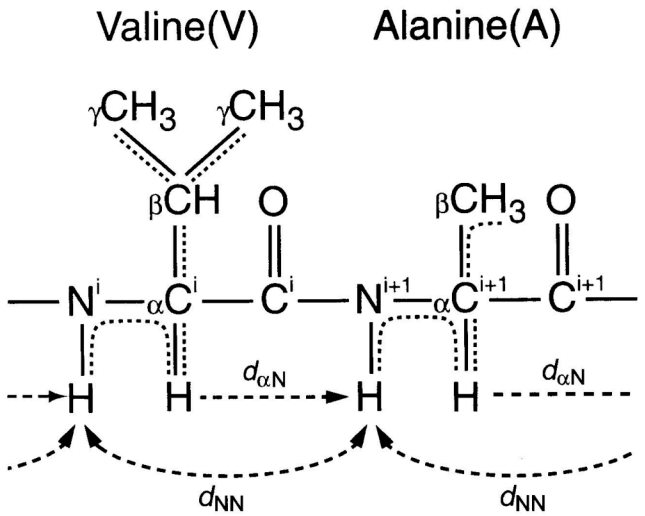


Figure 16 Sequential ^{1}H NMR assignment of proteins. The drawing shows the chemical structure of a – valine – alanine – dipeptide segment in a polypeptide chain. The dotted lines connect groups of hydrogen atoms that are separated by at most three chemical bonds and can therefore be connected using scalar spin–spin couplings. The broken arrows link pairs of hydrogen atoms in neighbouring amino acid residues that are separated by short through-space distances, $d_{\alpha N}$ and d_{NN}, and can therefore be connected by ‘sequential NOEs’.

Wüthrich, *J. Biomol. NMR,* **27:** 13-39, 2003

Sequential Assignment for proteins <10-15 kDa

▪Two-dimensional ^{1}H,^{1}H-COSY spectrum shows correlations between protons connected through three or fewer bonds (indicated by ·····, below left).

▪Each residue is a closed system in this experiment, called a "spin system", isolated by the carbonyl.

Can usually identify a spin system as a particular amino acid type based on the number of resonances and their chemical shifts.

▪Spin systems are connected sequentially using short-range NOE correlations from a 2D NOESY spectrum, usually $d_{\alpha N}$ and d_{NN} (indicated by ------, below left).

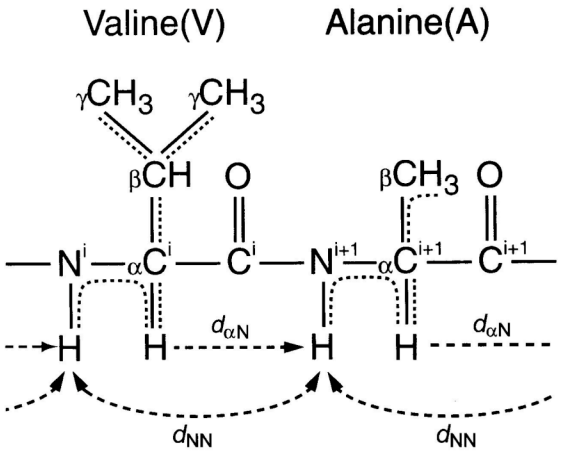


Figure 16 Sequential ^{1}H NMR assignment of proteins. The drawing shows the chemical structure of a – valine – alanine – dipeptide segment in a polypeptide chain. The dotted lines connect groups of hydrogen atoms that are separated by at most three chemical bonds and can therefore be connected using scalar spin–spin couplings. The broken arrows link pairs of hydrogen atoms in neighbouring amino acid residues that are separated by short through-space distances, $d_{\alpha N}$ and d_{NN}, and can therefore be connected by 'sequential NOEs'.

Wüthrich, *J. Biomol. NMR,* **27:** 13-39, 2003

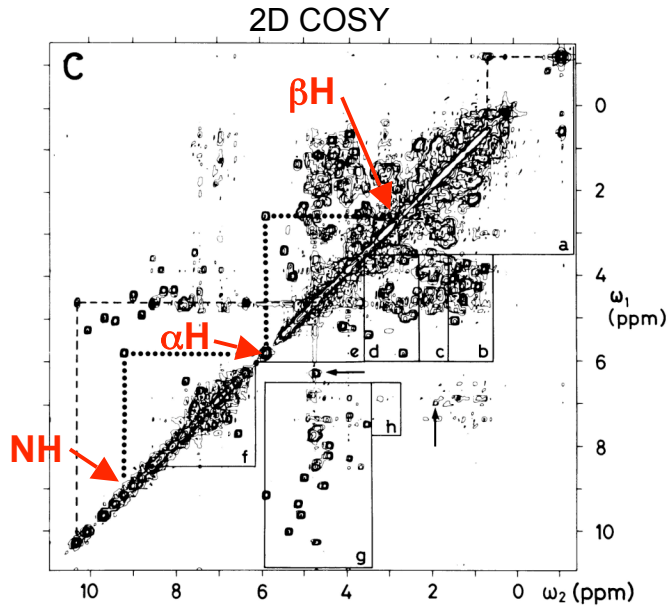


Figure 5.8. (*Continued*)

From "NMR of Proteins & Nucleic Acids, by K. Wüthrich, pp. 54-55

Sequential assignment for larger proteins (>15 kDa)

Two problems with larger proteins:

1. Many more protons lie in same spectral range, and peaks overlap.
2. Molecule tumbles more slowly as a whole, leading to broad peaks.

- Problem #1 can be overcome by labeling protein with other NMR-sensitive nuclei, such as ^{13}C and ^{15}N.

- Overcrowded spectra can then be spread out in additional dimensions.

- Accomplished by growing cells in a minimal growth medium with single carbon/nitrogen sources (*e.g.* ^{13}C-glucose and $^{15}NH_4Cl$ for *E. coli*).

- Disadvantage is the substantial cost of isotopic labeling.

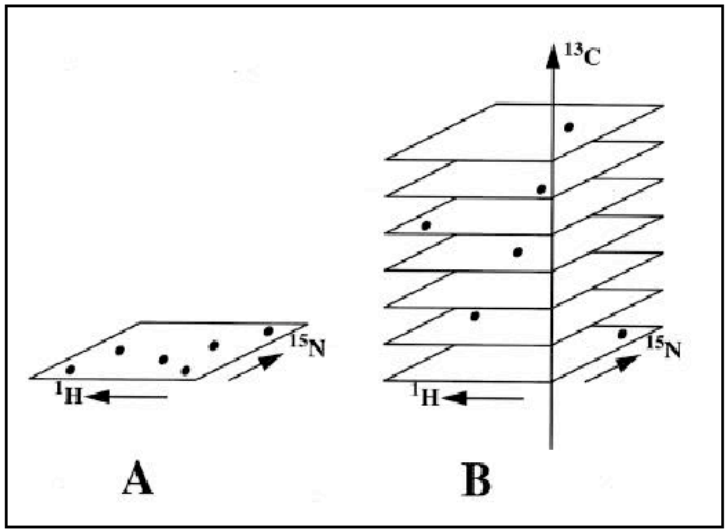


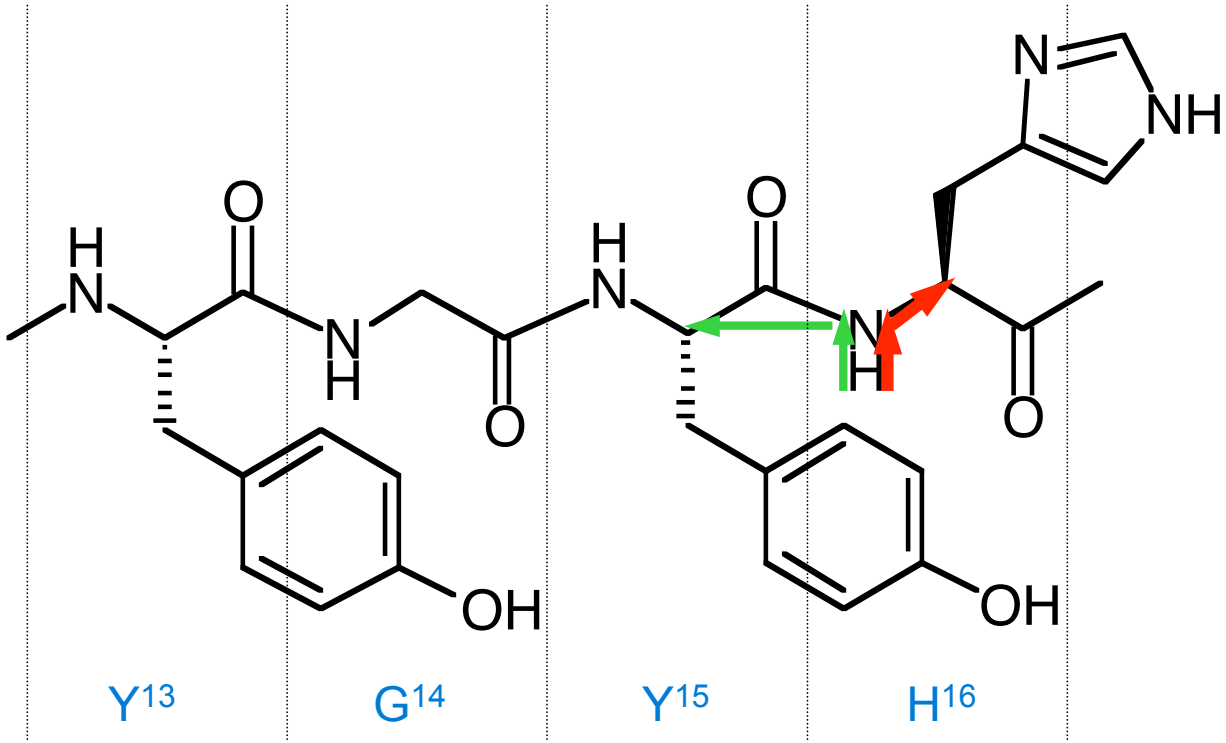
Figure 3. Schematic 2-D and 3-D NMR spectrum. The drawing illustrates the representation of a 3-D spectrum and the increased resolution obtained when going from 2-D to 3-D NMR spectra. The six peaks in the 2-D [^1H,^{15}N]-correlation spectrum (A) are separated in different planes of a 3-D spectrum (B) by an additional correlation with the α-carbon nuclei (^{13}C) attached to the nitrogen nuclei (^{15}N) in the same amino acid residue. The chemical shifts of the carbon nuclei are used to spread the resonances from the 2-D plane into a third dimension.

Wider, *Biotechniques,* **29:** 1278-1294, 2000

Sequential assignment for larger proteins (>10-15 kDa)

Carbon and nitrogen labeling enable tracing directly along backbone from one amino acid to the next via scalar “through-bond” couplings.

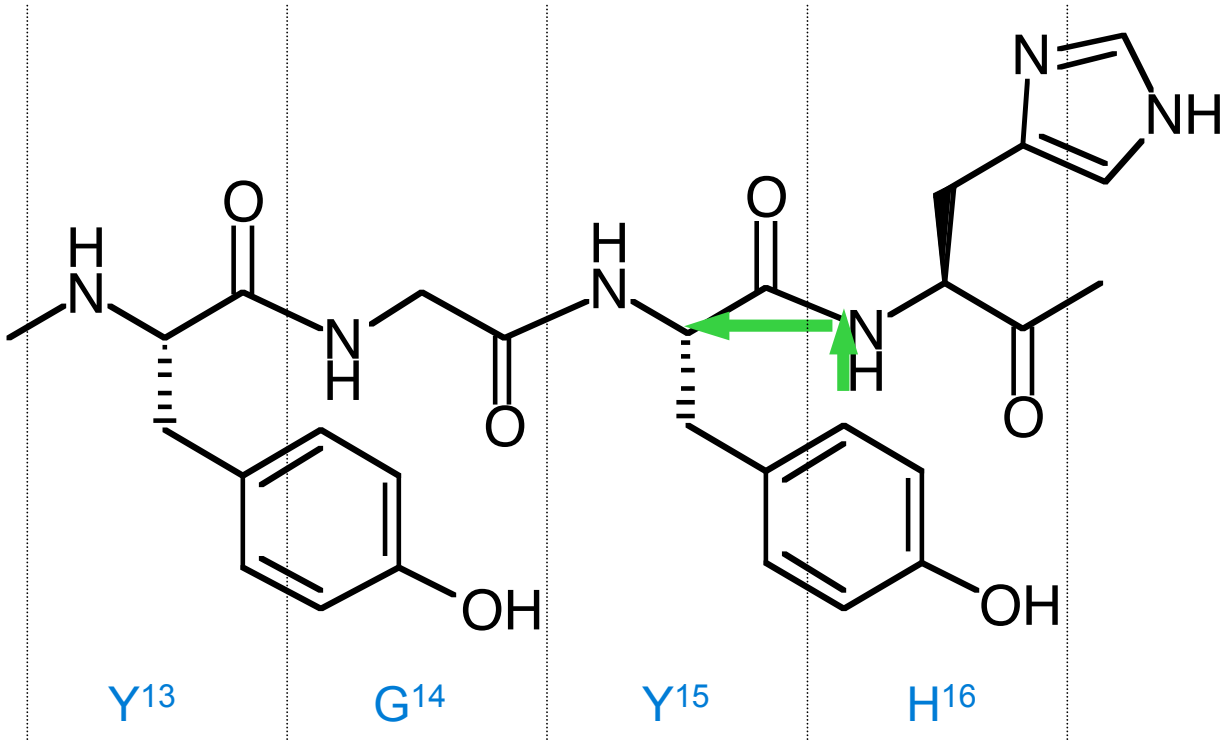
Example: An HNCA experiment yields a strong intra-residue correlation between the amide proton, nitrogen and alpha carbon, plus a weak correlation from the amide proton and nitrogen to the alpha carbon of the i-1 (preceding) residue.



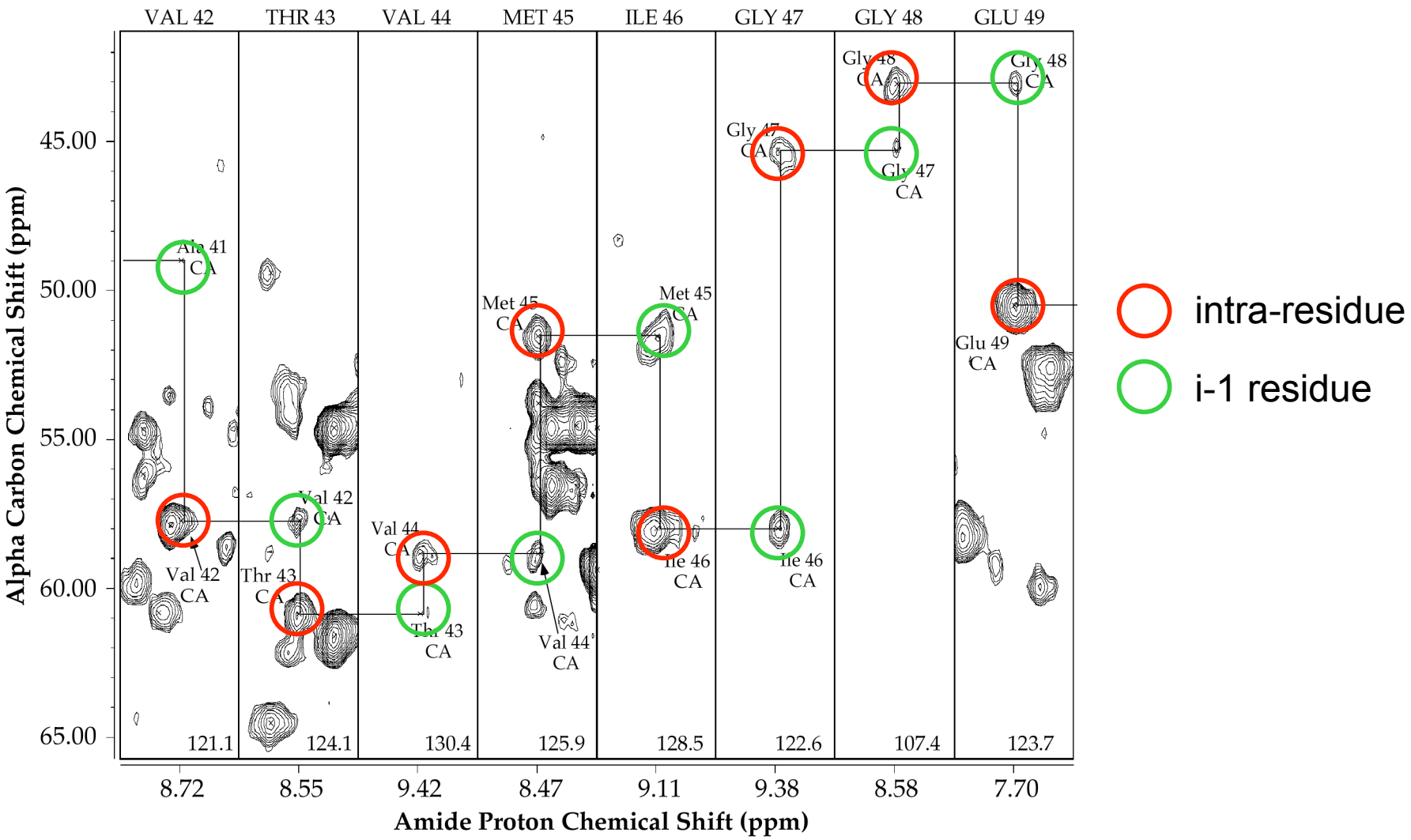
Sequential assignment for larger proteins (>10-15 kDa)

Carbon and nitrogen labeling enable tracing directly along backbone from one amino acid to the next via scalar “through-bond” couplings.

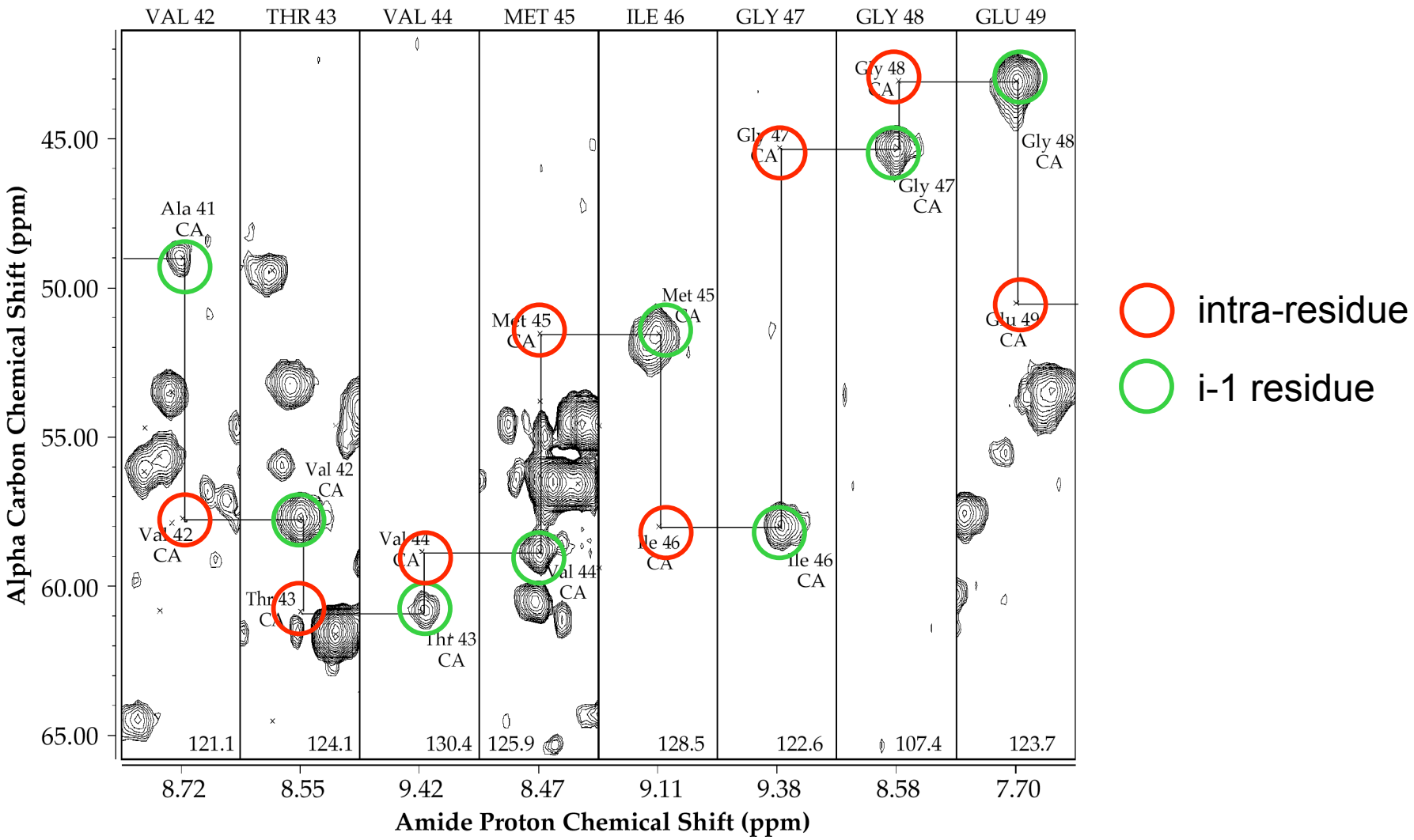
Example: An HN(CO)CA experiment yields only the correlation from the amide proton and nitrogen to the alpha carbon of the i-1 (preceding) residue.



Slices from a 3D HNCA experiment



Slices from a 3D HN(CO)CA experiment



Backbone sequential assignments proceed using these and other experiments like them (HNCO, CBCA(CO)NH, HNCACB etc.)

Next lecture:

-Side chain assignments
-Deriving structural information