Bio NMR spectroscopy

A. Spyros
aspyros@uoc.gr
Γ-207, Chemistry building

(Adapted from course material by Prof. Steven P. Smith,
School of Medicine, Department of
Biomedical and Molecular Sciences,
Queen's University, Canada.)

Protein NMR Spectroscopy

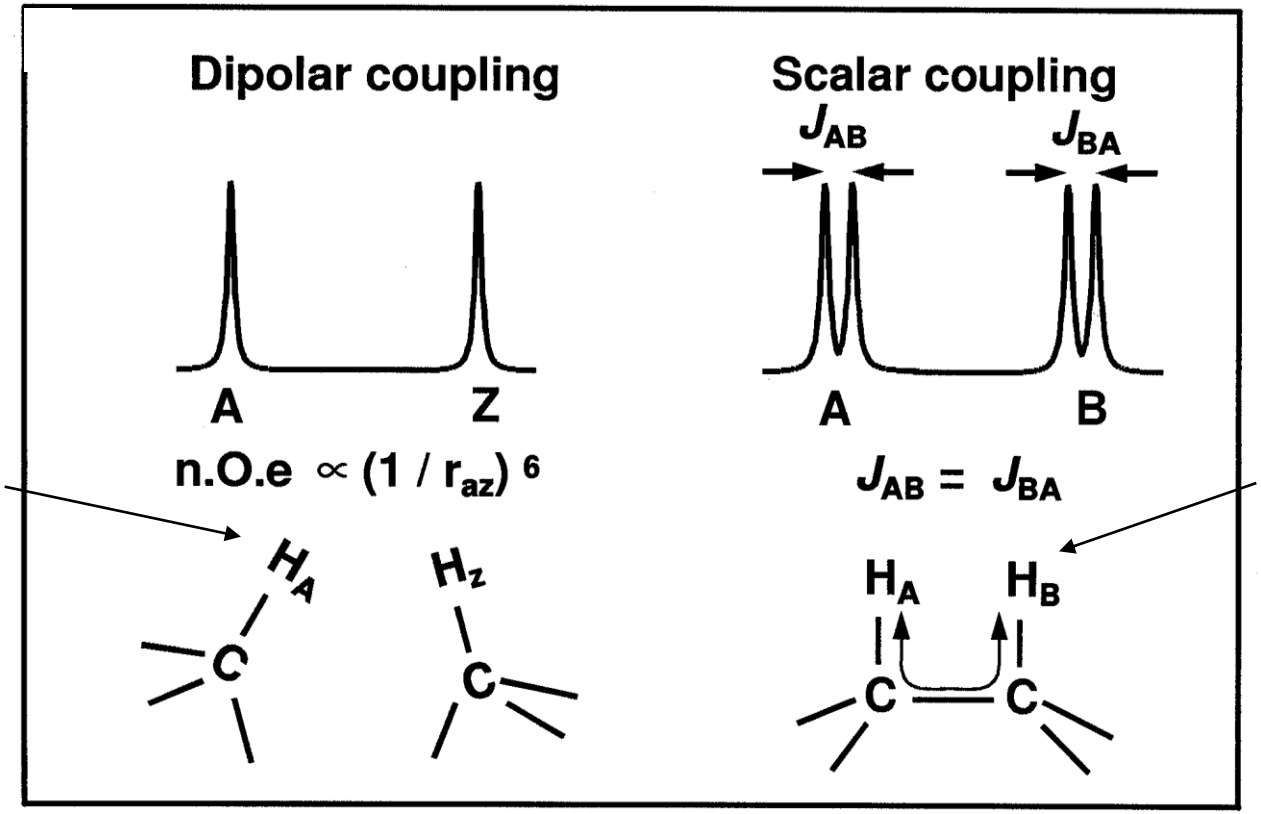
Determining three-dimensional structures and monitoring molecular interactions

Outline

- N-dimensional NMR
- Resonance assignment in proteins
- NMR-based structure determination
- Molecular interactions

NMR terminology

Scalar and Dipolar Coupling

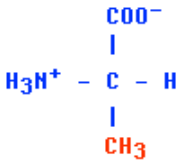


Through Space

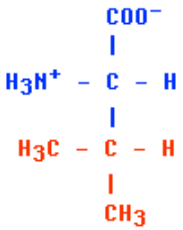
Through Bonds

Coupling of nuclei gives information on structure

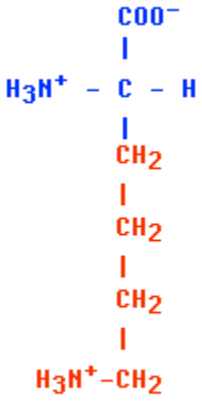
Examples of Amino Acids



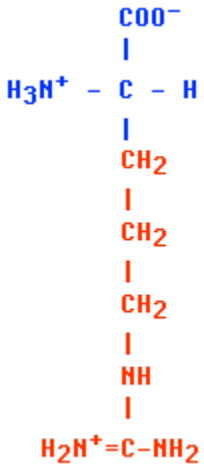
Alanine (Ala)[A]



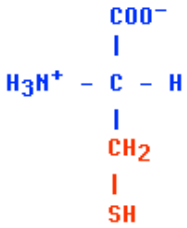
Valine (Val)[V]



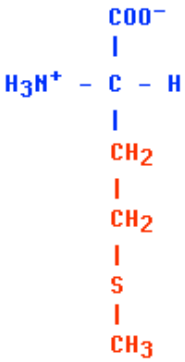
Lysine (Lys)[K]



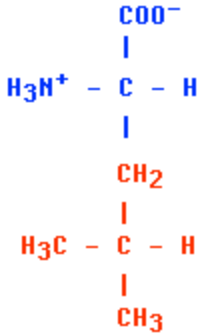
Arginine (Arg)[R]



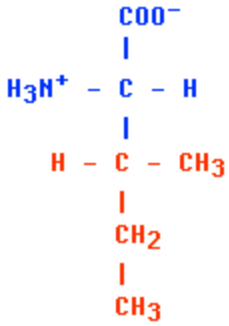
Cysteine (Cys)[C]



Methionine (Met)[M]



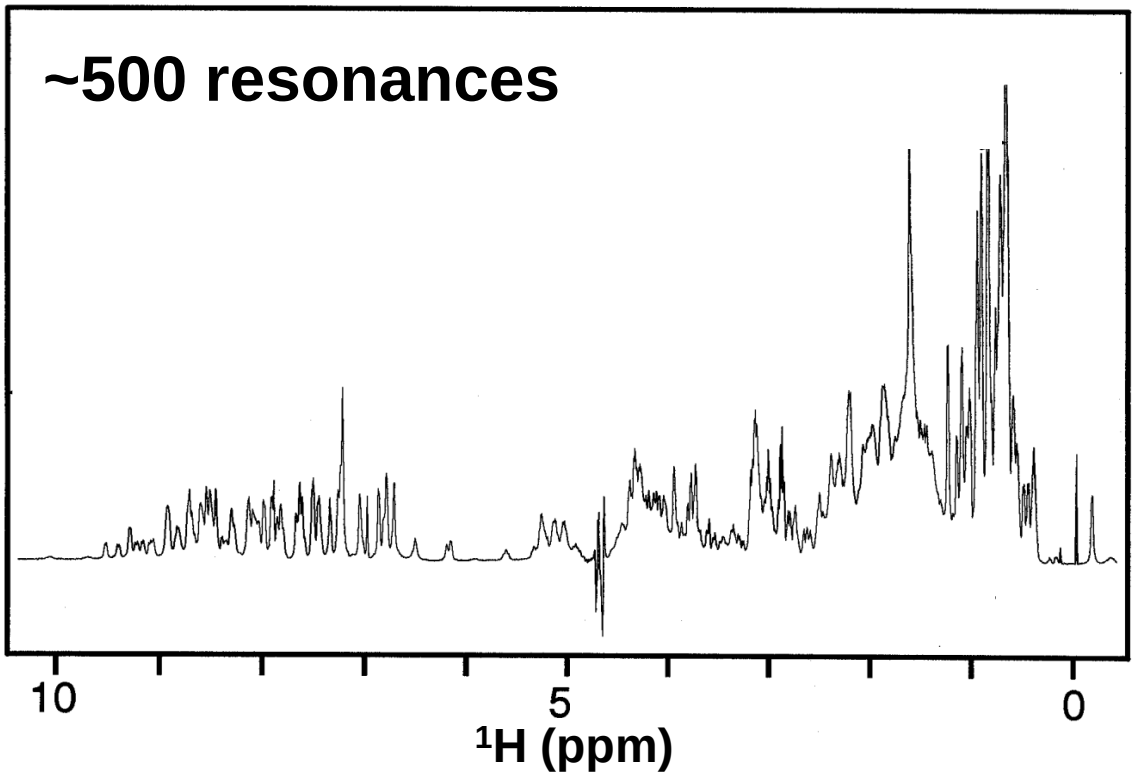
Leucine (Leu)[L]



Isoleucine (Ile)[I]

Proteins Have Too Many Signals!

^{1}H 1D NMR Spectrum of Ubiquitin



Resolve resonances by multi-dimensional experiments

Protein NMR: Practical Issues

Hardware:

- Magnet: homogeneous, high field - $$$$
- Electronics: stable, tunable
- Environment: temperature, pressure, humidity, stray fields

Sample Preparation:

- Recombinant protein expression (*E. coli*, *Pichia pastoris* etc)
- Volume: 300 μL – 600 μL
- Concentration: 1D ~ 50 μM, nD ~ 1mM ie. @ 20 kDa, 1mM = 10 mg
- Purity: > 95%, buffers
- Sensitivity (γ): isotope enrichment (^{15}N, ^{13}C)

Protein NMR: Practical Issues (cont.)

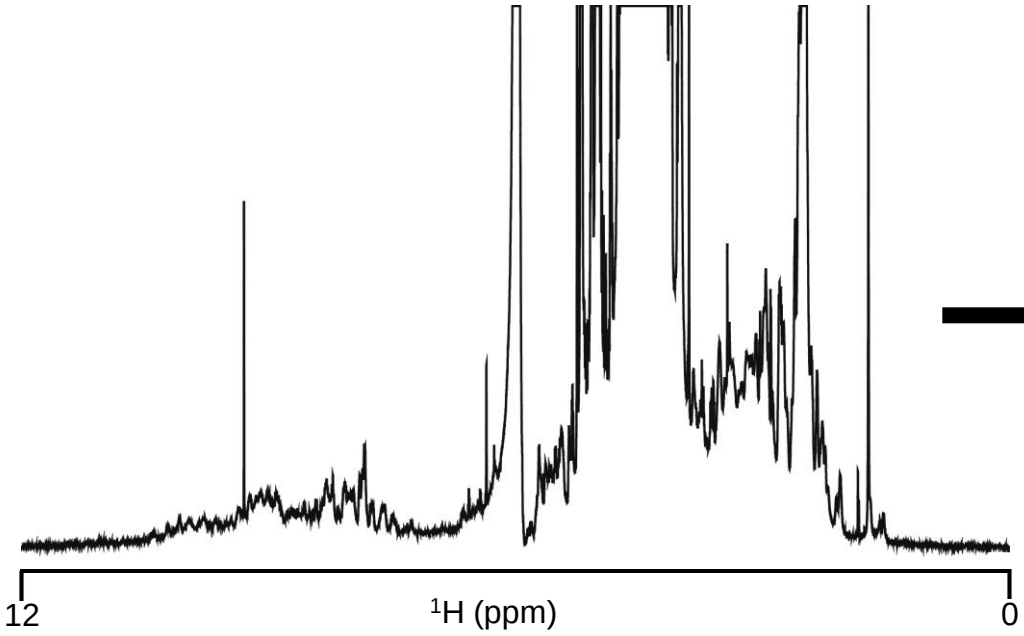
Solution Conditions:

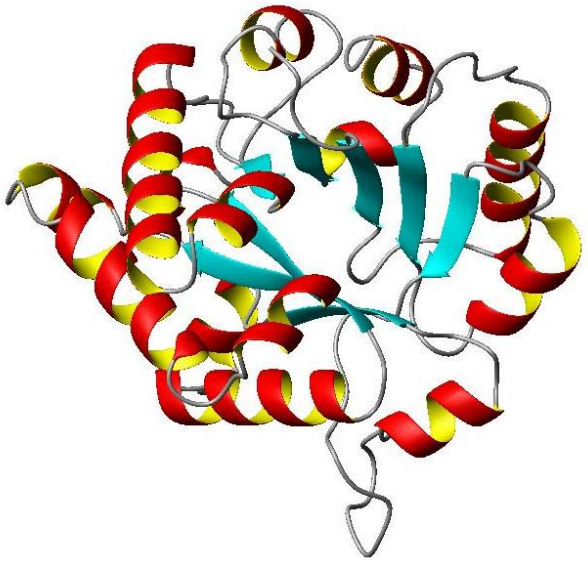
- Variables: buffer, ionic strength, pH, temperature
- Binding studies: co-factors, ligands
- No crystals!

Molecular Weight:

- up to 30 – 40 kDa for 3D structure determination
- > 100 kDa: uniform deuteration, residue and site-specific, atom-specific labeling
- Symmetry reduces complexity: 2 x10 kDa ≠ 20 kDa

NMR Spectrum to 3D structure?





Critical Features of Protein NMR Spectra

- The nuclei are not mutually coupled

> *Each amino acid gives rise to an independent NMR sub-spectrum, which is much simpler than the complete protein spectrum*

- Regions of the spectrum correspond to different parts of the amino acid

- Tertiary structure leads to increased dispersion of resonances
    - chemical shifts associated with each nucleus influenced by local chemical environment – nearby nuclei

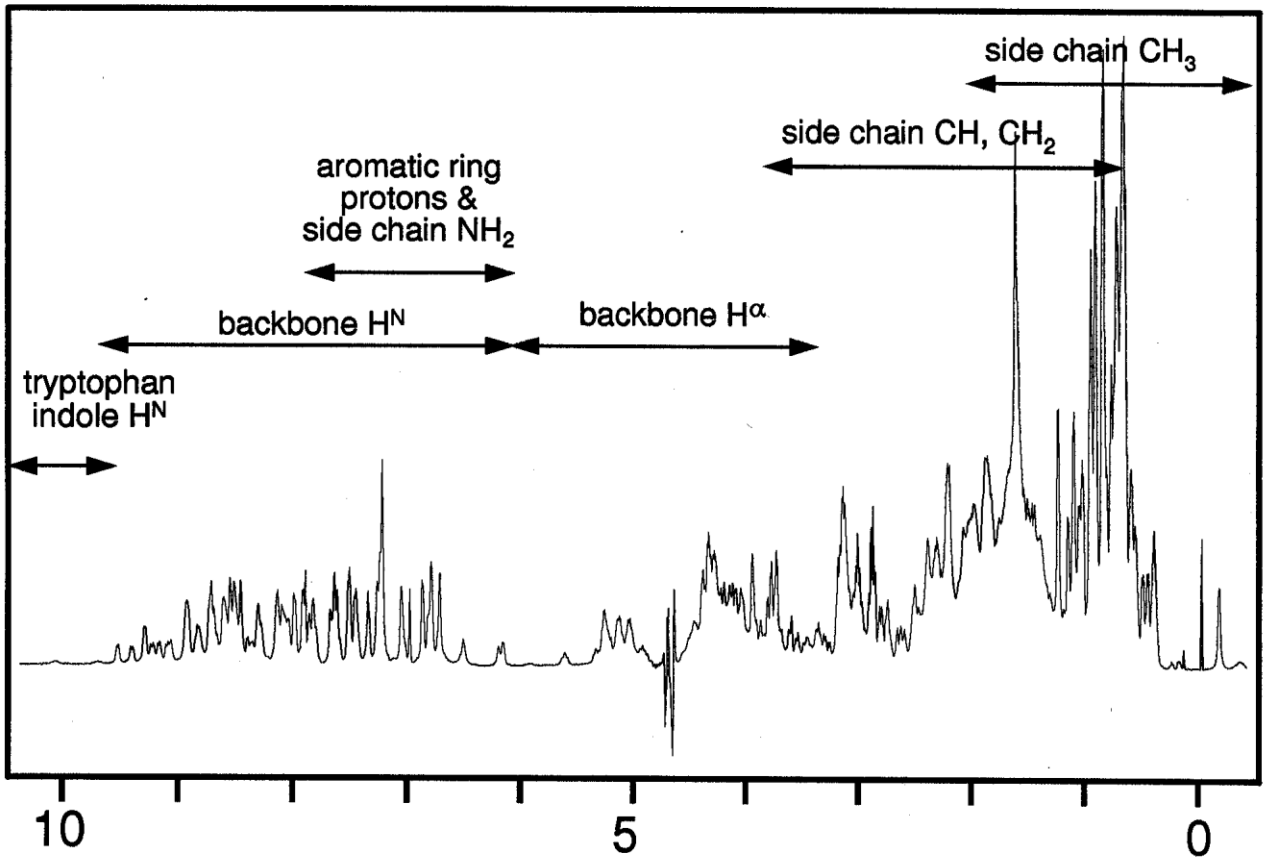
Solutions to the Challenges

1. Increase dimensionality of spectra to better resolve signals: $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$

2. Detect signals from heteronuclei (^{13}C, ^{15}N)

   - Better resolution of signals/chemical shifts not correlated nuclei
   - More information to identify signals
   - Lower sensitivity to MW of protein

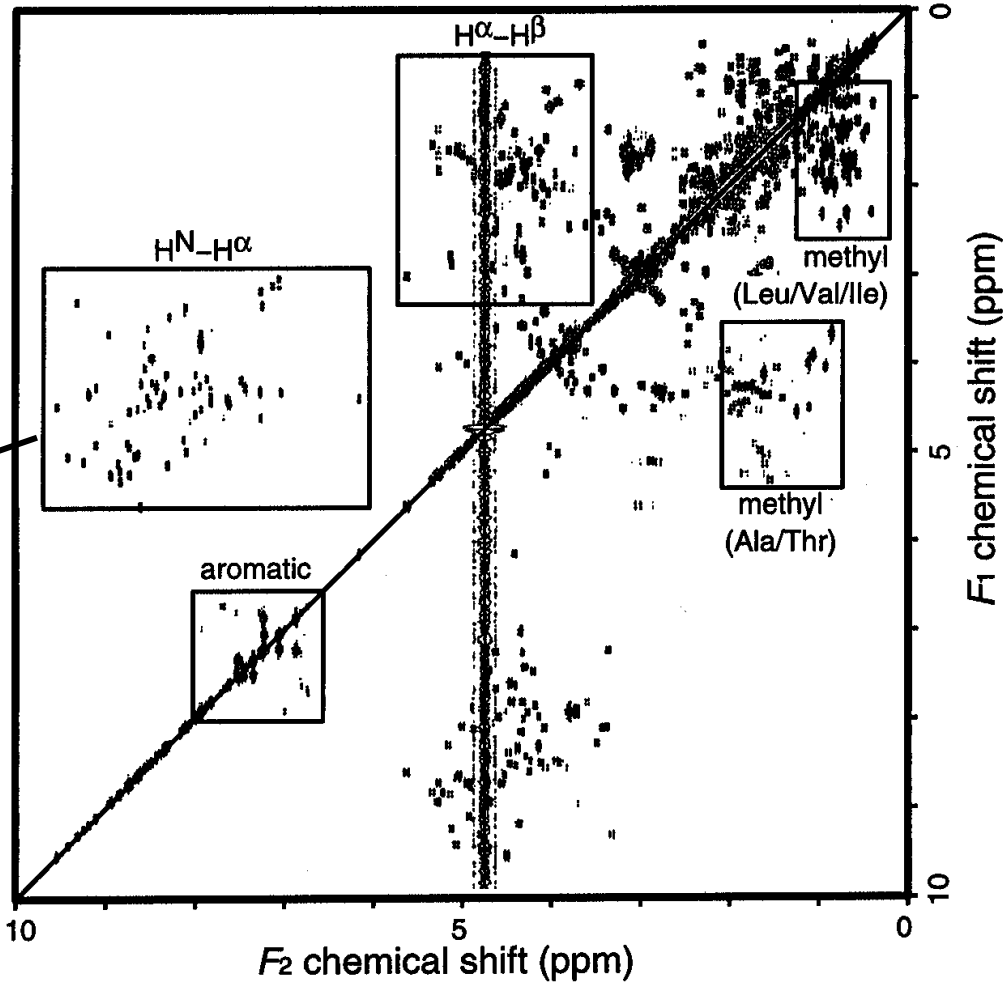
1D Protein ^{1}H NMR Spectrum



Resolve Peaks by Multi-D NMR

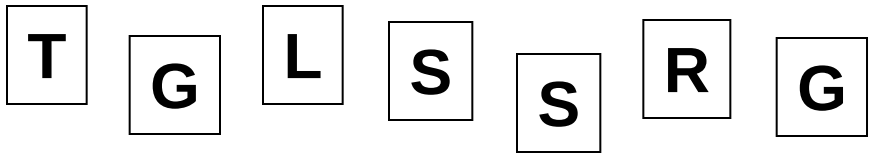
A BONUS→regions in 2D spectra provide protein fingerprints

If 2D cross peaks overlap→ go to 3D



Basic Strategy to Assign Resonances in Protein

1. Assign resonances for each amino acid



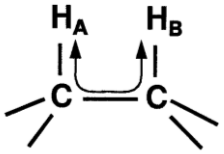
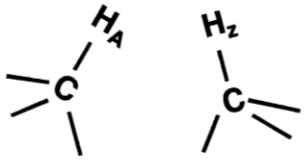
2. Put amino acids in order
   - Sequential assignment (**R-G-S**, **T-L-G-S**)
   - Sequence-specific assignment

1 2 3 4 5 6 7

R - G - S - T - L - G - S

Acronyms for Basic Experiments

Differ Only in the Nature of Mixing

	Homonuclear	***Hetero***nuclear
Scalar Coupling (thru-bond) 	**COSY** **CO**rrelation **S**pectroscop**Y**	**HSQC** **H**eteronuclear
	TOCSY **TO**tal **C**orrelation **S**pectroscop**Y**	**Hetero-TOCSY**
Dipolar Coupling (thru-space) 	**NOESY** **N**uclear **O**verhauser **E**ffect (**E**nhancement) **S**pectroscop**Y**	**NOESY-HSQC**

Homonuclear ^{1}H Assignment Strategy

- For proteins up to ~ 10 kDa

•Scalar couplings to identify resonances/spin systems/amino acids, dipolar couplings to place in sequence

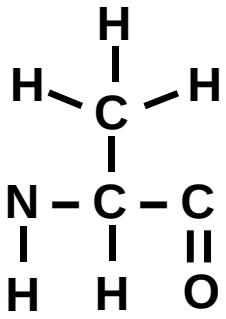
- Based on backbone H^N (unique region in ^{1}H spectrum, greatest dispersion of resonances, least overlap)

- ***Concept:*** Build out from the backbone to identify the side-chain resonances (unique spin systems)

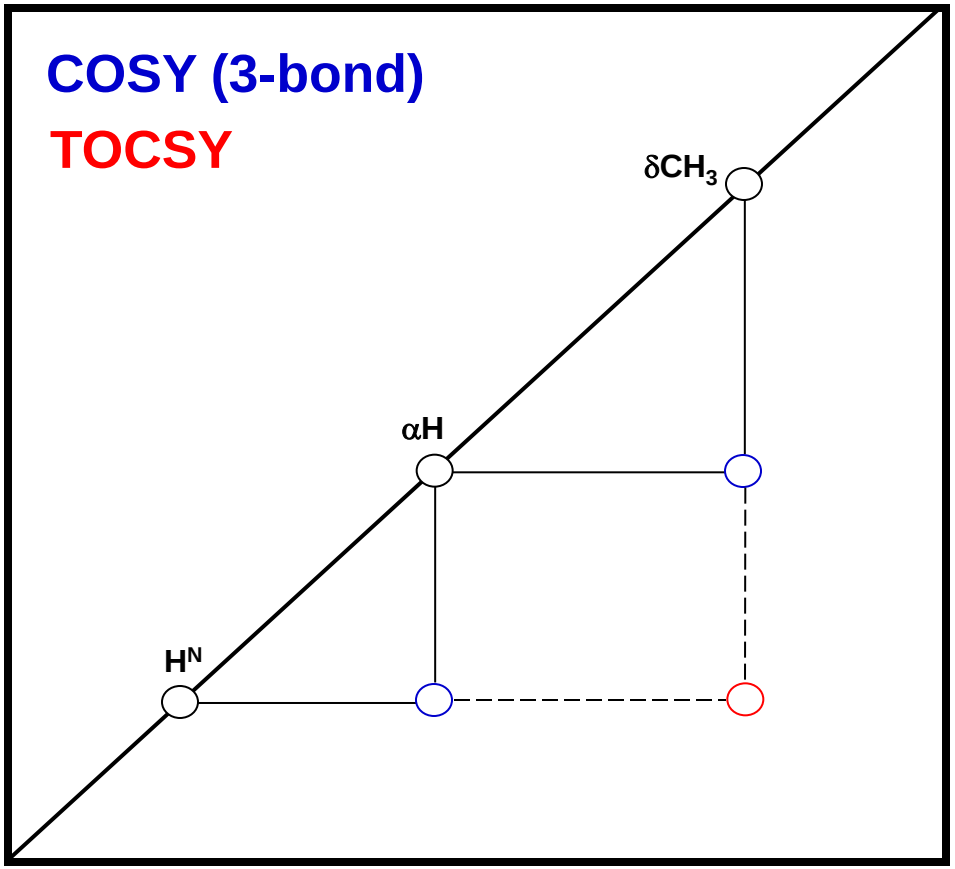
- 2^{nd} dimension resolves overlap, 3D rare

Homonuclear ^{1}H Assignment Strategy

Step 1: Identify Spin System

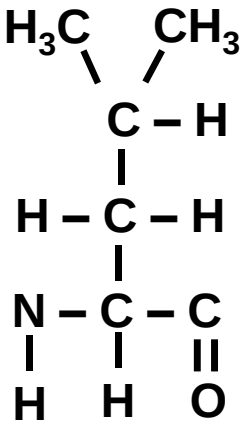


Alanine

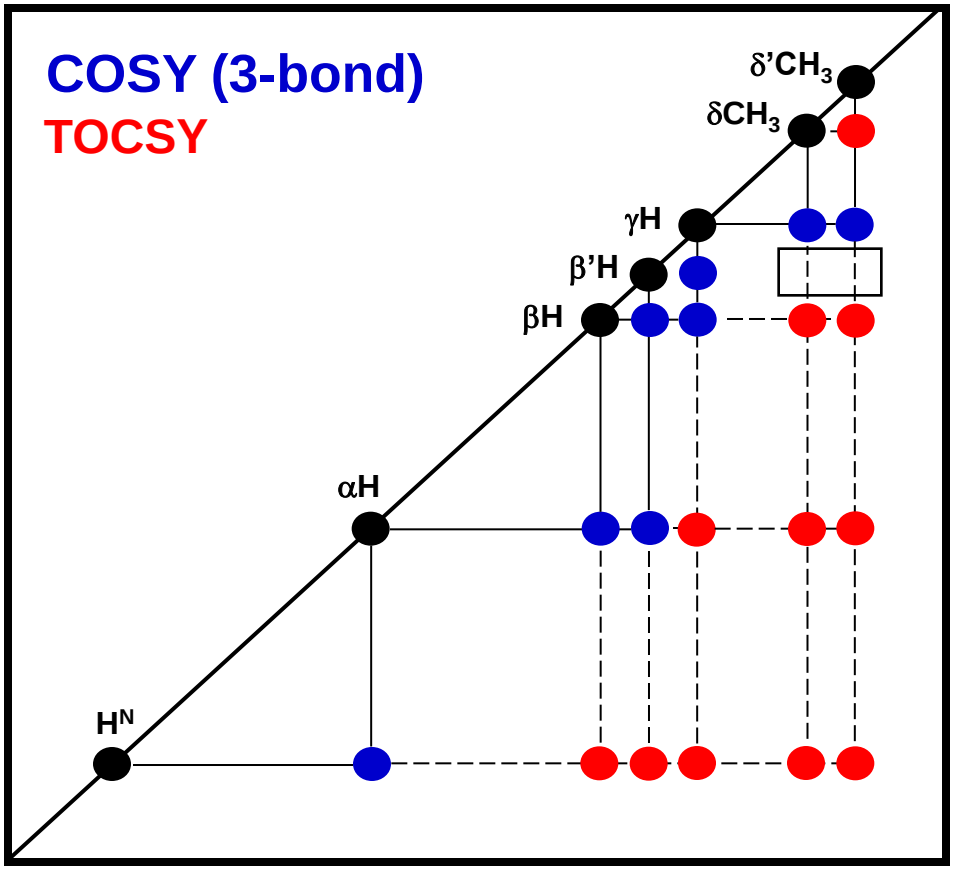


Homonuclear ^{1}H Assignment Strategy

Step 1: Identify Spin System

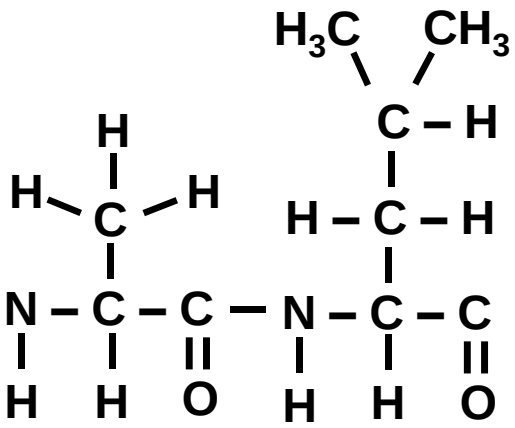


Leucine



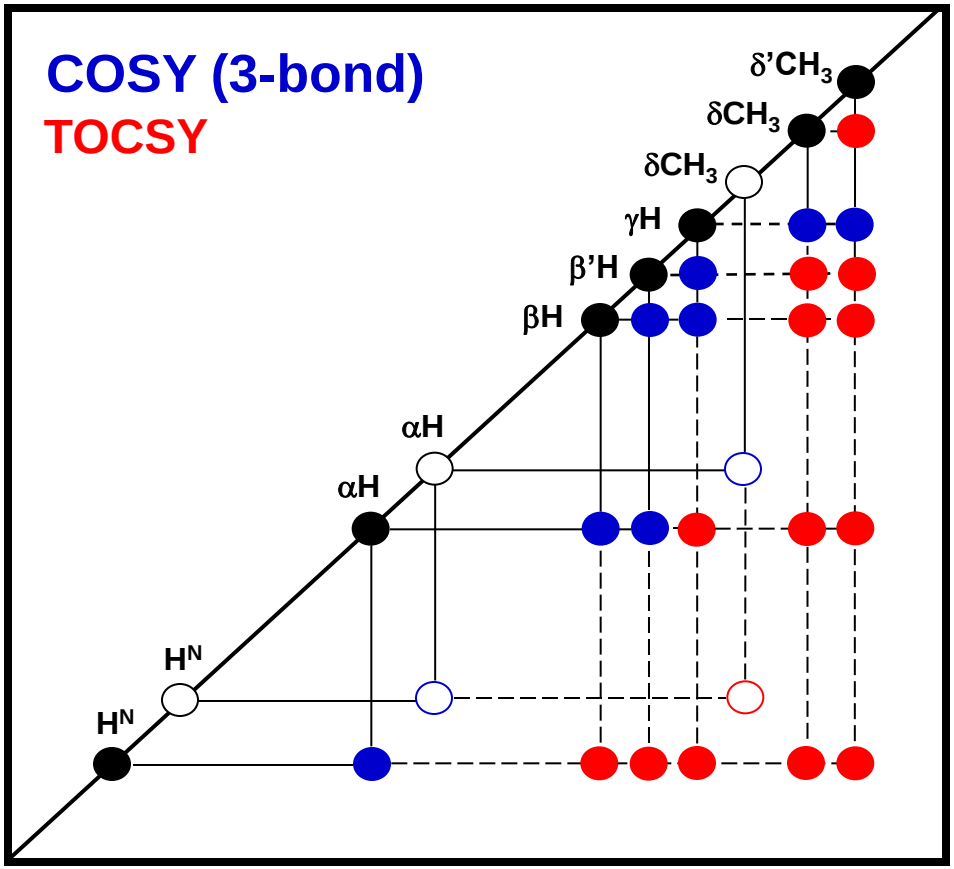
Homonuclear ^{1}H Assignment Strategy

Step 1: Identify Spin System



Alanine
- open circles

Leucine
- closed circles

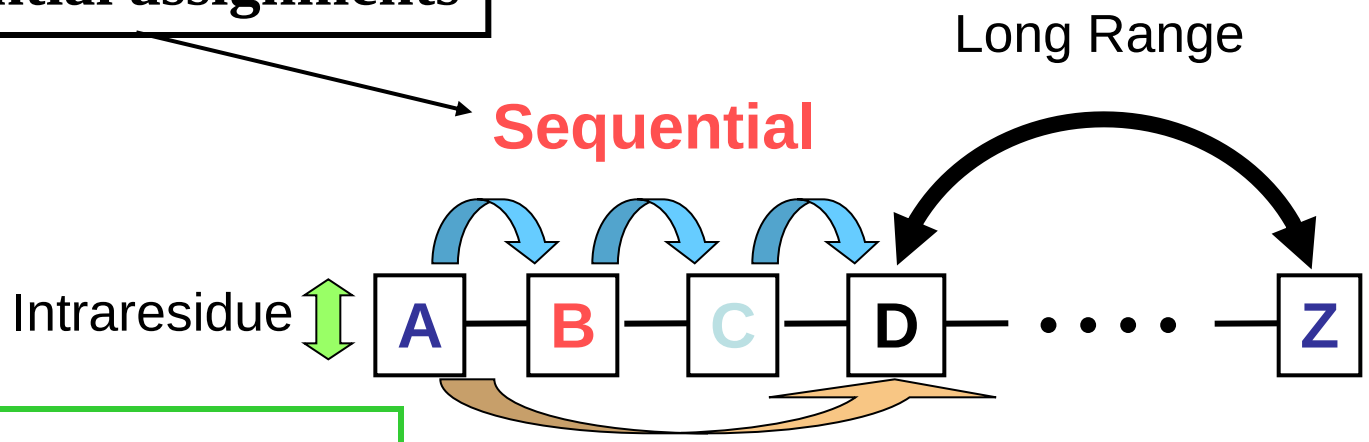


Homonuclear ^{1}H Assignment Strategy

Step 2: Fit residues in sequence

Minor Flaw: All NOEs mixed together!

Use only these to make sequential assignments

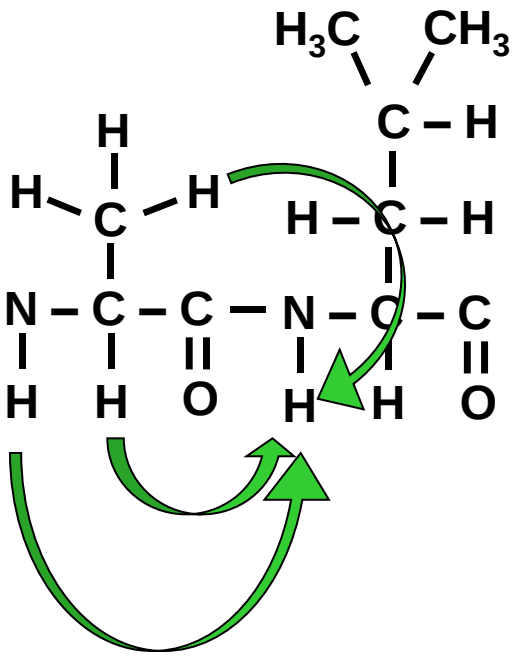


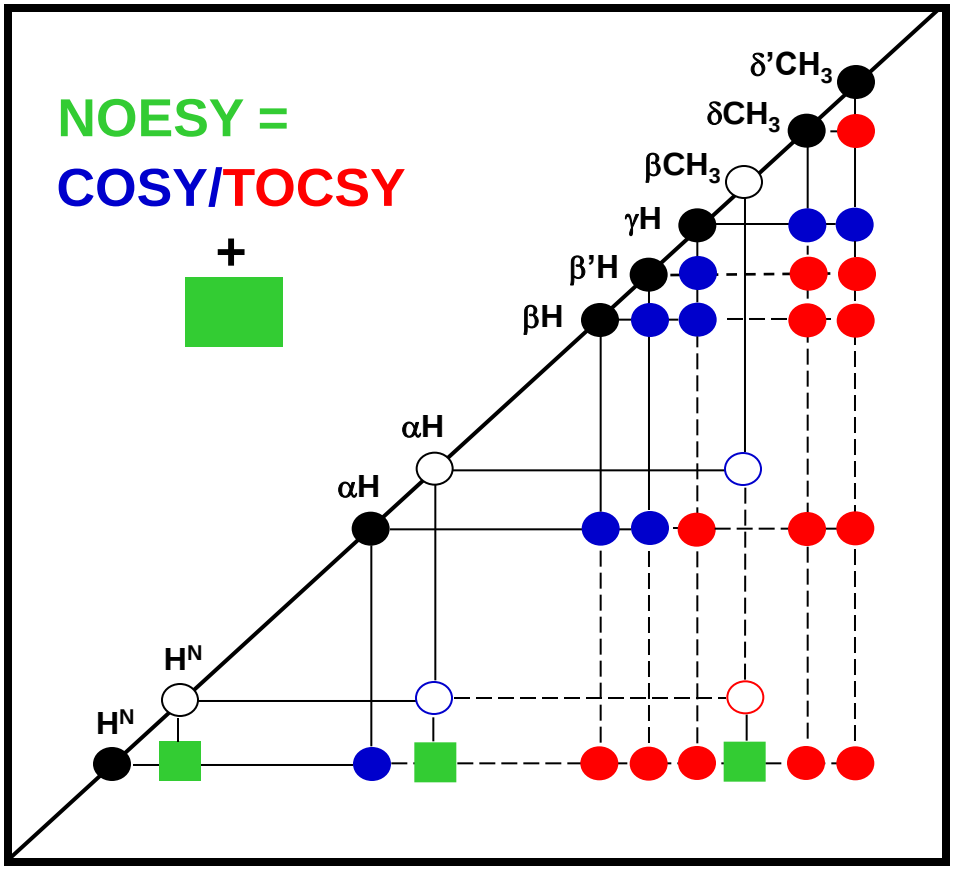
- Sequential NOEs
  - $\Rightarrow H^N$-H^N (i, i + 1)
  - $\Rightarrow H^\alpha$-H^N (i, i + 1)

Medium-range
(helices: Hα-HN (i, i + 3,4)))

Homonuclear ^{1}H Assignment Strategy

Step 2: Fit residues in sequence





Extended Homonuclear ^{1}H Strategy

- For proteins up to ~ 15 kDa

- Same basic idea as ^{1}H strategy: based on backbone H^{N}

- ***Concept:*** When backbone ^{1}H overlaps $\Rightarrow$ disperse with backbone ^{15}N

- Use heteronuclear 3D experiments to increase signal resolution

$$^{1}H \leftrightarrow {}^{1}H \leftrightarrow {}^{15}N$$

Solutions to the Challenges

1. Increase dimensionally of spectra to better resolve signals: $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$

2. Detect signals from heteronuclei (^{13}C, ^{15}N)

   - Labeling with NMR-observable ^{13}C, ^{15}N isotopes
   - Better resolution of signals/chemical shifts not correlated nuclei
   - More information to identify signals
   - Lower sensitivity to MW of protein

Isotopic Labeling

• Require uniform $^{15}N/^{13}C$ labeling ie. Every carbon and nitrogen isotopically labeled

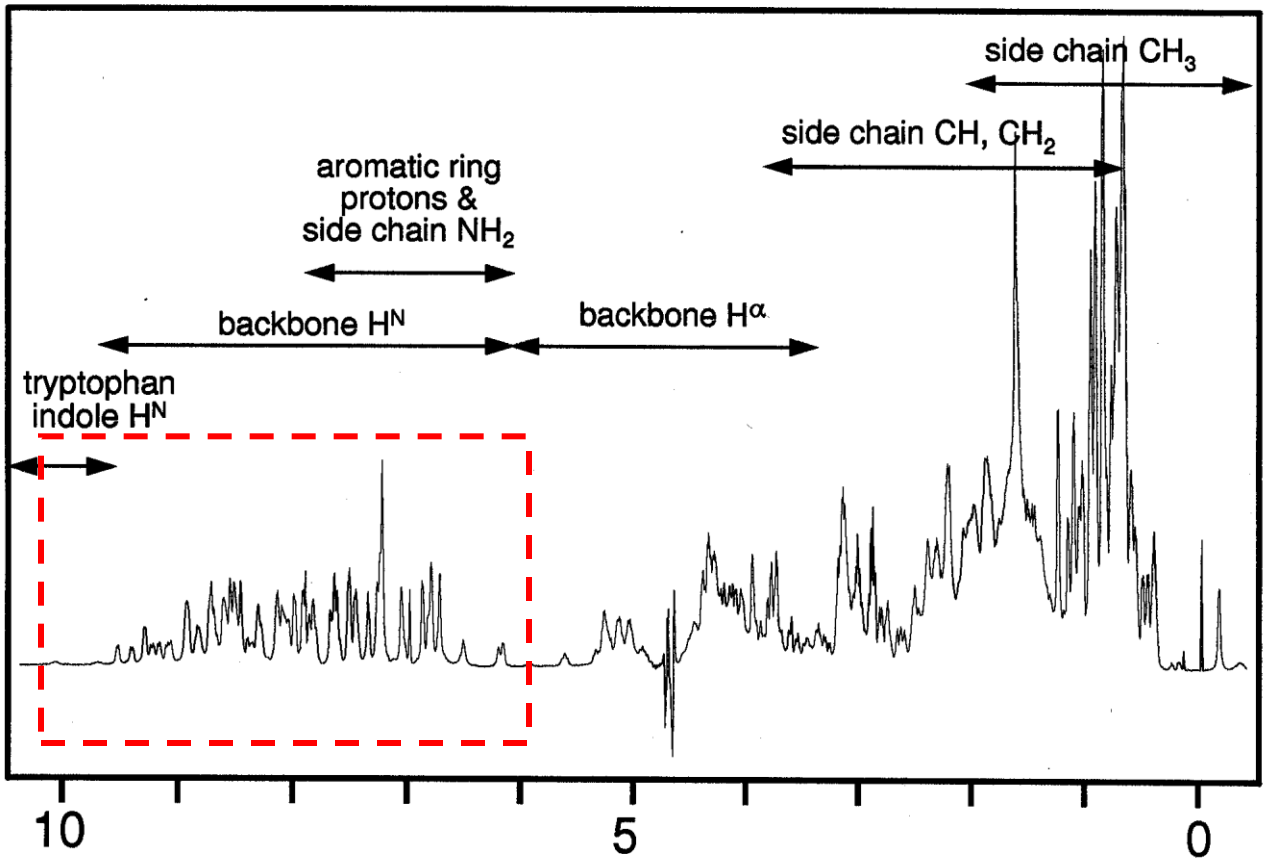
How?

• Grow bacteria on minimal media (salts) supplemented with ^{15}N-NH_4Cl and ^{13}C-glucose as soles sources of nitrogen and carbon

• lower yields than protein expression than on enriched media, therefore need very good recombinant expression system

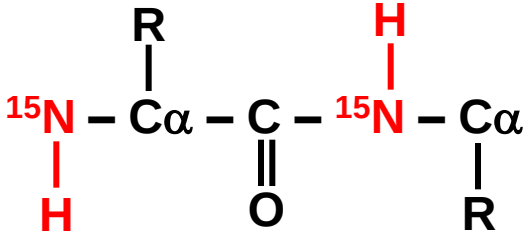
Double Resonance Experiments

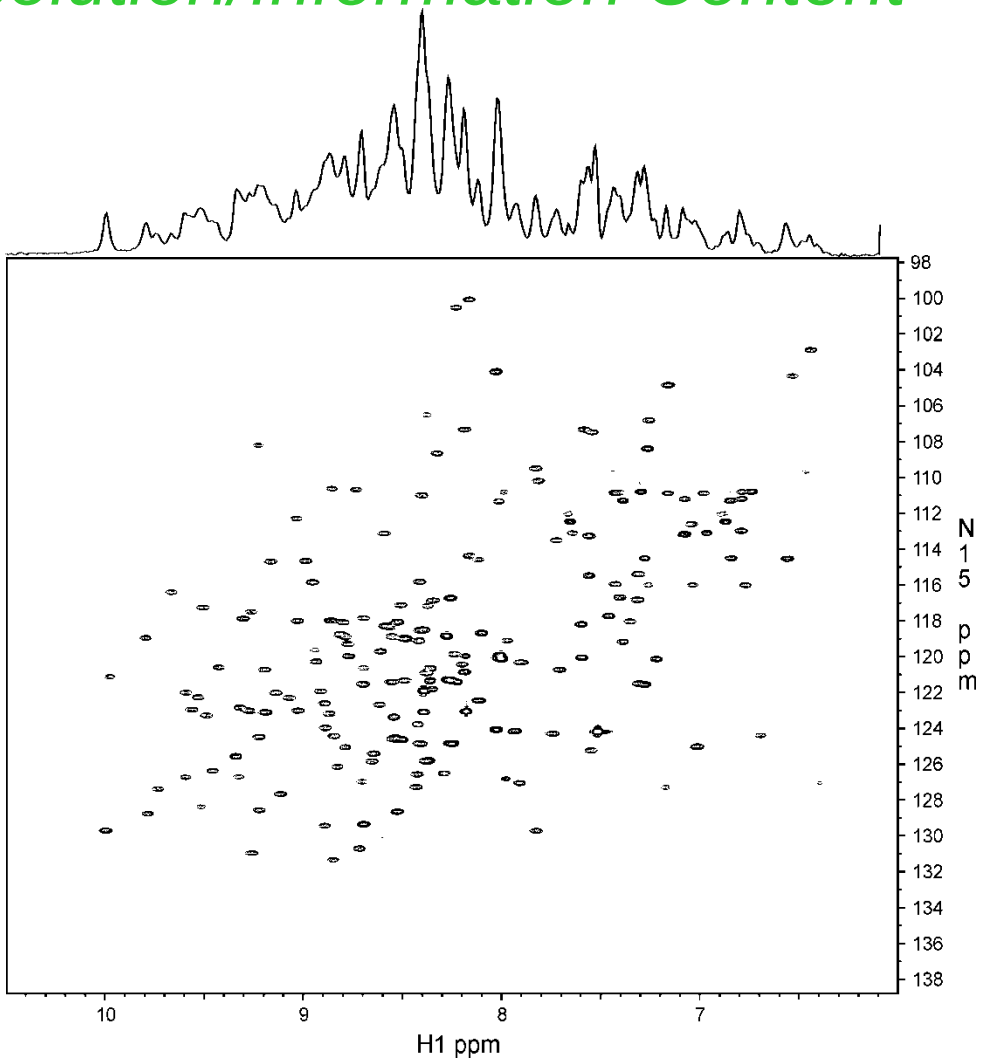
Increases Resolution/Information Content



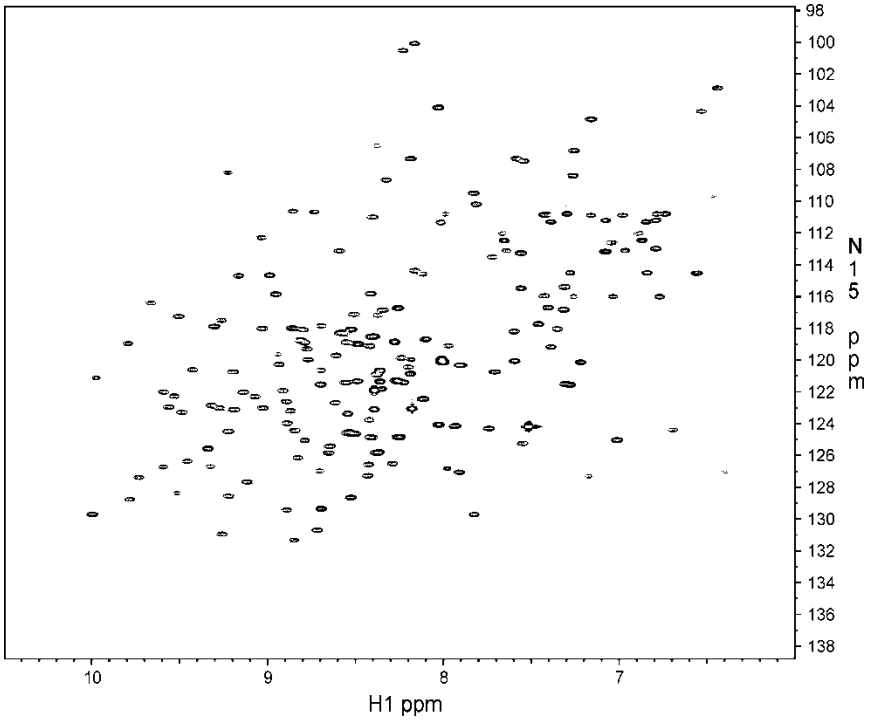
Heteronuclear NMR: ^{15}N-Edited Experiments

Increases Resolution/Information Content

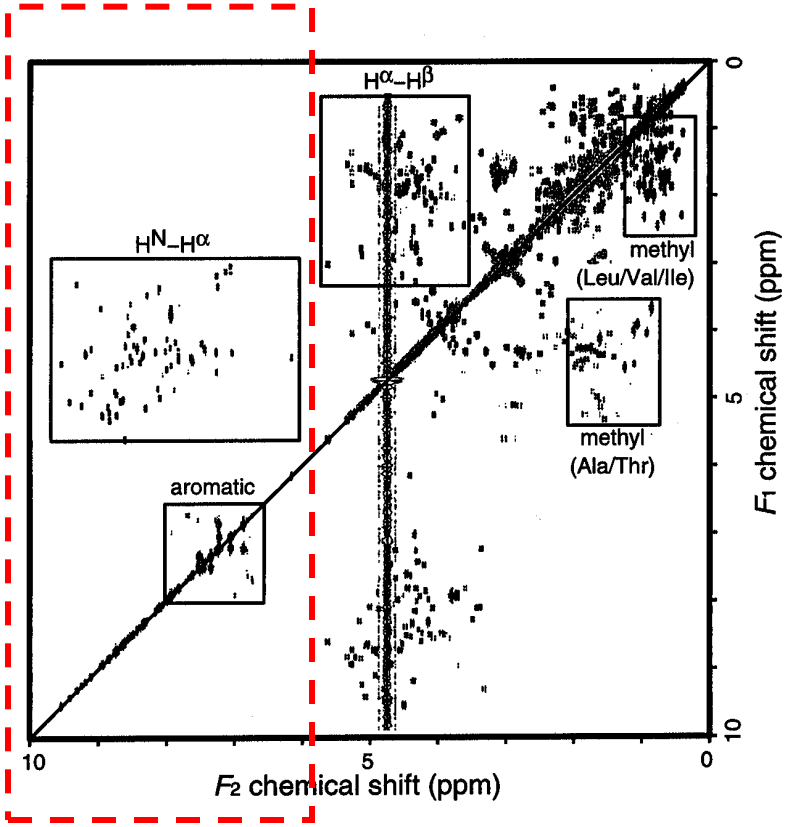




3D Heteronuclear NMR: ^{15}N-Edited Experiments

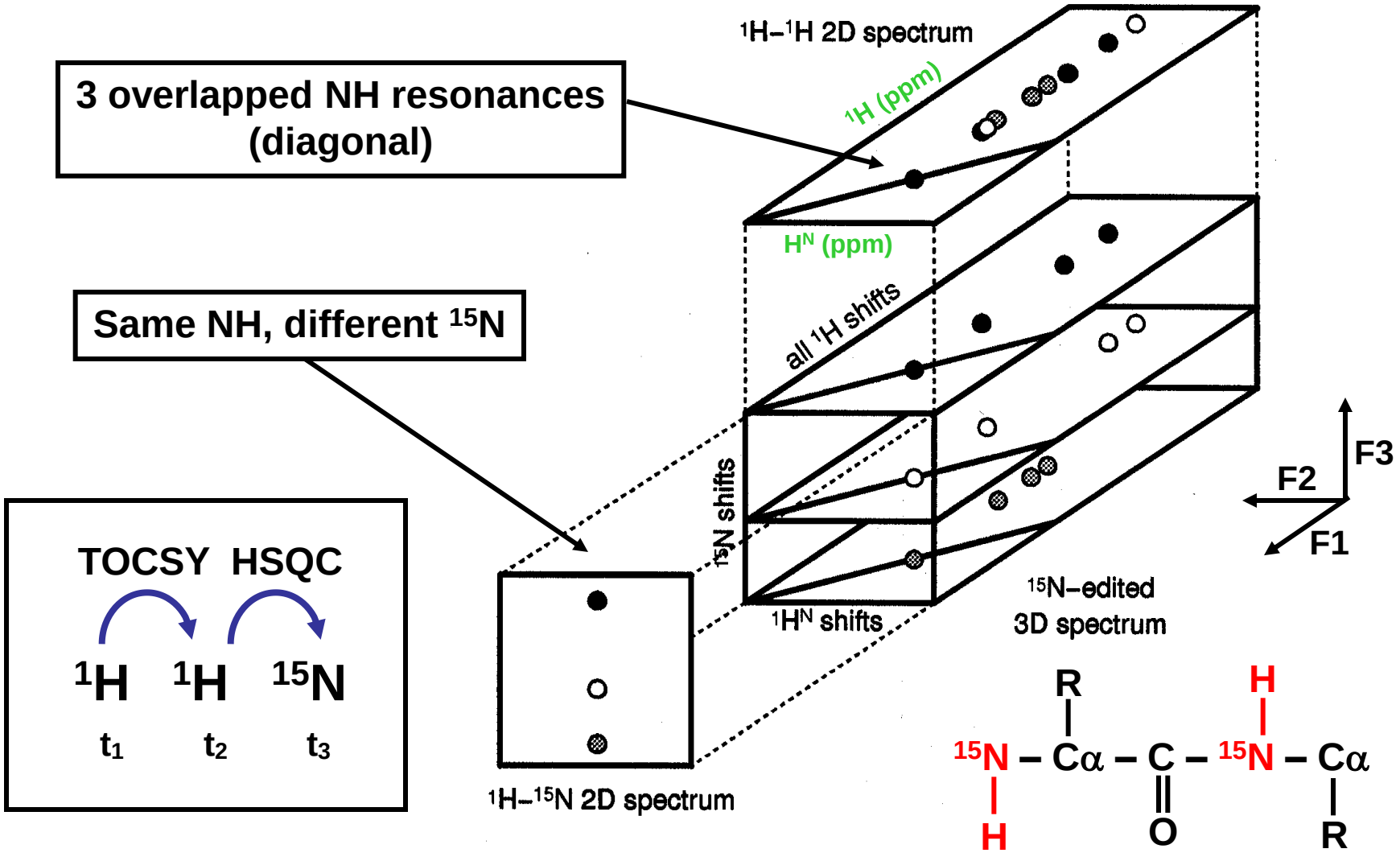


+



Extended Homonuclear ^{1}H Strategy

^{15}N dispersed ^{1}H-^{1}H TOCSY



Summary of Homonuclear Assignment Strategy

- for proteins up to ~10 kDa (2D homonuclear) and proteins up to ~ 15 kDa (^{15}N-labeling and 3D)

- using scalar coupling-type experiments (COSY, TOCSY) assign spin systems/side-chain resonances

- Connect amino acids (identified based on spin systems) sequentially using NOE-type experiments and characteristic sequential NOEs (H^N-H^N (i, i+1); Hα-H^N (i, i+1))

Heteronuclear (^{1}H, ^{13}C, ^{15}N) Strategy

- for larger proteins (backbone assignment: ~70 kDa; full structure determination: ~40 kDa)

- Assign resonances (chemical shifts) for all atoms (except O)

- Handles overlap in backbone H^{15N} region
  disperse with backbone **C', C^{α}, H^{α} ,C^{β}, H^{β}**

- Heteronuclear 3D/4D increases resolution

$$^{1}H \leftrightarrow {}^{13}C \leftrightarrow {}^{1}H \leftrightarrow {}^{15}N$$

- Works on bigger proteins because scalar couplings are larger

Heteronuclear (^{1}H, ^{13}C, ^{15}N) Strategy

Step 1: Sequence-specific backbone assignment

Assign backbone ^{1}H, ^{15}N, C^{α}, C^{β} resonances/chemical shifts and sequentially link amino acids using partner scalar coupling experiments

Step 2: Side-chain assignment

Assign side-chain ^{13}C & ^{1}H resonances/chemical shifts using TOCSY-type 3D scalar coupling experiments

**** Have complete list of chemical shifts for all ^{13}C, ^{15}N, ^{1}H atoms in protein ****

Heteronuclear (^{1}H, ^{13}C, ^{15}N) Assignments

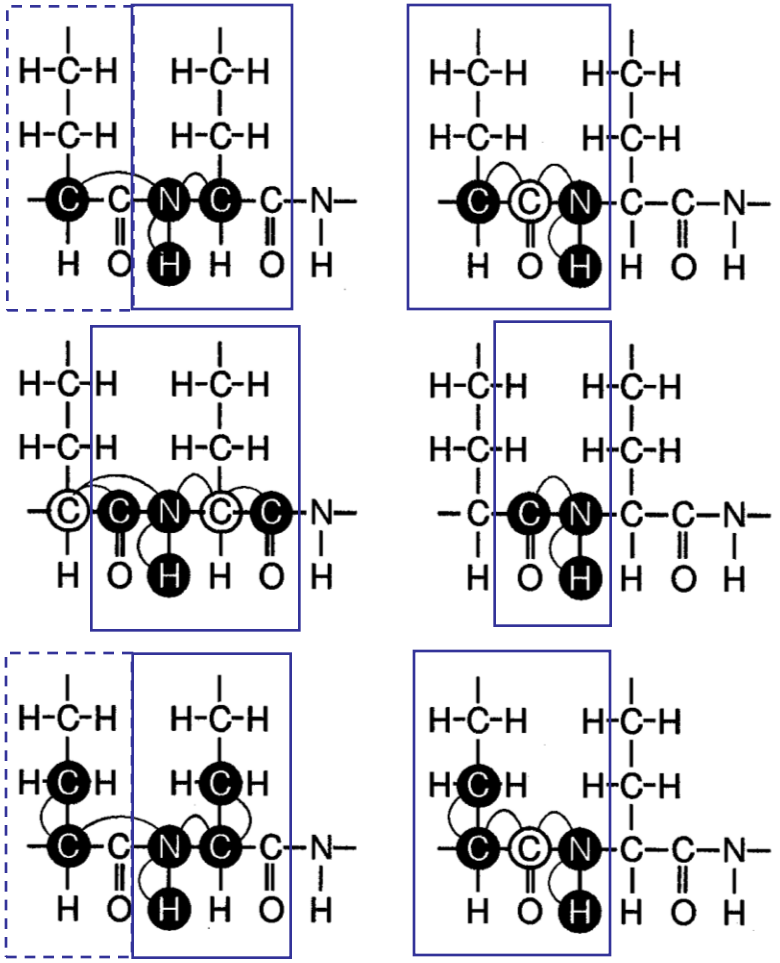
Backbone Experiments

Names of scalar experiments based on atoms detected

HNCA / HN(CO)CA

HN(CA)CO / HNCO

HNCACB / HN(CO)CACB



Consecutive residues!!

NOESY not needed

Heteronuclear (^{1}H, ^{13}C, ^{15}N) Assignments

Backbone Experiments

CBCA(CO)NH
- **inter-residue connectivity**

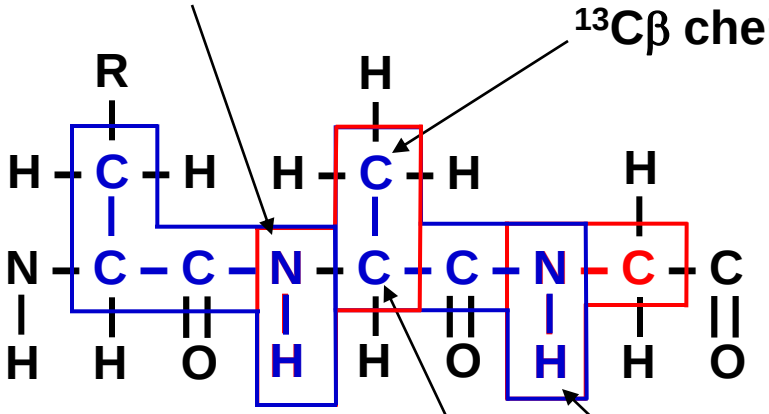
(HN to previous Cα, Cβ)

HNCACB

- intra-residue connectivity

(HN to own Cα, Cβ)

Search ^{15}N planes for $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts



$^{13}C\beta$ chemical shift

$^{13}C\alpha$ chemical shift

common ^{15}N and H^N chemical shift in both experiments (found on same ^{15}N plane)

Heteronuclear (^{1}H, ^{13}C, ^{15}N) Assignments

Backbone Experiments

CBCA(CO)NH
- **inter-residue connectivity**

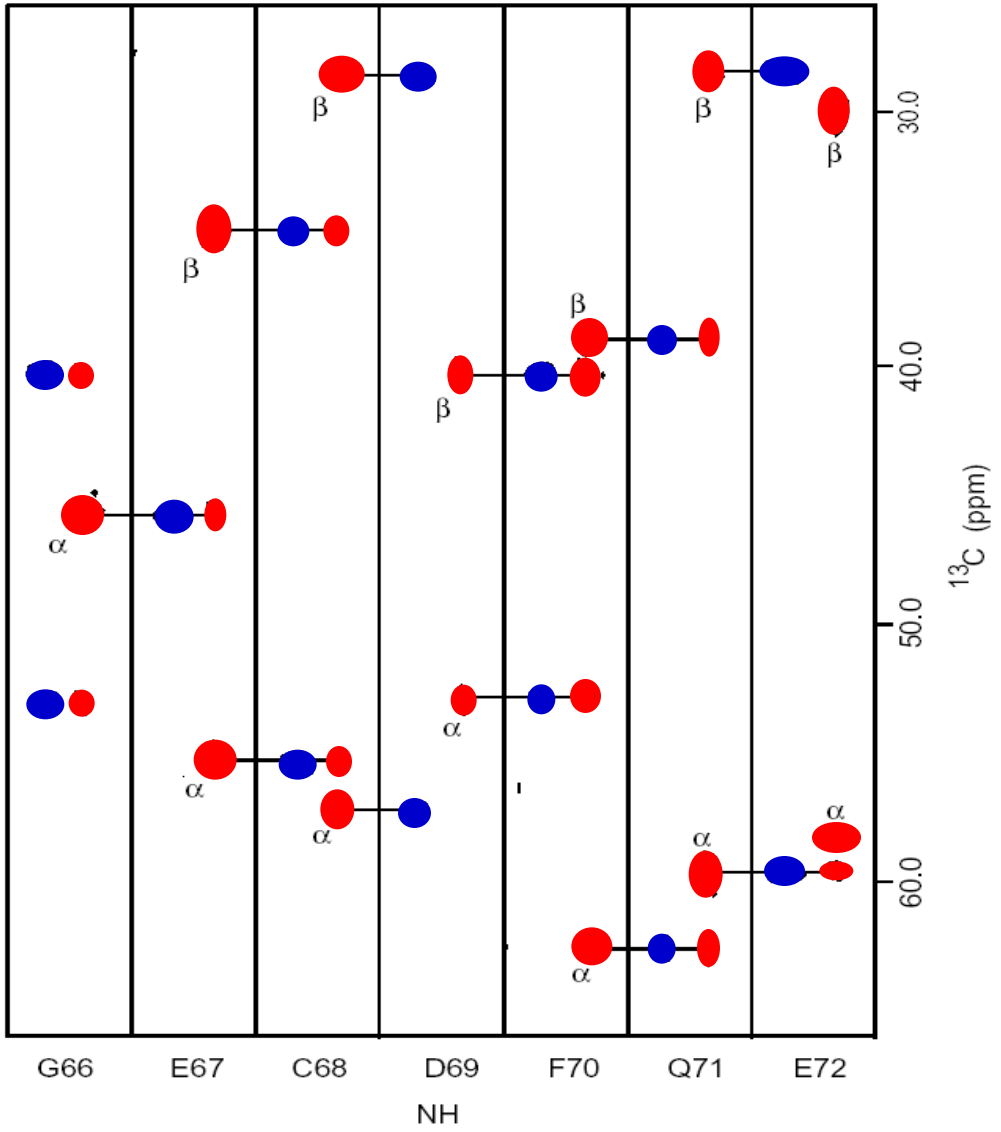
(HN to previous Cα, Cβ)

HNCACB
- **intra-residue connectivity and possibly inter-residue**

(HN to own Cα, Cβ)

Start with unique residue

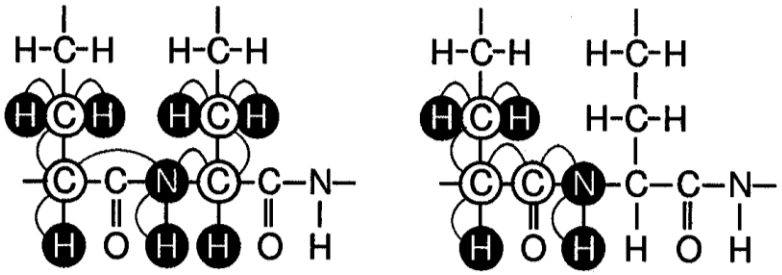
1. **Gly – only Cα**
2. **Ala – upfield-shifted Cβ (~18 ppm)**
3. **Thr/Ser – downfield-shifted Cα & Cβ which are close to each other**



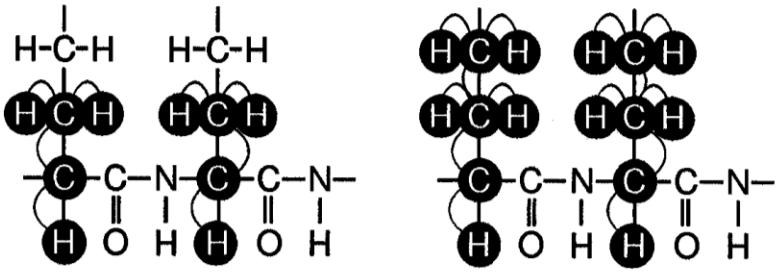
Heteronuclear (^{1}H, ^{13}C, ^{15}N) Assignments

Side-chain Experiments

HBHA(CBCA)NH /
HBHA(CBCACO)NH



HCCH–COSY /
HCCH–TOCSY



Multiple redundancies increase reliability

Heteronuclear (^{1}H, ^{13}C, ^{15}N) Assignments

Key Points

• Enables the study/assignment of much larger proteins (up to ~100 kDa)

•Scalar coupling-type 3-dimensional experiments only

•***Bonus:*** Amino acid identification and sequence-specific assignment all at once

• Most efficient but experiments are more complex

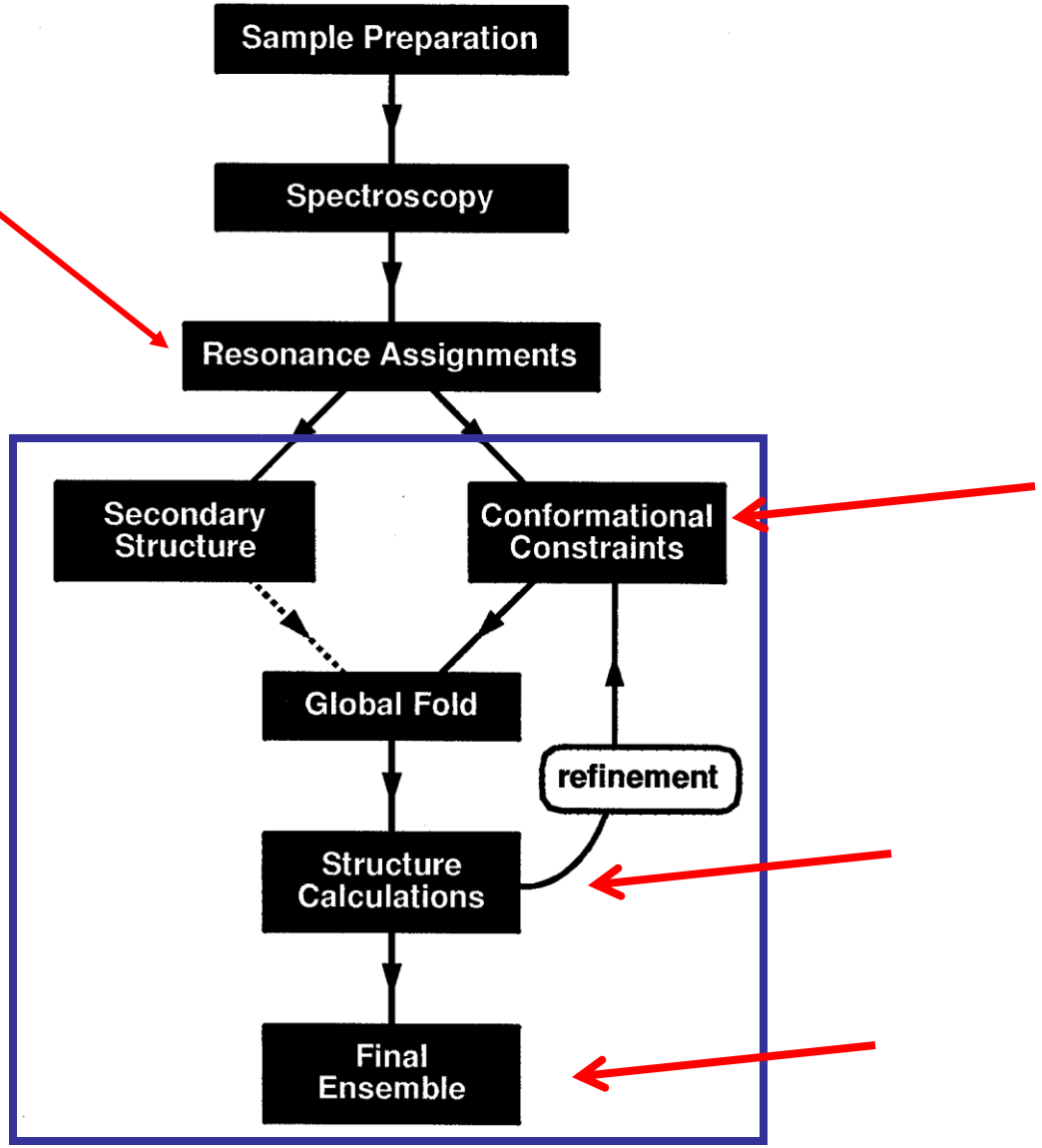
•Requires ^{13}C, ^{15}N enrichment (also ^{2}H)

$\Rightarrow$High expression levels on minimal media

$\Rightarrow$ Increased cost ($150/g ^{13}C-gluocose; $30/g $^{15}NH_4Cl$)

Structure Determination Overview

List of chemical shifts for all nuclei in protein (^{1}H, ^{13}C, ^{15}N)



NMR Experimental Observables Provide Structural Information

1. Backbone conformation from chemical shifts (Chemical Shift Index – CSI; H^{α}, C^{α}, C^{β}, C')

2. Hydrogen bond constraints

3. Backbone and side chain dihedral angle constraints from scalar couplings

4. Distant constraints from NOE connectivities

1. Chemical Shift Index

• Comparison of H^{α}, C^{α}, C^{β}, C' determined chemical shifts from protein to standard random coil chemical shift values

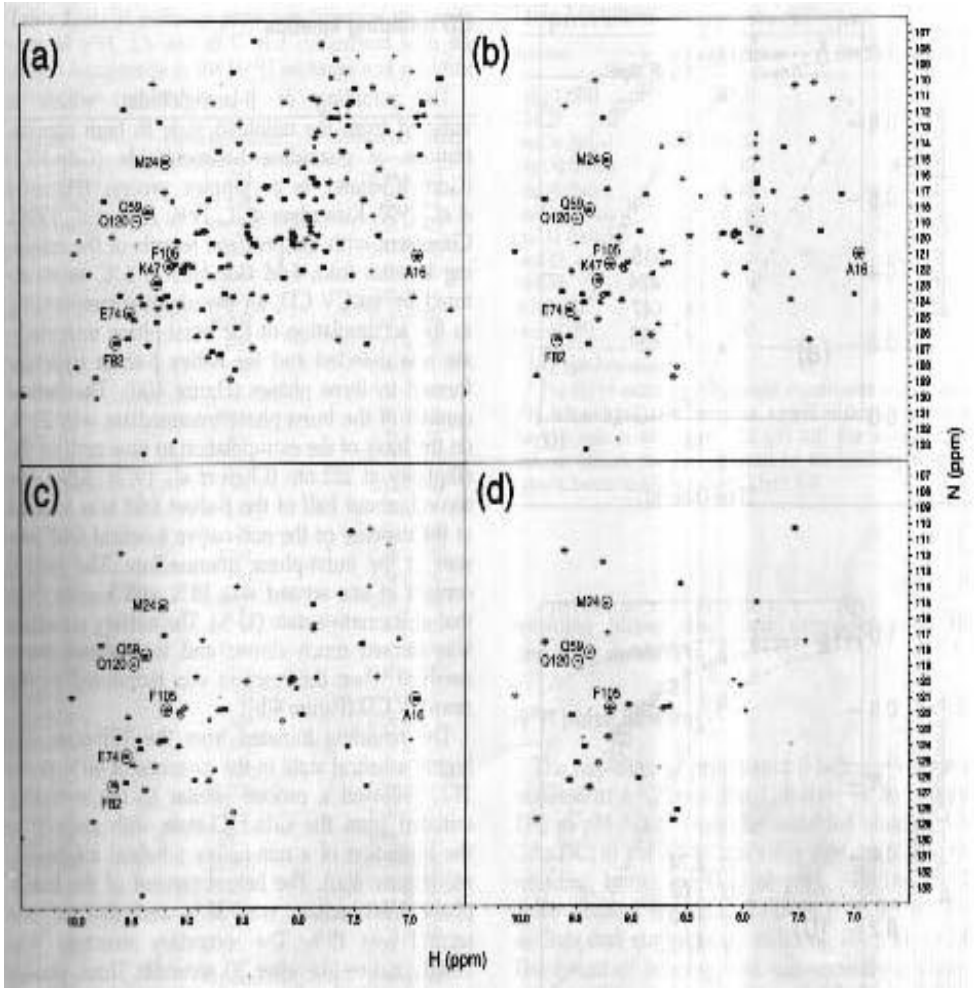
• Upfield-shifted H^{α} and C^{β} and downfield-shifted C^{α} and C' values indicate amino acid residues in an α-helical conformation (requires three consecutive residues displaying this pattern)

• Downfield-shifted H^{α} and C^{β} and upfield-shifted C^{α} and C' values indicate residues in an extended (β-strand) conformation

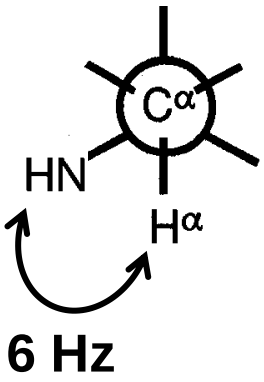
2. Hydrogen Bonds

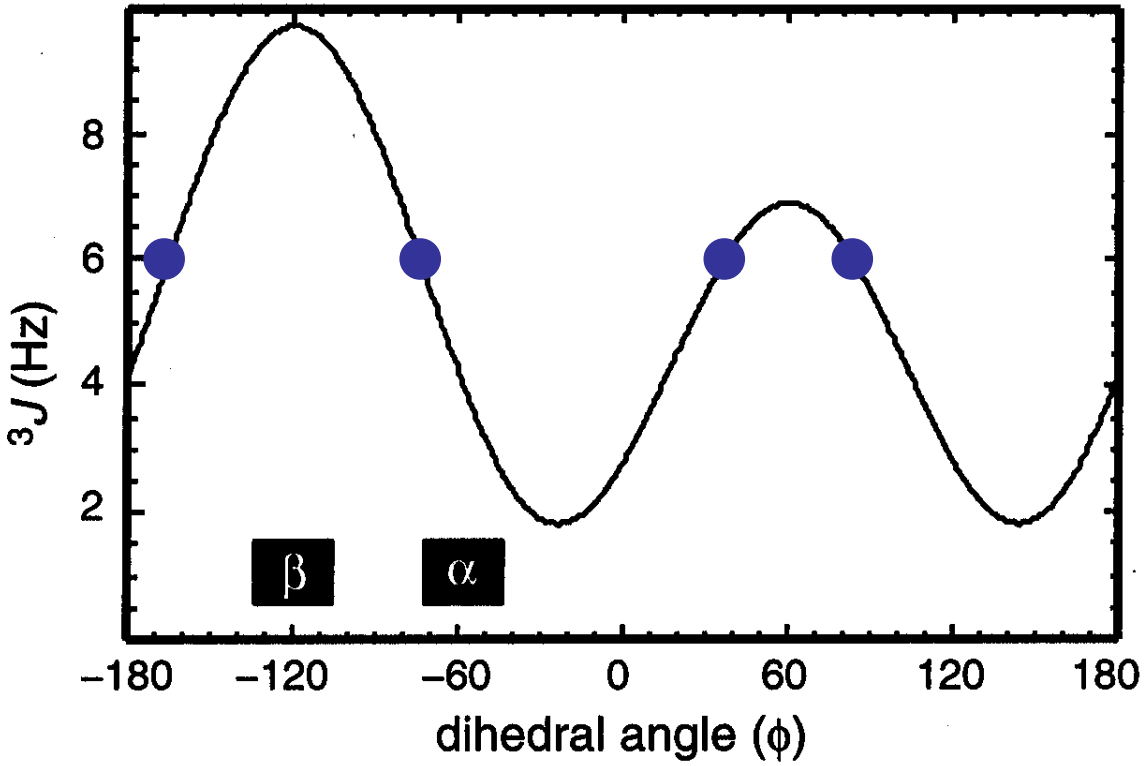
C=O·····H-N

- Slow rate of exchange of labile H^N with solvent
- Protein dissolved in ^2H_2O; H^N signals disappear with time
- H^N groups that are H-bonded (i.e. part of secondary structure) will exchange a lot slower than those in loops



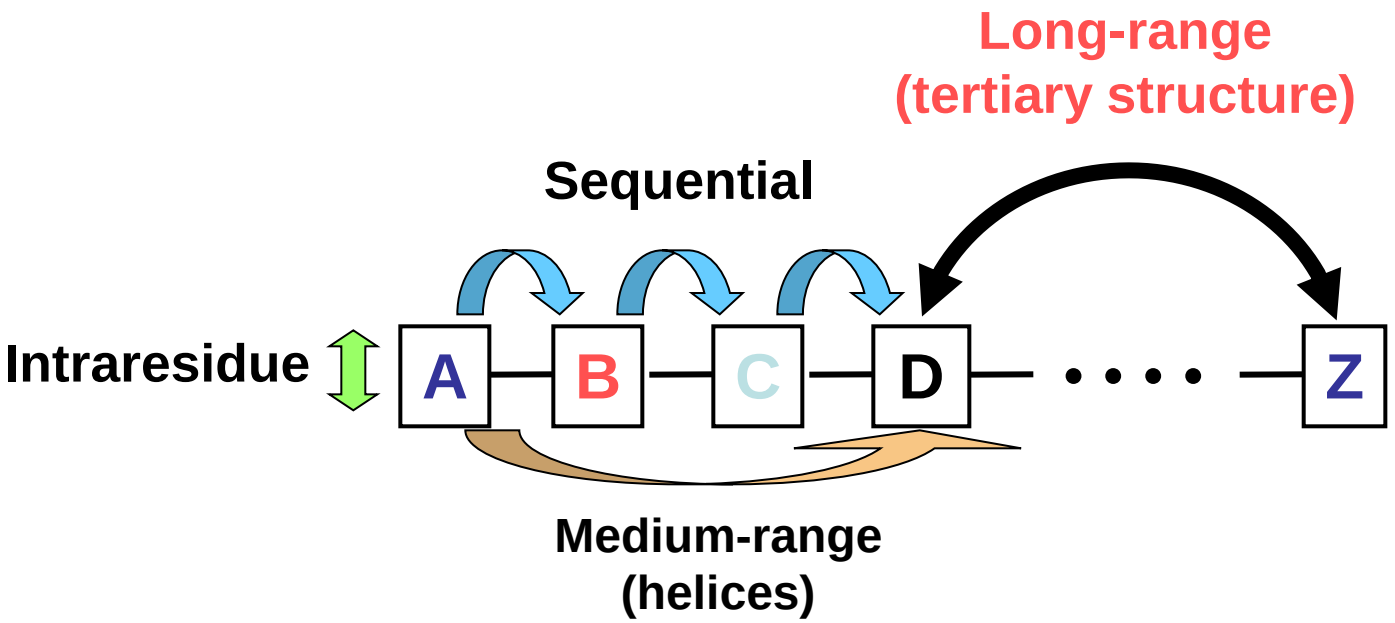
3. Dihedral Angles from Scalar Couplings





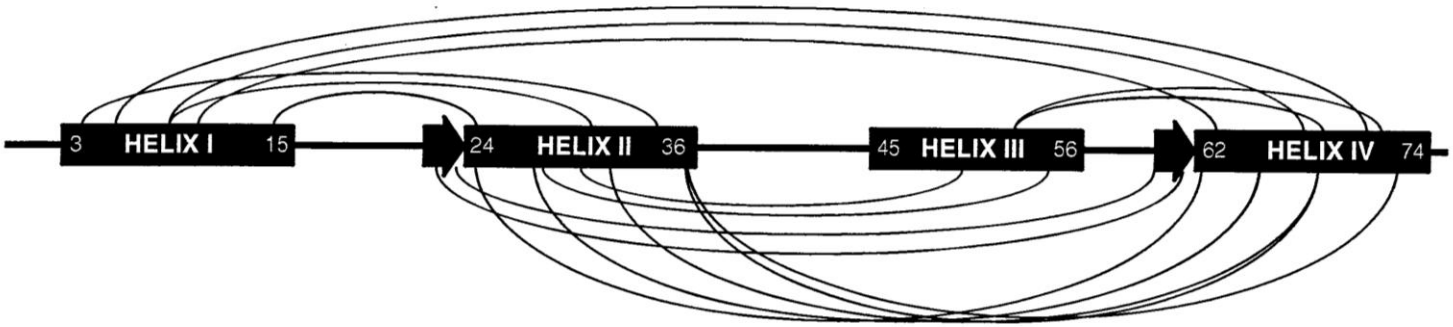
➢ *Must accommodate multiple solutions → multiple J values*

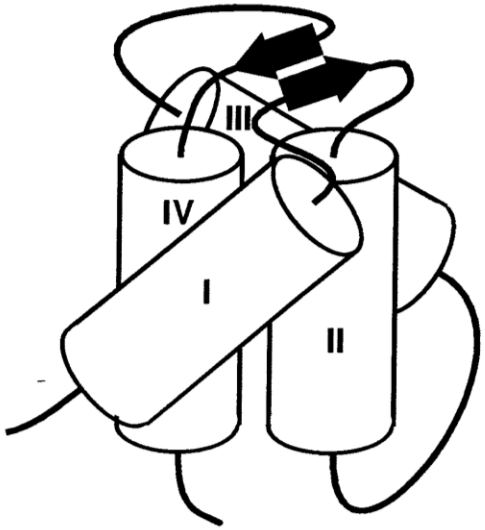
4. ^{1}H-^{1}H Distances from NOEs



Challenge is to assign all peaks in NOESY spectra - semi-automated processes for NOE assignment using NOESY data and table of chemical shifts yet still significant amount of human analysis

Protein Fold without Full Structure Calculations





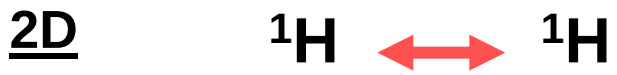
1. Determine secondary structure

- **CSI directly from assignments**
- **Medium-range NOEs**

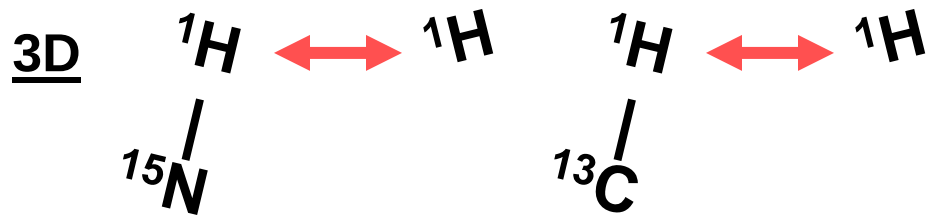
2. Add key long-range NOEs to fold

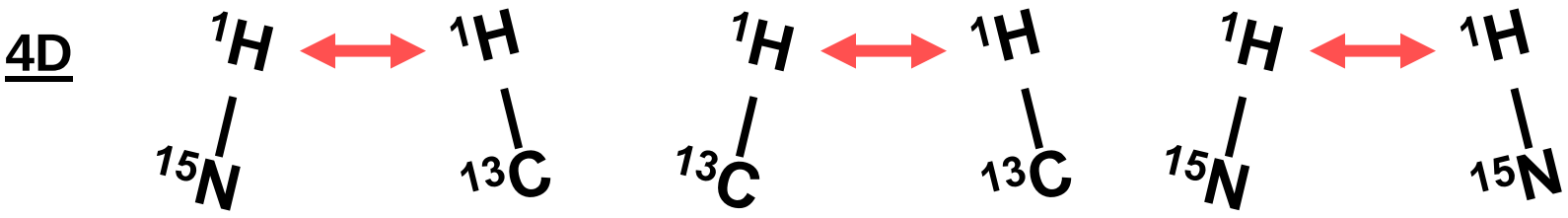
Approaches to Identifying NOEs

- ^{1}H-^{1}H NOESY



- ^{15}N- or ^{13}C dispersed ^{1}H-^{1}H NOESY





NMR Structure Calculations

Objective: Determine all conformations consistent with experimental data

• Programs that only do conformational search may lead to bad geometry → use simulations guided by experimental data
    • need a reasonable starting structure

•Distance restraints arrived at from NOE signal intensities → signal is an average of all conformations

NMR Structure Calculations (cont)

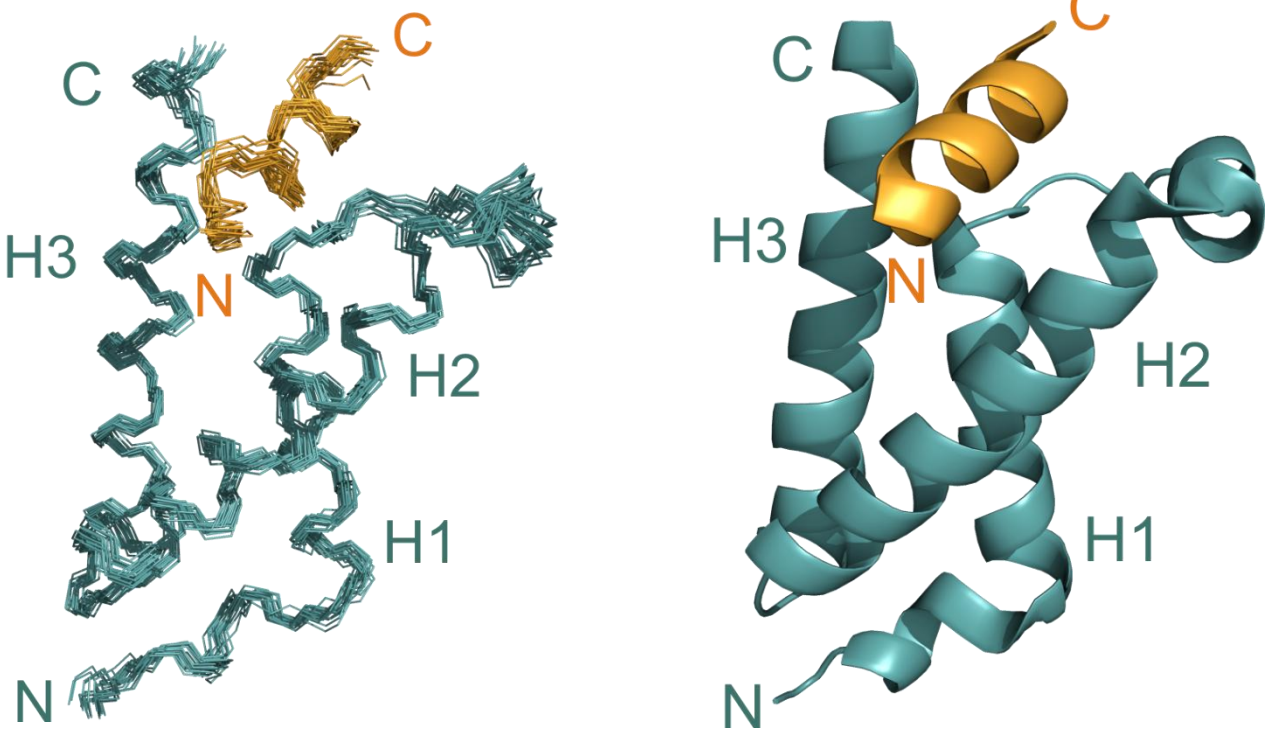
1. NOE signals are time & population-averaged (ie. measured on entire sample over period of time)

2. Intensity of NOE signal $\propto$ ^{1}H-^{1}H distance ($1/r^{6}$)

$\therefore$ NOE distance restraints are given a range of values
strong NOE: 0 - 2.8 Å
medium NOE: 2.8 – 3.5 Å
weak NOE: 3.5 – 5.0 Å

NMR data not perfect: Noise, incomplete data → multiple solutions (conformational ensemble unlike X-ray crystallography with one solution)

Variable Resolution of Structures



- Secondary structures well defined, loops variable
- Interiors well defined, surfaces more variable
- Trends the same for backbone and side chains
  - ➢ More dynamics at loops/surface
  - ➢ Constraints in all directions in the interior

Assessing the Quality of NMR Structures

- Number of experimental constraints
- RMSD of structural ensemble (subjective!)
- Violation of constraints- number, magnitude
- Molecular energies
- Comparison to known structures: ***PROCHECK***
- Back-calculation of experimental parameters

Summary of Protein NMR Structure Determination

Sample preparation with possible isotope labeling

↓

Data collection (scalar coupling and dipolar coupling expts.

↓

Resonance and sequence-specific assignments

↓

Identification and quantification of NOE peaks and intensities and conversion to approx. ^{1}H-^{1}H distances

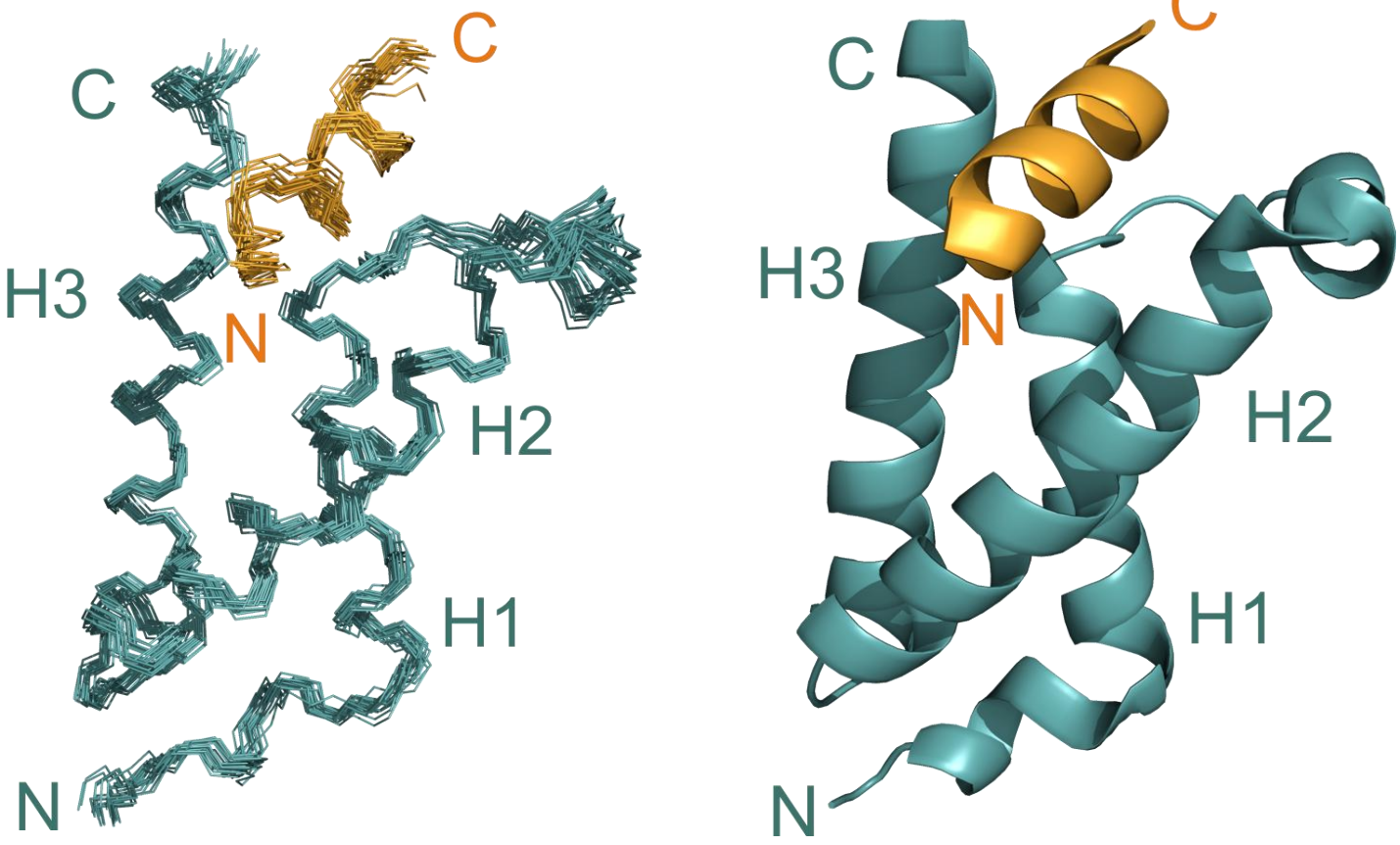
↓

Generation of models consistent with NOE distance constraints, dihedral angle ranges, H-bond distances

↓

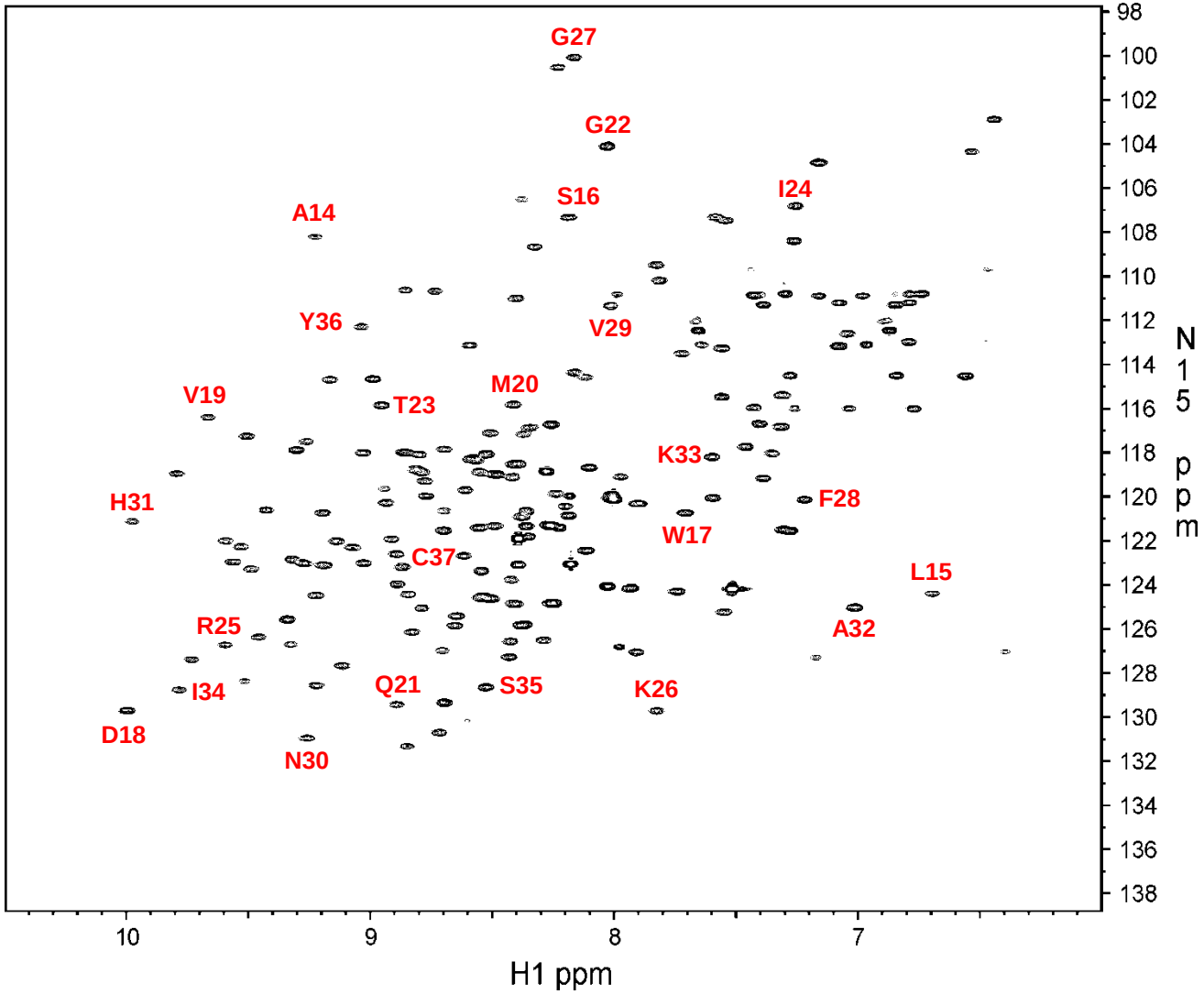
Model improvement by inclusion of newly identified NOES using above mentioned models

NMR Structures – Now what?



Monitoring Molecular Interactions

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

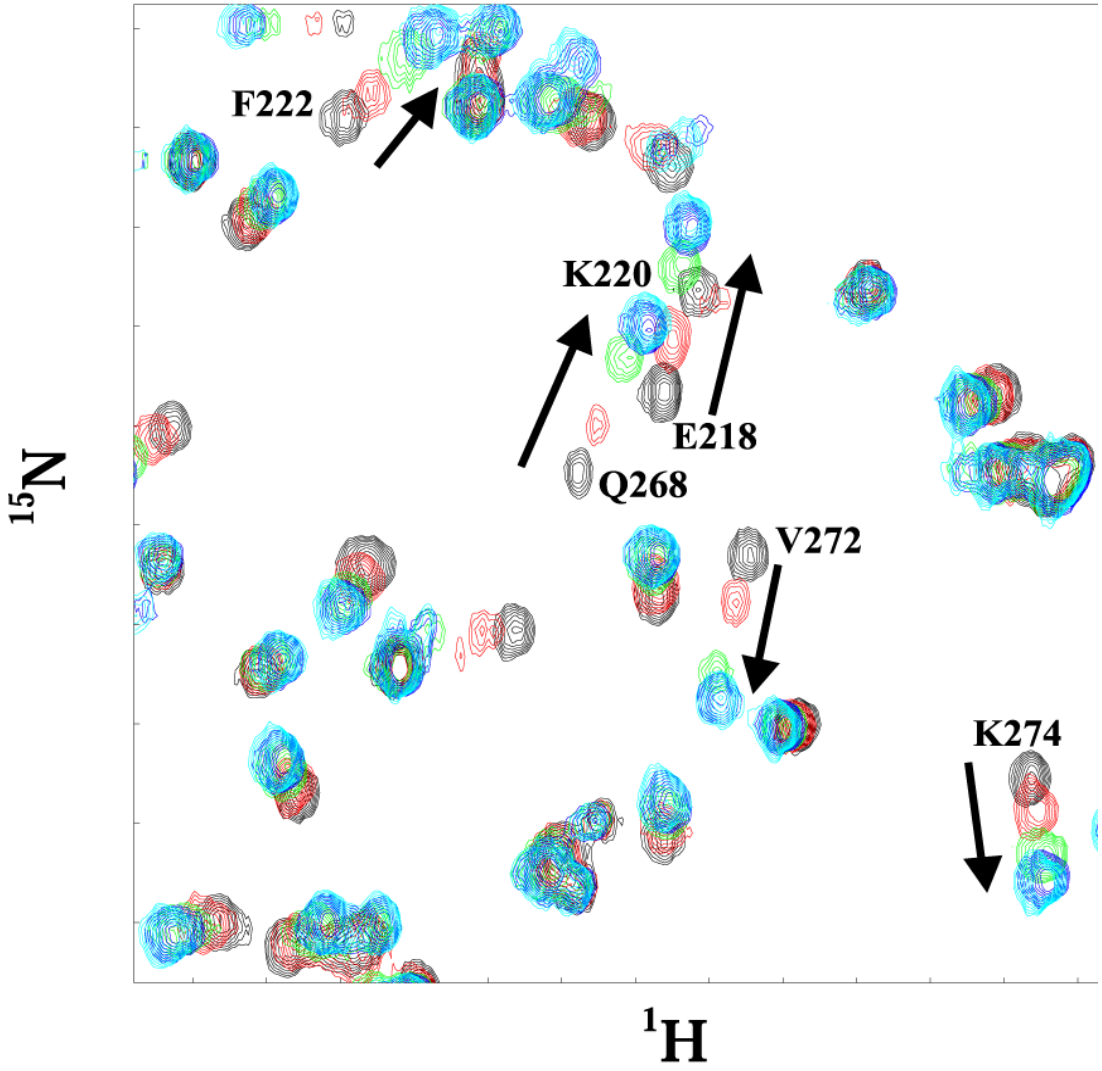


NMR Provides

- **Site-specific**
- **Multiple probes**
- **In-depth info**
- **Spatial distribution of responses can be mapped on structure**

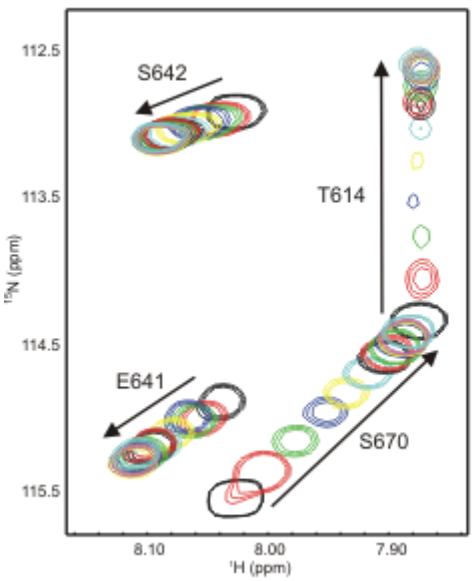
Monitoring Molecular Interactions

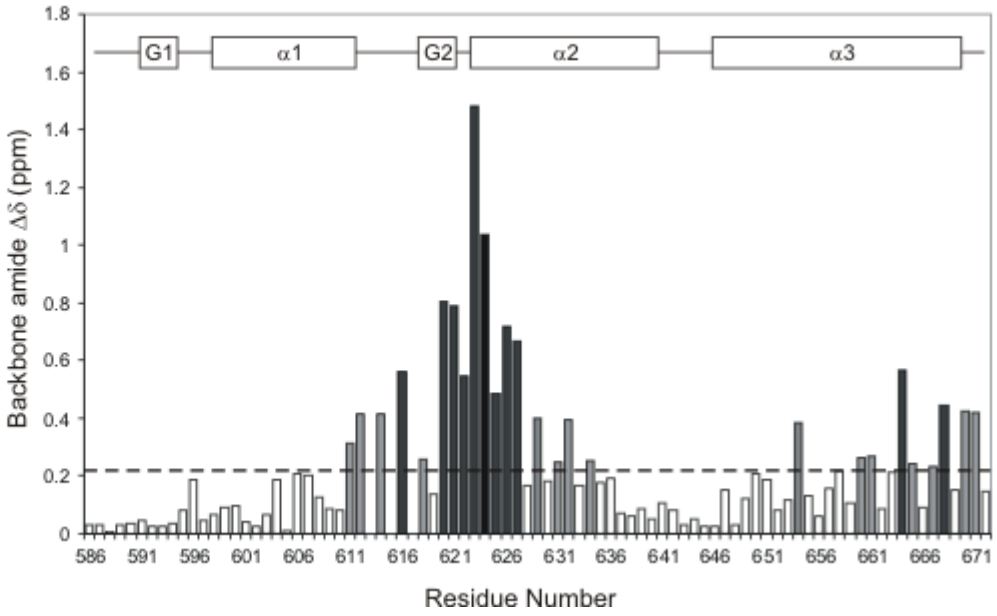
Titration followed by ^{15}N-^{1}H HSQC



Monitoring Molecular Interactions

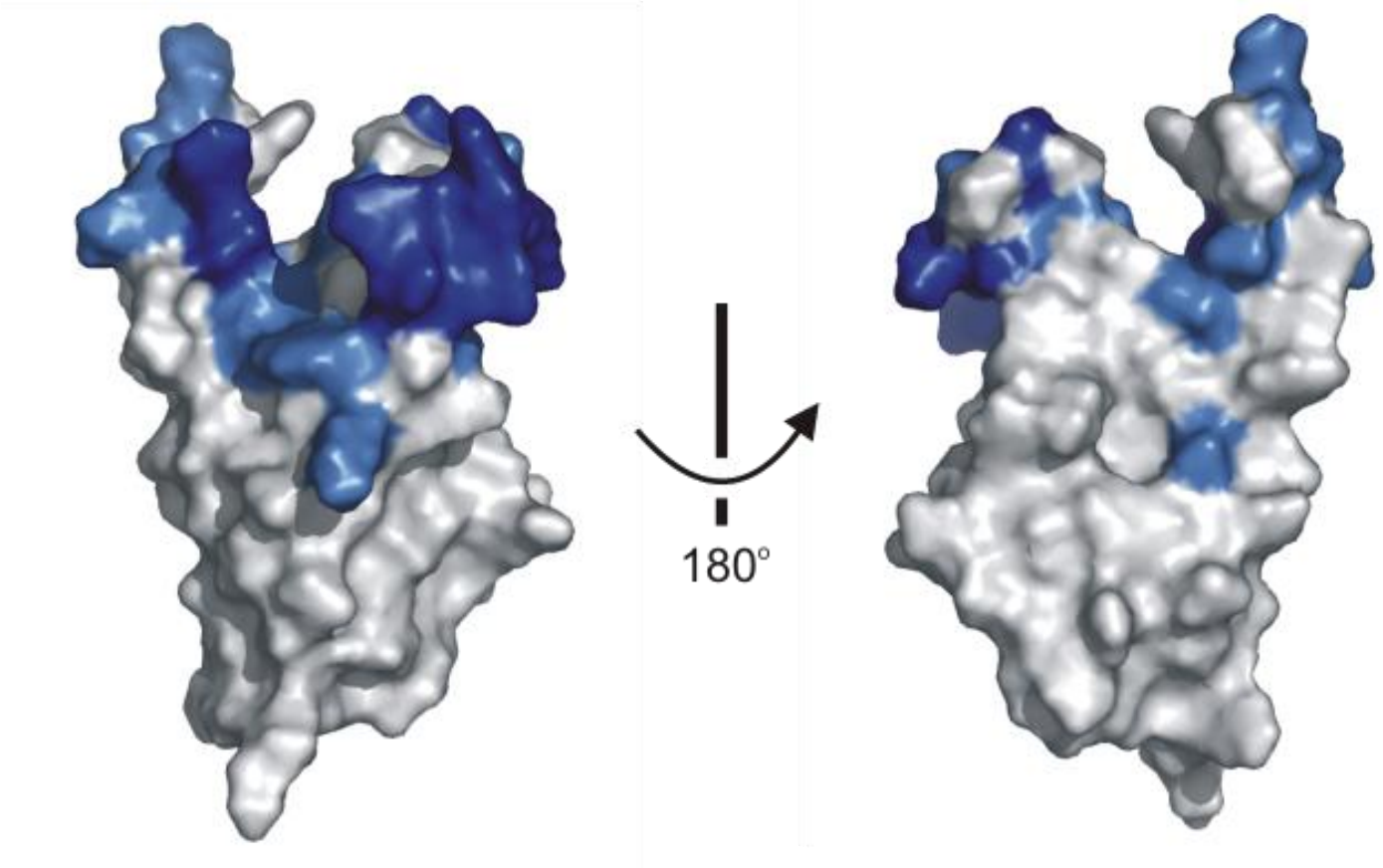
Transcription factor (CBP) -oncoprotein (E2A) interaction





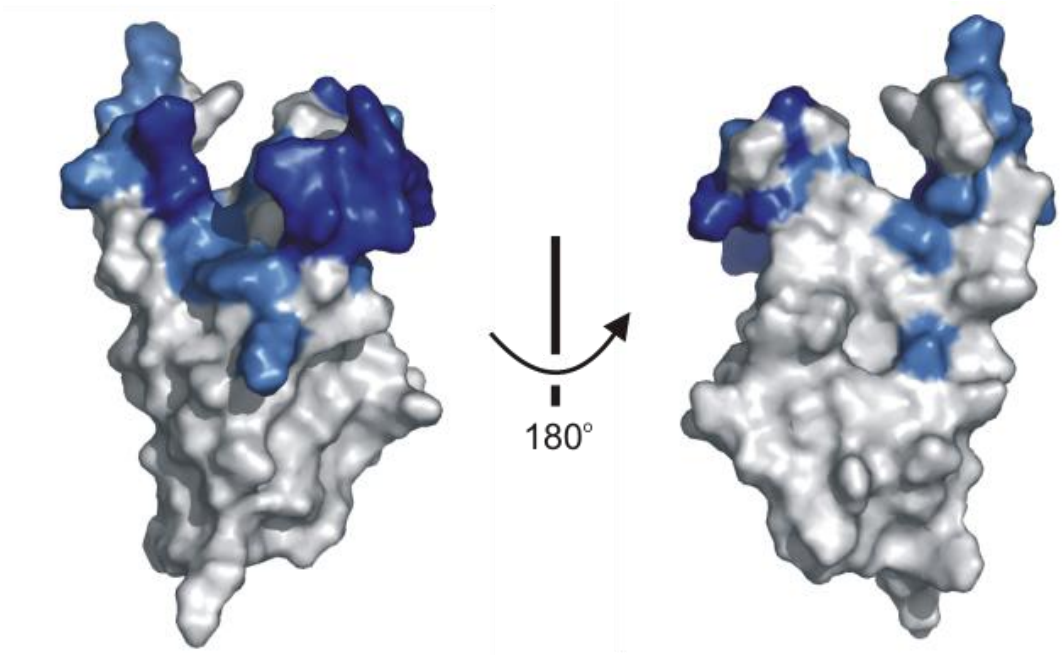
Map of chemical shift perturbations on the structure of protein?

Monitoring Molecular Interactions

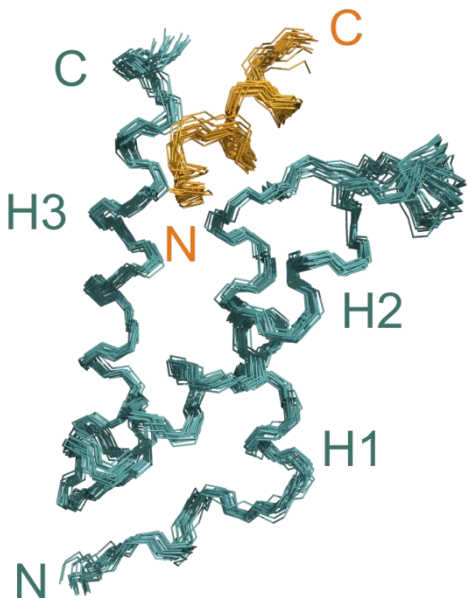


- Identification of ligand (E2A)-binding site on the structure of the KIX domain of CBP

Monitoring Molecular Interactions



Chemical Perturbation Mapping



Structure

Ligand Binding

NMR timescale – 1 sec to 1 x 10^{-6} sec

$1/k_{off}$ = t >> 1 sec $\Rightarrow$ slow exchange, superposition of spectra

$1/k_{off}$ = t << 1 x 10^{-6} sec $\Rightarrow$ fast exchange, weighted average

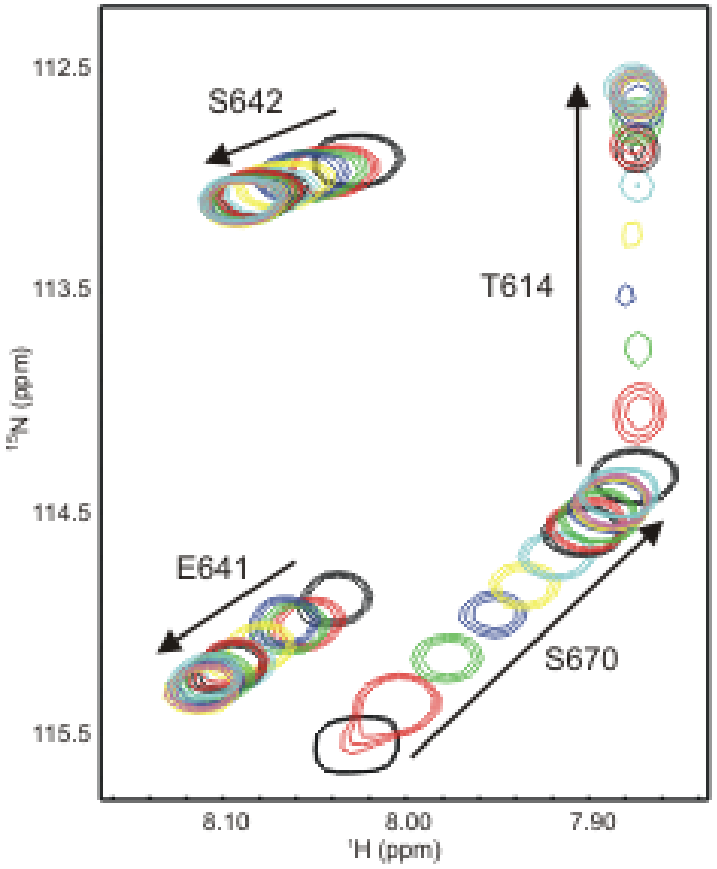
$$A \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} B \qquad K_{diss} = [A]/[B] = k_{off}/k_{on}$$

Ligand Binding

- Another protein
- Metal ion
- Drug or chemical

$$P + L = PL \qquad K_{diss} = \frac{[P]\,[L]}{[PL]}$$

Ligand Binding - exchange



<u>E641, S642, and S670</u>

- Fast exchange
(weighted average of free and bound populations)

<u>T614</u>

- Intermediate-fast exchange

Ligand Binding

$$P_{tot} = P + PL$$

$$L_{tot} = L + PL$$

$$\text{So……. } K_{diss} = \frac{[P_{tot} - PL]\,[L_{tot} - PL]}{[PL]}$$

Plot $[L_{tot}]/[P_{tot}]$ vs "*change*" in NMR spectra

For fast exchange (weak binding):

$$\text{Change} = \frac{\delta_{obs} - \delta_{init}}{\delta_{sat} - \delta_{init}} = \frac{[PL]}{[P_{tot}]}$$

shifting of resonances in spectra

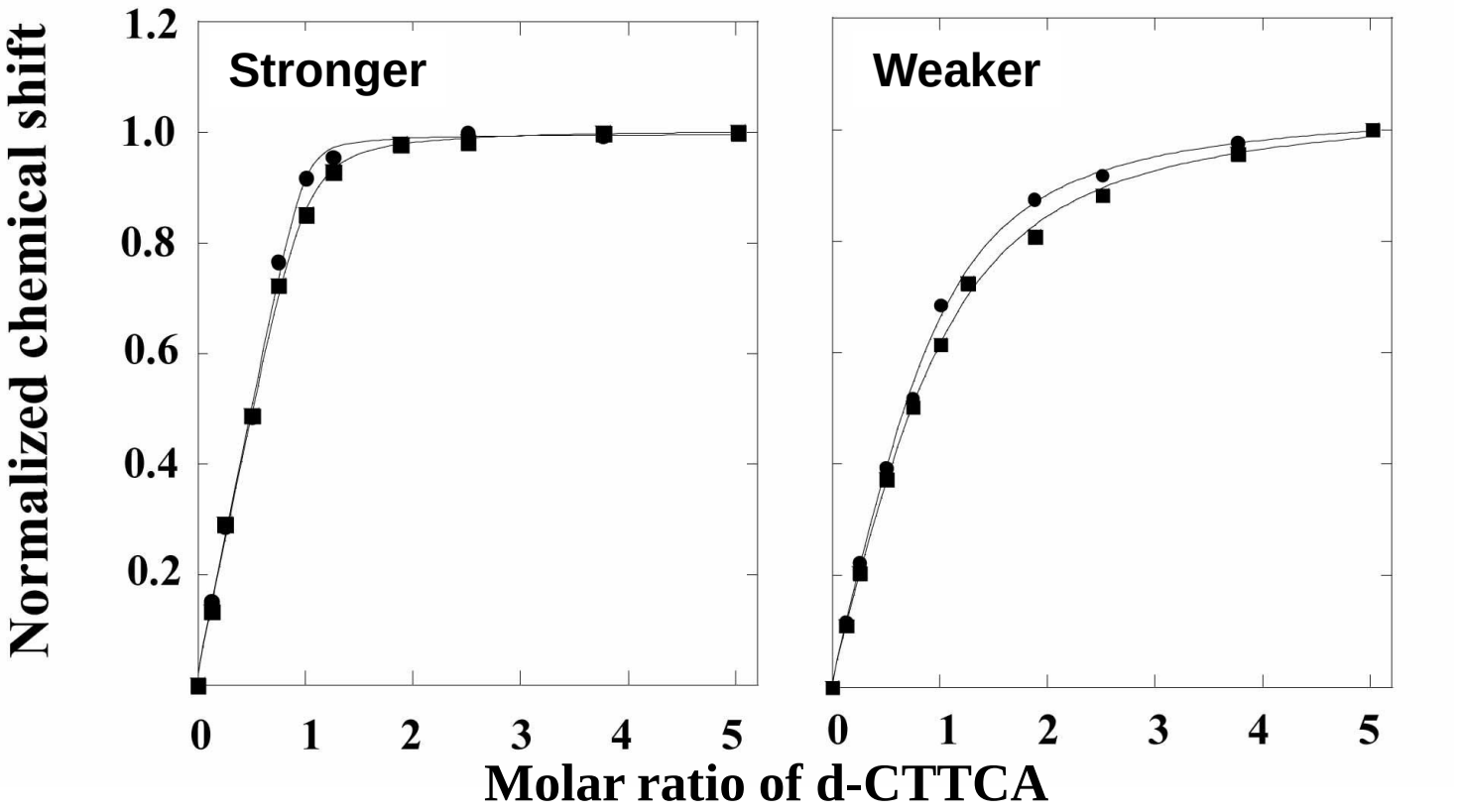
For slow exchange (tight binding):

$$\text{Change} = \frac{\text{Integral of peak}_{obs}}{\text{Integral of peak}_{max}} = \frac{[PL]}{[P_{tot}]}$$

intensity changes in peaks of free and bound forms

Monitoring Molecular Interactions

Binding Constants by NMR



Fit change in chemical shift to binding equation

NMR and Crystallography

NMR	***X-ray***
• Can mimic biological conditions - pH, temp, salt	• Highly automated with more objective interpretation of data
• information on dynamics	• Quality indicators (resolution, *R*)
• monitor conformational change on ligand binding	• Surface residues and water molecules well defined
• 2° structure derived from limited experimental data	• Huge molecules and assemblies can be determined
• need concentrated sample - lots of protein; aggregation issues	• non-physiological conditions – crystallization difficult
• size limited – ~40kDa for full structure determination	• need heavy-atom derivatives – production not always trivial
• more subjective interpretation of data	• snap-shot of protein in time – less indication of mobility
• lack of quality factors - resolution and *R*-factor	• flexible proteins difficult to crystallize