TMM 3102: Protein Structure, Function and Disease

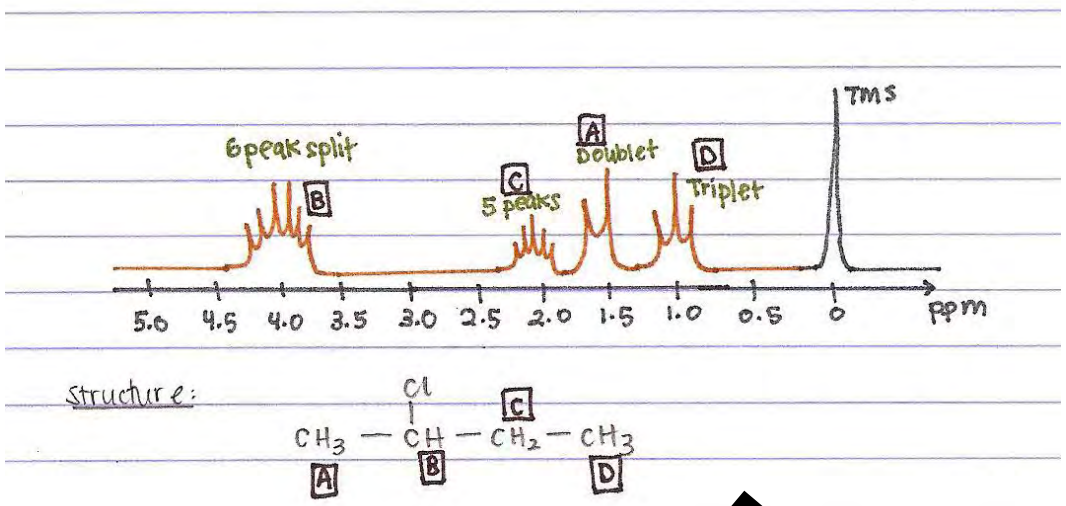
- Structural Biology Methods: Nuclear Magnetic Resonance (NMR) Spectroscopy

(October 12th, 2021)

Jyh-Yeuan (Eric) Lee, Assistant Professor, BMI

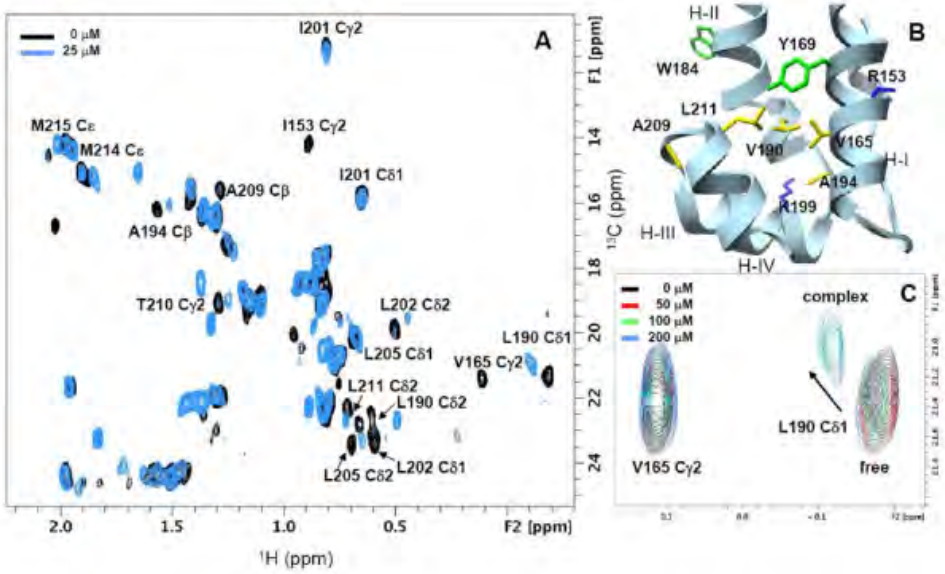
(Partially adopted from former lectures by Dr. John Baenziger)

Nuclear Magnetic Resonance (NMR) Spectroscopy



NMR uses a special kind of electromagnetic waves and monitors physical and chemical changes of the atomic nuclei in the molecules at the atomic scale.

The figures show the typical NMR spectra, 1-D (top) and 2-D (bottom). This course talks about what these spectra mean and how we use them to study the structure-function relationship of proteins



Applications

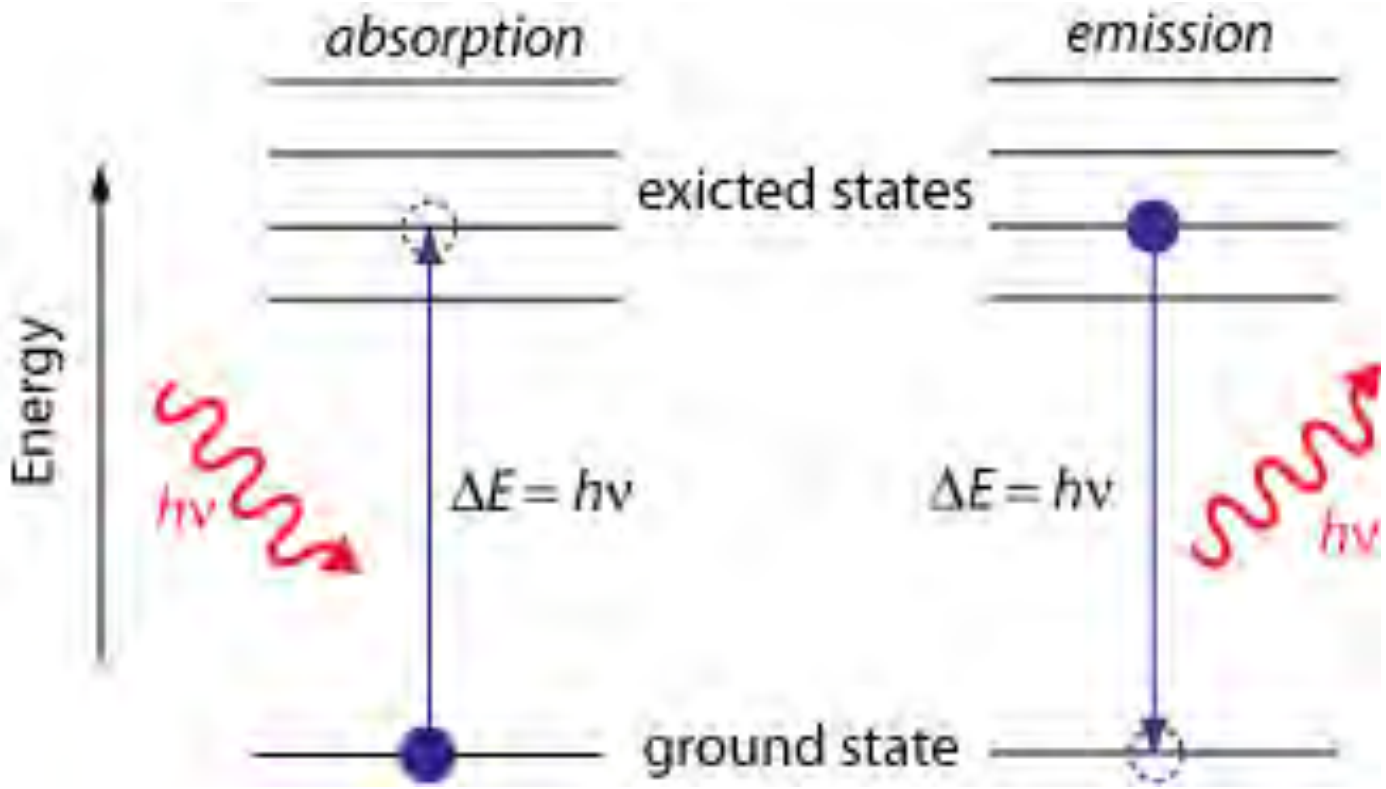
- Chemistry: structures and dynamics of small compounds
- Medicine: magnetic resonance imaging (MRI)
- Metabolite Identification: metabolomics & systems biology
- Segmental and molecular motion: dynamic information of 3-D protein structures
- Determination of macromolecular structures
- Others

Spectroscopy: a quick reminder "excitation *v.s.* emission"

To start out, it's good to take a look at what we know about energy travel in microscopic particles in general.

A particle (solid blue circle) normally is stable at the ground state. When absorbing some energy ($h\nu$), the particle jumps into a higher energy state, excited state.

However, the particle at a higher energy state is not stable. The energy has to be released to allow the particle return back to the ground state. If 100% energy release, it would be $h\nu$ again.



General Terms:

- Nuclear spin (L), magnetic moment (μ), gyromagnetic ratio (γ)

$$\mu = \gamma L$$

The nuclei of the atoms can be seen as the central axis of spin tops.

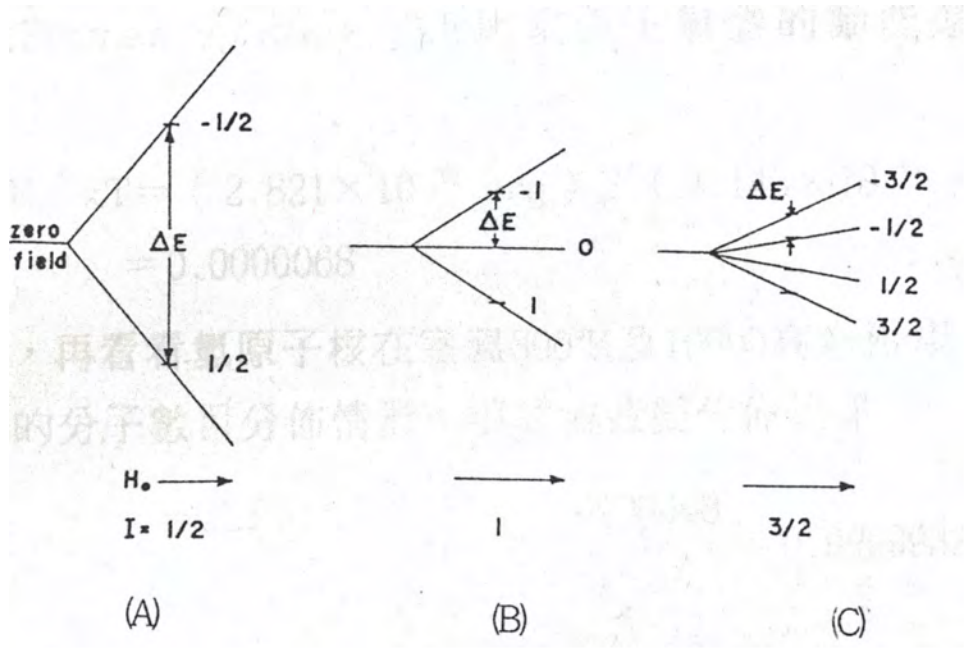
The nuclei of each kind of atoms have unique “nuclear spins”, **L**, which is defined by a constant called “gyromagnetic ratio”, γ.

When nuclei spin, a magnetic moment is generated, μ.

I: nuclear spin quantum number

$$|L| = h[I(I+1)]^{1/2}/2\pi$$

h: Planck constant



General Terms:

- Nuclear spin (L), magnetic moment (μ), gyromagnetic ratio (γ)

$$\mu = \gamma L$$

- Zeeman splitting

$E = -\mu H_o$ (or $-\mu B_o$), H_o (B_o): external magnetic field

$\Delta E = \gamma(h/2\pi)H_o$

When atoms are placed in the middle of a big external magnetic field, **H_0 (or B_0)**, almost equal amount of such imaginary nuclear spins are separated into energy populations of two, three, four, etc, depending on the type of atoms.

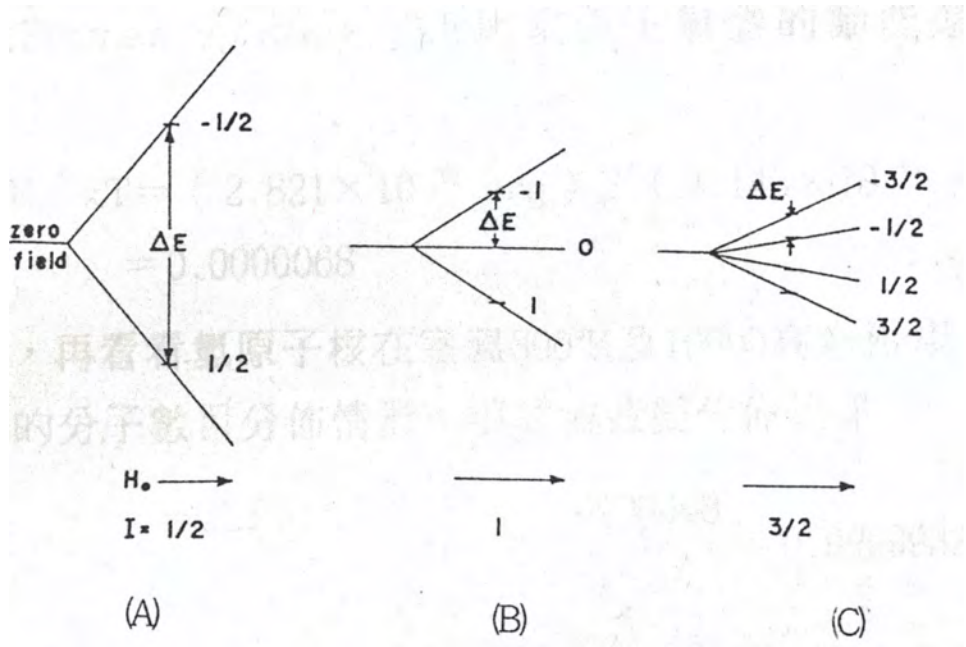
The process of energy splitting the adjacent energy states is called "Zeeman splitting".

Proton (^{1}H), the most common atom in biological systems, has a nuclear spin quantum number of 1. This leads to two states, -1/2 and 1/2, seem in (A).

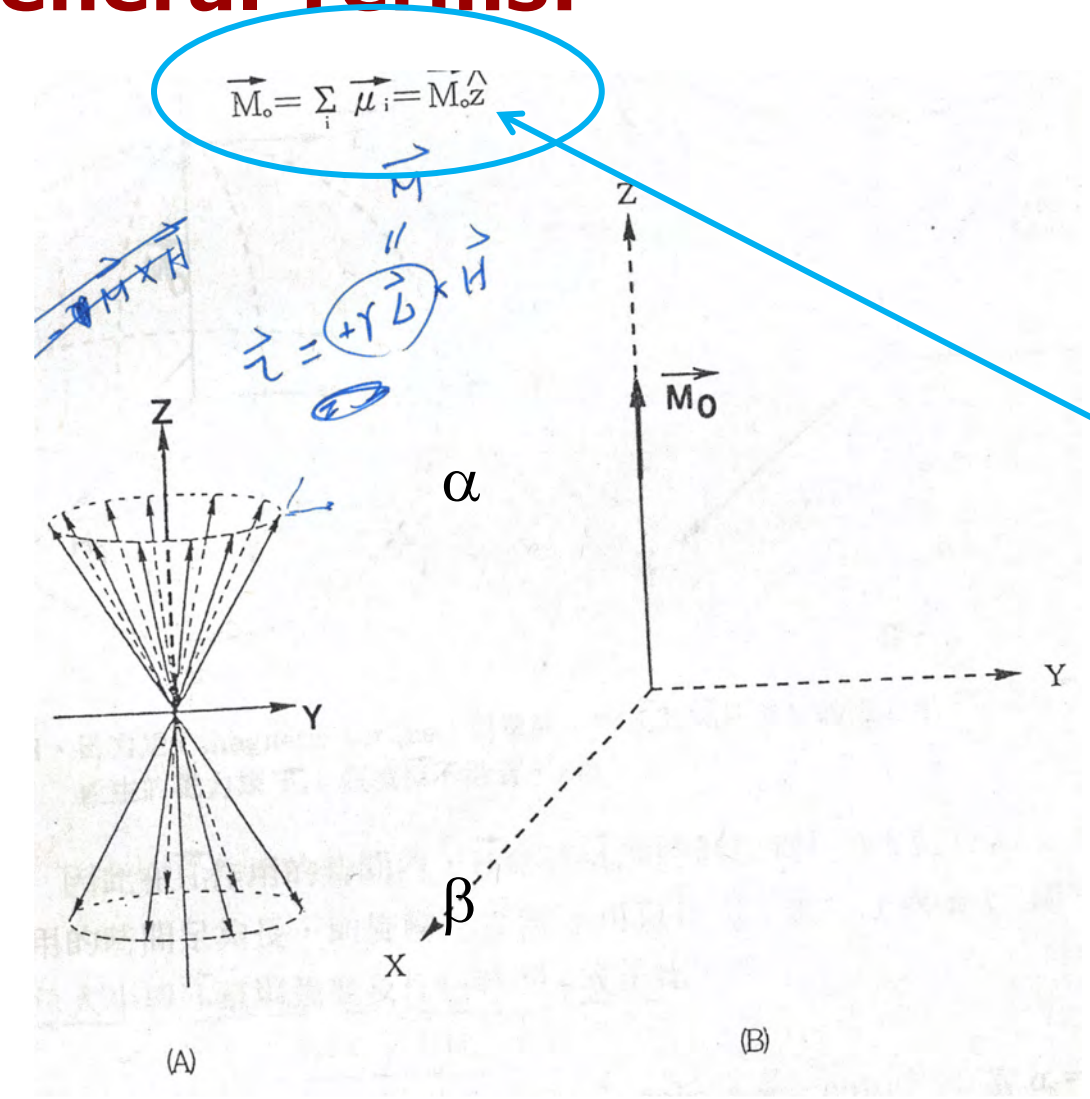
I: nuclear spin quantum number

$|L| = h[I(I+1)]^{1/2}/2\pi$

h: Planck constant



General Terms:



Taking ^1H for example:

Usually there is a little more at α state than β state. Hence, the resonance!

This little population difference between the two splitting energy states allows anything containing the same energy potential to resonance with it.

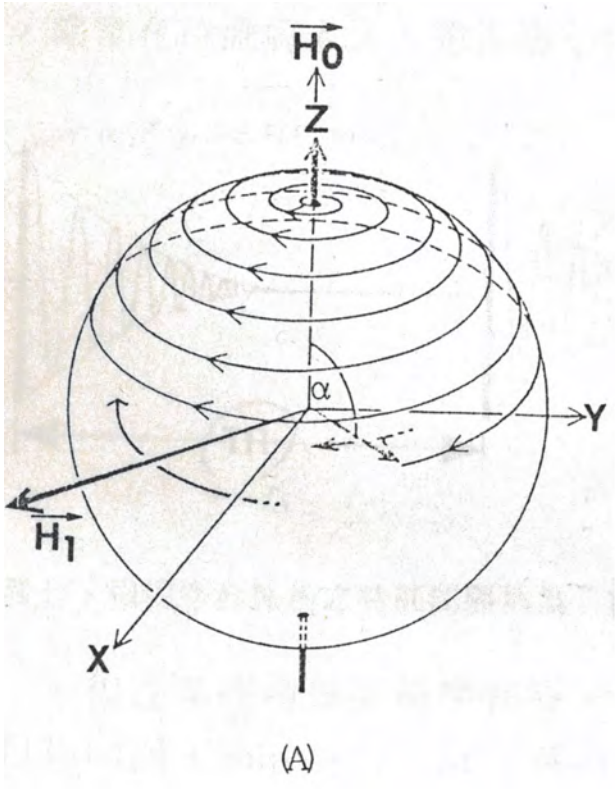
Think of an analogy to using a metal fork to tune the pianos.

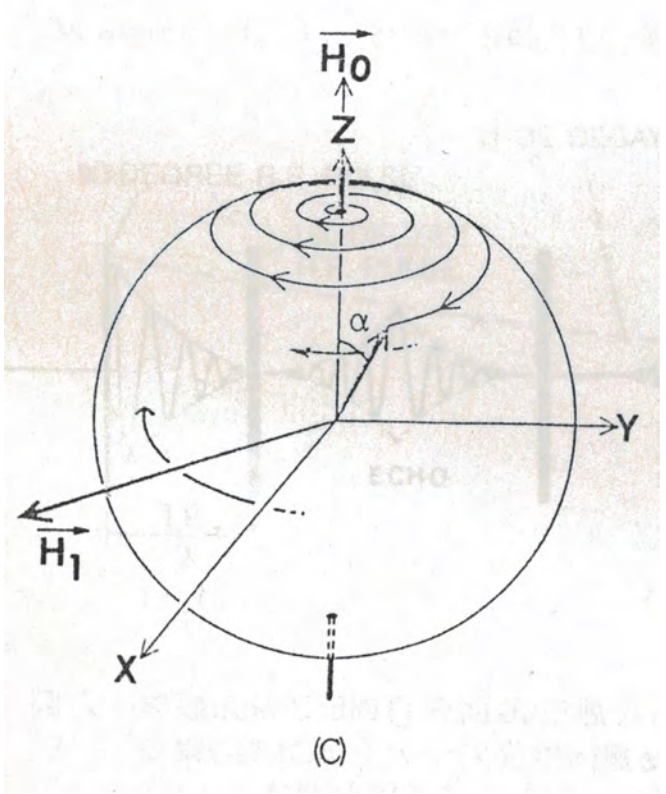
Nuclear relaxation

So under the big external magnetic field, H_0, the little extra energy state population line up with the direction of the magnetic field.

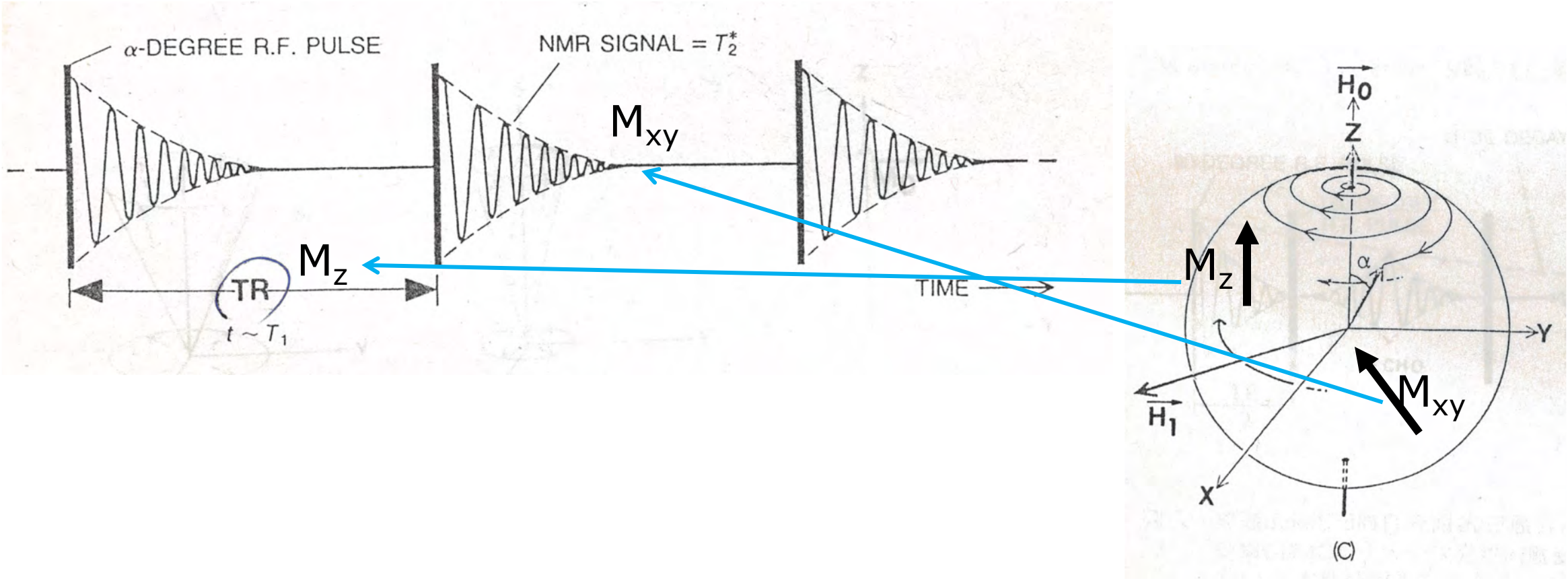
Say this energy is presented as a vector pointing to the +Z (**A**)

When another external force is applied, say H_1, the vector bends to where H_1 goes. Upon release of this extra force, the vector spins back to where the most stable state under H_0 (**the spiral arrows**).





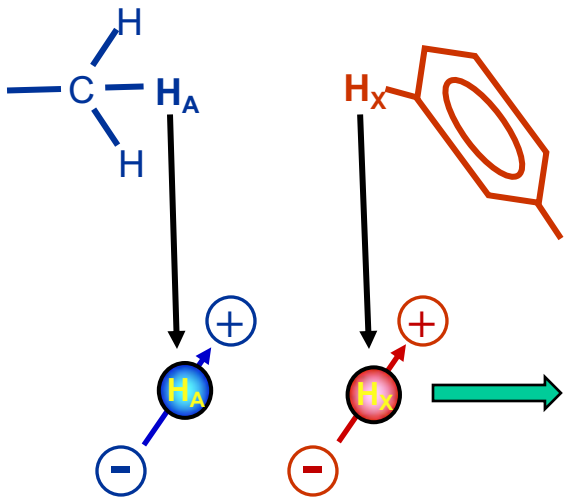
Making of an NMR spectrum



1) The physical basis for NMR spectroscopy?

magnetic field

B_o



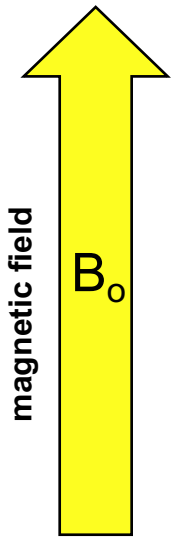
nucleus

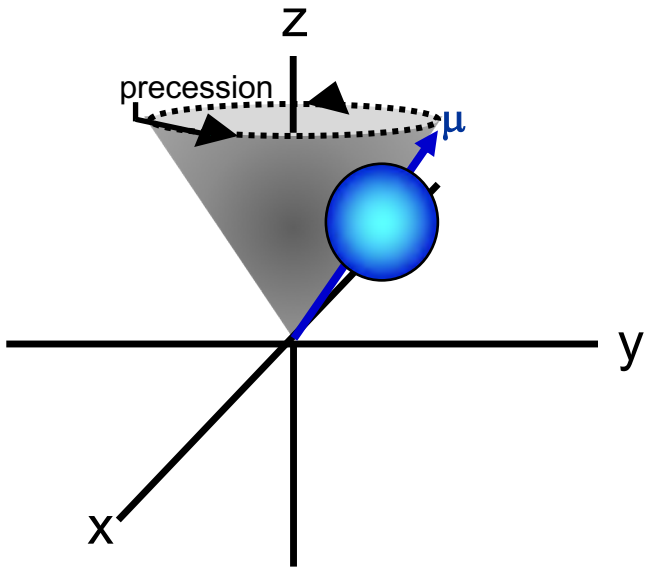
magnetic moment, μ

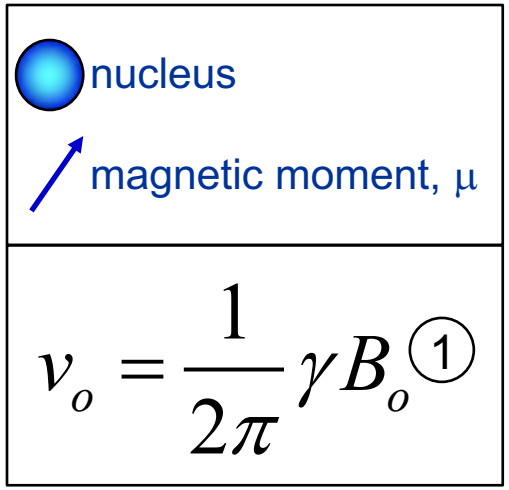
Think of magnetic moments as tiny dipole moments with a positive and a negative pole – the dipoles interact with each other!

The NMR "phenomenon" arises because some nuclei (^{1}H, ^{13}C, ^{15}N, ^{31}P) possess what is called a magnetic moment. This means they behave like tiny bar magnets when placed in a magnetic field (similar to a compass in the earth's magnetic field). NMR can tell us about protein structure because these bar magnets "talk" to each other providing distance information. By understanding how these magnets communicate, we can solve protein structures!

Nuclear Spin

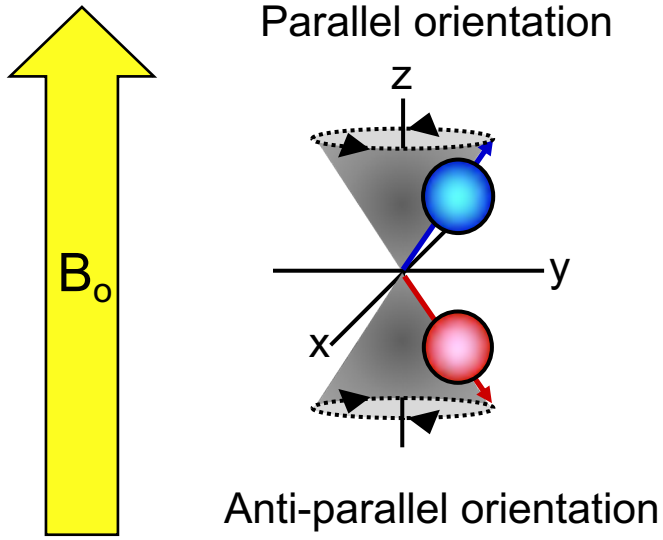






The nuclei of atoms, such as ^{1}H, ^{13}C (1% natural abundance), ^{31}P, and ^{15}N (0.4% natural abundance), possess a magnetic moment, μ, - also called a nuclear spin. A magnetic field (B_o) causes μ to precess (orbit) about B_o (z-axis) at a frequency, ν_o, which is proportional to the strength of B_o and the gyromagnetic ratio, γ, a property of the nucleus (equation 1). The bigger the γ and the larger the B_o, the faster the precession frequency.

Nuclear spins align parallel or anti-parallel to B_o

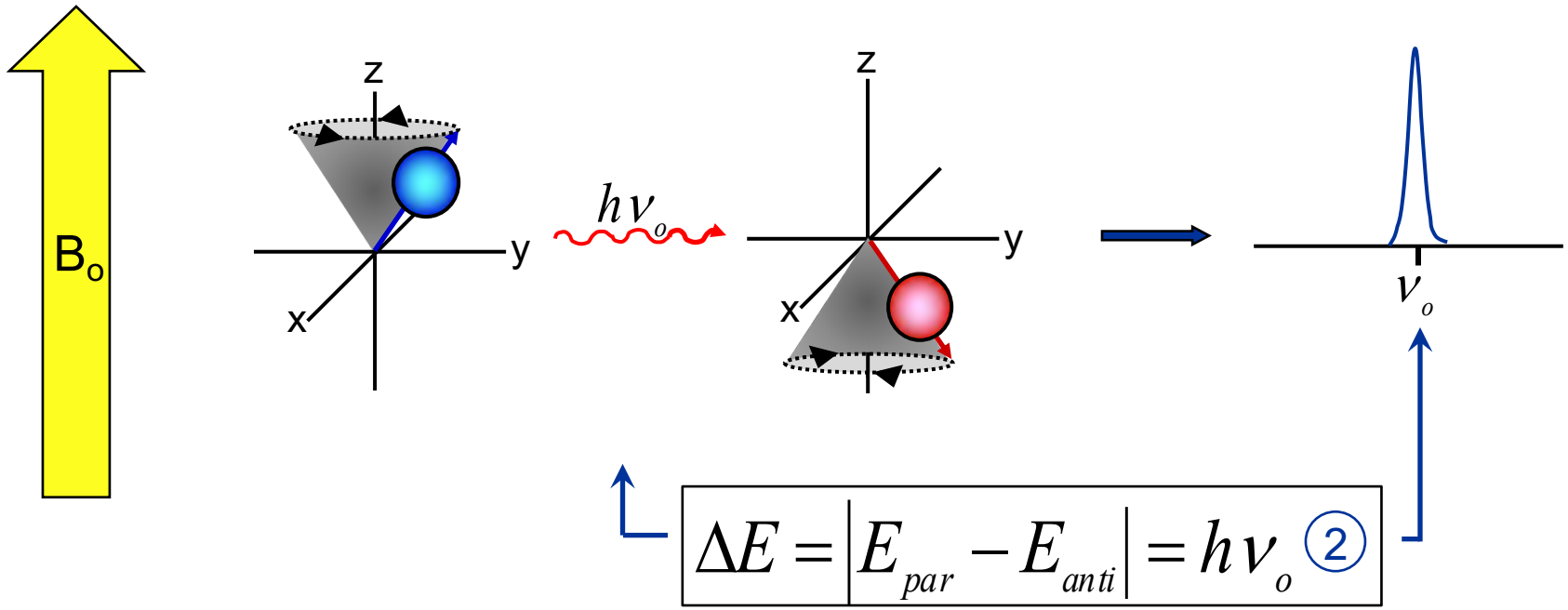


$$\Delta E = \left| E_{par} - E_{anti} \right| = h\nu_o \quad (2)$$

Nuclear spins align either parallel or anti-parallel to B_o. There is a slight energy difference between the two orientations, called ΔE (equation 2). *The parallel orientation is slightly preferred because it has a slightly lower energy than anti-parallel* – for now we need only think about the surplus parallel spins (the other spins cancel each other out).

Note that the energy difference is related to the precession frequency of the protons, which depends on the strength of the magnetic field as defined in equation 1.

An applied radio-wave pulse will flip the nuclear spins



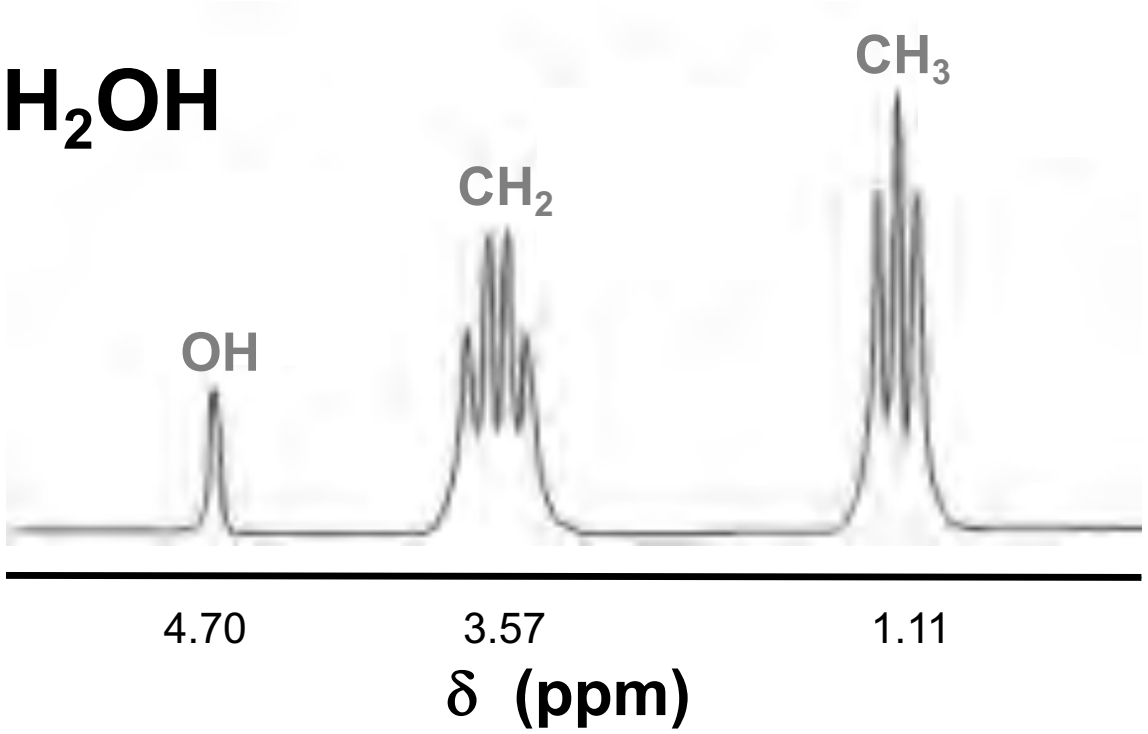
A radio-wave pulse with a ν_o matching the frequency of precession (related to the energy difference between spin states)* will flip the spin from parallel to anti-parallel. NMR detects the absorption of energy at ν_o.

The utility of NMR stems from the sensitivity of ΔE and thus ν_o to chemical environment/structure, but this relationship is complex.

**because radio-wave frequency required to flip spins depends on the B_o, we use a ppm (parts per million) scale*

^{1}H NMR spectrum of ethanol

CH_3CH_2OH

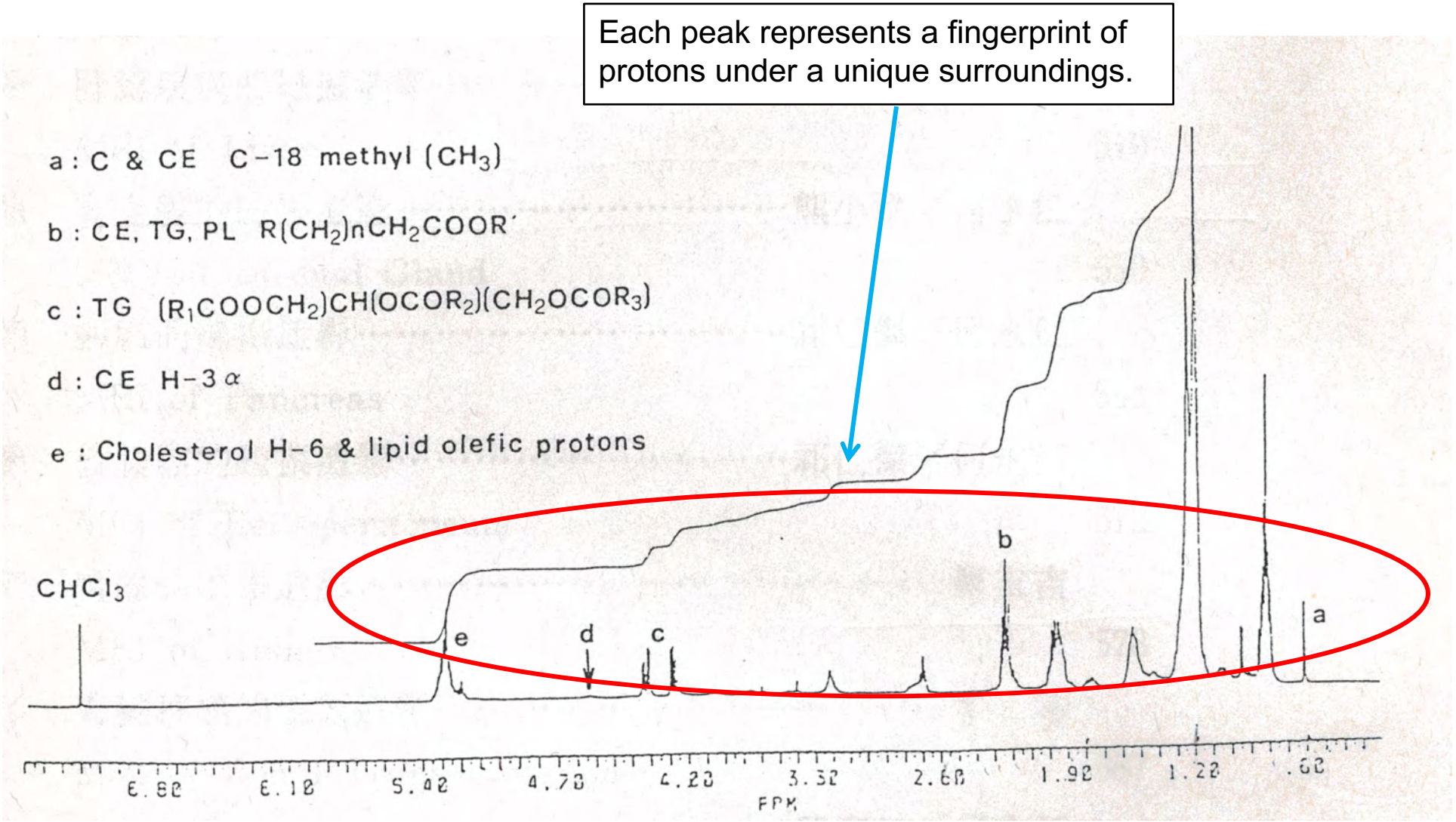


A proton NMR spectrum exhibits a different peak for each set of protons in the molecule – in this case the methylene, the methyl and the hydroxyl protons.

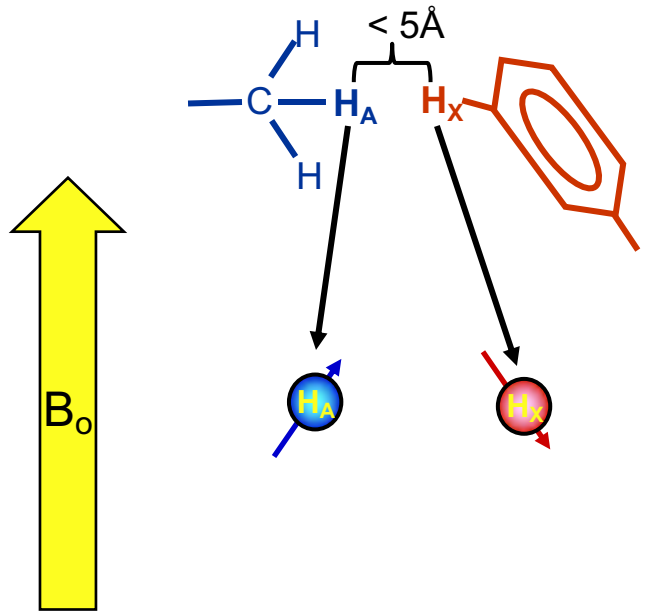
Each signal occurs at a different position (ppm), which depends on the chemical environment surrounding the proton. Each signal has fine structure depending on the # of protons attached to adjacent atoms – called *spin-spin coupling*.

This means that every amino acid in a protein will have a distinct NMR fingerprint – so we can assign each peak in an NMR spectrum to a specific proton in specific amino acid.

^{1}H NMR spectra of the plasma extracts from a patient with hypertriglycerolemia

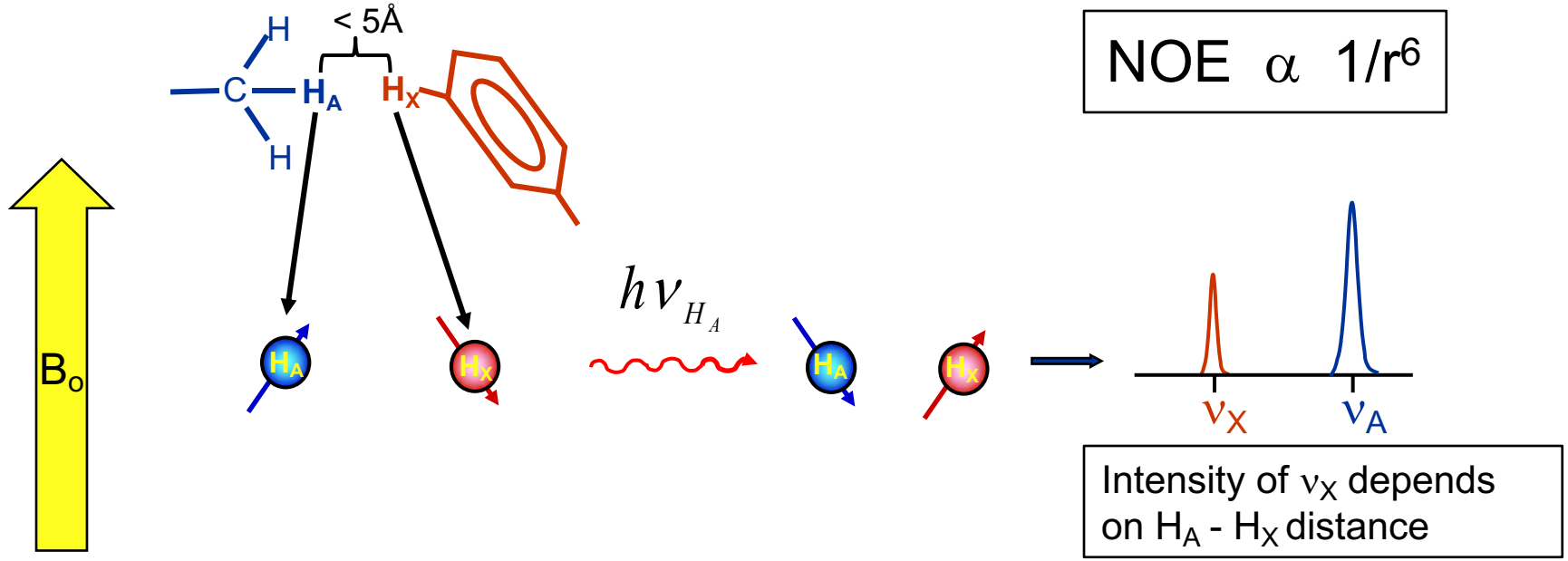


2) Nuclear Overhauser effect (NOE) and distance constraints



The NOE is the key parameter that allows one to solve a protein structure. It occurs when protons talk to each other through space – not through covalent bonds. An NOE will occur between two protons, such as H_A and H_X that are <5Å apart. The NOE is an effect that occurs because the spins interact with each other via "dipolar coupling" – essentially dipole-dipole interactions.

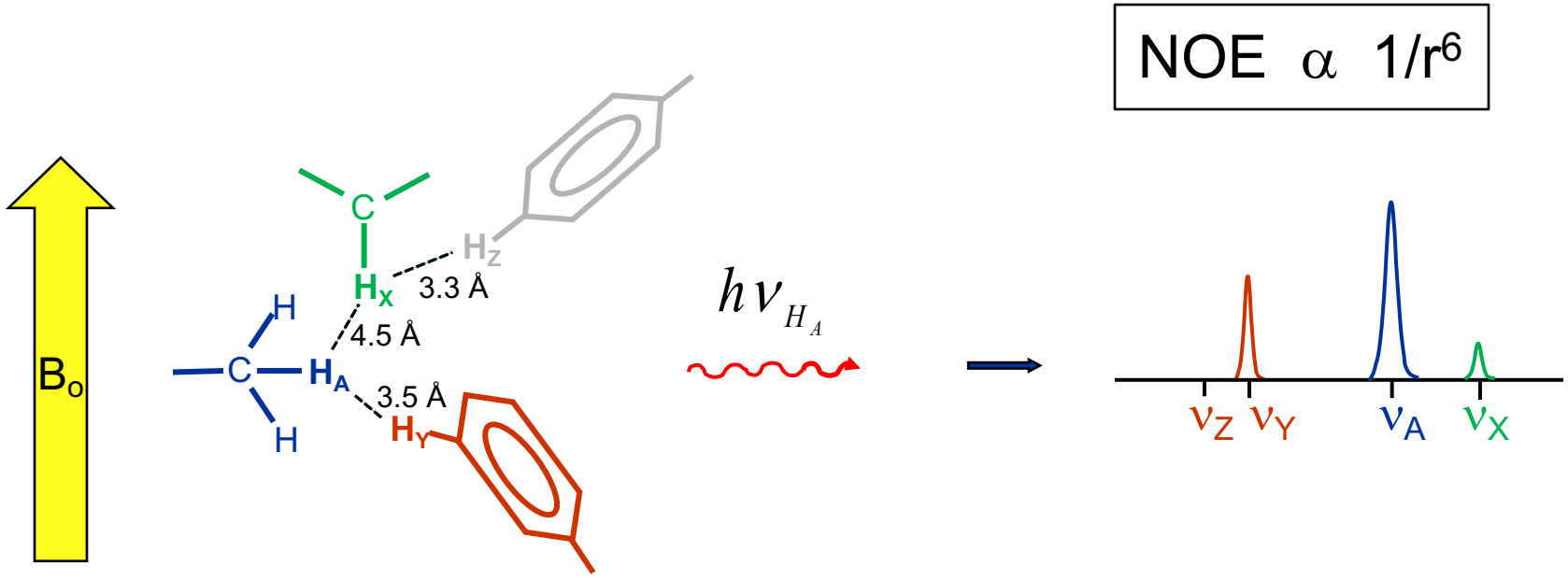
2) Nuclear Overhauser effect (NOE) and distance constraints



The NOE works as follows; assume that H_A is parallel and H_X anti-parallel. If we apply a radiofrequency pulse that flips H_A from parallel to anti-parallel, the dipolar coupling between H_A and H_X may cause H_X to flip as well, so an absorption of energy is observed for H_X, as well as H_A.

Note that the probability of proton A causing proton X to flip depends on how tightly they are coupled - i.e. how close the protons are to each other. *The intensity of the ν_X peak is thus related to the distance between the spins!*

2) Nuclear Overhauser effect (NOE) and distance constraints

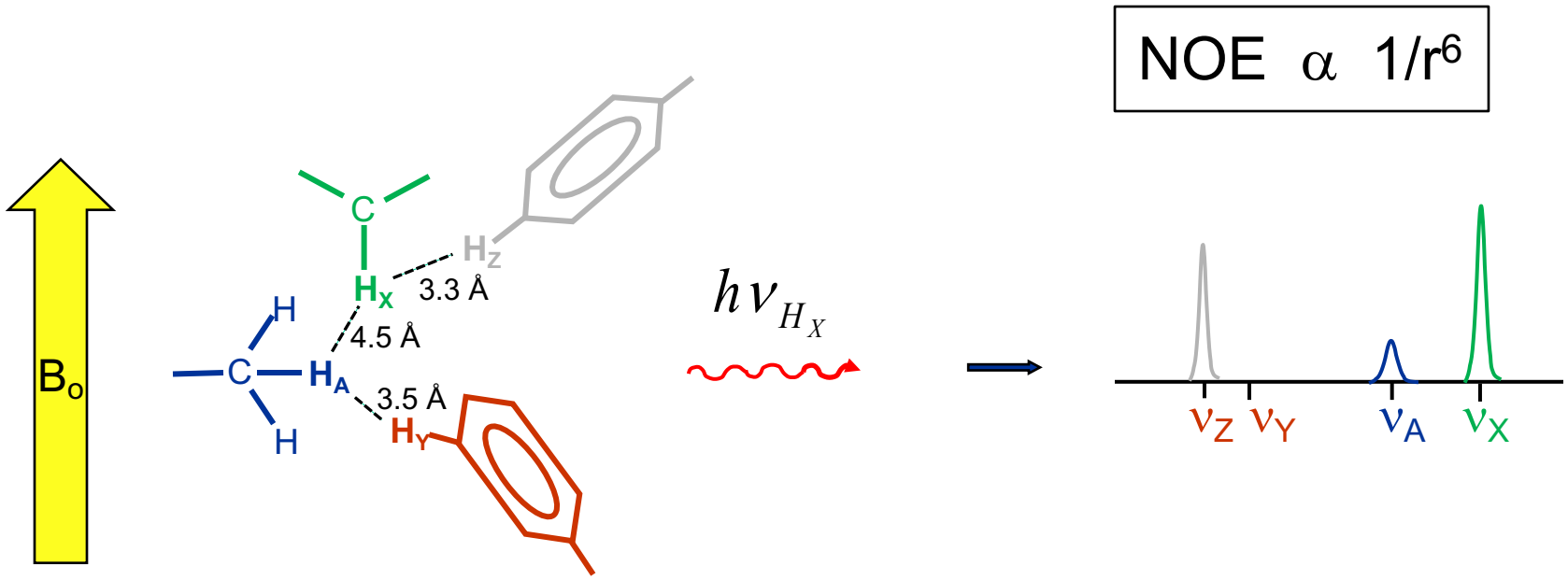


H_A is close to H_x and H_Y, but not H_Z. If we apply a radiofrequency pulse that flips H_A from parallel to anti-parallel, dipolar coupling will cause some H_X and H_Y protons to flip, so we will see peaks for both H_X and H_Y – but not for H_Z.

Note that the intensity of the ν_Y peak is greater than the intensity of the ν_X peak because H_X is closer to H_A than H_Y!

By measuring the height of the peaks – the NOE – we determine the distance between H_A and both H_X and H_Y to be 4.5 and 3.5 Å, respectively.

2) Nuclear Overhauser effect (NOE) and distance constraints



Next, we determine the protons that are in close proximity to H_X, and so on. In this case, H_X is within 5Å of H_A and H_Z, but is too far from H_Y to yield a measurable NOE.

We apply a radiofrequency pulse that flips H_X from parallel to anti-parallel, dipolar coupling will cause some H_A and H_Z protons to flip, so we will see peaks for both.

By measuring the height of the peaks – the NOE – we determine the distance between H_X and both H_A and H_Z to be 4.5 and 3.3 Å, respectively.

2) Nuclear Overhauser effect (NOE) and distance constraints

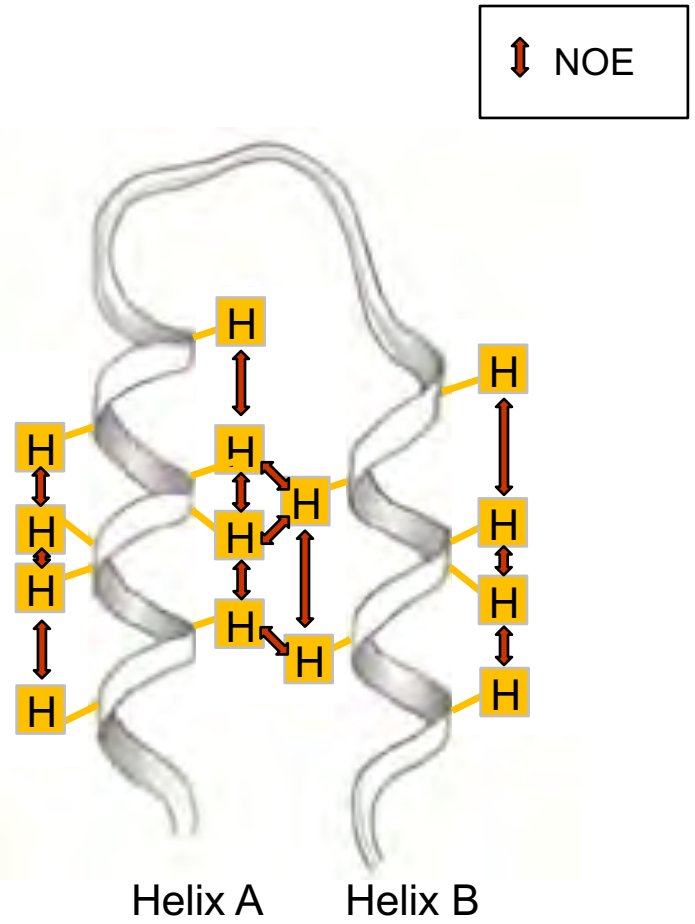
The NOE experiment tells us which protons are close in space.

Consider protons on adjacent faces of two α-helices:

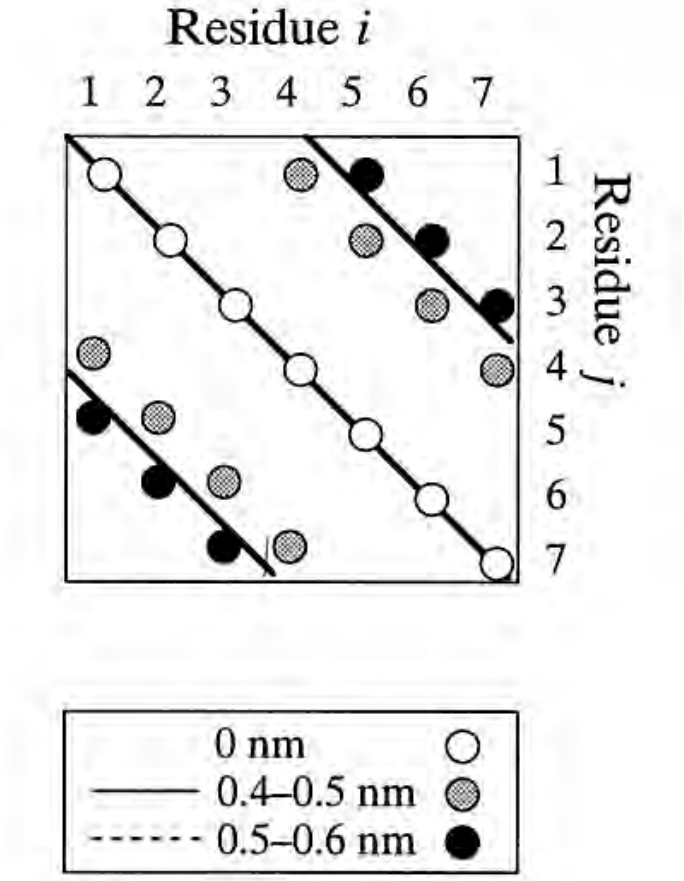
The rise of an α-helix is 1.5 Å per residue, so protons on residues 3/4 apart within an α-helix should be < 5 Å apart and should give an NOE.

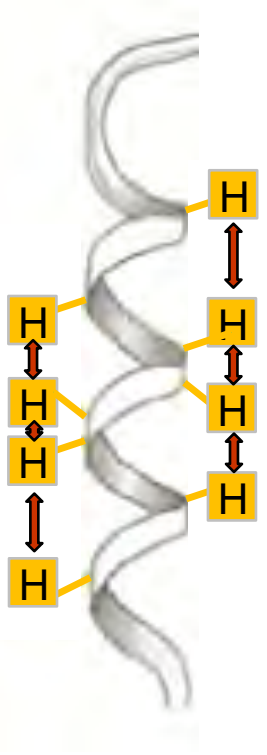
Residues on two facing but adjacent α-helices should be very close due to "ridges into grooves" packing.

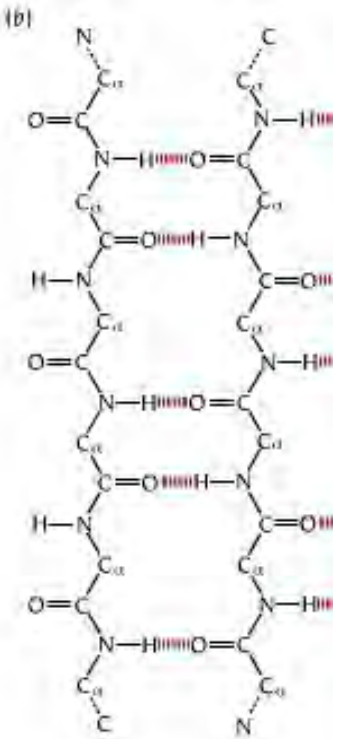
NOEs can tell us about secondary structure and how different secondary structures pack together in 3D – i.e. protein structure!



3) Solving and assessing the accuracy of a 3D structure

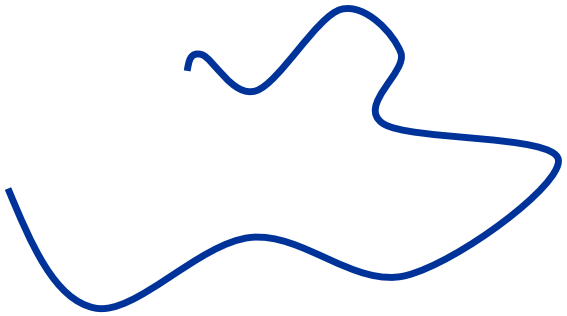






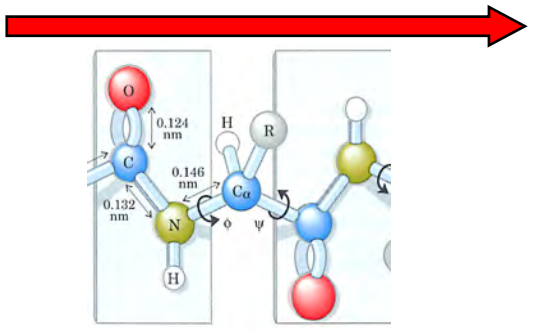
The NOE data is placed in a grid listing the distance between one set of protons in the protein and every other set of protons in the protein. For example, in the *simplified* distance grid above, distances between residues (not protons) are shown. Residue 1 is close to residues 4 & 5, but not 2, 3, 6, and 7, whereas residue 3 is close to residues 6 and 7. *What is the structure of this short peptide?*

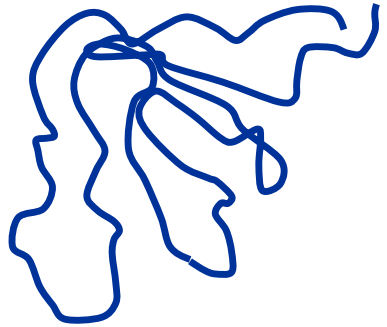
Distance constraints are used to solve the 3D protein structure



Random starting structure

Iterative adjustment of ψ and ϕ to get a better fit b/w structure and NMR data



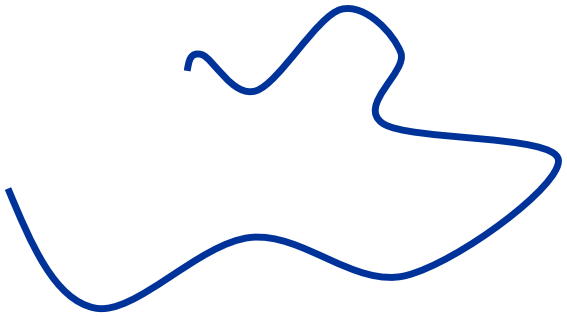


One best-fit structure

All the measured distance *constraints* (and other data) are placed into a computer. The computer randomly sets the ψ and ϕ angle for each residue in the protein to generate a random starting structure (*left*).

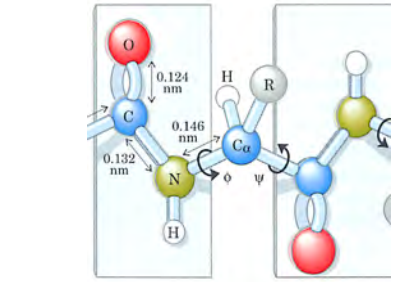
The ψ and ϕ angles for each residue are then iteratively changed. At each step or iteration, a test is performed to see if there is an improved match between the structure and the constraints. Eventually, the computer reaches a “best fit”, which corresponds to one possible structure

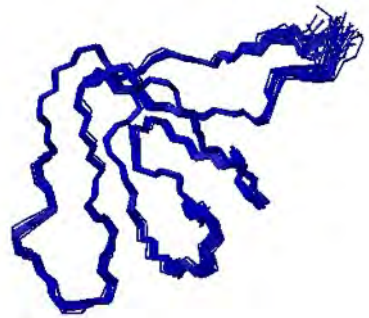
Many 3D structures are generated based on different starting conformations



Random starting structure

Iterative adjustment of ψ and ϕ to get a better fit b/w structure and NMR data



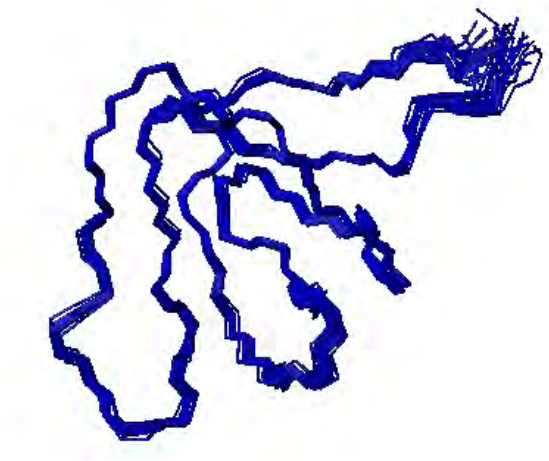


20 Best-fit structures

Final solution = average of 20 best-fits

The best-fit structure depends on the randomly chosen starting structure. For this reason, 100 different randomly chosen starting structures are selected, each leading to a best-fit structure. The 20 structures that best fit the NMR data are typically presented (*top right*), along with the average of these 20 structures – the final solution (*bottom right*). Energy minimization is part of this process.

Precision of NMR structures

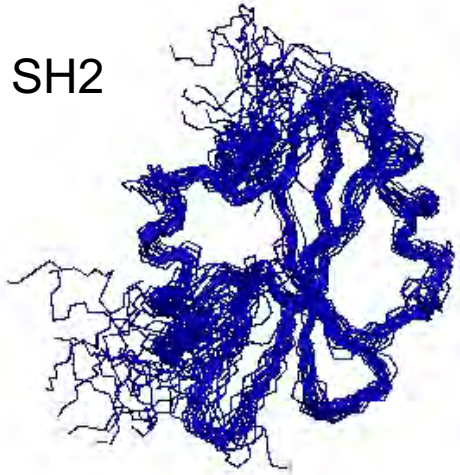


$$rmsd = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(x_i - \bar{x}\right)^2}$$

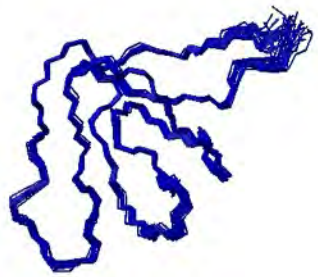
The precision of the determined structure is evaluated by comparing the position of each backbone atom in every best-fit structure to the position of the same atom in the final structure (i.e. the average of the 20 best-fits). The root mean squared deviation (RMSD) is analogous to standard deviation. It is a measure of the average deviation of each individual amino acid position (x_i) relative its average position in the final solution (x). The lower the RMSD, the more precise the determined structure.

Precision of SH2 versus SH3 domain structures

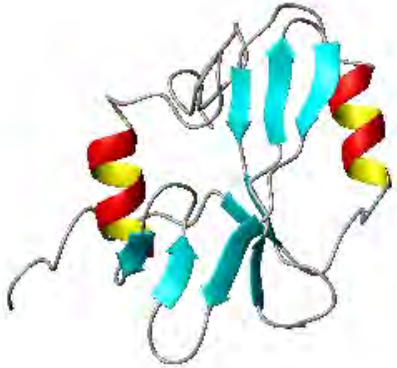
SH2



SH3



Backbone rmsd: 1.17 Å 0.43Å



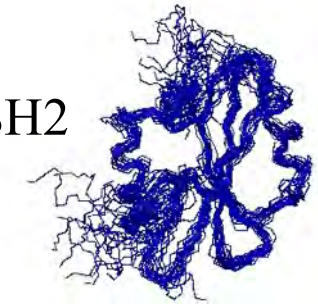
Pascal et. al (1994) Cell

Liu et al (2003) Mol Cell

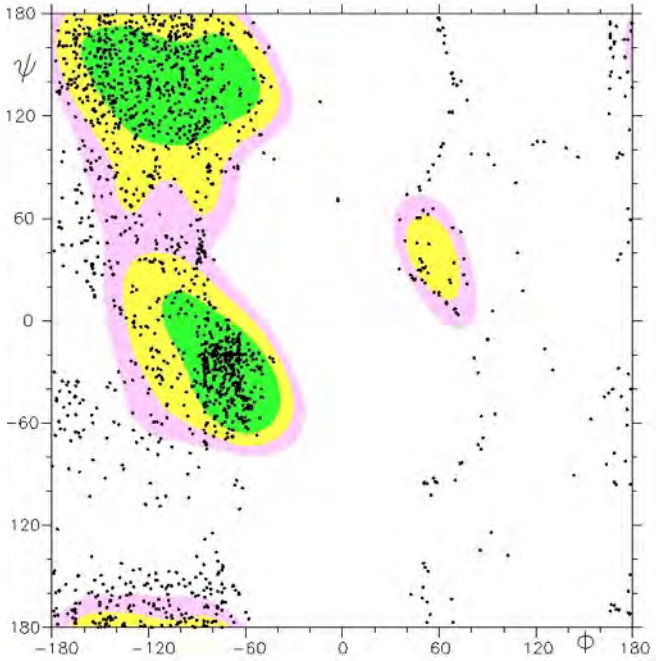
Two NMR structures of peptide binding domains have different backbone RMSD values. The SH3 structure is more precise – there is little variability amongst the 20 best-fit solutions – the RMSD is lower. Note that variability typically increases at the N- and C-terminus.

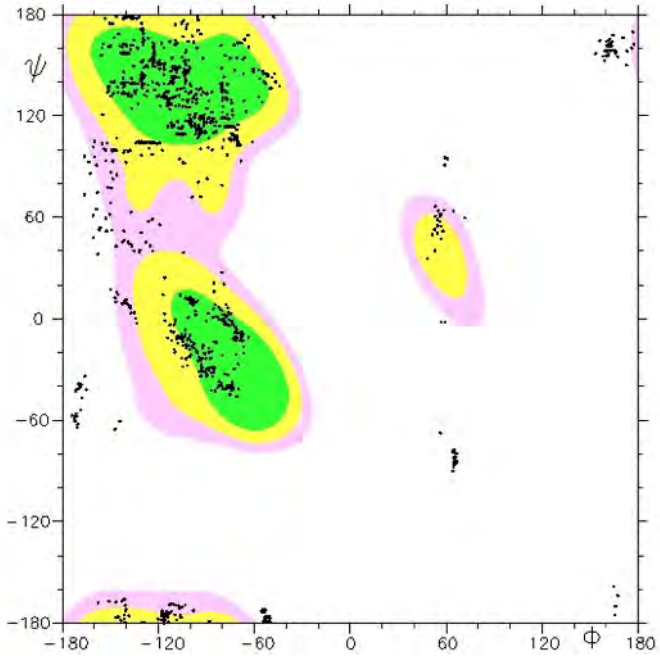
Ramachandran plots for SH2 vs SH3 domains

SH2



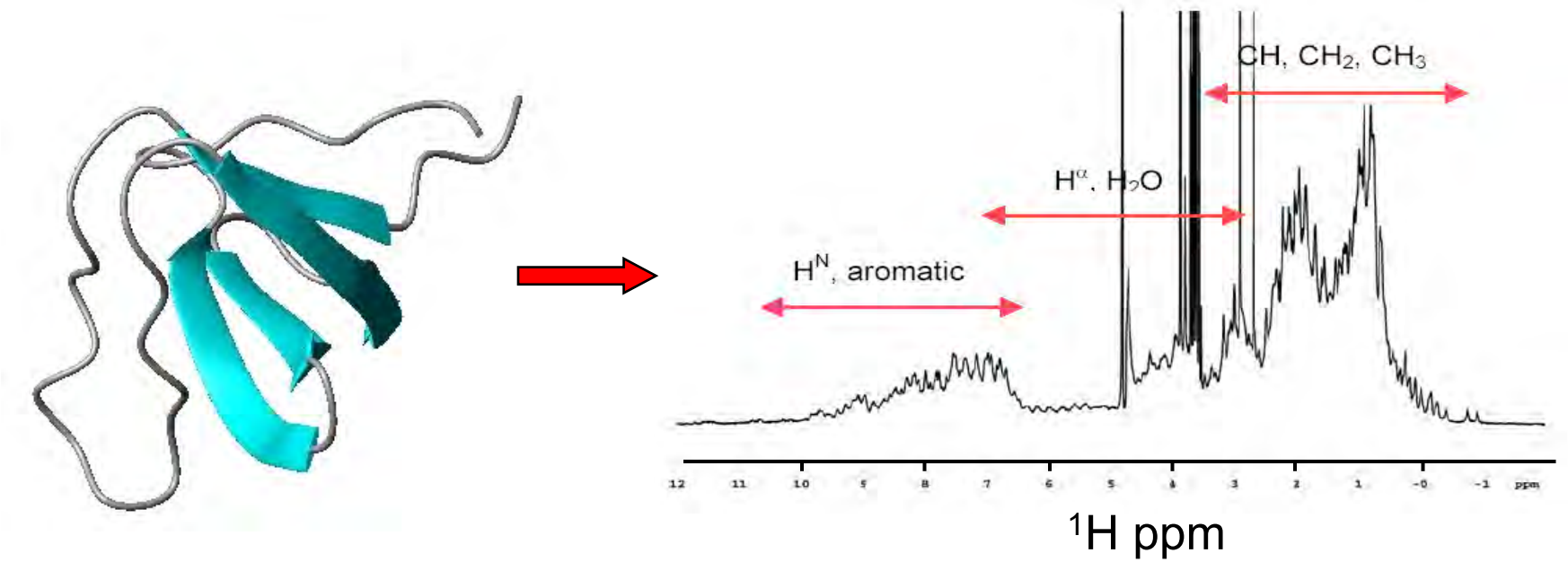
SH3





Many dihedrals in regions that are not allowed	Only a small number in non-allowed regions

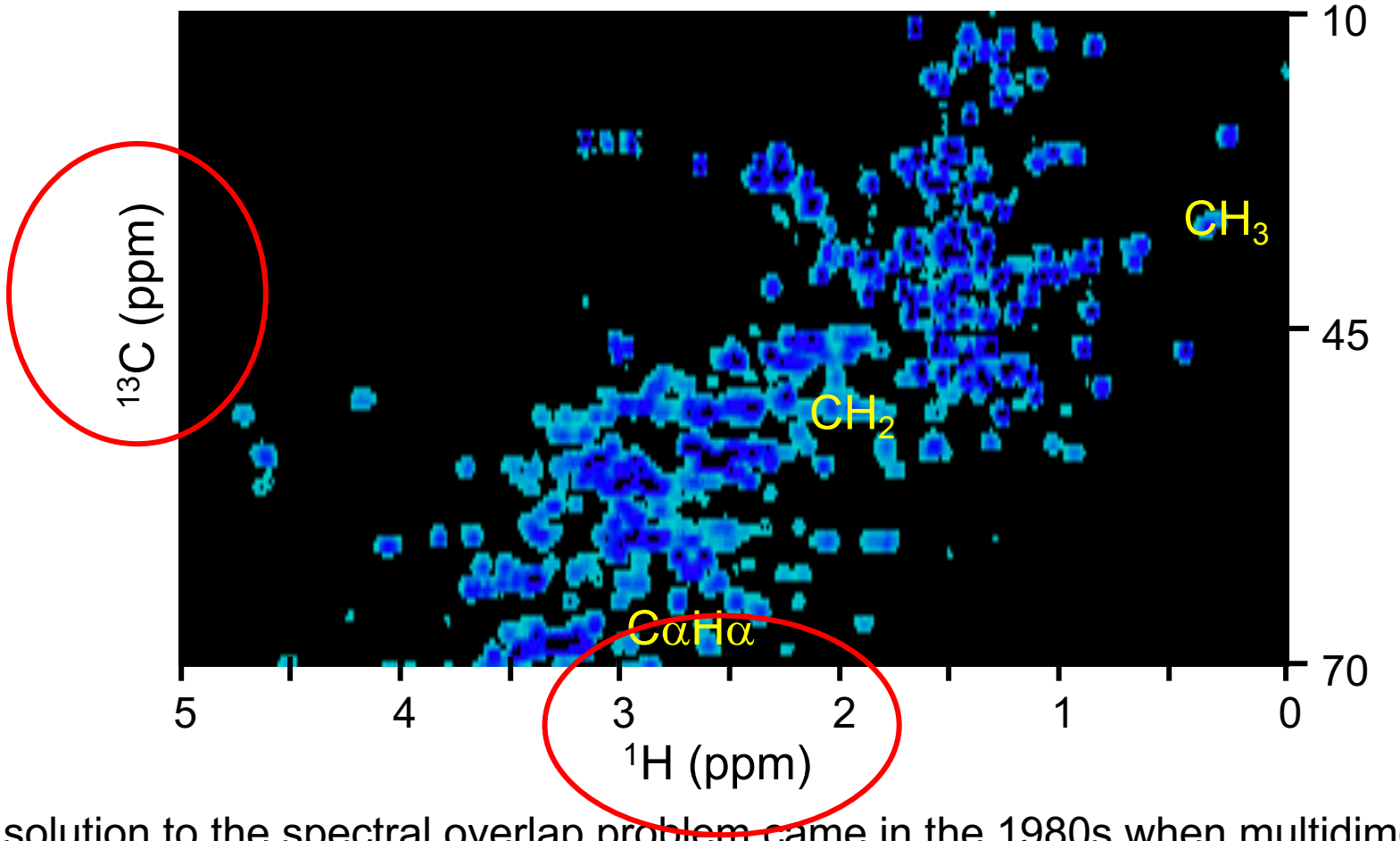
4) Multidimensional NMR & ^{1}H-^{15}N HSQC spectra



Even a small protein of 100 residues (~11 kDa) will have hundreds of protons. Because there are so many protons, protein NMR spectra are extremely crowded. It is difficult to obtain information from such spectra because there is so much peak overlap that we cannot assign individual peaks to individual protons.

To study proteins, we need a combination of multi-nuclear (where we use nuclei other than protons) and multi-dimensional NMR.

Multidimensional NMR

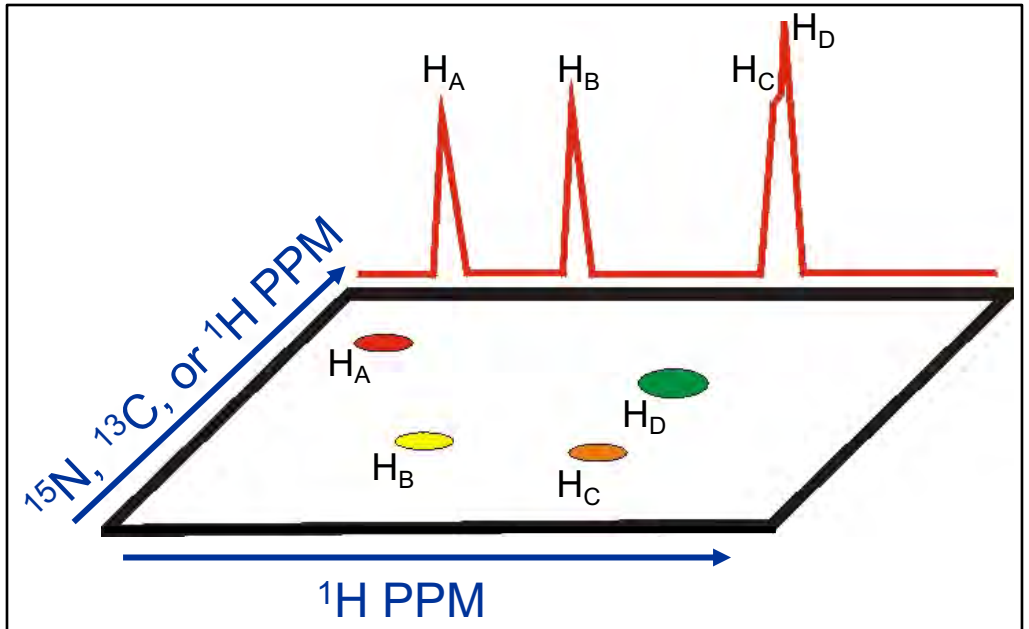


A solution to the spectral overlap problem came in the 1980s when multidimensional NMR was used to spread the NMR signals in multiple dimensions as shown above. Above is a 2 dimensional (2-D) NMR spectrum, but one can also generate 3-D and 4-D spectra. There are many types of multi-dimensional NMR spectra

HSQC spectra

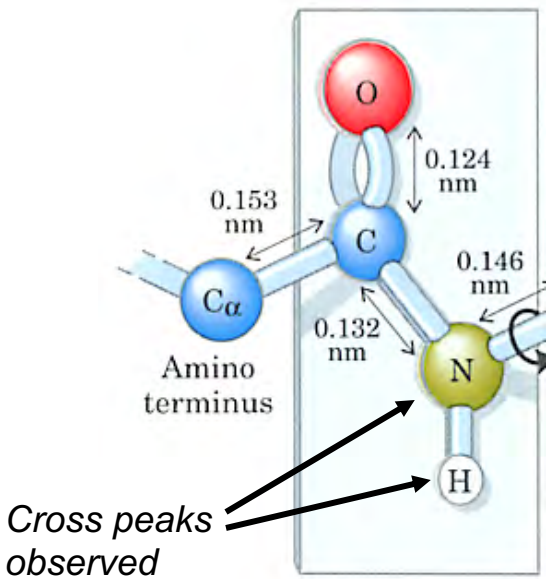
One of the most published types of 2D NMR experiment is the *Heteronuclear Single Quantum Coherence (HSQC)* plots the ^{1}H ppm on one axis and either the ^{15}N, ^{13}C, or ^{1}H ppm on the other.

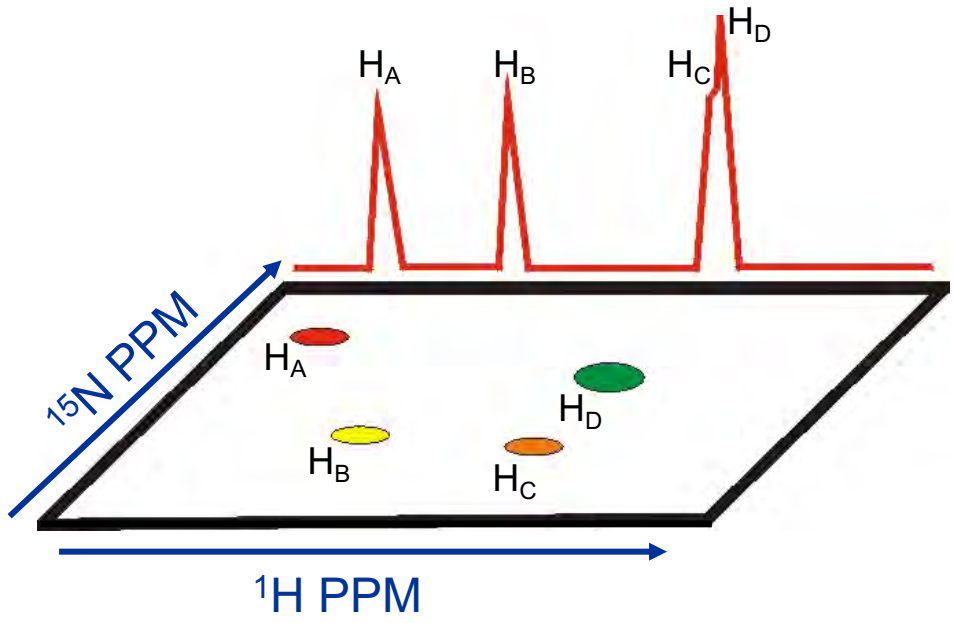
The experiment is set up so that peaks (called “cross peak” in a 2D experiment) are observed between hetero-nuclei that are *directly covalently bonded*.



2D HSQC spectrum showing the cross peaks between covalently bonded nuclei.

^{1}H-^{15}N HSQC spectrum



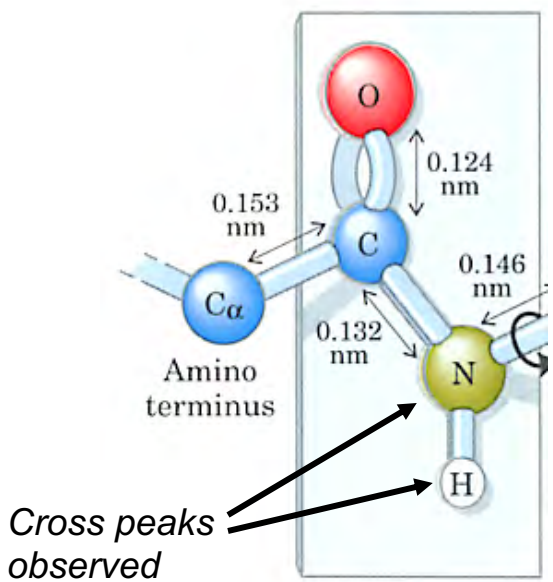


2D NMR spectrum showing the cross peaks as circles.

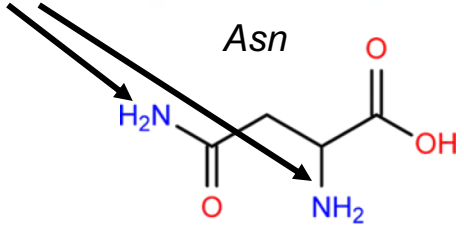
In a ^{1}H-^{15}N HSQC spectrum, you record both a ^{1}H-NMR spectrum and a ^{15}N-NMR spectrum (protein must be labelled with ^{15}N), but do so in a way that you explore the links between ^{1}H and ^{15}N nuclei that are covalently attached.

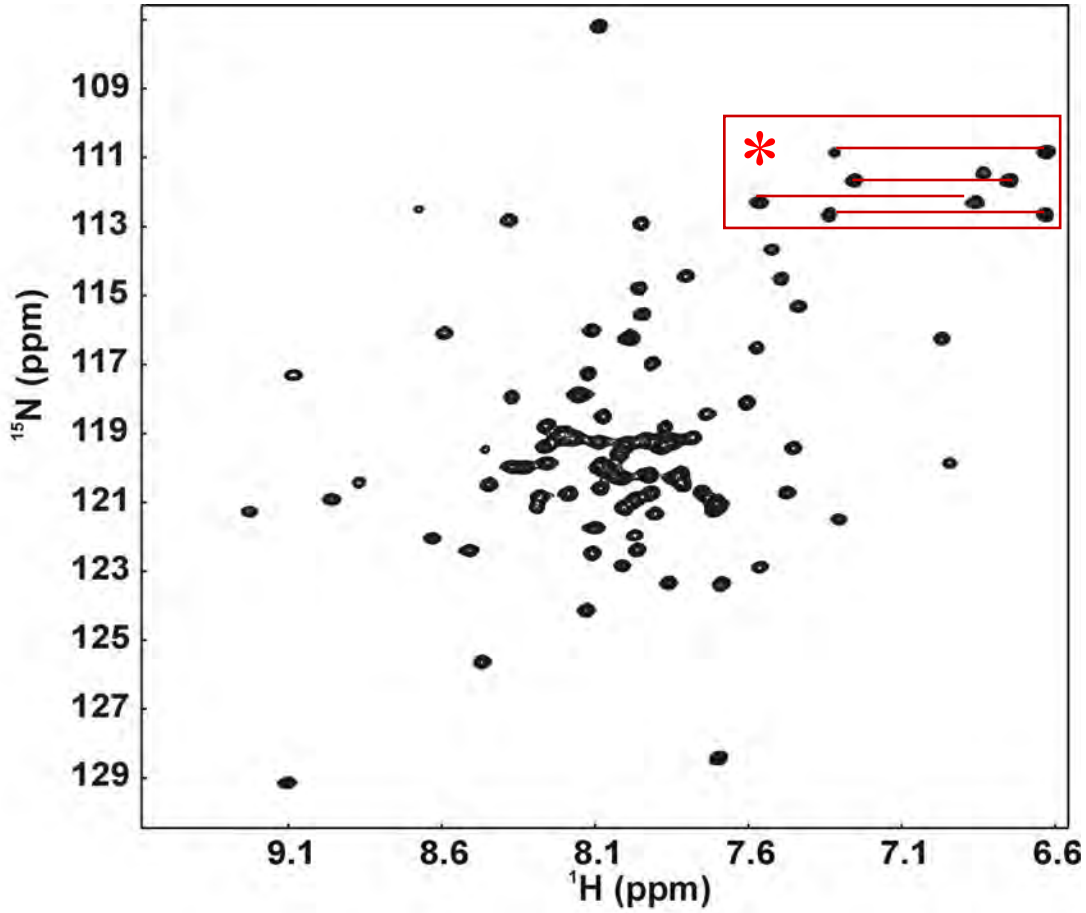
The result is plotted so that you have an axis for the ^{1}H-NMR spectrum and one for the ^{15}N-NMR spectrum. You get a "cross peak" for every hetero-nuclei (^{1}H-^{15}N) pair that are covalently bonded – *but the position of each depends on local environment!*

^{1}H-^{15}N HSQC spectrum – a protein "fingerprint"



Cross peaks observed b/w ^{15}N and ^{1}H.



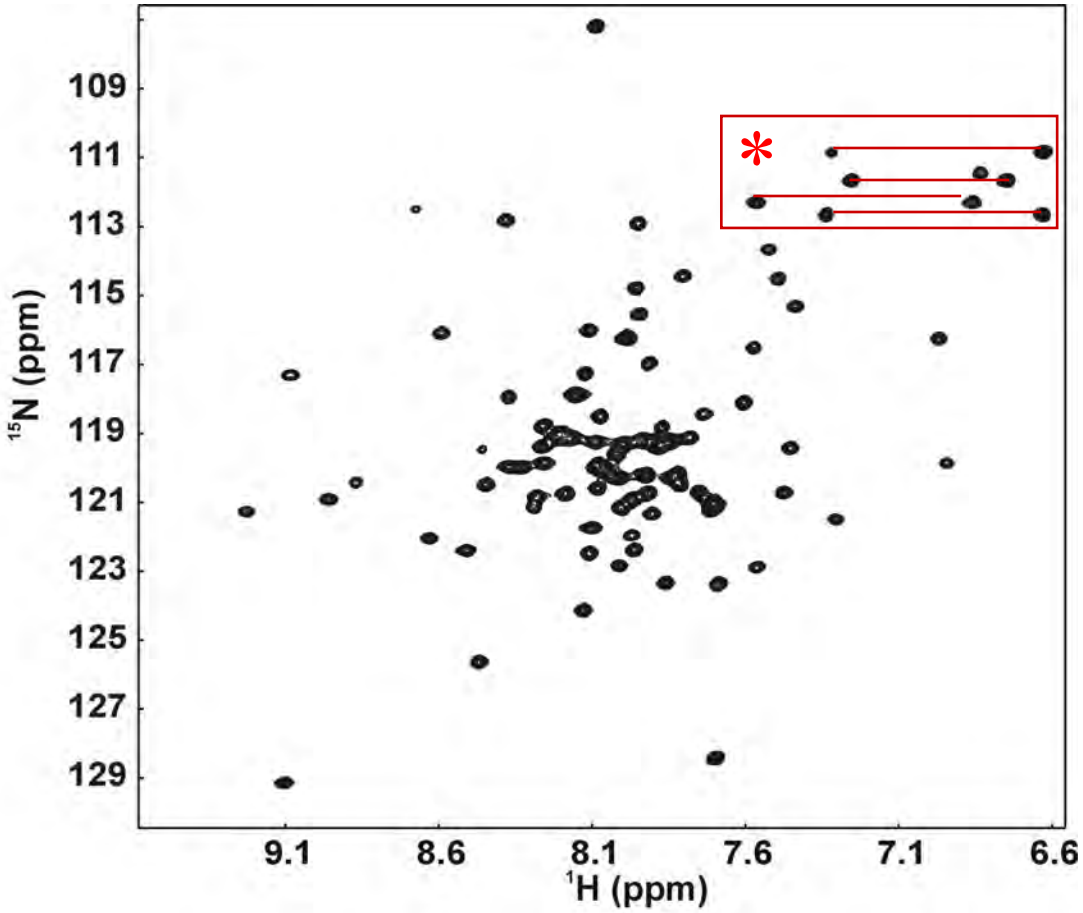


An ^{15}N-^{1}H HSQC shows cross peaks for every directly bonded ^{15}N and ^{1}H, such as the ^{15}N-^{1}H of the polypeptide backbone. Each residue gives rise to one ^{15}N-^{1}H cross peak, but two are observed for Gln and Asn (*). *Why?*

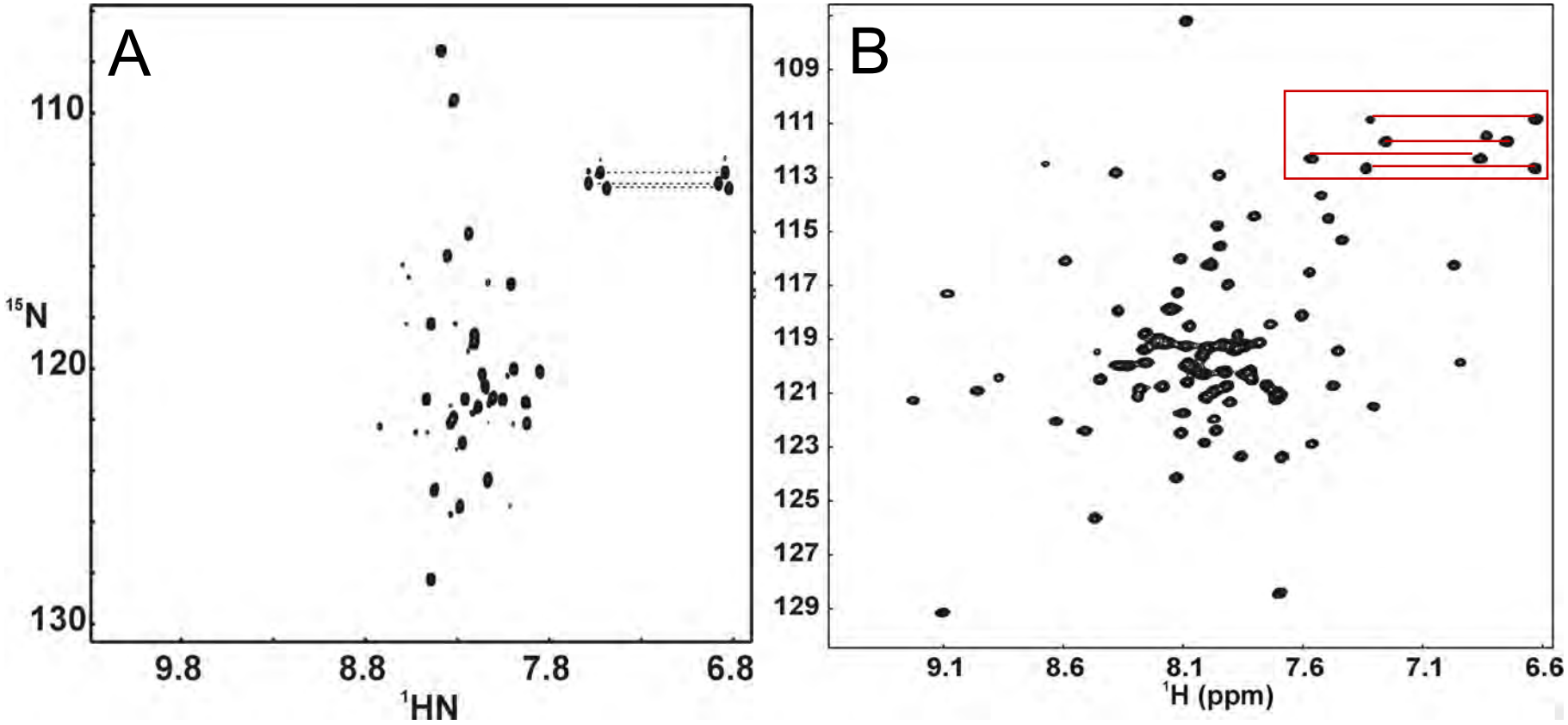
^{1}H-^{15}N HSQC spectrum – one cross peak for each residue

The frequency of each cross peak is sensitive to chemical environment (i.e. the structure of the protein).

An ^{15}N-^{1}H HSQC spectrum is a sensitive probe of overall protein structure and can be used for many things, including testing whether a sample is suitable for detailed NMR studies...



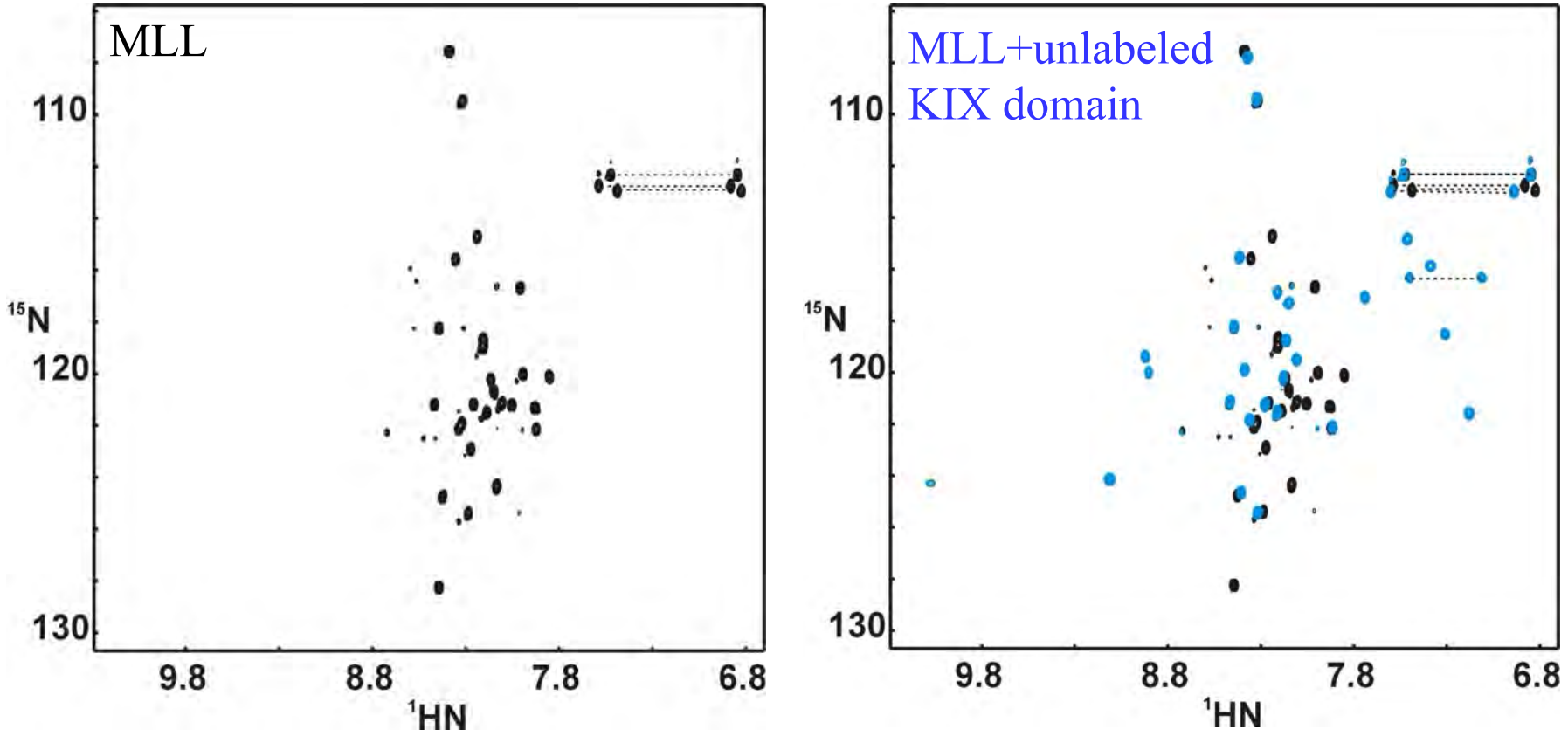
^{1}H-^{15}N HSQC tells you if protein is folded



The ^{15}N-^{1}H HSQC spectra on the left (*A*) has very weak dispersion along the ^{1}H axis (only 7.7-8.8 ppm) compared to (*B*) (6.8-9.2). Why is the dispersion so small?

Different chemical shifts reflect different environments, so a small dispersion implies uniformity of environment – *protein (A) is unfolded!*

^{1}H-^{15}N HSQC can monitor protein folding



The spectra in black (left and right) of MLL suggests it is unfolded.

The spectrum of MLL + KIX domain (blue, right) shows increased chemical shift dispersion suggestive that KIX binding leads to folding (J. Mol. Biol. (2006) **355**:1005). (KIX is not labeled with ^{15}N, so it is invisible in this experiment)

5) NMR versus X-ray structures

- Agreement between x-ray and NMR structures is excellent.
- A good NMR structure with backbone rmsd < 1.0 Å can be considered to be roughly equivalent to a 2 Å resolution x-ray structure
- Side chain conformations in NMR structures are well defined in the core, but not on the surface

NMR: Pros and Cons

- Pros:
  - NMR is ideal for proteins < 15 kDa
  - Proteins with unstructured regions can be difficult to crystallize, but can still be studied by NMR
  - There are no crystal packing forces (i.e. contacts between proteins) which can sometimes distort multi-domain proteins.
  - Protein-protein or protein-ligand interactions are easy to study.
  - Protein dynamics and conformational changes can be probed.

- Cons:
  - Crystal structures are generally better quality – i.e. better defined structures.
  - NMR is tedious: determine a lot of constraints.
  - Surface side-chain conformations are sometimes not resolved.
  - Size limitation to NMR (< 30kDa).
  - Isotope labeling (^{15}N, ^{13}C) is required for NMR.