R. A BRIEF INTRODUCTION TO STRUCTURE SOLUTION BY NMR

NMR is commonly used in structure solution for smaller biological molecules – proteins of 100-120 residues and RNA molecules of up to 50 nucleotides. The technique becomes more complicated when used with larger species. The strengths of the technique relative to crystallography are that (1) you don't need a crystal, (2) there are no crystal contacts to alter the conformation of the protein, (3) conditions are closer to those found in the cell and (4) there are a variety of advanced techniques that can address macromolecular dynamics that are not accessible by crystallography.

NMR Fundamentals

Recall that the NMR spectrum arises from an energy gap between two possible spin states for a nucleus (spin = ½ for our purposes) that arises in a strong external magnetic field. That gap corresponds to the energy delivered by photons in the radiofrequency range (Figure R.1). The larger the magnetic field, the larger then energy gap. In addition, the local chemical environment plays an important role in determining the frequency of radiation that is capable of promoting the transition between the two spin states, providing a window into the structure of a molecule based on the distinct influences experienced by a given nucleus. Because of the familiarity of ^{1}H NMR, we will focus on that nucleus, though ^{13}C and ^{15}N nuclei are also important for structure determination in biochemistry.

Chemical shift

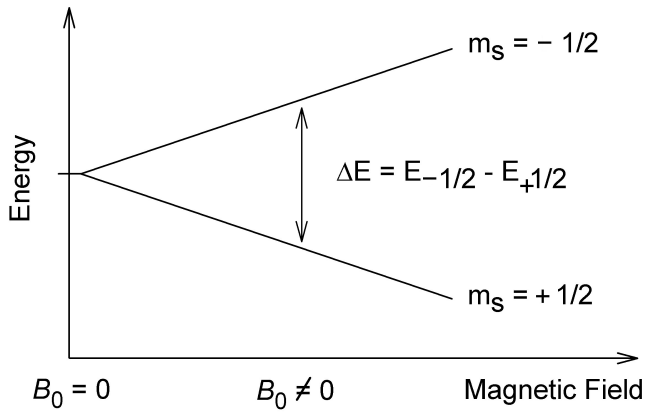

Figure R.1. Spin ½ nuclei have identical energies (are degenerate) in the absence of an external magnetic field (B_o), but as the field increases, the energy difference increases as well, requiring a more energetic photon to excite the transition. (Image taken from http://en.wikipedia.org/wiki/Nuclear_magnetic_resonance).

In Figure R.1, the resonant frequency of a given nucleus (the frequency that excites a transition) is shown to vary with the external magnetic field. In addition, the resonant frequency also varies due to the influence of electrons surrounding the nucleus. The less electron density surrounding the

nucleus, the more the external magnetic field is sensed by it and the larger the energy gap between spin states. Typically, we use a **reference compound** that places a lot of electron density around the nucleus of interest (tetramethyl silane is commonly used for protons) and define that arbitrarily as the nucleus resonating at the nominal frequency of the instrument (ie 400 MHz on our Bruker). Other protons will be less **shielded** by electrons and will resonate at a higher energy/frequency. The difference in energy is quite small. For example, if the nominal frequency is 400 MHz, sample protons will resonate at frequencies 4-40 Hz (note the lack of the "Mega" prefix) higher than the standard. These differences are 1-10 parts per million higher in frequency, and we define chemical shift by the parameter "δ" which gives the increase in frequency of resonance in parts per million (ppm).

For proton NMR spectroscopy in proteins, there are two significant determinants of chemical shift – the presence of heteroatoms (N, O and S) in the vicinity and the presence of π electrons in aromatic ring systems (the latter amplify the external magnetic field for protons due to a "ring current", which I'm not going to get into here). Both lead to a larger magnetic field at the nucleus and a larger energy gap. In an NMR spectrum that is displayed as a "downfield" (leftwards) shift from the reference standard. A list of shifts associated with commonly encountered proton functionalities in proteins is given in Table R.1.

Table R.1. Typical chemical shifts for ^{1}H in various common functional groups.

Functional group	**Typical chemical shift (ppm)**
aliphatic protons	0-2
protons α to a carbonyl	2-3
protons adjacent to a hydroxyl	4-5
protons on α carbon	4-6
aromatic protons	7-8
amide protons	7-9

Through Bound (J) Coupling

Another important feature of the NMR spectrum that you have used in Organic Chemistry is spin-spin coupling between nuclei. Chemically non-equivalent nuclei within three (and sometimes four) bonds will interact with each other magnetically, so that the magnetic field of one is experienced by the other. So long as we are restricting the conversation to spin ½ nuclei (^{1}H, ^{13}C and ^{15}N), the "splitting" of the signal from one nucleus by the possible orientations of the magnetic field of nearby nuclei is given by the n+1 rule. If a given proton is adjacent to a methyl group, the three protons on the methyl group (n=3) will split the resonance of the proton of interest into a quartet. Two methyl groups, as in an isopropyl group, will split the resonance into a heptuplet. And so on.

As a side note, in organic chemistry one doesn't often observe coupling between protons on heteroatoms (like N and O) and protons on carbon. For example, the spectrum of methanol

typically includes two singlets, a singlet from the methyl group and a broad singlet from the hydroxyl group. The reason no coupling is observed is because of rapid exchange of the hydroxyl proton in the sample. In the course of a single pulse and acquisition sequence, a proton may exchange millions of times from a single oxygen meaning that no fixed set of nuclear spin states will be associated with a single molecule. If the rate of exchange can be slowed, by lowering the temperature for example, then the coupling will be observed. The same issue applies to N-H protons in amides. Proton exchange in amides is base catalyzed, so lowering the pH slightly (pH 6 is enough) allows the coupling pattern between N-H protons and C_α protons to be observed in protein samples.

We won't dwell on splitting patterns, but it is worth noting that the magnitude of splitting (referred to as the J value). The magnitude of this interaction is referred to as the coupling constant (J) and is typically about 7 Hz for vicinal protons on carbons that rotate freely with respect to each other, roughly 1 in 10^8 as strong as the external magnetic field and it is independent of the magnetic field, since it is an internal effect. In molecules without freely rotating carbon-carbon bonds, the magnitude of the coupling constant can be related to the dihedral angle between nuclei. For protons, this relationship is depicted in Figure R.2 and is often described as the Karplus correlation (for Martin Karplus, who provided the theoretical basis for it).

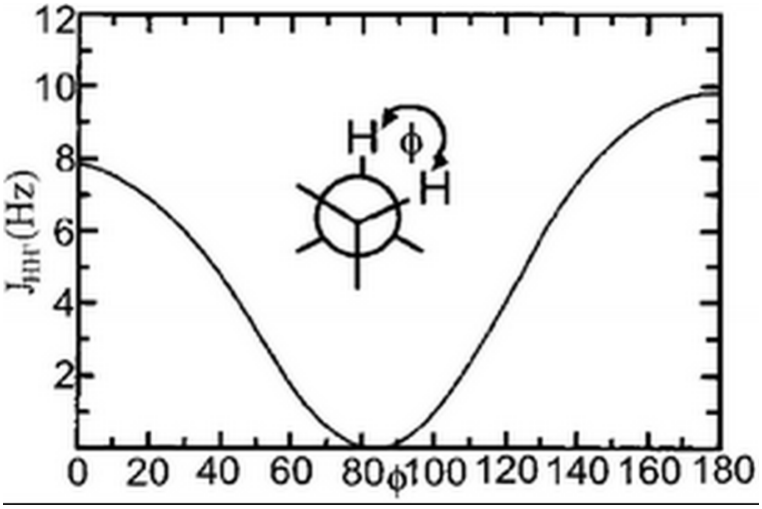

Figure R.2. Plot of the coupling constant vs. dihedral angle. This relationship allows one to extract conformational information from a simple NMR spectrum.

While pairs nuclei engaged in J coupling can often be identified because of a share J value (for example, in ethanol the J values for the triplet and quartet are identical), connectivity can also be identified via techniques we will describe qualitatively below (see COSY spectrum). For now it is enough to know that J coupling is a powerful way to identify protons closely linked to each other through bonding

Through Space Coupling

It is not discussed in OChem, but another form of nuclear coupling takes place through space, rather than through bonds. The **Nuclear Overhauser Effect** (nOe; named for Al Overhauser) refers to a dipole-dipole interaction between nuclei that permits energy transfer between nuclei that are 5 Å or less apart from one another. The strength of this signal decreases as $1/r^6$ (as do all dipole-dipole interactions) and is therefore extremely weak at longer distances. Steady irradiation at the resonant frequency of nucleus B in close proximity to nucleus A causes rapid spin relaxation of nucleus A, a larger ground state population of nucleus A, and therefore a stronger resonant signal for nucleus A during the NMR experiment. That strengthening of the signal provides important structural evidence for the proximity of two nuclei.

Two-Dimensional Spectra

One of the big problems associated with the NMR spectra of proteins is the number of resonating nuclei present in a sample. There are hundreds to thousands of protons creating a forest of signals that are nearly impossible to tease apart. One way to counter this problem is to use a larger magnetic field in collecting the spectrum. Since the chemical shift observed for a resonance is independent of magnetic field (when measured in ppm) the absolute energy difference between two peaks 1 ppm apart is twice as large in a 600 MHz spectrum as in a 300 MHz instrument. However, J values are constant and so two triplets that overlap in a 300 MHz experiment might be teased apart at 600 MHz. Protein NMR spectra are typically collected on instruments that operate between 600 to 1000 MHz. They are huge. And expensive. A 1000 MHz instrument in France is over 5 m high, weighs more than 12 tons and cost well over $10 million. But even that kind of magnetic field is insufficient to truly expand the reach of NMR techniques to protein structure using typical spectra. Instead, multi-dimensional NMR spectra are collected to expand the spectrum, normally presented along a single axis, to cover a 2-dimensional plane (or 3D or even 4D...). We will look at two basic forms of 2-dimensional NMR spectroscopy, one that relies on J coupling (COSY) and one that relies on nOe's (NOESY).

Correlated Spectroscopy (COSY and TOCSY)

The physics behind COSY spectra is complicated but the interpretation of the spectra is relatively simple. For a ^{1}H-^{1}H COSY spectrum, the proton spectrum for a compound is plotted on each of two axes, creating a square plot. Peaks on a 2D plot are like mountains on a topographical map – they are displayed using contour lines. So it is like looking down on a mountain range from an airplane, where the peaks are easily separated, instead of from the ground, where they all pile in front of each other.

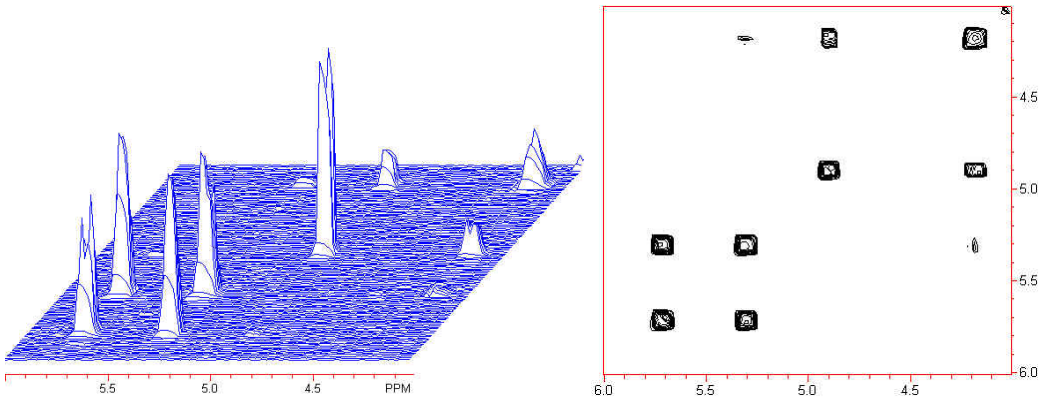

Figure R.3. Two different displays of a portion of the 2D spectrum of codeine. At the left, the spectrum is displayed "side on" and at right from a "top down" view. The black "dots" are the peaks shown at left. (http://www.acornnmr.com/codeine/cosy.htm)

In a COSY spectrum, **cross-peaks** arise whenever protons giving rise to a resonance on one axis are within three bonds of protons giving rise to a resonance on the opposite axis. A simple 2D COSY spectrum, for 2-butanone, is given in Figure R.4. The 1D spectrum (plotted for convenience here on both axes) is readily interpretable based on coupling constants. The ethyl group gives rise to the triplet and quartet (resonances "b" and "c") while the acyl methyl group gives rise to a singlet "a". The only J coupling between protons in this simple molecule is between the protons in the ethyl group. We see that as two cross-peaks (they are mirror images) labeled "b-c" on the spectrum. Those peaks arise because of coupling between the protons giving rise to the peaks labeled b and c. There is no cross-peak for the singlet "a" because that methyl group has no coupling to other protons in the molecule. On the other hand there are peaks for all three resonances **on the diagonal**. Each set of protons couples with itself so we three cross-peaks related to self-correlation. These are trivial peaks in the 2D spectrum, but they are useful because it helps confirm the presence of three resonances in the 1D spectrum.

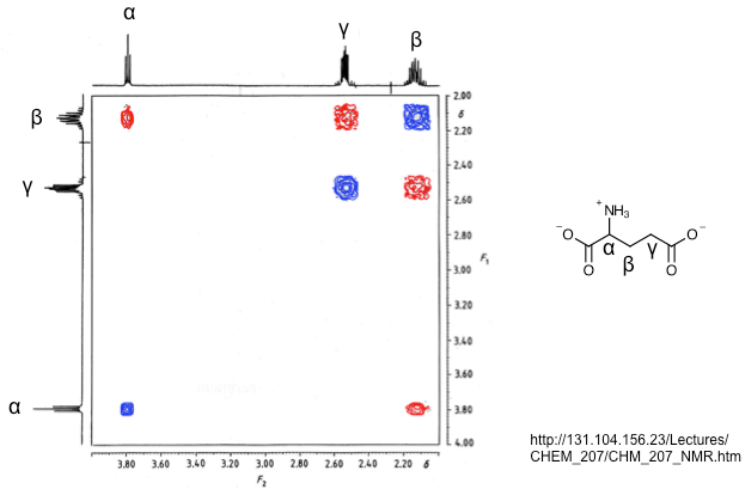

Figure R.4 2D COSY spectrum of glutamate. The diagonal peaks are highlighted in blue. (http://131.104.156.23/Lectures/CHEM_207/CHM_207_NMR.htm)

Before we move on, consider a slightly more complex situation. Figure R.5 displays a simulated 2D COSY for N-acetyl valine. Note that the acetyl group, as in 2-butanone, has no cross peaks off the diagonal, but in a more complex molecule, some protons give rise to multiple cross peaks (for example, the C_α proton couples to both the N-H proton and the C_β proton).

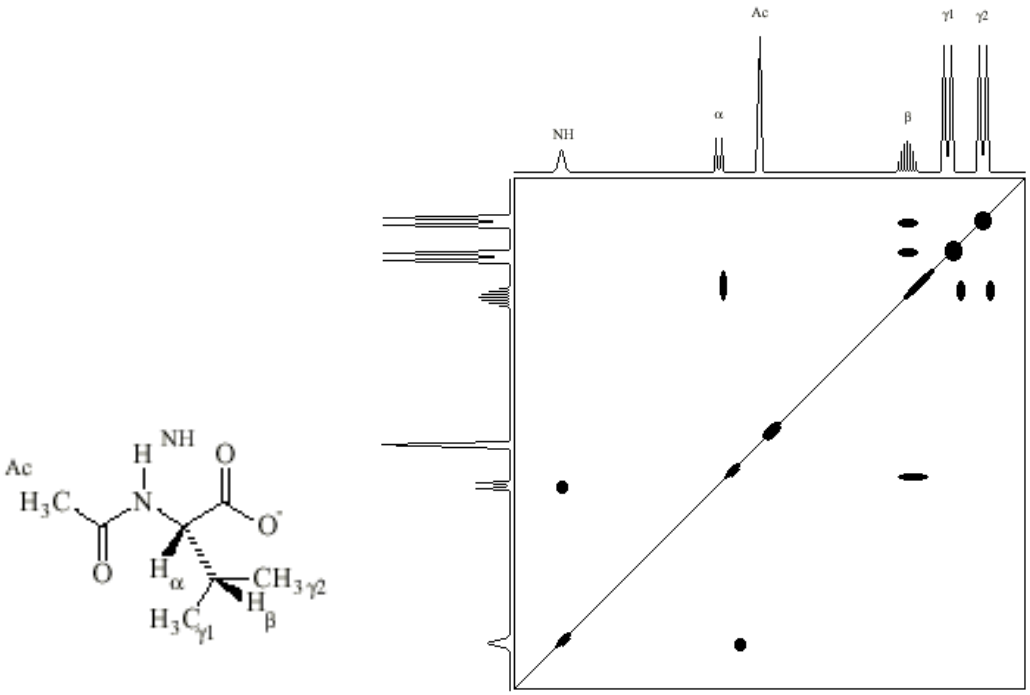

Figure R.5. Simulated COSY spectrum for N-acetylvaline. The top spectrum is labeled as shown for the molecule.

Total correlated spectroscopy (TOCSY) is an alternate strategy that relies on through bond coupling to derive cross-peaks in 2D NMR. This form of spectroscopy investigates *spin systems* rather than pairwise coupling between nuclei. The spin system is comprised of all nuclei that can be linked to each other through a chain of pairwise coupling. For example, consider butanone and butyraldehyde (Figure R.6). Butyraldehyde has one spin system while butanone has two. Why? In the aldehyde, the protons of the methyl group are coupled to the adjacent methylene. That methylene group is in turn coupled to the second methylene group. Finally the second methylene group is coupled to the aldehyde proton (albeit weakly). That chain of linkages places all protons in one spin system. In 2-butanone, however, the ethyl and methyl groups are separate spin systems because there is no coupling taking place between any set of protons in the ethyl group and any other set of protons in them methyl group.

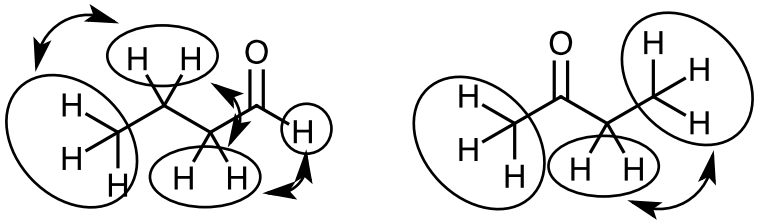

Figure R.6. Spin systems in butyraldehyde (one spin system) and 2-butanone (two spin systems). Note that butyraldehyde presents a chain of coupling that ultimately links the left hand methyl group to the aldehydic proton on the right. In butanone, there is no coupling between the methyl group on the left to the methylene group on the right. Hence, two spin systems.

TOCSY shows a cross peak between sets of protons that are part of the same spin system. Returning to N-acetylvaline, a simulated TOCSY spectrum (Figure R.7) shows cross peaks not present in the COSY spectrum (Figure R.5). Those new peaks, colored green, are between protons not on adjacent carbons (or nitrogens) but between protons that are all part of the valine spin system. Only the acetyl group is part of a separate spin system, since the acetyl methyl is coupled to no other protons. While this is confusing to look at in comparison to the COSY spectrum, it is a powerful technique because it allows the spectroscopist (or her trusted software) to rapidly identify protons coming from a single side chain (or at least part of that single side chain).

NOESY Spectra

NOESY spectra go one step further than TOCSY by completely ignoring issues of through bond coupling and looking for protons that are within 5 Å of each other. This will often create cross-peaks between protons that are in the same spin system (they are often close to each other), but importantly it will also reveal proximity of protons not in the same spin system. In N-acetylvaline (Figure R.8) we now see cross-peaks between the acetyl methyl group and the amide NH and Cα proton.

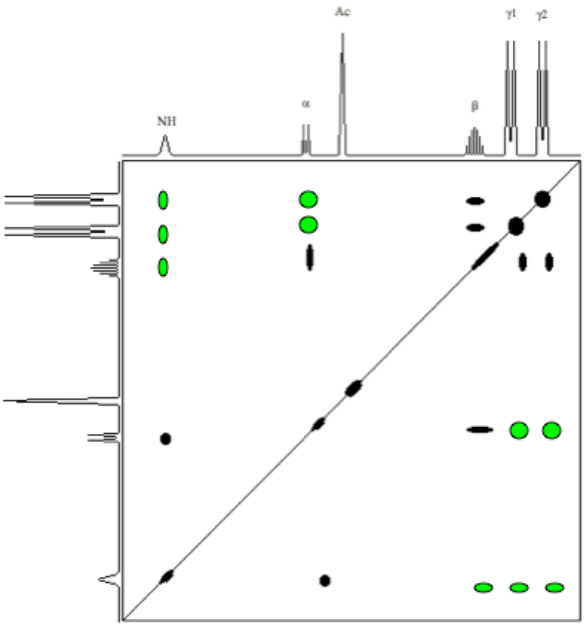

Figure R.7. TOCSY spectrum of N-acetylvaline. New peaks, not seen in the COSY spectrum are shown in green. They occur between protons that are in a common spin system, but not on adjacent carbons. For example, cross-peaks can now be found between the Cα and the Cγ protons.

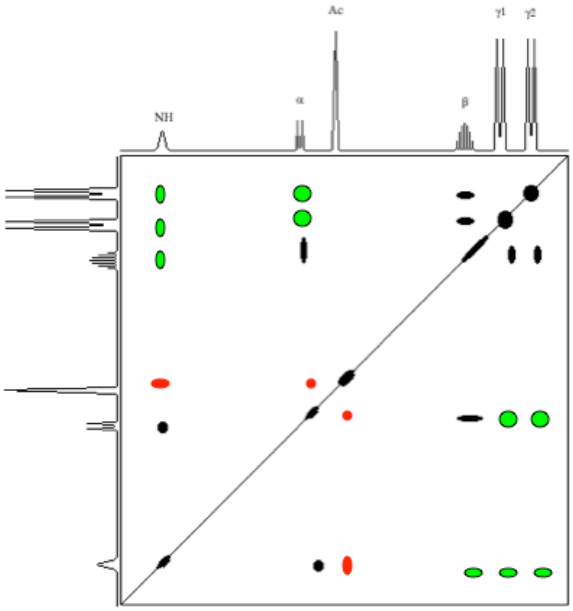

Figure R.8. NOESY spectrum of N-acetylvaline. The red cross-peaks show unique NOESY interactions between the acetyl methyl protons and the NH and Cα protons. No through bond coupling links those protons.

Using 2D NMR to Obtain 3D Structures

I will present a *highly* simplified scheme for using COSY and NOESY spectra to determine the three-dimensional structure of a protein. The methodology using takes place in two steps: sequence assignment and conformational restraint. For this example we will examine a short 10-residue peptide that folds into an anti-parallel β-sheet. In any NMR experiment, one knows the sequence of the protein one is working with from the start. The goal is not to get the covalent structure of the protein, but rather to see how that covalent structure is folded in solution.

Sequence Assignment

Let's say that our peptide has the sequence SLAMGPVDAT. COSY and TOCSY spectra will allow us to identify sets of resonances as belonging to a given residue. The serine will be obvious because of the chemical shift of the alcohol side chain, the leucine is unique in the coupling pattern associated with an isobutyl side chain, etc. The one difficulty is that the spectrum will show two sets of resonances that belong to alanine (Ala3 and Ala9 in the sequence). It's like having lunch bags in a kindergarden classroom all labeled with names, but there are two kids named Dana in the room. You need more information to figure out which lunch goes with which kid.

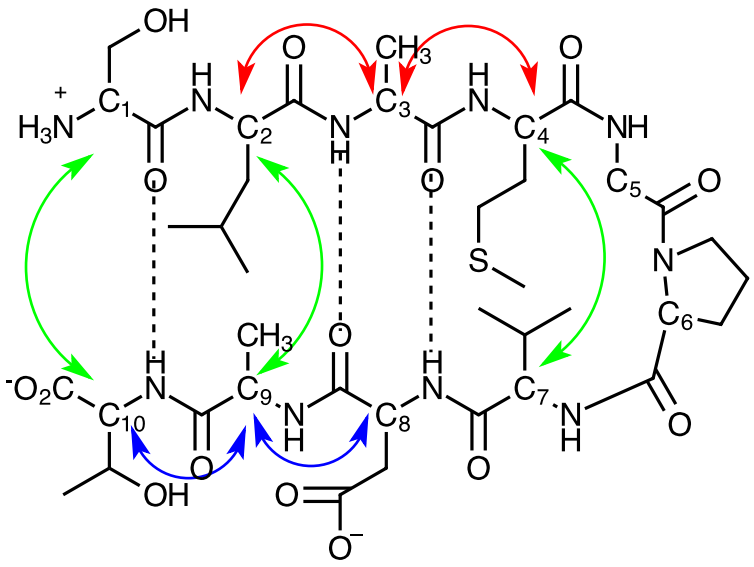

Figure R.9. Sequence and structure assignment by NOESY. Sequential NOESY cross-peaks between Ala3 and Leu2/Met4 and between Ala9 and Asp8/Thr10 can be used to tell which resonances belong to each of the two alanine residues. Long range NOE's indicate a pattern of proximities that strongly suggest an antiparallel pair of β-strands for this peptide.

To determine which set of alanine resonances belong to each alanine, one uses short range NOE interactions (Figure R.9).[1] The NOE signal decreases in strength as distance increases, but the

[1] Nowadays, one can also do the neighbor analysis by COSY. Spectroscopists often label protein with ^{13}C and ^{15}N so that J coupling can be observed across the peptide bond.

strongest NOE's will take place between adjacent residues in the peptide. Thus the strongest cross-peaks made between each of the two alanine residues will likely be to neighboring residues. Ala3 will show a close correlation to the Leu and Met protons while Ala9 will show sequential cross-peaks with Asp and Thr. Thus the ambiguity in sequence assignment is resolved.

Getting the Three Dimensional Structure

The final job is to see what conformation is adopted by the peptide. In my make-believe world, this peptide forms an anti-parallel β sheet. As a result NOESY cross-peaks (less intense than the sequential cross-peaks) will arise between residues brought into proximity by the folding of the peptide. Ser1 will be close to Thr10, Leu2 next to Ala9, Ala3, near Asp8, and so on (Figure R.9). Those NOESY cross-peaks will be critical in constraining our interpretation of the peptide structure.

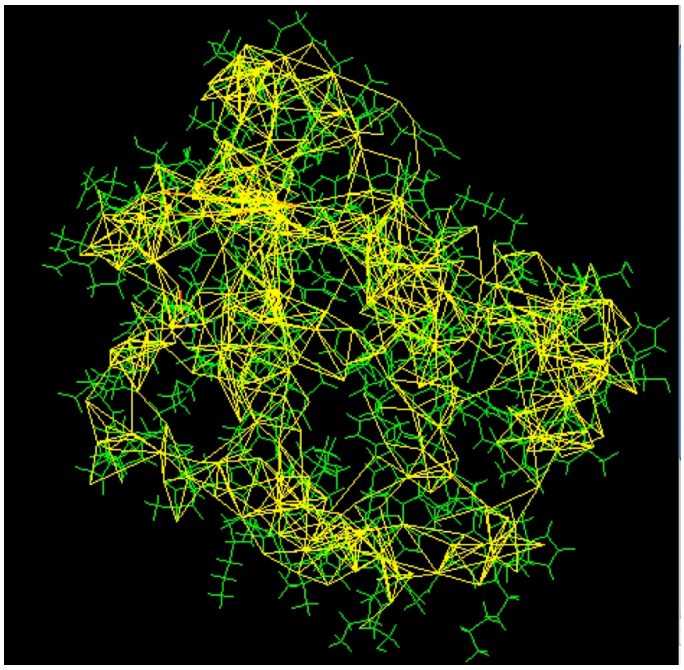

Figure R.10 Structure of a monomeric hemoglobin solved by NMR. Bonds are shown in green and restraints are shown in yellow. (http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/methods.html)

Of course, six cross-peaks do not a structure make. Well-determined NMR structures rely ≥ 15 "restraints" per residue (Figure R.10). A restraint is any datum that defines a structural feature of the protein. NOESY cross-peaks are important in defining spatial relationships between nuclei, but coupling constants can define dihedral angles and chemical shift can be used to define secondary structure. Restraints do just that – they restrain the imagination. In the absence of any restraint, one can take a model of the known polypeptide, an extended polymer with a specified sequence of amino acid residues, and then fold it up into any shape the heart desires. Restraints confine your folding to those shapes that satisfy the data obtained in the NMR experiment.

Actually, it's not imagination that governs folding, it's computer modeling. NMR models are prepared by taking a 3D model of the polypeptide (a virtual stick model) and "heating" it to about

3000 K. Now the stick molecule is a rapidly vibrating/rotating ball of residues sampling a large number of conformations rapidly. The computer then allows the model to "cool" slowly under the influence of the physical parameters that restrain all molecules (bond distances, bond angles and atomic radii) as well as the NMR restraints that are specific to the protein of interest. When the cooling is complete, a single model now exists. This process is typically repeated many times (100 is not uncommon) and then some number, say 20, of the models that most closely resemble each other are selected as an ensemble that represent the possibilities apparent given the restraints that are known. In a good ensemble, the backbone atoms and interior side chains will typically align quite closely, while the solvent exposed side chains will typically be less well defined and the ends of the peptide may also be mobile (Figure R.11). That may either simply reflect a lack of data, or equally likely, reflect the natural conformational flexibility of the protein in solution.

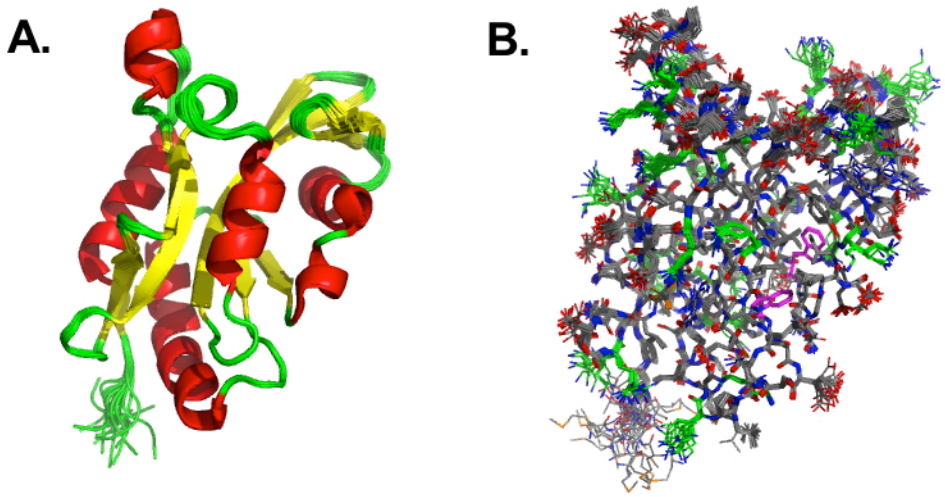

Figure R.11. NMR ensemble of 20 models for chicken cofilin (PDB ID 1TVJ). (A) A cartoon of the backbone, colored by secondary structure. Note the small region of poorly defined structure at the lower left of the figure. This is a nice set of models. (B) A wire model of all non-hydrogen atoms. Note close overlap of the interior Trp residues (in purple) and the less well-defined positions of the surface Lys residues, in green.

How good is the model?

When NMR structural data are reported, authors are typically concerned with making a few important points: (1) They were able to obtain a lot of restraints, (2) their models conform to those restraints and (3) to standard stereochemical libraries of bond lengths and angles and (4) the models of the ensemble closely resemble each other. In this way they show quality of the experiment, reasonableness of their results and, importantly, reproducibility. Consider the data in Table R.2.

Table R.2. Structural Statistics for an ensemble of 20 NMR structures of chicken cofilin (PDB ID 1TVJ).[2]

[2] Data taken from *Molecular Cell* **24**, 511-522 (2006)

(1) **Experimental restraints**

Interproton distances (2533 per 166 residues)	
Intraresidue	465
Sequential	665
Medium range	459
Long range	944
H-bond	92
Dihedral angles	365

(2) **Mean root mean square deviations (rmsd) from experimental restraints**

NOE distances (Å)	0.0321 ± 0.0004 Å
Dihedral angles (°)	0.54 ± 0.02

(3) **Mean rmsd from idealized geometry**

Bonds (Å)	0.00323 ± 0.00003
Angles (°)	0.454 ± 0.003

(4) **Mean rmsd between models of ensemble**

Backbone well-defined region (5-166)	0.25 Å

Note that this is a well-determined structure. The points to make: (1) there are adequate restraints, with (2533+92+365) for 166 residues, a ratio of 18:1. (2) The models are in close agreement with experimental data – note the small deviations in distances and angles expected from the NMR experiment of 0.032 Å and 0.54°. (3) The models have good stereochemistry, and look like real molecules – no funny bond lengths or angles. (4) The models agree well with each other over most of the range with a 0.25 Å rmsd between models.