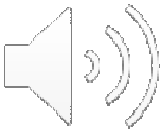
What NMR can tell about proteins

For Application to Protein Characterization

by
Radovan Fiala, Karel Kubicek, and Pavel Kaderavek
CEITEC, Masaryk University

This project received funding by the European Commission under the GA n° 2019-1-IT02-KA203-063391



How to prepare and what to expect from NMR spectroscopy

A] Scientific question / I'd like to obtain:

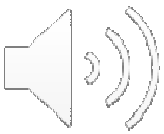
1) Structure
2) Relaxation properties
3) Interaction at atomic level resolution
4) Analysis (NMR is also analytical method)
5) Image (MRI)

B] What do I need to provide / prepare?

C] How much would it cost?

D] How long would it take?

E] Can I do it on my own (or a sudent/colleague of mine)

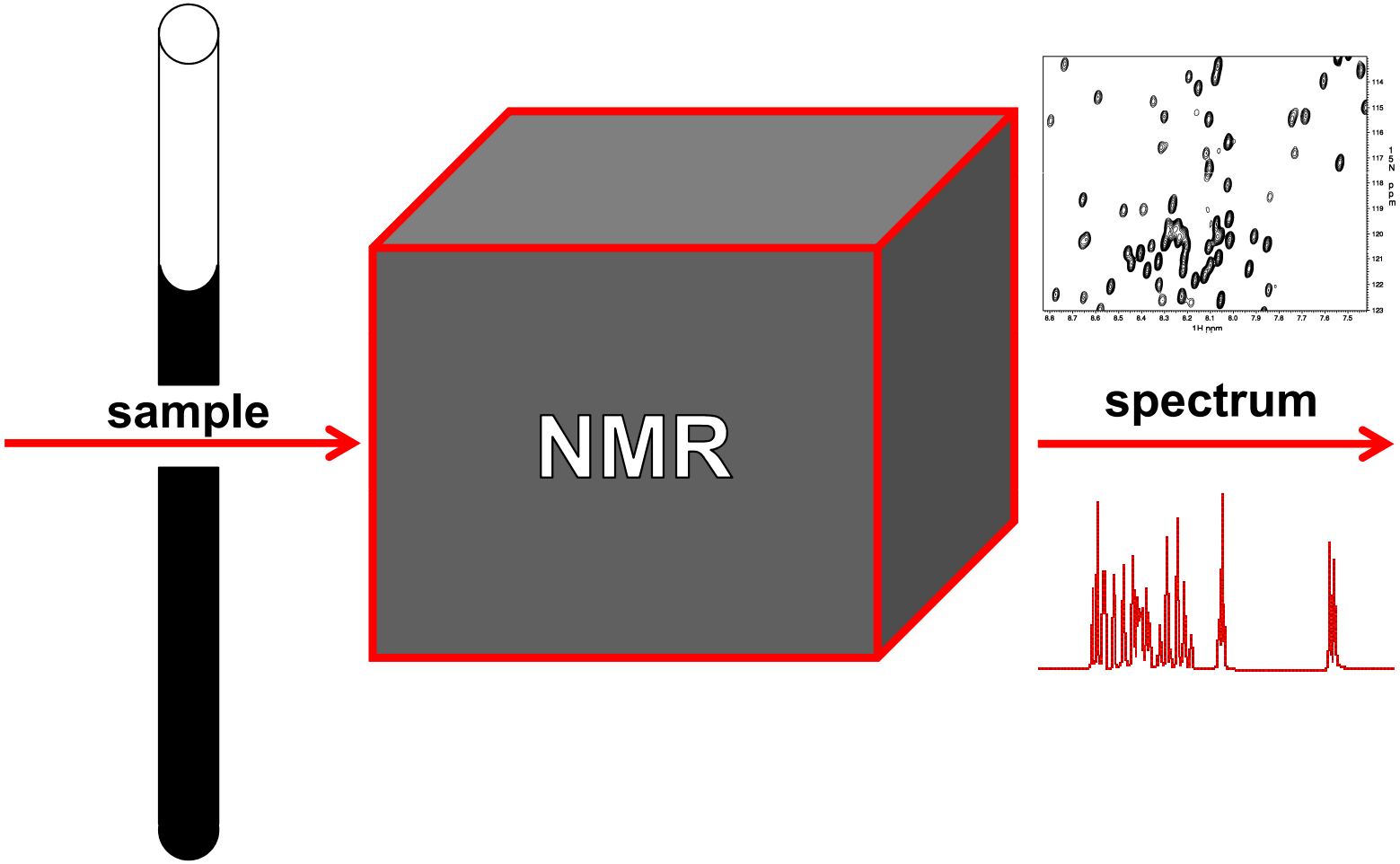


200-500 µl of
100-1000 µM
compound

Method of choice

Data to be analyzed

Results

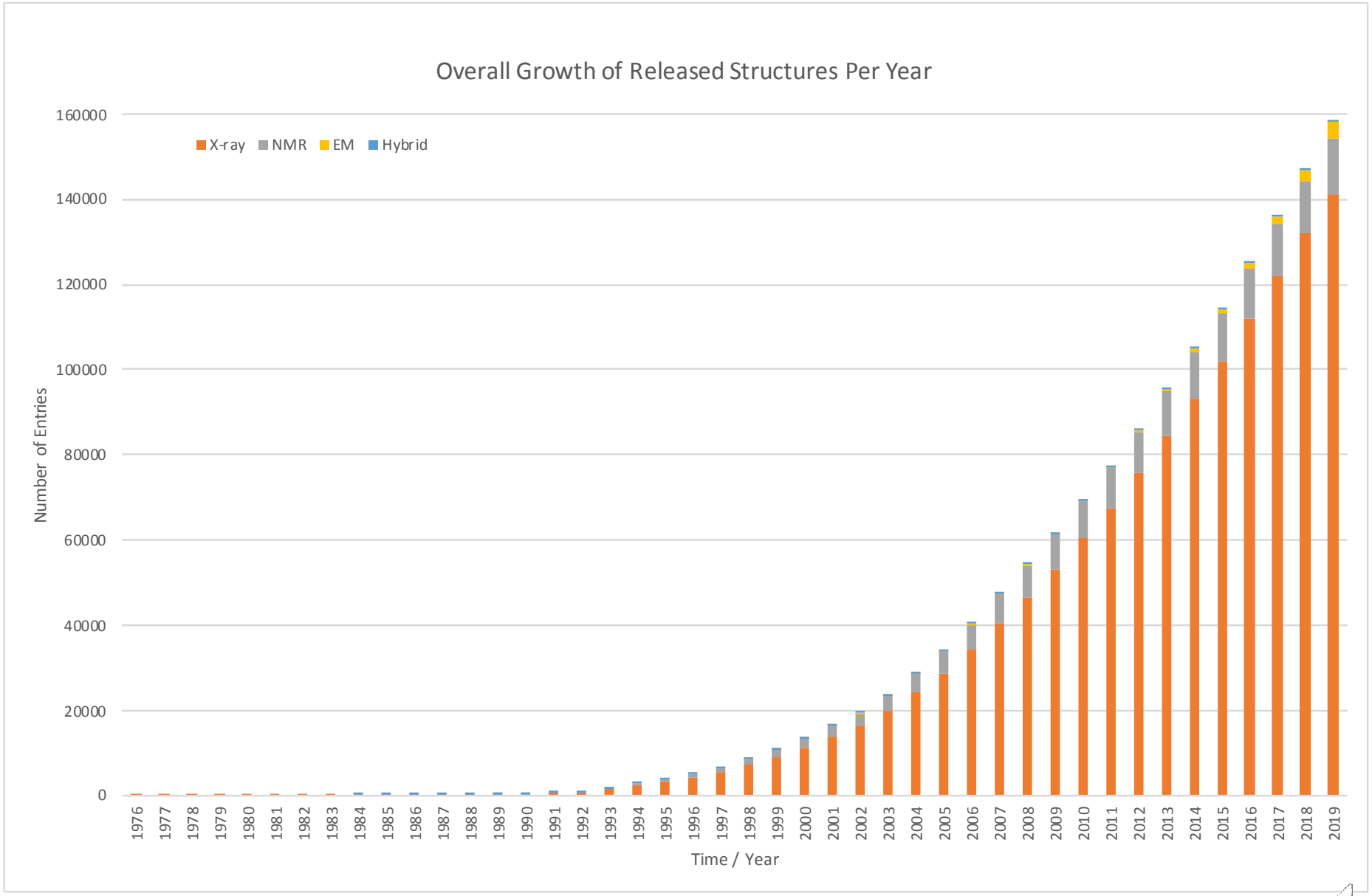


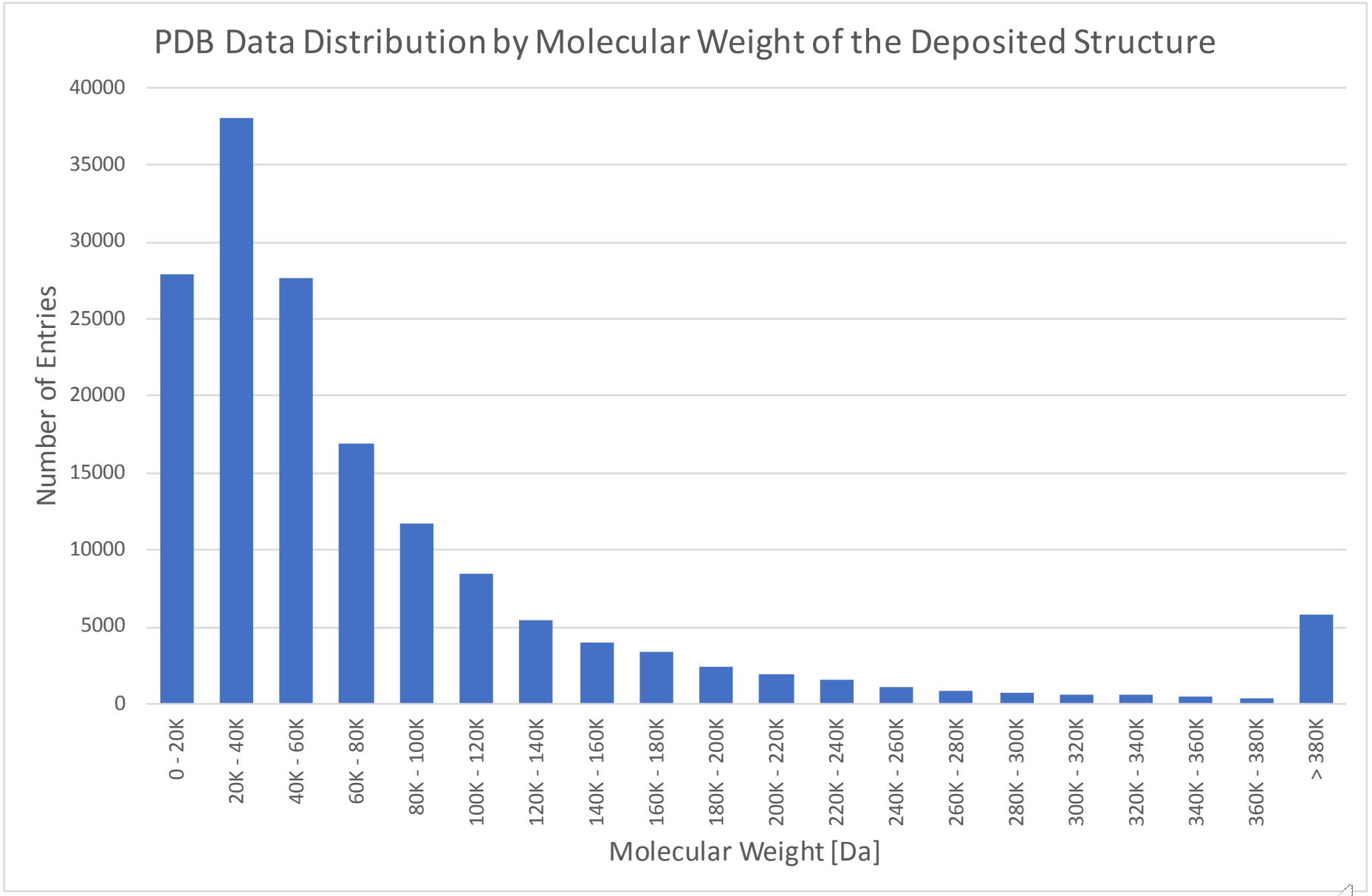
1) Structure
2) Relaxation properties
3) Interaction at atomic level resolution
4) Analysis
5) Image

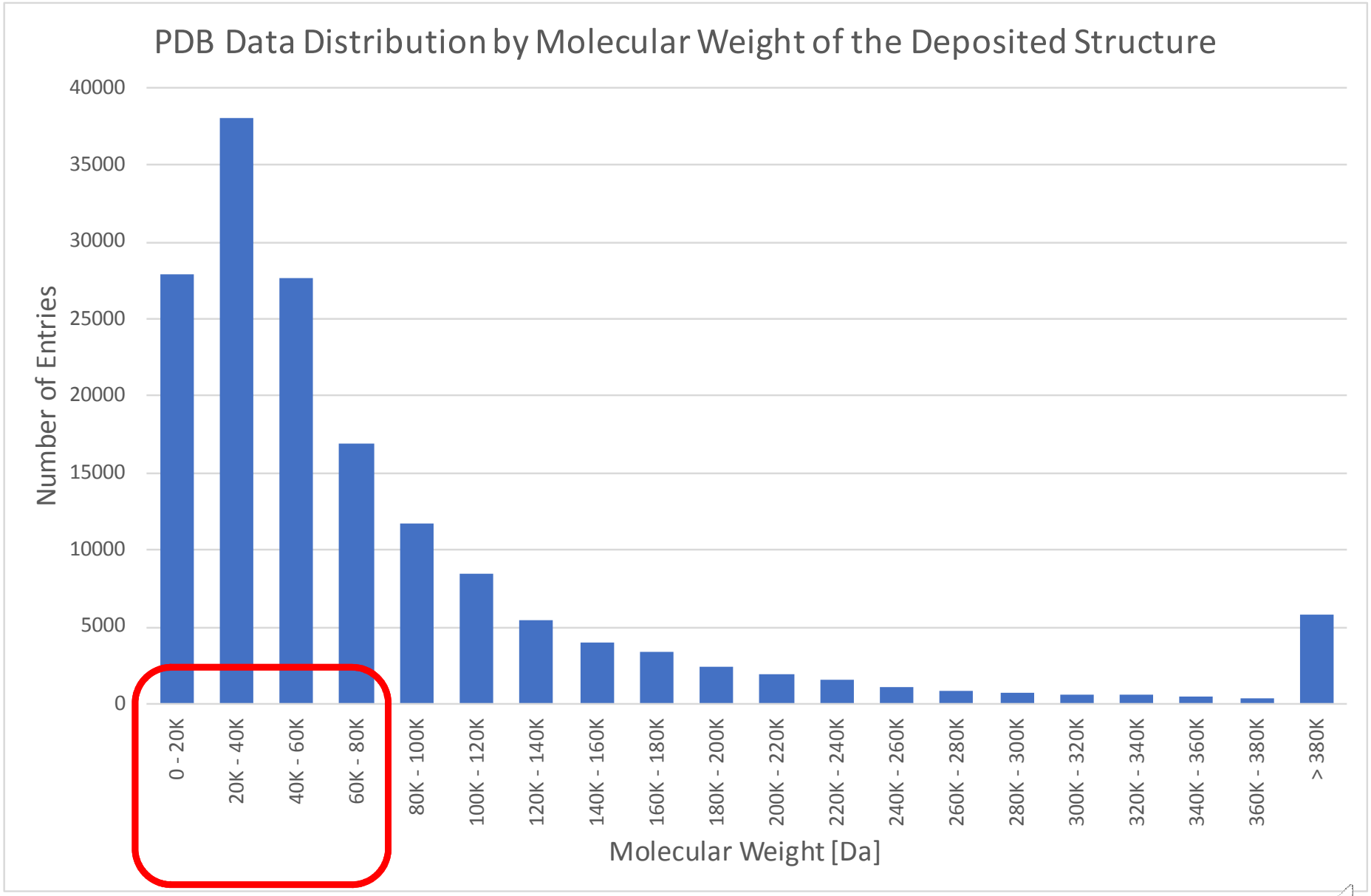
PDB Data Distribution by Experimental Method and Molecular Type

Exp. Method	Proteins	NA	Protein/NA Complex	Other	Total
X-Ray	133461	2098	7091	8082	**150732**
NMR	11493	1309	267	92	**13161**
EM	4181	47	1478	501	**6207**
Other	32	1	0	4	**37**
Multi Method	162	6	3	6	**177**

Source: http://www.rcsb.org as of 03Nov2020


Overall Growth of Released Structures Per Year
X-ray
NMR
EM
Hybrid
Number of Entries
160000
140000
120000
100000
80000
60000
40000
20000
0
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Time / Year


PDB Data Distribution by Molecular Weight of the Deposited Structure
Number of Entries
40000
35000
30000
25000
20000
15000
10000
5000
0
0 - 20K
20K - 40K
40K - 60K
60K - 80K
80K - 100K
100K - 120K
120K - 140K
140K - 160K
160K - 180K
180K - 200K
200K - 220K
220K - 240K
240K - 260K
260K - 280K
280K - 300K
300K - 320K
320K - 340K
340K - 360K
360K - 380K
> 380K
Molecular Weight [Da]


PDB Data Distribution by Molecular Weight of the Deposited Structure
Number of Entries
40000
35000
30000
25000
20000
15000
10000
5000
0
0 - 20K
20K - 40K
40K - 60K
60K - 80K
80K - 100K
100K - 120K
120K - 140K
140K - 160K
160K - 180K
180K - 200K
200K - 220K
220K - 240K
240K - 260K
260K - 280K
280K - 300K
300K - 320K
320K - 340K
340K - 360K
360K - 380K
> 380K
Molecular Weight [Da]

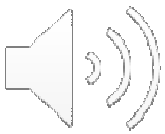
NMR
Safety First

Composition of the Earth's Crust, Seawater, and the Human Body*

Earth's Crust		Seawater		Human Body†	
Element	%	Compound	m*M*	Element	%
O	47	Cl^-	548	H	63
Si	28	Na^+	470	O	25.5
Al	7.9	Mg^{2+}	54	C	9.5
Fe	4.5	SO_4^{2-}	28	N	1.4
Ca	3.5	Ca^{2+}	10	Ca	0.31
Na	2.5	K^+	10	P	0.22
K	2.5	HCO_3^-	2.3	Cl	0.08
Mg	2.2	NO_3^-	0.01	K	0.06
Ti	0.46	HPO_4^{2-}	<0.001	S	0.05
H	0.22			Na	0.03
C	0.19			Mg	0.01

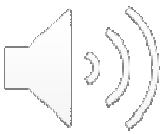
*Figures for the earth's crust and the human body are presented as percentages of the total number of atoms; seawater data are millimoles per liter. Figures for the earth's crust do *not* include water, whereas figures for the human body do.

†Trace elements found in the human body serving essential biological functions include Mn, Fe, Co, Cu, Zn, Mo, I, Ni, and Se.



For most of modern NMR applications, ^{13}C, ^{15}N, and often ^{2}H needed

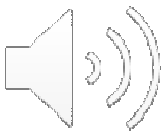
Isotope	Ground state spin	Natural abundance [%]	Rel. Sensitivity
^{1}H	½	~100	$1.00x10^{+0}$
^{13}C	½	1.10	$1.59x10^{-2}$
^{15}N	½	0.37	$1.04x10^{-3}$
^{19}F	½	100	$8.30x10^{-1}$
^{31}P	½	~100	$6.63x10^{-2}$
^{12}C	0	98.90	-
^{16}O	0	~100	-



Samples:

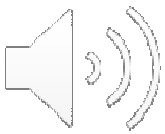
1) Small organic molecules - synthesis
2) Peptides (10-40 aa) - synthesis
3) Proteins (40-200 aa) - $^{13}C/^{15}N$ enriched media
4) Large proteins > 200 aa - $^{2}H/^{13}C/^{15}N$ enriched media

If high concentrations (>1 mM) multidimensional spectra in natural abundance can be performed



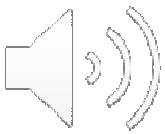
NMR as a tool for structure determination of proteins and biological complexes:

From single proteins to large systems



Conventional scheme for structure determination by NMR

1) Sample preparation / protein expression
2) NMR data acquisition and processing
3) **Backbone chemical shifts assignment**
4) Assignment of side-chain chemical shifts
5) Peak-picking of NOESY spectra
6) Structure calculation
7) (iterative improvement of 5 and 6)



Before we start with assignment of backbone chemical shifts we need to know:

1) Protein primary sequence

MQQDDDFQNF VATLESFKDL KSGISGSRIK KLTTYALDHI DIESKIISLI
IDYSRLCPDS HKLGSLYIID SIGRAYLDET RSNSNSSSNK PGTCAHAINT
LGEVIQELLS DAIAKSNQDH KEKIRMLLDI WDRSGLFQKS YLNAIRSKCF
AMDLEHHHHHH

2) Standard chemical shifts of Cα and Cβ (*vide infra*)

3) Secondary structure prediction (values of 2) will be affected accordingly)

4) Precise peak-picking (manually or semi-automatically)

5) $^{13}C/^{15}N$ uniformly labeled protein (6-30kDa)

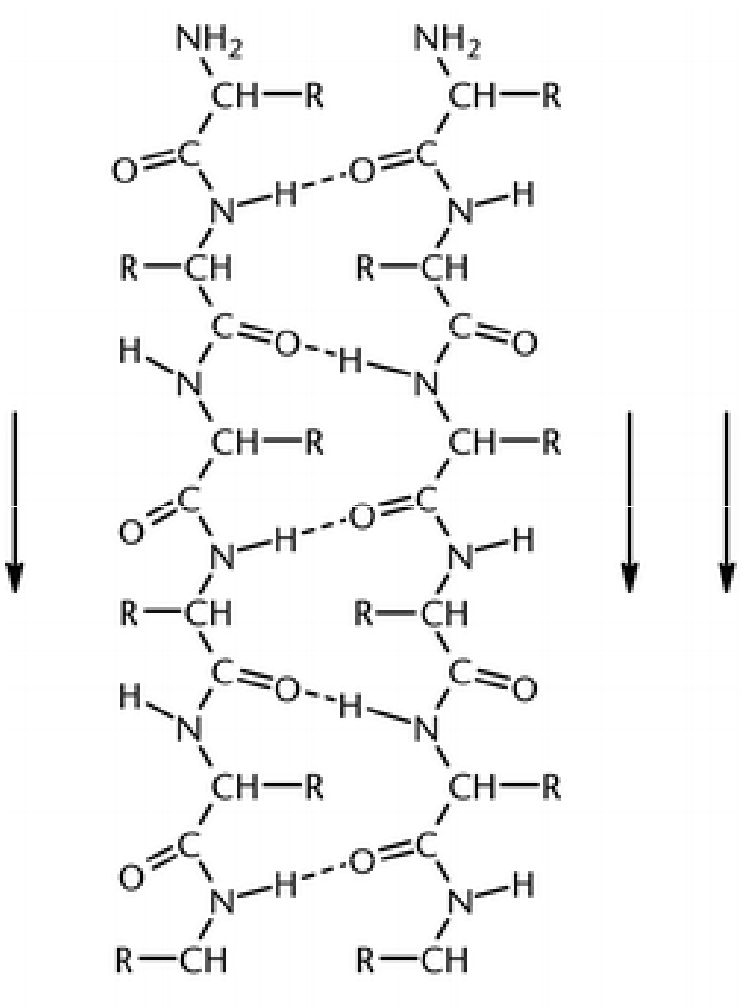
Protein backbone and secondary structure

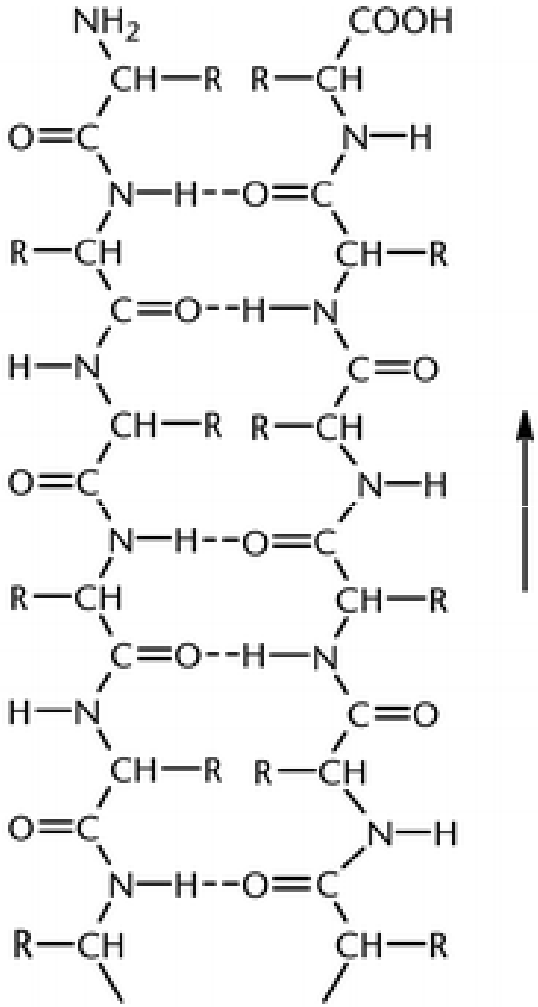
β-strand

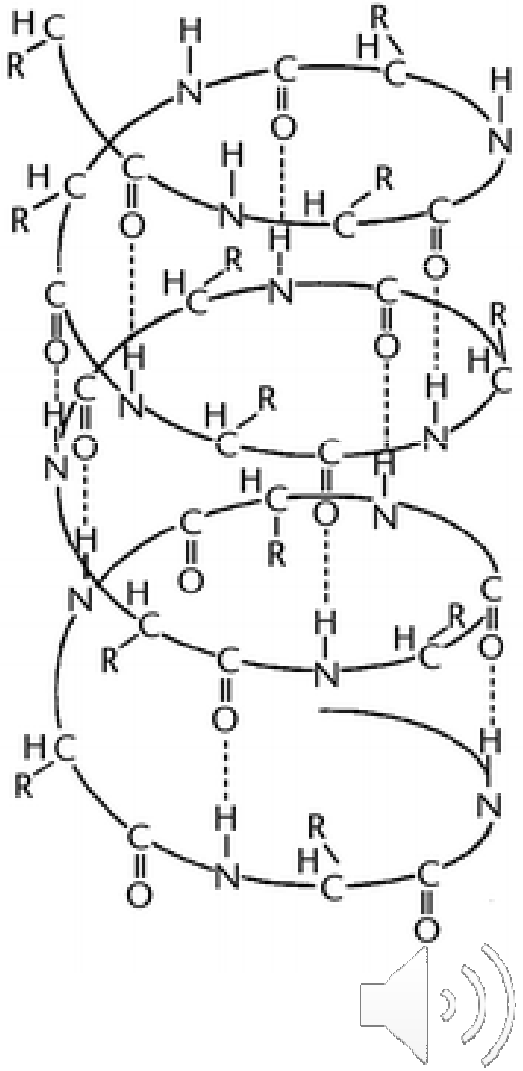
parallel

antiparallel

α-helix



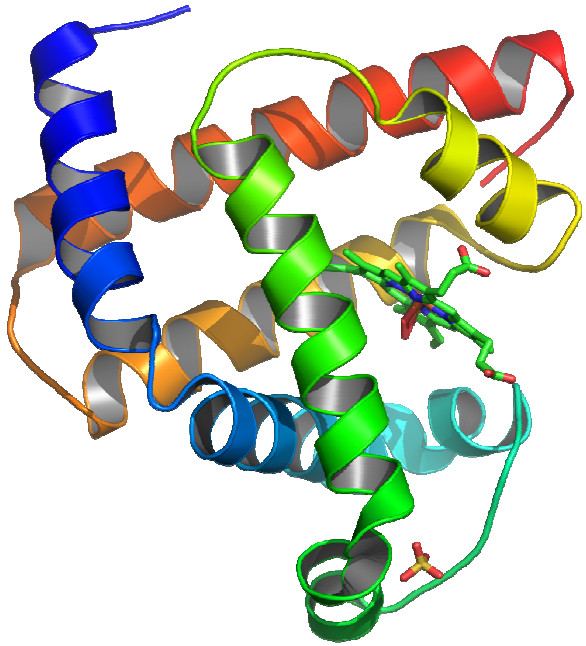




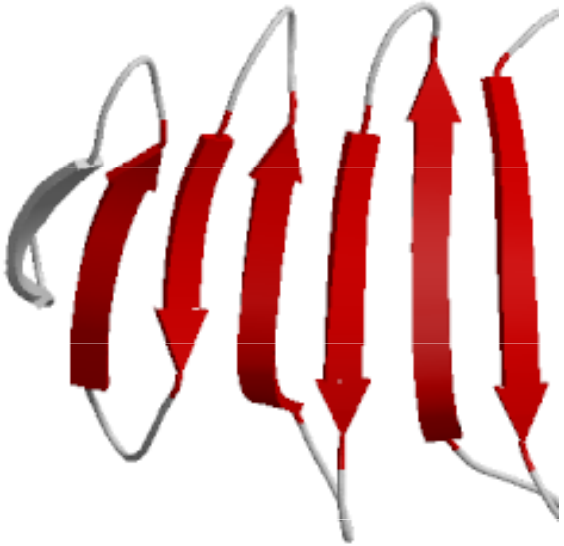
N- ⟶ C-terminus

Hydrogen bond

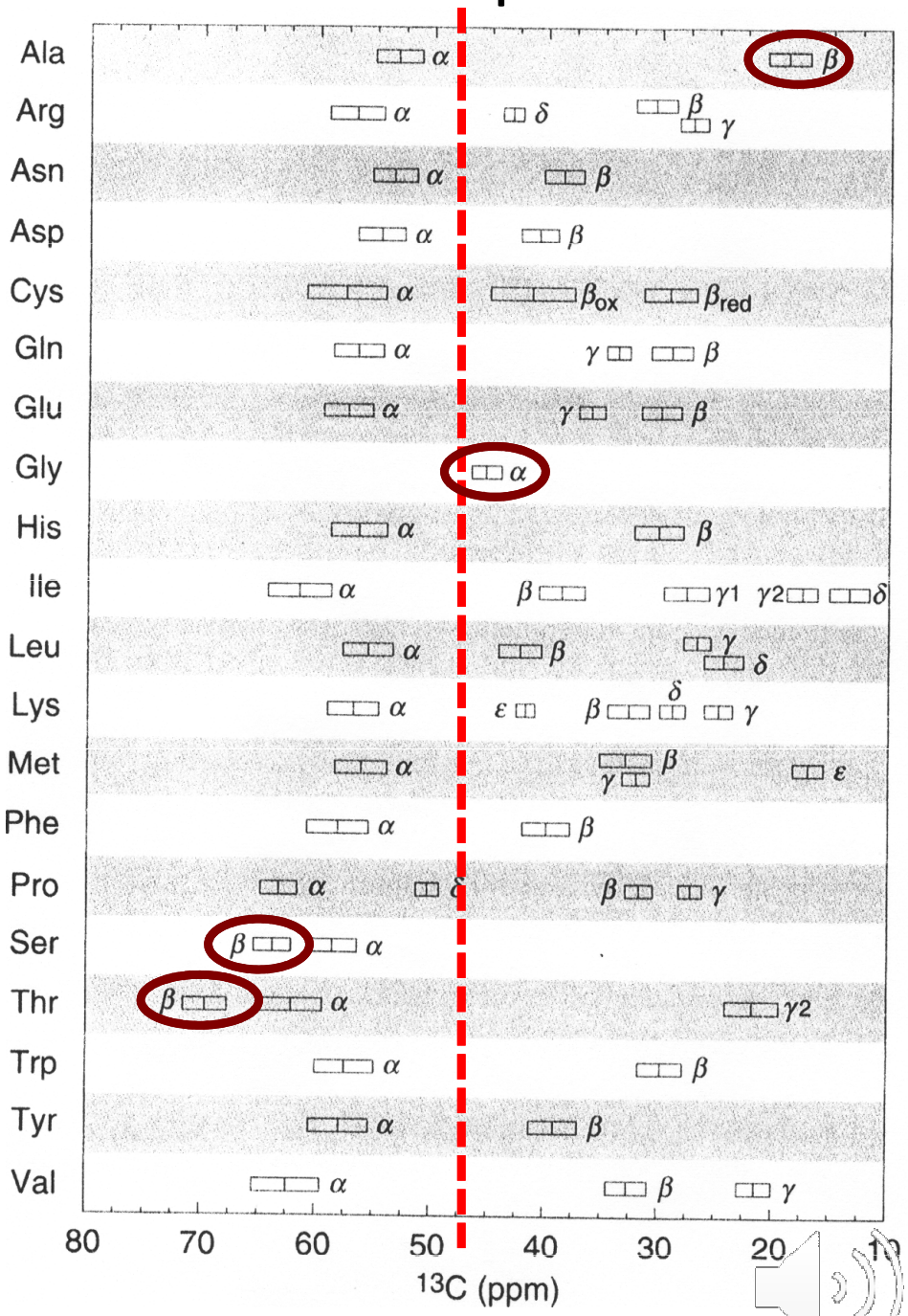
α-helix



β-strand

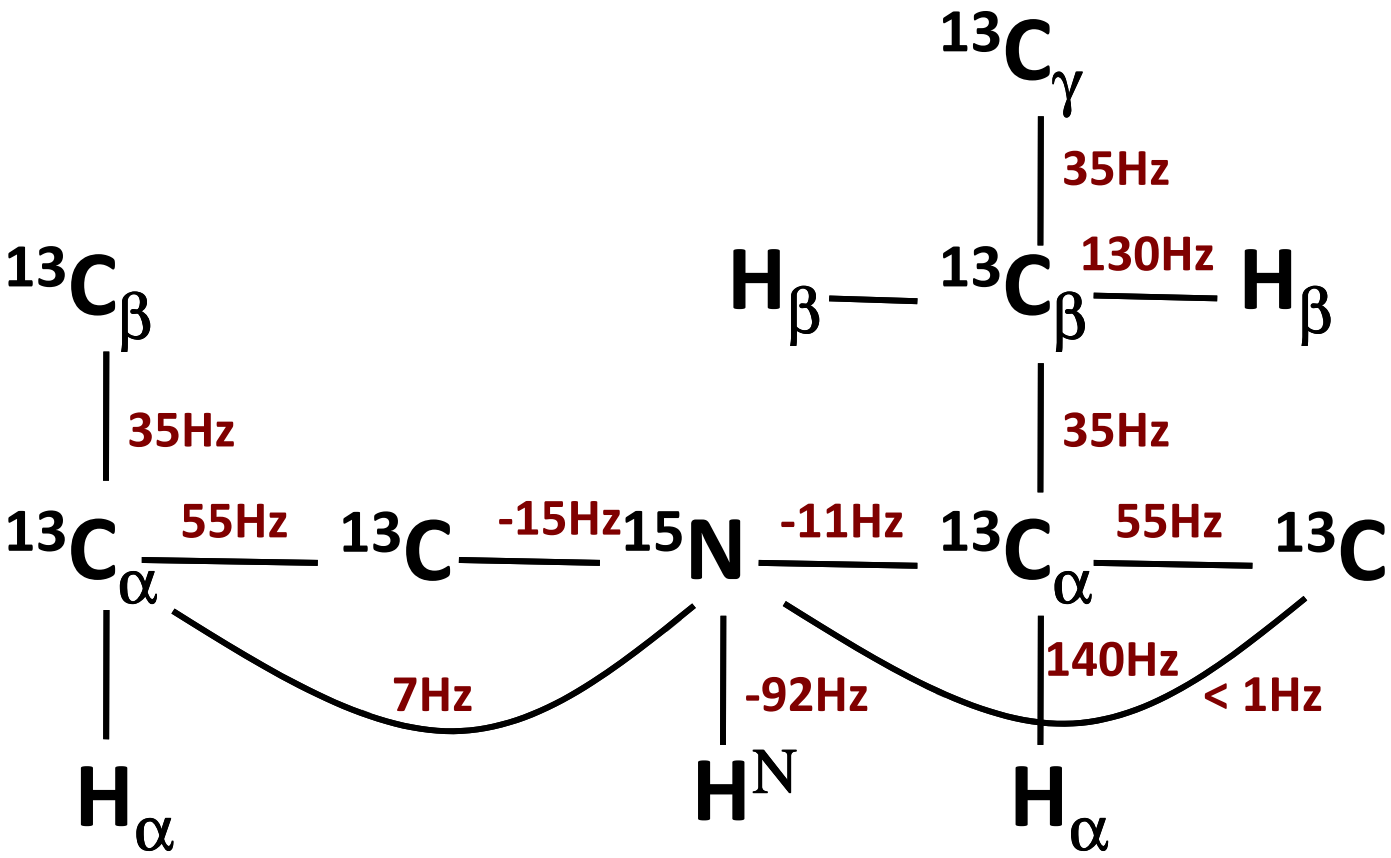


^{13}C chem. shift in proteins

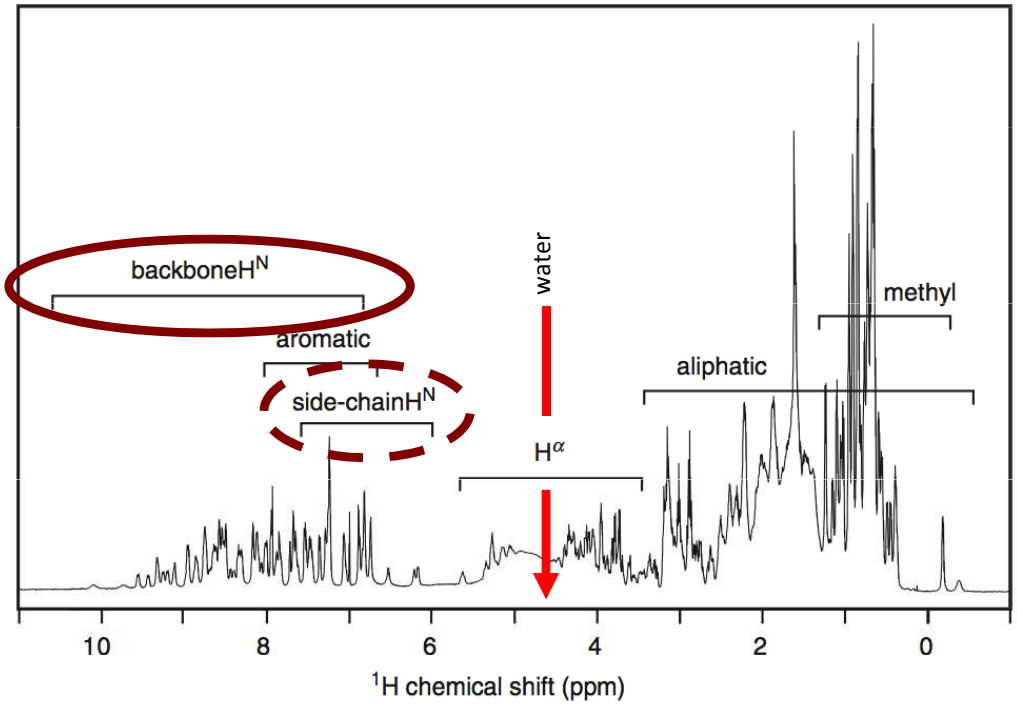


Cavanagh et al. Protein NMR Spectroscopy 2nd ed.

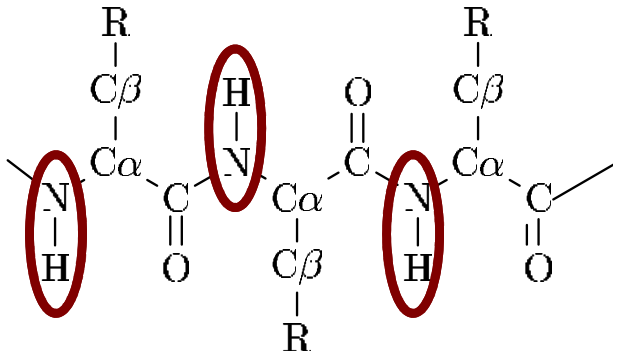
In a $^{13}C/^{15}N$ uniformly labeled protein, correlation spectra can be measured through single- and double-bond couplings

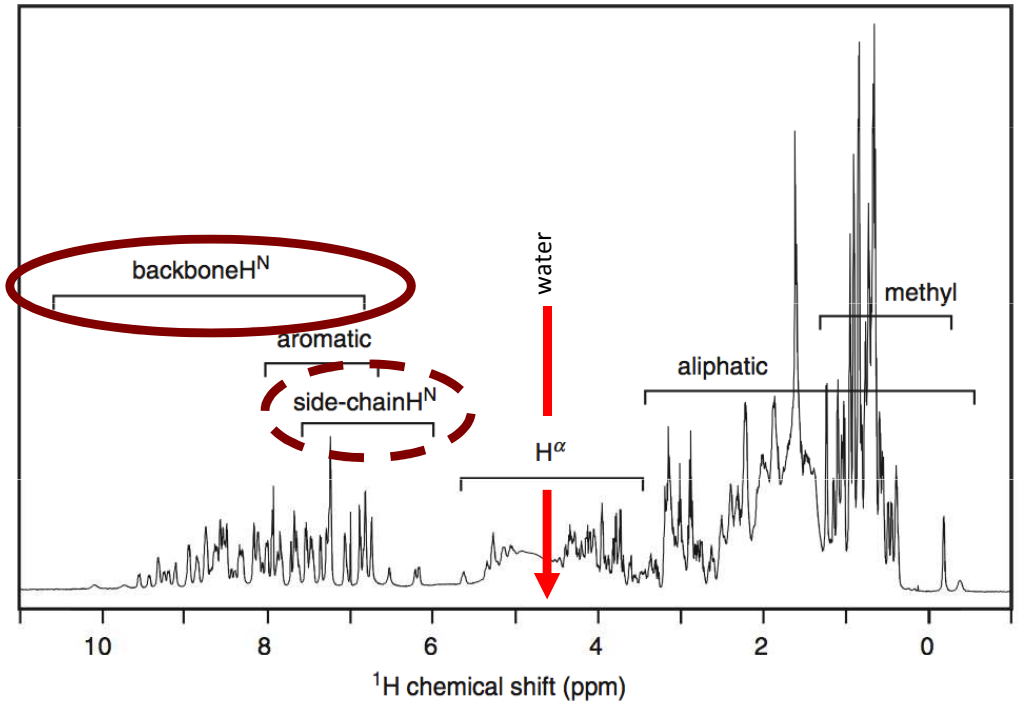


Sattler et al. Prog. NMR Spectrosc. (**1999**) 34, 93-158

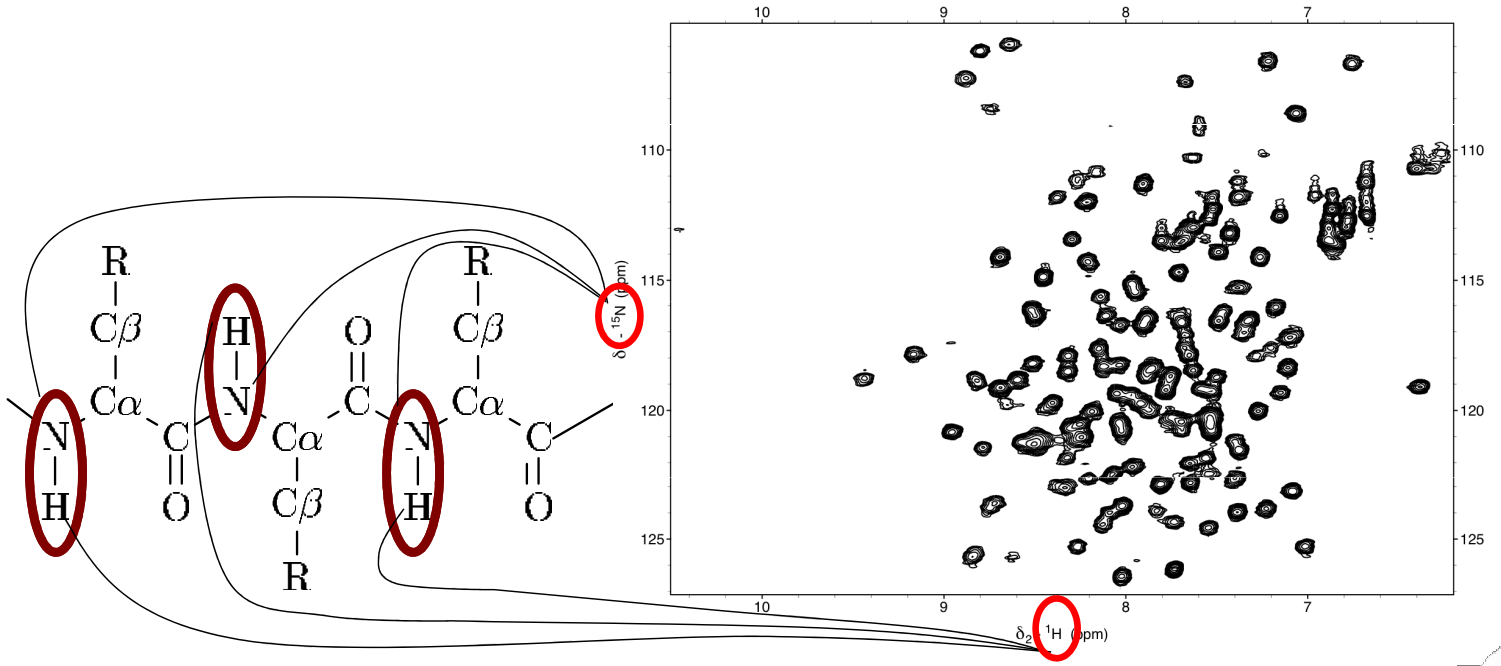


^{1}H 1D, Cavanagh et al., 2007





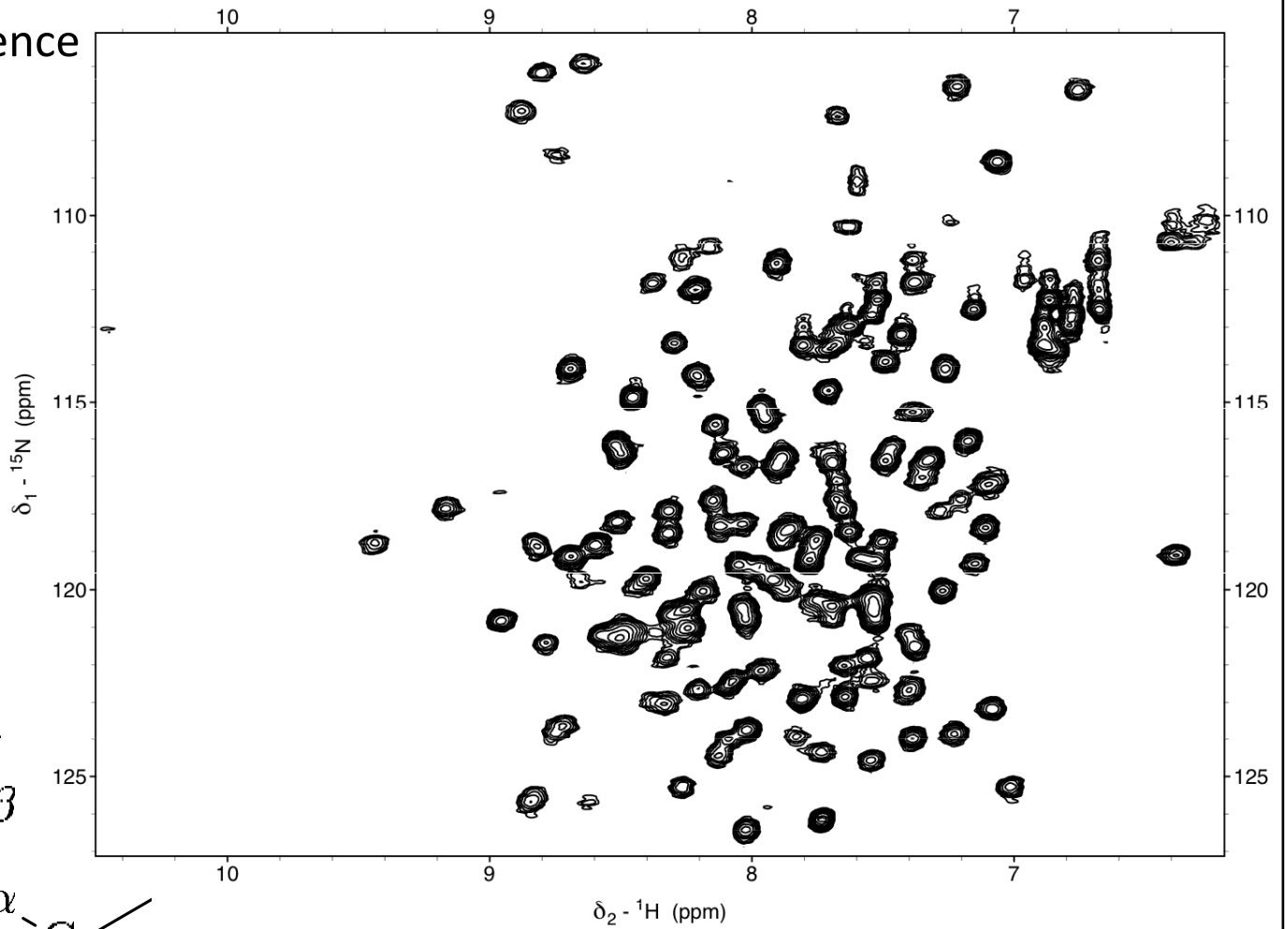
^{1}H 1D, Cavanagh et al., 2007

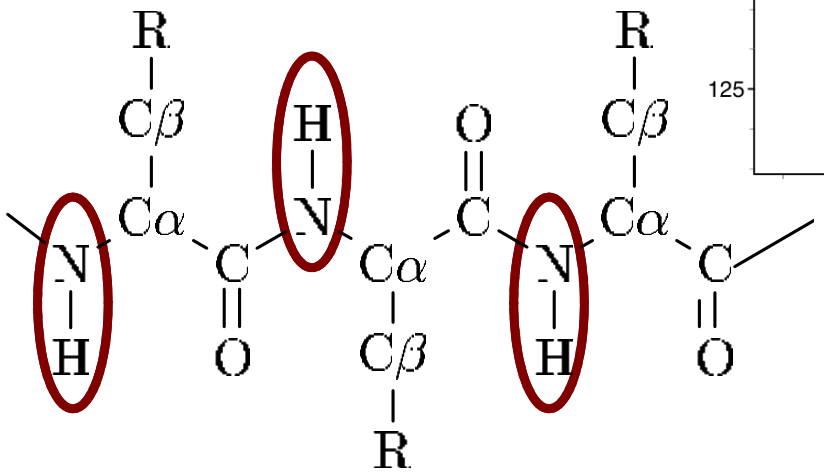


^{1}H-^{15}N HSQC, cca 155 aa, well folded, 600MHz, 29

^{15}N-^{1}H HSQC – **H**eteronuclear **S**ingle**Q**uantum **C**orrelation

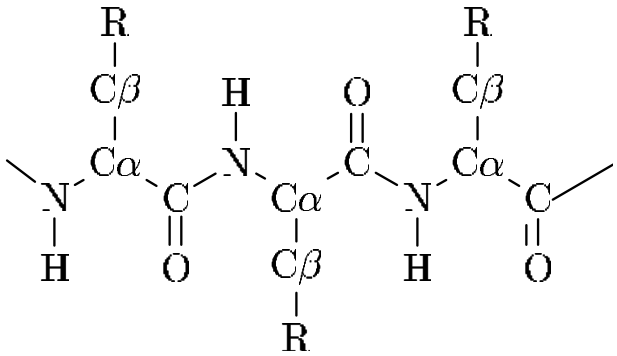
1) 1 peak ≅ 1 amino acid
2) Excellent info about the protein fold
3) No info about primary sequence

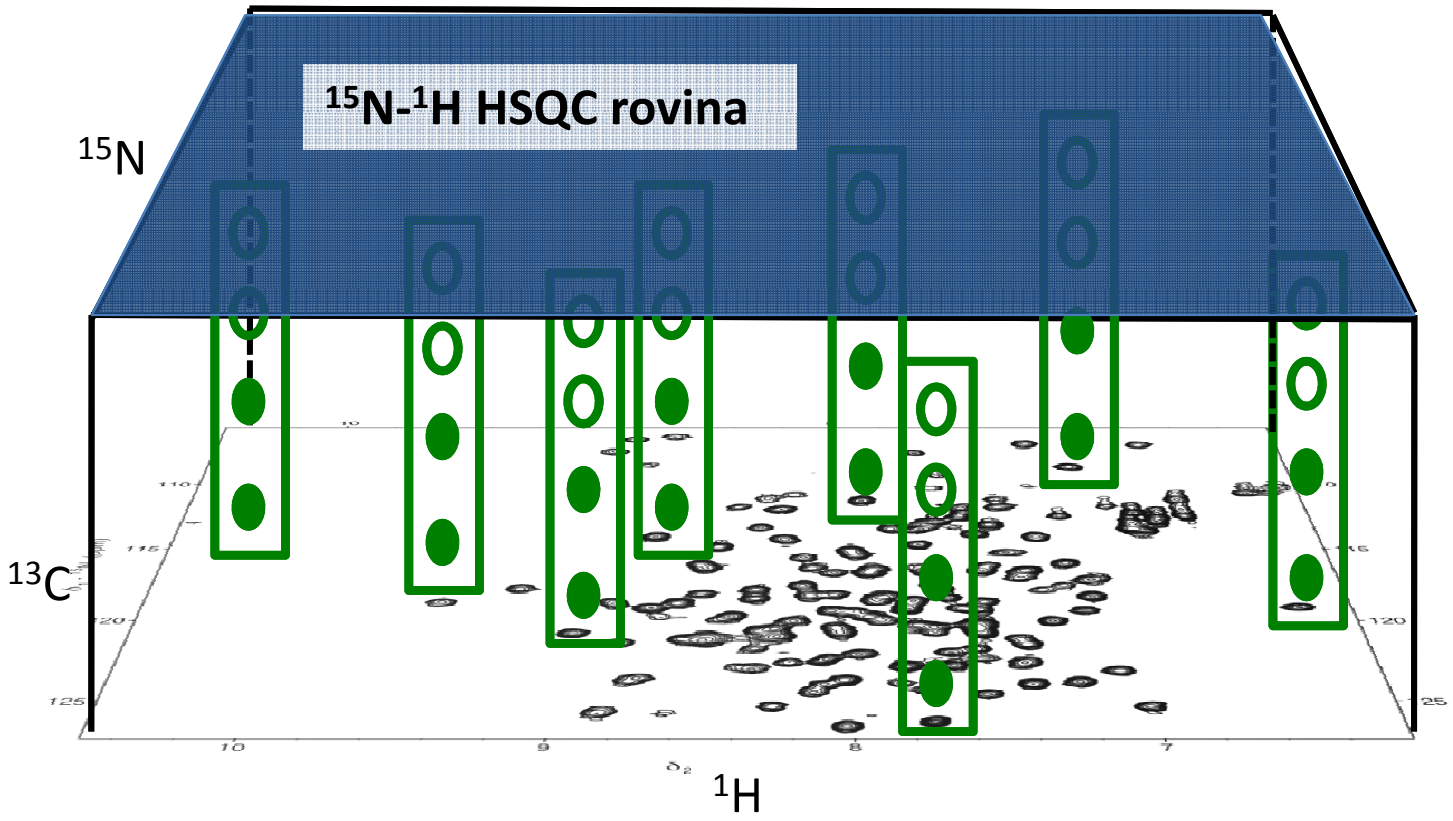




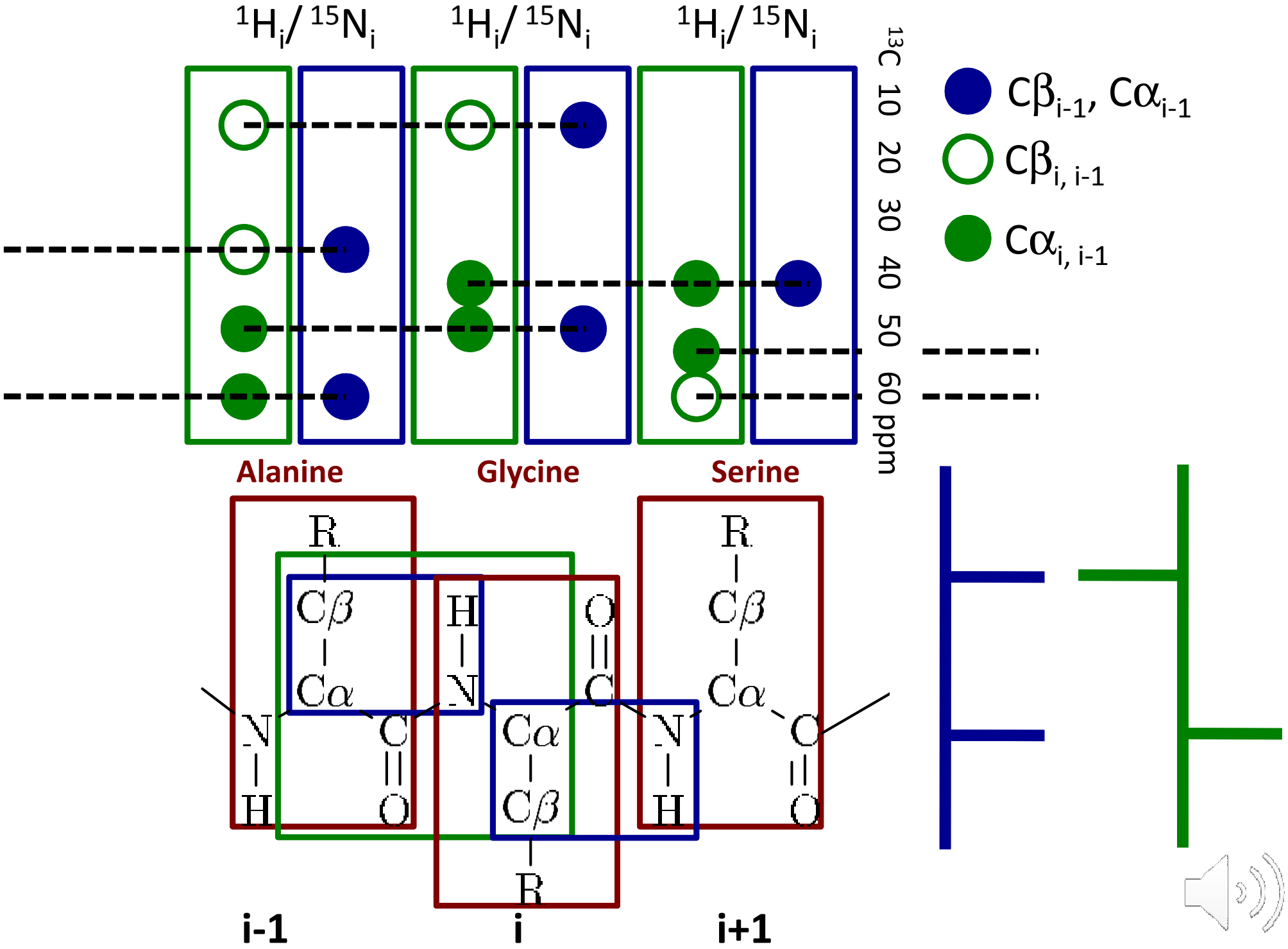
Traditional triple resonance NMR spectra for protein backbone assignment

1) HNCO – chem. shifts of C=O
2) HNCA – chem. shifts of $C_{\alpha,i}$, $C_{\alpha,i-1}$
3) HNCOCA – chem. shifts of $C_{\alpha,i-1}$
4) **HNCACB** – chem. shifts of $C_{\alpha,i}$, $C_{\alpha,i-1}$, $C_{\beta,i}$, $C_{\beta,i-1}$
5) **HNCOCACB** – chem. shifts of $C_{\alpha,i-1}$, $C_{\beta,i-1}$



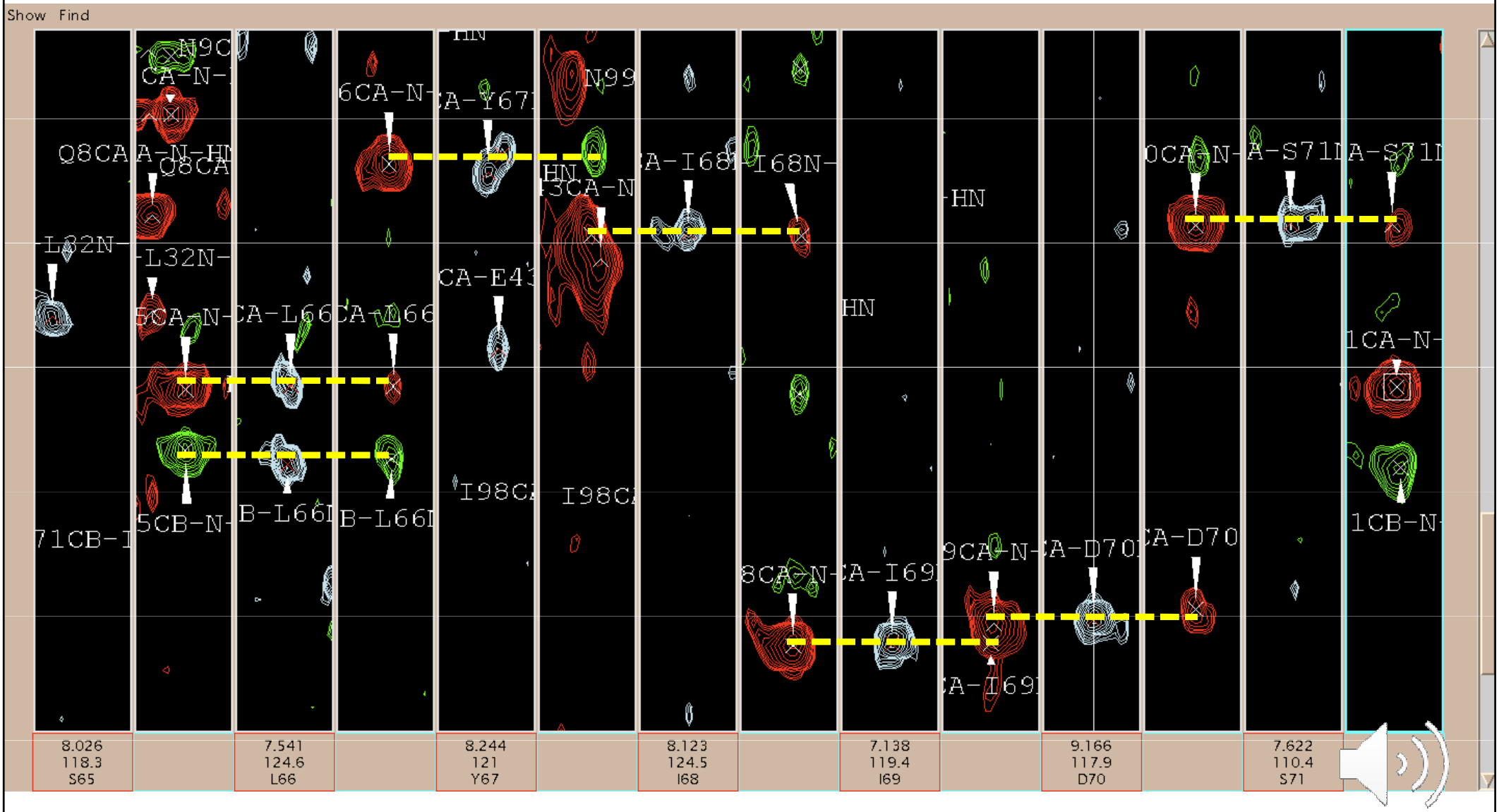


Two of the above spectra needed


$^1H_i/^{15}N_i$
$^1H_i/^{15}N_i$
$^1H_i/^{15}N_i$
^{13}C 10 20 30 40 50 60 ppm
$C\beta_{i-1}$, $C\alpha_{i-1}$
$C\beta_{i,\ i-1}$
$C\alpha_{i,\ i-1}$
Alanine
Glycine
Serine
R
$C\beta$
$C\alpha$
N
H
C
O
i-1
i
i+1

Backbone chemical shifts assignment example:

- fragment S65-S71 (50-66ppm v ^{13}C**)**



Automatic chemical shifts assignment

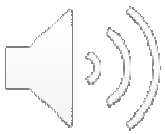
1) With a known 3D protein structure (e.g. X-ray, Modeller)
2) Without prior knowledge of a structure
3) **e.g. A**utomated **P**rojection **S**pectroscop**Y** APSY (S. Hiller, F. Fiorito, K. Wüthrich & G. Wider, *PNAS* (**2005**) 102, 10876-81)

Advantages

1) Time

Disadavantages

1) With increasing size of the protein (increasing spectral overlap) the reliability is decreasing



How to prepare and what to expect from NMR spectroscopy

A] Scientific question / I'd like to obtain:

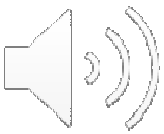
1) Structure
2) Relaxation properties
3) Interaction at atomic level resolution
4) Analysis (NMR is also analytical method)
5) Image (MRI)

B] What do I need to provide / prepare?

C] How much would it cost?

D] How long would it take?

E] Can I do it on my own (or a sudent/colleague of mine)



Thank you for your attention

For Application to Protein Characterization

by
Radovan Fiala, Karel Kubicek, and Pavel Kaderavek
CEITEC, Masaryk University

This project received funding by the European Commission under the GA n° 2019-1-IT02-KA203-063391