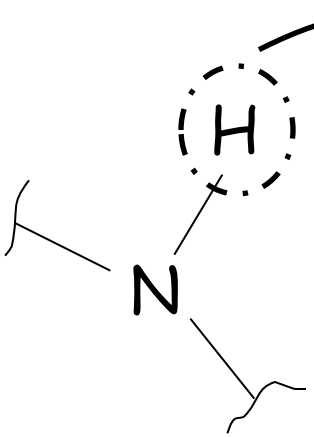
Using NMR to study Macromolecular Interactions

John Gross, BP204A
UCSF
presented by Stephen Floor

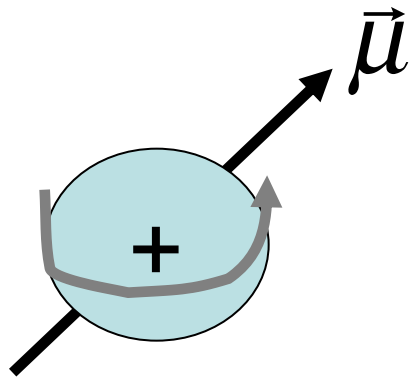
Outline

- Multidimensional NMR
- Macromolecular Interactions
- Dynamics
- Dealing with large complexes
- Structure Determination

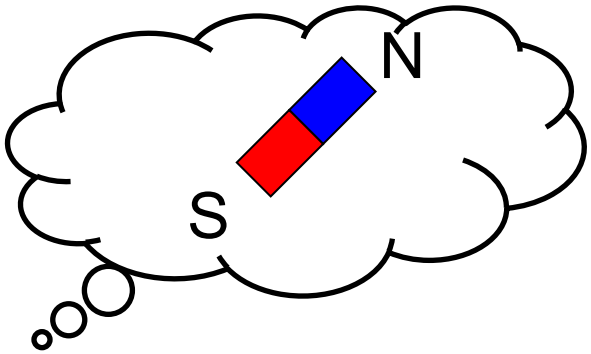
Review: Nuclear Spins are Microscopic Bar Magnets



Protein Fragment



Magnetic Moment



Bar Magnet

Magnetic moment $\vec{\mu} = \gamma \vec{S}$ Angular Momentum

The proportionality constant γ: strength of bar magnet

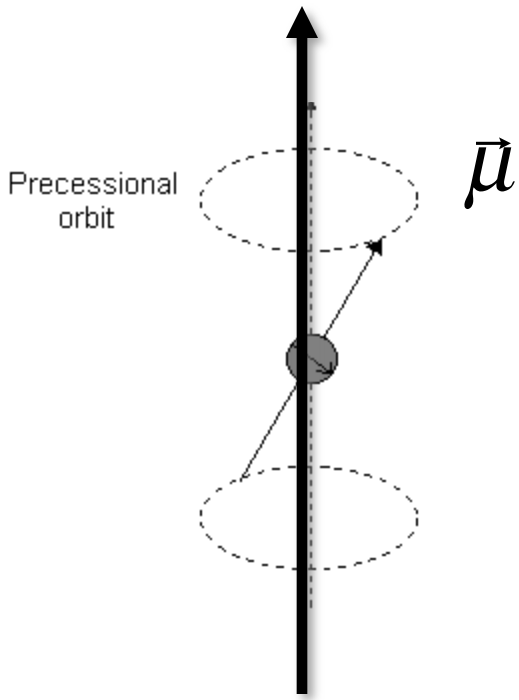
Equation of Motion

$$\frac{d\vec{\mu}}{dt} = \gamma \vec{B} \text{ x } \vec{\mu}$$

Based on magnetic torque:

$$\frac{d\vec{L}}{dt} = \vec{B} \times \vec{L}$$

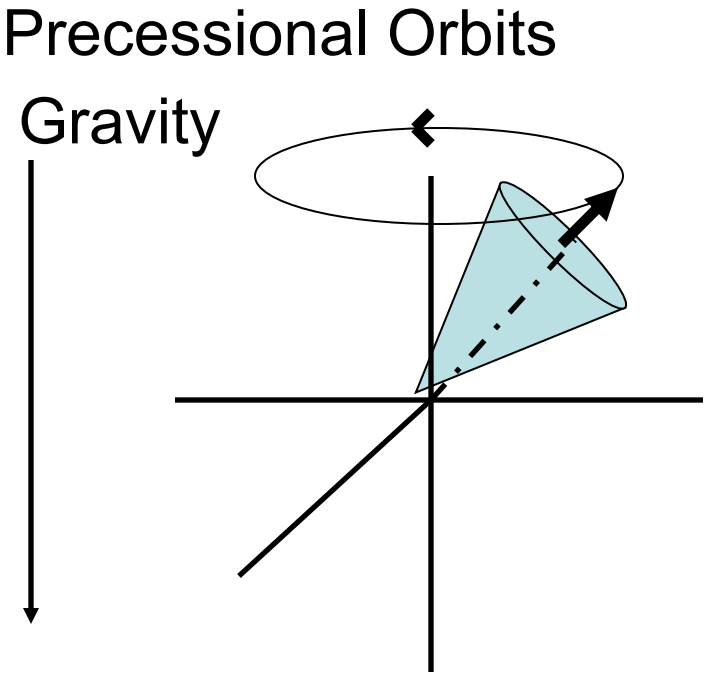
Spin Precession



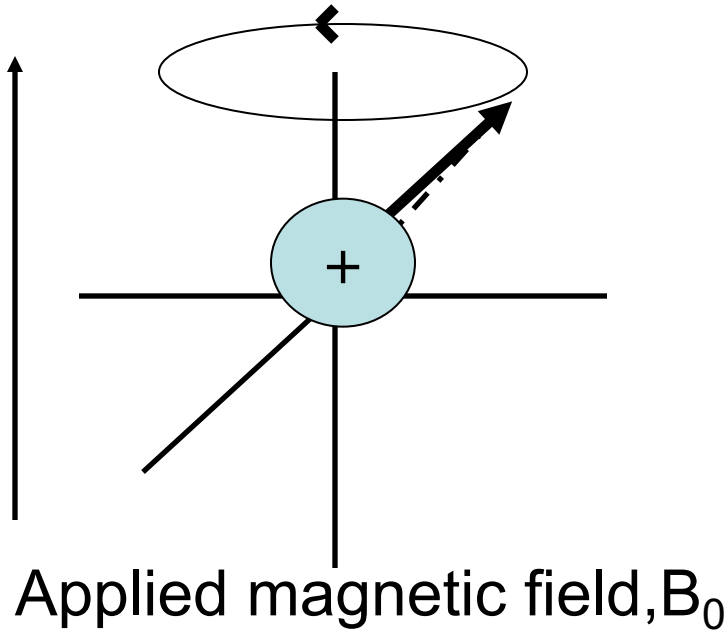
Magnetic Field, Bo

Precession frequency: $\gamma B_0 = \omega_0$

Driving Forces for Precession

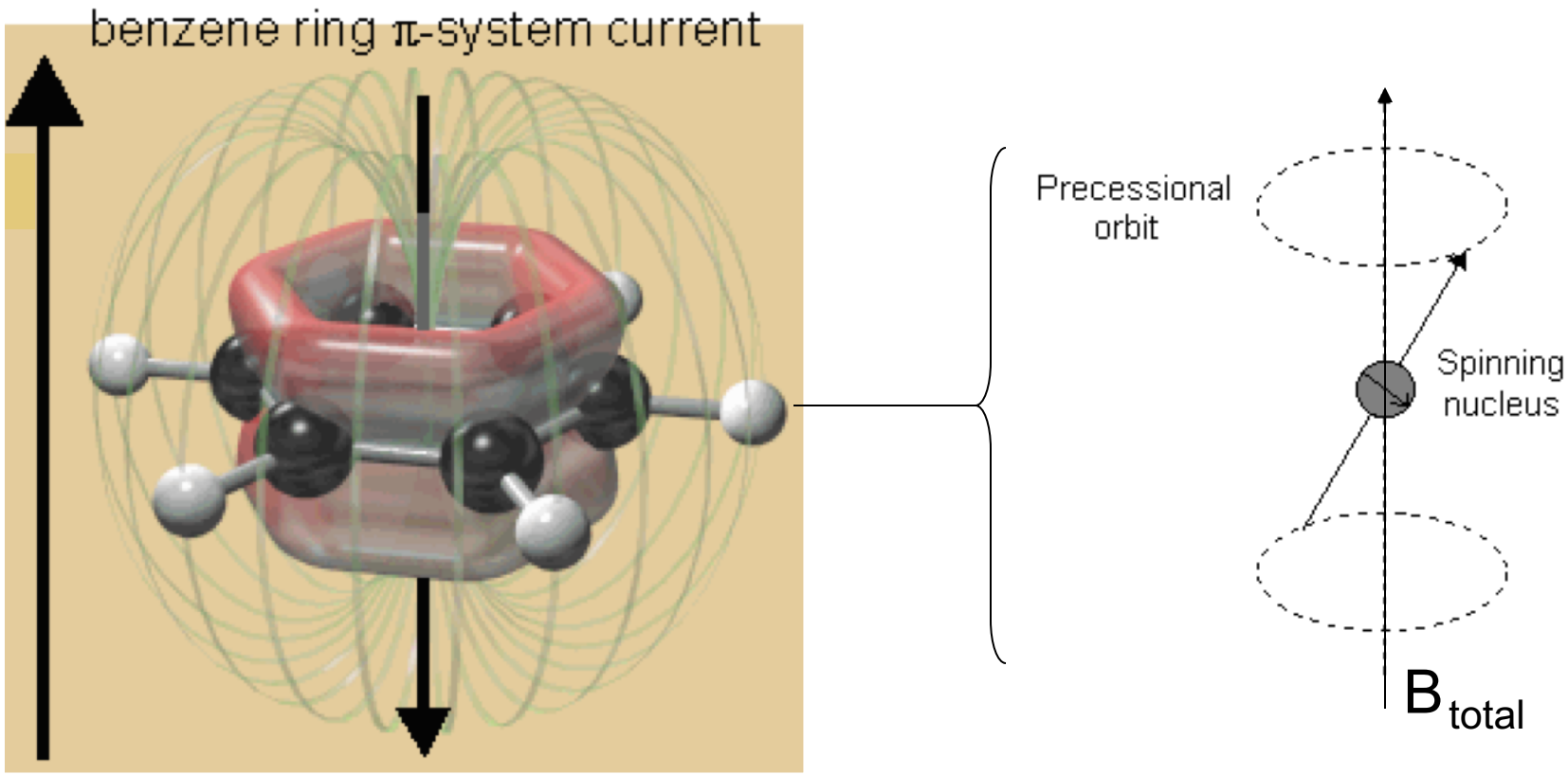


Spinning Top



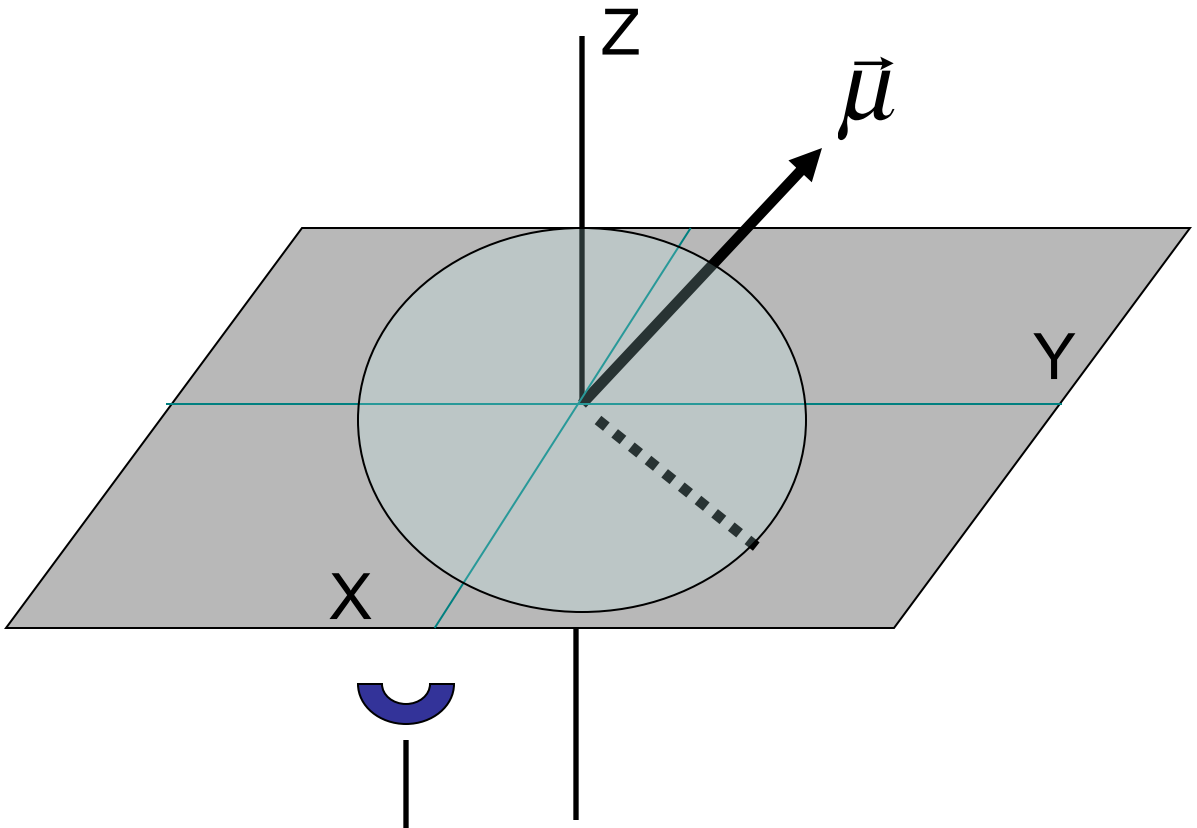
Spinning Nucleus

Nuclear Spins Report Local Environment

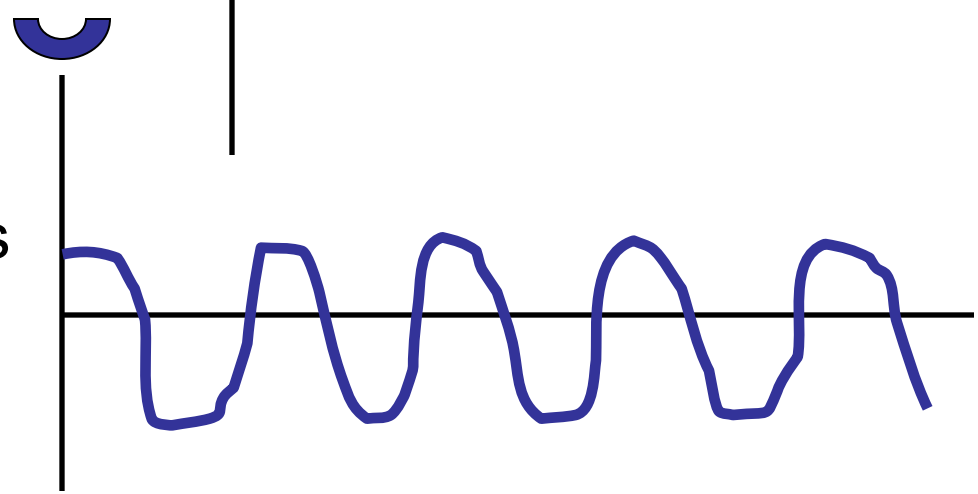


$B_{applied}$ + B_{local} = B_{total} **determines precession**

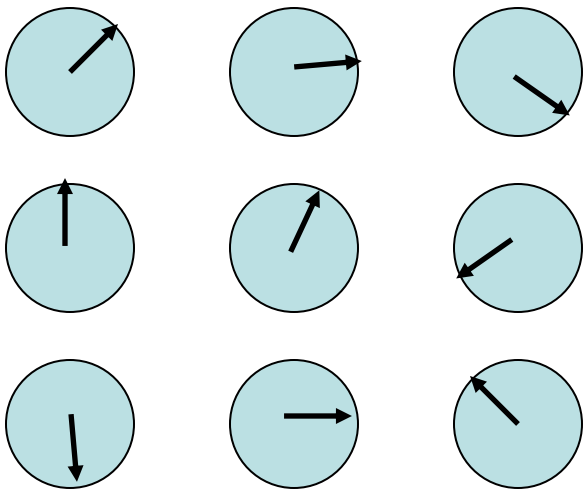
Detection of Spin Precession



Detector measures magnetic field on x-axis



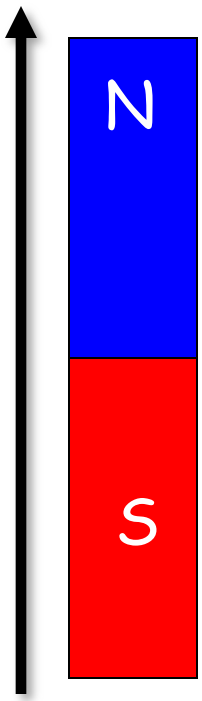
Net Magnetization

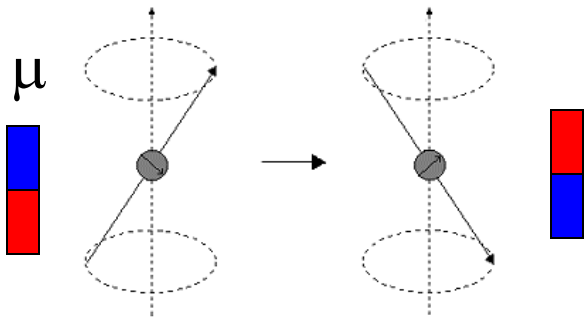


$$\left.\begin{aligned} \mathrm{M_x} &= \sum_j \mu_{\mathrm{x}}^j = 0 \\ \mathrm{M_y} &= \sum_j \mu_y^j = 0 \end{aligned}\right\} \text{No Transverse Magnetization at equilibrium}$$

Magnetic Energy

$$E = -\vec{\mu} \bullet \vec{B}$$

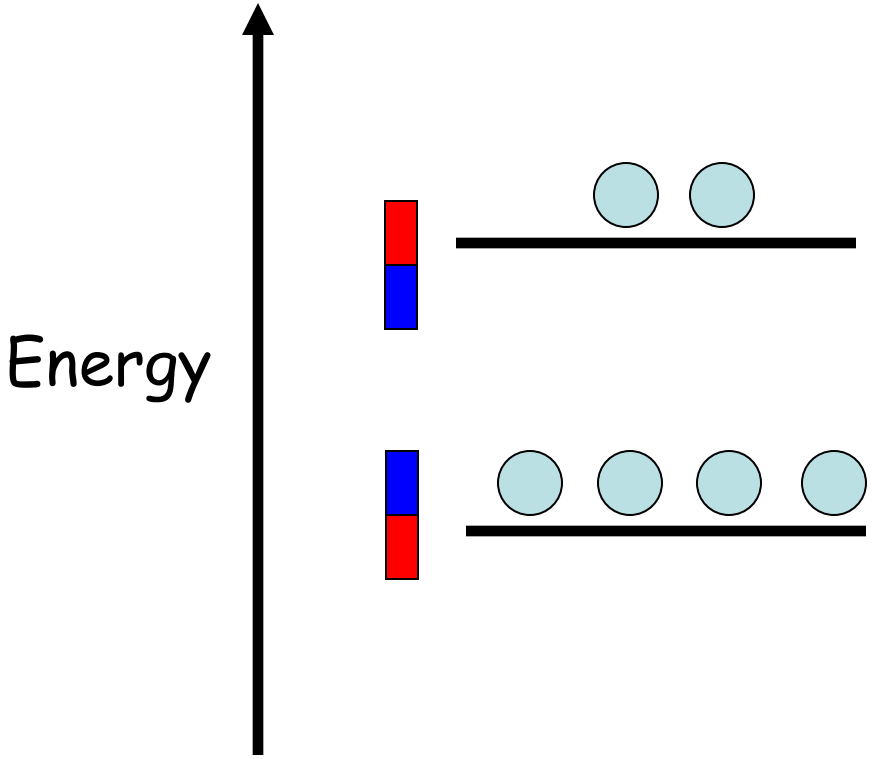




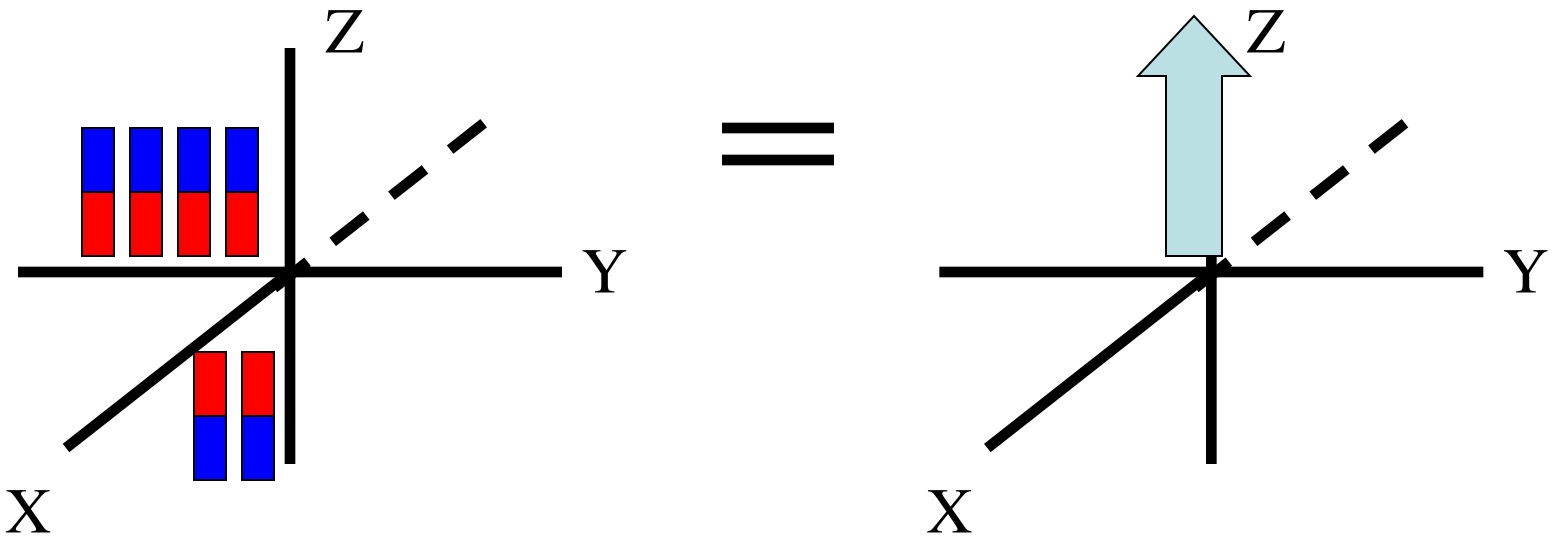
$$E = -\mu_z B_z$$

Static Magnetic Field
Oriented Along Z-Axis

Energy States (spin-1/2 nucleus)

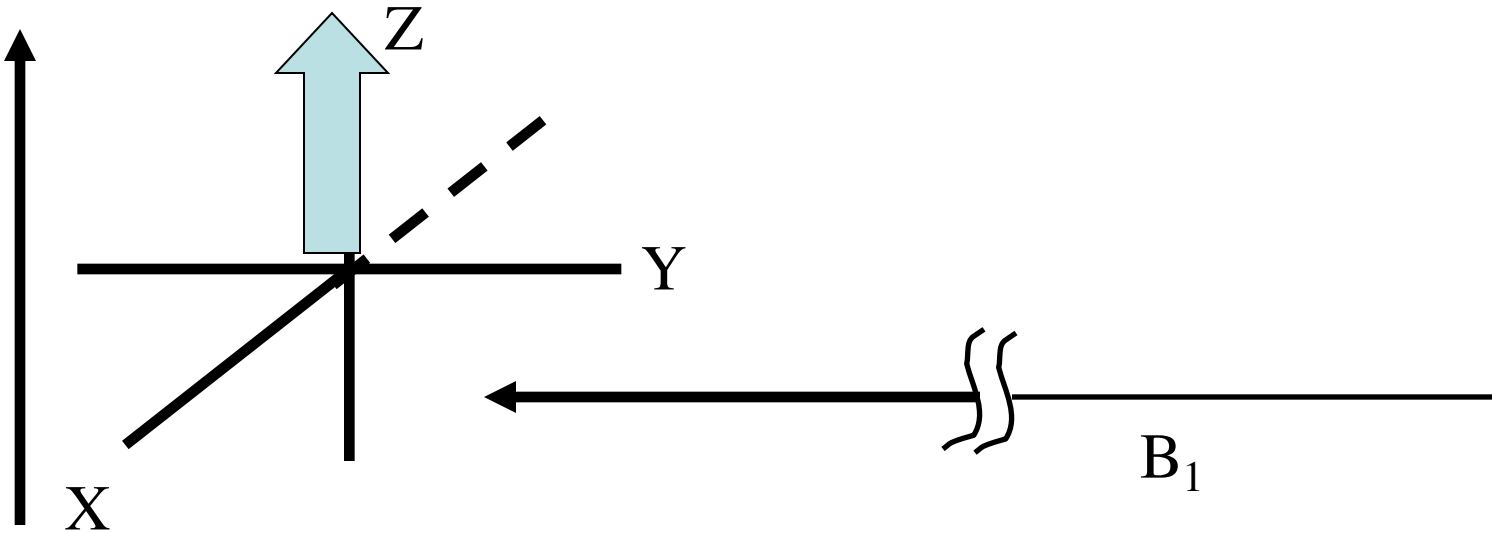


Net Magnetization along Z Axis



$$\sum_j \mu_z^j = M_z$$

Thought experiment: apply 2nd field along Y Axis

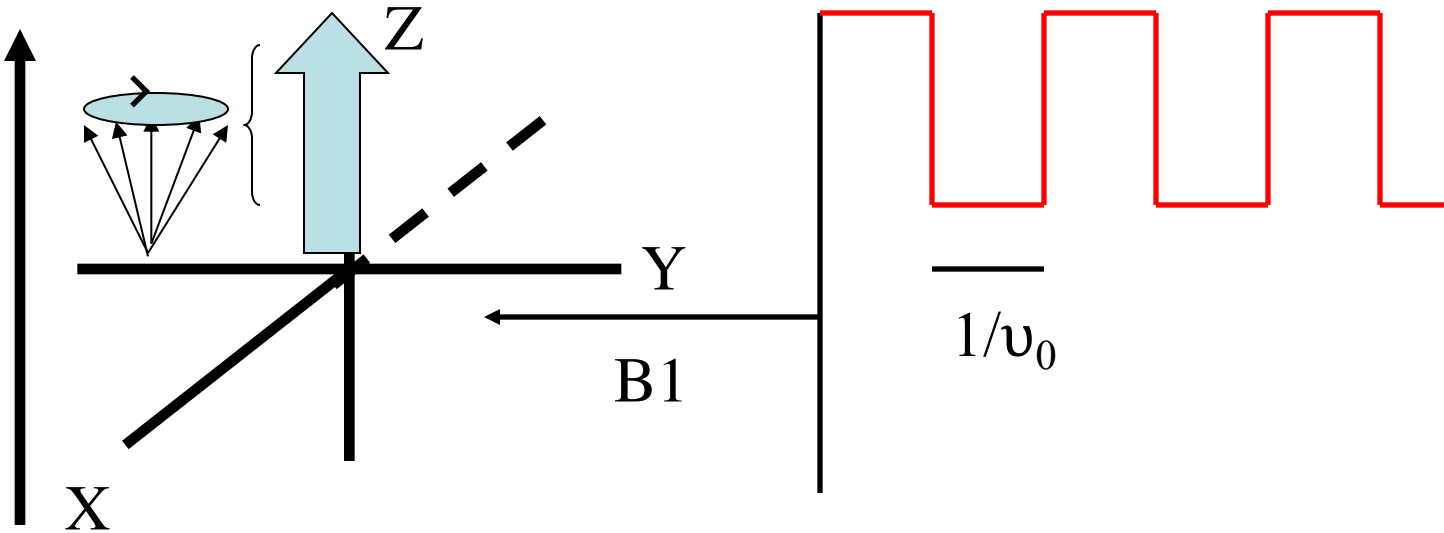


Bo

If $B_1 >> B_0$, M_Z would rotate about B_1.
Leave B_1 on until X axis reached ----> transverse magnetization
Approach is not practical.

Same effect achieved with weak, resonant oscillating field

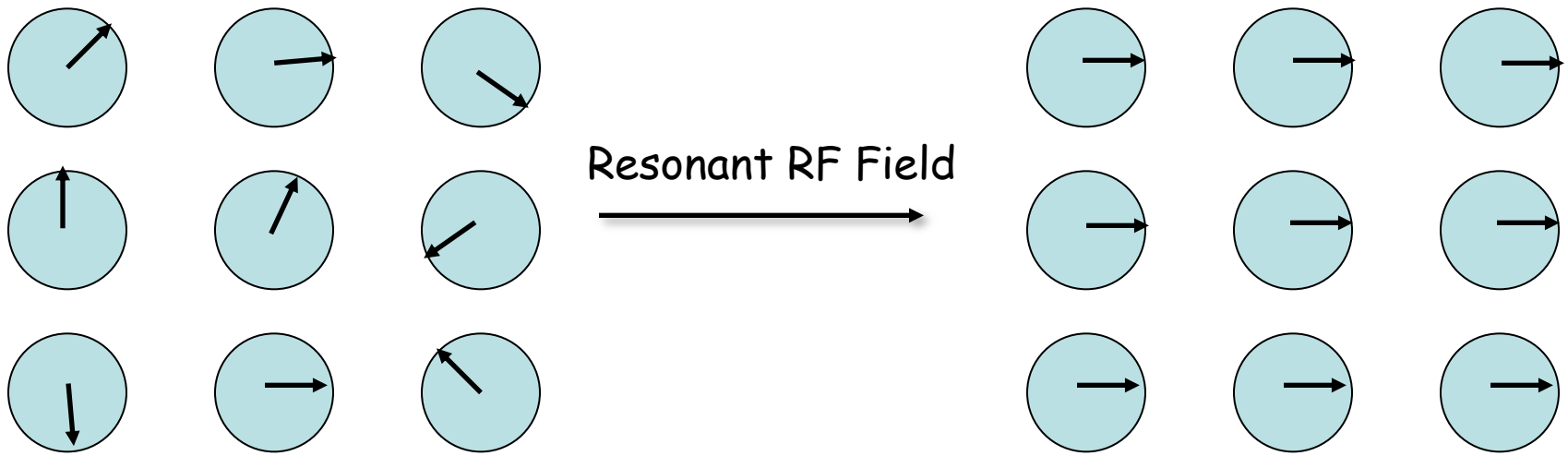
Bo



Turn B_1 on and off with a frequency matching the precessional frequency

Resonance

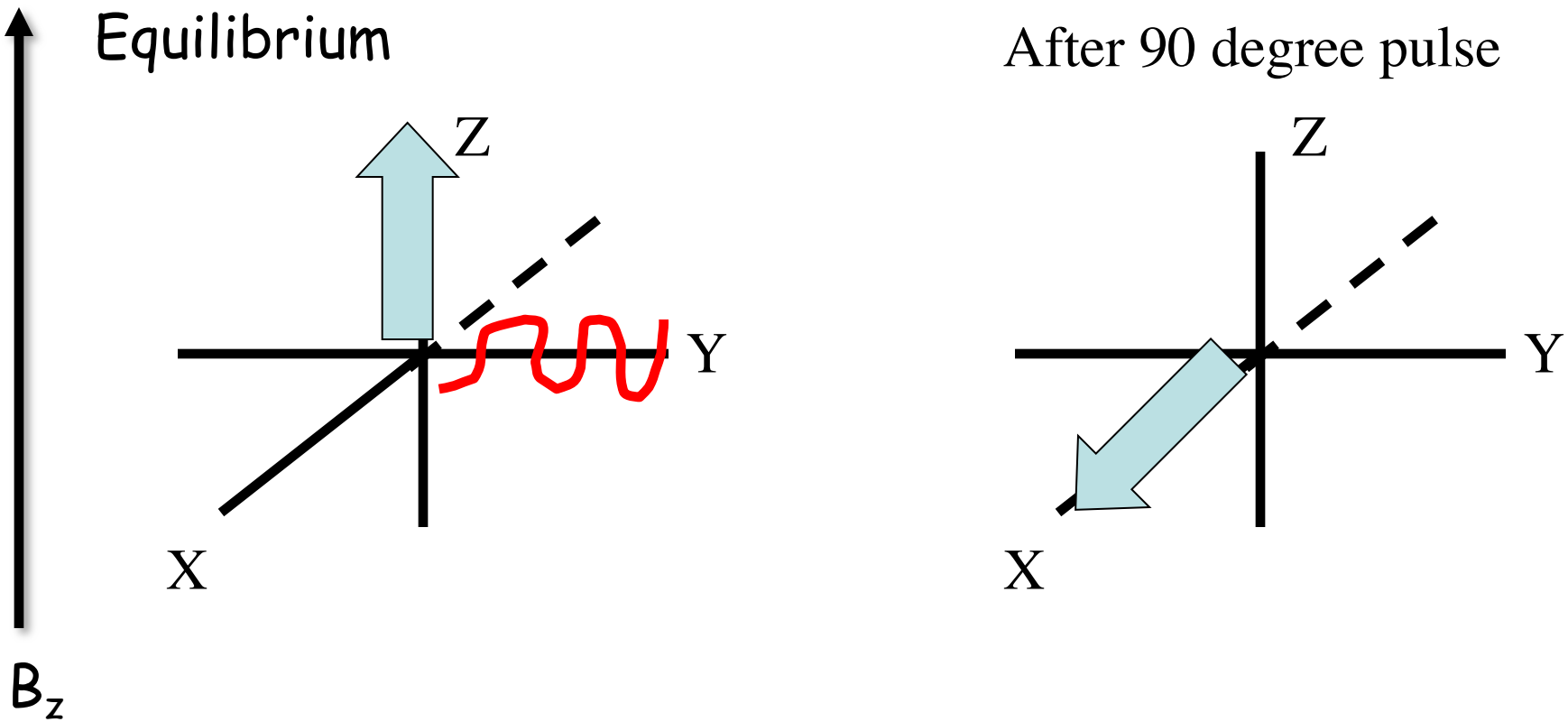
Ensemble of Nuclear Spins



Random Phase
No NMR Signal

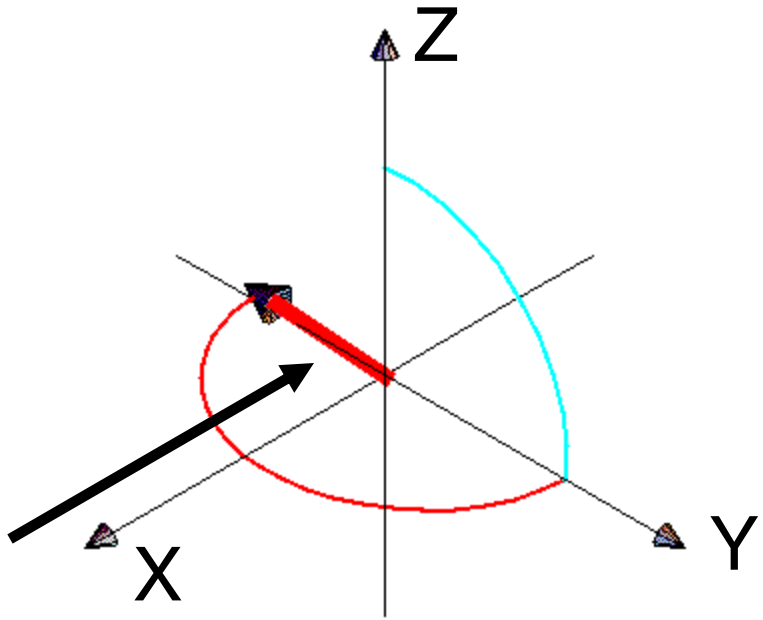
Phase Synchronization
NMR Signal!

Magnetization Vector Model



90y: Resonant 90 Degree Pulse

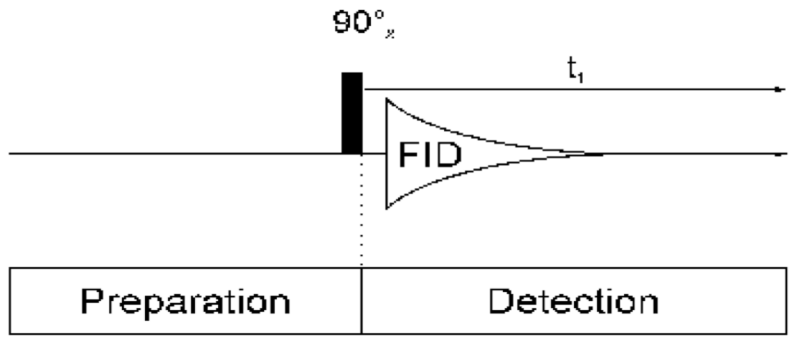
Resonant Pulse in Real Time



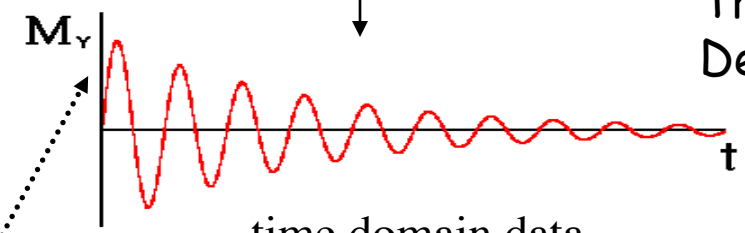
R.F. Field (applied at precession frequency)

Net magnetization rotated into transverse plane
Rotates due to static and local fields

Summary of 1D Experiment



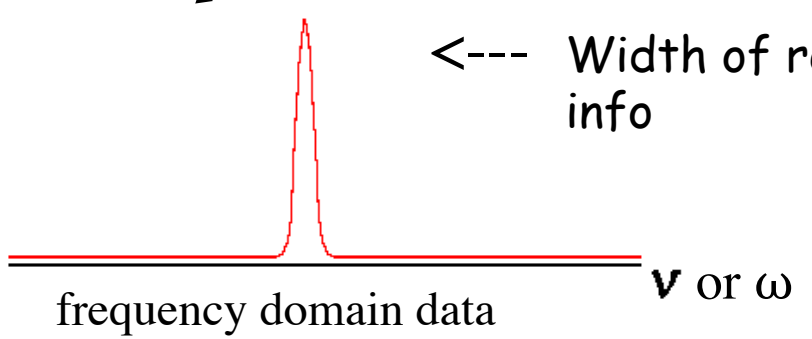
Transverse Magnetization
Decay constant T2



time domain data

Amplitude propotional
to amount magnetization
prior to pulse

Fourier Transform (FT)



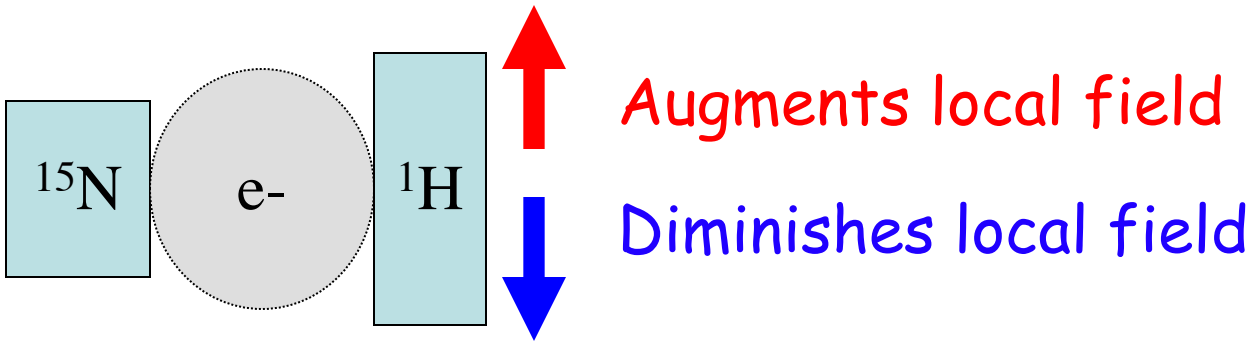
<--- Width of resonance--dynamical
info

Position of resonance --->
local magnetic environment

frequency domain data

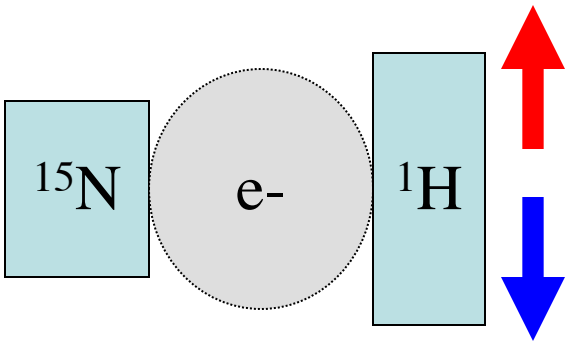
The J Coupling

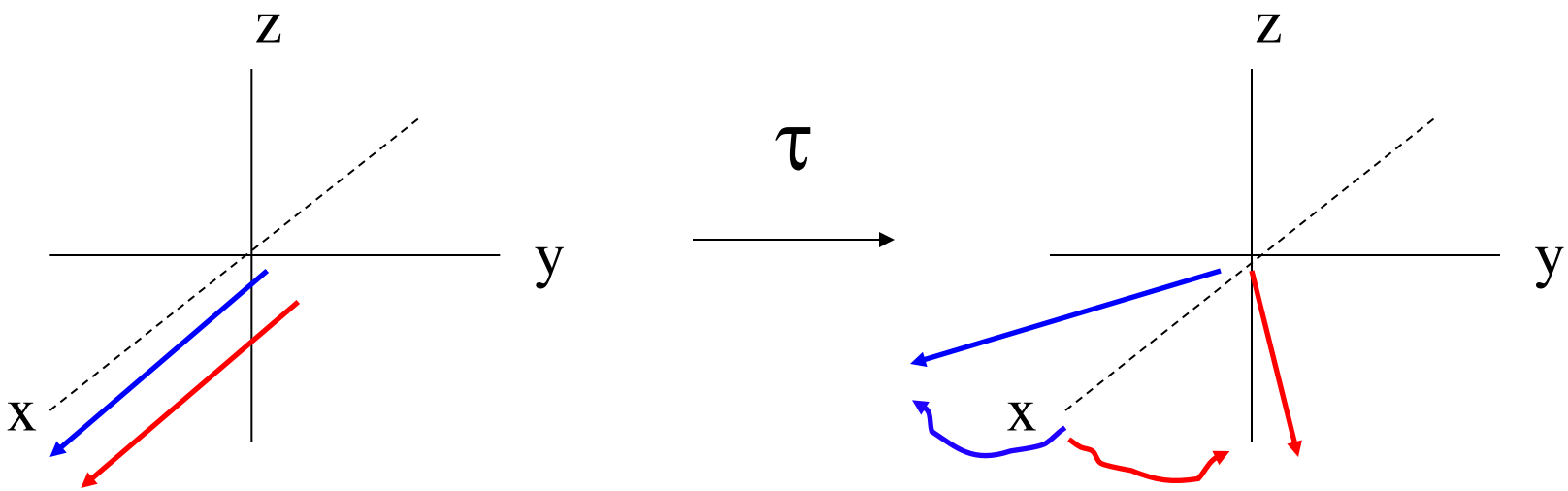
Consider two spin-1/2 nuclei (ie, ^{1}H and ^{15}N):



Effect transmitted through electrons in intervening bonds

Vector View

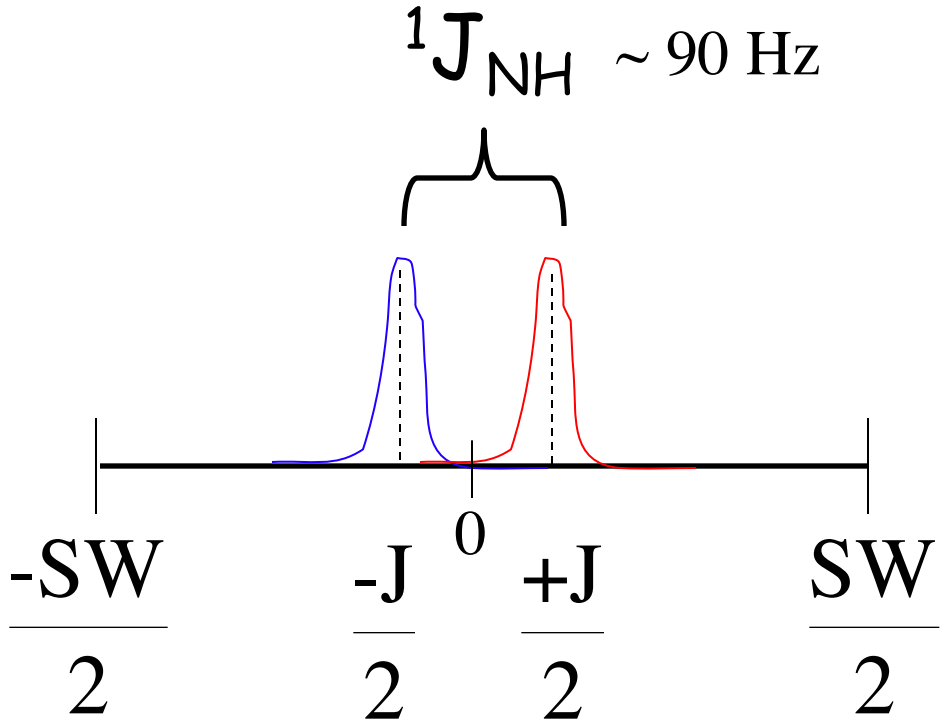




(After 90y pulse)

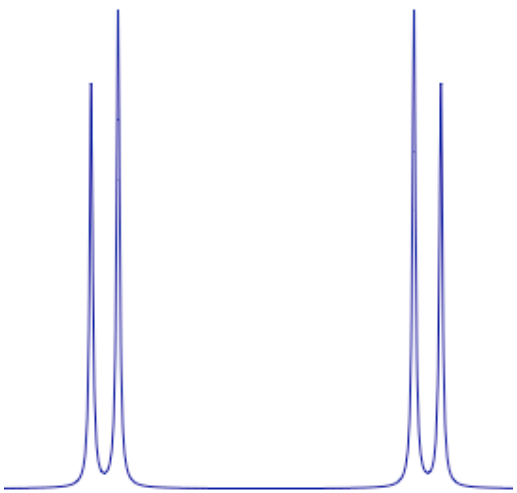
Components rotate faster or slower than rotating frame by +- J/2

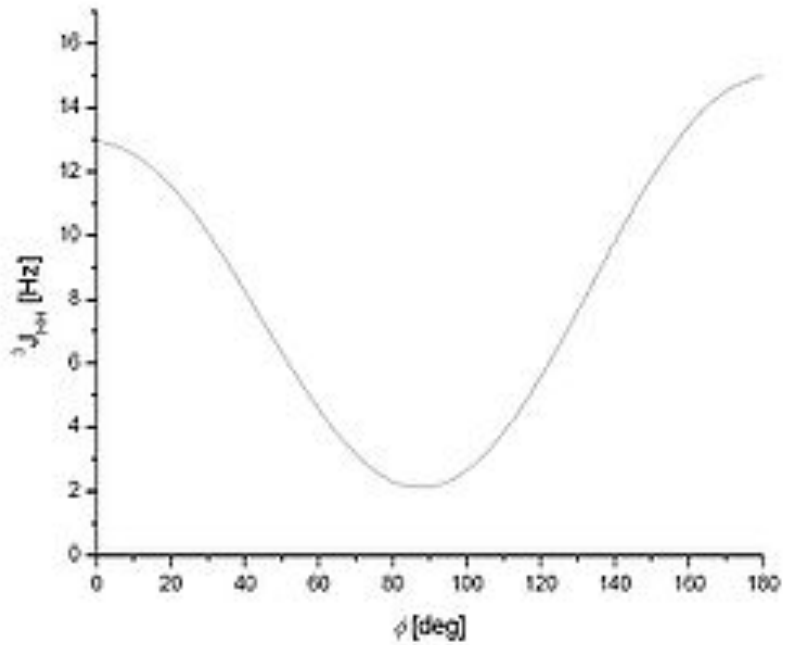
Spectrum with J coupling



^{15}N Detected Spectrum

J couplings contain information on structure





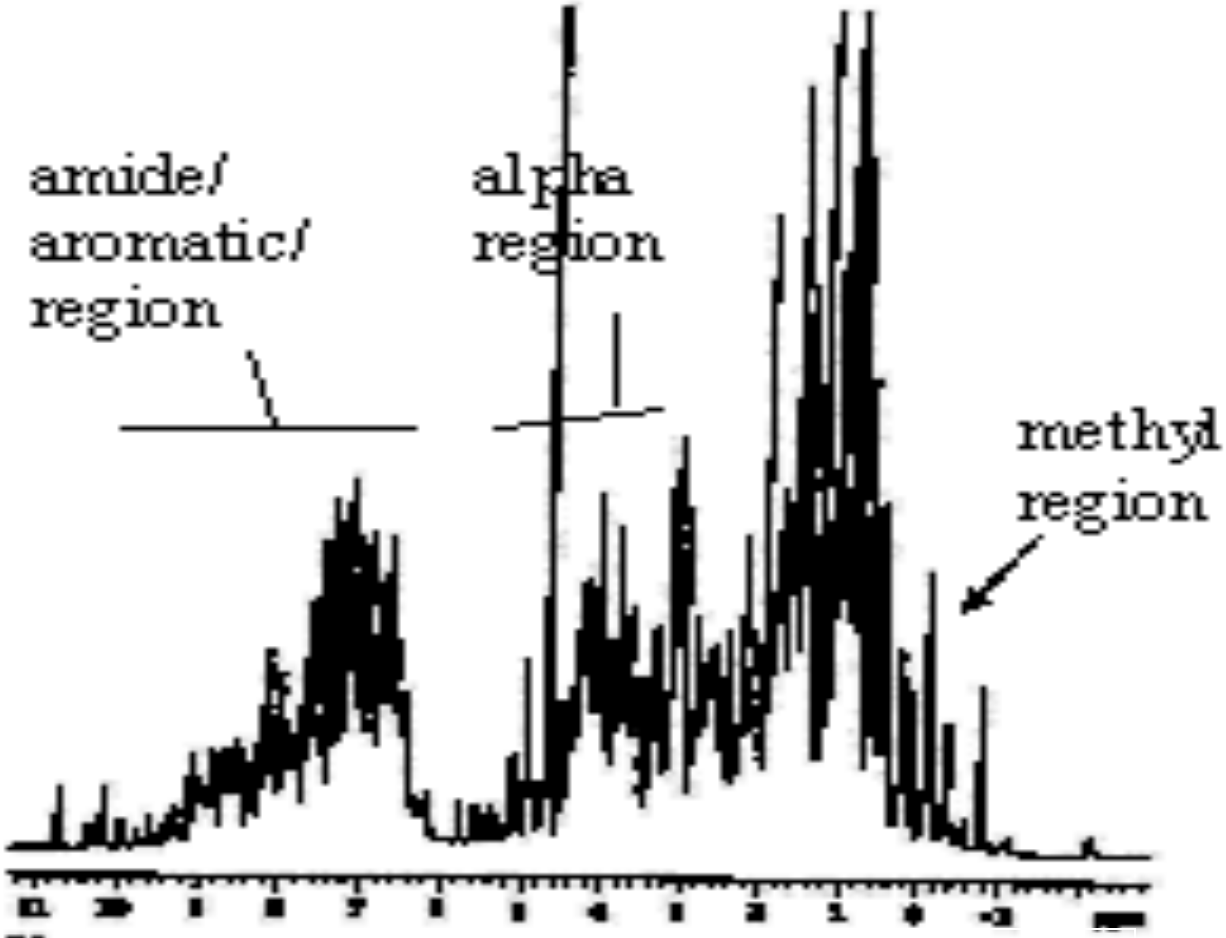
$$J(\phi) = A\cos^2\phi + B\cos\phi + C$$

Important Observables

Chemical shift is a reporter of magnetic environment

The J coupling can inform torsion angles

Protein NMR Spectroscopy



Periodic Table of NMR active Nuclei

Spin = $\frac{1}{2}$

Spin > $\frac{1}{2}$

IA	IIA	IIIB	IVB	VB	VIB	VIIB	VIIIB			IB	IIB	IIIA	IVA	VA	VIA	VIIA	VIIIA
H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Rd	Ac															

Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Isotopic Labeling Proteins for NMR

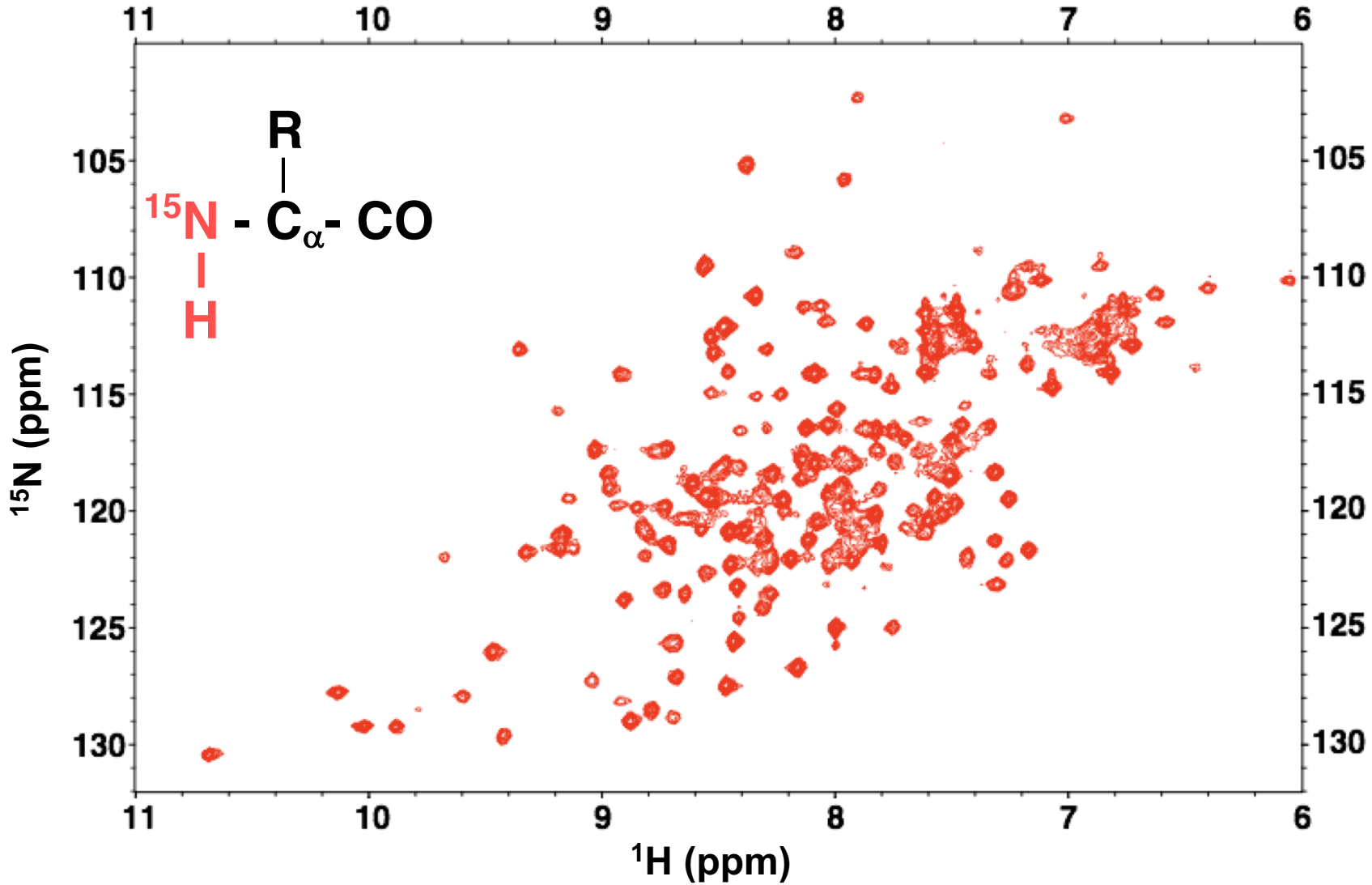
Bacterial expression:
Minimal media, ^{15}N NH_4Cl or
^{13}C glucose as sole nitrogen and
carbon source

Amino acid-type labeling
Auxotrophic or standard strains
(ei, BL21(DE3) depending on scheme

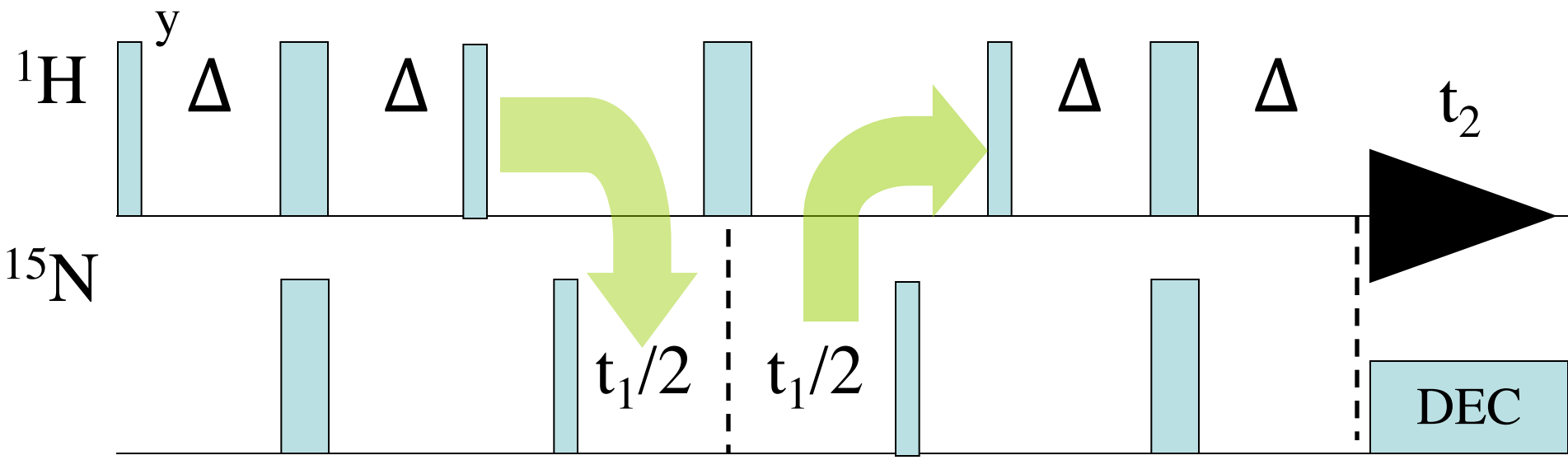
Labeling post purification ; reductive
methylation of lysines

Results in additional spin-1/2 nuclei which can be used as probes

The HSQC is an NH chemical shift correlation map



An overview of the HSQC



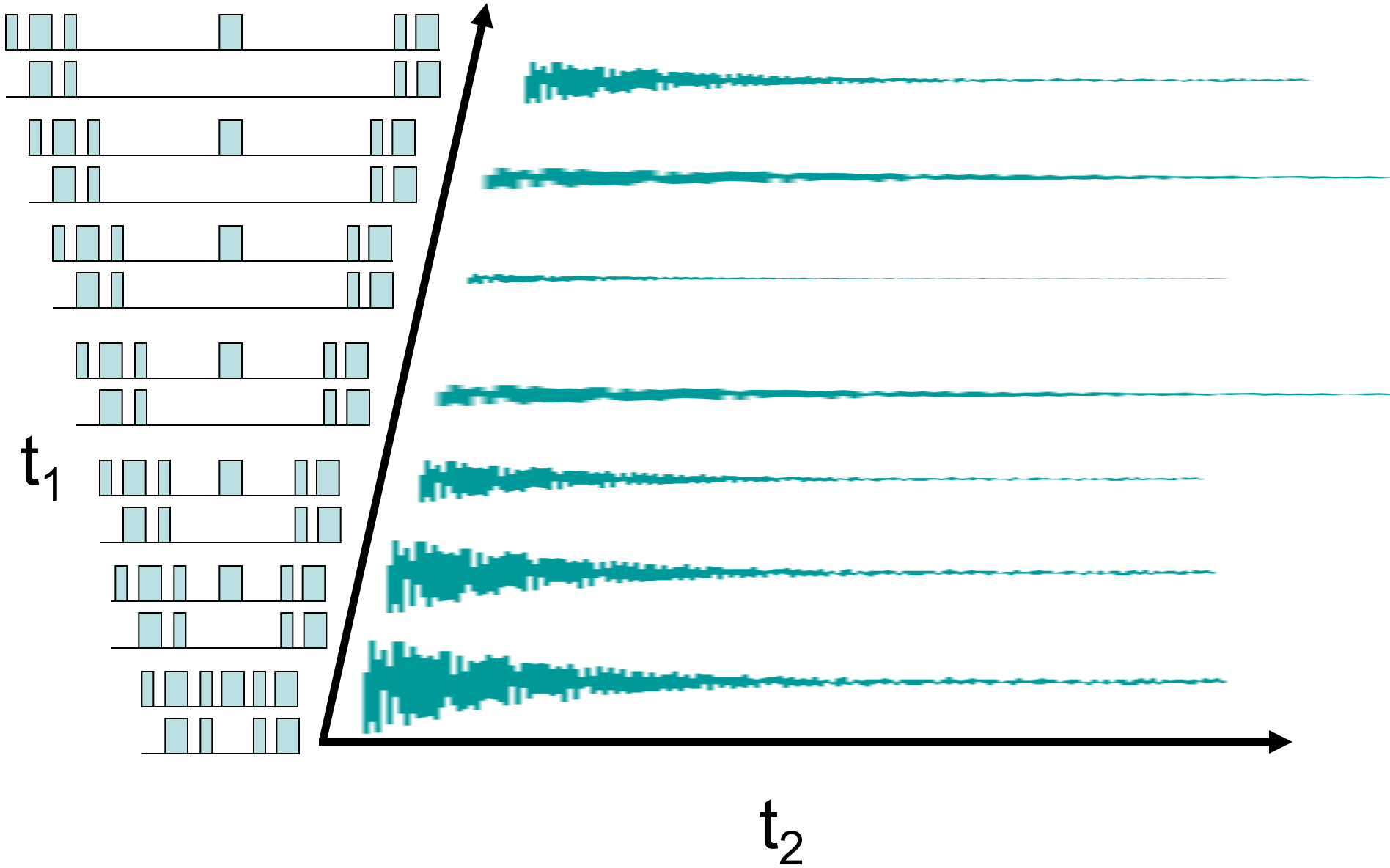
Transfer to ^{15}N

Transfer back to ^{1}H

Encode ^{15}N chemical shift for time t_1

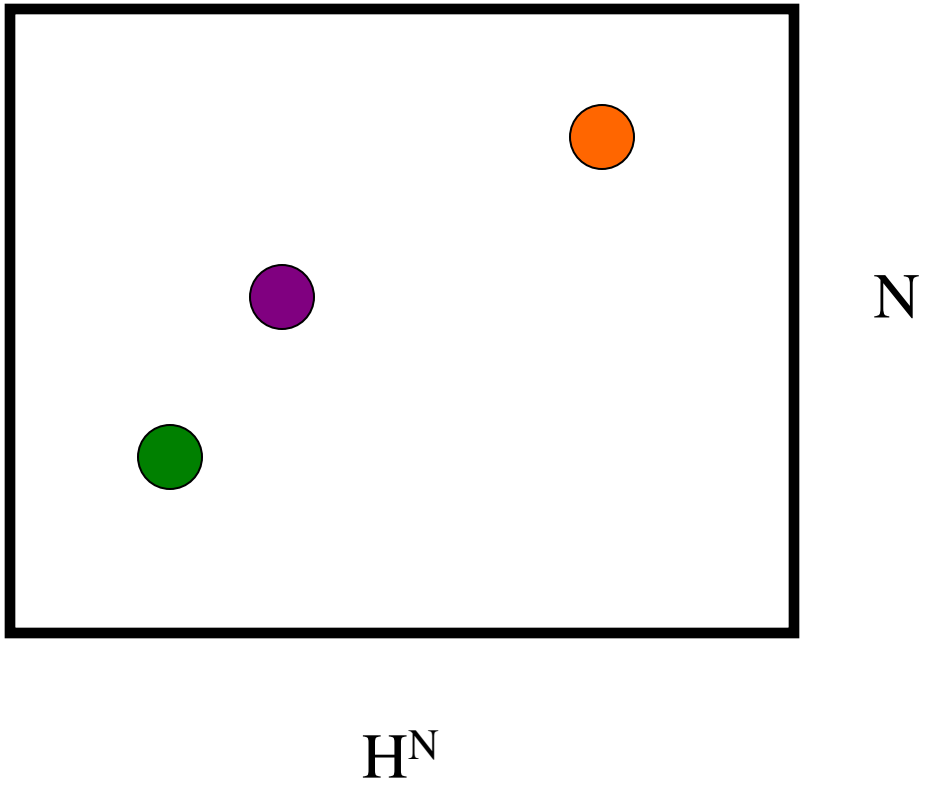
Bodenhausen & Ruben

2D Time-Domain Data

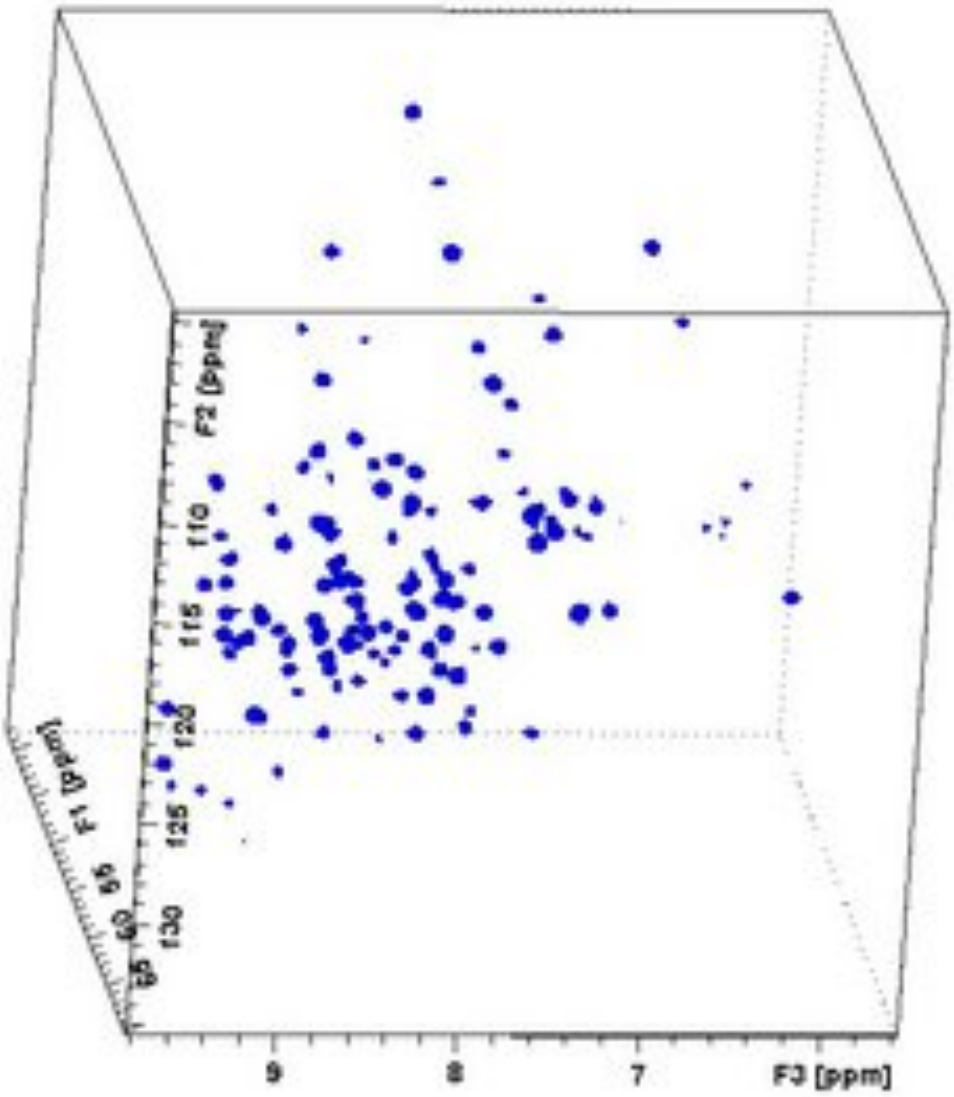


Some data shuffling then 2D FT =the HSQC Spectrum

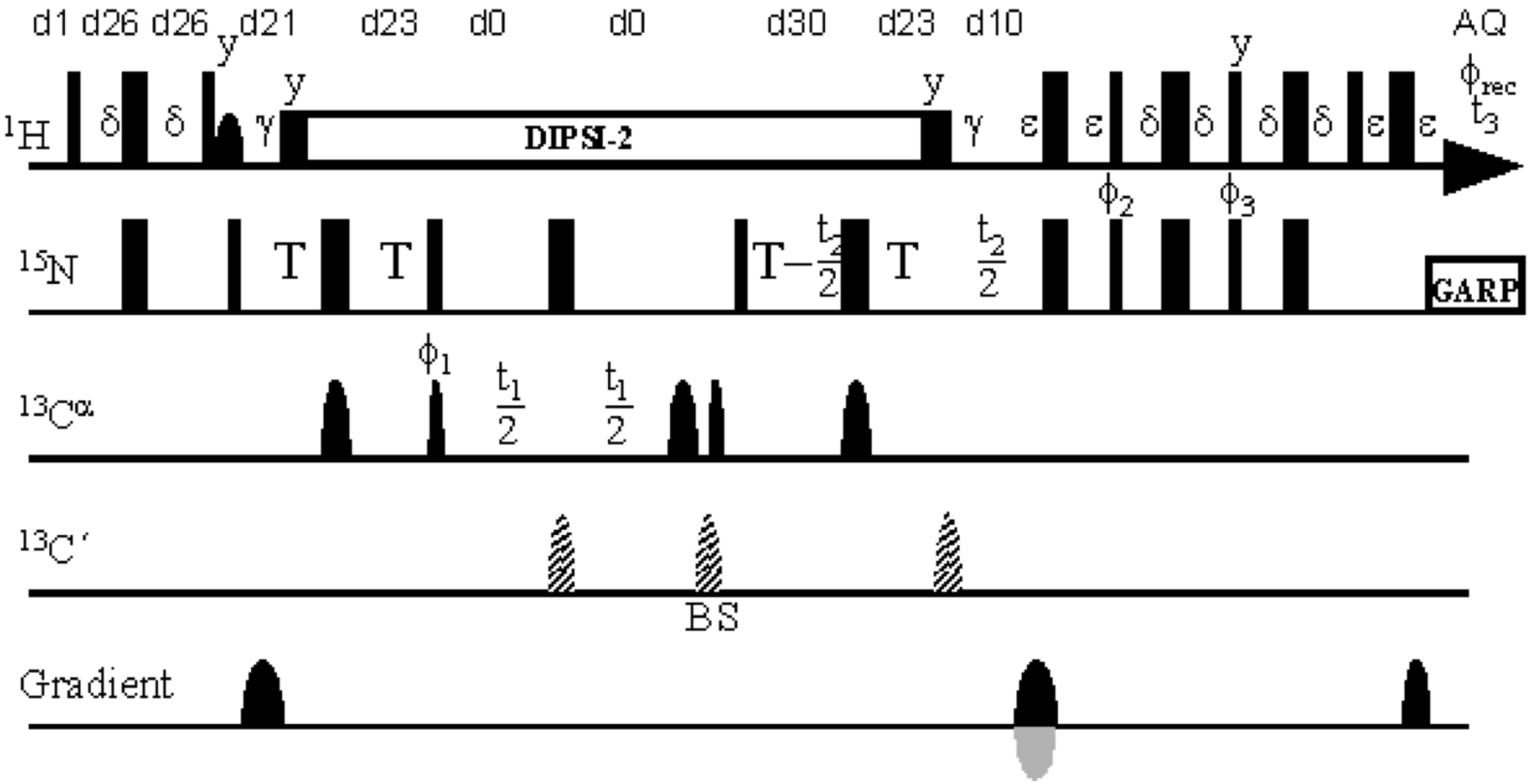
$$\mathrm{Re}\,[S'(\nu_1, \nu_2)] = A_1^N A_2^H$$



3D Dimensional NMR

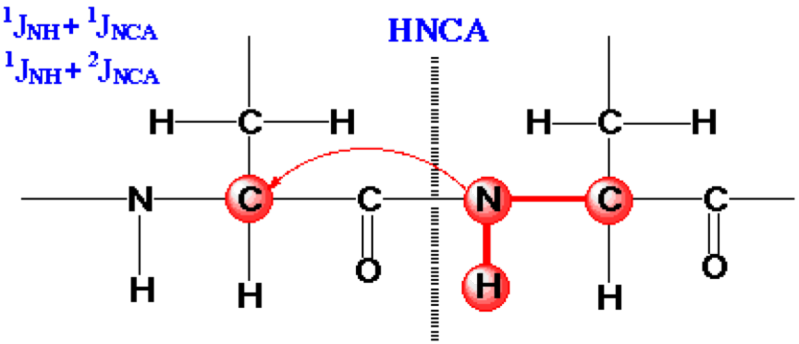


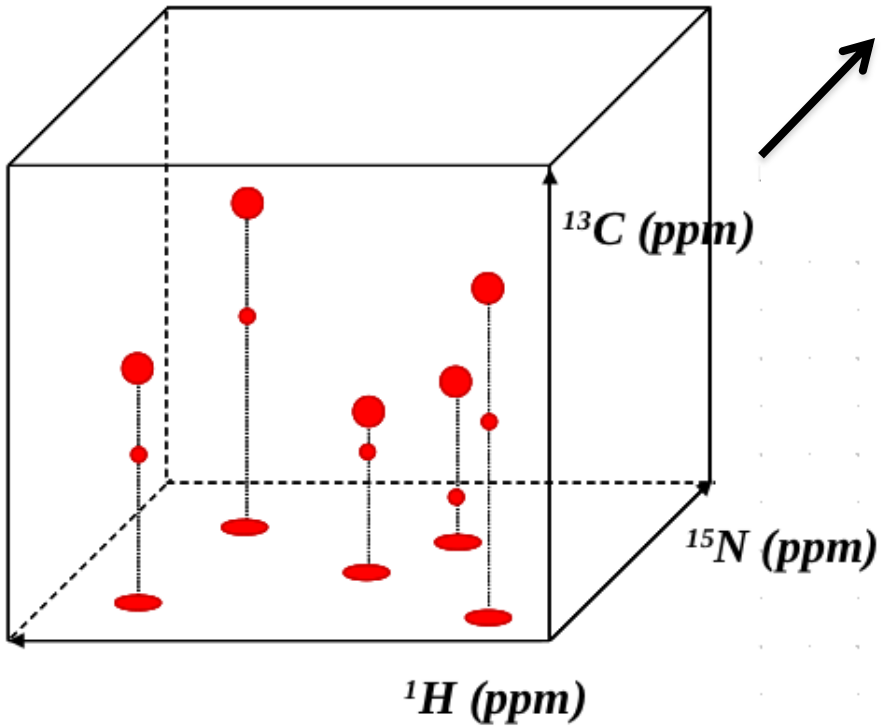
Resonance Assignments from Triple Resonance Experiments

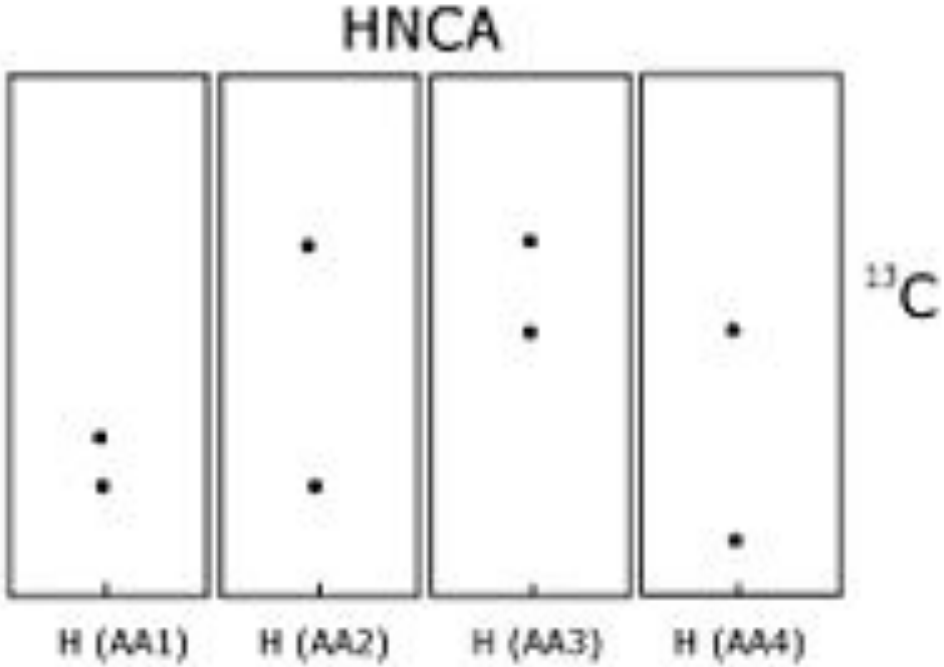


The 3D HNCA Experiment

Backbone Resonance Assignments from HNCA

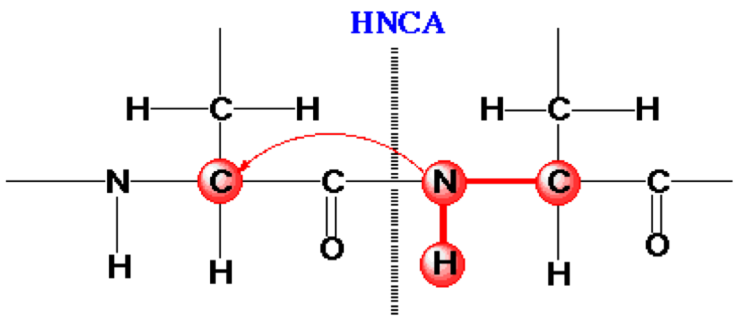


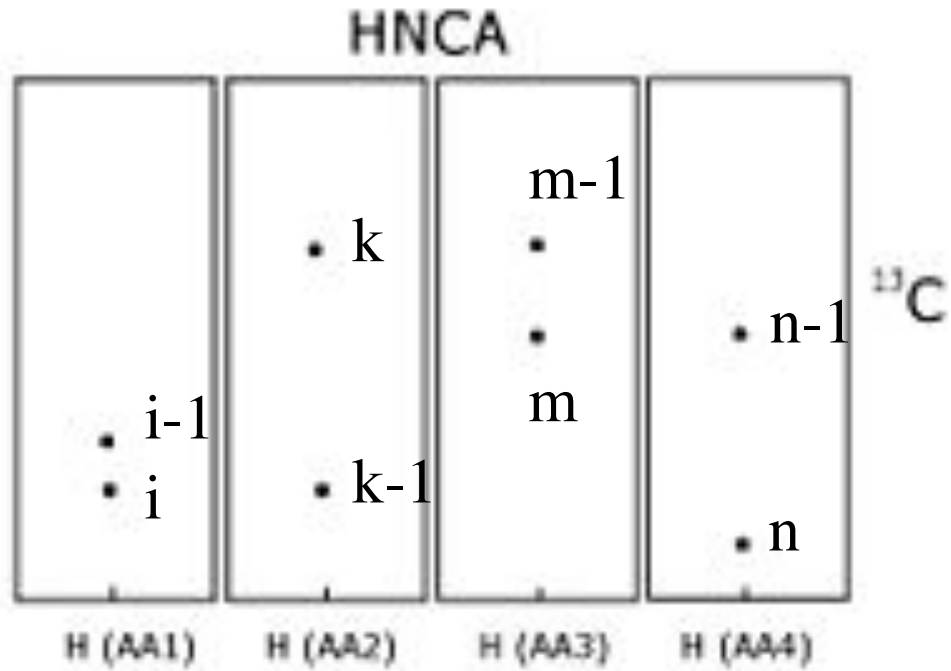




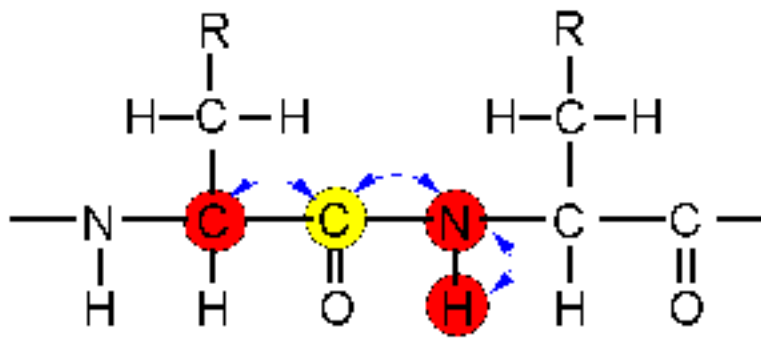
Triple Resonance Pairs

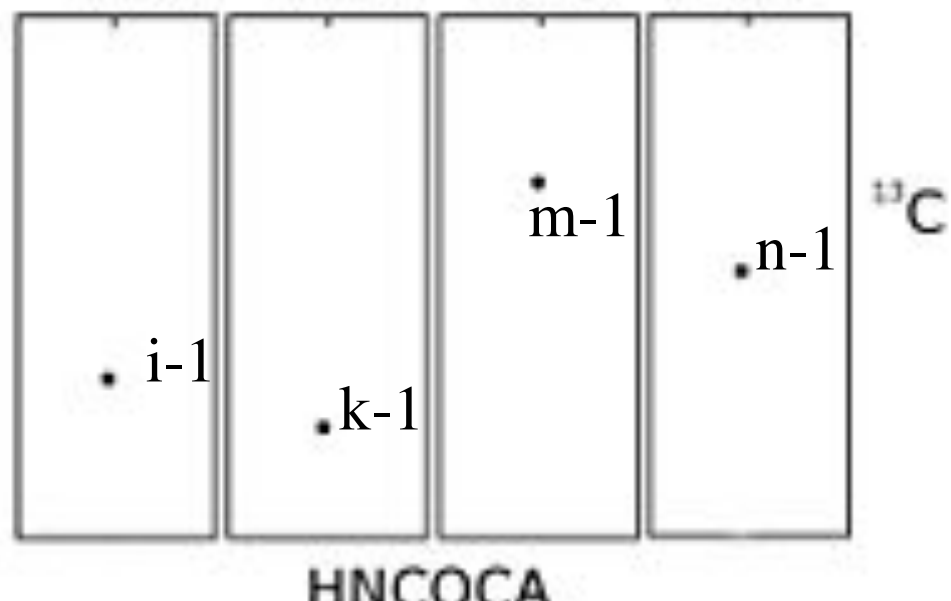
Residue i-1 Residue i





HN(CO)CA





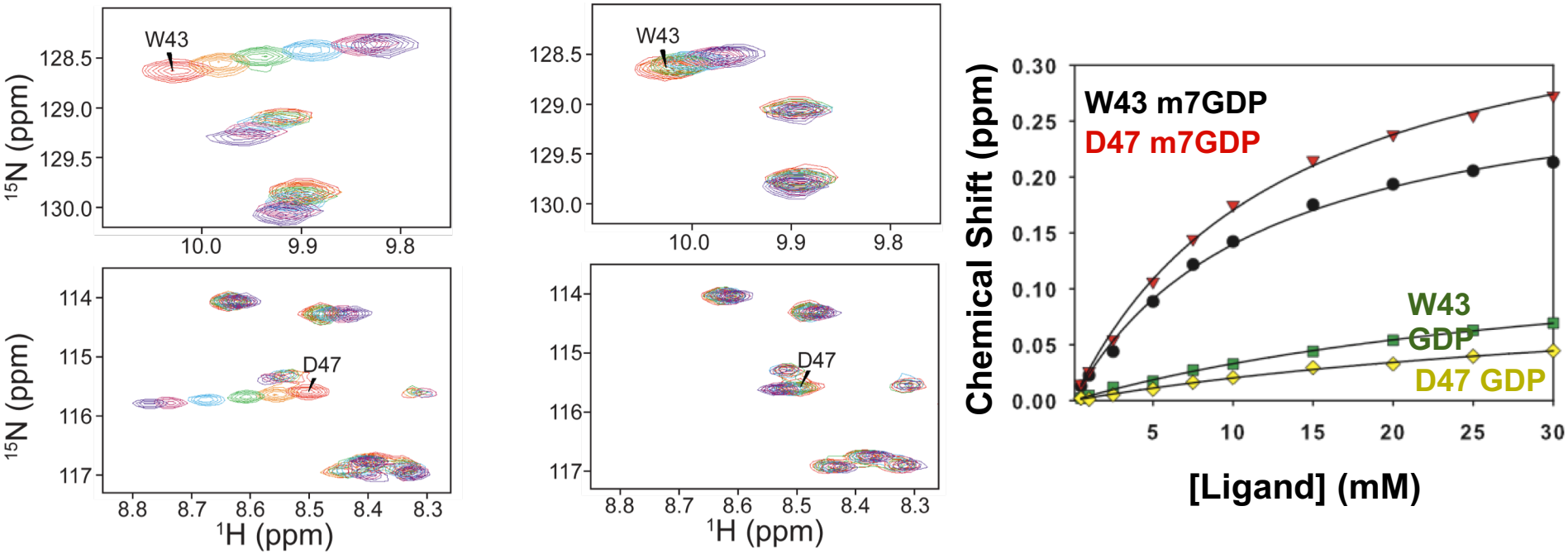
sequential correlation only

Applications for NMR

- Mapping protein interactions
- Fragment based drug discovery, SAR-by-NMR
- Protein folding , allostery and dynamics
- TROSY: deuteration and Methyl labeling to do this on large assemblies (~1 MDa)
- Structure Determination (<40 kDa)

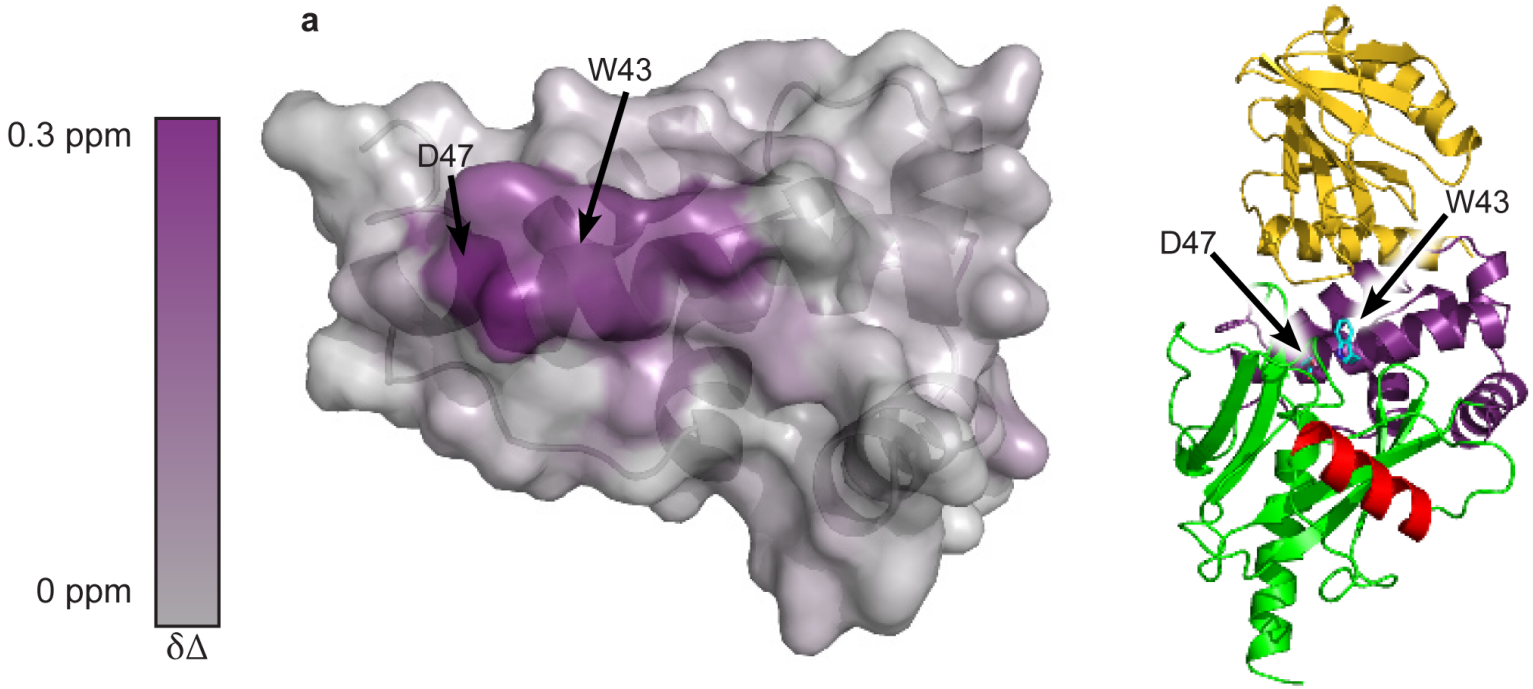
Part II:
Macromolecular Interactions Detected by NMR

Binding of nucleotide to protein



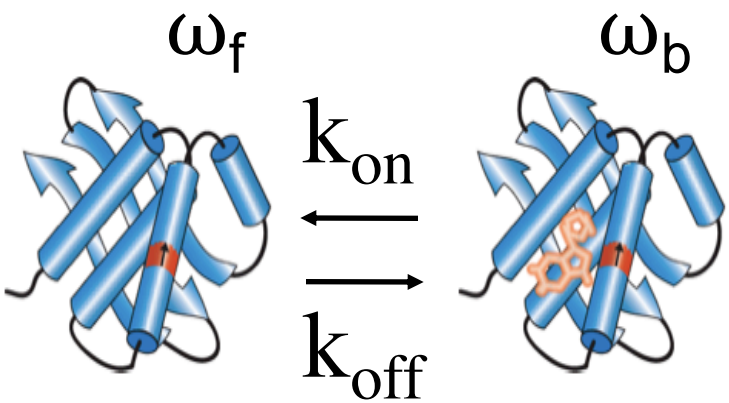
Dose dependent resonance shifts can be fit to obtain Kd

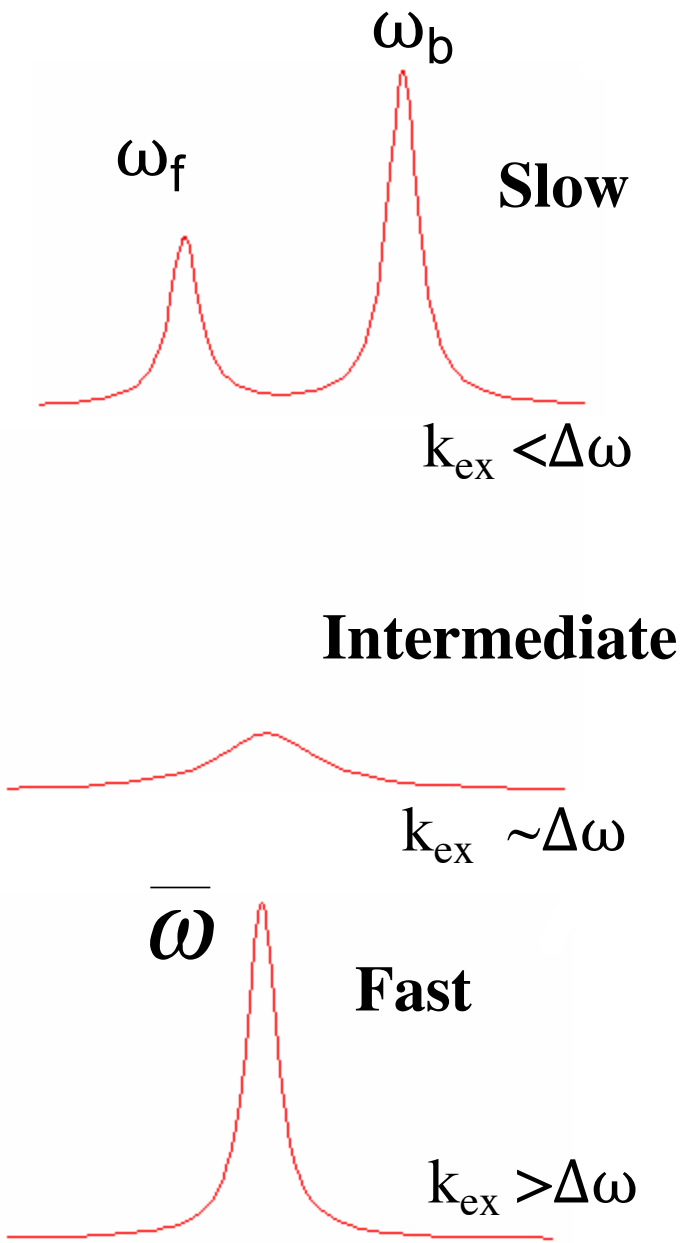
Shifts may be color coded onto surface to identify ligand binding site



Caveats?

NMR to monitor ligand binding





$$k_{ex} = k_{on}[L] + k_{off}$$

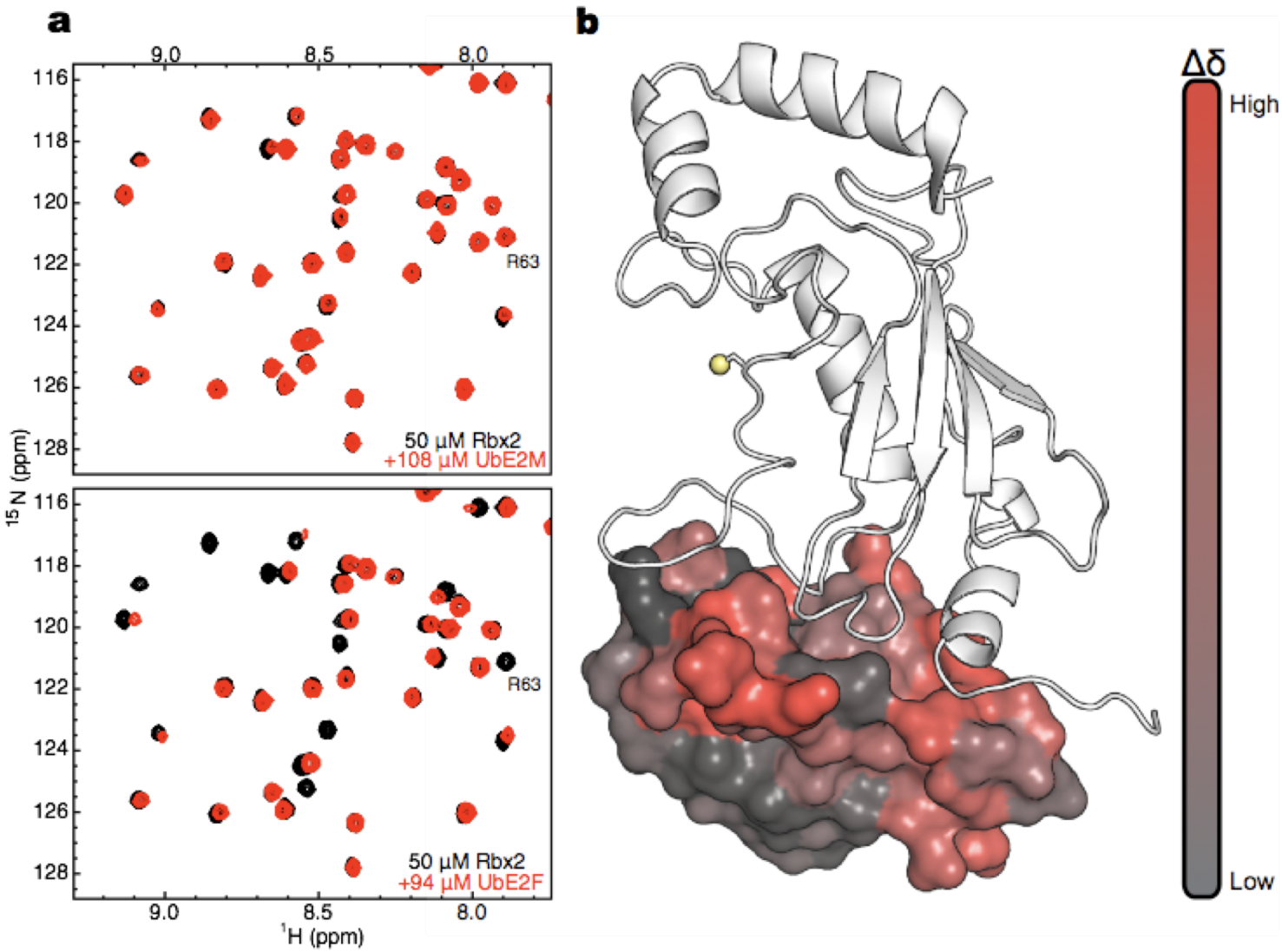
$$\Delta\omega = \omega_f - \omega_b$$

Fraction bound of labeled protein

$$P_b = \frac{\overline{\omega} - \omega_f}{\omega_b - \omega_f} = \frac{[L]}{[L] + K_d}$$

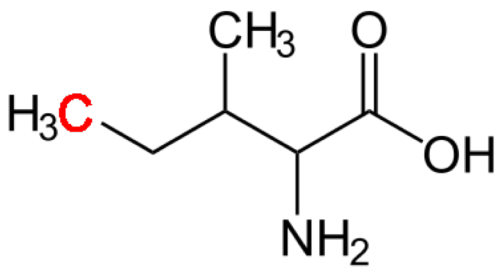
$\overline{\omega}$: observed chemical shift

Monitoring Protein/Protein Interactions by HSQC

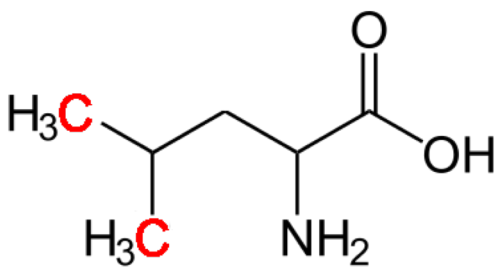


Sparse Labeling to Simplify Spectra

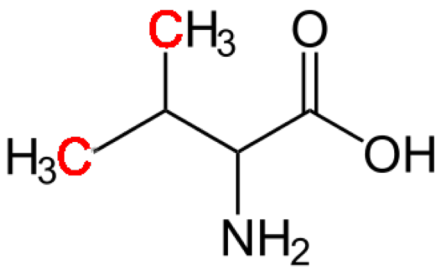
Selectively label R group methyls with C-13 (NMR visible)



Isoleucine



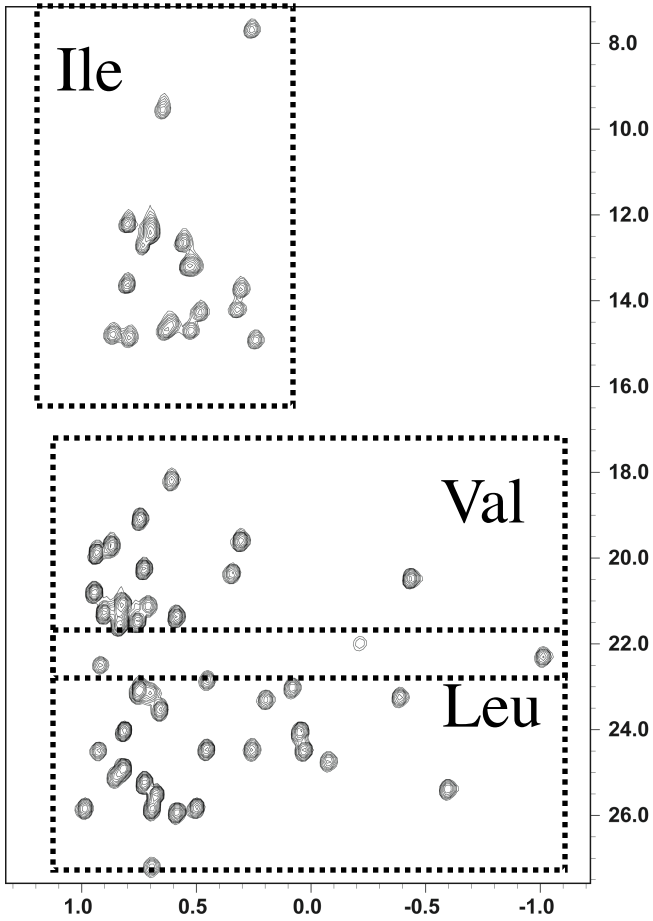
Leucine



Valine

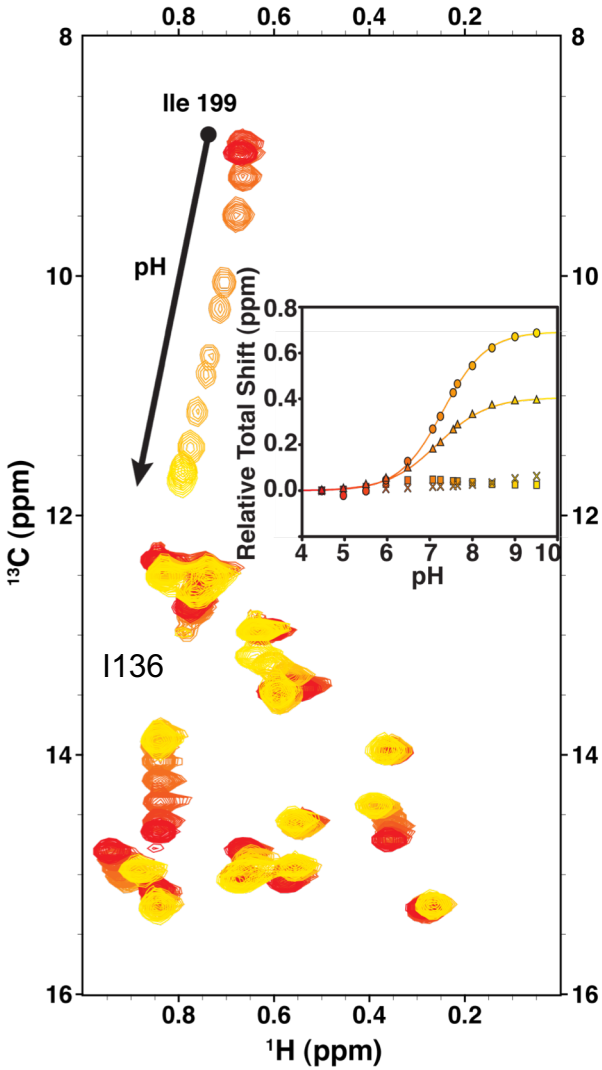
(add alpha-ketoacid precursors to ILV 30 minutes prior to induction)

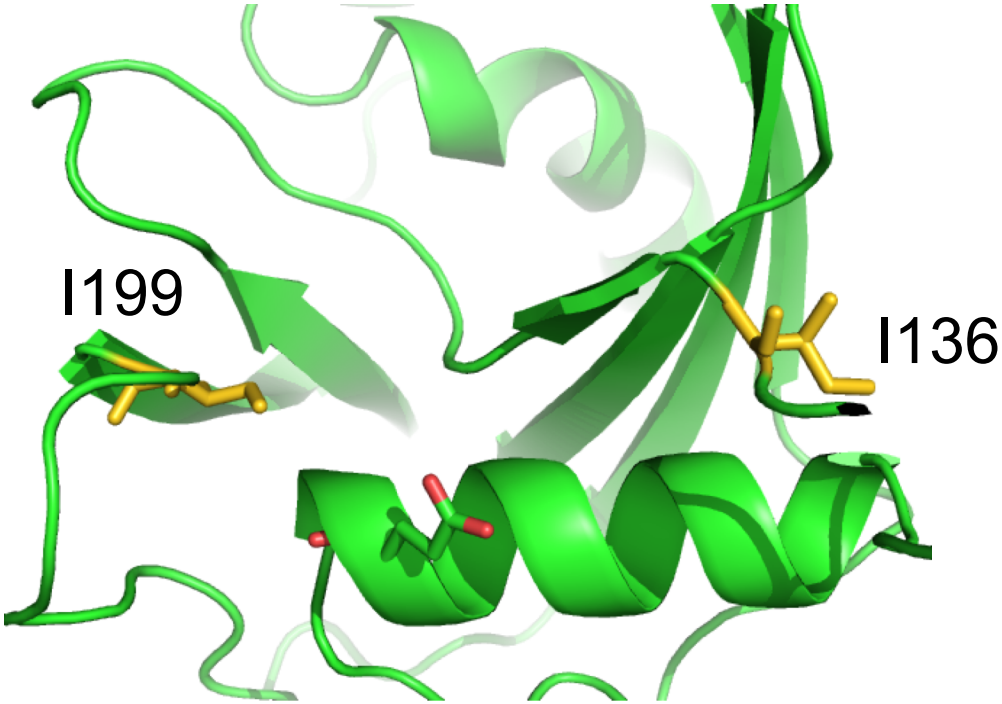
^{13}C HSQC of ILV labeled protein



Measuring pKa by NMR

(pH 4.5 - 9.5)





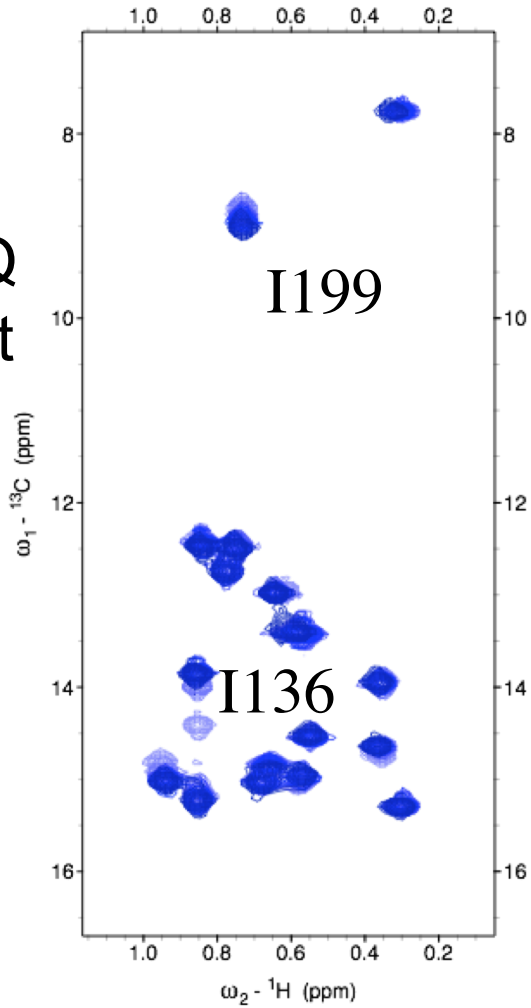
pKa of 7.2, elevated for Glu

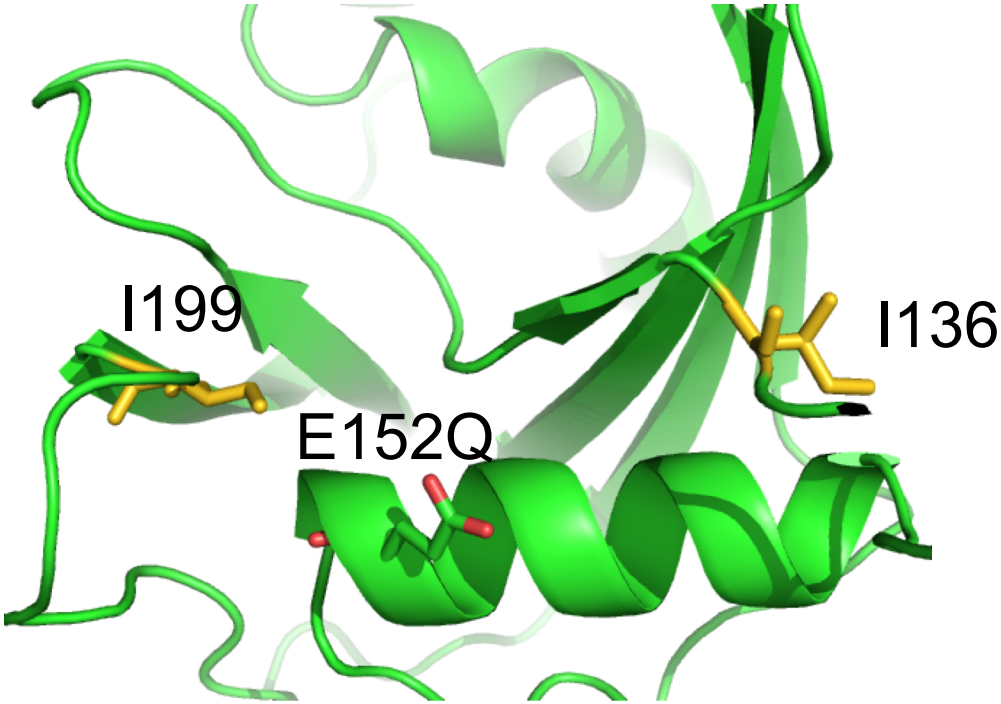
Aglietti et al, Structure 2013

Identification of titratable residue by site-directed mutagenesis and NMR

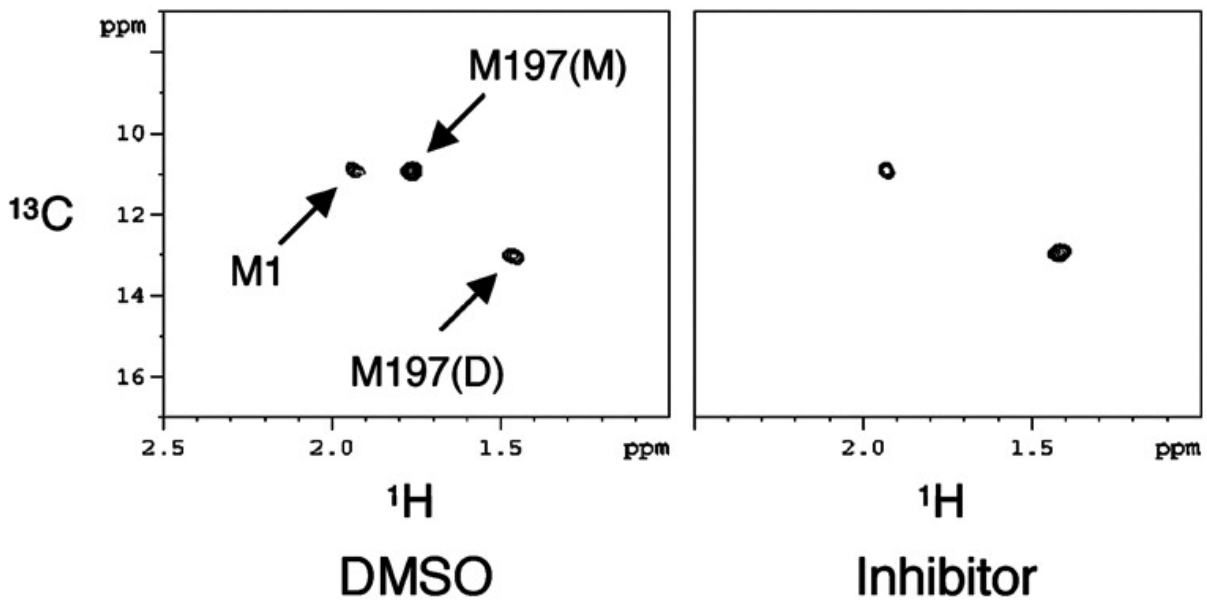
(pH 4.5 - 9.5)

E152Q mutant





Example of slow exchange: monomer-dimer equilibrium

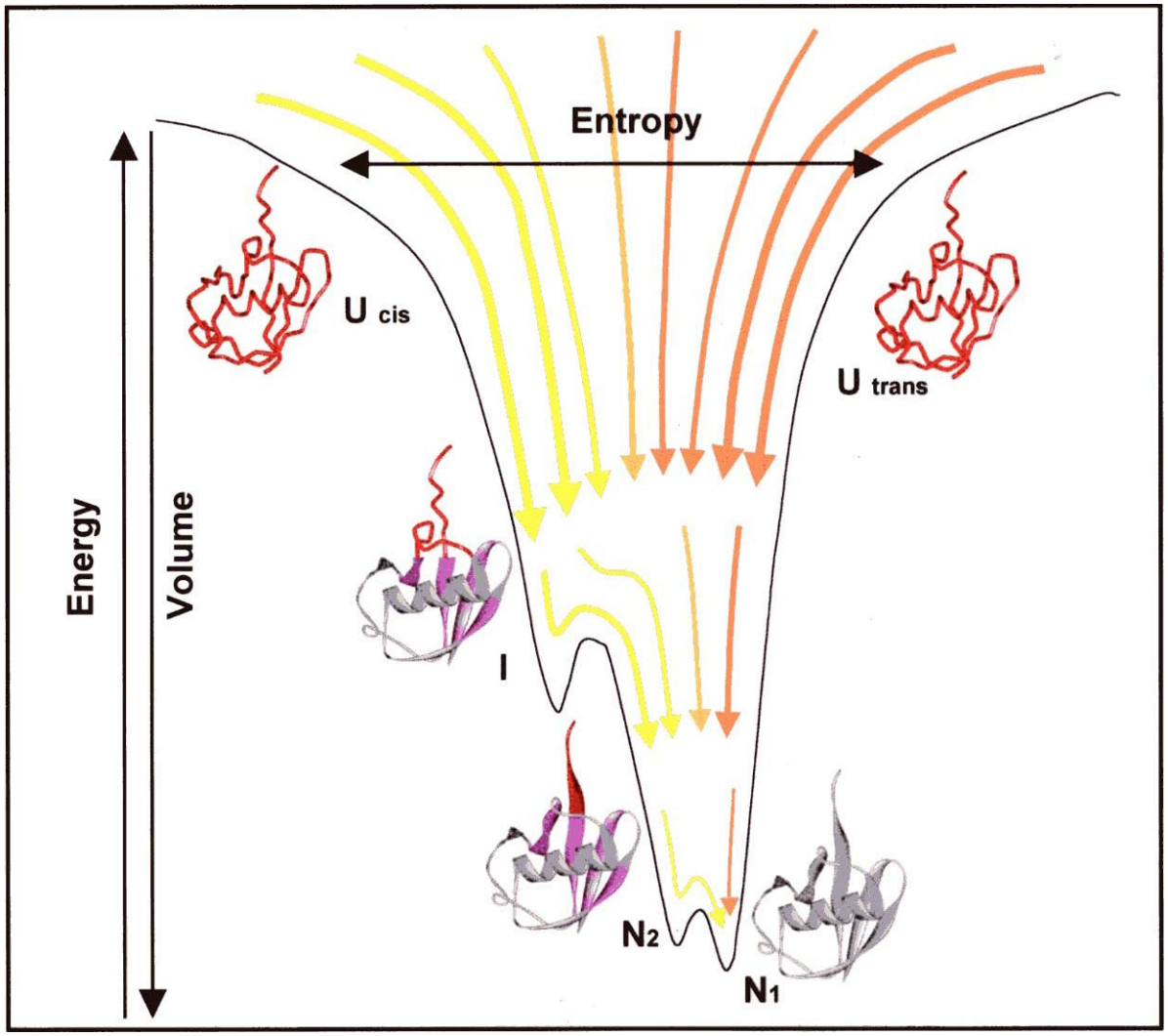


Inhibition of KSHV Pr stabilizes the dimeric conformation

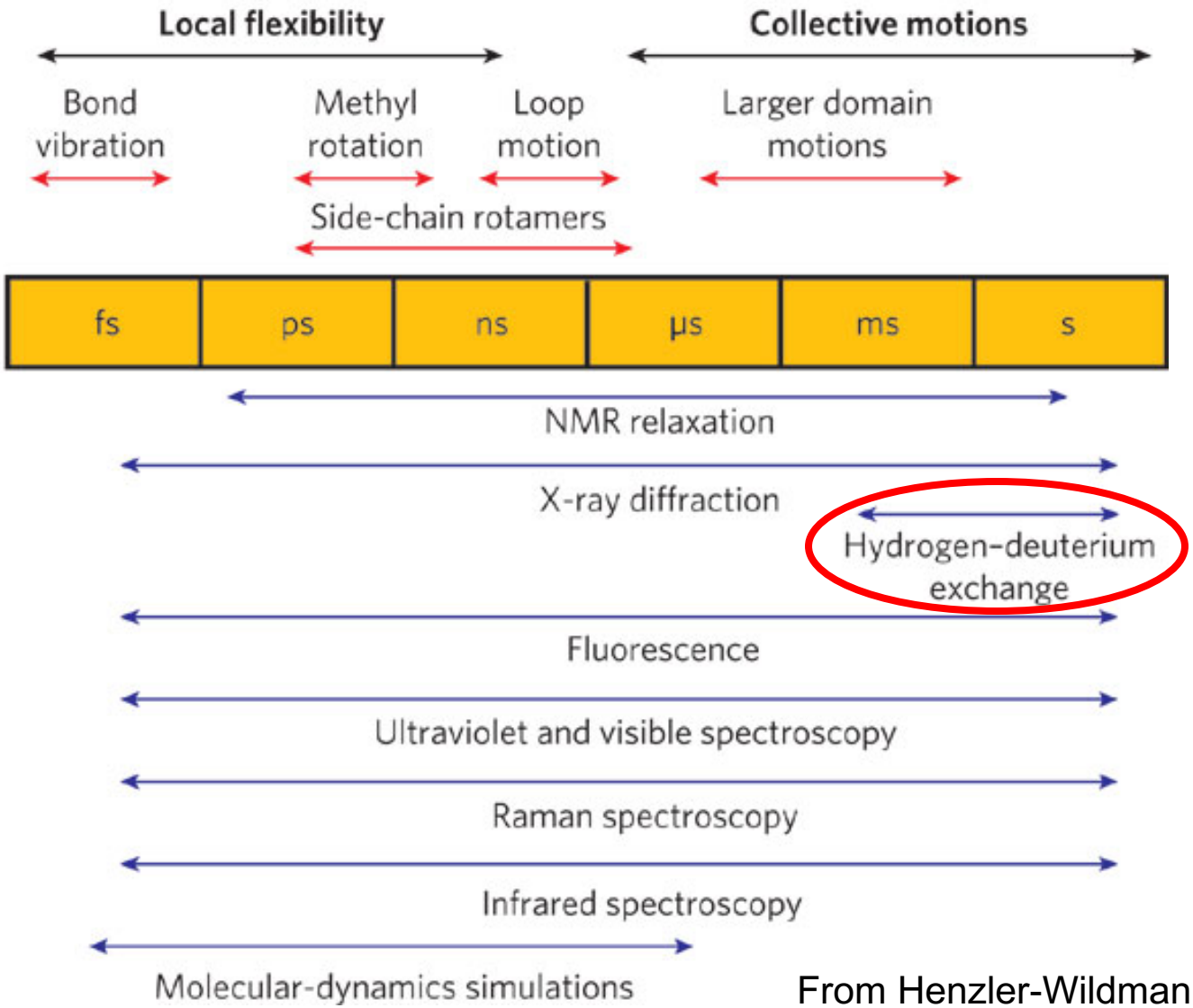
Methionine specific labeling simplifies analysis

Marnett A. B. et.al. PNAS 2004;101:6870-6875

Part III: Dynamics by NMR



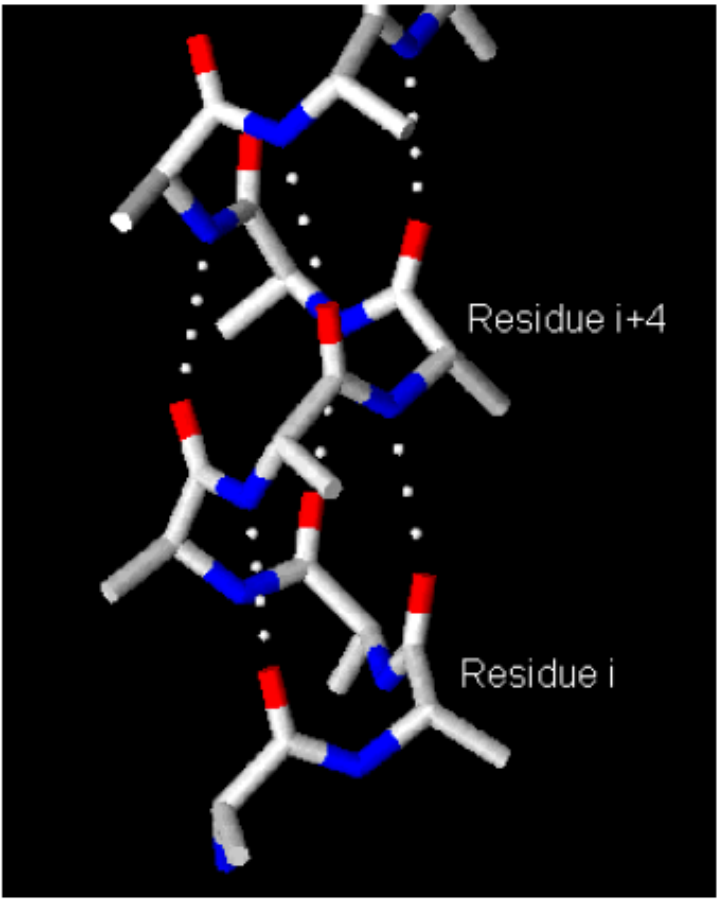
Timescales of Protein Dynamics

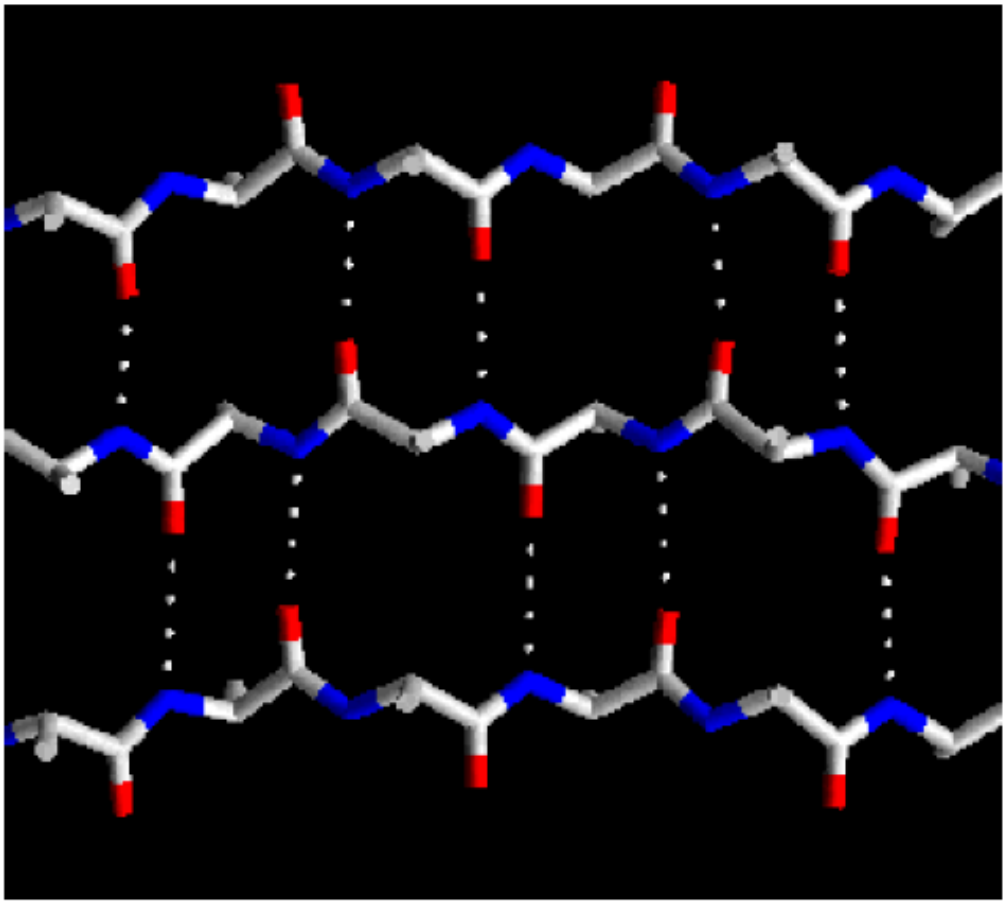


From Henzler-Wildman and Kern, Nature 2007

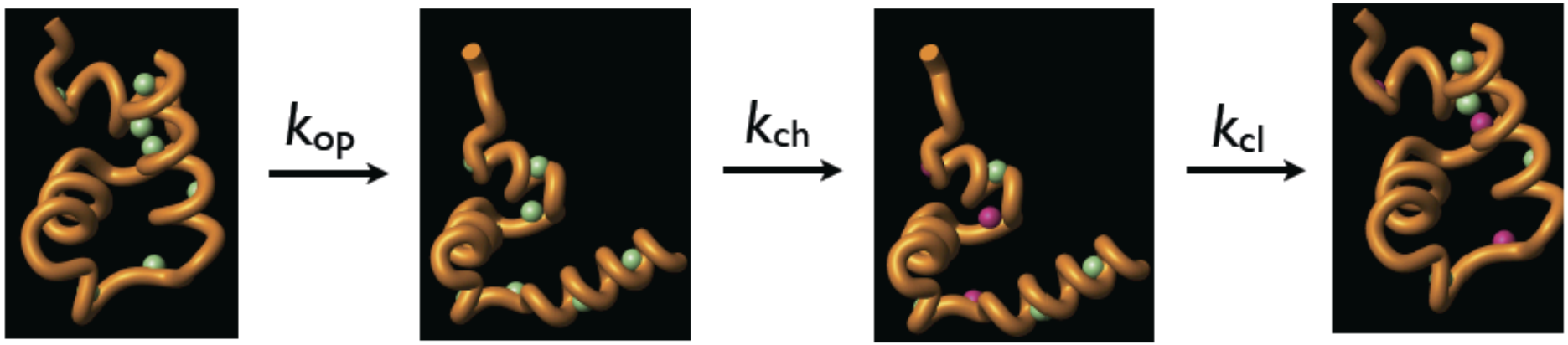
H/D exchange for measuring stability of H-bonds

Long lasting hydrogen bonds in proteins are typically part of secondary structure





Exchange of protons in the open conformation



EX1: $k_{cl} << k_{ch}$ $\qquad$ $k_{obs} = k_{op}$

EX2: $k_{cl} >> k_{ch}$ $\qquad$ $k_{obs} = k_{op}\, k_{ch}/(k_{cl}) = K_{op}\, k_{ch}$

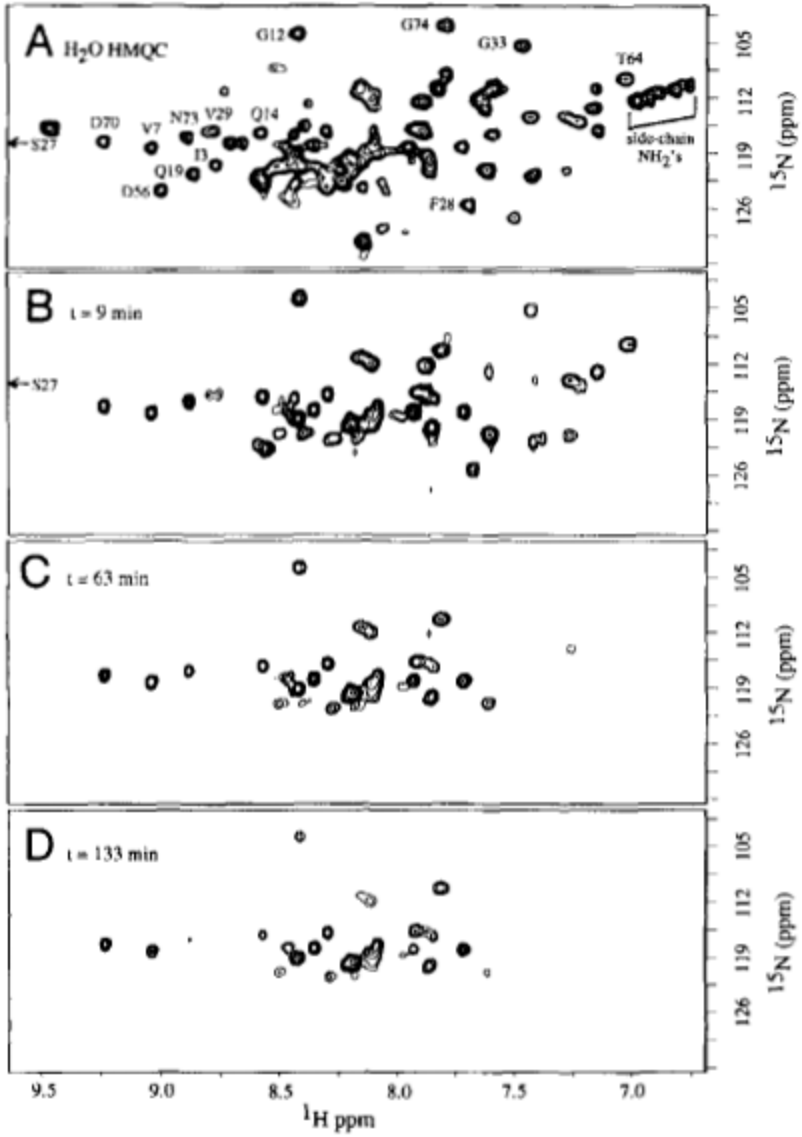
K_{op} is referred to as the protection factor, P

$$\Delta G_{op} = -RT\ln K_{op}$$

NMR Analysis of Protein Dynamics

Hydrogen-Deuterium Exchange

- As we saw before, slow exchanging NHs allowed us to identify NHs involved in hydrogen-bonds.
- Similarly, slow exchanging NHs are protected from the solvent and imply low dynamic regions.
- Fast exchanging NHs are accesible to the solvent and imply dynamic residues, especially if not solvent exposed.



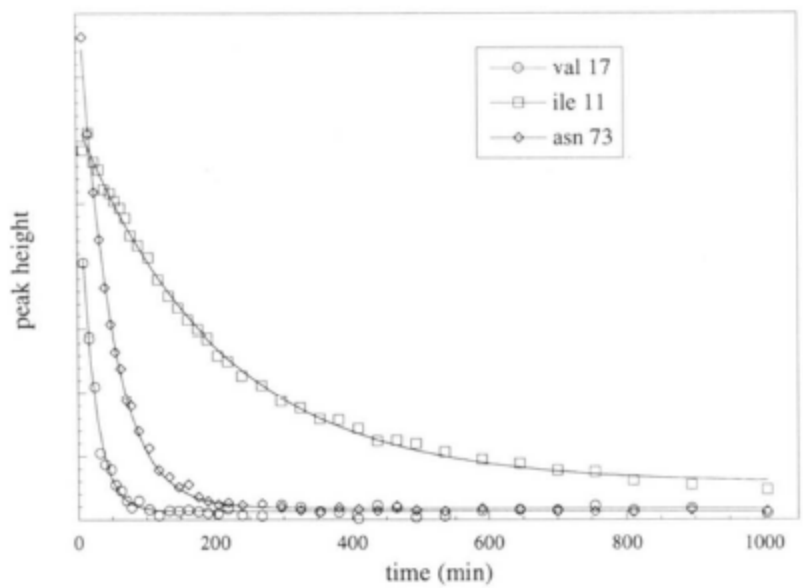
Protein sample is exchanged into D_2O and the disappearance of NHs peaks in a 2D ^1H-^{15}NH spectra is monitored.

Protein Science (1995), 4:983-993.

NMR Analysis of Protein Dynamics

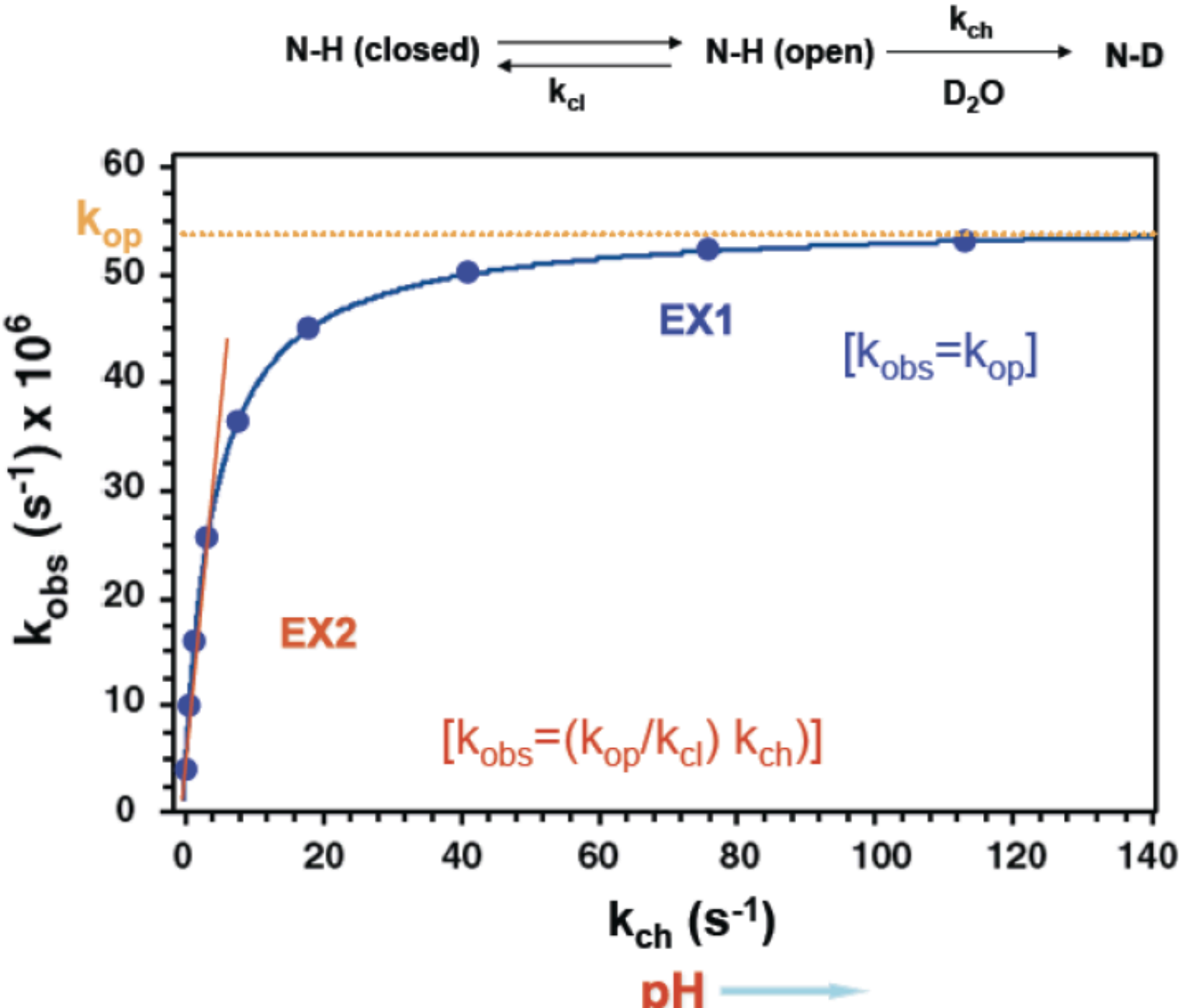
Hydrogen-Deuterium Exchange

- The observed NH intensity loss can be fit to a simple exponential to measure an exchange rate (k_{ex})
- These exchange rates may range from minutes to months!
  - NHs with long exchange rates indicate stable or low mobility regions of the protein
  - NHs with short exchange rates indicate regions of high mobility in the protein

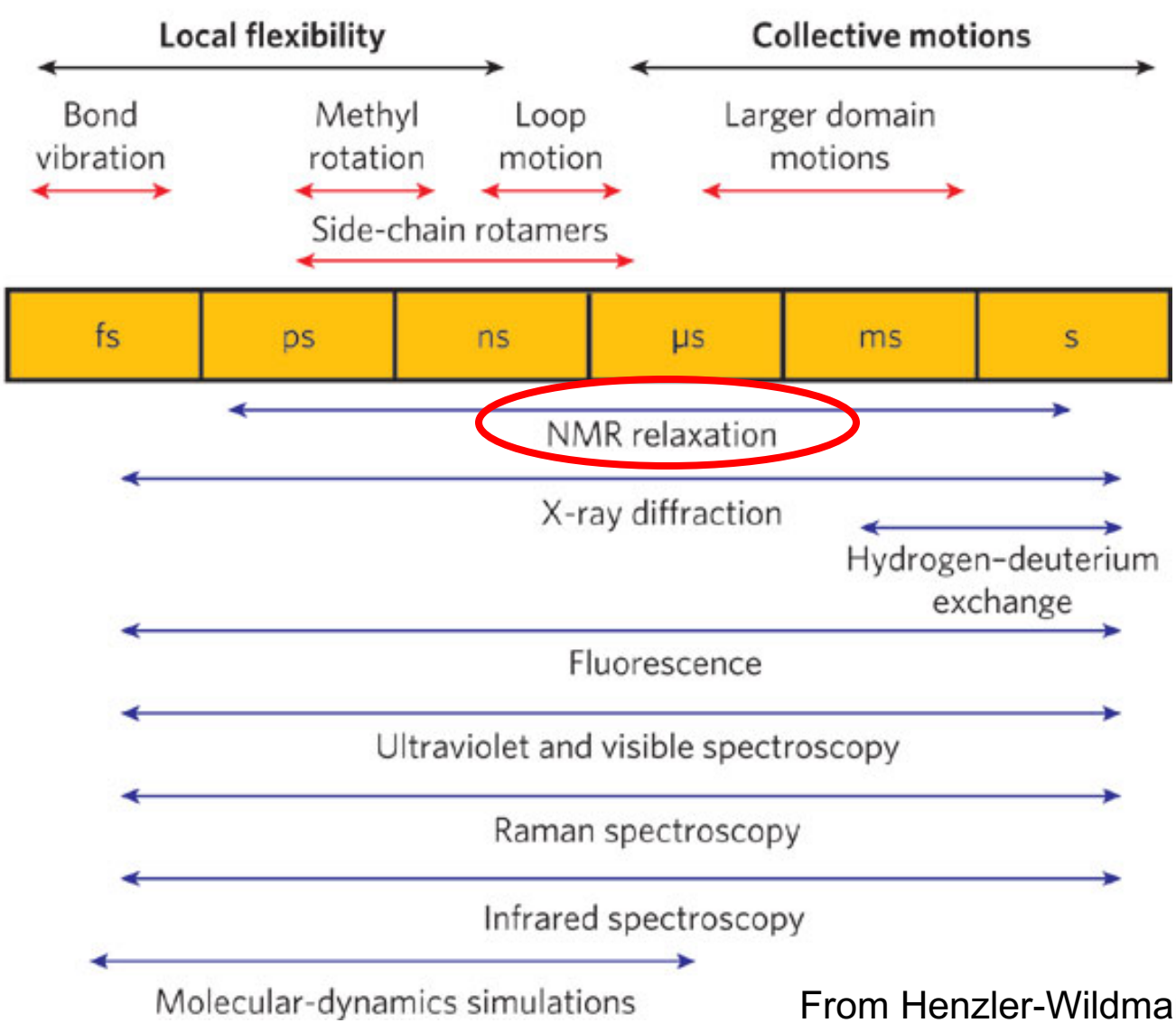


$$I = \alpha e^{-k_{obs}t}$$

The exchange mechansim depends on pH

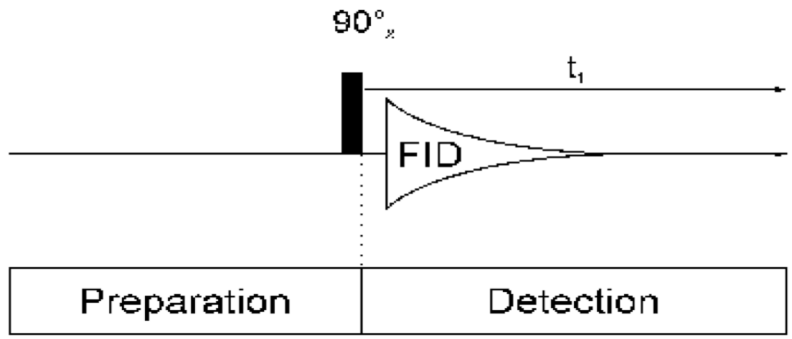


Timescales of Protein Dynamics

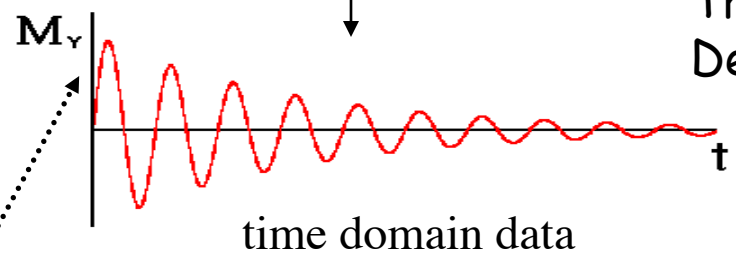


From Henzler-Wildman and Kern, Nature 2007

Summary of 1D Experiment



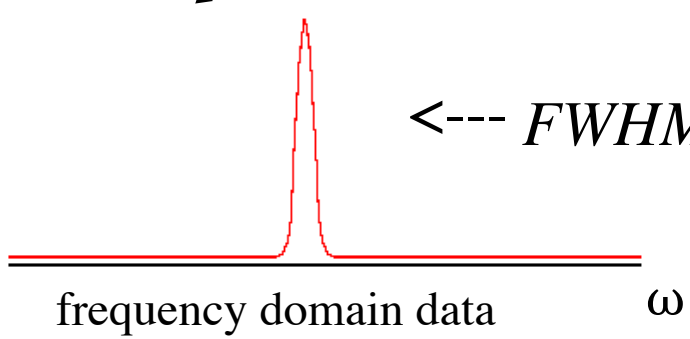
Transverse Magnetization
Decay constant T2



$$R_2 = \frac{1}{T_2}$$

Amplitude propotional
to amount magnetization
prior to pulse

Fourier Transform (FT)

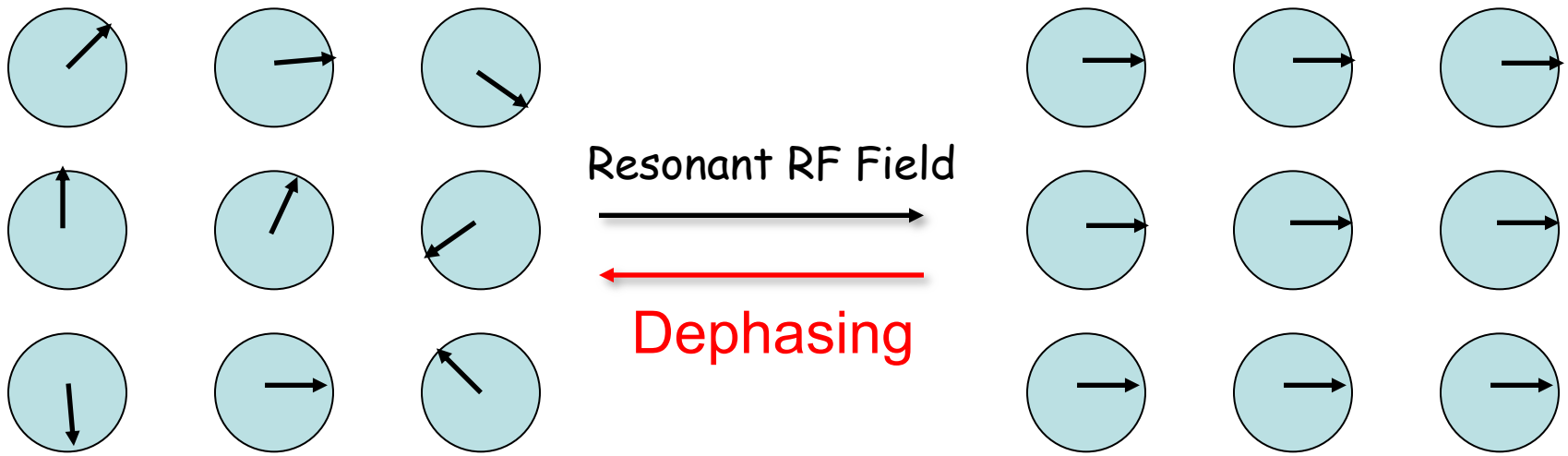


<--- $FWHM = \frac{R_2}{\pi}$

Position of resonance --->
local magnetic environment

R_2 is a Measure of Dephasing

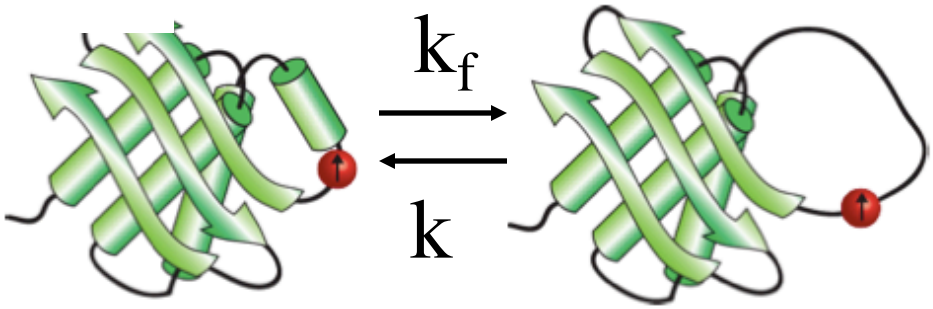
Ensemble of Nuclear Spins

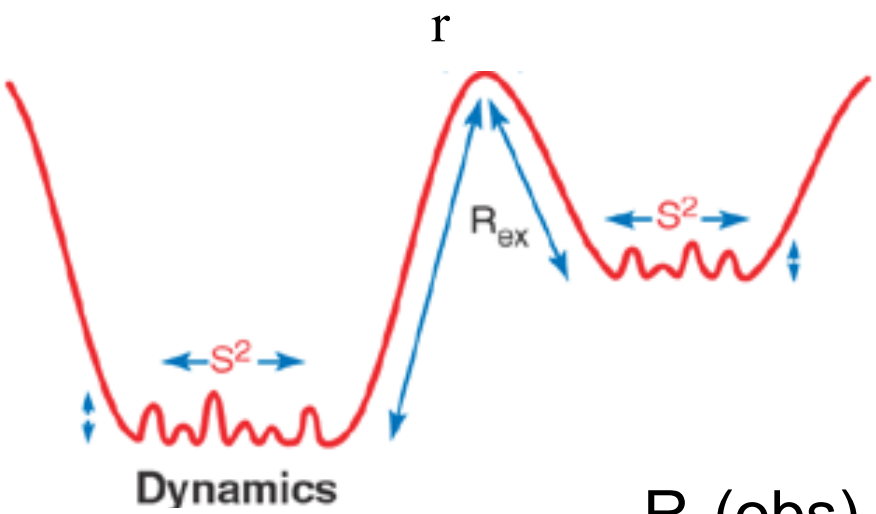


Random Phase
No NMR Signal

Phase Synchronization
NMR Signal!

Contributions to R_2 from Conformational Dynamics on the Chemical Shift timescale



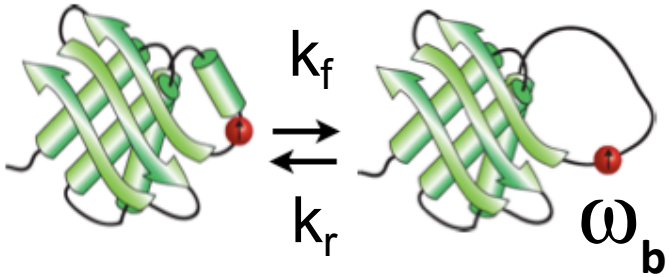


Tumbling,
Vibration,
Libration

Binding,
Conf Change,
Allostery

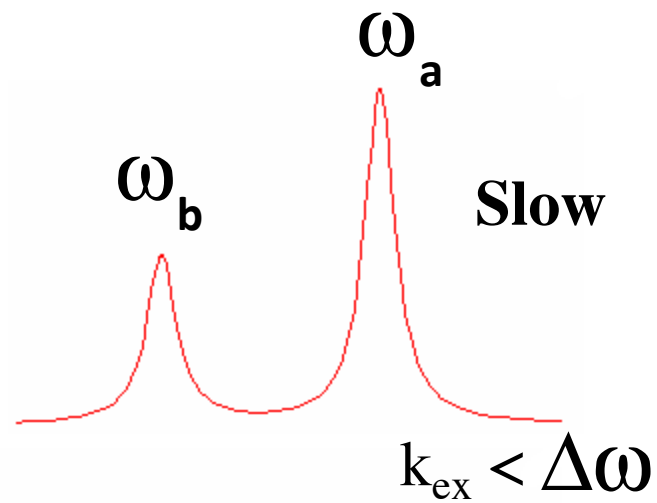
$$R_2(\text{obs}) = R_2(\text{ns-ps}) + R_{ex}(\text{ms-}\mu\text{s})$$

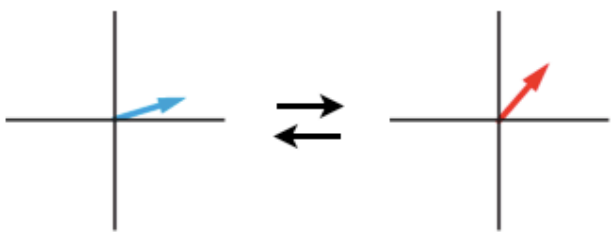
Spectral Manifestations of Exchange

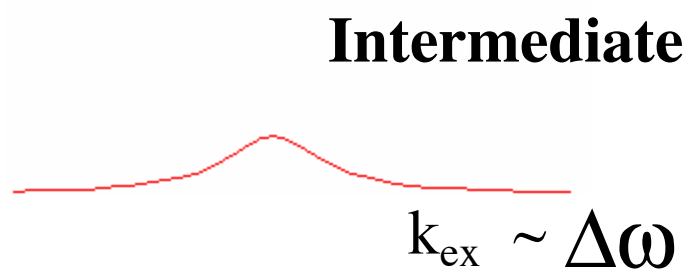


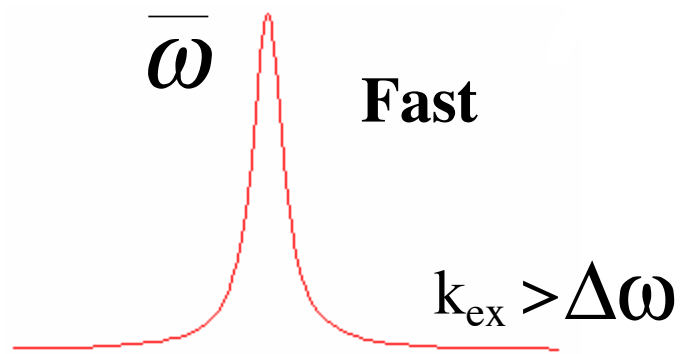
$$k_{ex} = k_f + k_r$$

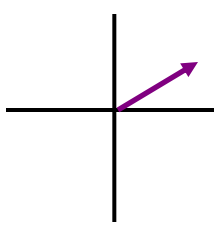
$$\Delta\omega = \omega_a - \omega_b$$



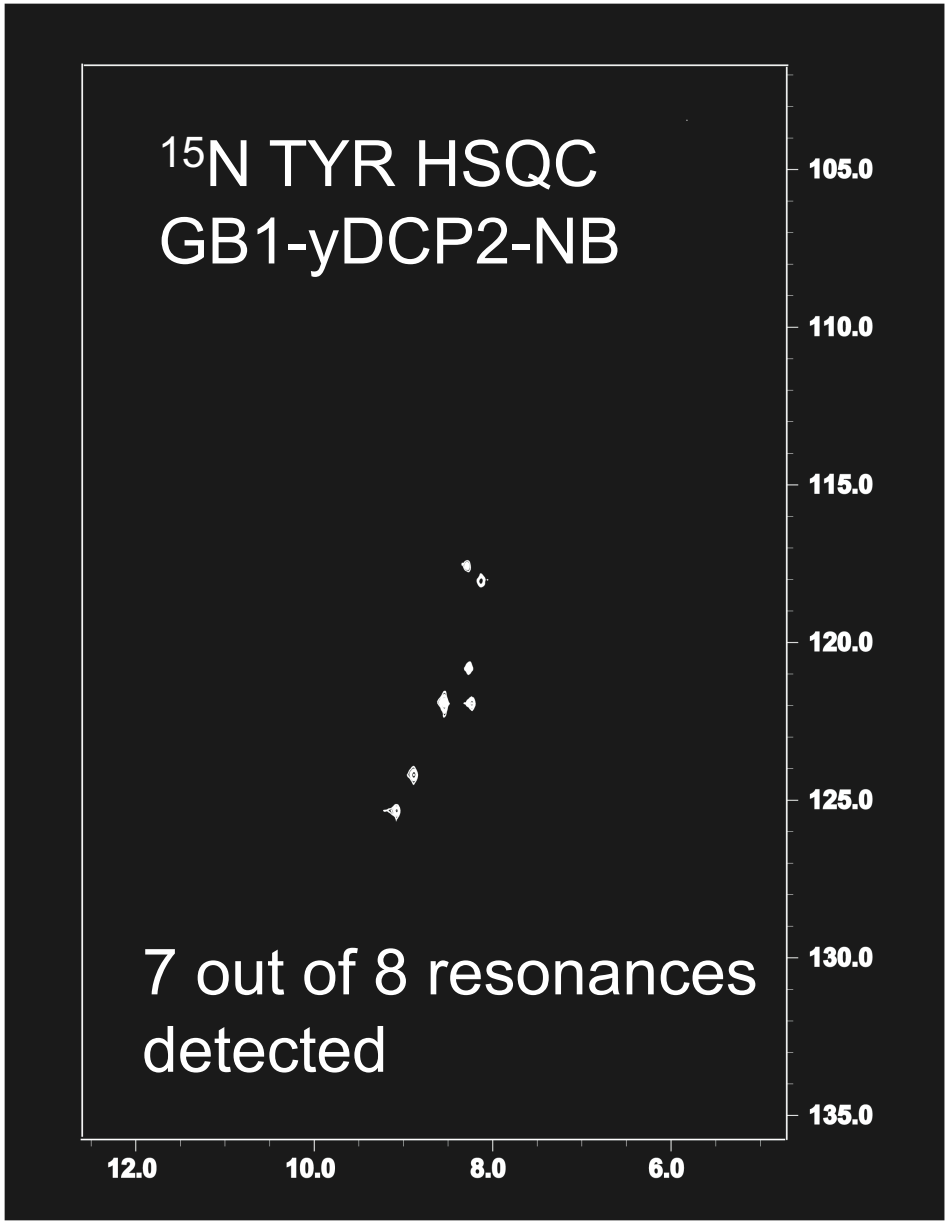




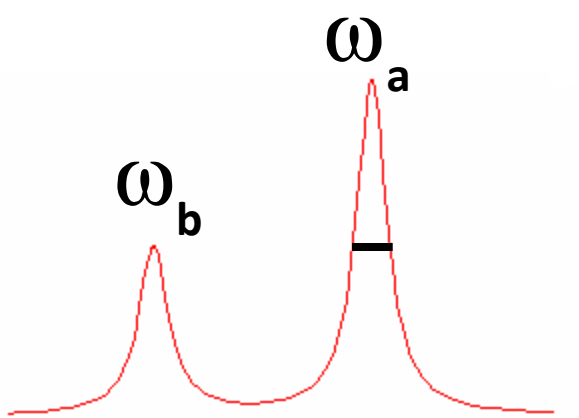




$$\overline{\omega} = p_a\omega_a + p_b\omega_b$$


15N TYR HSQC
GB1-yDCP2-NB
7 out of 8 resonances detected
105.0
110.0
115.0
120.0
125.0
130.0
135.0
12.0
10.0
8.0
6.0

Expressions for the Linewidth in the Slow Exchange Limit ($k_{ex} < \Delta\omega$)



$$A \underset{k_{rev}}{\overset{k_{for}}{\rightleftharpoons}} B$$

Populations p_a, p_b

$k_{ex} = k_{for} + k_{rev}$

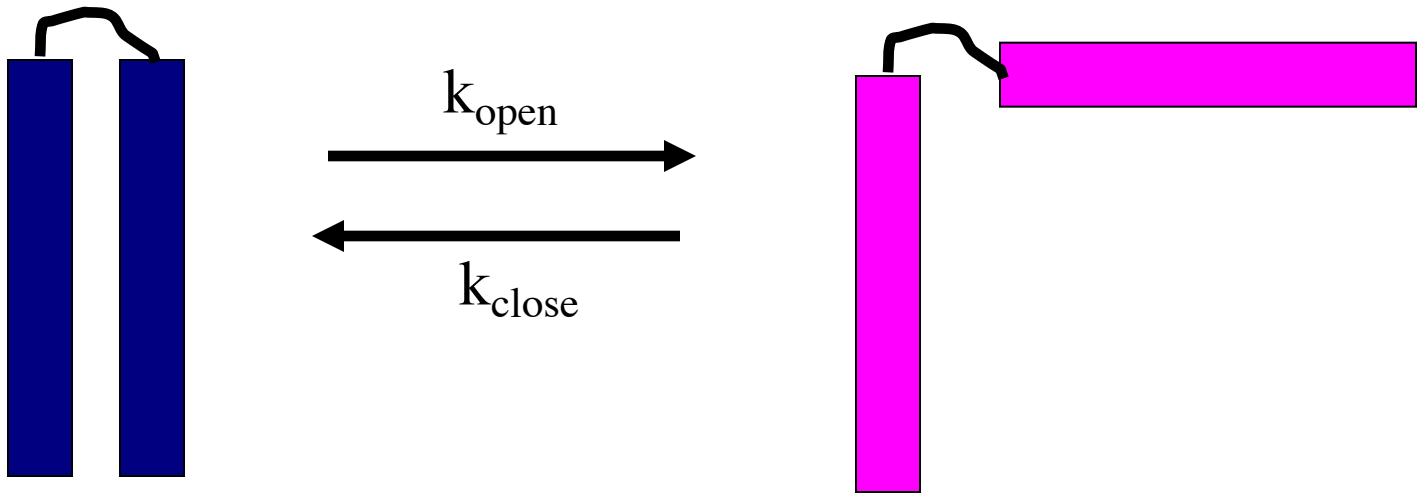
$\Delta\omega = \omega_a - \omega_b$

$$R_2^a = R_2^a(\text{ns} - \text{ps}) + p_b k_{for}$$

$$R_2^b = R_2^b(\text{ns} - \text{ps}) + p_a k_{rev}$$

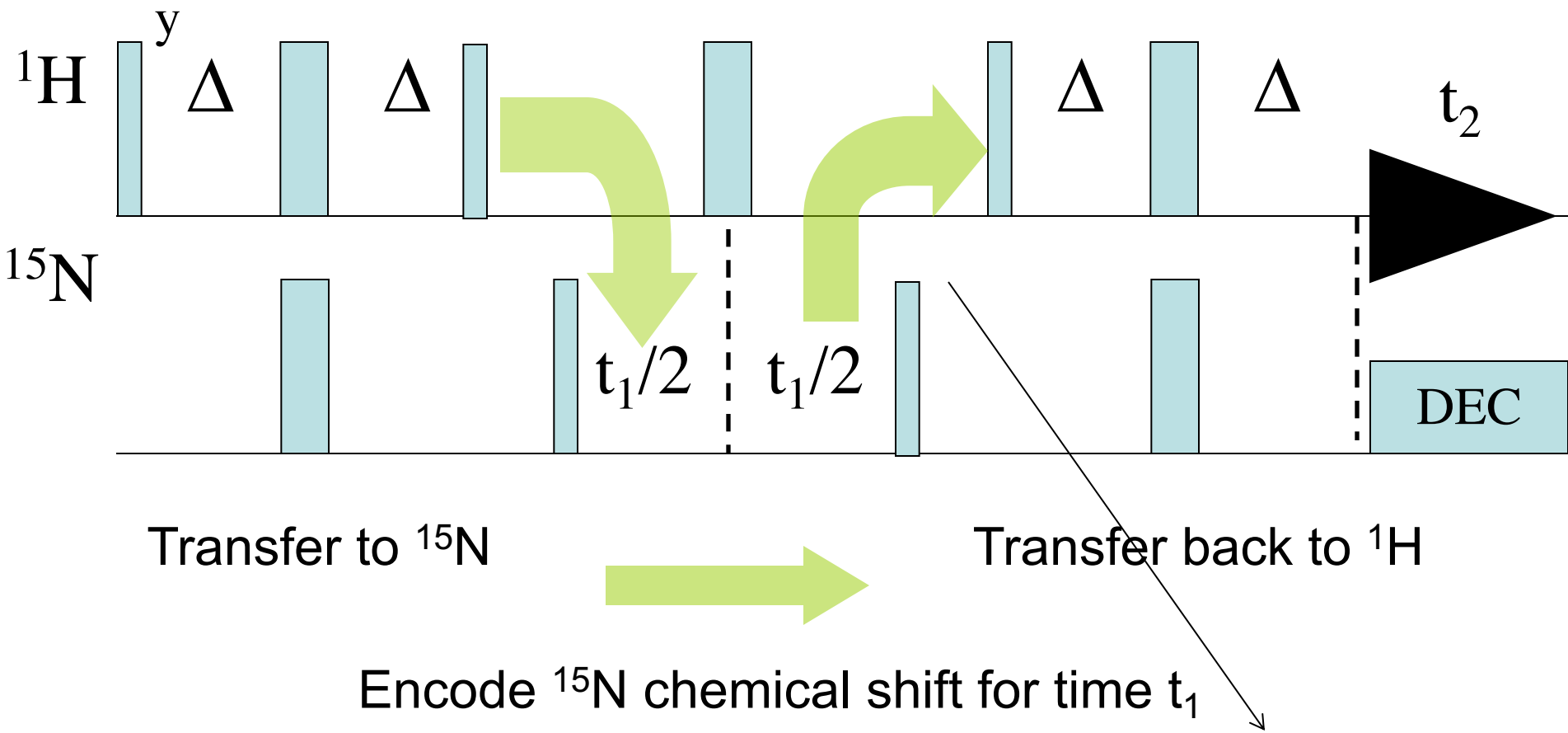
R_{ex}

Slow Exchange Between Two States



$$k_{ex}=k_{open}+k_{close}$$

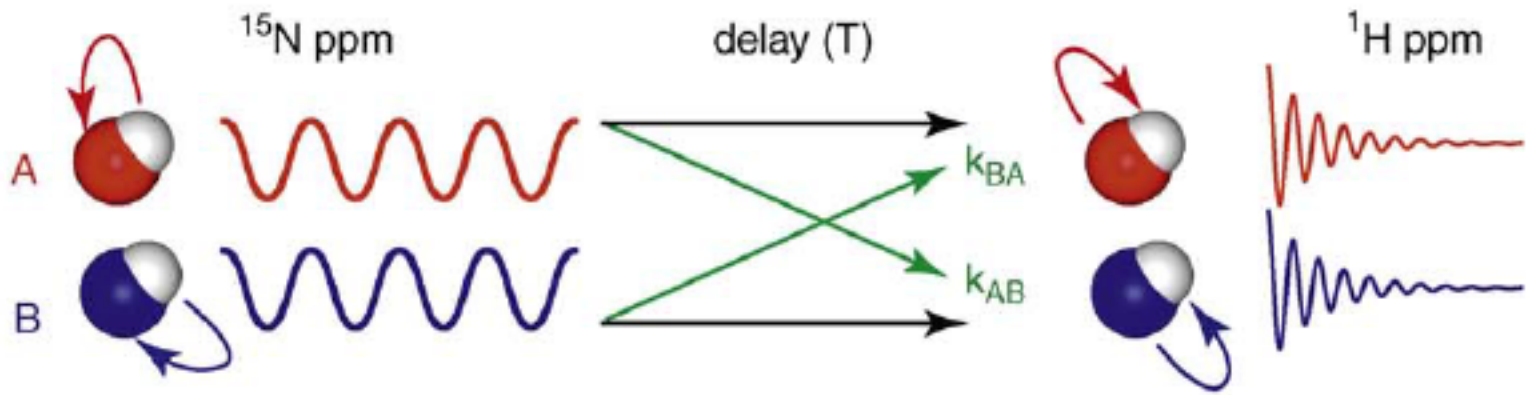
Using HSQC to measure slow exchange



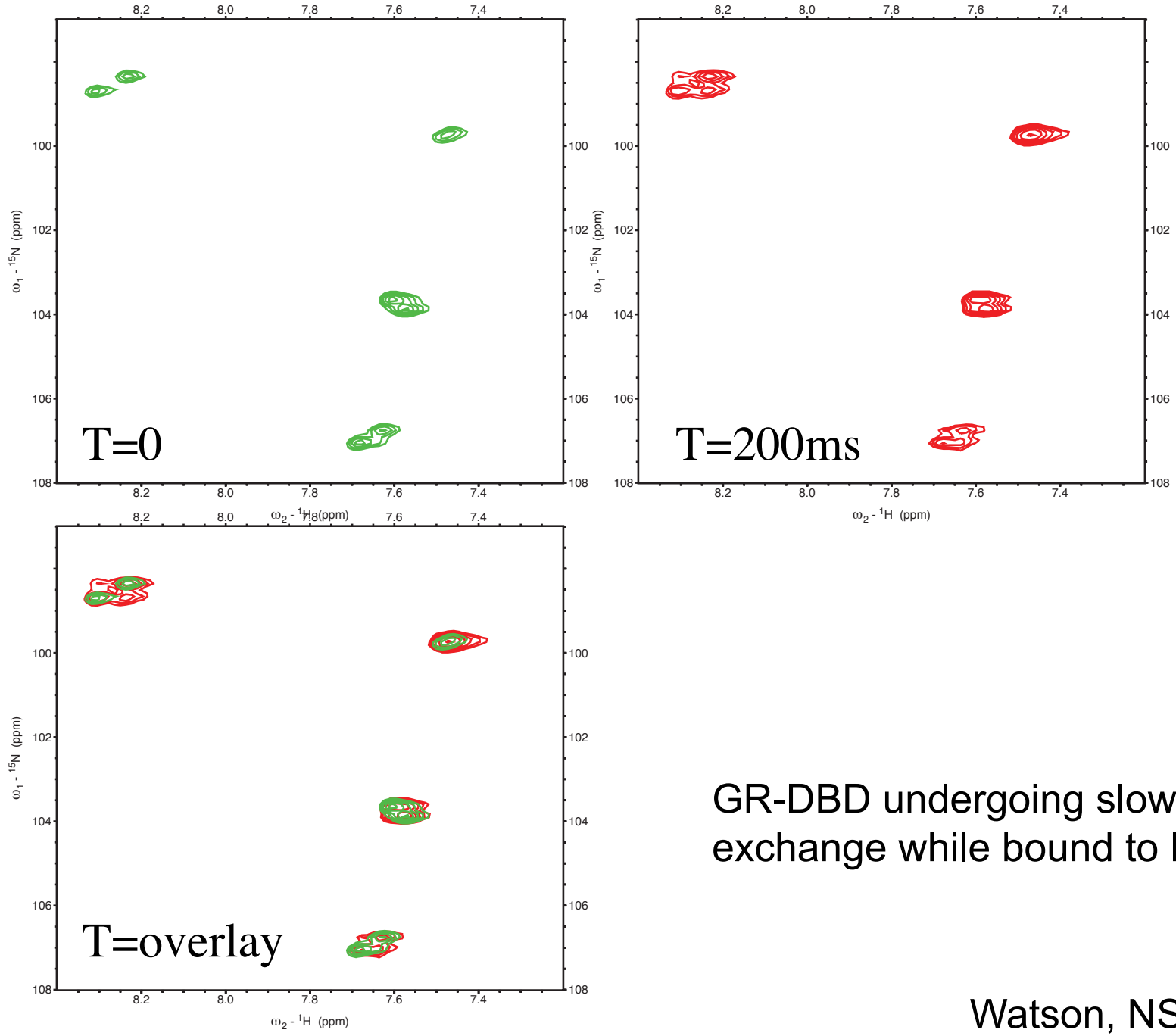
Insert delay to allow exchange

ZZ-exchange, Montelione and Wagner

Exchange Cross-Peaks



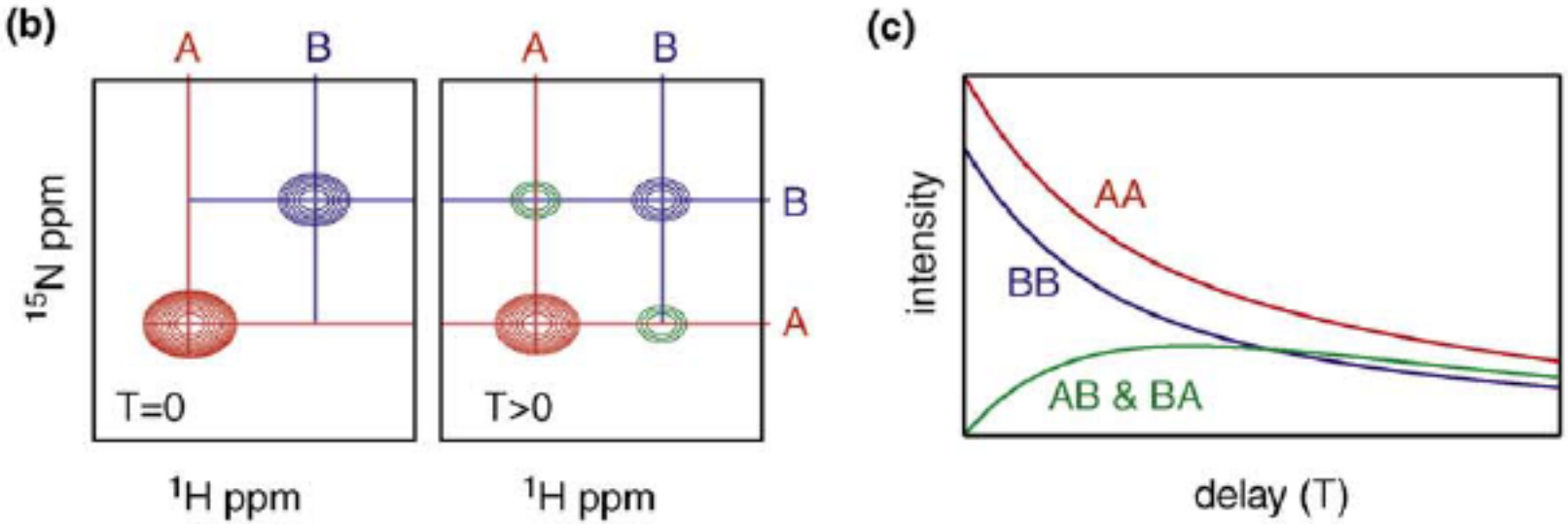
Cross-peaks from a conformational change during delay
e.i.-red to blue



GR-DBD undergoing slow exchange while bound to DNA

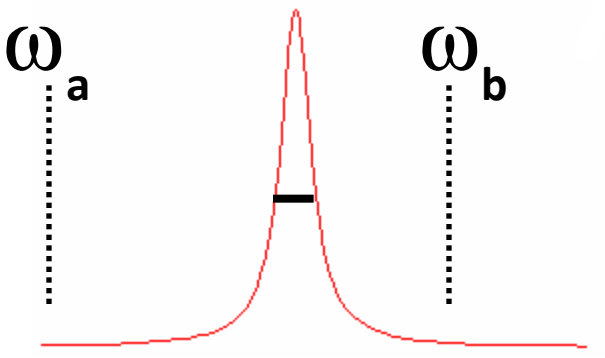
Watson, NSMB 2013

ZZ-exchange peak intensity dependence on delay



Fit to obtain populations and rate constants

Expressions for the linewidth in the Fast Exchange Limit ($k_{ex} > \Delta\omega$)



$$A \underset{k_{rev}}{\overset{k_{for}}{\rightleftharpoons}} B$$

$$\bar{\omega} = p_a \omega_a + p_b \omega_b$$

Populations p_a, p_b

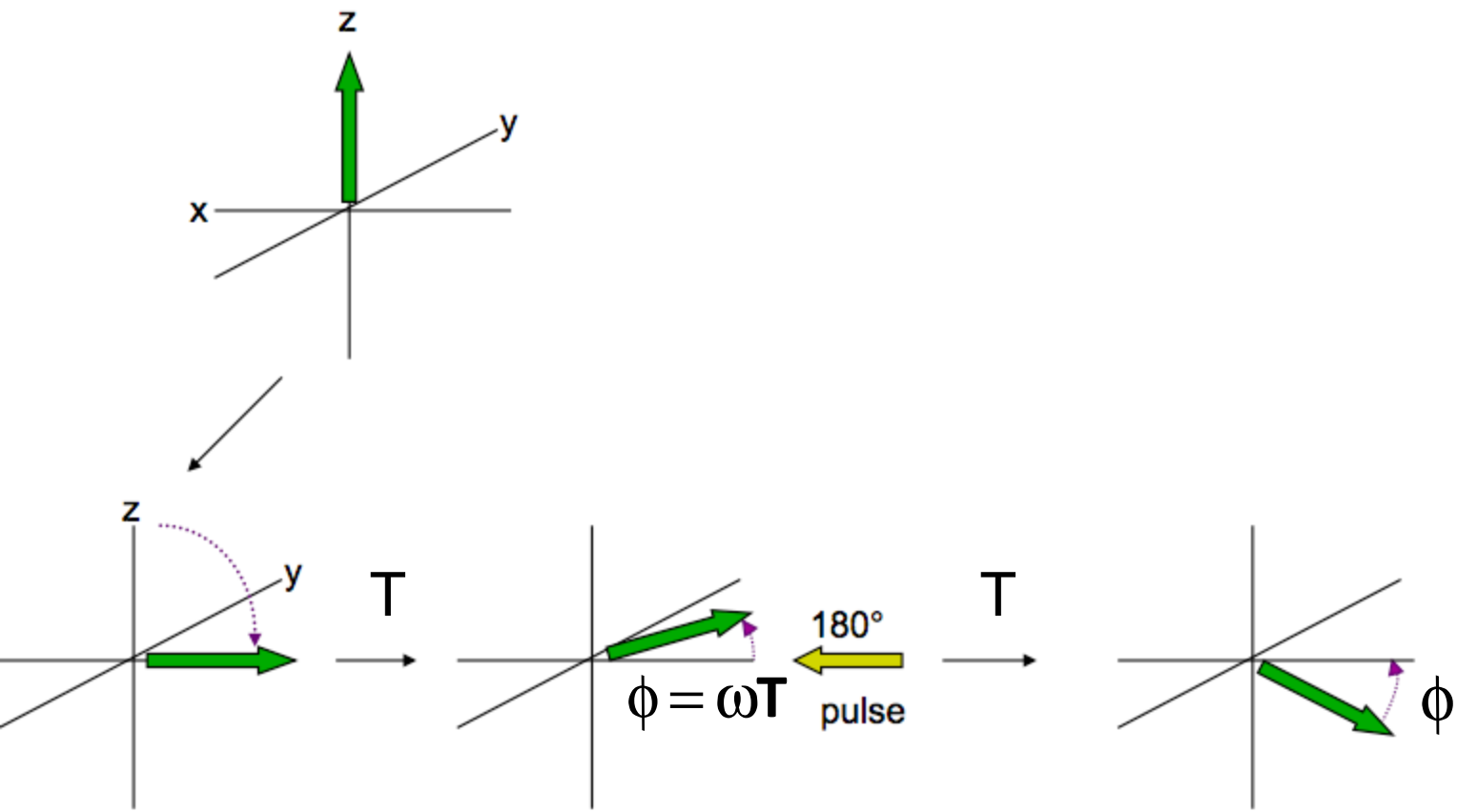
$k_{ex} = k_{for} + k_{rev}$

$$R_2 = \bar{R}_{2\,(ns-ps)} + \left.\frac{p_a p_b \Delta\omega^2}{k_{ex}}\right\} R_{ex}$$

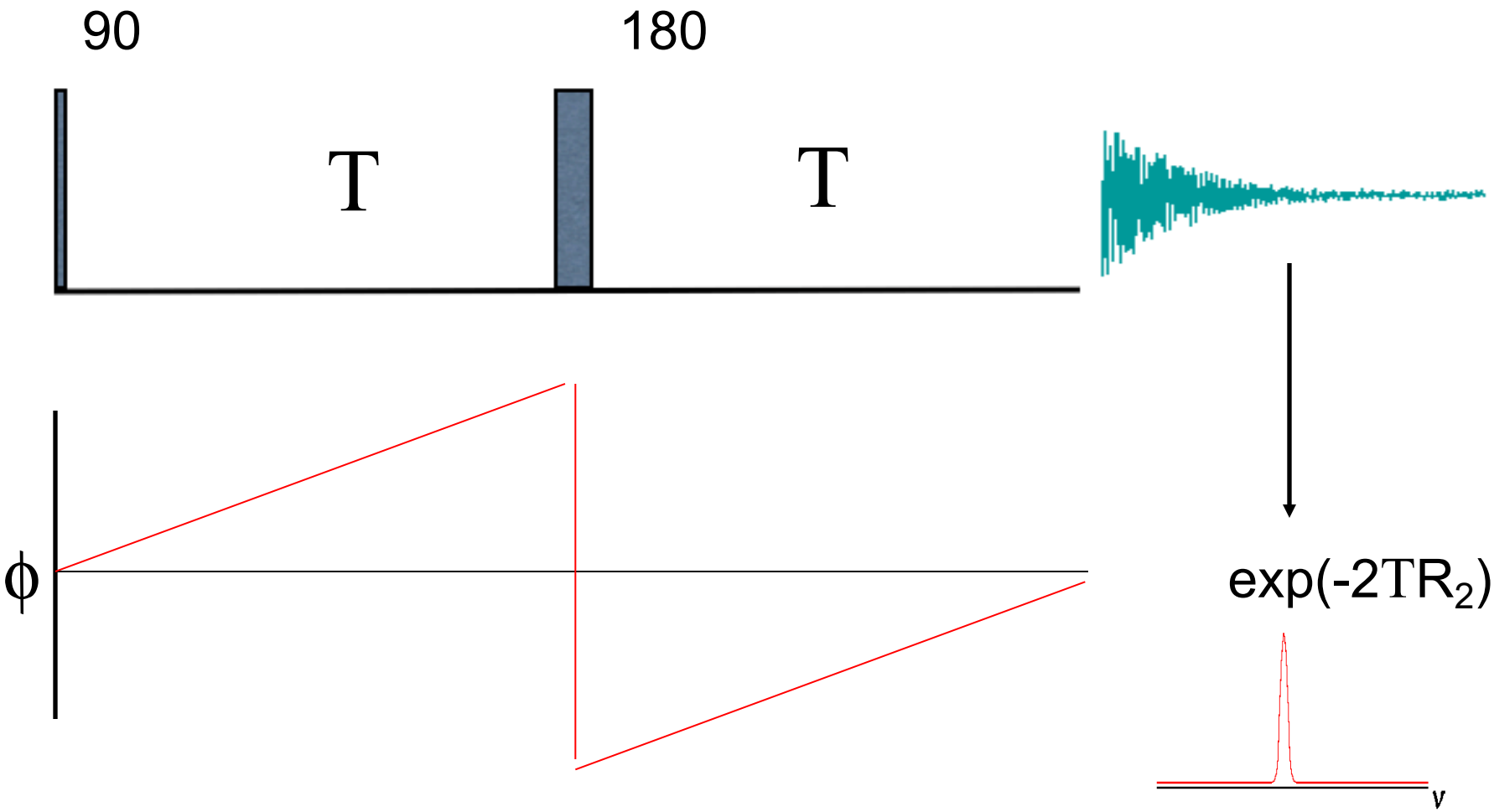
$$\Delta\omega = \omega_a - \omega_b$$

Spin Echo

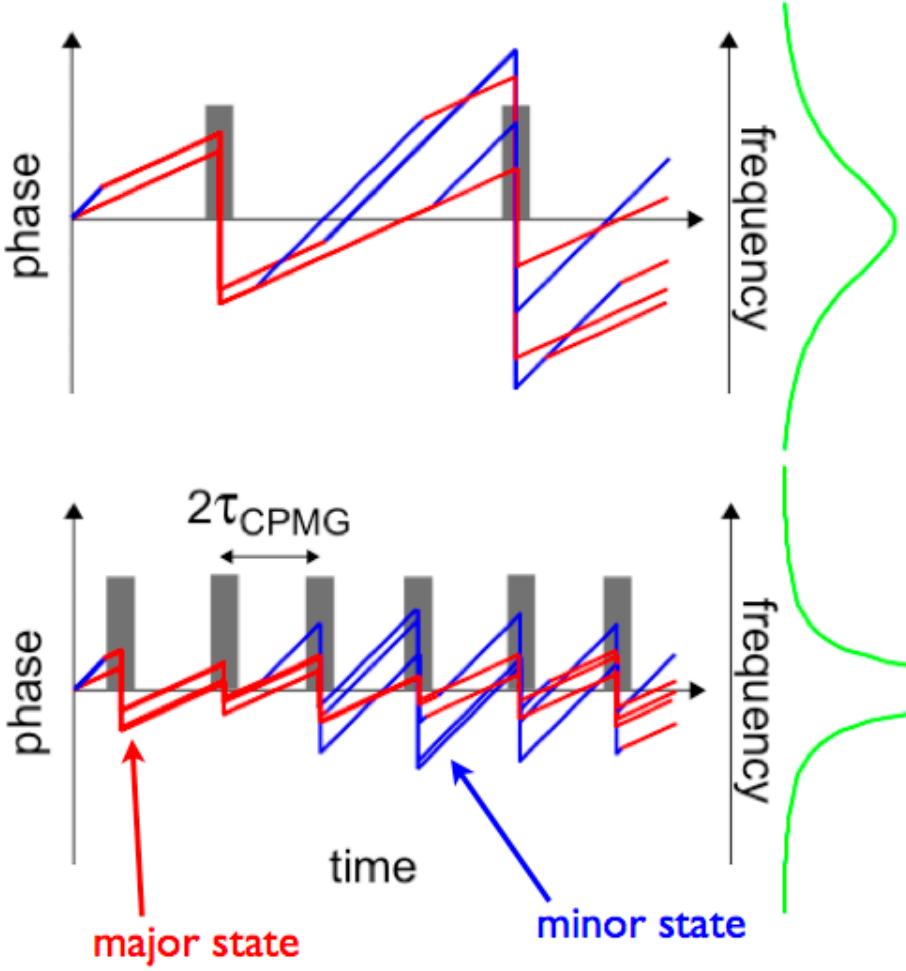
to measure ms-usec dynamics



Spin Echo to Measure R2



cpmg experiment



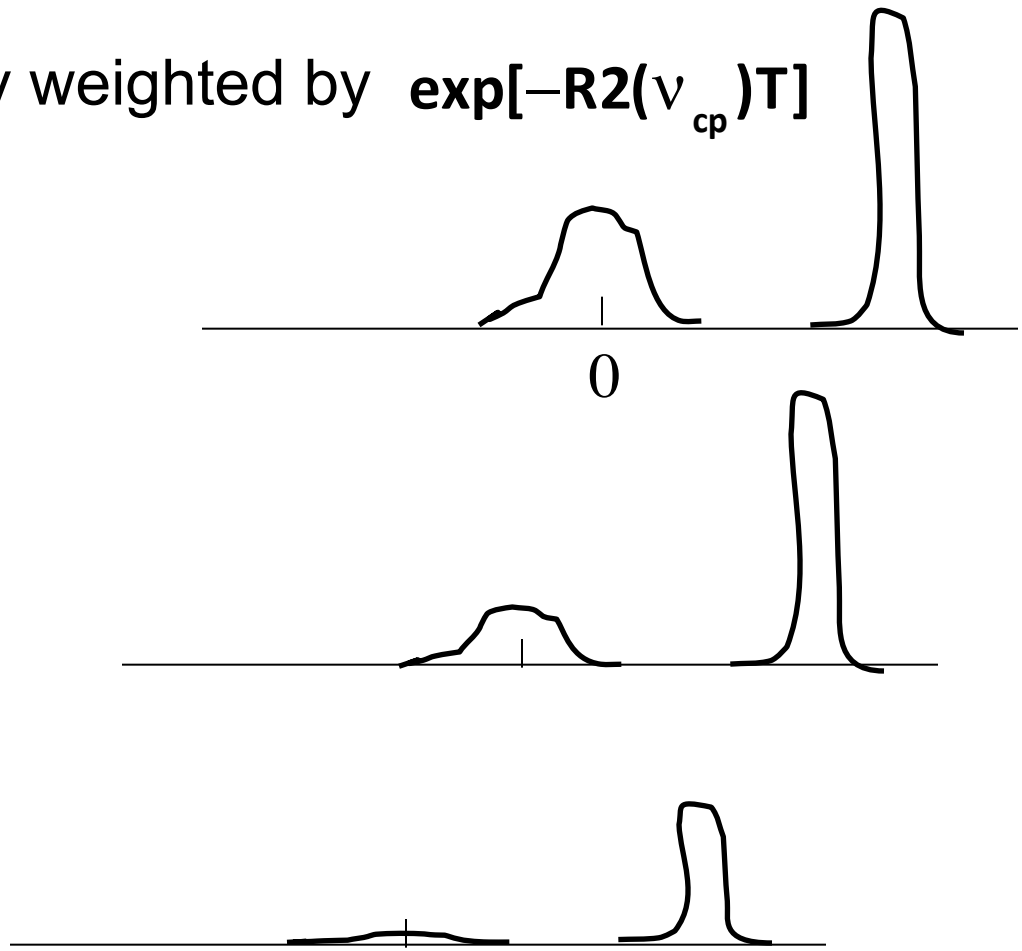
- red and blue lines = phased evolution of spins
- gray = RF pulses
  - refocus magnetization and quench effect of exchange
- Rex proportional to linewidth (green)
- no exchange = perfect refocusing

image: Mikael Akke (Lund University)

CPMG Protocol

Intensity weighted by **exp[−R2(ν_{cp})T]**

$$\nu_{cp} = \frac{1}{2\tau_{cp}}$$

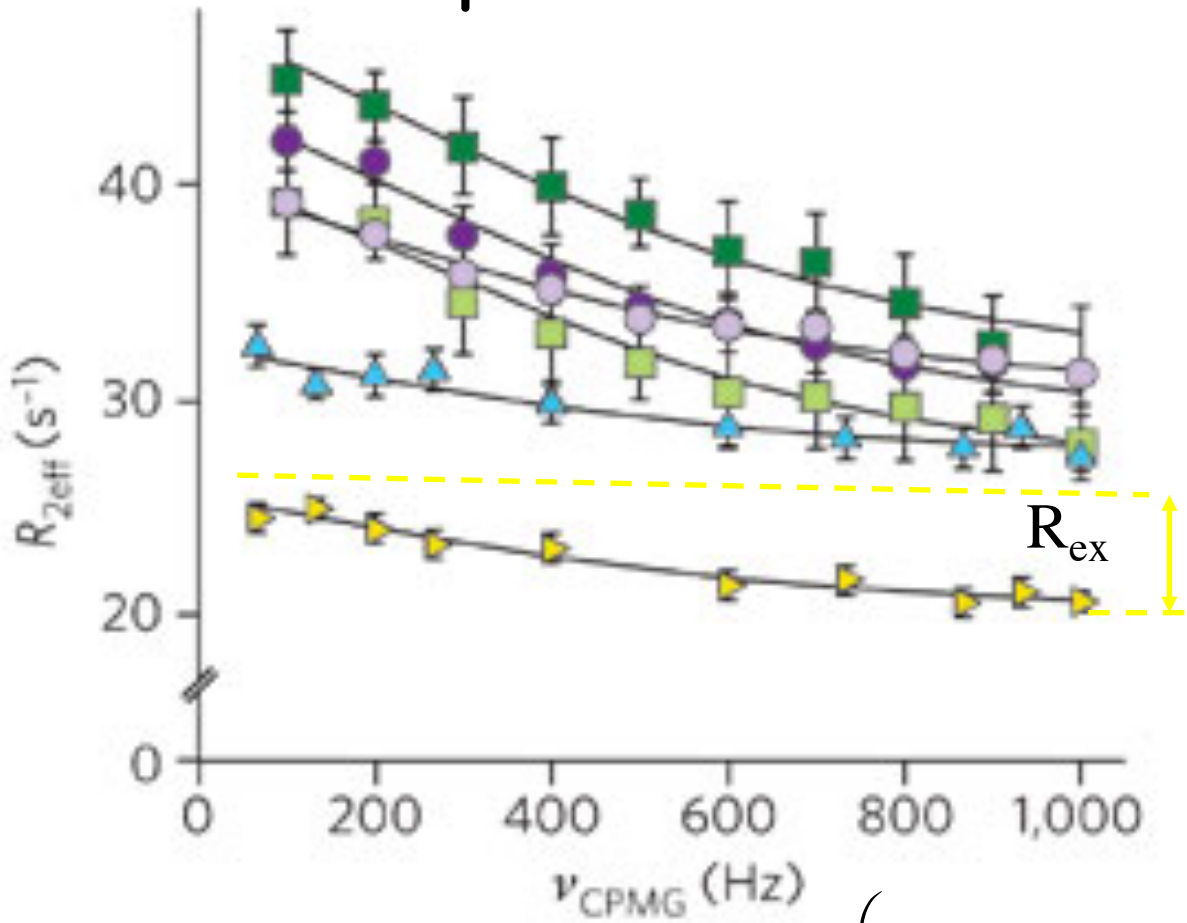


$$\nu_{cp} = 1000\text{Hz}$$

$$\nu_{cp} = 500\text{Hz}$$

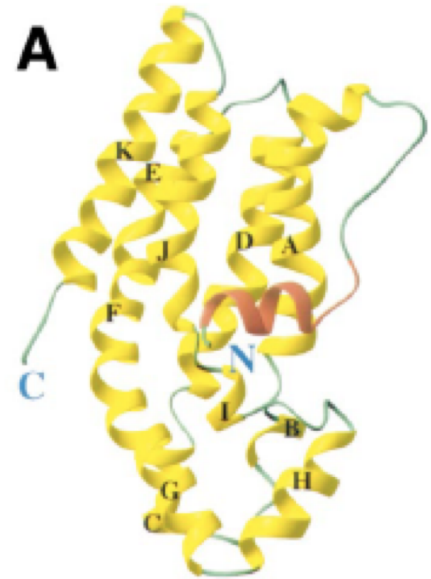
$$\nu_{cp} = 50\text{Hz}$$

Dynamics Constants from Relaxation Dispersion Curves

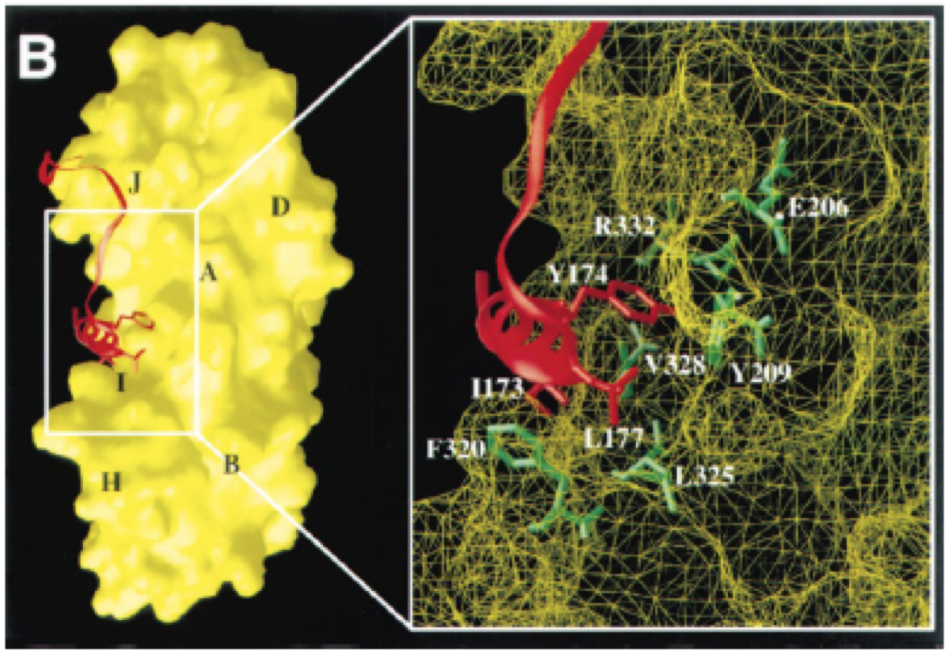


$$R_2^{eff}(\upsilon_{CPMG}) = R_2^a(ns\text{-}ps) + p_a p_b \frac{\Delta\omega^2}{k_{ex}}\left(1 - \frac{4\upsilon_{CPMG}}{k_{ex}}\tanh\frac{k_{ex}}{4\upsilon_{CPMG}}\right)$$

Regulation of Vav1 activity by autoinhibition



How is this GEF activated?



Aghazhadeh, Cell 2000

Example from literature

Internal dynamics control activation and activity of the autoinhibited Vav DH domain

Pilong Li[1,3], Ilídio R S Martins[1–3], Gaya K Amarasinghe[1,3,4] & Michael K Rosen[1]

Nature Structural and Molecular Biology, 15:6 (2008)

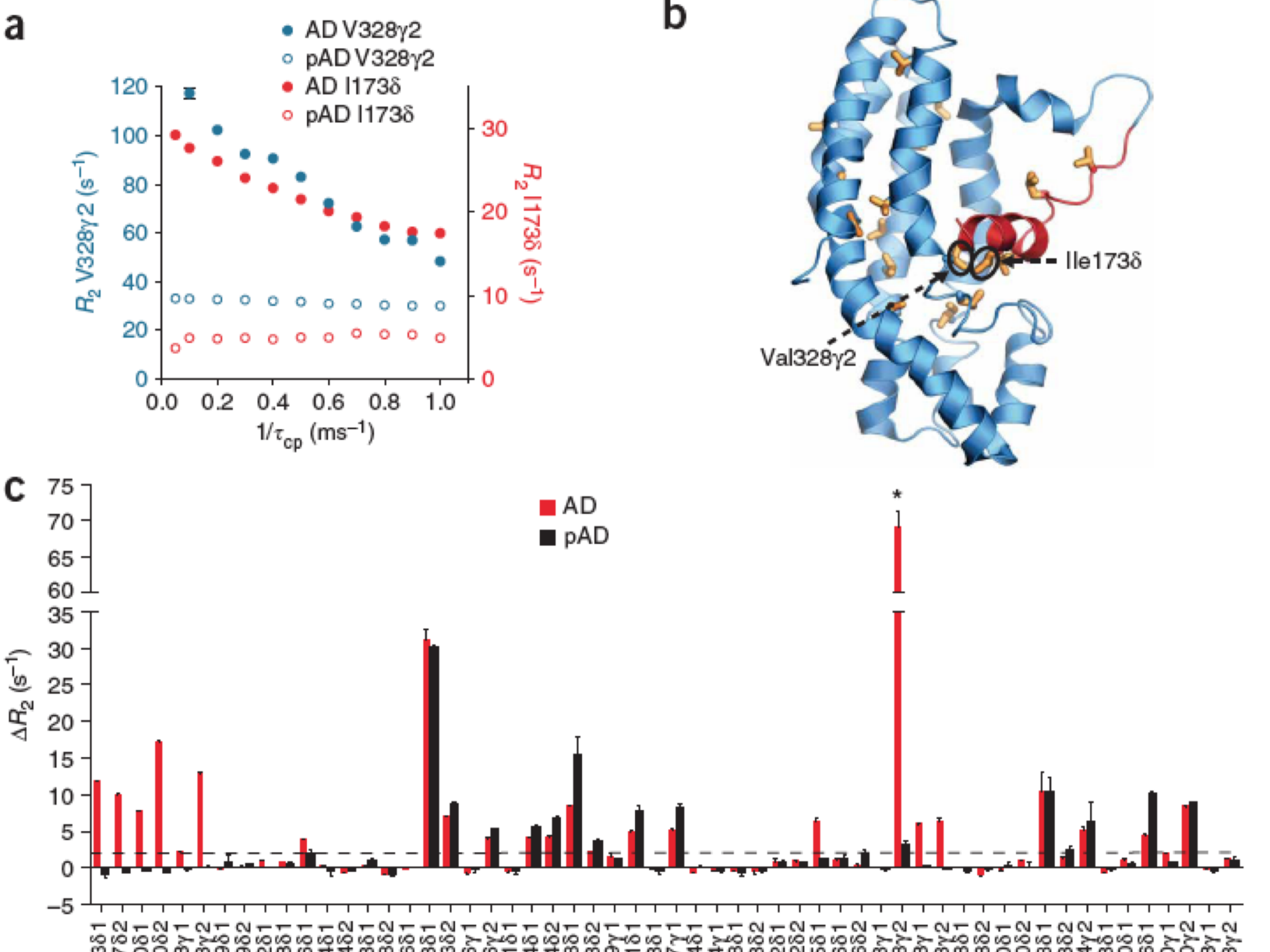
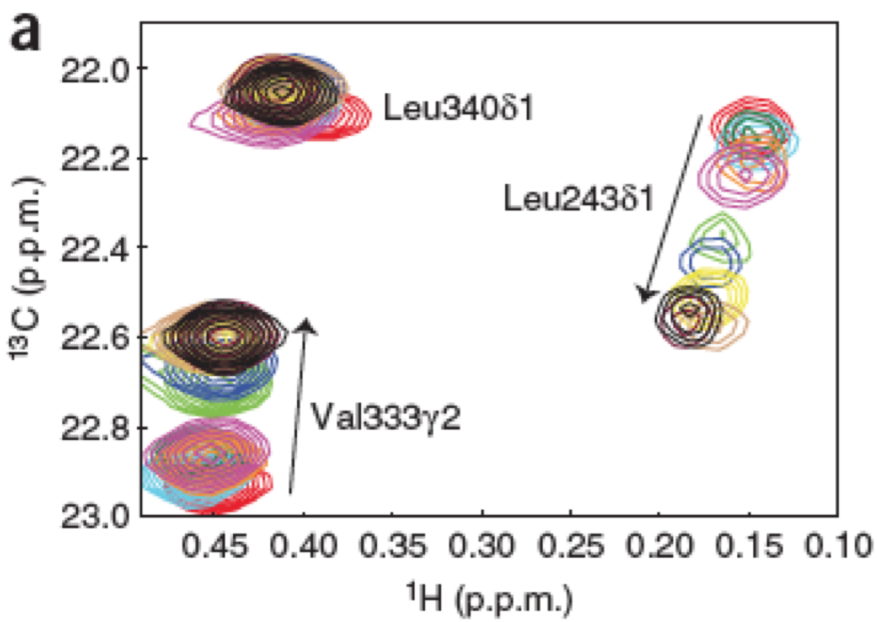

a
AD V328γ2
pAD V328γ2
AD I173δ
pAD I173δ
R_2 V328γ2 (s^{-1})
R_2 I173δ (s^{-1})
1/τ_{cp} (ms^{-1})
b
Ile173δ
Val328γ2
c
AD
pAD
ΔR_2 (s^{-1})

Figure 2



b

Protein	p_O
pWT	~1
K208E	1.00 ± 0.06
Y174F	0.99 ± 0.03
K208S	0.86 ± 0.06
K208A	0.77 ± 0.12
E169K K208A	0.64 ± 0.06
K208V	0.26 ± 0.10
E169K K208V	0.14 ± 0.08
Wild type	0.09 ± 0.03
E169K	0.05 ± 0.06
E169K E207K	~0

Figure 3

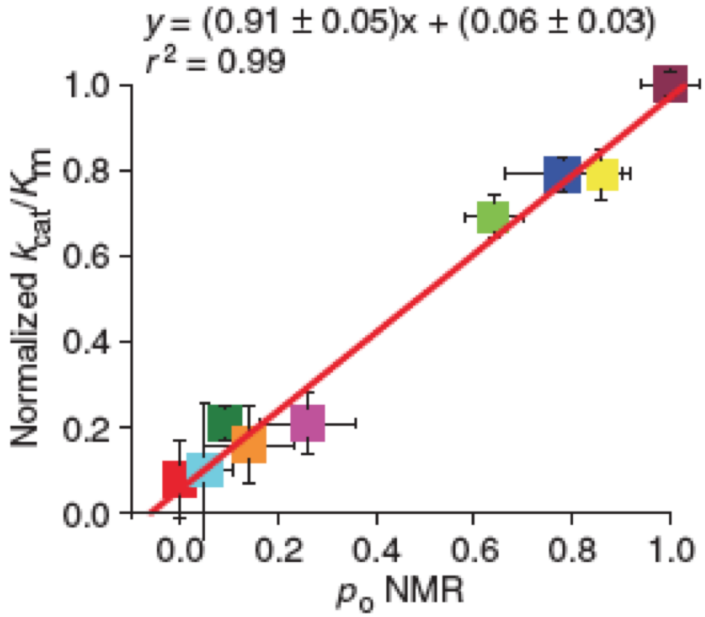
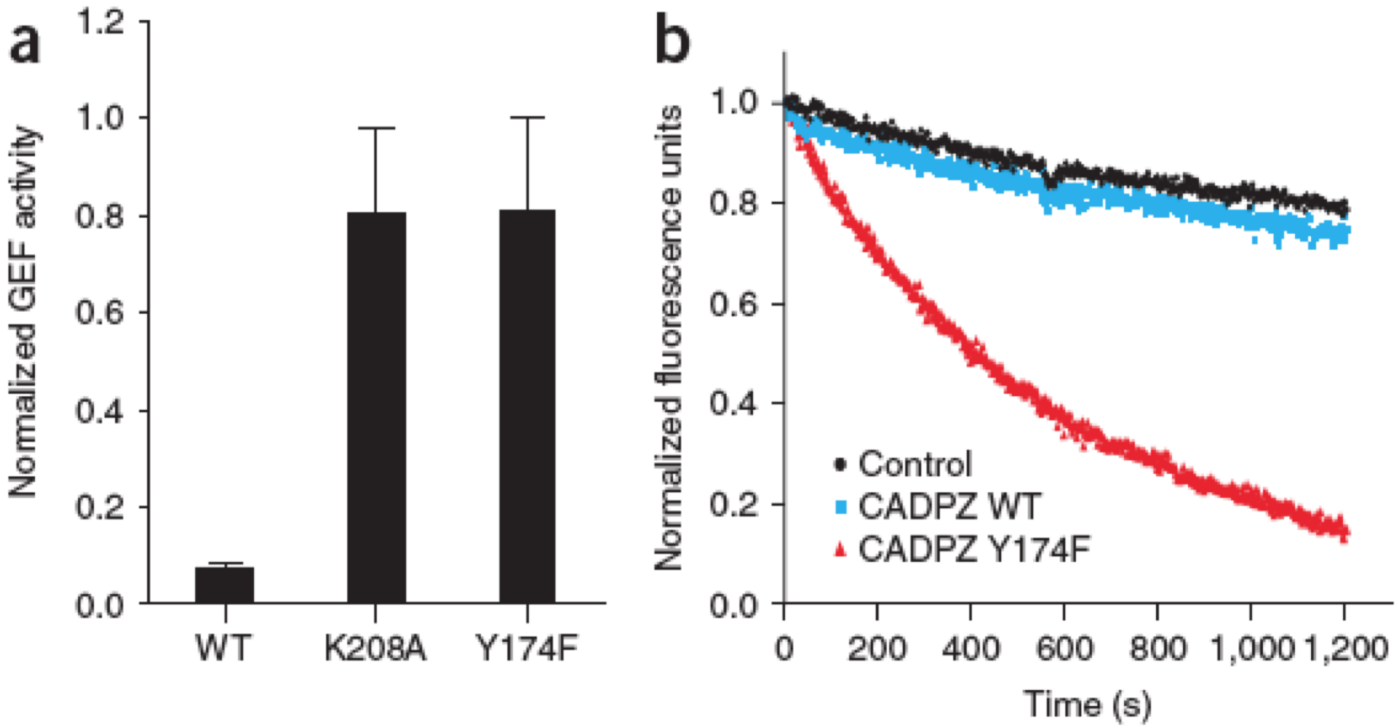
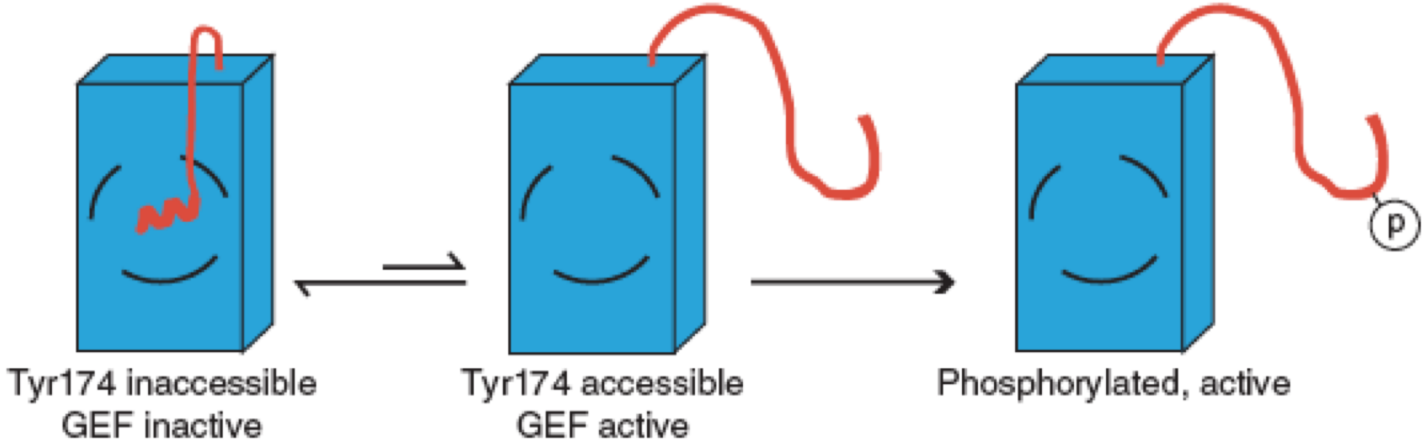

y = (0.91 ± 0.05)x + (0.06 ± 0.03)
r² = 0.99
Normalized kcat/Km
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
po NMR

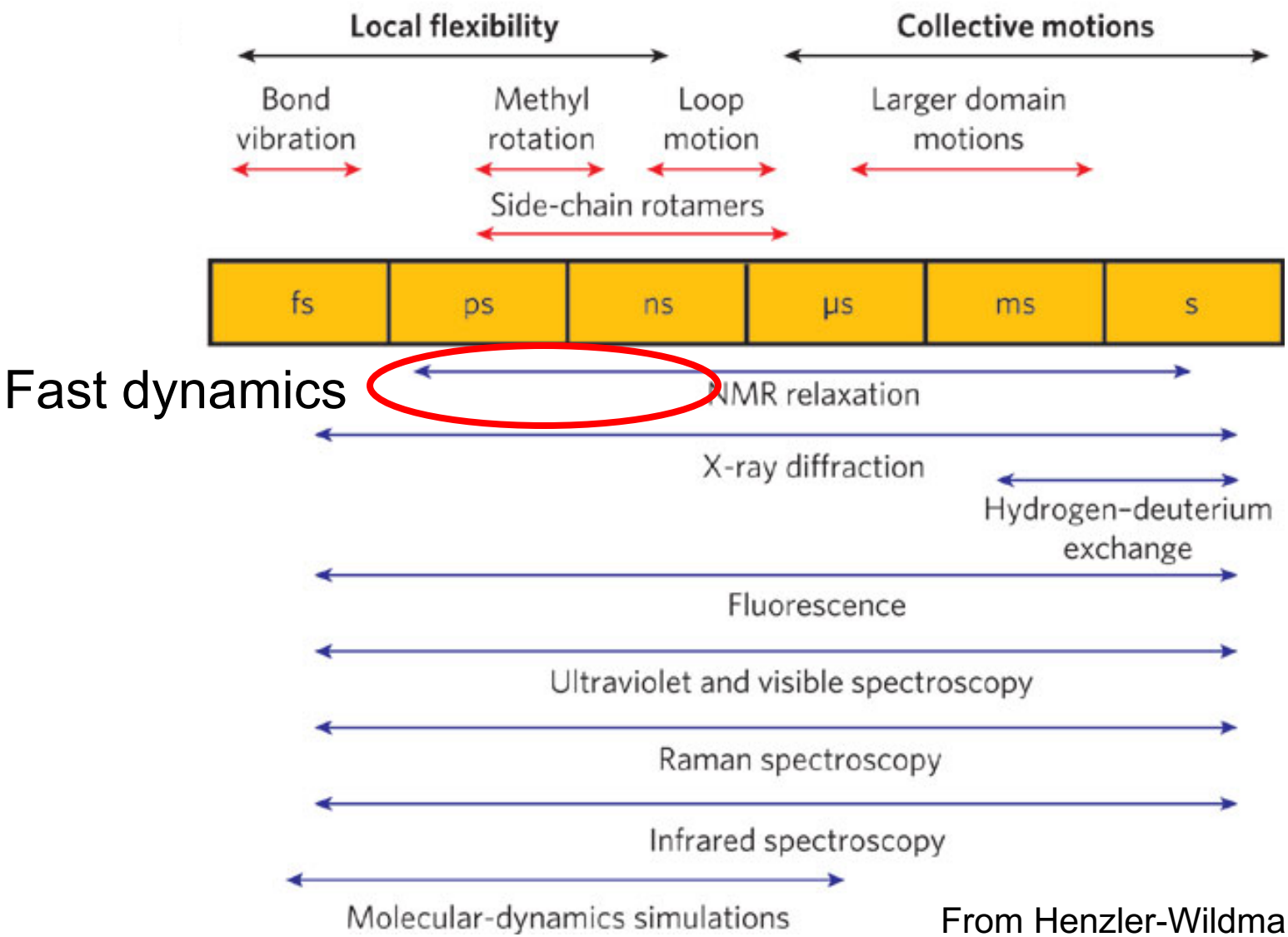
Figure 4


a
Normalized GEF activity
1.2
1.0
0.8
0.6
0.4
0.2
0.0
WT
K208A
Y174F
b
Normalized fluorescence units
1.0
0.8
0.6
0.4
0.2
0.0
Control
CADPZ WT
CADPZ Y174F
0
200
400
600
800
1,000
1,200
Time (s)

Model

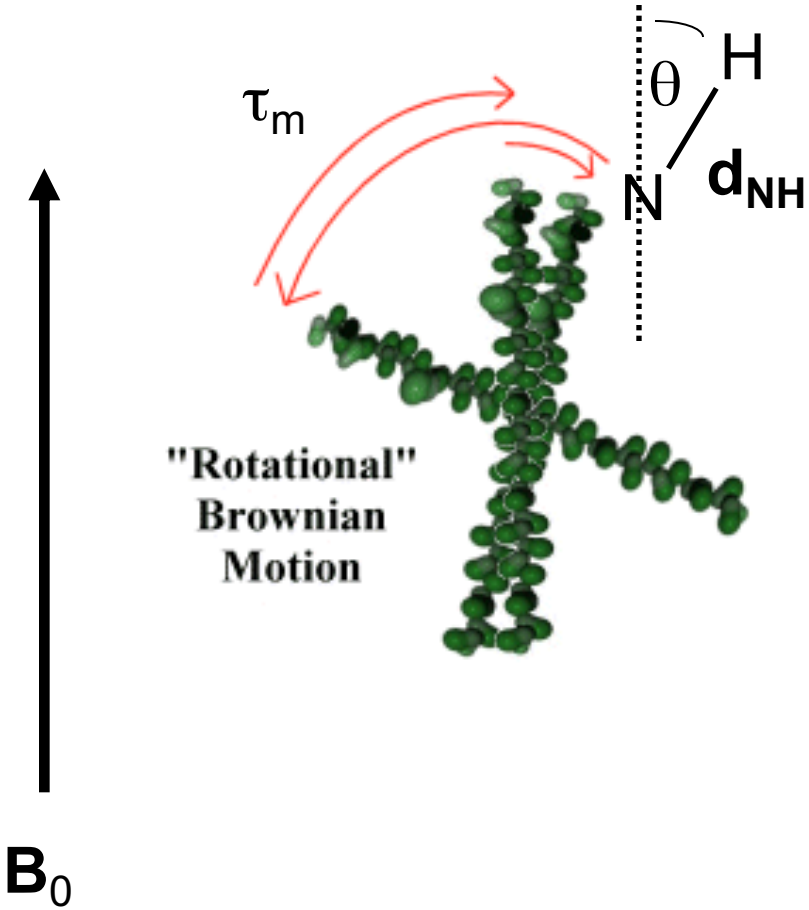

P
Tyr174 inaccessible
GEF inactive
Tyr174 accessible
GEF active
Phosphorylated, active

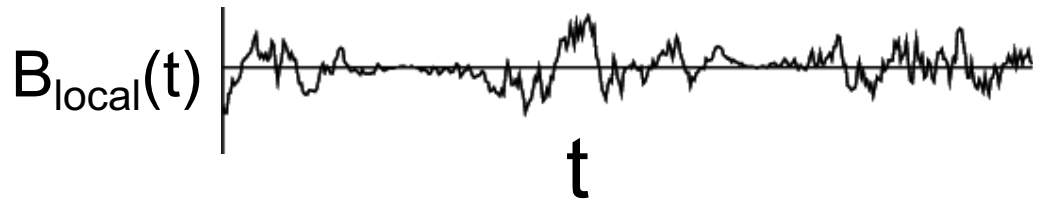
Timescales of Protein Dynamics



From Henzler-Wildman and Kern, Nature 2007

A Major Source of Relaxation is Brownian Rotational Diffusion

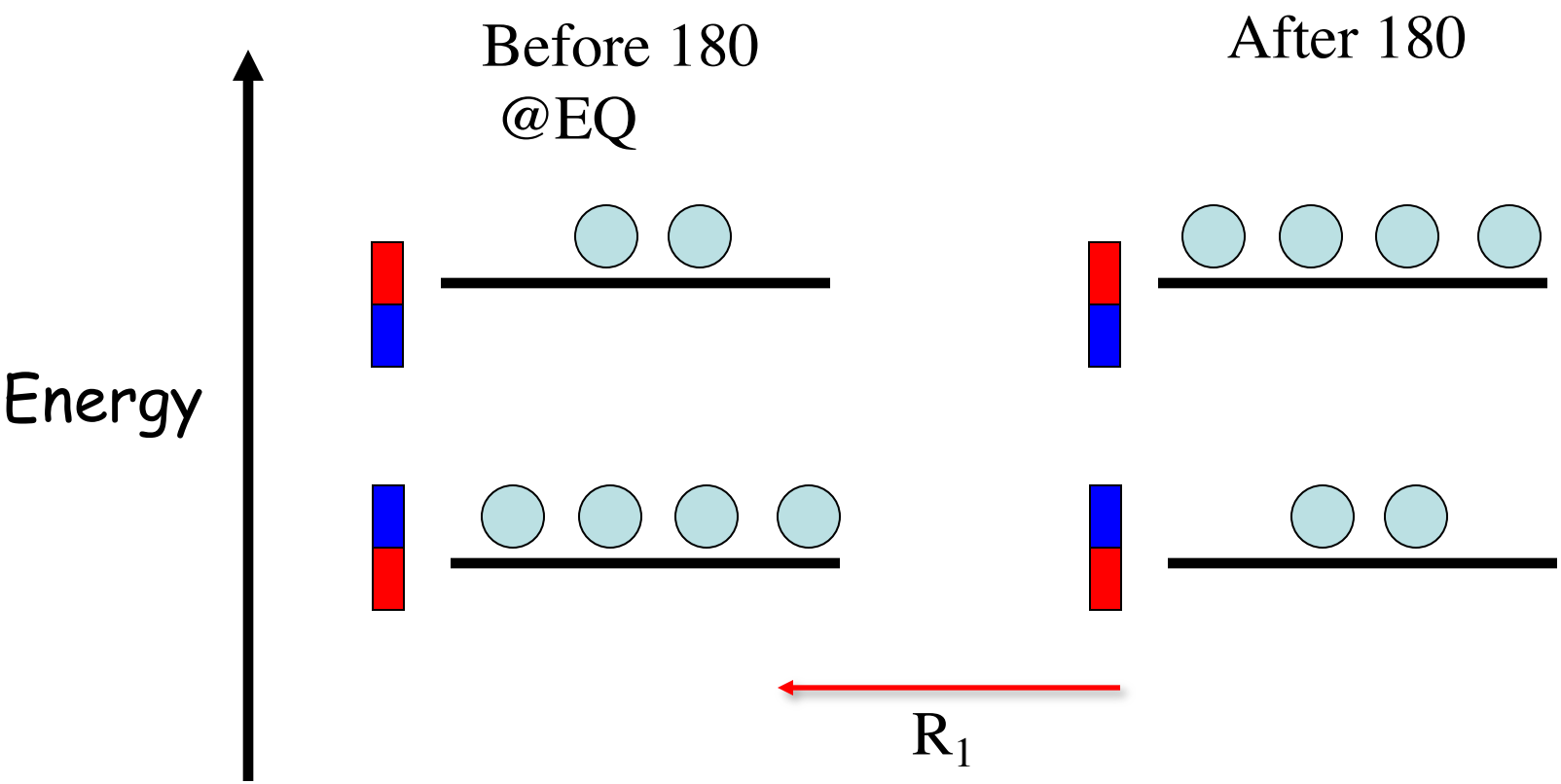




τ_m : rotational correlation time--the time to rotate through one radian.

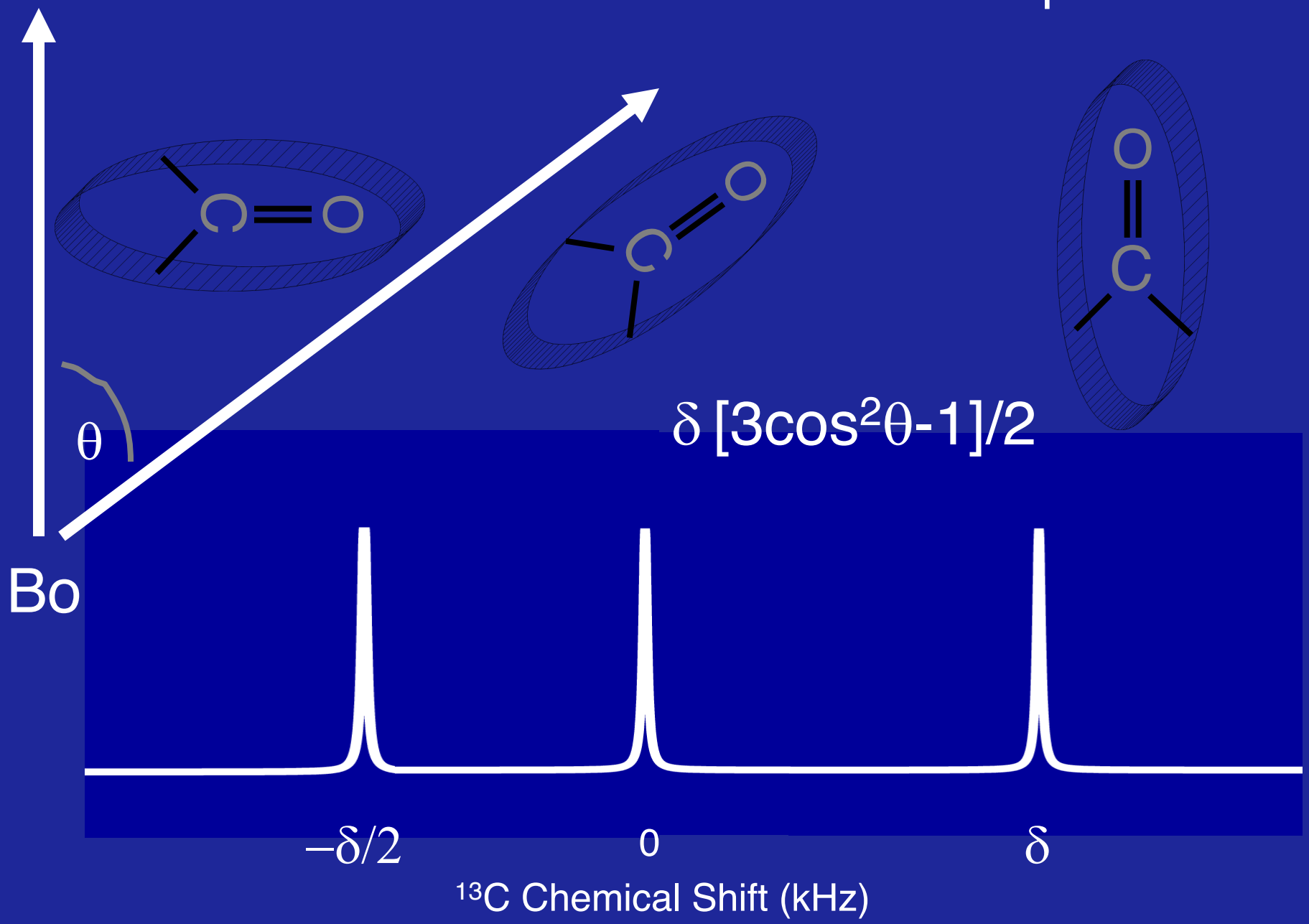
e.g. 20 ns for 40 kDa protein

A 180 degree pulse inverts the population distribution

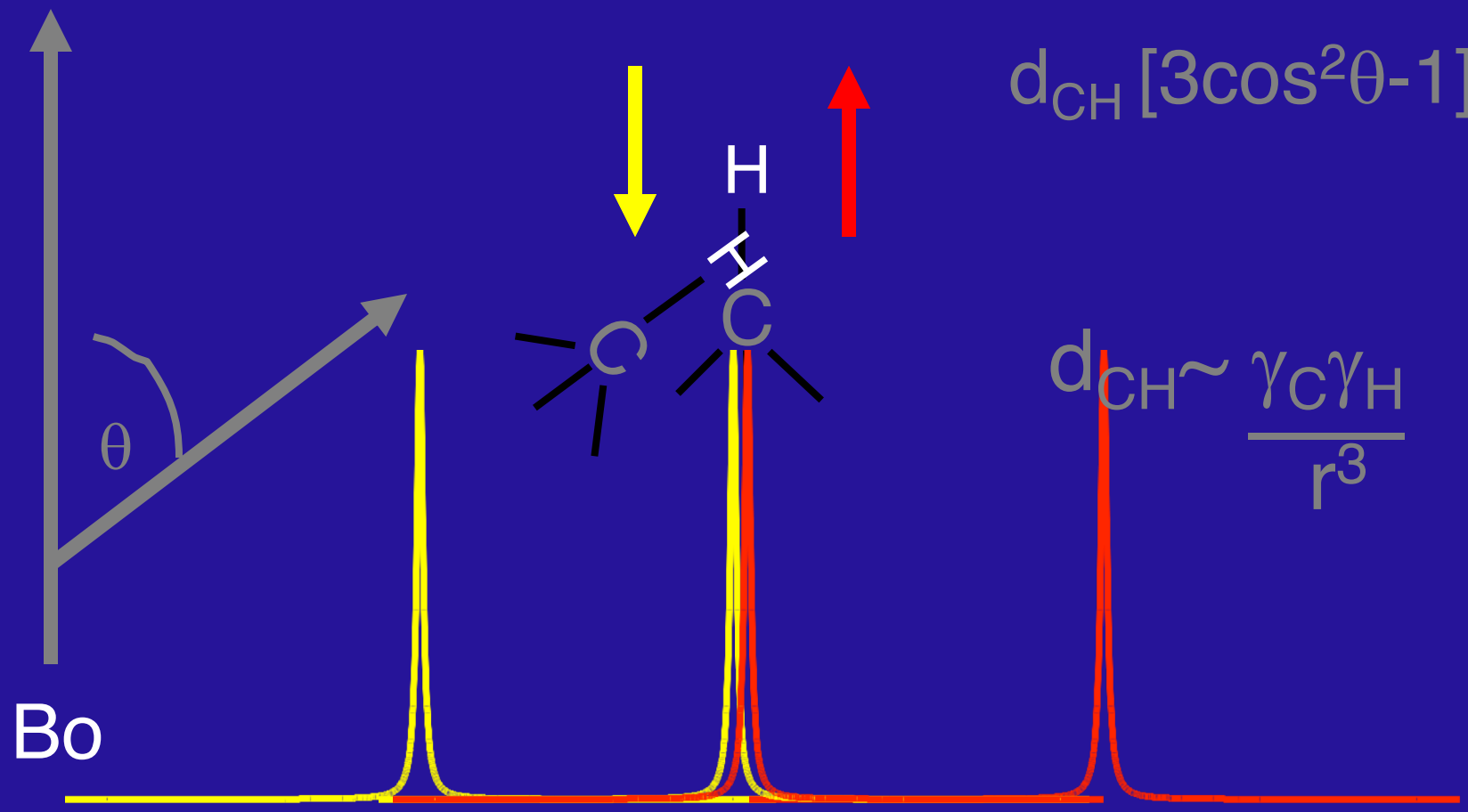


is a measure of how populations return to equilibrium

Chemical Shift Interaction is Anisotropic



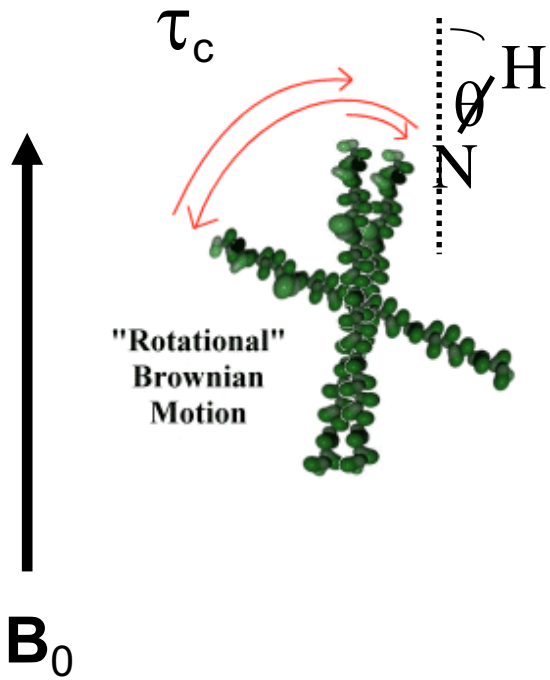
Dipolar Coupling Interaction is Anisotropic

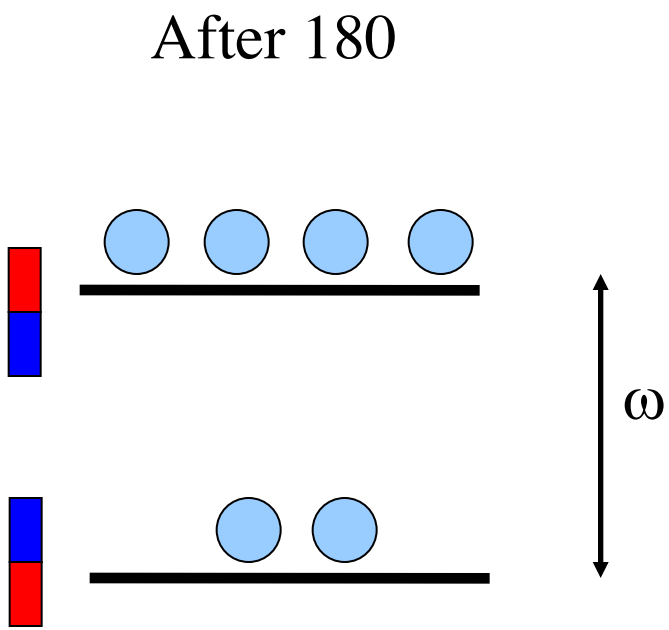


$$d_{CH}\,[3\cos^2\theta - 1]/2$$

$$d_{CH} \sim \frac{\gamma_C \gamma_H}{r^3}$$

The "Frequency" Dependence of Relexation Rates, R1 example





Efficient relaxation if $1/\tau_c=\omega$

Relaxation Rates Depend on Amplitude and Frequency of Local Field Fluctuations

$$R_1(N) = c^2 J(\omega)$$

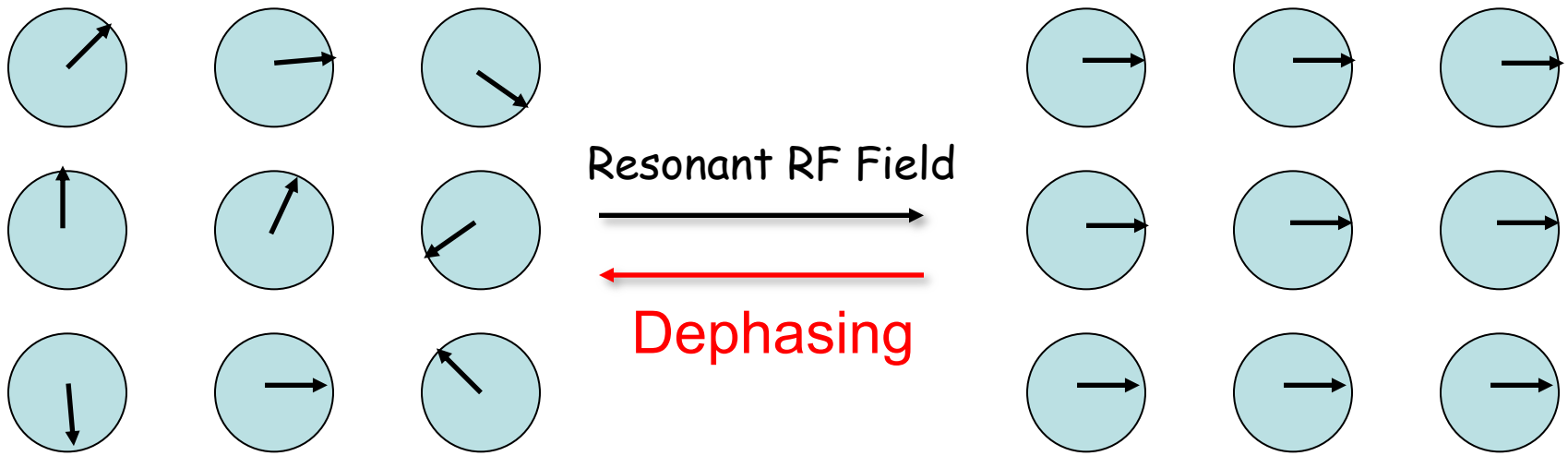
Square of fluctuating local field

Spectral Density Function

$$J(\omega) = \frac{\tau_m}{1 + (\omega\tau_m)^2}$$

Dephasing depends on ns-ps dynamics too

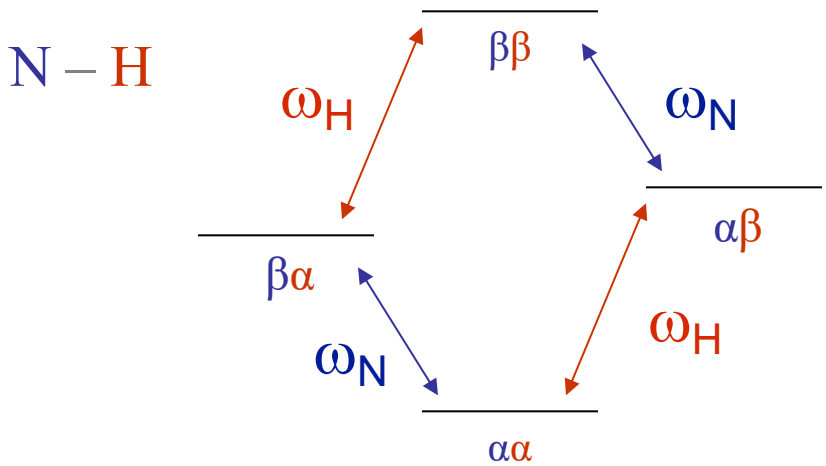
Ensemble of Nuclear Spins



Random Phase
No NMR Signal

Phase Synchronization
NMR Signal!

^{15}N-^{1}H spin pair has four states



Spectral Density Functions

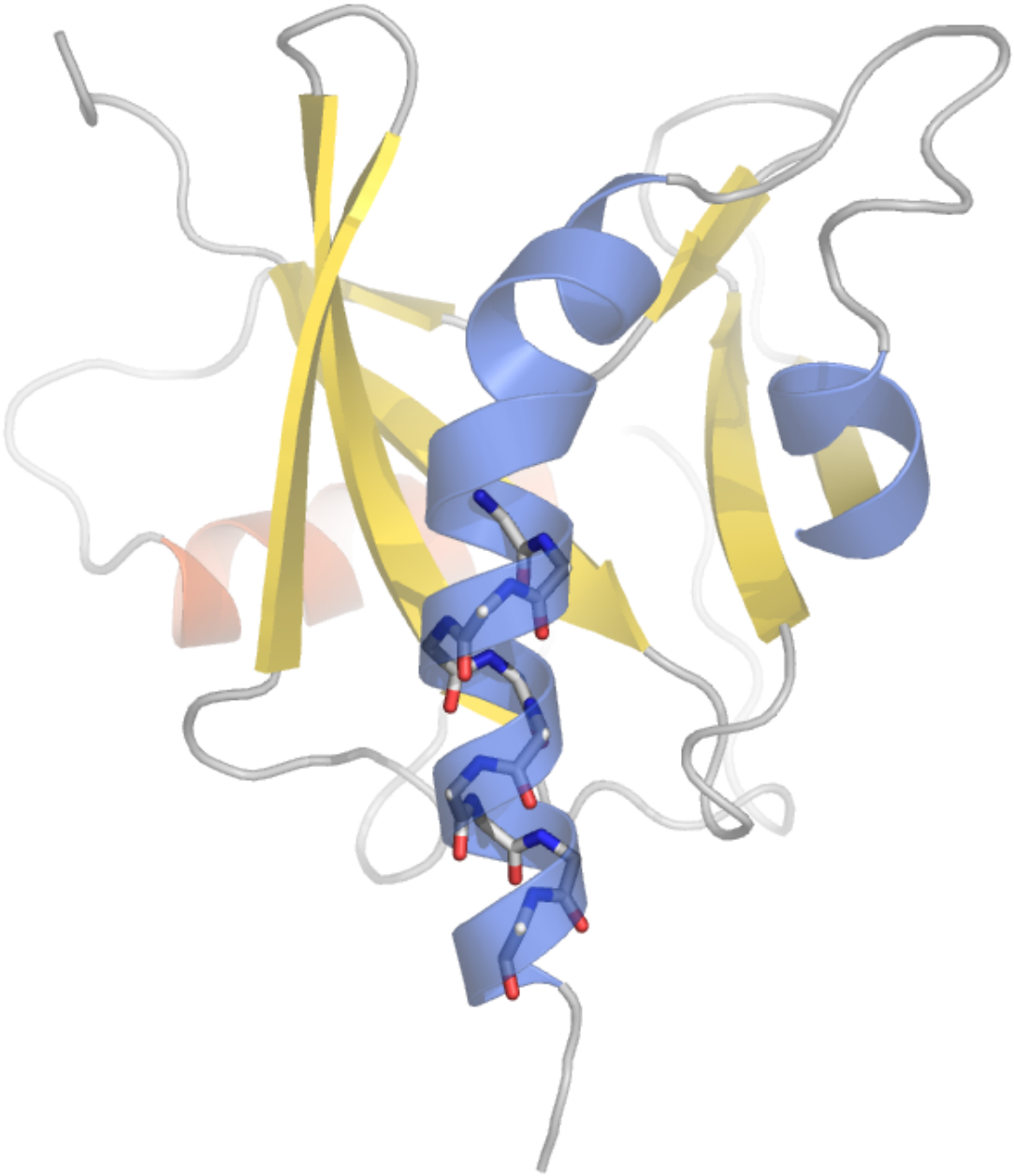
$$R_1 = \frac{d^2}{4}\left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\right] + c^2 J(\omega_N)$$

$$R_2 = \frac{d^2}{8}\left[4J(0) + J(\omega_H + \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H - \omega_N)\right]$$
$$+ \frac{c^2}{6}\left[3J(\omega_N) + 4J(0)\right]$$

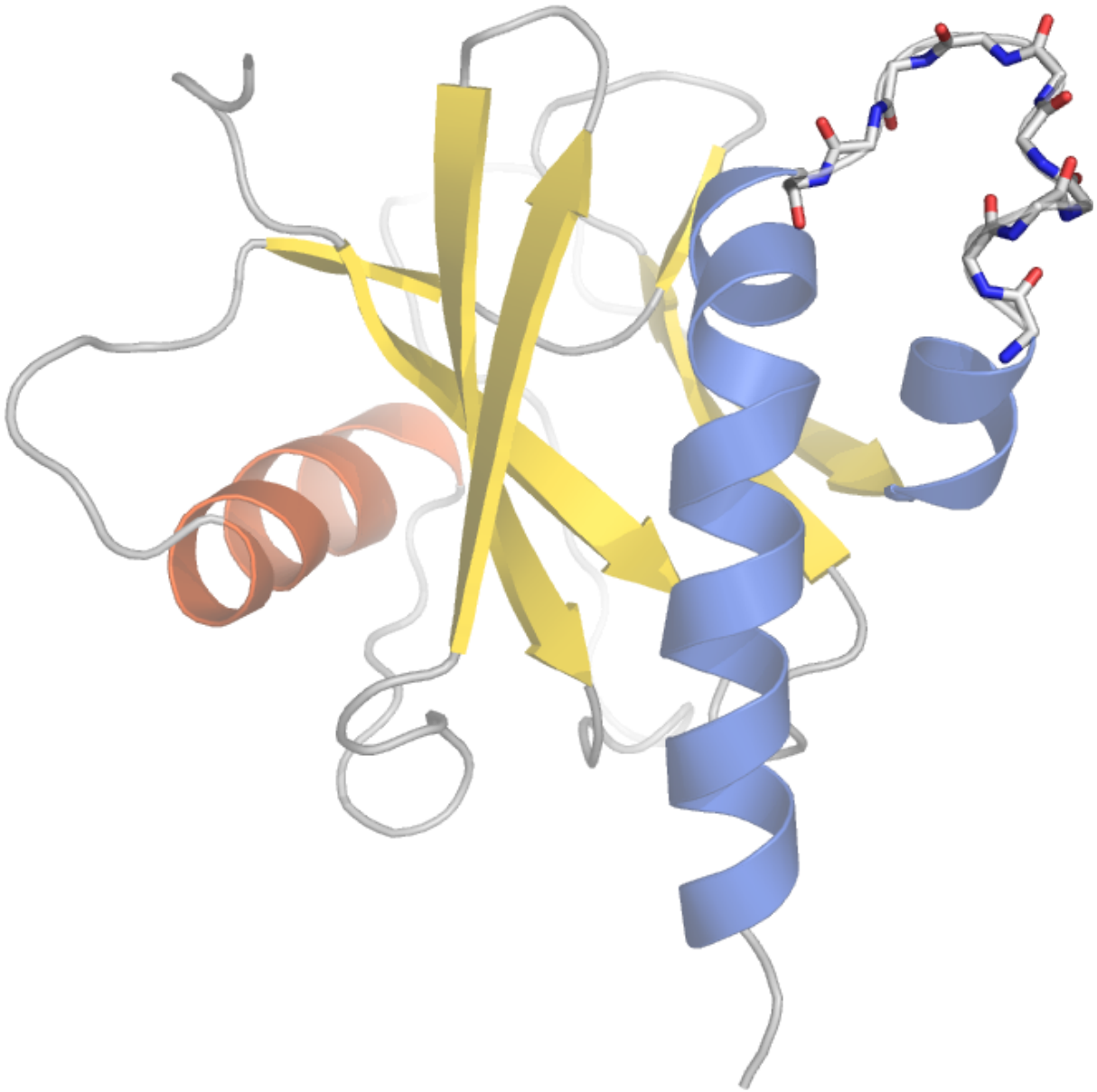
where $d = \left(\frac{\mu_0 h \gamma_N \gamma_H}{8\pi^2}\right)\left\langle \frac{1}{r_{NH}^3} \right\rangle$ $\qquad c = \Delta\left(\frac{\omega_N}{\sqrt{3}}\right)$

Farrow et.al, (1995) *J. Biomol. NMR* **6**, 153

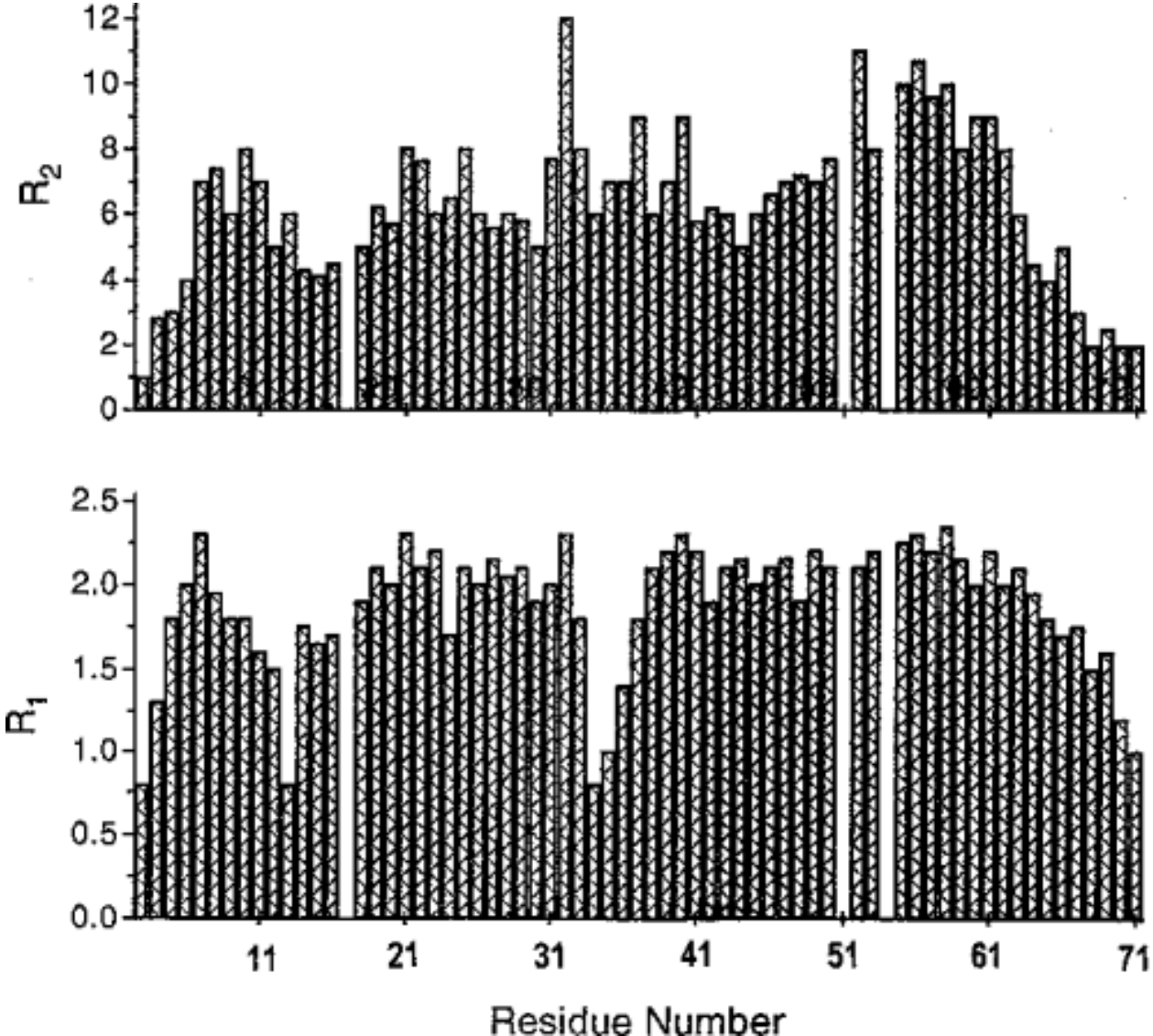
Rigid amide groups



Detecting mobile amide groups



R1 and R2 are not uniform

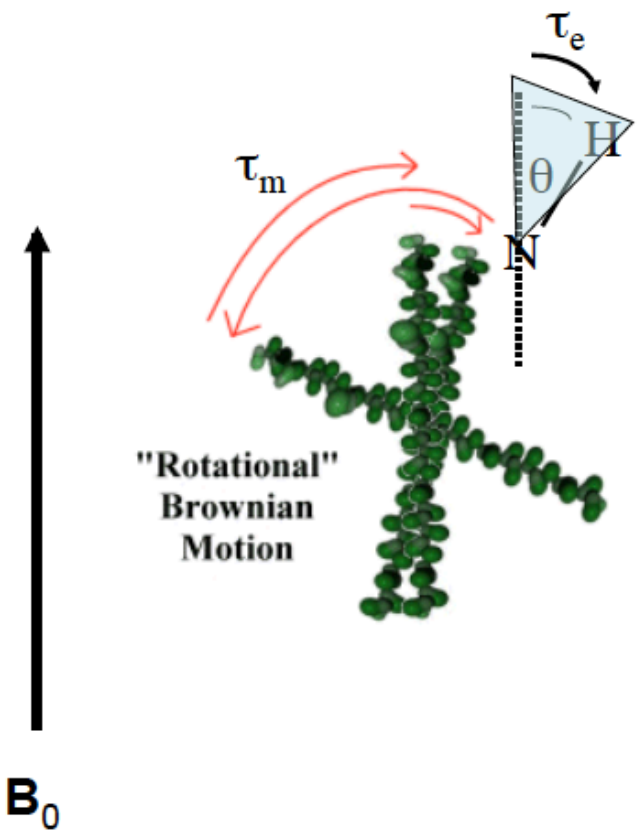


Model Free formalism accounts for internal motions

Lipari-Szabo (Model Free)

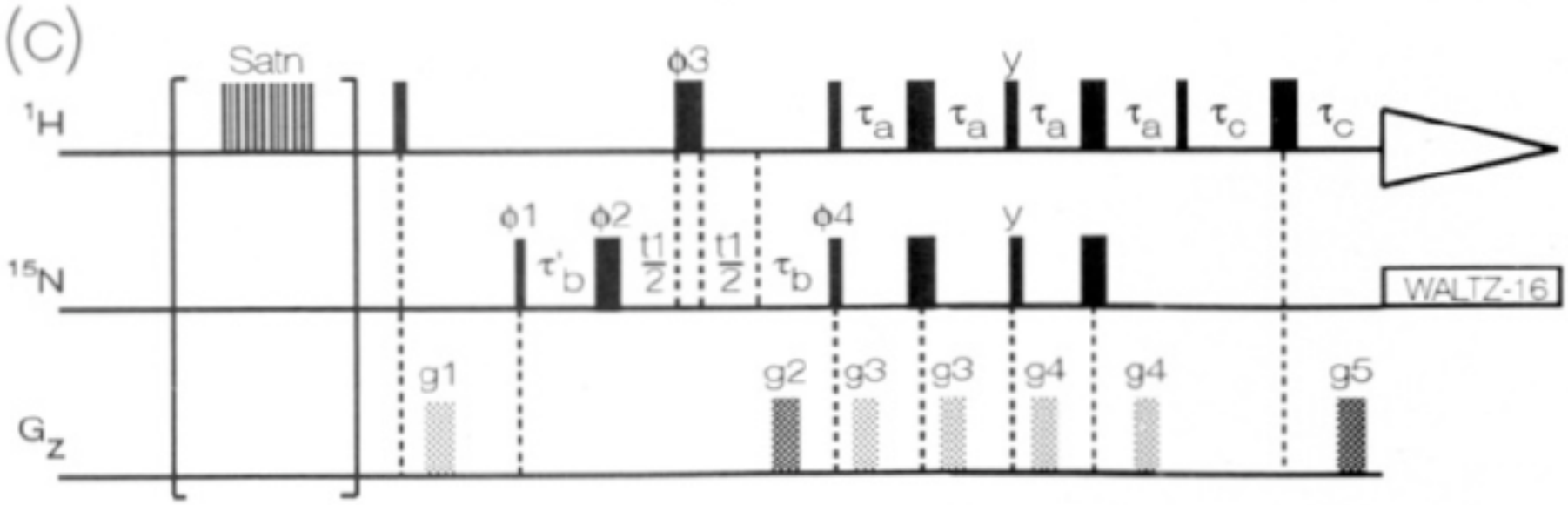
$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{\left(1+\omega^2\tau_m^2\right)} + \frac{\left(1-S^2\right)\tau}{\left(1+\omega^2\tau^2\right)}\right]$$

where $$\frac{1}{\tau} = \frac{1}{\tau_e} + \frac{1}{\tau_m}$$



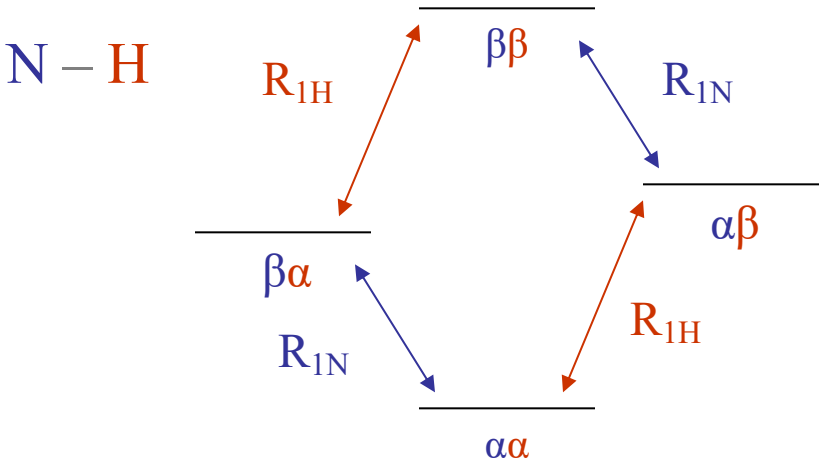
hNOE measurements

- Measure saturated and unsaturated experiments and take the intensity ratio for each peak



Farrow and Kay, Biochemistry, 1993

The heteronuclear NOE



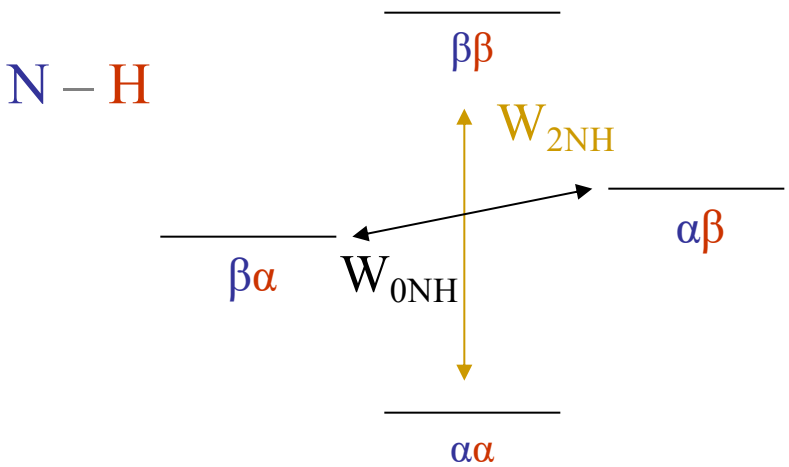
$$M_N \propto (N_{\alpha\alpha} - N_{\beta\alpha}) + (N_{\alpha\beta} - N_{\beta\beta})$$

$$M_H \propto (N_{\alpha\alpha} - N_{\alpha\beta}) + (N_{\beta\alpha} - N_{\beta\beta})$$

Saturation equalizes ββ and βα, αβ and αα → $M_H = 0$

R_1 transitions are an independent return to equilibrium

The heteronuclear NOE



$$M_N \propto (N_{\alpha\alpha} - N_{\beta\alpha}) + (N_{\alpha\beta} - N_{\beta\beta})$$

W_2 transitions increase $N_{\alpha\alpha}$ and decrease $N_{\beta\beta}$

→ M increases (positive NOE)

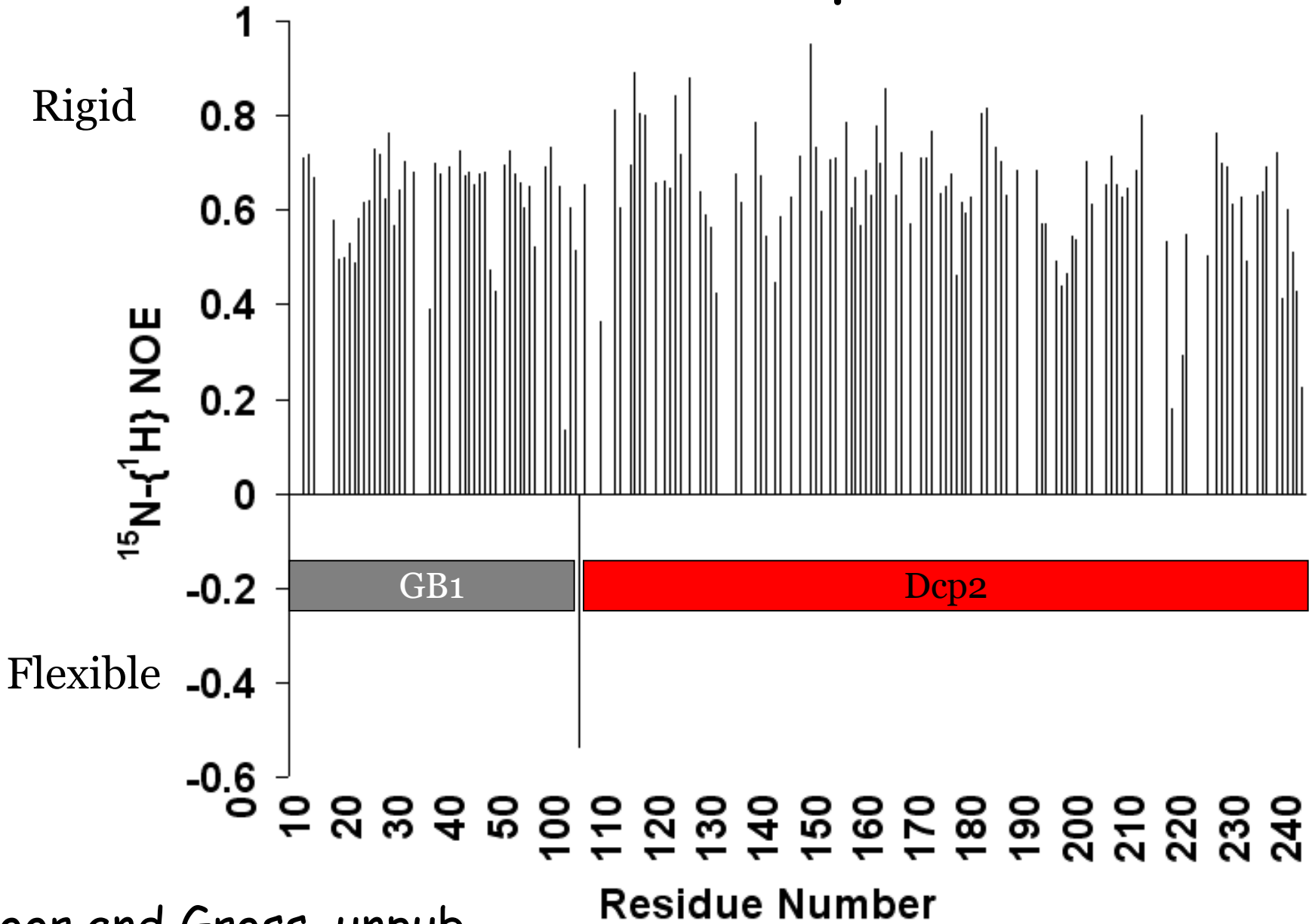
M_N decreases (negative NOE)

W_0 transitions increase $N_{\beta\alpha}$ and decrease $N_{\alpha\beta}$

→ M decreases (negative NOE)

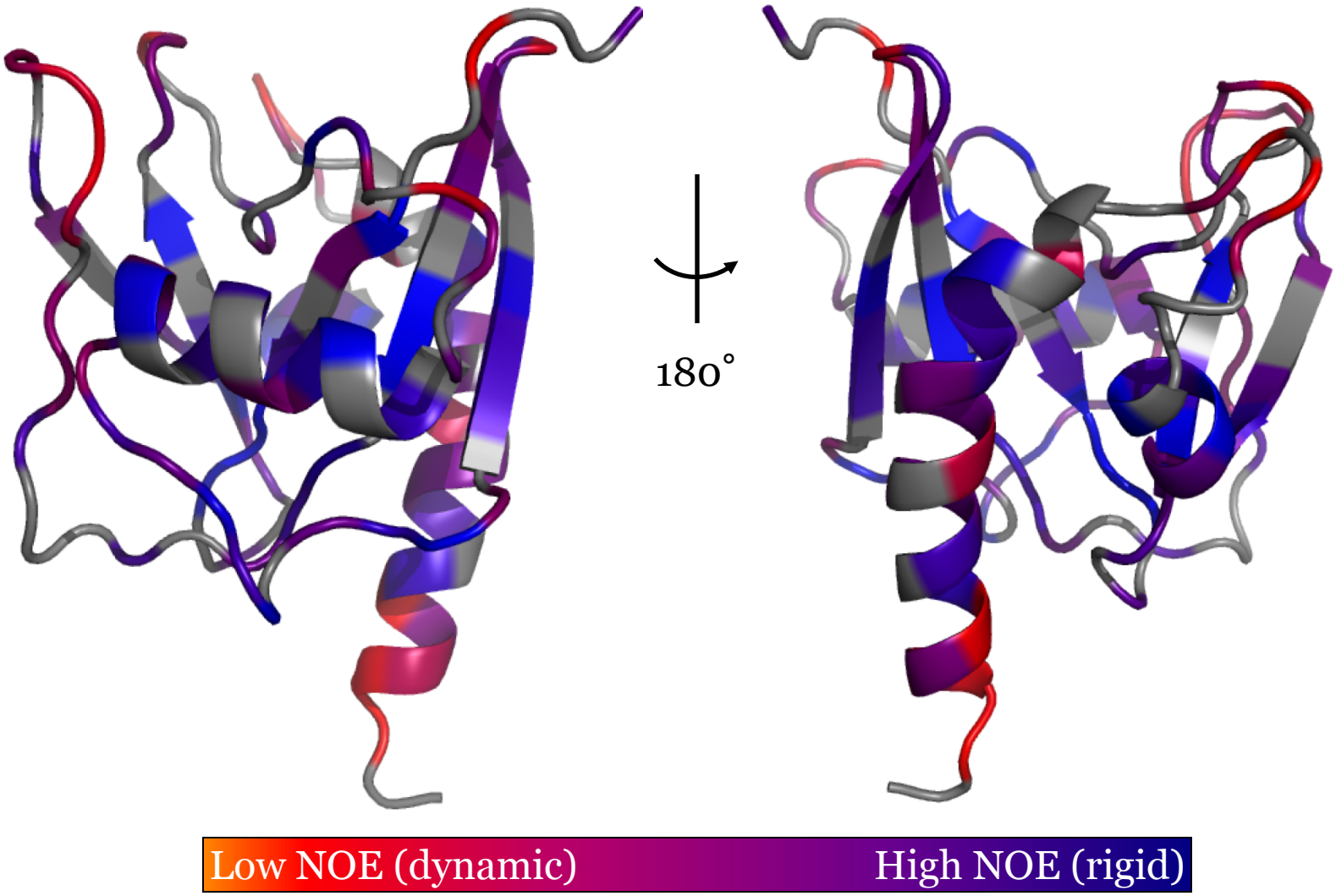
M_N increases (positive NOE)

hNOE and Dcp2

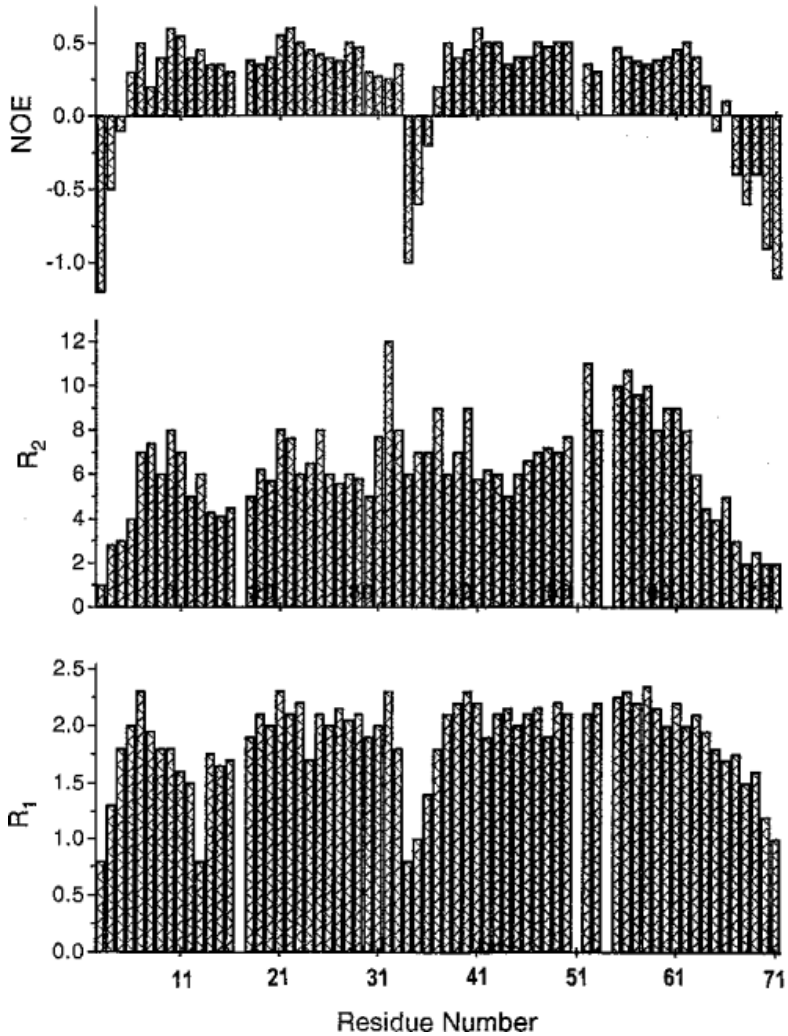


Floor and Gross, unpub.

hNOE versus structure



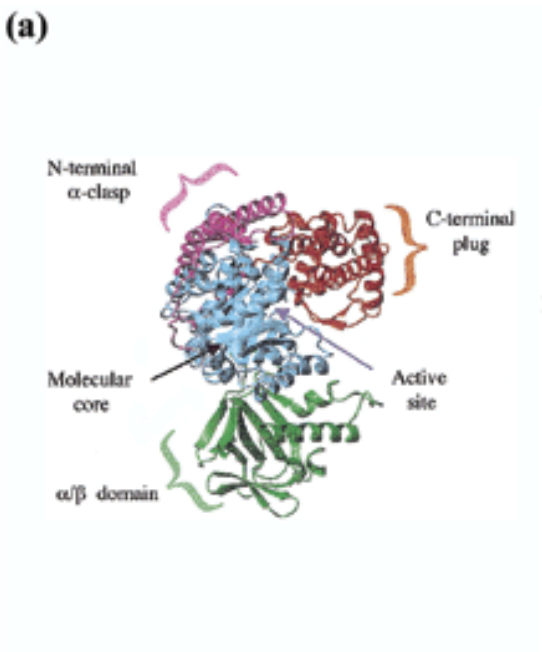
R1, R2 and NH-NOE: three relaxation rates

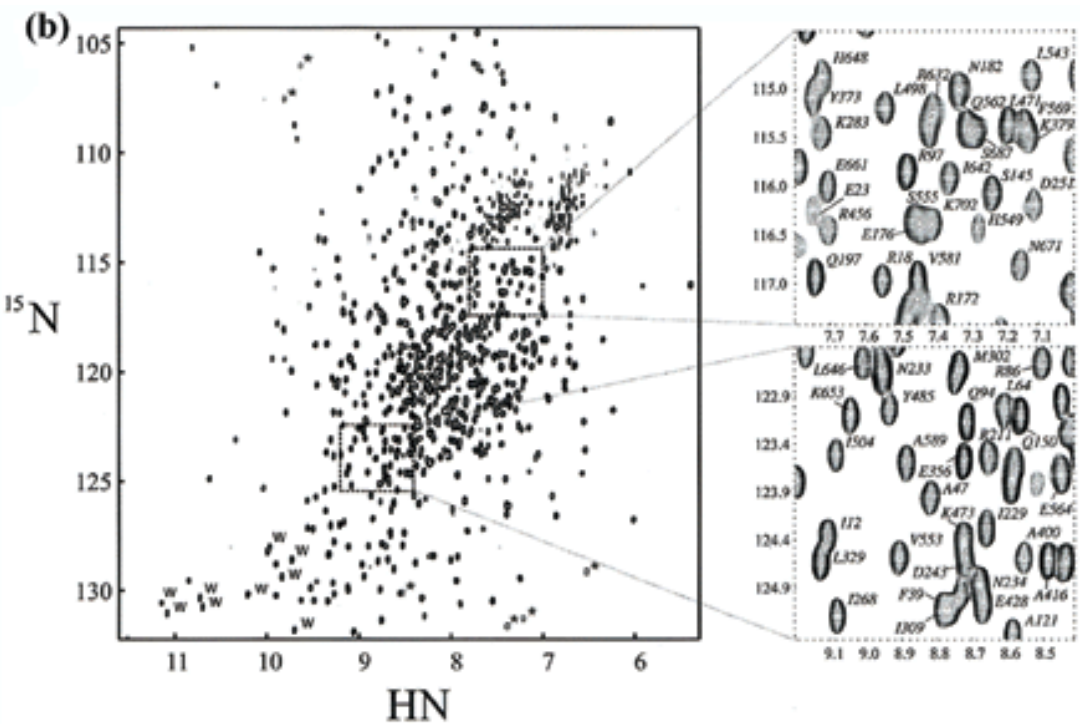


-> three fit parameters: τ_m,τ_e, S^2

Part IV:NMR of Assemblies

TROSY



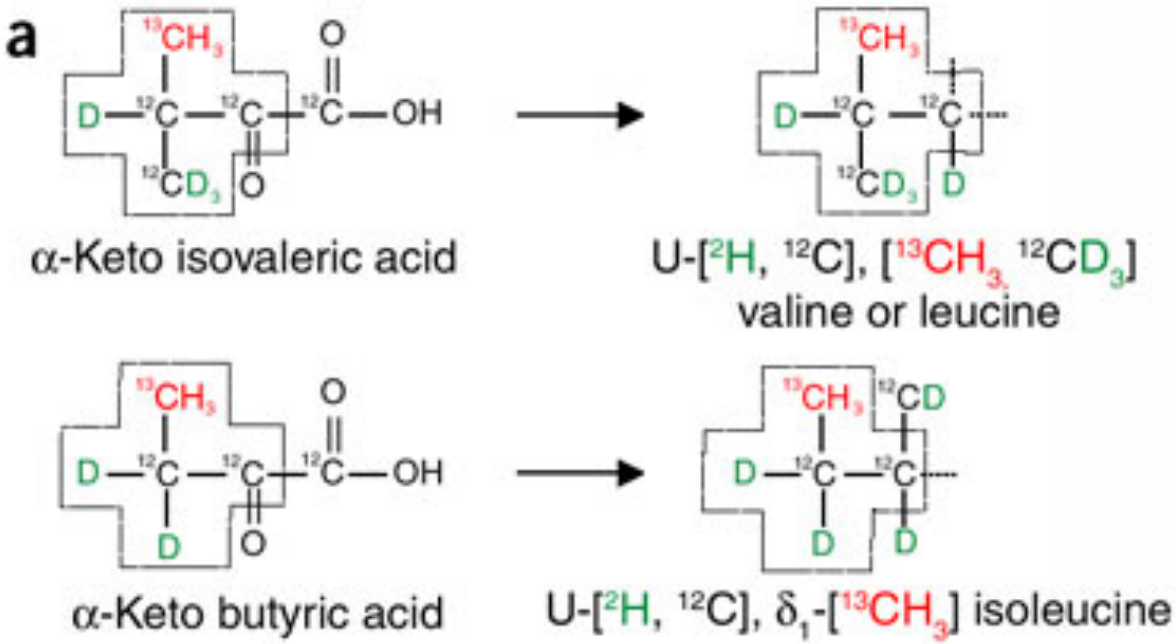


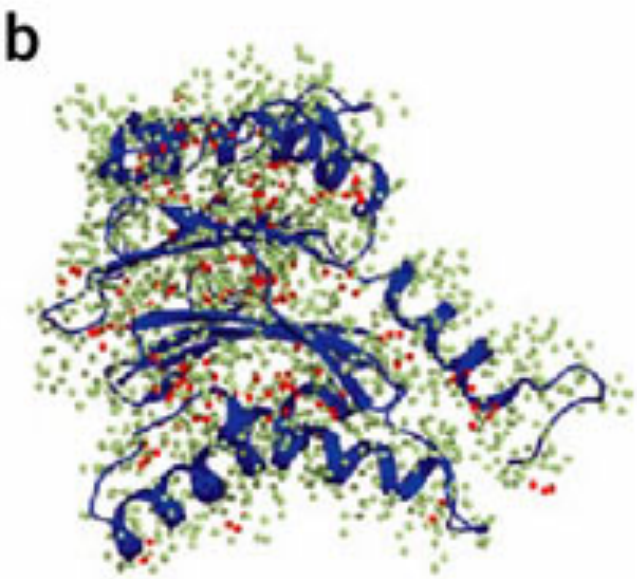
Malate Synthase G
80 kDa

Turgarinov et al, JACS 2002

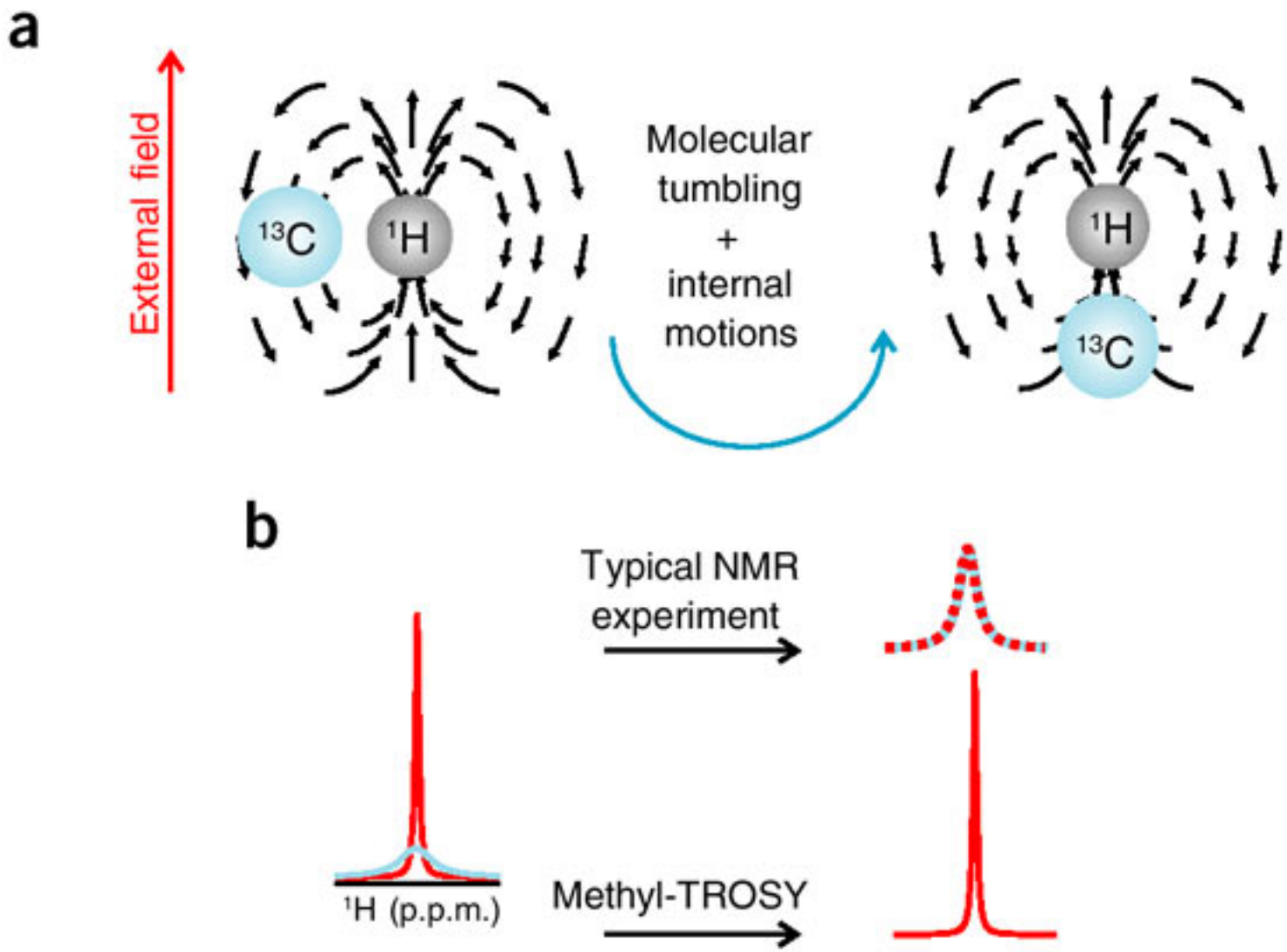
Same info as ^{15}N HSQC

Methyl-group labeling





Methyl-TROSY



doi:10.1038/nature05512

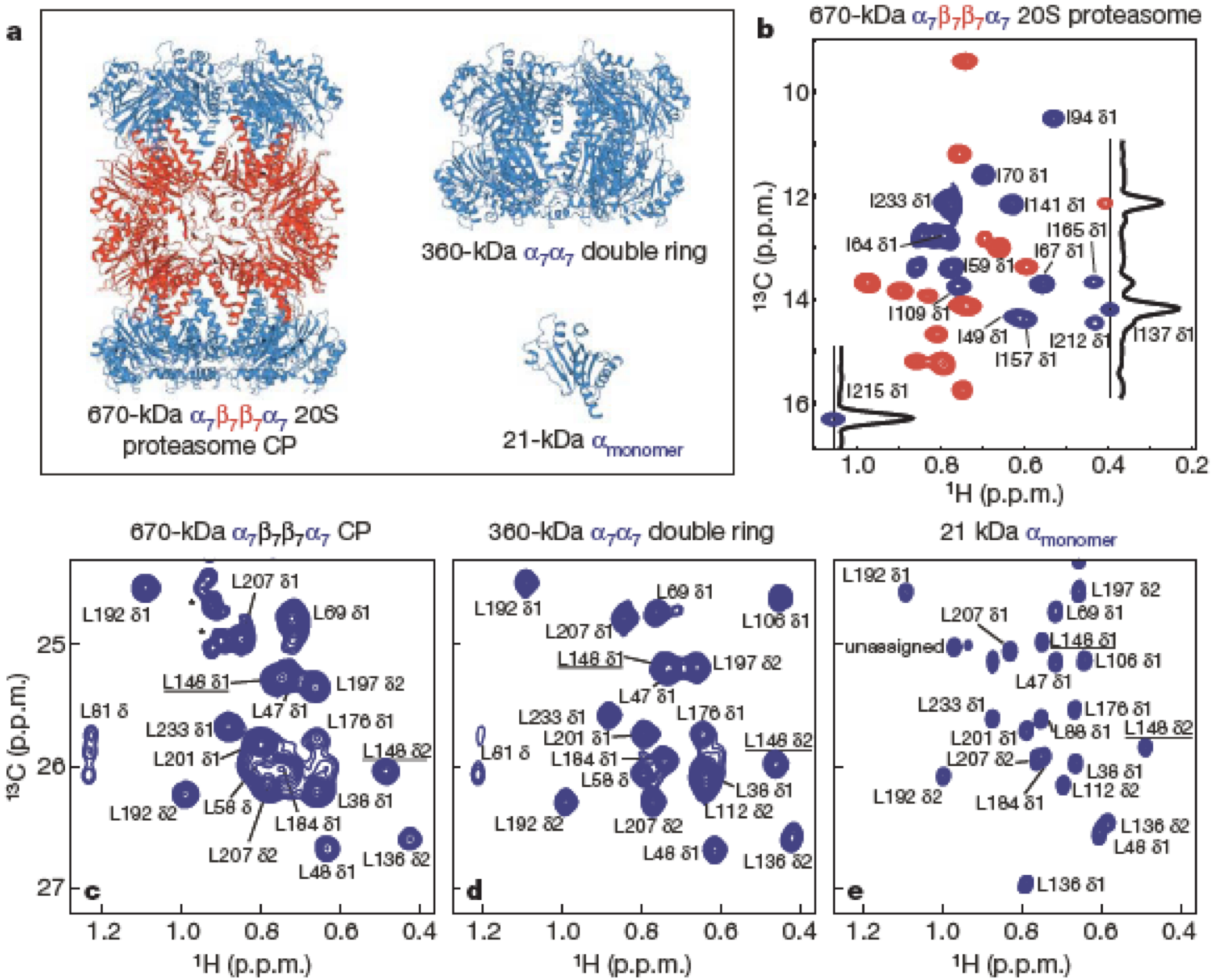
nature

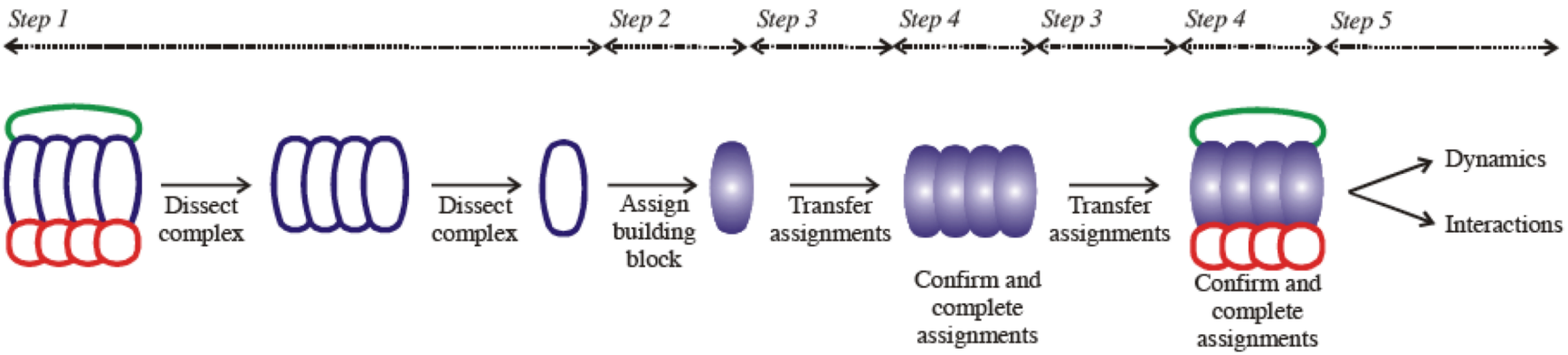
ARTICLES

Quantitative dynamics and binding studies of the 20S proteasome by NMR

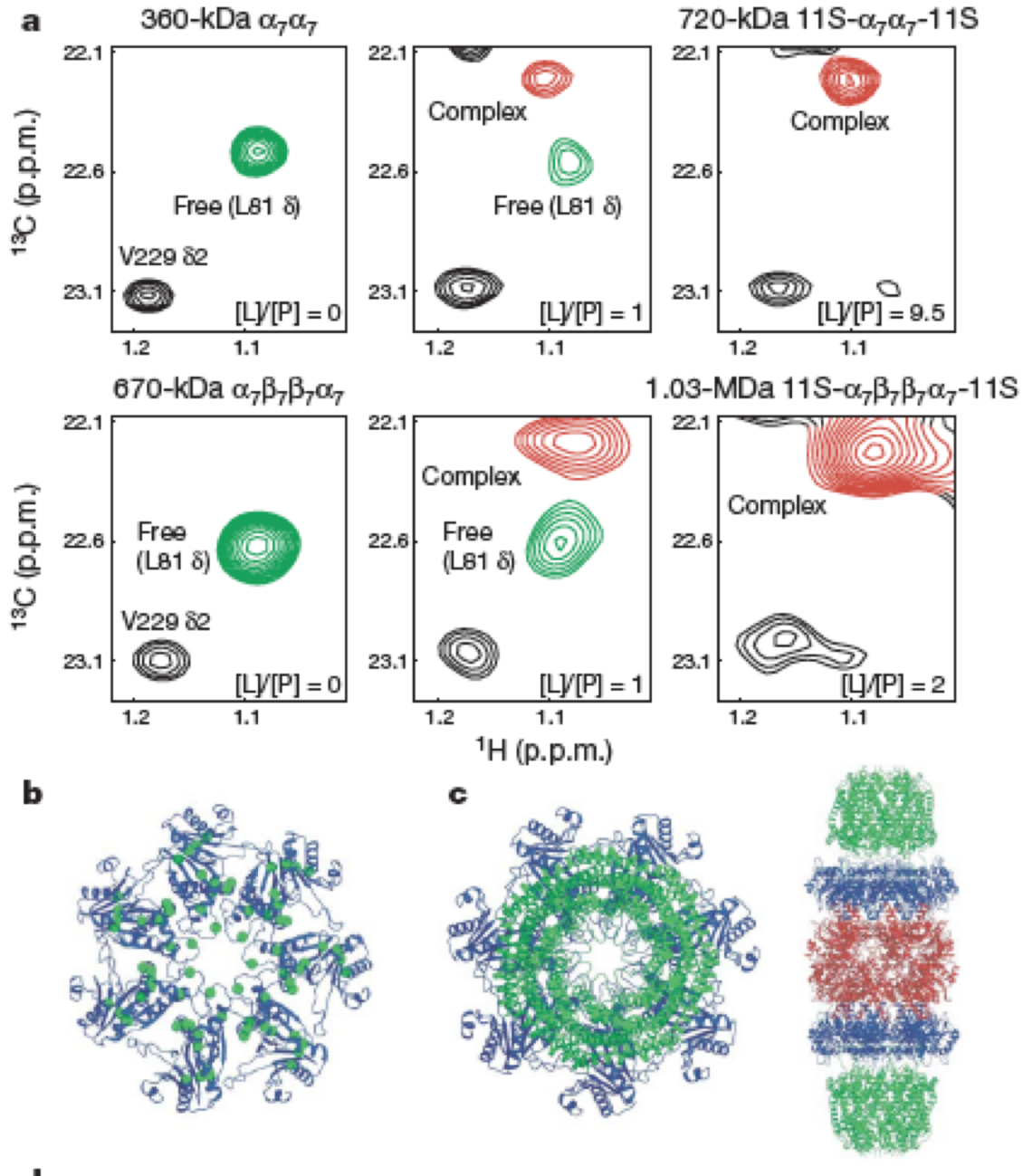
Remco Sprangers[1] & Lewis E. Kay[1]

ILV METHYL ASSIGNMENTS OF 670 KDA COMPLEX

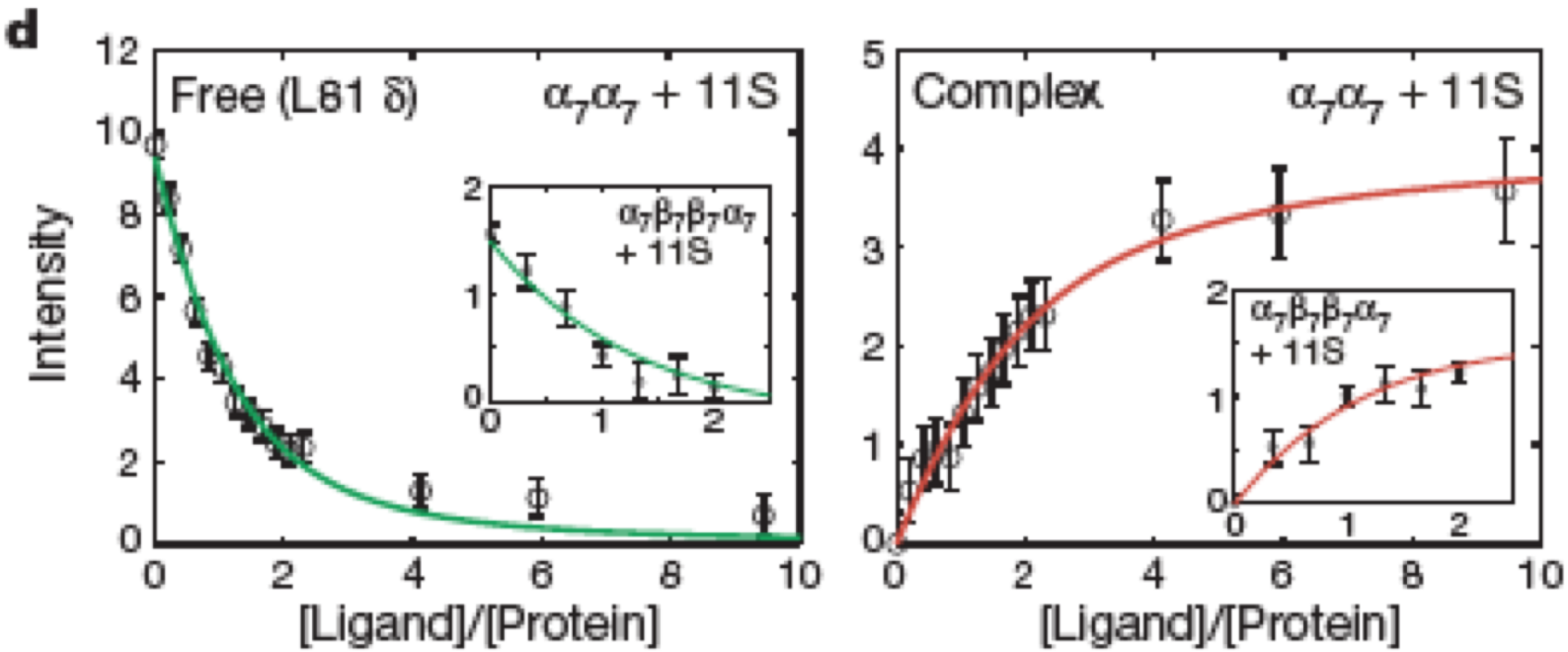



Step 1
Step 2
Step 3
Step 4
Step 3
Step 4
Step 5
Dissect complex
Dissect complex
Assign building block
Transfer assignments
Confirm and complete assignments
Transfer assignments
Confirm and complete assignments
Dynamics
Interactions

11S ACTIVATOR BINDING



11S BINDING CURVES



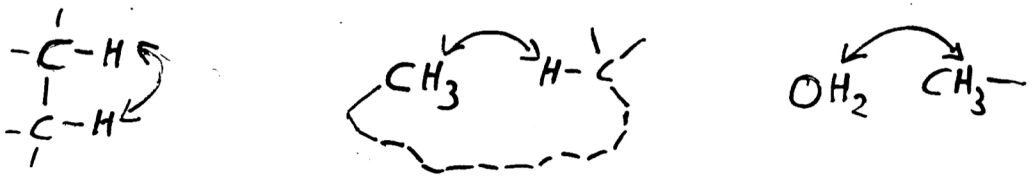
Part V: Structure Determination by NMR

Calculating 3D structures: we need distance measurements & assignments

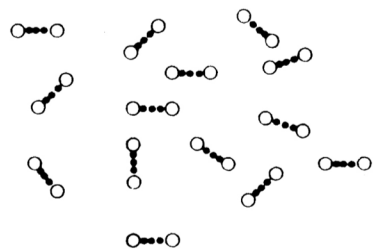
$$NOE = f(\tau_c) \cdot \frac{1}{r^6}$$

We can see distances that are

$$\leq 5\,\text{Å}$$

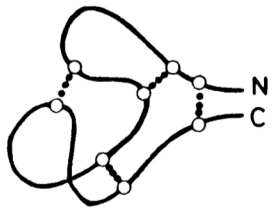


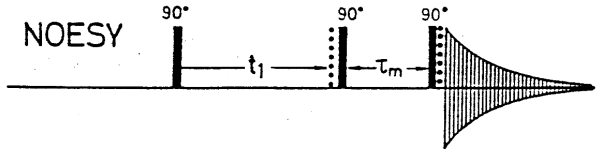
NO ASSIGNMENTS

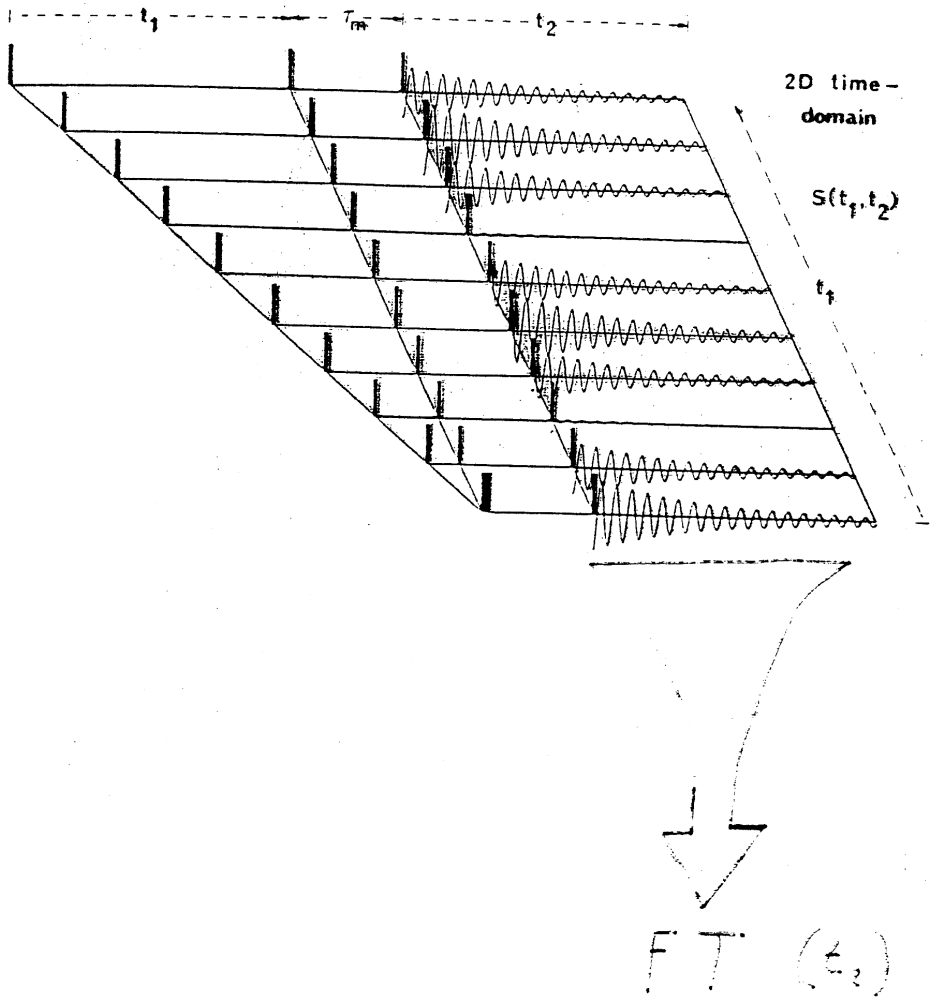


N ——— C

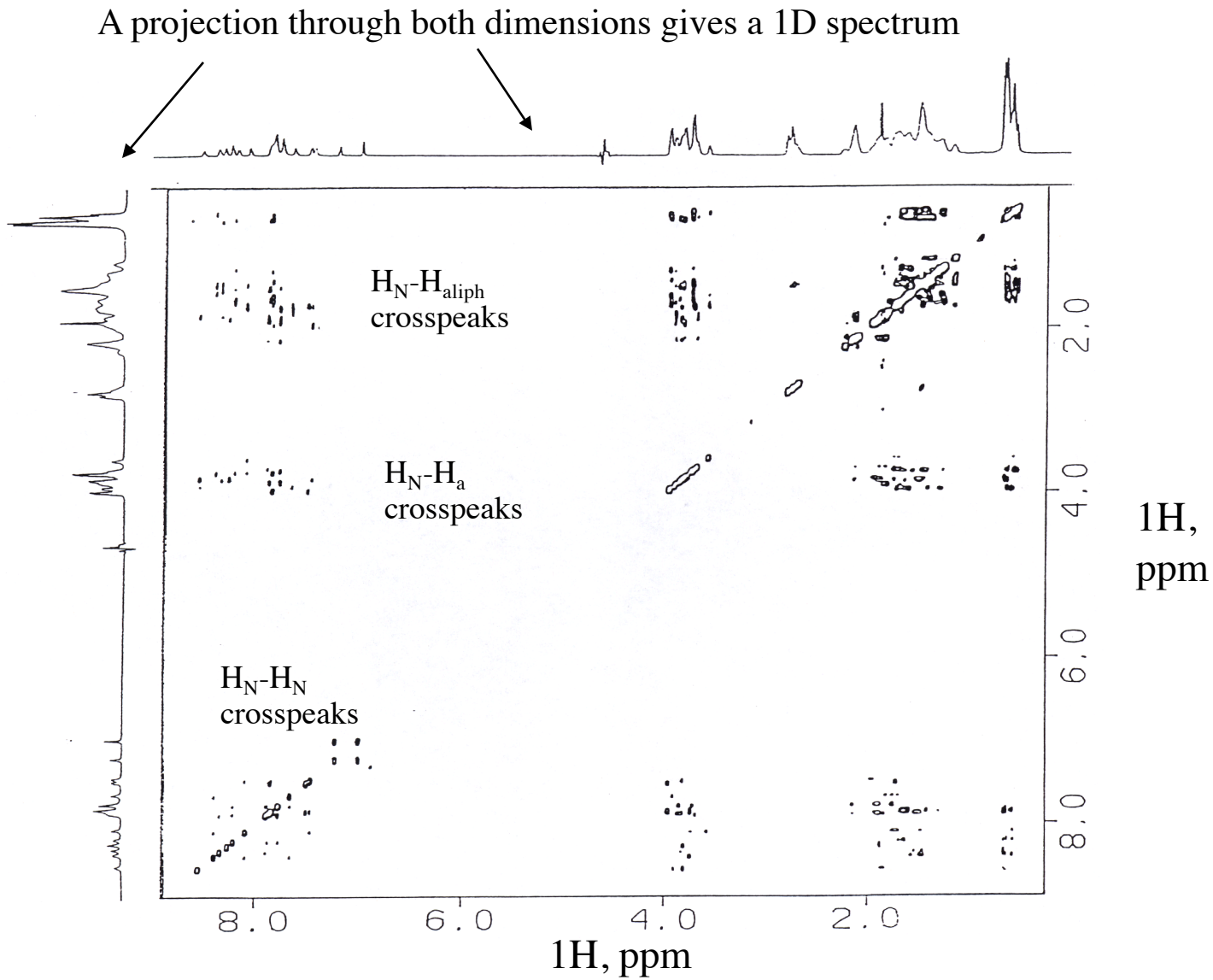
WITH ASSIGNMENTS




NOESY
90°
90°
90°
t_1
τ_m


t_1
τ_m
t_2
2D time-domain
$S(t_1,t_2)$
t_1
FT (t_2)

A Real 2D NOE Experiment of a Small Peptide

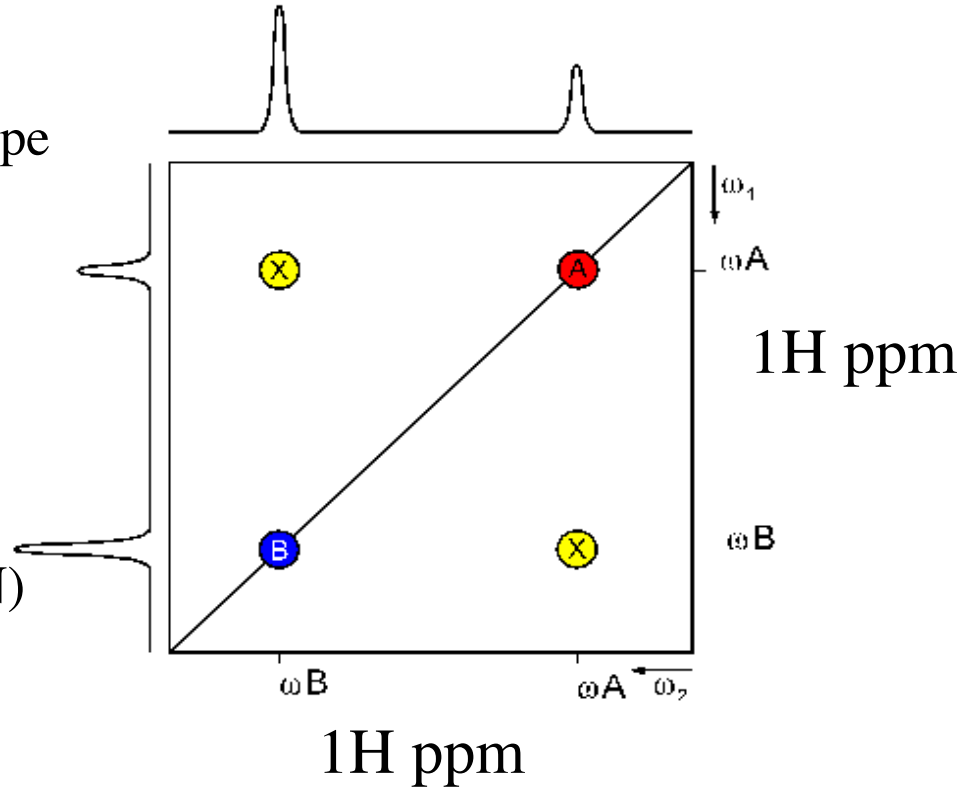


Interpretation of 2D NMR Spectra

- Crosspeaks are a measure of some type of interaction between 2 spins (NOE, J-coupling....)

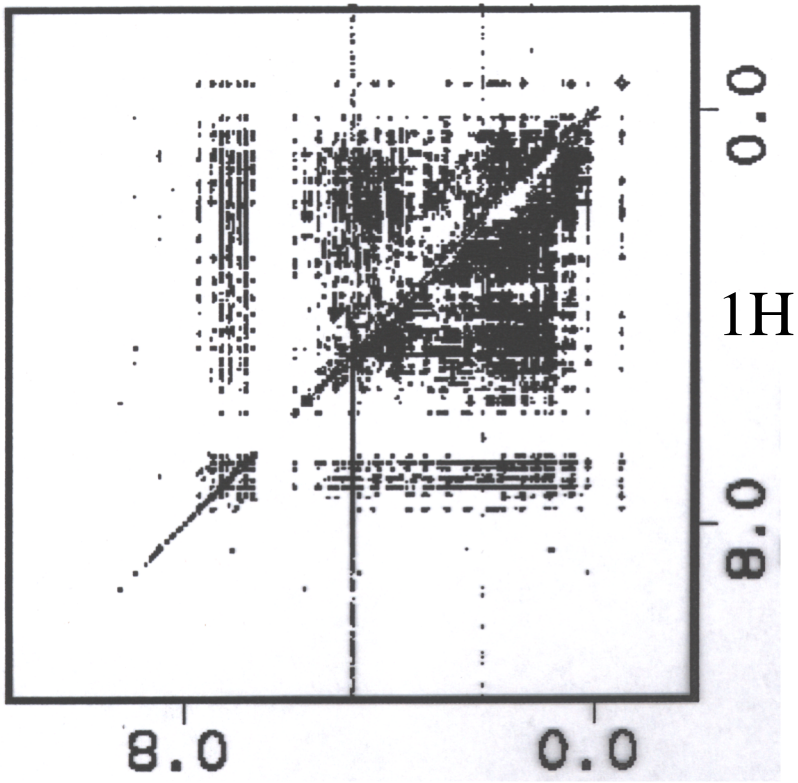
- The intensity of the crosspeak often quantifies the interaction.

- A heteronuclear experiment (1H-15N) would not have diagonal crosspeaks.



Higher Dimensionality 3 and 4D Heteronuclear Experiments on Isotopically Labeled (15N-13C) Proteins

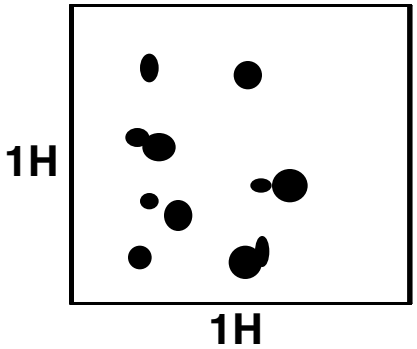
2D NOESY of a 76 residue protein homodimer (effectively 18kD) in D_2O



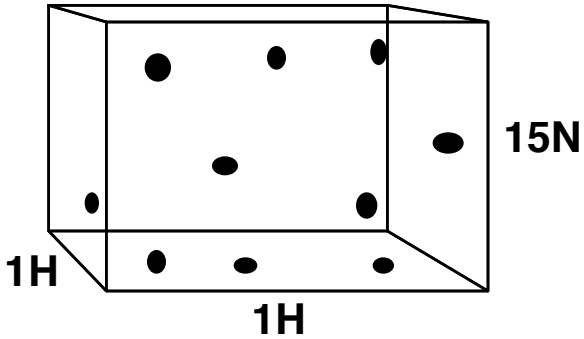
In practice, even small proteins have very crowded 2D spectra making assignment very difficult. In this case the fact that it is in D2O simplifies the spectra because the amide protons exchange for deuterium and are not visible.

Benefit of C13 and N15 labeling of Proteins for NMR

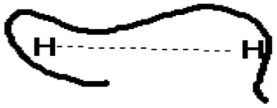
Higher Dimensionality (3 and 4D) Experiments Reduce Overlap Compared to 2D Experiments

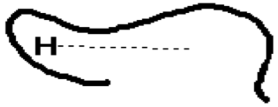


2D nOe Expt. on unlabeled protein

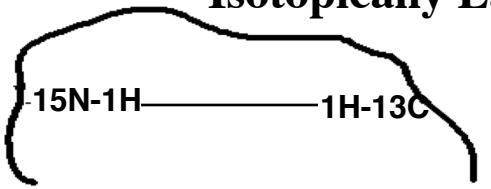


3D nOe Expt. on N15-labeled protein

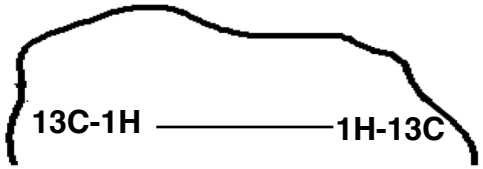




Many More Types of Experiments Can be Done on Isotopically Labeled Protein

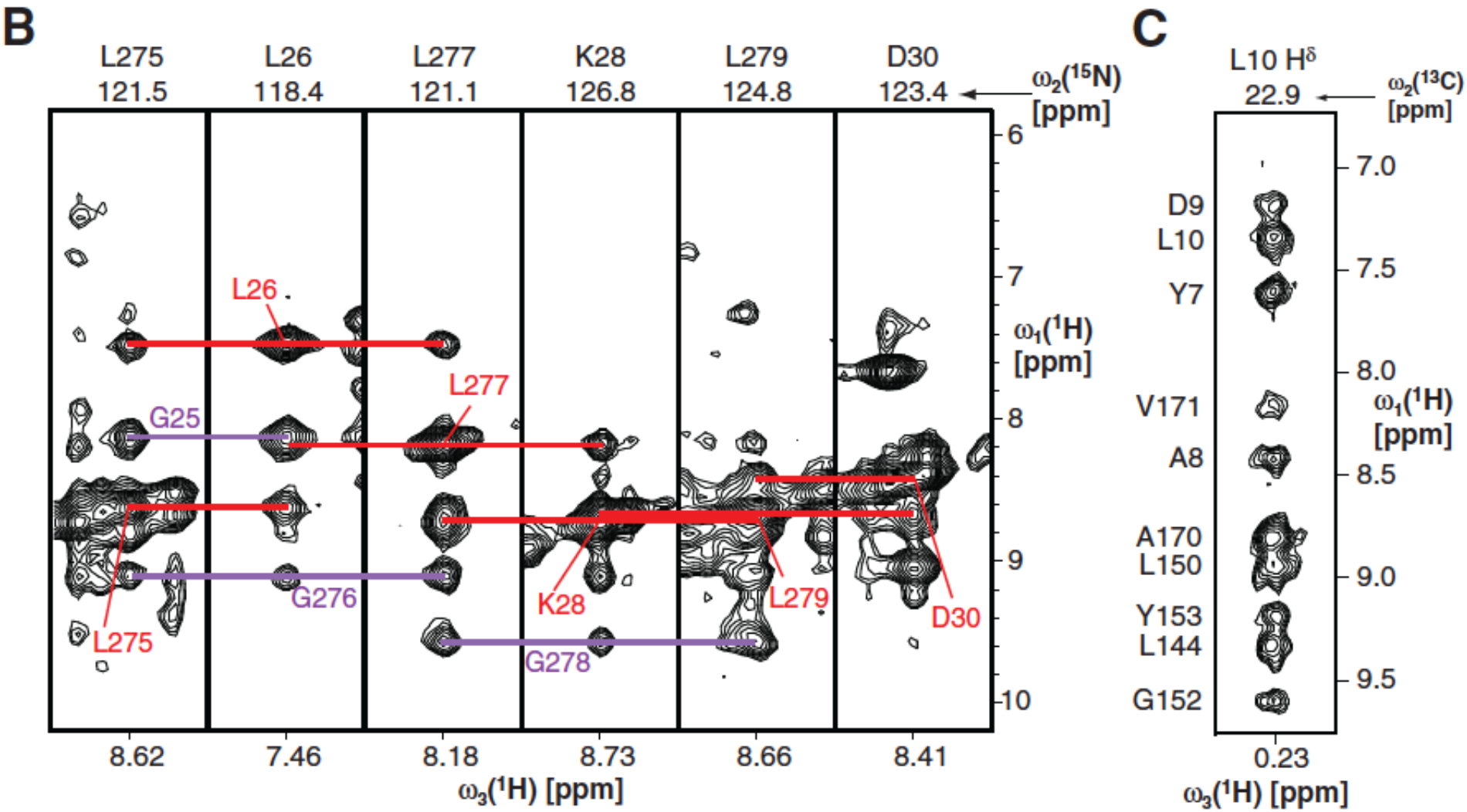


nOes between Protons Attached to N15 and Protons Attached to 13C



nOes between Protons Attached to 13C and Protons and Attached to 13C

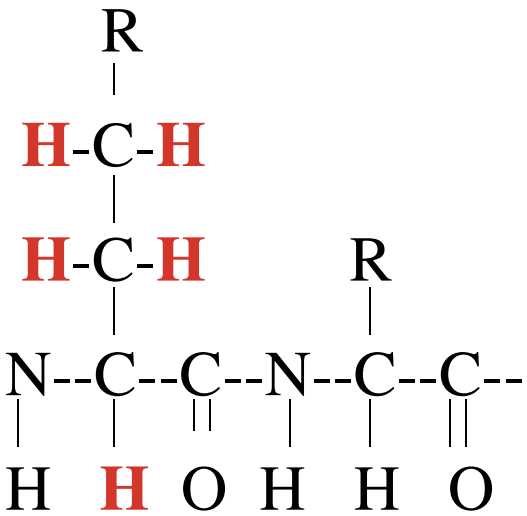
Examples of 15N and 13C dispersed NOESY



15N NOESY-HSQC

13C NOESY-HSQC

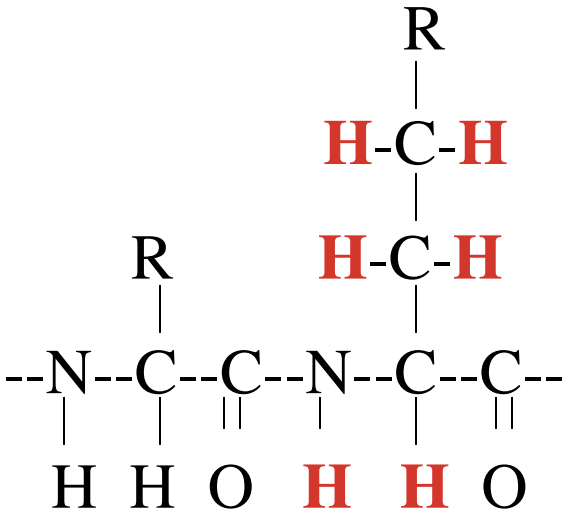
Side-chain protein assignments



H(CCO)NH-TOCSY

i - 1 res.

All Carbon's H's at i-1 to N-H pair.

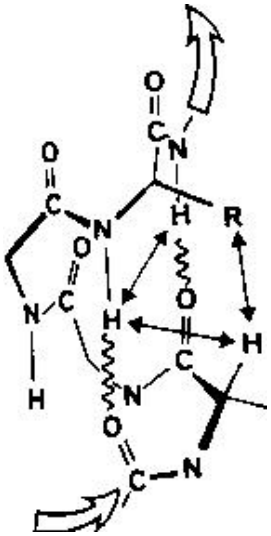


15N-TOCSY i res.

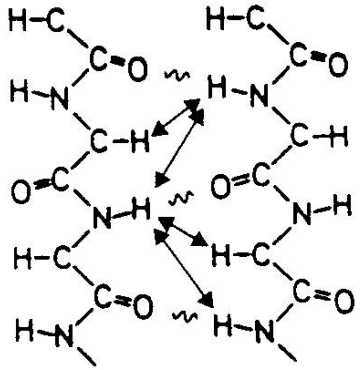
All H's at i to N-H pair.

TOCSY methods relies on through-bond J Couplings

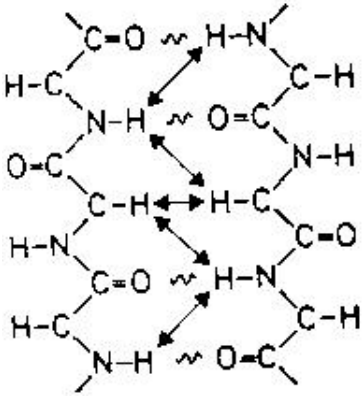
Close interatomic distances in secondary structures



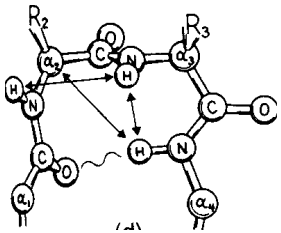
alpha-helix



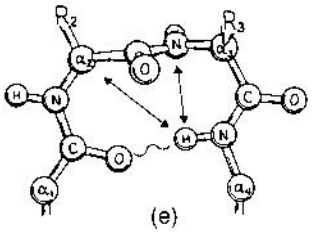
parallel beta-sheet



antiparallel beta-sheet



type I turn



type II turn

THE RANGE OF ^{13}C CHEMICAL SHIFTS OBSERVED FOR EIGHT DIFFERENT PROTEINS

Res.	α	β	γ		δ	ε
Gly	42-48					
Ala	49-56	18-24				
Ser	55-62	61-67				
Thr	58-68	66-73	19-26			
Val	57-67	30-37	16-25			
Leu	51-60	39-48	22-29		21-28	
Ile	55-66	34-47	25-31	14-22	9-16	
Lys	52-61	29-37	21-26		27-34	40-43
Arg	50-60	28-35	25-30		41-45	
Pro	60-67	27-35	24-29		49-53	
Glu	52-62	27-34	32-38			
.	.	.	.		.	.
.	.	.	.		.	.
.	.	.	.		.	.

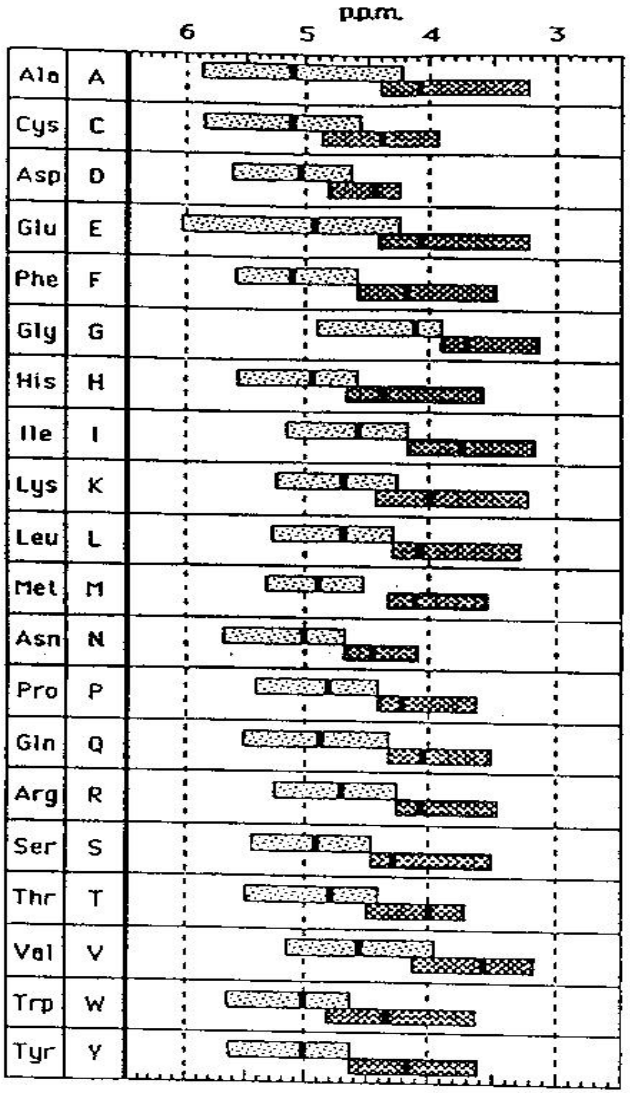
WAGNER AND BRUHWILER, 1986…et al.

Or http://www.bmrb.wisc.edu/ref_info/statsel.htm

H^{α} chemical shifts and secondary structure

- **the figure at right shows distributions of H^{α} chemical shifts observed in sheets (lighter bars) and helices (darker bars).**

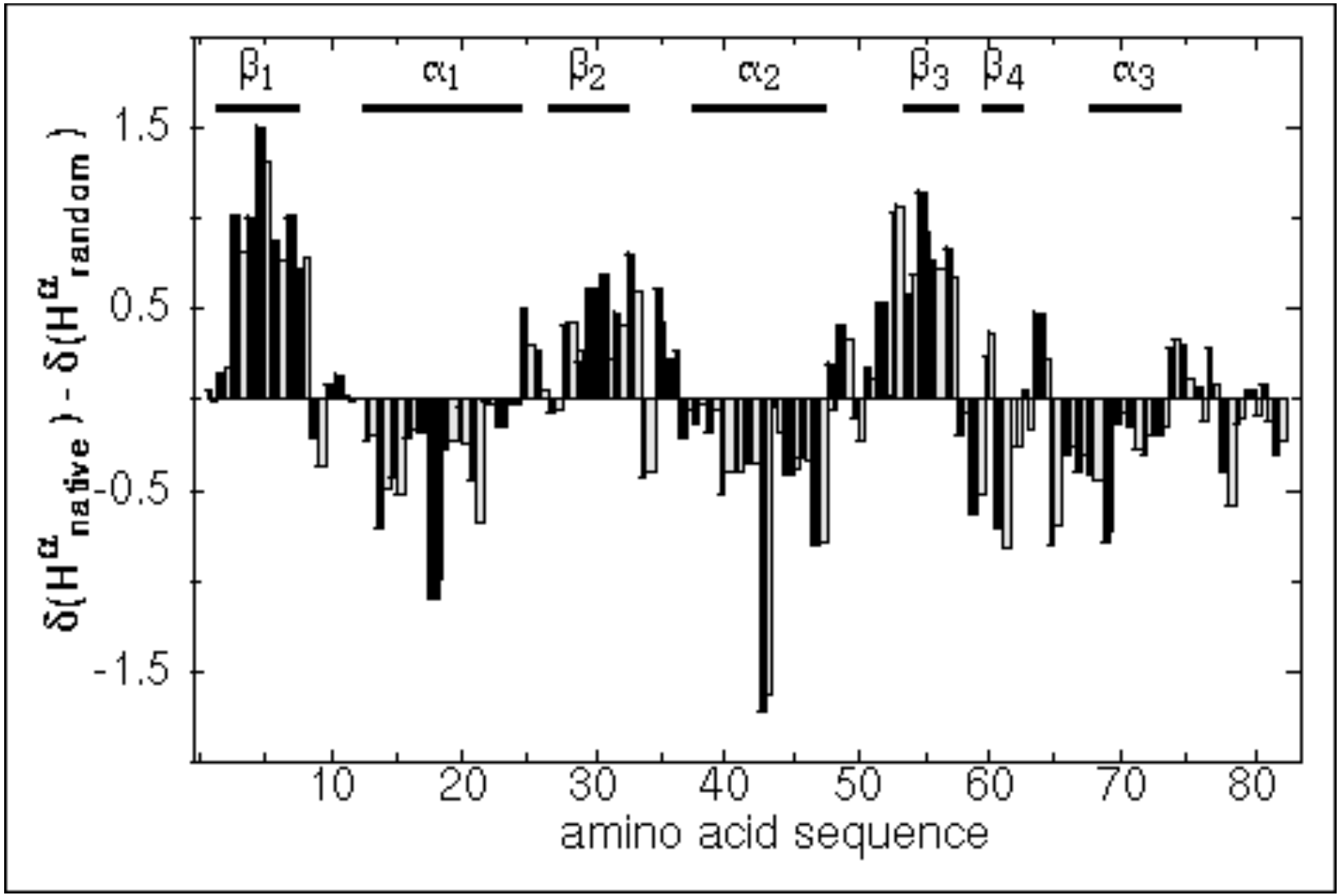
- **H^{α} chemical shifts in α–helices are on average 0.39 ppm below "random coil" values, while β-sheet values are 0.37 ppm above random coil values.**



Wishart, Sykes & Richards

J Mol Biol (1991) **222**, 311 .

Secondary Shift vs Sequence



Reveals secondary structure !

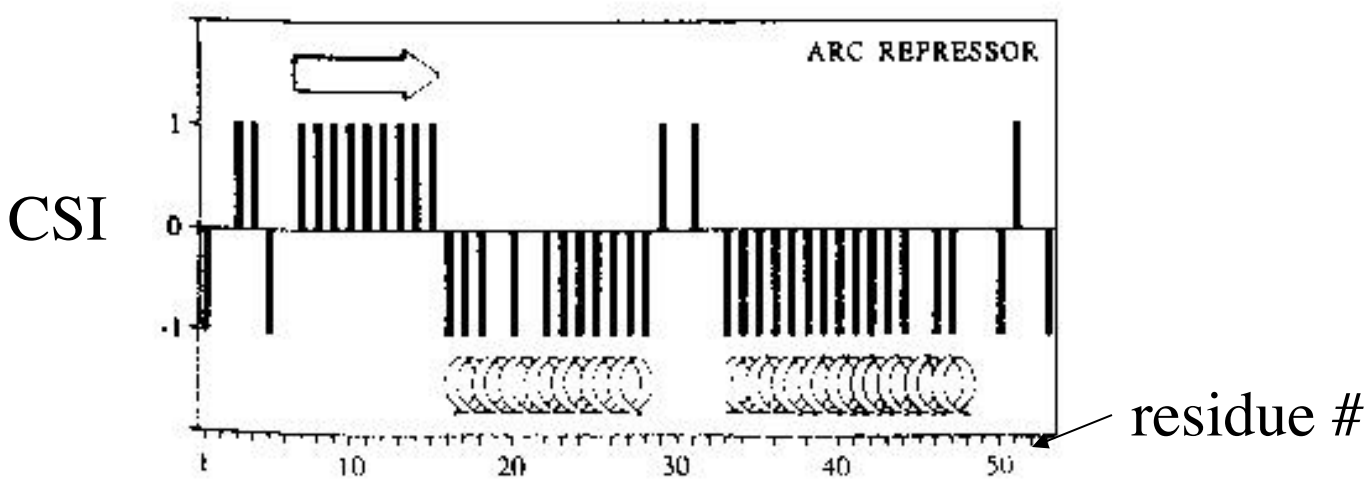
Chemical shift index (CSI)

- trends like these led to the development of the concept of the *chemical shift index** as a tool for assigning secondary structure using chemical shift values.
- one starts with a table of reference values for each amino-acid type, which is essentially a table of "random coil" H^{α} values
- CSI's are then assigned as follows:

exp'tl H^{α} shift rel. to reference	assigned CSI
within ± 0.1 ppm	0
>0.1 ppm lower	-1
>0.1 ppm higher	+1

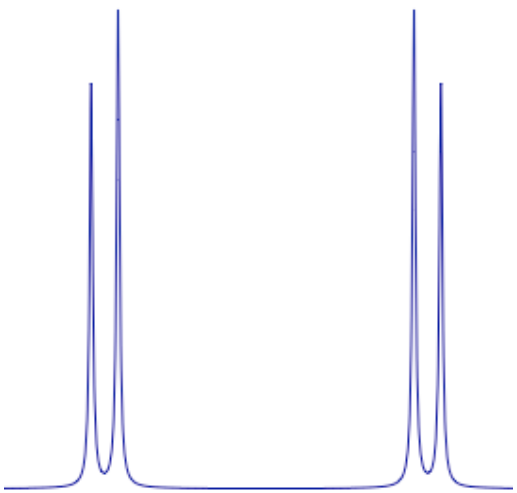
*Wishart, Sykes & Richards *Biochemistry* (1992) **31**, 1647-51.

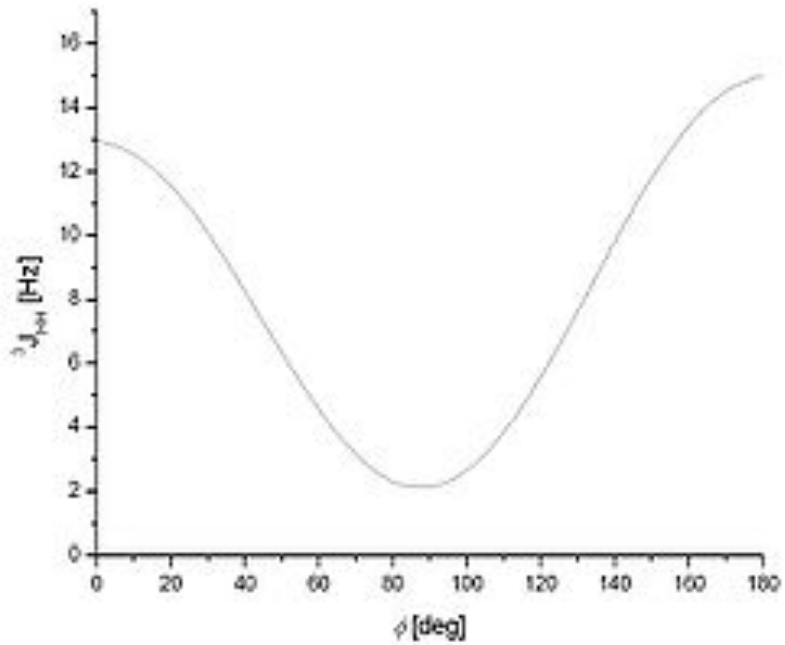
Chemical shift indices



- any "dense" grouping of four or more "-1's", uninterrupted by "1's" is assigned as a helix, while any "dense" grouping of three or more "1's", uninterrupted by "-1's", is assigned as a sheet.
- a "dense" grouping means at least 70% nonzero CSI's.
- other one regions are assigned as "coil"
- **this simple technique assigns 2ndary structure w/90-95% accuracy**
- similar useful relationships exist for $^{13}C^{\alpha}$, $^{13}C^{C=O}$ shifts

J couplings contain information on structure



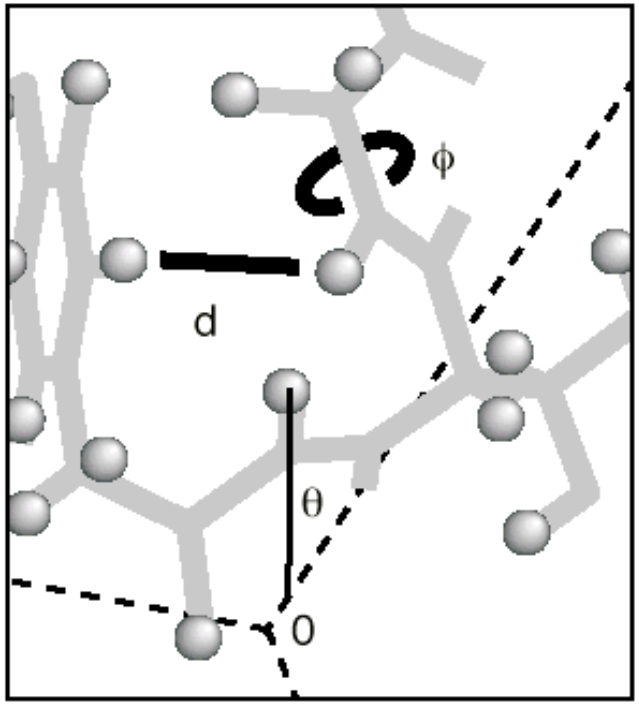


$$J(\phi) = A\cos^2\phi + B\cos\phi + C$$

3D structure calculation

- NMR provides information about structure
    - chemical shifts <=> local electronic environment
    - coupling constants <=> torsion angles
    - NOE, ROE <=> interproton distances
    - residual dipolar couplings <=> bond orientation
- and dynamics
    - relaxation times
    - NOE, ROE
- Most of the data describe
    - local environment of the protons
    - relative to each other
    - not the global conformation of the molecule

Experimental data from NMR



• Distance

NOE: The distance between i and j is a function of the NOE intensity $\mathbf{D_{ij} \sim C(NOE_{ij})^{-6}}$

H-bonds: Identified by slowly exchanging amide H_N protons

•Angles

Side Chain χ and backbone torsion identified from J-coupling experiments

Chemical Shift also gives Angular Information

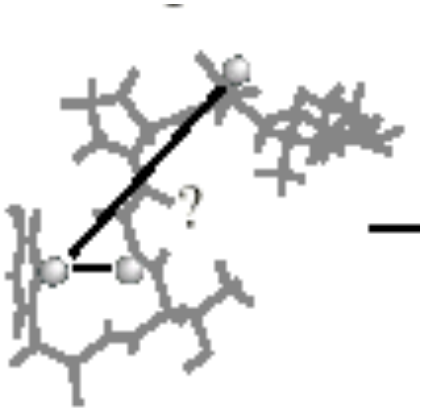
• Residual Dipolar Couplings

Bond Orientations Relative to an Alignment Tensor

3D structure calculation

- Molecular Dynamics with Simulated Annealing starting from random coordinates

- Goal is to minimize the hybrid energy function



Additional
Unambiguous
Experimental
Restraints

E-ForceField

E-NOEs
E-Angles
E-H_bonds
E-Chemical_shift
E-Dipolar_couplings

The hybrid energy function

- Structure calculation = minimization of hybrid energy function (target function) which combines

  1. different experimental data
  2. *a priori* information (force field)

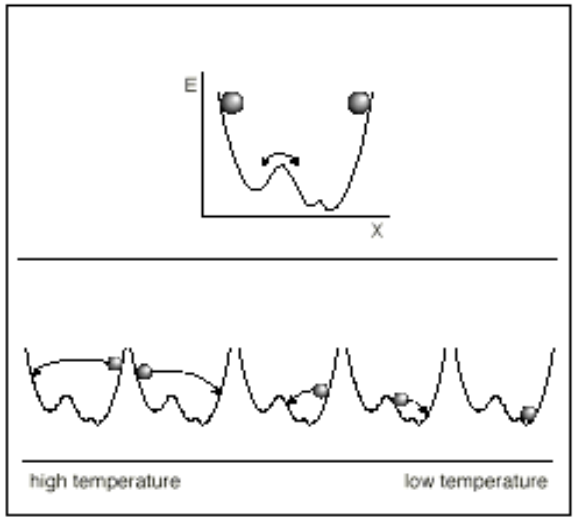
$$
\begin{aligned}
E_{hybrid} &= \textstyle\sum_l w_l E_l = \\
& \quad w_{bond} E_{bond} \\
&+ w_{angle} E_{angle} \\
&+ w_{improper} E_{improper} \\
&+ w_{nonbonded} E_{nonbonded} \\
&+ w_{unambig} E_{unambig} + w_{ambig} E_{ambig} + \ldots \\
&+ w_{torsion} E_{torsion} \\
&+ w_{Jcoup} E_{Jcoup} \\
&+ w_{RDC} E_{RDC} + \ldots
\end{aligned}
$$

Minimization by molecular dynamics

- MD solves Newton's eqns. of motion:

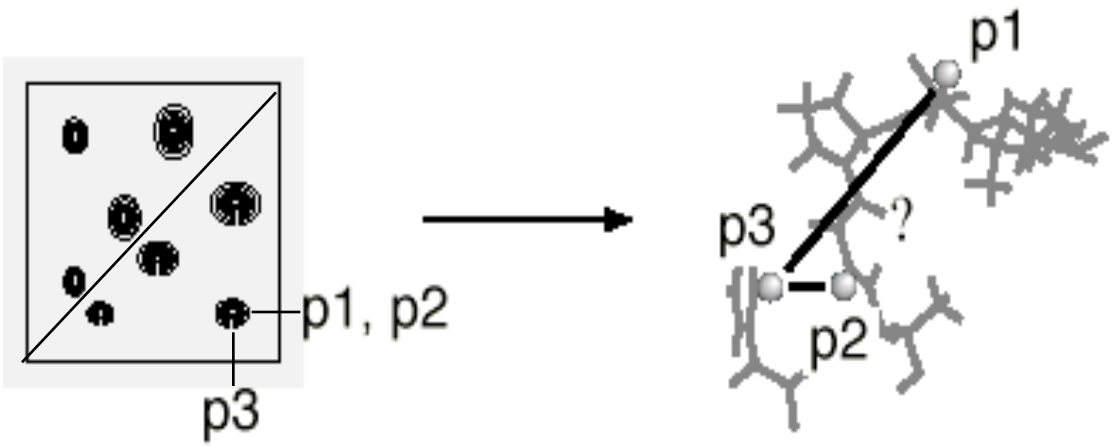
$$\frac{d^2\vec{r}_i}{dt^2} = -\frac{c}{m_i}\frac{\partial}{\vec{r}_i}E_{hybrid}$$

- Molecular dynamics can overcome local energy barriers



- Temperature control and variation: minimization by simulated annealing

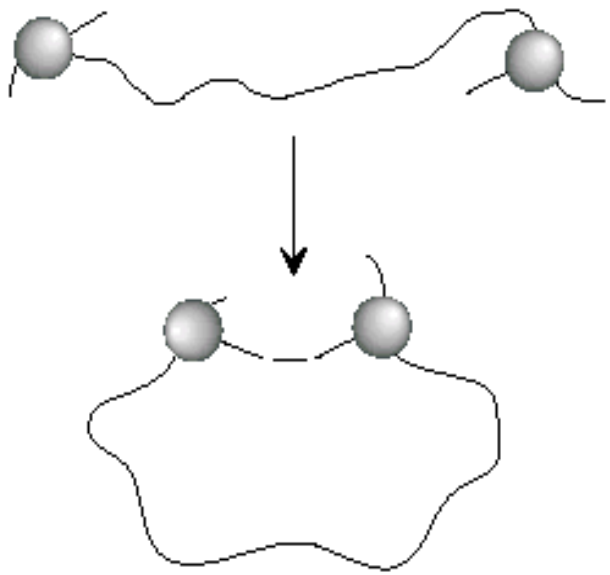
NOE ambiguities



- Key problem is ambiguity in NOE assignments
- Need for higher dimensional data: 3D & 4D
- Need for heteronuclear data
- Need for better calculational strategies that can deal with ambiguous data

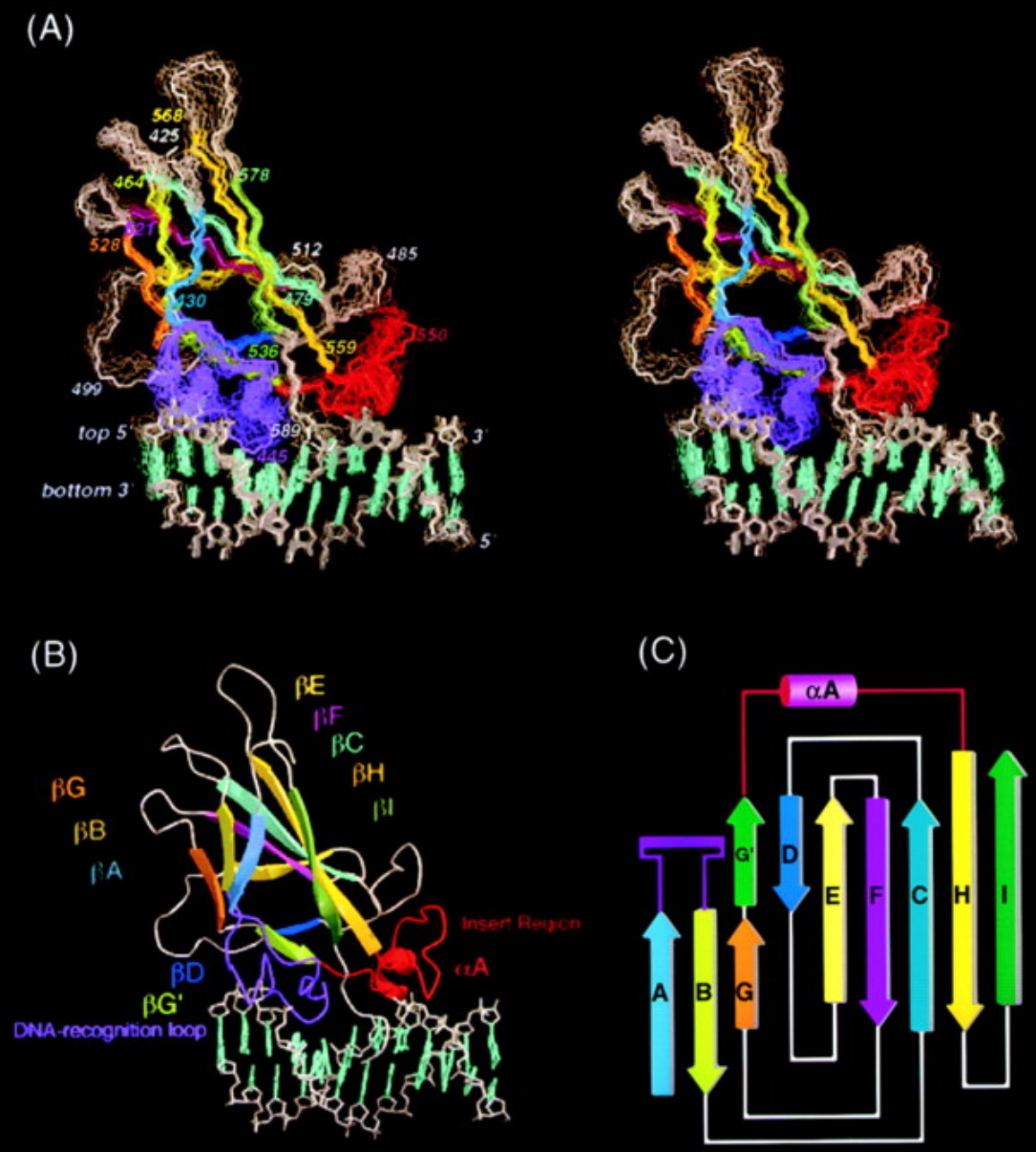
Errors in data: error bounds

- Cumulative error in D_{ij} is treated by using loose error bounds $L \ldots U$
- Precise value not (too) critical:
  loose bounds restrict conformational space



- However, consequences for:
  - precision of structure

Solution Structure of the Core NFATC1/DNA Complex



Zhou et al, Cell , 1998

Need to evaluate restraint numbers, violations and precision

Table 1. Structural Statistics for NFATC1-DBD*/DNA Complex

Protein		
Amino acids residues sequentially assigned (nonproline)		162 of 167
Effective distance constraints		1050
Intraresidue		236
Sequential (|i − j| = 1)		374
Medium-range (|i − j| ≤4)		86
Long-range (|i −j| ≥5)		272
H bonds		82
Dihedral angle constraints		363
DNA		
Effective distance constraints		276
Intraresidue		146
Sequential		131
Interstrand		1
H bond		58
Protein–DNA interface		
Effective distance constraints		56
Distance constraint violations > 0.2 Å (per structure)		1.81 ± 2.07
Dihedral constraint violations > 3.0° (per structure)		1.44 ± 0.96
X-PLOR potential energy (E_{L-J}, Kcal/mol, avg. per structure)		−501 ± 28.9
R.m.s.d. to the mean for backbone heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		0.62 ± 0.07
R.m.s.d. to the mean for heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		1.20 ± 0.14
R.m.s.d. to the mean for backbone heavy atoms (residues 425–590)		1.14 ± 0.11
R.m.s.d. to the mean for heavy atoms (residues 425–590)		1.58 ± 0.10
R.m.s.d. to the mean for DNA heavy atoms (superposition of all DNA heavy atoms)		0.83 ± 0.21
R.m.s.d. to the mean for protein + DNA heavy atoms (superposition of all β strand heavy atoms and DNA heavy atoms)		1.20 ± 0.16
Ramachandran plot[a]		
	Residues 425–590	Secondary Structures
Most favorable region	60.4%	83.7%
Additionally allowed region	38.9%	14.0%
Generously allowed region	0.7%	2.3%
Disallowed region	0.0%	0.0%

[a] Laskowski et al., 1996.

How were restraints measured?

Analysis of Table

Table 1. Structural Statistics for NFATC1-DBD*/DNA Complex

Protein		
Amino acids residues sequentially assigned (nonproline)		162 of 167
Effective distance constraints		1050
Intraresidue		236
Sequential (|i − j| = 1)		374
Medium-range (|i − j| ≤4)		86
Long-range (|i −j| ≥5)		272
H bonds		82
Dihedral angle constraints		363
DNA		
Effective distance constraints		276
Intraresidue		146
Sequential		131
Interstrand		1
H bond		58
Protein–DNA interface		
Effective distance constraints		56
Distance constraint violations > 0.2 Å (per structure)		1.81 ± 2.07
Dihedral constraint violations > 3.0° (per structure)		1.44 ± 0.96
X-PLOR potential energy (E_{L-J}, Kcal/mol, avg. per structure)		−501 ± 28.9
R.m.s.d. to the mean for backbone heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		0.62 ± 0.07
R.m.s.d. to the mean for heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		1.20 ± 0.14
R.m.s.d. to the mean for backbone heavy atoms (residues 425–590)		1.14 ± 0.11
R.m.s.d. to the mean for heavy atoms (residues 425–590)		1.58 ± 0.10
R.m.s.d. to the mean for DNA heavy atoms (superposition of all DNA heavy atoms)		0.83 ± 0.21
R.m.s.d. to the mean for protein + DNA heavy atoms (superposition of all β strand heavy atoms and DNA heavy atoms)		1.20 ± 0.16
Ramachandran plot[a]		
	Residues 425–590	Secondary Structures
Most favorable region	60.4%	83.7%
Additionally allowed region	38.9%	14.0%
Generously allowed region	0.7%	2.3%
Disallowed region	0.0%	0.0%

[a] Laskowski et al., 1996.

How many NOE restraints?

>6 per residue is acceptable

Analysis of Table

Table 1. Structural Statistics for NFATC1-DBD*/DNA Complex

Protein		
Amino acids residues sequentially assigned (nonproline)		162 of 167
Effective distance constraints		1050
Intraresidue		236
Sequential (|i − j| = 1)		374
Medium-range (|i − j| ≤4)		86
Long-range (|i −j| ≥5)		272
H bonds		82
Dihedral angle constraints		363
DNA		
Effective distance constraints		276
Intraresidue		146
Sequential		131
Interstrand		1
H bond		58
Protein–DNA interface		
Effective distance constraints		56
Distance constraint violations > 0.2 Å (per structure)		1.81 ± 2.07
Dihedral constraint violations > 3.0° (per structure)		1.44 ± 0.96
X-PLOR potential energy (E_{L-J}, Kcal/mol, avg. per structure)		−501 ± 28.9
R.m.s.d. to the mean for backbone heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		0.62 ± 0.07
R.m.s.d. to the mean for heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		1.20 ± 0.14
R.m.s.d. to the mean for backbone heavy atoms (residues 425–590)		1.14 ± 0.11
R.m.s.d. to the mean for heavy atoms (residues 425–590)		1.58 ± 0.10
R.m.s.d. to the mean for DNA heavy atoms (superposition of all DNA heavy atoms)		0.83 ± 0.21
R.m.s.d. to the mean for protein + DNA heavy atoms (superposition of all β strand heavy atoms and DNA heavy atoms)		1.20 ± 0.16
Ramachandran plot[a]		
	Residues 425–590	Secondary Structures
Most favorable region	60.4%	83.7%
Additionally allowed region	38.9%	14.0%
Generously allowed region	0.7%	2.3%
Disallowed region	0.0%	0.0%

[a] Laskowski et al., 1996.

How are NOE restraints distributed?

Reading the NMR Statistics Structure Table

Table 1. Structural Statistics for NFATC1-DBD*/DNA Complex

Protein		
Amino acids residues sequentially assigned (nonproline)		162 of 167
Effective distance constraints		1050
Intraresidue		236
Sequential (|i − j| = 1)		374
Medium-range (|i − j| ≤4)		86
Long-range (|i −j| ≥5)		272
H bonds		82
Dihedral angle constraints		363
DNA		
Effective distance constraints		276
Intraresidue		146
Sequential		131
Interstrand		1
H bond		58
Protein–DNA interface		
Effective distance constraints		56
Distance constraint violations > 0.2 Å (per structure)		1.81 ± 2.07
Dihedral constraint violations > 3.0° (per structure)		1.44 ± 0.96
X-PLOR potential energy (E_{L-J}, Kcal/mol, avg. per structure)		−501 ± 28.9
R.m.s.d. to the mean for backbone heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		0.62 ± 0.07
R.m.s.d. to the mean for heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		1.20 ± 0.14
R.m.s.d. to the mean for backbone heavy atoms (residues 425–590)		1.14 ± 0.11
R.m.s.d. to the mean for heavy atoms (residues 425–590)		1.58 ± 0.10
R.m.s.d. to the mean for DNA heavy atoms (superposition of all DNA heavy atoms)		0.83 ± 0.21
R.m.s.d. to the mean for protein + DNA heavy atoms (superposition of all β strand heavy atoms and DNA heavy atoms)		1.20 ± 0.16
Ramachandran plot[a]		
	Residues 425–590	Secondary Structures
Most favorable region	60.4%	83.7%
Additionally allowed region	38.9%	14.0%
Generously allowed region	0.7%	2.3%
Disallowed region	0.0%	0.0%

[a] Laskowski et al., 1996.

How were H Bonds Determined?

Reading the NMR Statistics Structure Table

Table 1. Structural Statistics for NFATC1-DBD*/DNA Complex

Protein		
Amino acids residues sequentially assigned (nonproline)		162 of 167
Effective distance constraints		1050
Intraresidue		236
Sequential (|i − j| = 1)		374
Medium-range (|i − j| ≤4)		86
Long-range (|i −j| ≥5)		272
H bonds		82
Dihedral angle constraints		363
DNA		
Effective distance constraints		276
Intraresidue		146
Sequential		131
Interstrand		1
H bond		58
Protein–DNA interface		
Effective distance constraints		56
Distance constraint violations > 0.2 Å (per structure)		1.81 ± 2.07
Dihedral constraint violations > 3.0° (per structure)		1.44 ± 0.96
X-PLOR potential energy (E_{L-J}, Kcal/mol, avg. per structure)		−501 ± 28.9
R.m.s.d. to the mean for backbone heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		0.62 ± 0.07
R.m.s.d. to the mean for heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		1.20 ± 0.14
R.m.s.d. to the mean for backbone heavy atoms (residues 425–590)		1.14 ± 0.11
R.m.s.d. to the mean for heavy atoms (residues 425–590)		1.58 ± 0.10
R.m.s.d. to the mean for DNA heavy atoms (superposition of all DNA heavy atoms)		0.83 ± 0.21
R.m.s.d. to the mean for protein + DNA heavy atoms (superposition of all β strand heavy atoms and DNA heavy atoms)		1.20 ± 0.16
Ramachandran plot[a]		
	Residues 425–590	Secondary Structures
Most favorable region	60.4%	83.7%
Additionally allowed region	38.9%	14.0%
Generously allowed region	0.7%	2.3%
Disallowed region	0.0%	0.0%

[a] Laskowski et al., 1996.

How many intermolecular NOEs?

Reading the NMR Statistics Structure Table

Table 1. Structural Statistics for NFATC1-DBD*/DNA Complex

Protein		
Amino acids residues sequentially assigned (nonproline)		162 of 167
Effective distance constraints		1050
Intraresidue		236
Sequential (|i − j| = 1)		374
Medium-range (|i − j| ≤4)		86
Long-range (|i −j| ≥5)		272
H bonds		82
Dihedral angle constraints		363
DNA		
Effective distance constraints		276
Intraresidue		146
Sequential		131
Interstrand		1
H bond		58
Protein–DNA interface		
Effective distance constraints		56
Distance constraint violations > 0.2 Å (per structure)		1.81 ± 2.07
Dihedral constraint violations > 3.0° (per structure)		1.44 ± 0.96
X-PLOR potential energy (E_{L-J}, Kcal/mol, avg. per structure)		−501 ± 28.9
R.m.s.d. to the mean for backbone heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		0.62 ± 0.07
R.m.s.d. to the mean for heavy atoms of all β strands (residues 428–432, 460–464, 472–479, 493–495, 506–510, 515–521, 528–538, 559–568, 578–583)		1.20 ± 0.14
R.m.s.d. to the mean for backbone heavy atoms (residues 425–590)		1.14 ± 0.11
R.m.s.d. to the mean for heavy atoms (residues 425–590)		1.58 ± 0.10
R.m.s.d. to the mean for DNA heavy atoms (superposition of all DNA heavy atoms)		0.83 ± 0.21
R.m.s.d. to the mean for protein + DNA heavy atoms (superposition of all β strand heavy atoms and DNA heavy atoms)		1.20 ± 0.16
Ramachandran plot[a]		
	Residues 425–590	Secondary Structures
Most favorable region	60.4%	83.7%
Additionally allowed region	38.9%	14.0%
Generously allowed region	0.7%	2.3%
Disallowed region	0.0%	0.0%

[a] Laskowski et al., 1996.

Criteria for publication:
No distance violations >0.5 Å
No dihedral angle violations >5°

How many violations ?

Read methods section to evaluate table

Resonance Assignments

1-HNCA/HN(CO)CA
2-Amino-acid specific labeling
3-NOESY experiments:
Homonuclear and heteronuclear

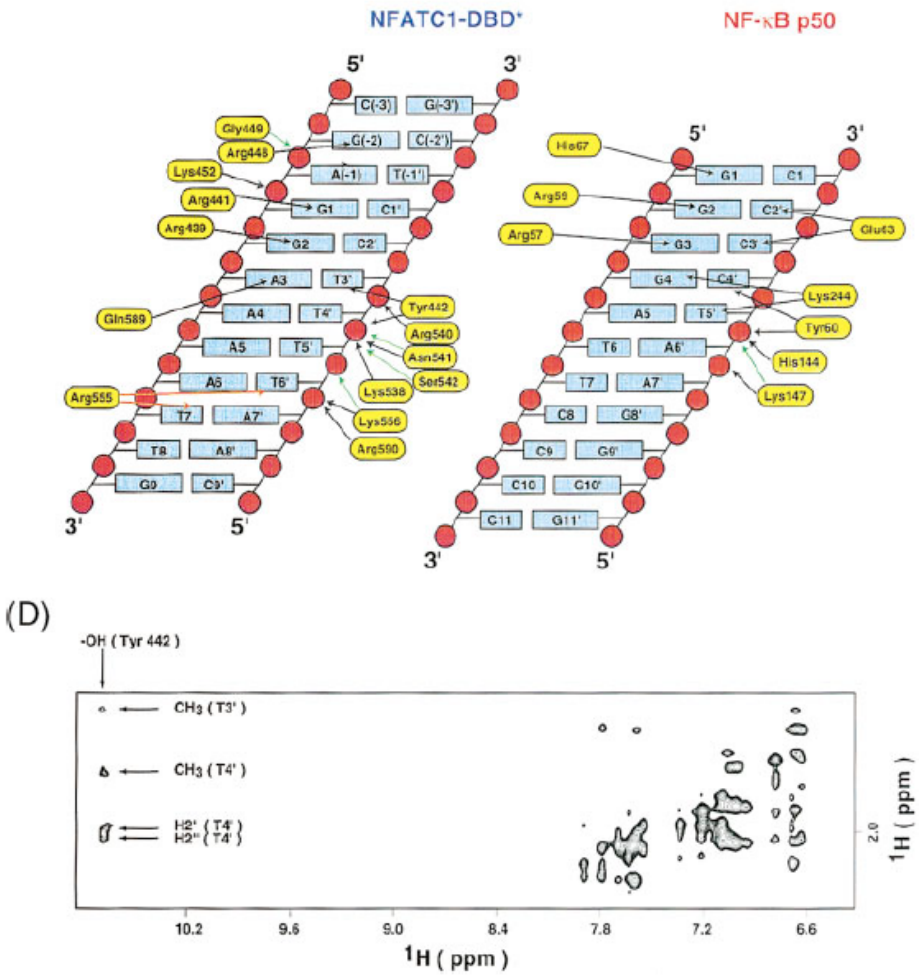
Distance Restraints

1-2D and 3D NOESY at longer mixing times to get weak NOEs
2-HD exchange to get hydrogen bonds

Many of these experiments performed with a specific labeling scheme to facilitate NOEs assignment

Side chain assignments from NOESY, caveats?

Assymetric isotope labelling to get intermolecular NOEs



Protein is deuterated, DNA protonated experiment done in D2O solvent

Caveats?

Structures of larger proteins and complexes

Sample Deuteration Increases Sensitivity and Resolution

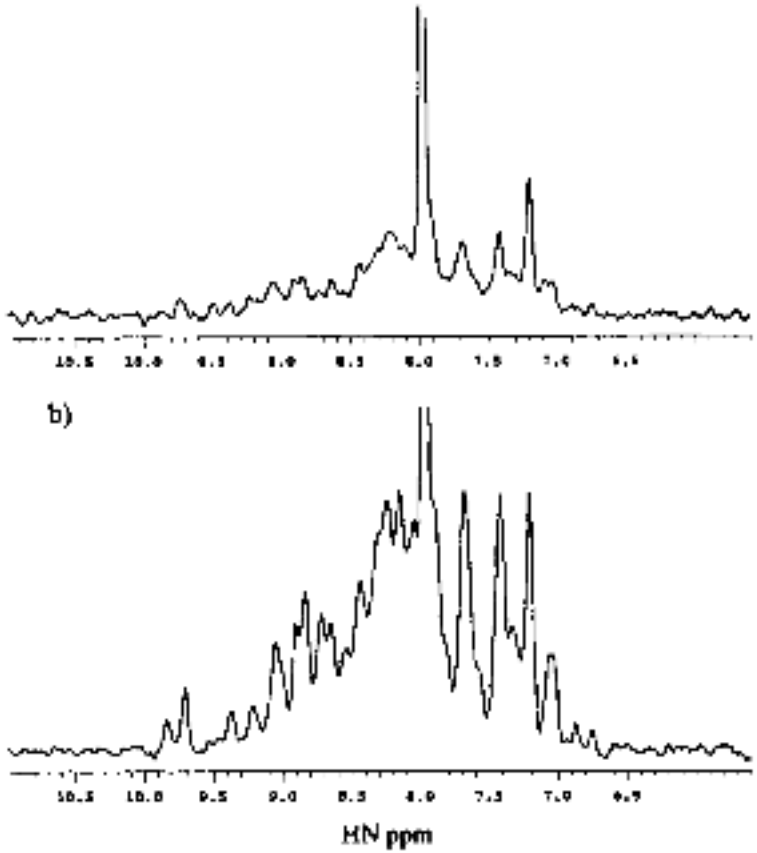
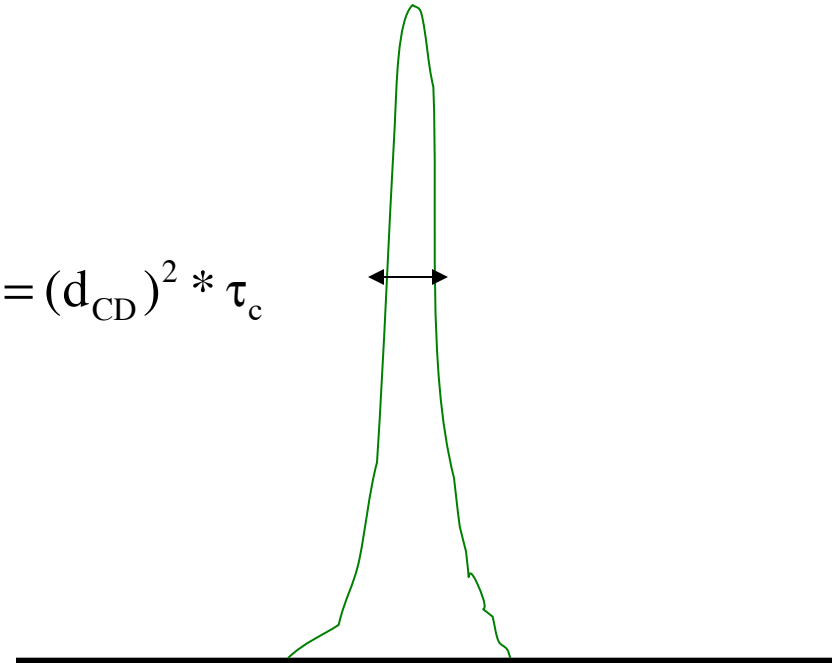


Figure 14. Comparative 1D spectra of fully protonated (a) and ~70% deuterated (b) ^{15}N, ^{13}C trpR using the HN(CA)CB scheme (Figure 2c). A total of 1024 transients were recorded for the ^{2}H sample and 6928 transients (1024×2.6^2) for the ^{1}H sample which is a factor of 2.6 more dilute. The rms noise levels are normalized in the plots.

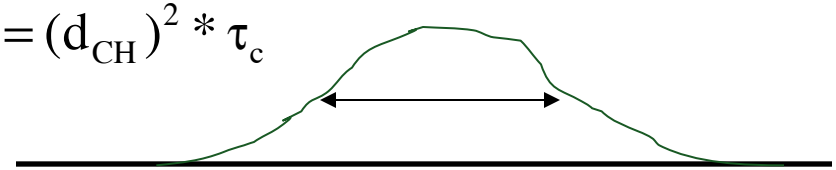
Why does deuteration help?

$$d_{CH} = \frac{\gamma_C \gamma_H}{r^3} \frac{3\mathrm{Cos}^2(\theta) - 1}{2}$$

$$\frac{1}{T_2} = (d_{CD})^2 * \tau_c$$



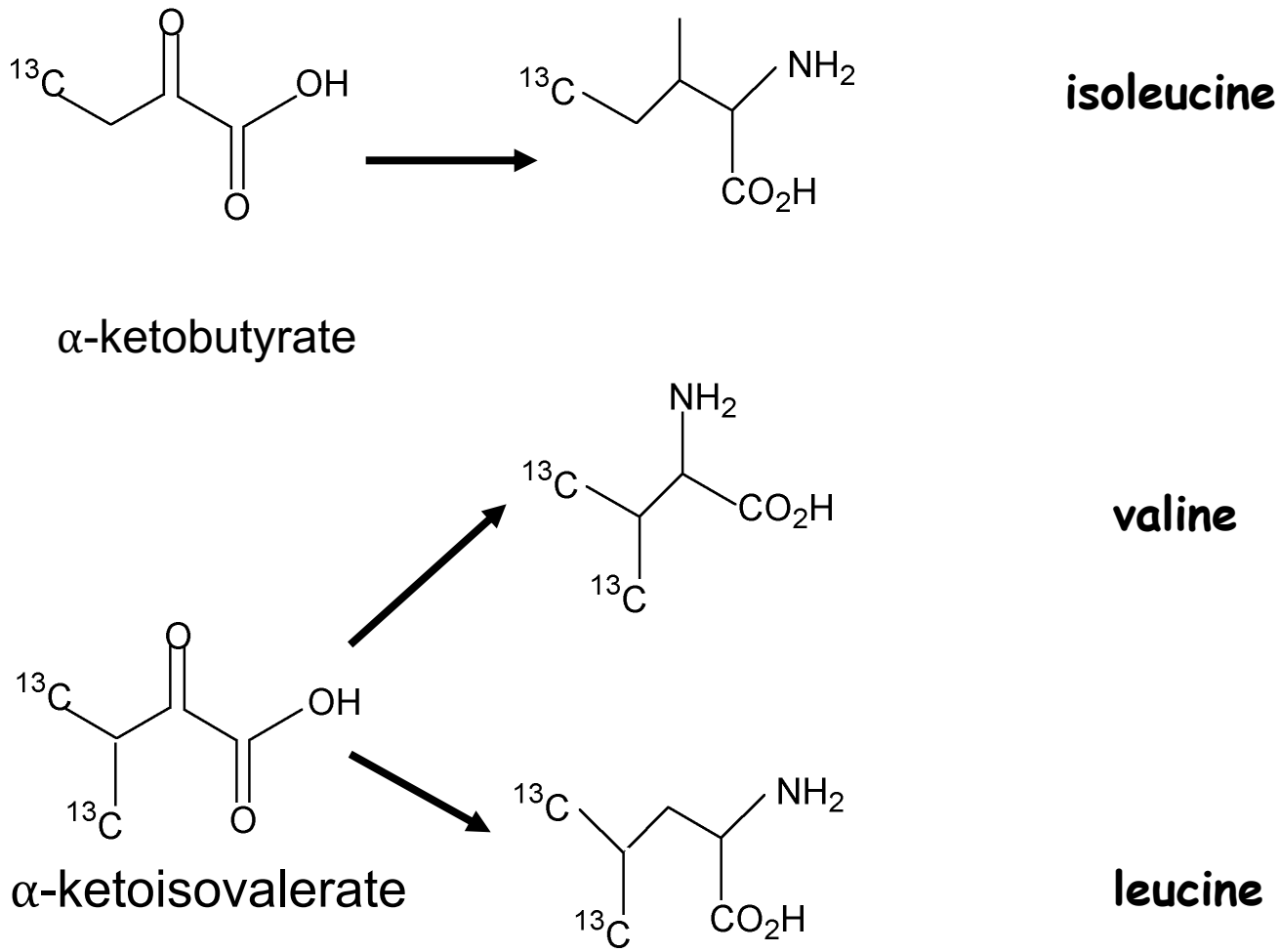
$$\frac{1}{T_2} = (d_{CH})^2 * \tau_c$$



Dipolar coupling for CH spin pair 6.5 times stronger than for CD--->roughly 50 fold reduction in linewidth with increase in S:N.

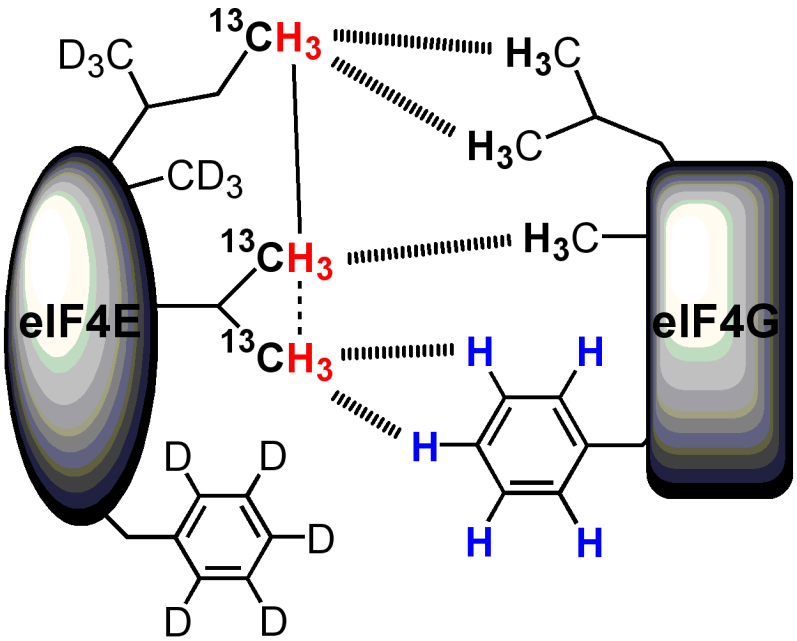
Need methods for measuring distance restraints in sparse ^{1}H environment

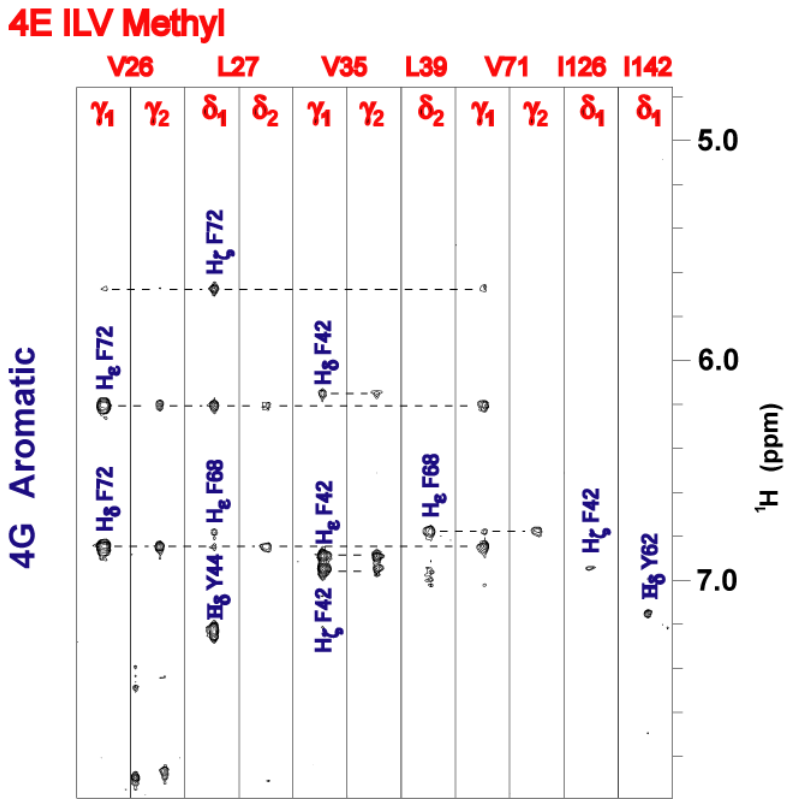
Selective Re-introduction of Protons for NOE experiments



Grow E. coli in D2O and deuterated glucose, add precursors to introduce 1H/13C methyl labels

Measurement of Intermolecular NOEs using Asymmetric Deuteration with ILV Labeling

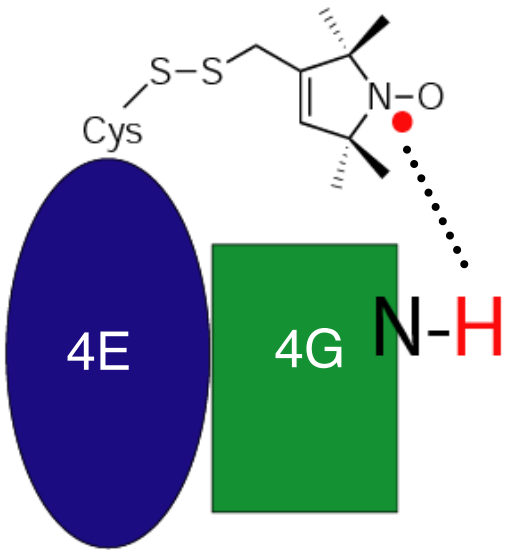




Aromatic/methyl NOEs are unambiguously identified

Gross , Gelev and Wagner, J Biomol NMR 2003

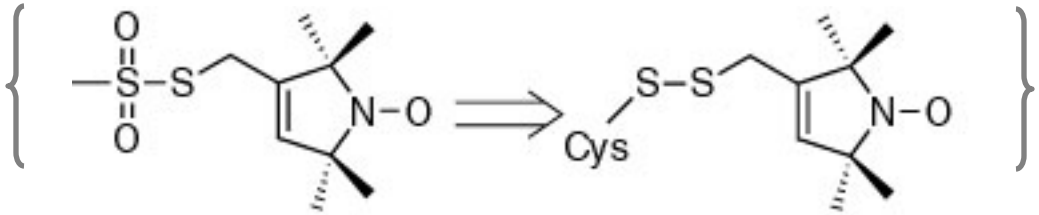
Determining Long Range Distances through Paramagnetic Relaxation Enhancements



Dipolar broadening between unpaired electron and ^{1}H: $1/r^{6}$ dependence.

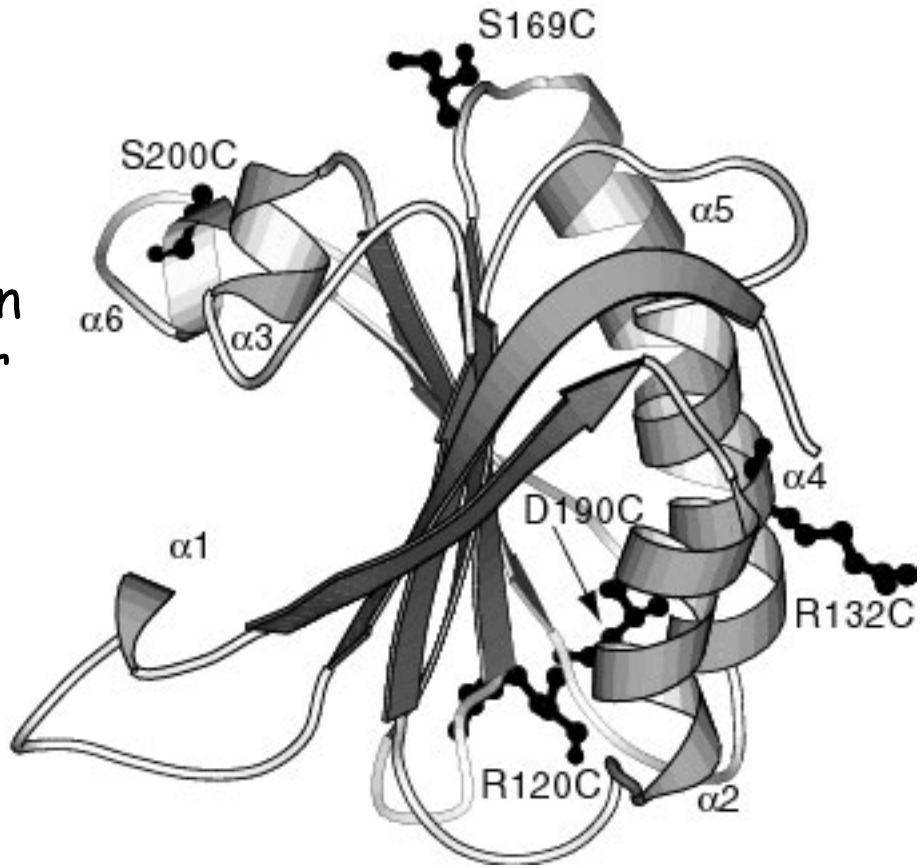
Provides long range distance information (15-20 angstroms)

Site-Directed Spin Labeling for PRE



MTSL-Cysteine Adduct

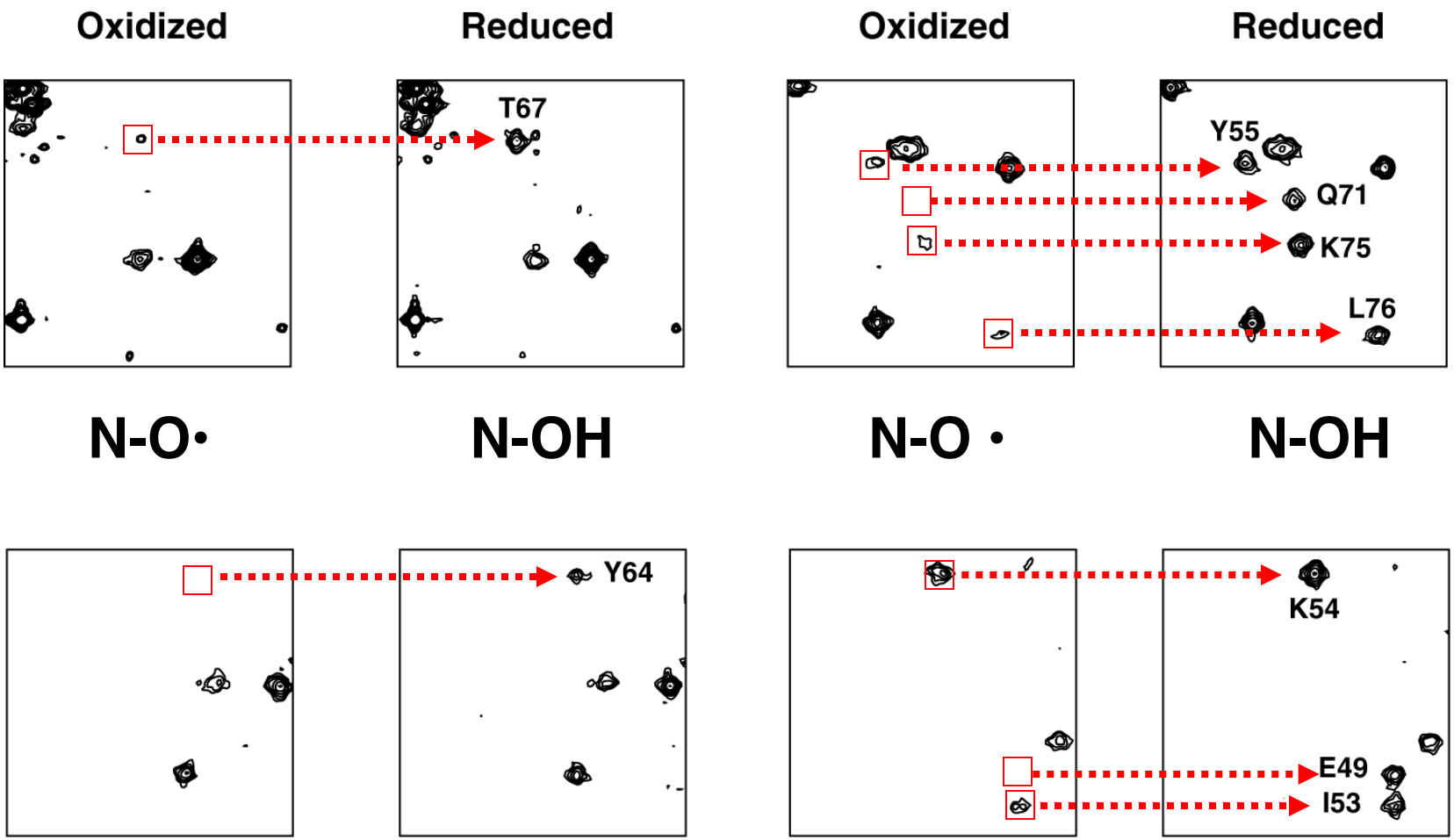
Cysteine mutation in surface loop or helix



Record HSQC In absence and presence of reductant

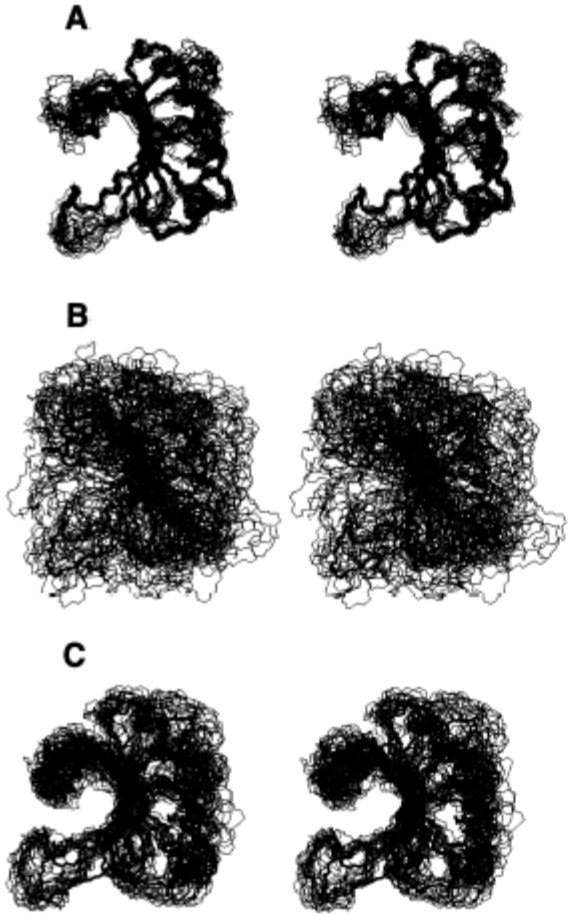
Battiste and Wagner , Biochemistry, 2000

Paramagnet Relaxation Enhancement from Site-Directed Spin Labelling



Intensity reduction from dipolar coupling ($1/r^6$), so distances can be extracted

Impact of PRE restraints for structure determination



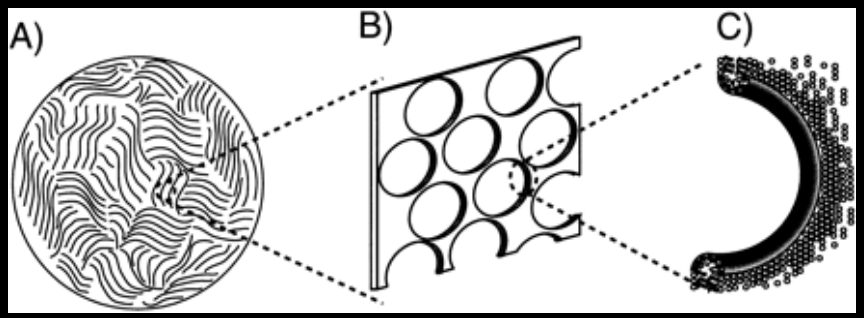
ALL NOEsm (2014)

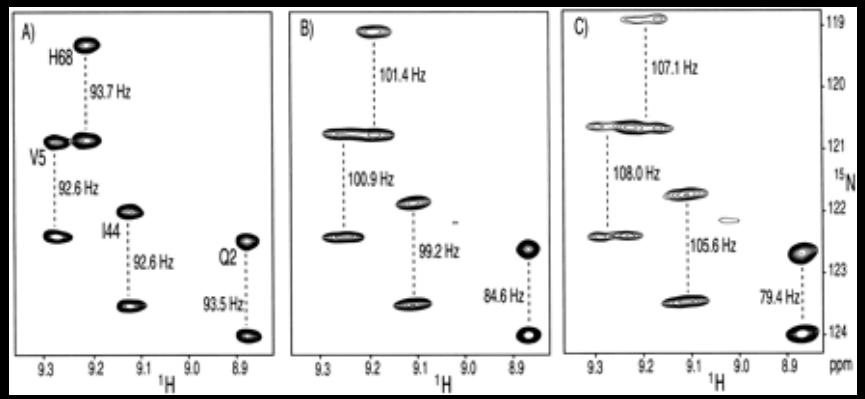
ONLY HN-HN NOEs (403)

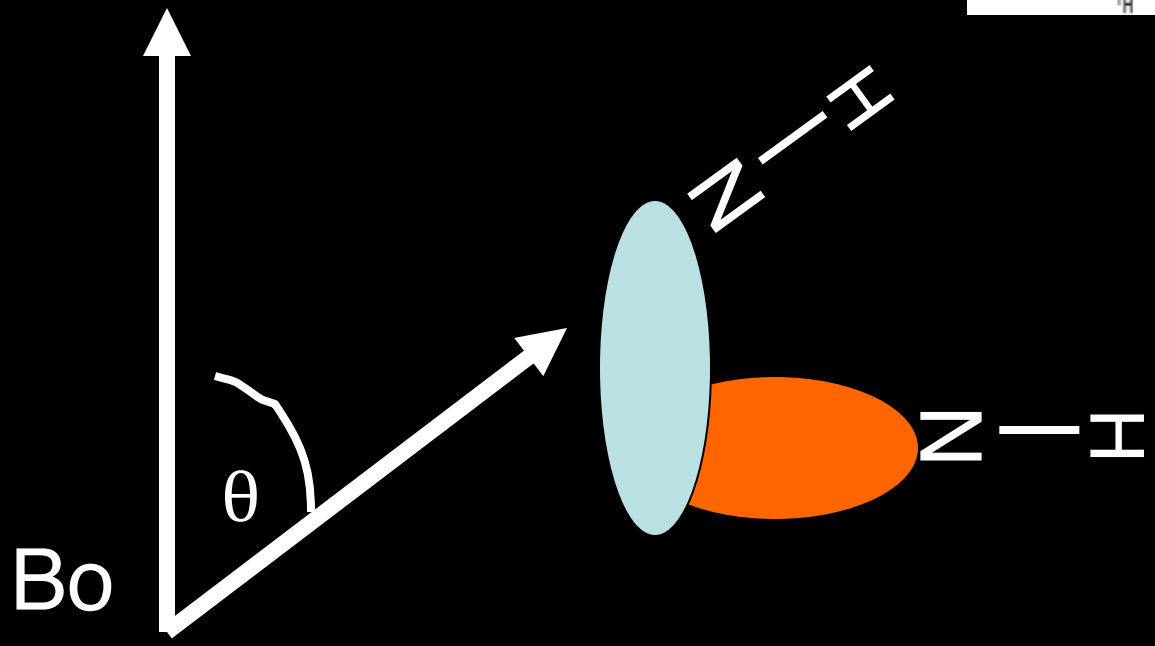
HN-HN NOEs + PRE (515)

Battiste and Wagner , Biochemistry, 2000

Residual Dipolar Couplings



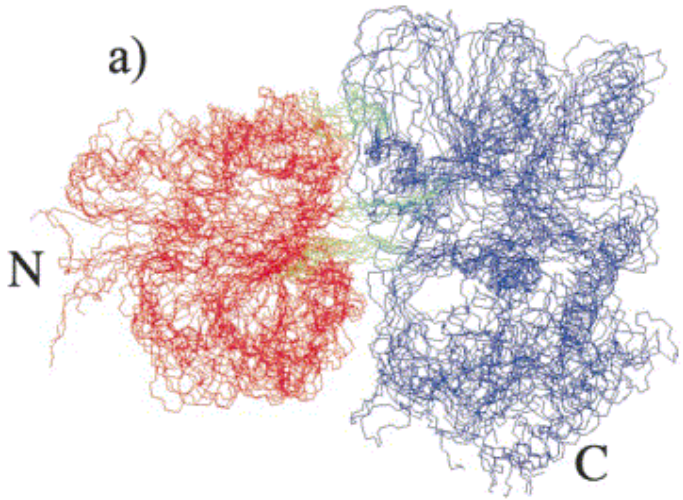


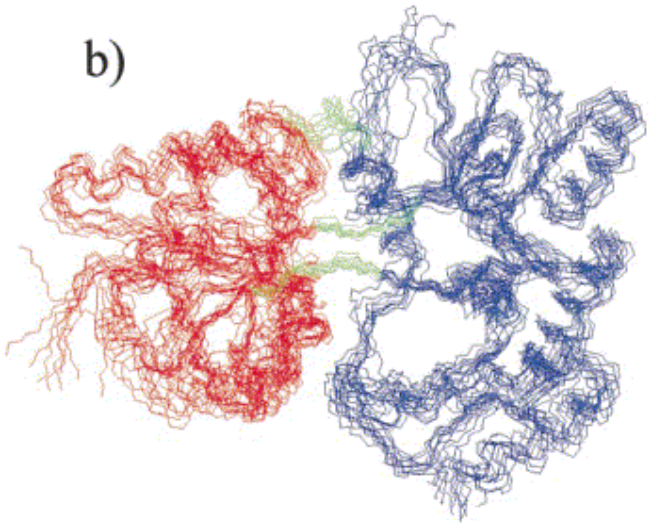


$$d_{NH}\,[3\cos^2\theta-1]/2$$

$$d_{NH} \sim \frac{\gamma_C \gamma_H}{r^3}$$

Impact of RDCs on Precision and Accuracy: MBP, a 42 kDa test Case





RMSD

Precision:	5.5Å	2.2 Å
Accuracy:	5.1Å	3.3 Å

From Mueller et. Al. JMB 300(1) 197-212 2000

Prospects for even Larger Proteins

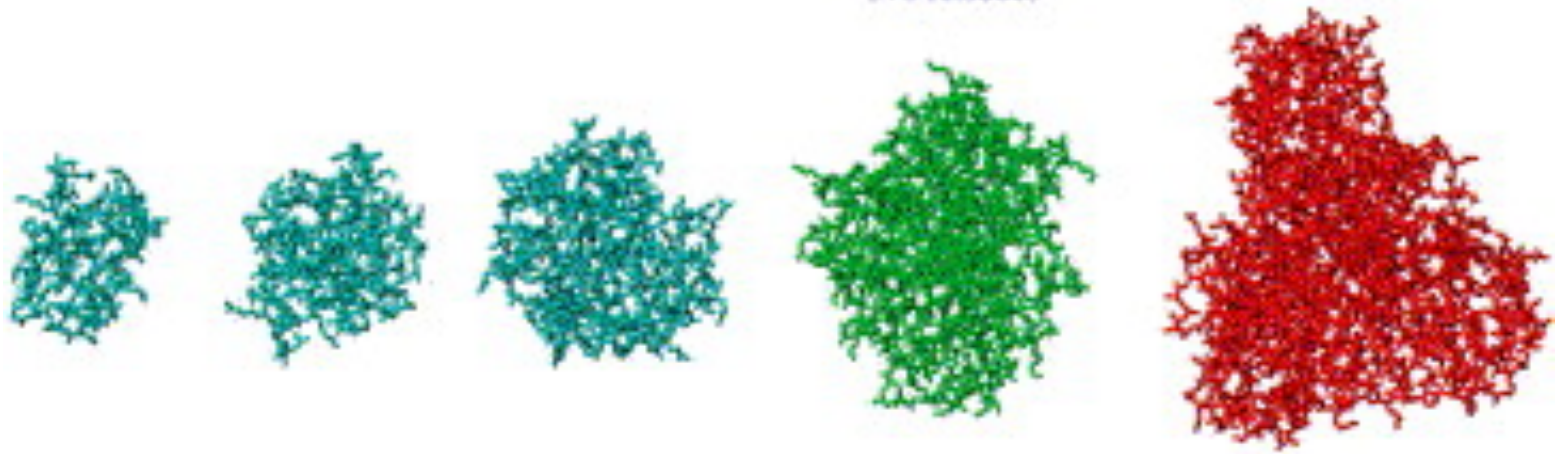
Ubiquitin
76 residues

Lysozyme
156 residues

Adenylate Kinase
214 residues

Maltose binding
protein (MBP)
370 residues

Malate synthase G
723 residues



i. NMR assignments
ii. Structure determination
iii. Dynamics

requires ^{15}N, ^{13}C labeling

i. NMR assignments
ii. Global fold/structure determination
iii. Domain orientation
iv. Dynamics

requires ^{15}N, ^{13}C, ^{2}H labeling

?

X-ray vs NMR

- crystal
- single structure-best fit to electron density

- solution
- ensemble of 20-50 structures that equally fit experimental data

Limitations of NMR

- small proteins (20-30 kD max, although this is changing)
- must be soluble and nonaggregating at 1-3 mM conc
- lots of protein needed

Advantages of NMR

- don't need crystal
- observe protein in solution
- more than a method for determining structure

dynamics
ligand binding (drug/protein/DNA/etc)
protein folding
conformational change
chemistry, chemical reactions, protonation states.....