CBP
CENTRE BLAISE PASCAL

Les journees du Centre Blaise Pascal
L'analyse et la modelisation de données issues du monde vivant

NMR experiment-driven modeling of biological macromolecules

Torsten HERRMANN

Centre de RMN à très Hauts Champs
Institut des Sciences Analytiques
Université de Lyon/ UMR 5280 CNRS / ENS Lyon / UCB Lyon 1

Outline

- General introduction & overview
  - Experimental techniques for Structural Biology
  - Strength & weakness, complementarity
  - Experimental data content & modeling
- State-of-the-art liquid and solid-state NMR structure determination
  - Multi-purpose UNIO platform
  - Recent progress in liquid-state NMR
  - Recent progress in solid-state NMR

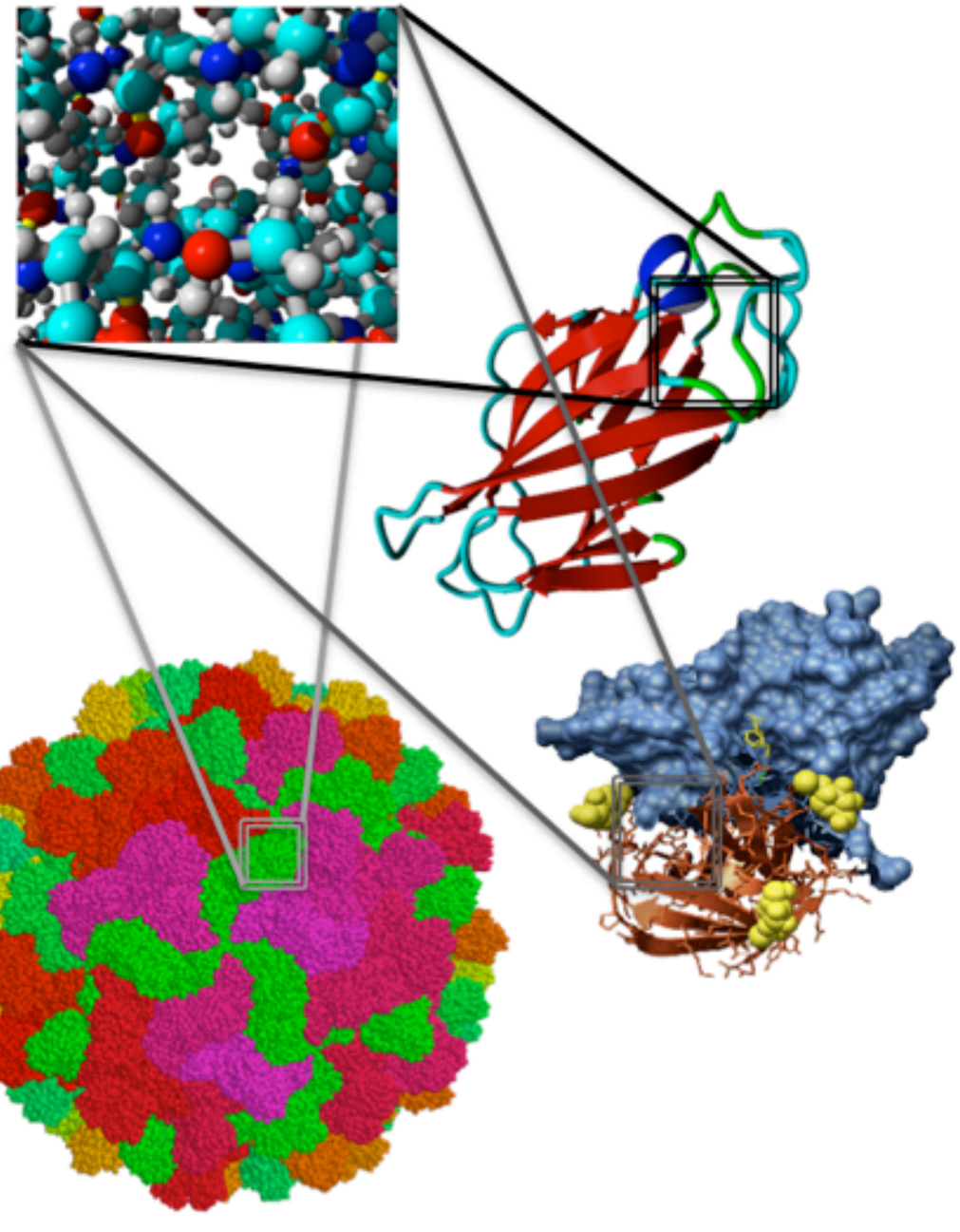
Experimental techniques & atomic resolution

DTSGTVCL
SALPPEAT
DTLNLIAS
D Q
L
GEATQEDY
YTGDHYAT
FSLIDQTC

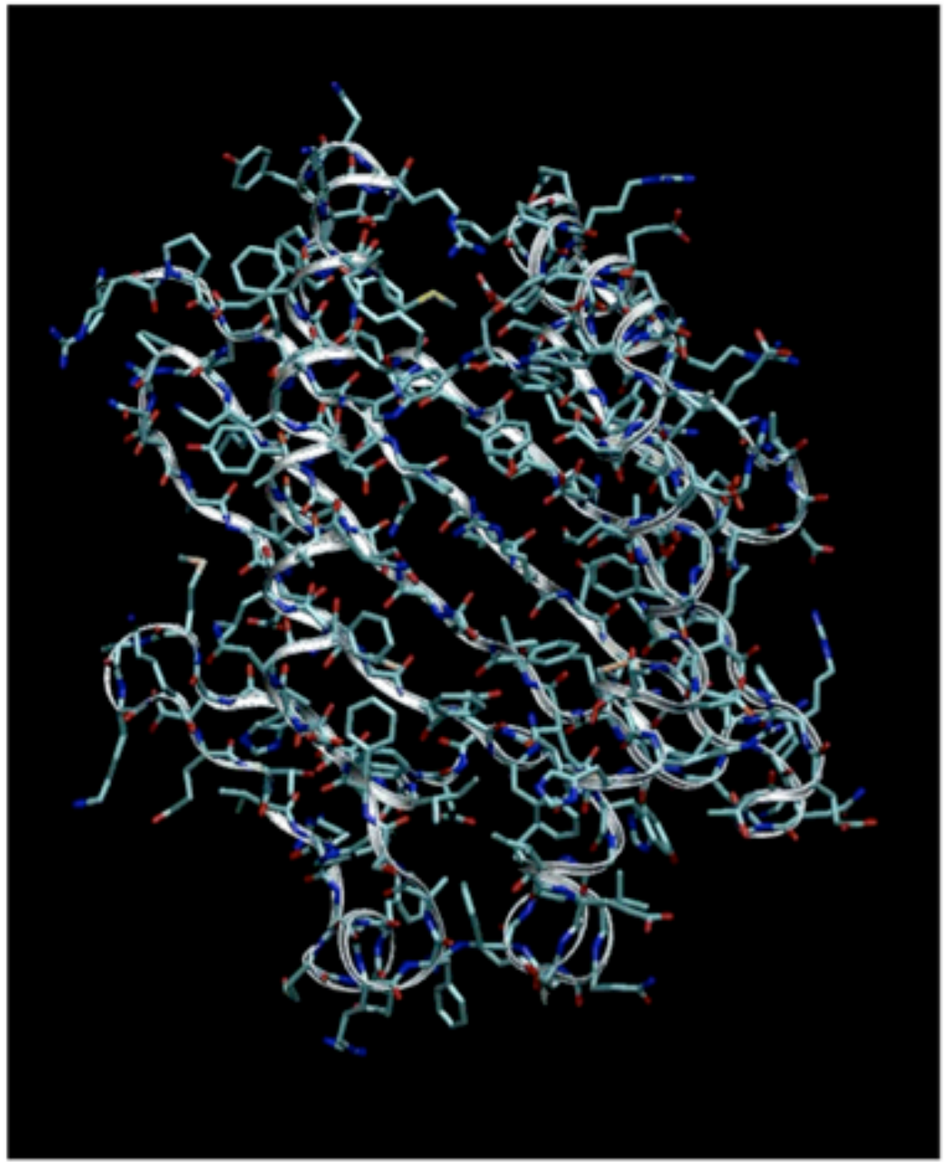
Higher resolution

X-ray

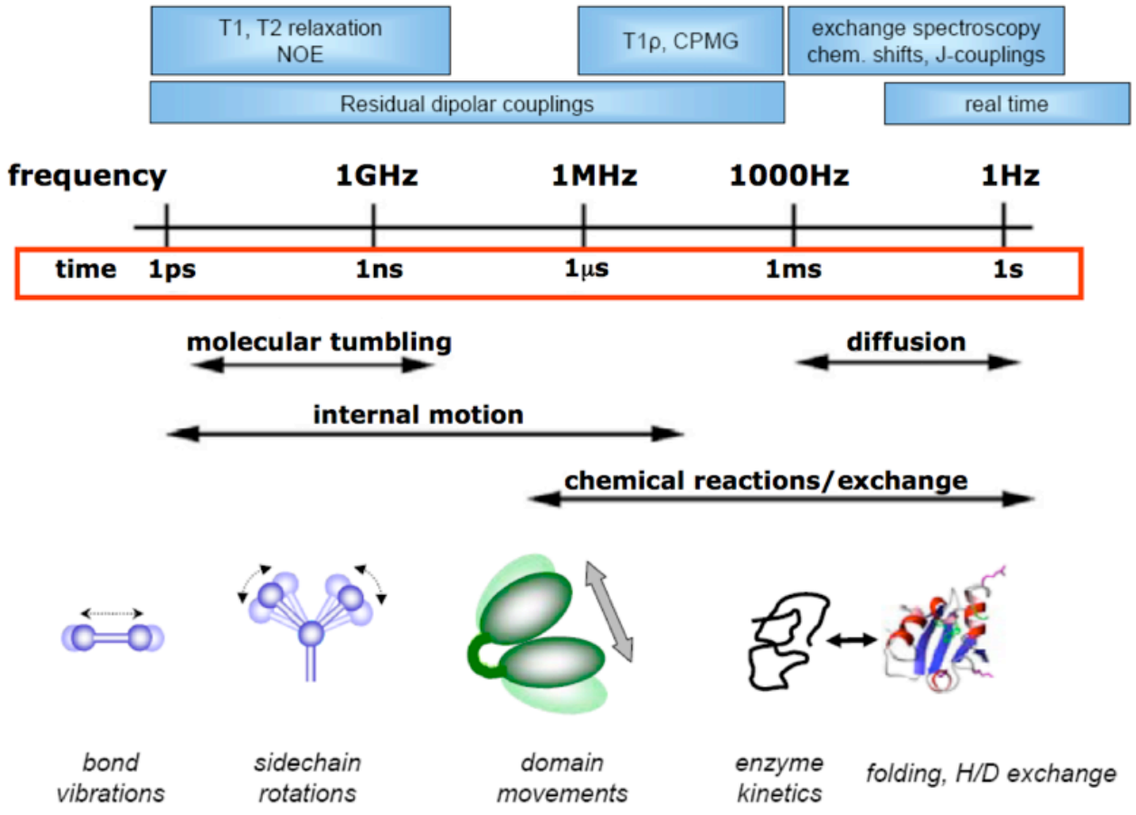

```
ATOM      1  N   ASP A   1      -10.341   -9.922    9.398
TRANSFORMATION
ATOM      3  C   ASP A   1      -10.288   -9.894    6.915
```

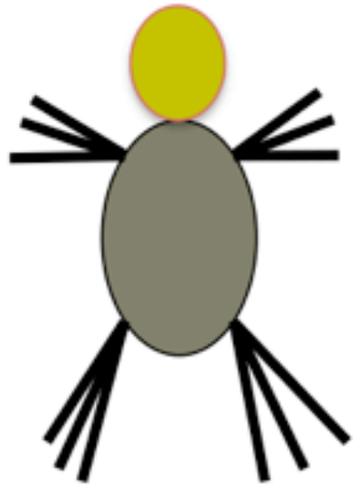


*but there is **dynamics**...*

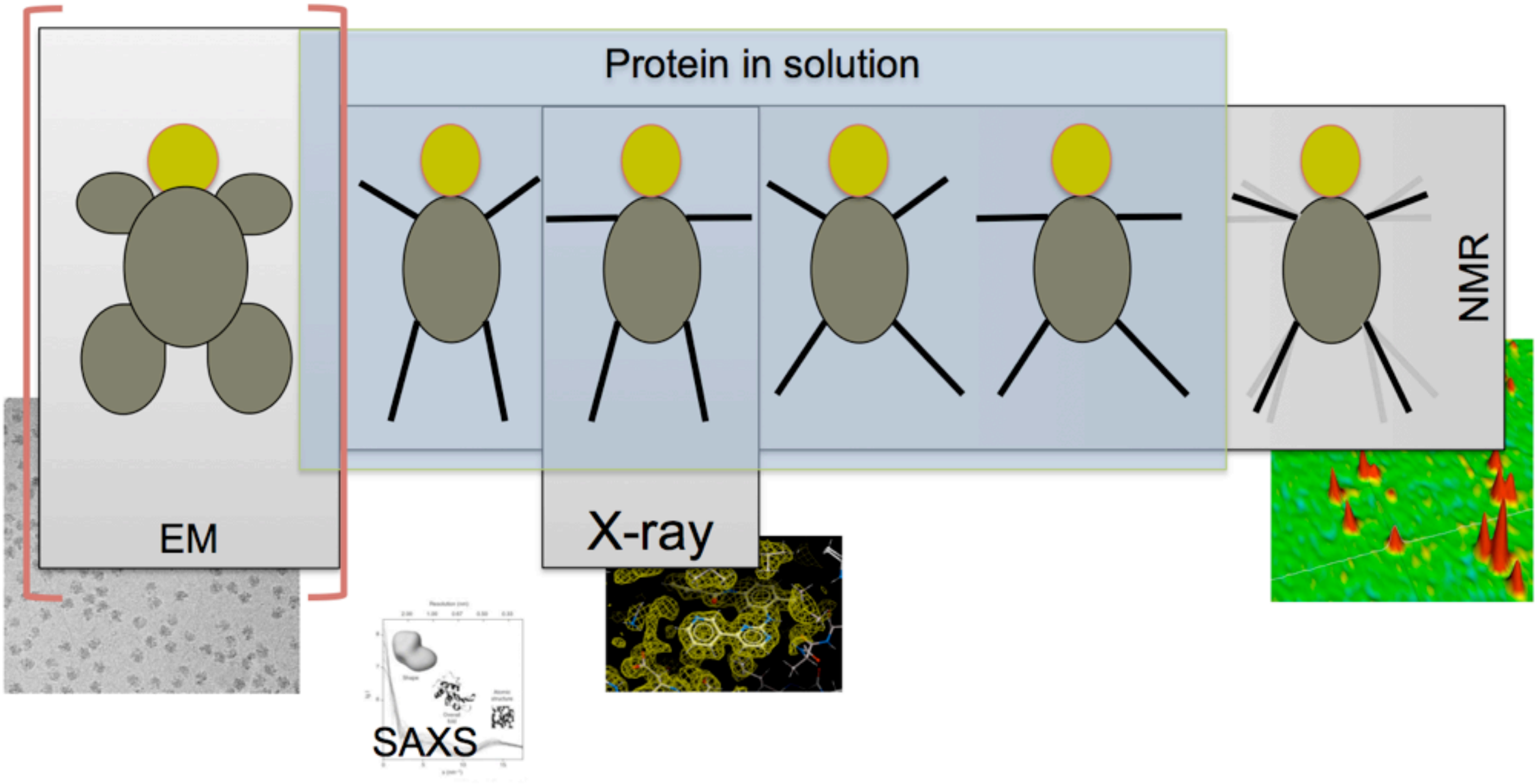


Protein dynamics probed by NMR

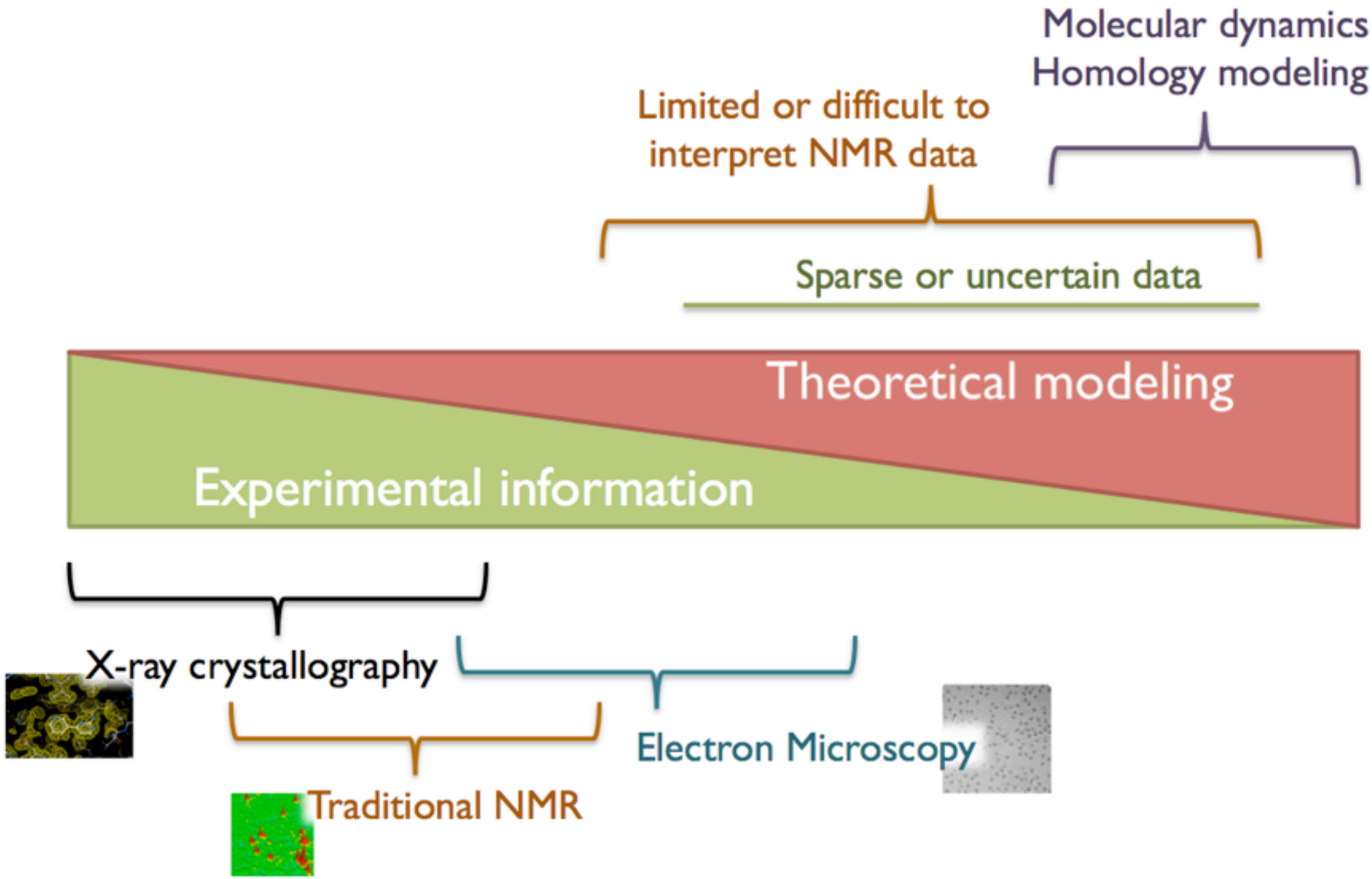




Experimental techniques & dynamics



What is a macromolecular structure?



History of NMR spectroscopy

Nobel prizes

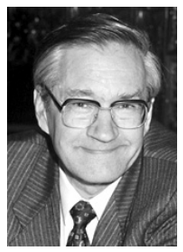
1944 *Physics* Rabi (Columbia)

"for his resonance method for recording the magnetic properties of atomic nuclei"

1952 *Physics* Bloch (Stanford), Purcell (Harvard)

"for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith"

1991 *Chemistry* Ernst (ETH)



"for his contributions to the development of the methodology of high resolution nuclear magnetic resonance (NMR) spectroscopy"

2002 *Chemistry* Wüthrich (ETH)

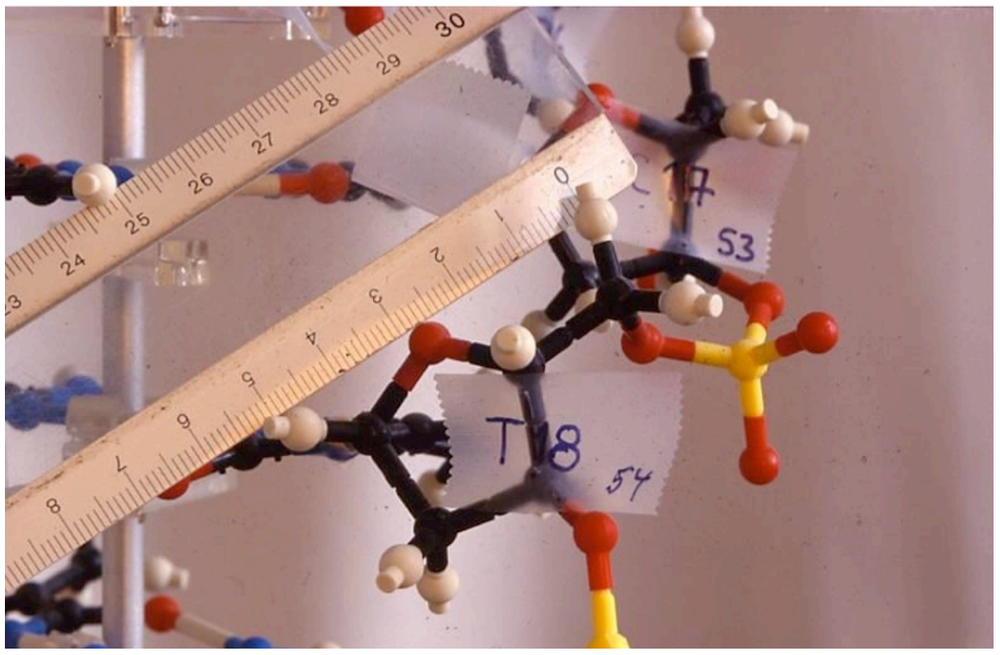
"for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution"

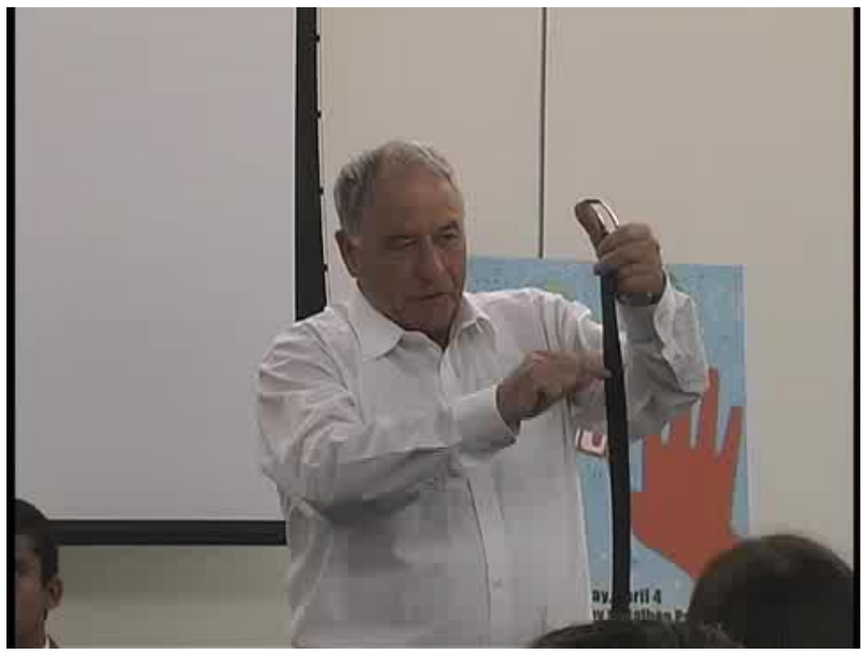
2003 *Medicine* Lauterbur (University of Illinois in Urbana), Mansfield (University of Nottingham)

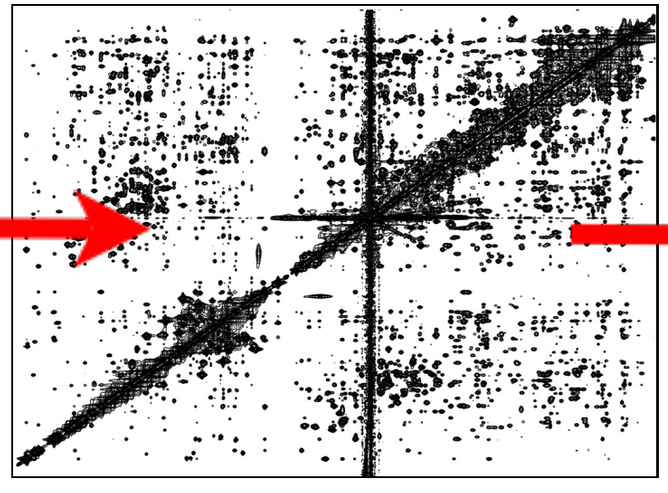
"for their discoveries concerning magnetic resonance imaging"

The basic idea of protein NMR structure determination

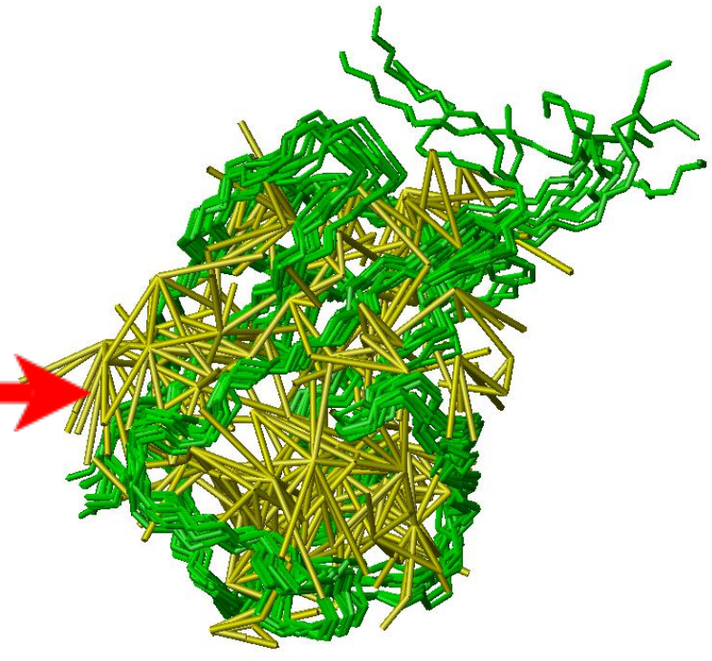
It's all about distance measurements between atoms..



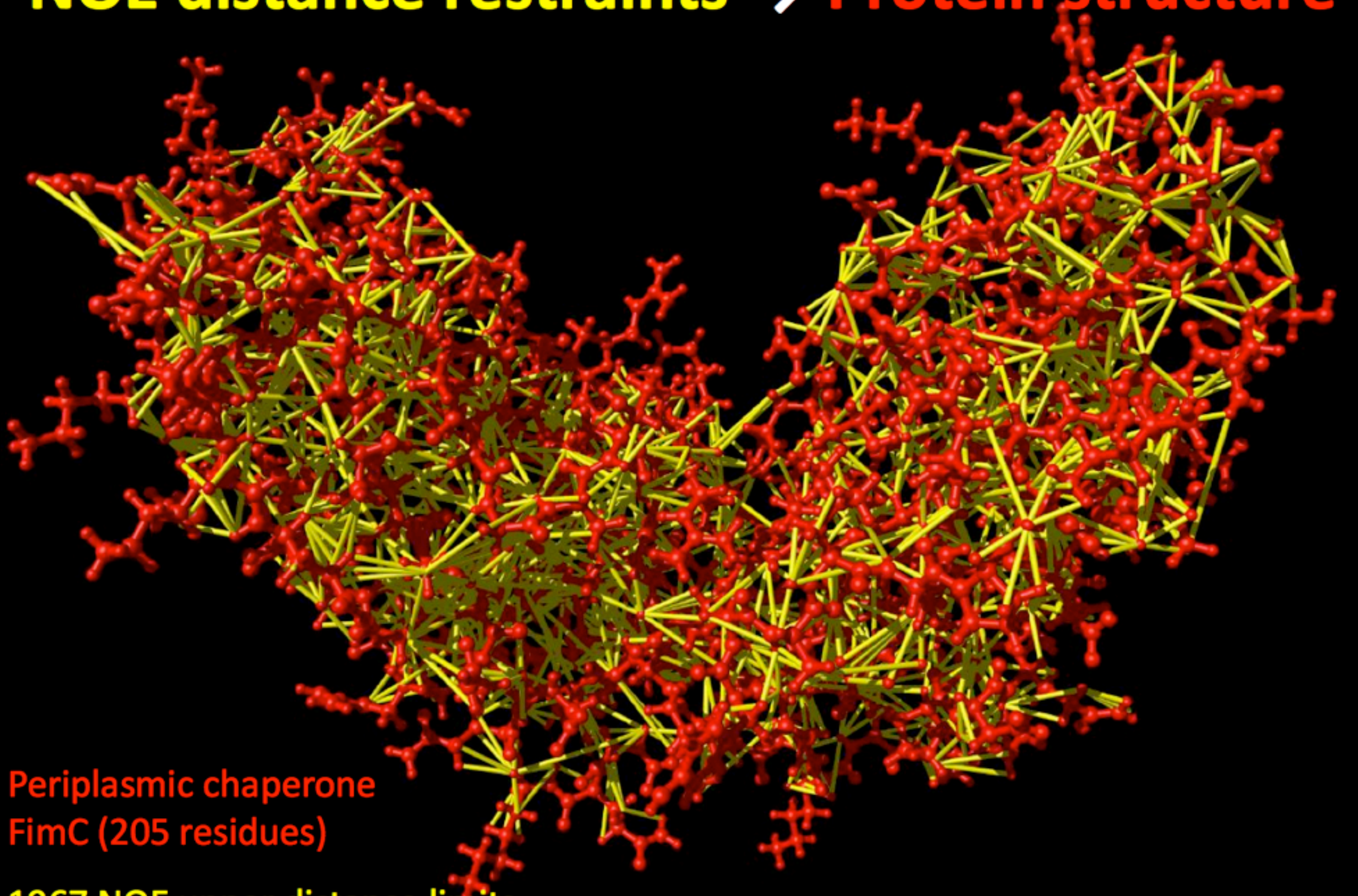




NOESY



NOE distance restraints → Protein structure

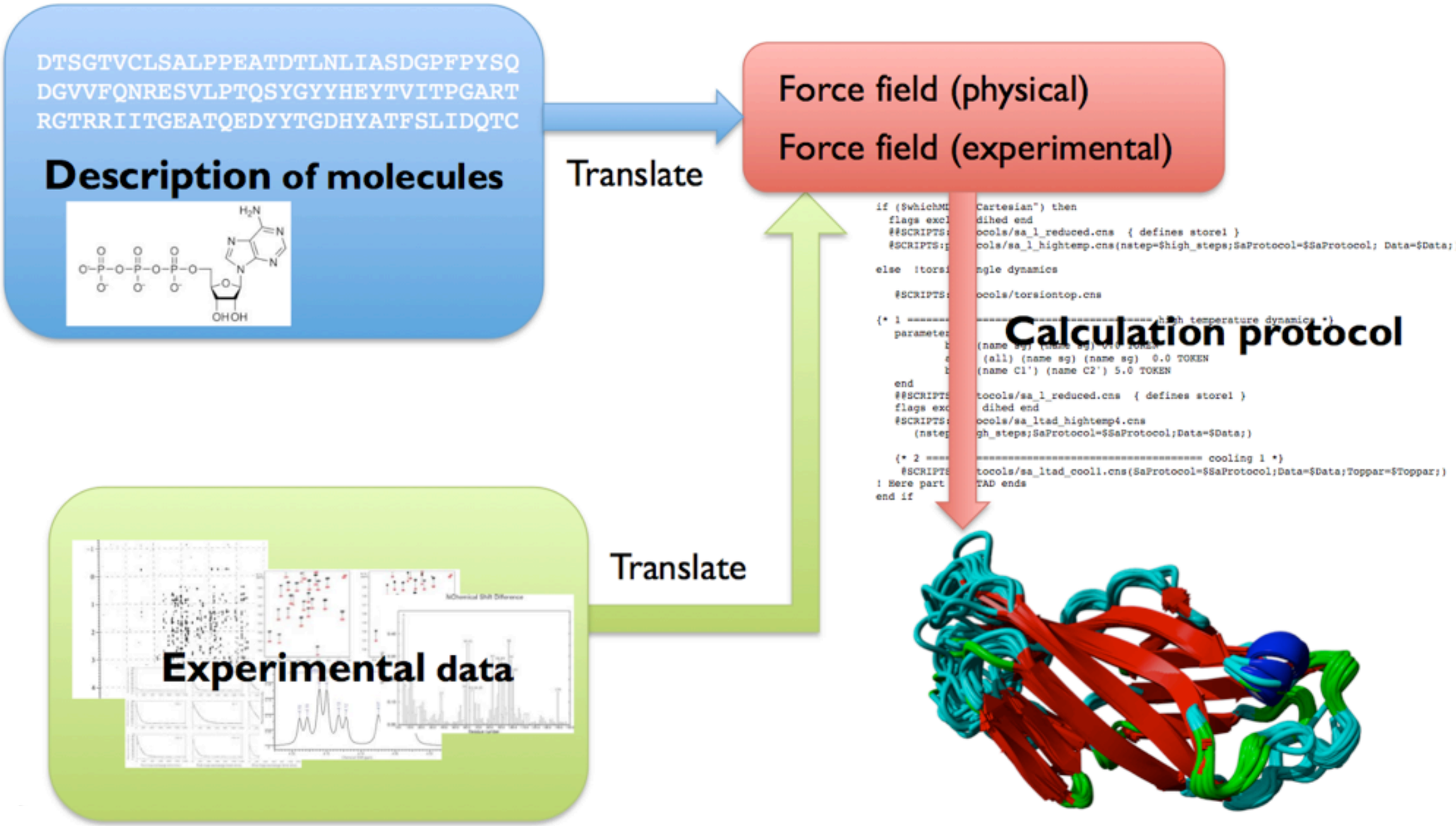


Periplasmic chaperone
FimC (205 residues)

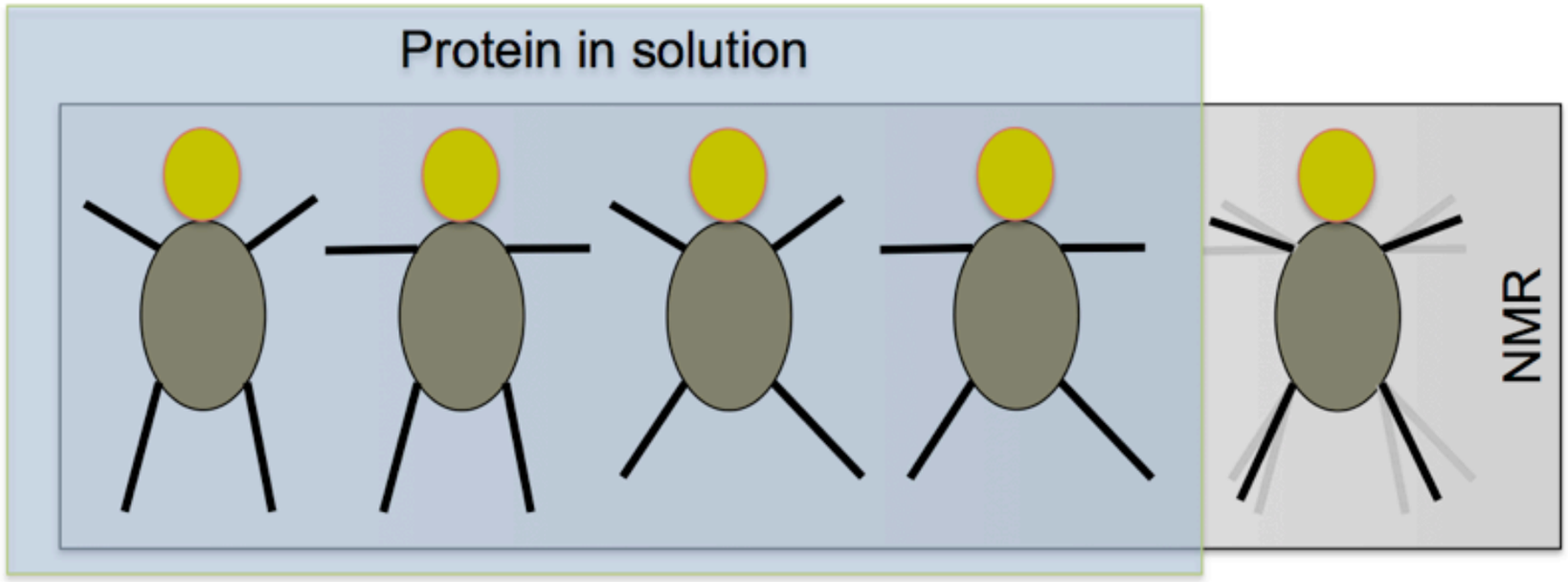
1967 NOE upper distance limits

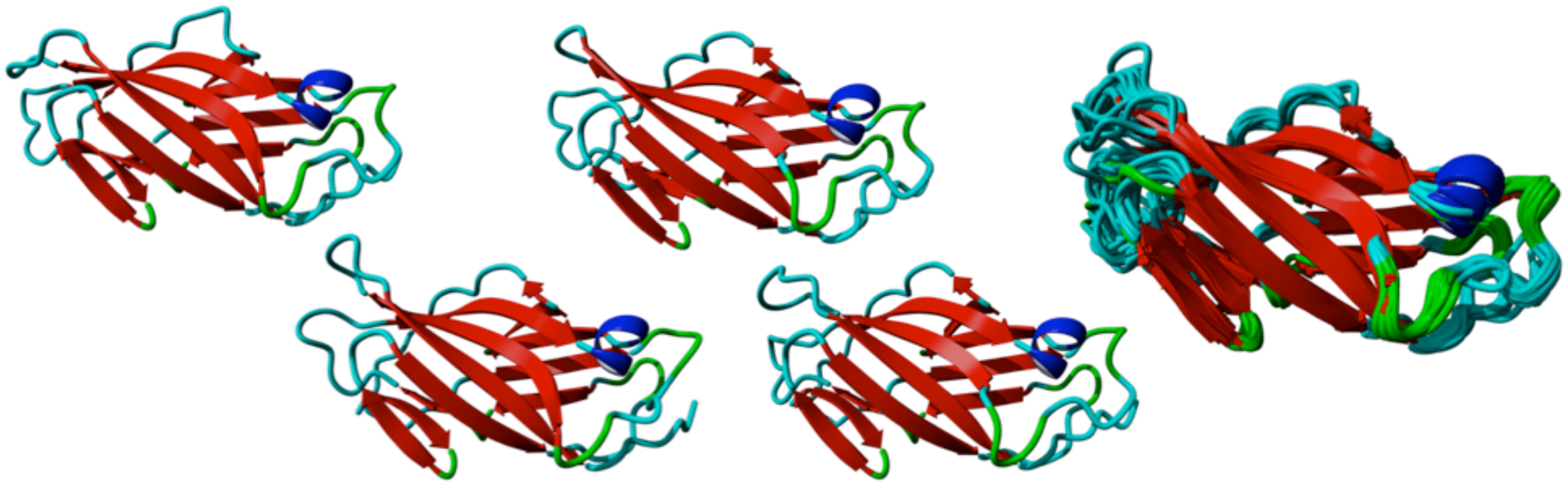
M. Pellecchia et al. *Nature Struct. Biol.* 5, 885-890 (1998)

NMR experiment-driven modeling of macromolecules

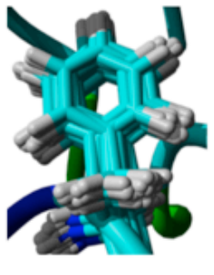


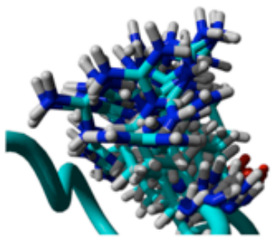
Averaged NMR data & dynamics





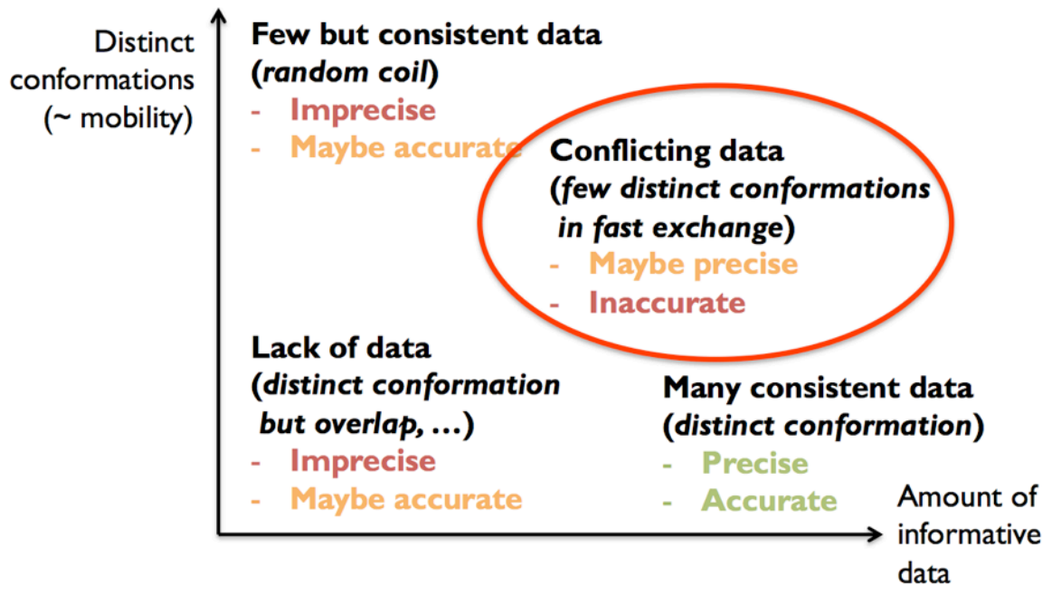
Accuracy vs. precision



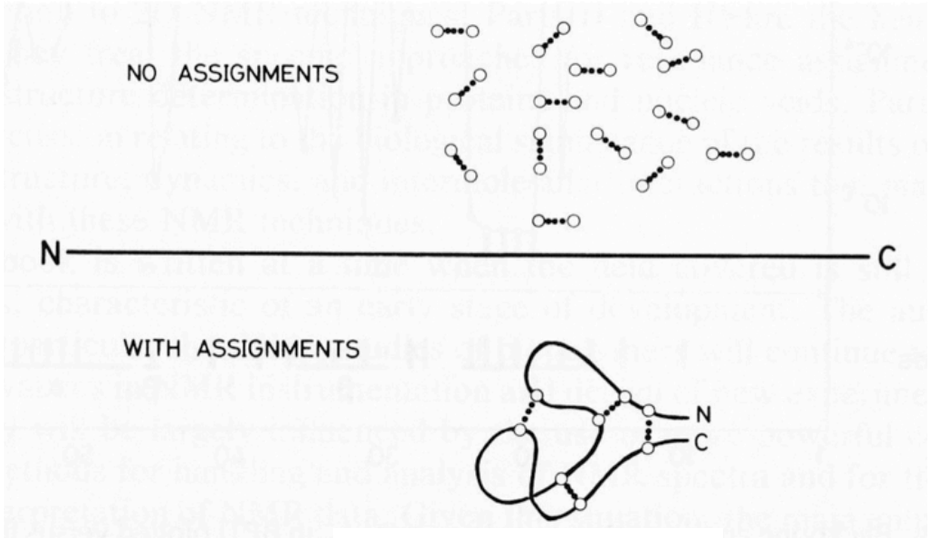


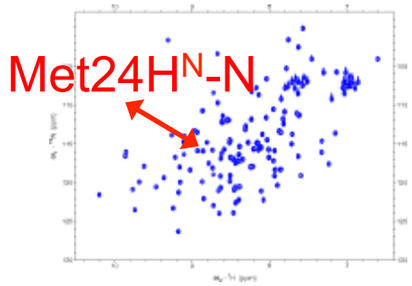
	Accurate	Inaccurate (systematic error)
Precise **Low RMSD**		
Imprecise (reproducibility error) **High RMSD**		

- Amount of data available vs. conformations

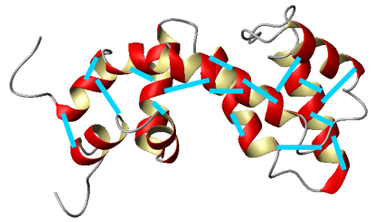


Protocol for protein NMR structure determination

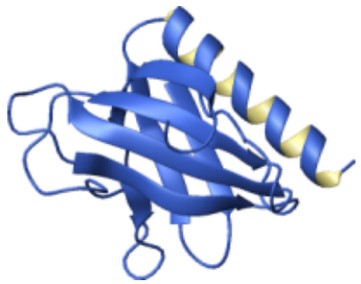




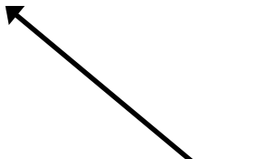
Assignment



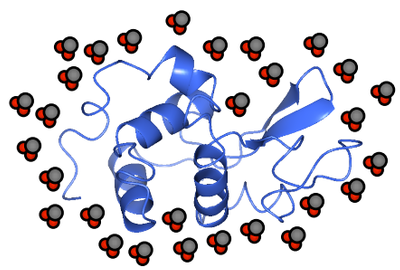
Restraint Generation



Structure Calculation



Validation



Water Refinement

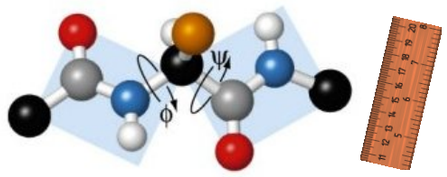
WORLDWIDE PDB PROTEIN DATA BANK

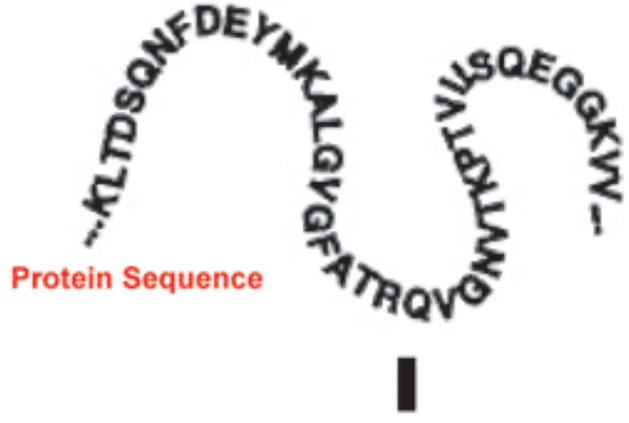
Deposition

Outline

- General introduction & overview
  - Experimental techniques for Structural Biology
  - Strength & weakness, complementarity
  - Experimental data content & modeling
- State-of-the-art liquid and solid-state NMR structure determination
  - Multi-purpose UNIO platform
  - Recent progress in liquid-state NMR
  - Recent progress in solid-state NMR

Research overview: unsupervised molecular modeling





Computational methods

Protein Structure Model

What is the challenge?

NMR resonance assignment is like solving a puzzle…

…with missing pieces
(incomplete signals)

…with additional pieces
(artifacts)

…in the mist
(low signal-to-noise, line-broadening)

What is the challenge?

NMR - Giant Jigsaw Puzzle

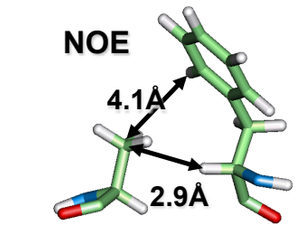
NMR Structure Determination

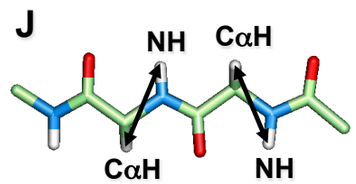
NOE
- a through space correlation (<5Å)
- distance constraint

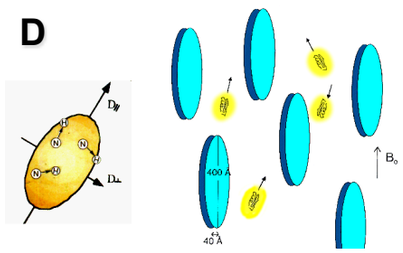
Coupling Constant (J)
- through bond correlation
- dihedral angle constraint

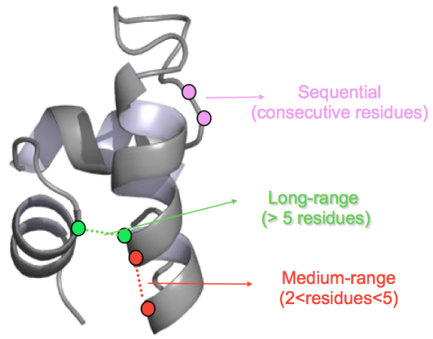
Chemical Shift
- very sensitive to local changes in environment
- dihedral angle constraint

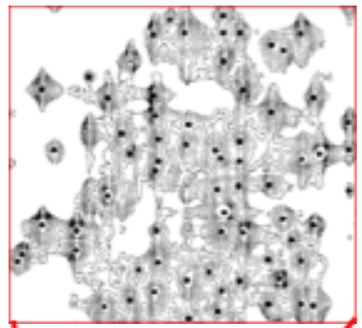
Dipolar coupling constants (D)
- bond vector orientation relative to magnetic field
- alignment with bicelles or viruses

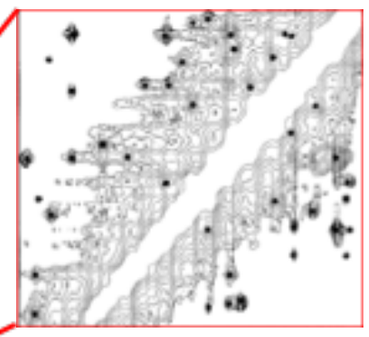


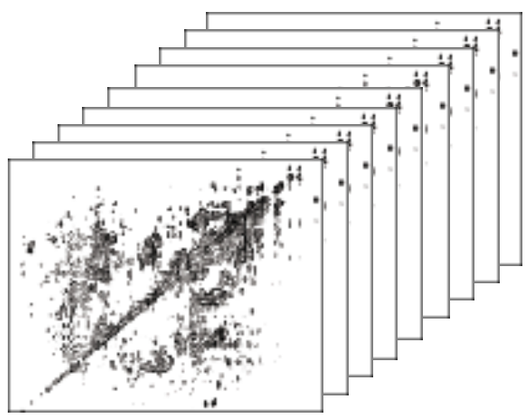


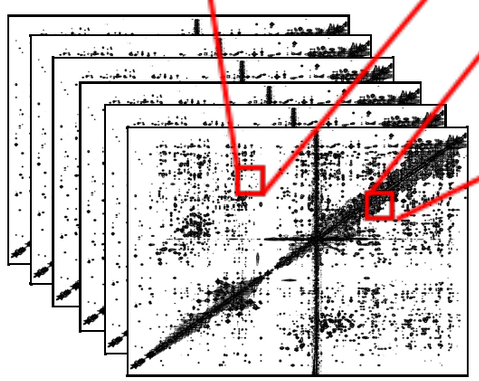








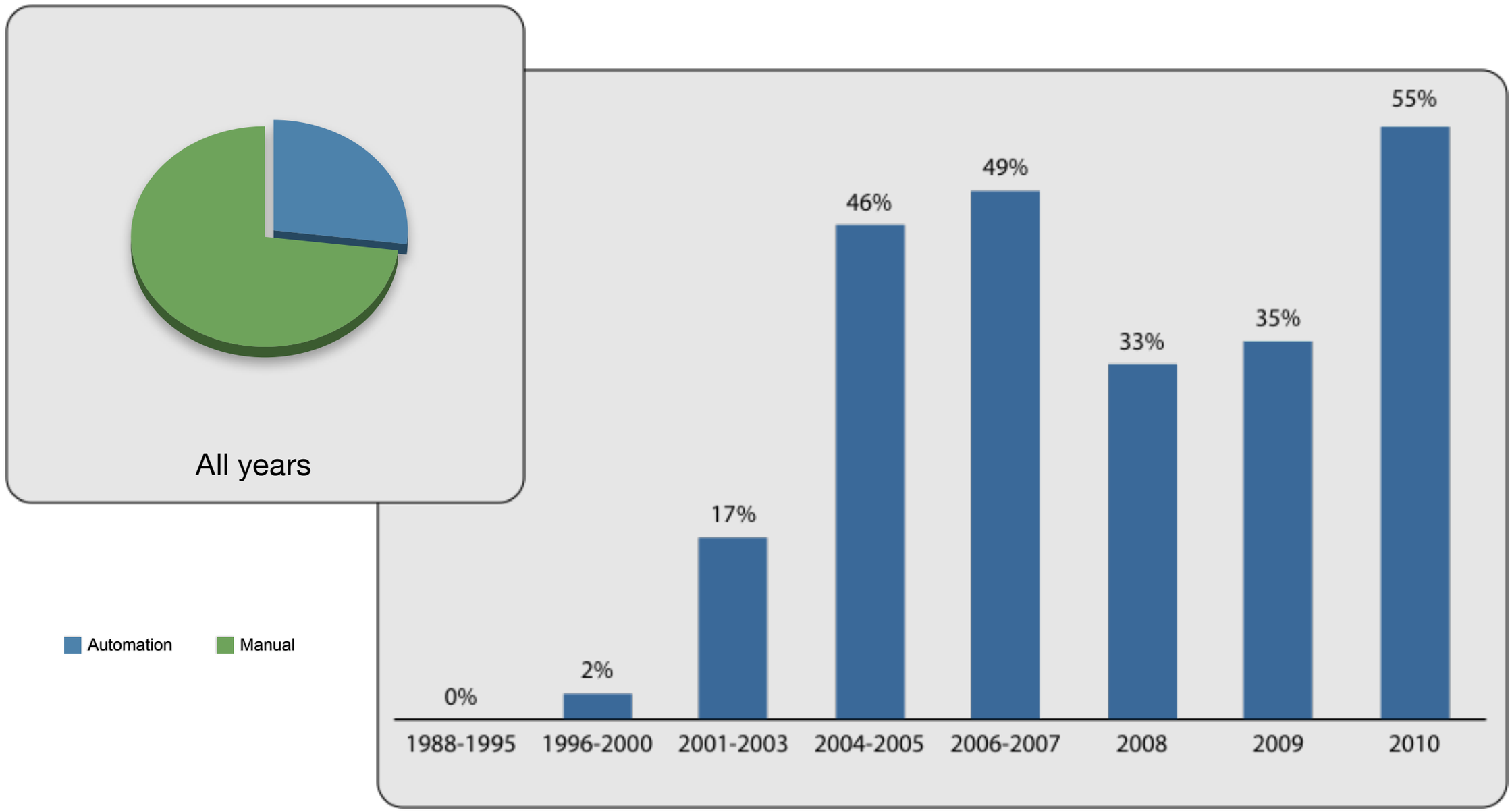




Often 15-20 NMR spectra need to be consistently and jointly analyzed.

What is the challenge?

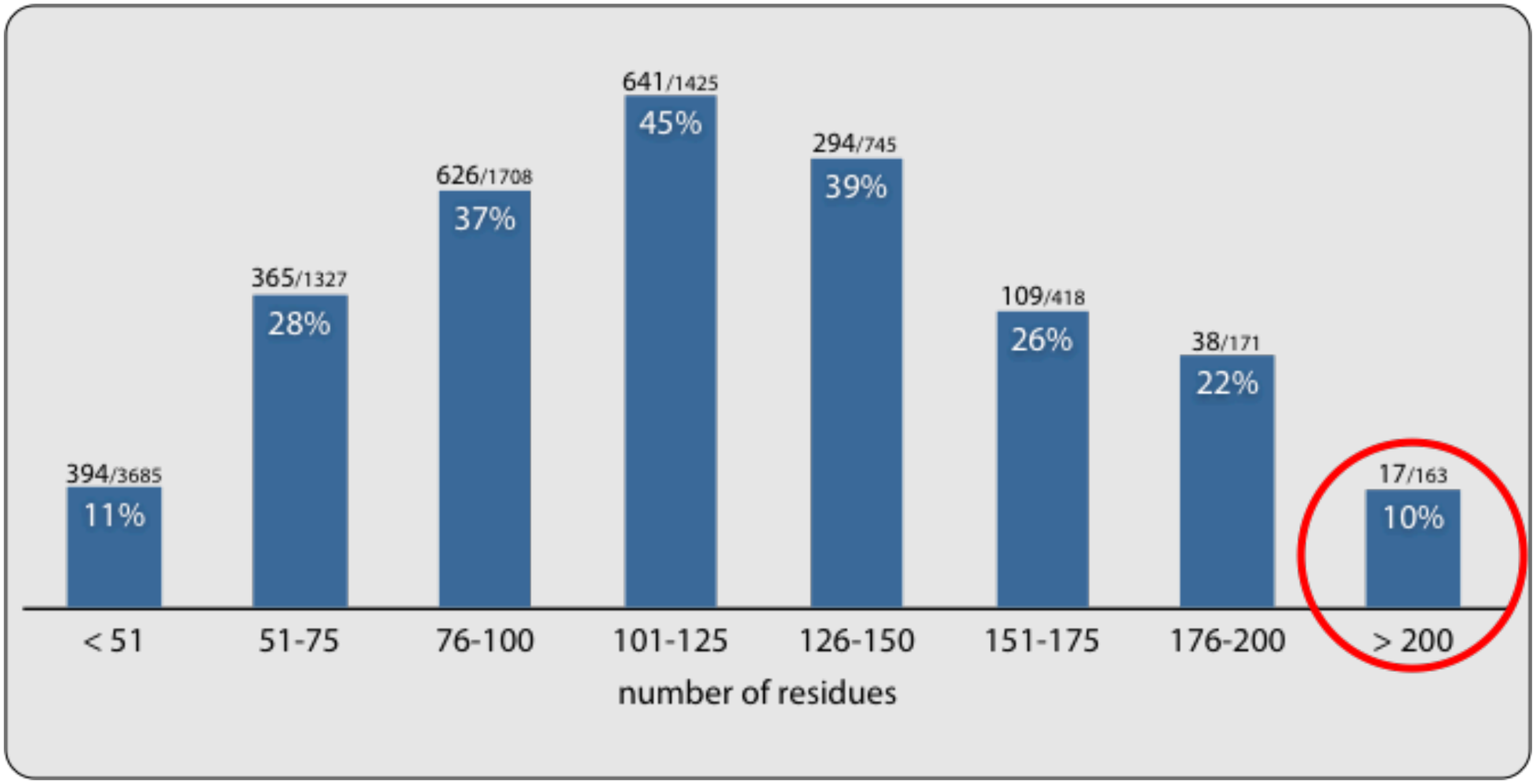
Percentage of "automated" PDB-deposited NMR structures



Guerry & Herrmann, *Quart. Rev. Biophys*. **2011**, Aug., 44(3): 257-309.

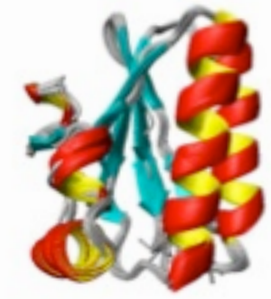
What is the challenge?

Size of "automated" PDB-deposited NMR structures



! Need of improvement for 'traditional' algorithms !

Outline - Part 1

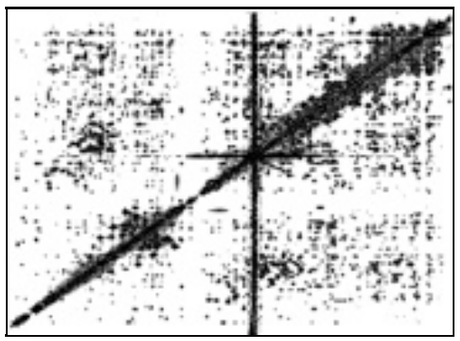


UNIO

Aim of UNIO: Overcome current weaknesses …

- **Automated approaches operate on intermediate textual peak lists.**

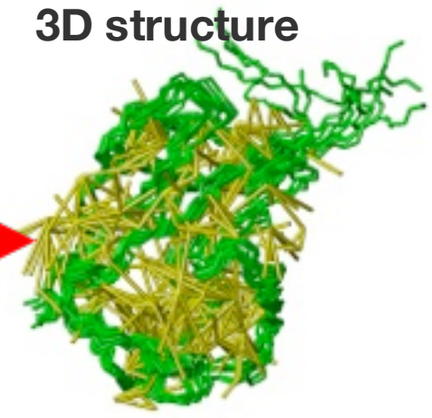
Spectrum



Peak lists

1	2.288	8.907	1	U	2.743E+05	3.047E+04	e
2	4.556	8.884	1	U	1.262E+06	3.324E+04	e
3	4.122	8.885	1	U	6.448E+05	2.616E+04	e
4	4.454	8.452	1	U	2.651E+05	2.639E+04	e
5	4.134	8.444	1	U	7.291E+05	2.528E+04	e
6	2.191	8.447	1	U	2.433E+05	1.839E+04	e
7	4.539	8.440	1	U	7.925E+05	3.504E+04	e
8	2.598	8.443	1	U	3.101E+05	2.311E+04	e
9	0.637	8.442	1	U	4.169E+05	4.216E+04	e
10	7.426	8.435	1	U	2.256E+05	1.719E+04	e
11	1.756	8.431	1	U	1.979E+06	2.529E+04	e
12	7.356	8.338	1	U	2.031E+05	1.719E+04	e

3D structure



Reduced efficiency and lack on reproducibility.

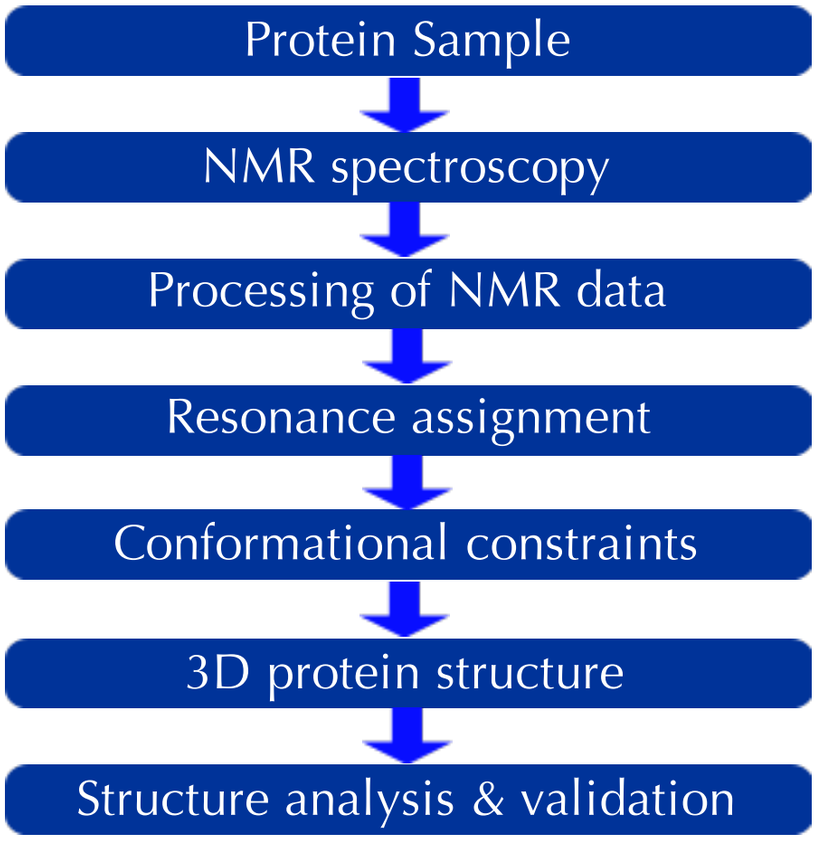
- **Automated approaches target individual stages.**

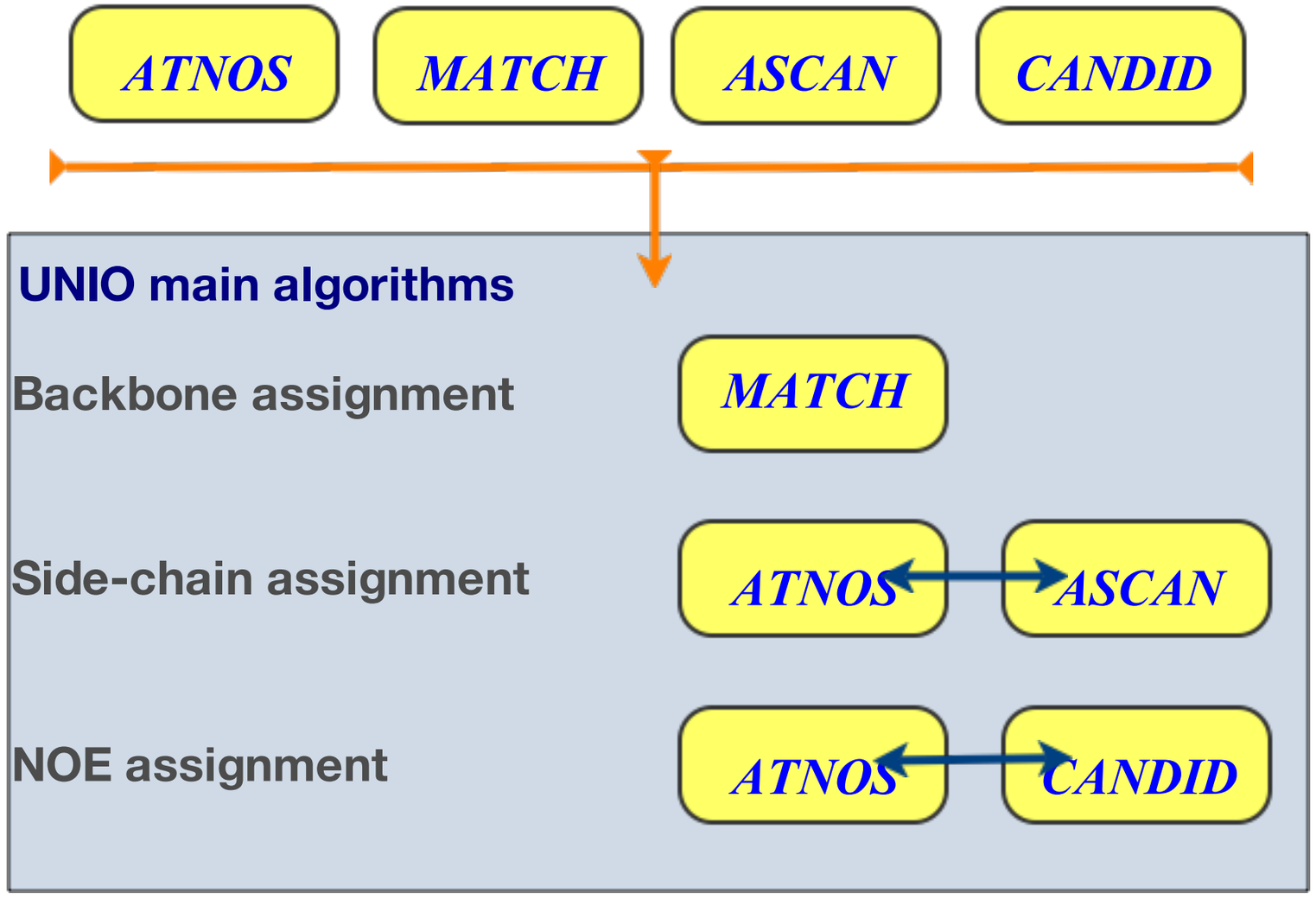
NMR is a multi-stage data analysis process

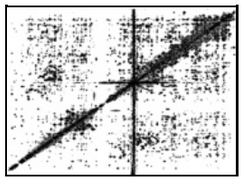
Brute force automation is likely to fail.

Smooth integration is key for laboratory efficiency.

UNIO: a multi-purpose application suite



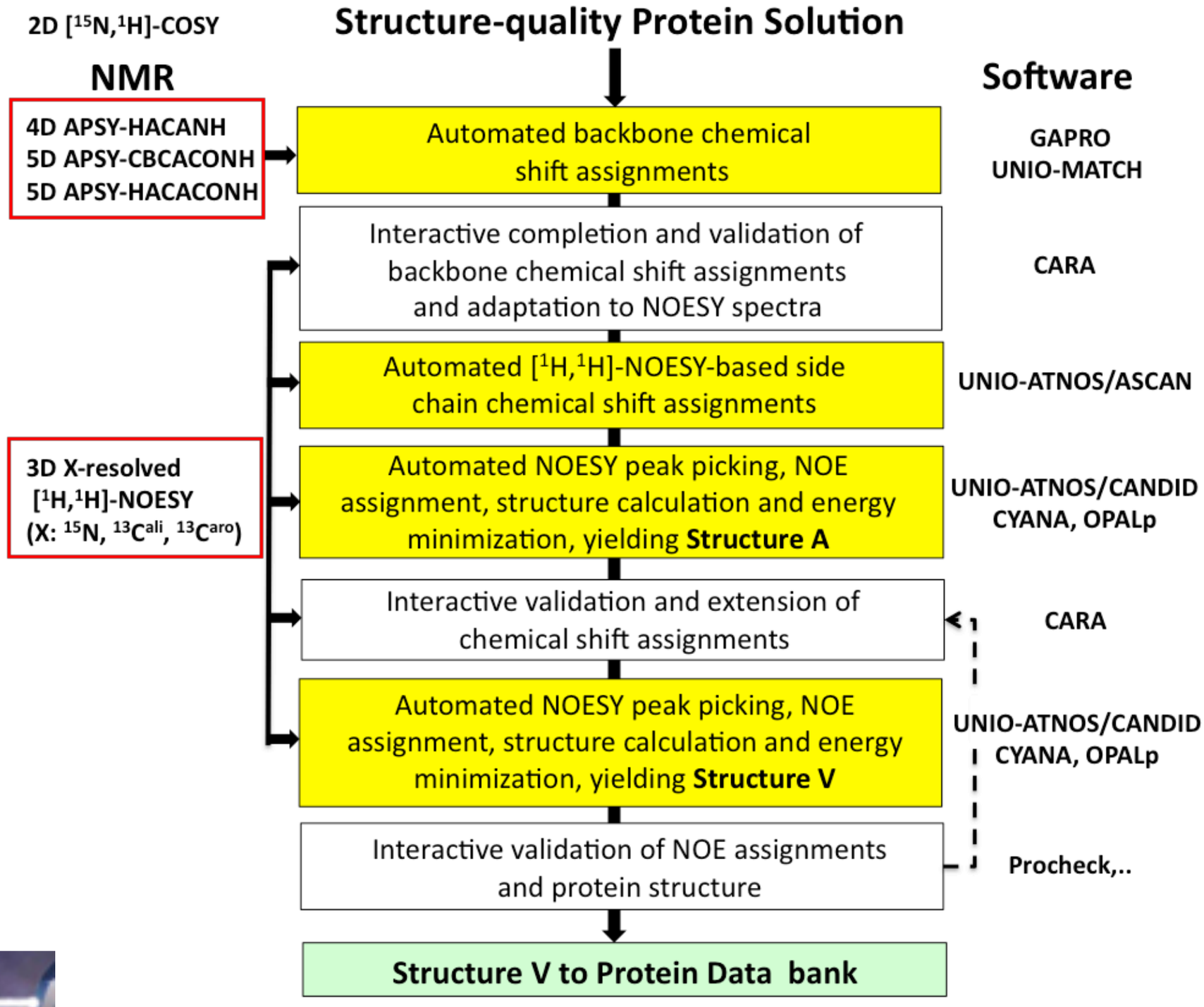




- UNIO assembles four previously developed **expert algorithms** into three principal data analysis modules, each one **operating directly on the NMR spectra**.

- Correct and consistent chemical shift referencing of all NMR spectra.
- Adaptation of previous chemical shifts to subsequently used set of NMR spectra.

Application example: The J-UNIO protocol



Serrano, Pedrini, Mohanty, Geralt, Herrmann, Wüthrich. *J. Biomol. NMR* **2012** Aug. 53(4): 341-354

J-UNIO protocol: a representative example

Initial Phase 1

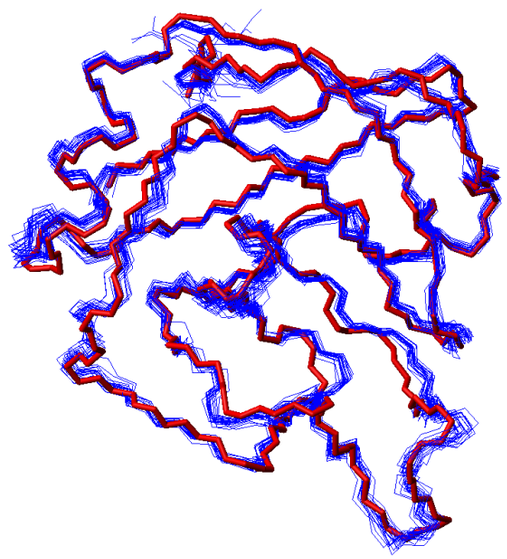
TM1367
124 aa

NMR data collection and automated NMR structure determination

7d

RMSD to X-ray structure
1.11 ± 0.10 Å (bb)

Starting Structure A



Real-world UNIO application: an example

Initial Phase 1

TM1367
124 aa

NMR data collection and automated NMR structure determination — 7d

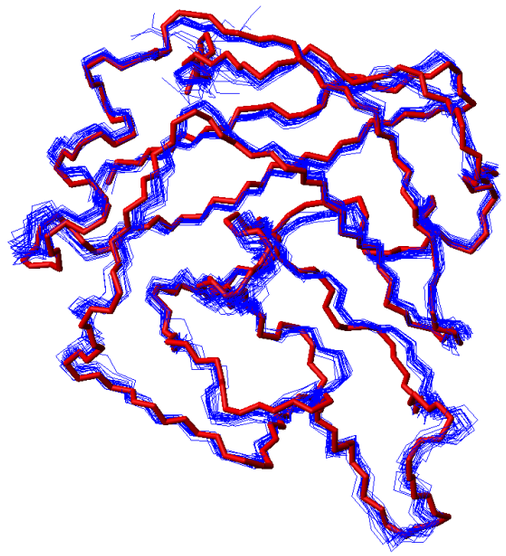
RMSD to X-ray structure
1.11 ± 0.10 Å (bb)

Refinement Phase 2

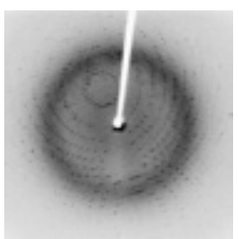
Interactive NMR structure refinement — 6d

RMSD to X-ray structure
0.92 ± 0.06 Å (bb)
1.54 ± 0.06 Å (ha)

Final Structure V



Similar to the work-flow seen in X-ray

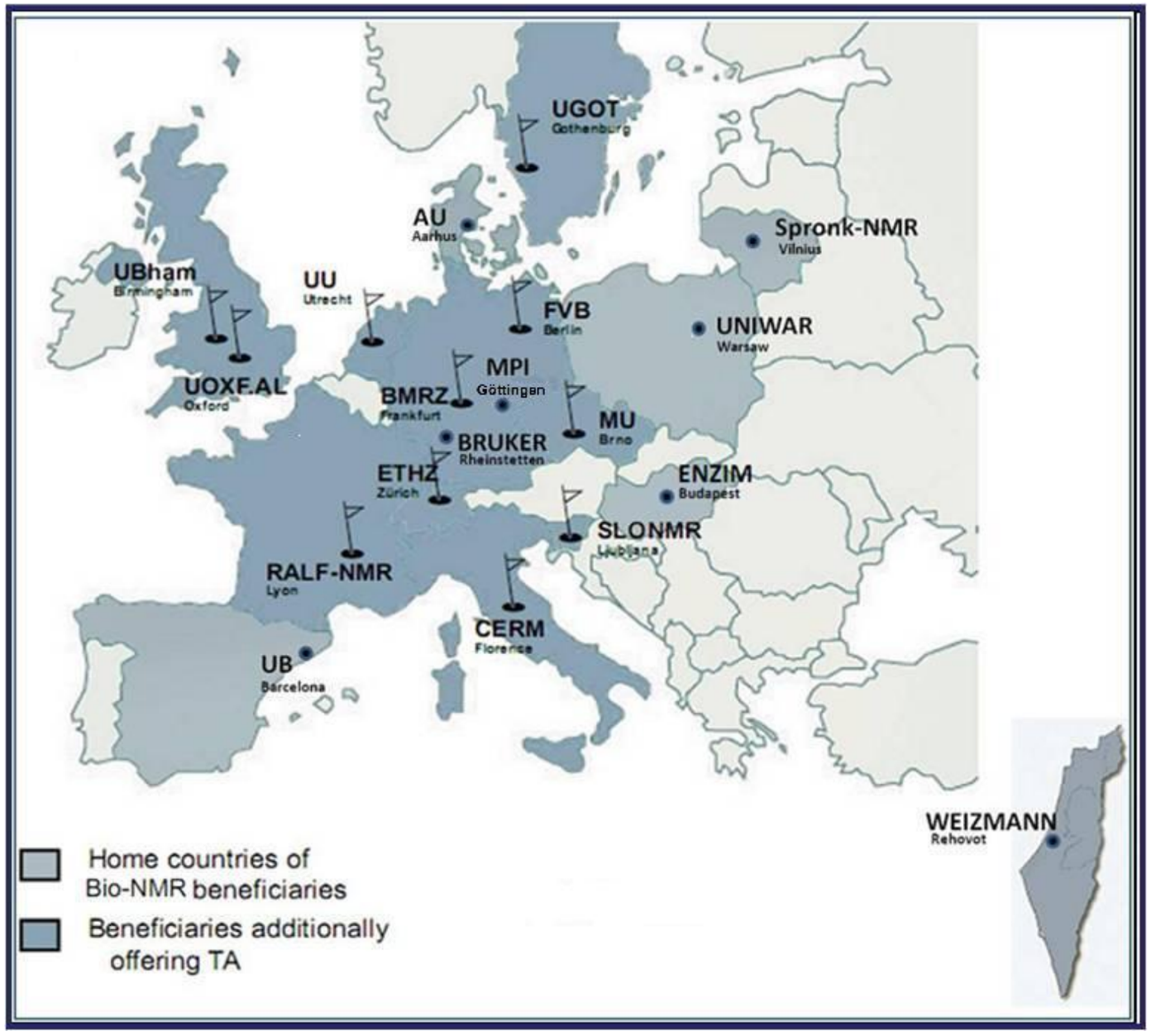


PDB structures determined by the J-UNIO protocol

Initial Phase 1 versus Refinement Phase 2

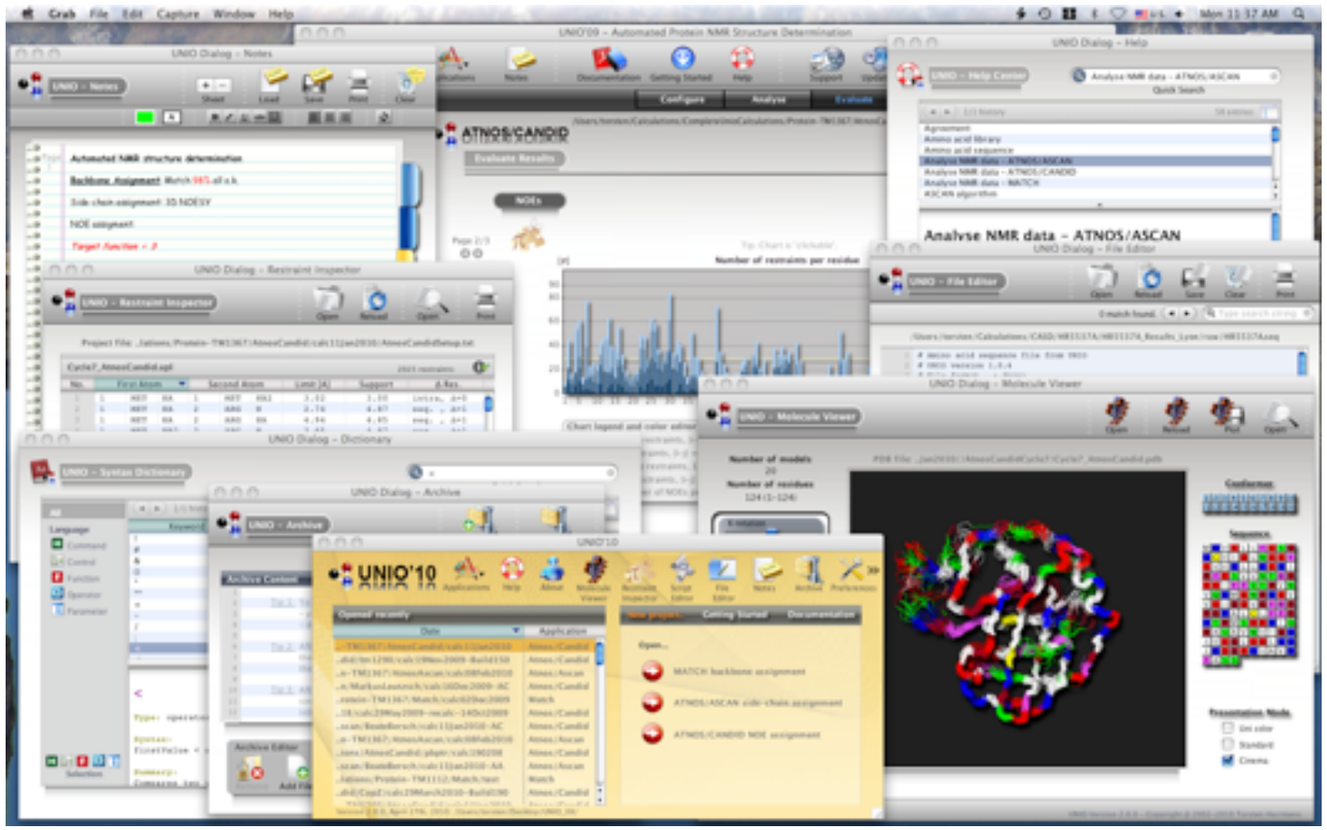
Protein sample (PDB id)[a] 17 (37)	Size (aa)	Chemical shift assignments				Precision (bb RMSD, Å)			$RMSD_{AV}$[i]
		$H^{\alpha},C^{\alpha},H^{N},N,C',C^{\beta}$ (%)		All-atoms[b] (%)					
		MATCH[c]	Final[d]	ASCAN[e]	Final[d]	Residues[f]	ASCAN[g]	Final[h]	
TM1112 (2k9z)	89	96	100	72	90	2–89	0.79	0.43	1.45
TM0212 (2ka7)	124	100	100	66	92	1–110	0.69	0.51	1.16
TM1367 (2ka0)	124	92	98	80	92	2–123	0.80	0.44	2.28
A2LD1 (2kl2)	149	85	97			2–100,106–144			
YP_001336205.1 (2l1s)	83	82	99	89	95	4–82	0.73	0.44	1.23
TM0320 (2kyz)	67	91	97	80	96	1-67	0.57	0.45	0.81
YP_510488.1 (2kzc)	85	93	96	76	94	1–85	1.17	0.68	1.41
NP_415897.1 (2kts)	117	81	100	76	93	3–117	1.65	0.62	1.64
YP_399305.1 (2l1n)	120	82	99	67	94	1–34,43–92,97-117	2.03	0.58	1.75
NP_954075.1 (2l1t)	109	91	98	78	95	7–103	0.80	0.64	1.14
NP_253742.1 (2l6p)	124	81	95	76	91	2–38,49–117	1.08	0.61	1.17
YP_001092504.1 (2l6n)	132	76	96	78	93	8–44, 56–120	2.42	0.70	2.82
YP_926445.1 (2l6o)	114	92	98	67	95	10–40,44–113	1.98	0.64	2.32
NP_888769.1 (2l25)	141	87	99	78	93	3–50, 66–136	1.78	0.71	2.01
YP_546394.1 (2l9d)	108	95	99	75	95	8–108	1.89	0.67	2.01
YP_557733.1 (2la7)	145	75	100	80	96	18–144	1.37	0.58	1.97
YP_001302112.1 (2lg7)	129	92	100	81	93	11–81, 91–129	1.45	0.75	2.20

RALF-NMR offers the J-UNIO protocol



Financial support by the Access to Research Infrastructures activity in the 7th Framework Programme of the EC (Project number: 261863, Bio-NMR) for conducting the research is gratefully acknowledged

*But UNIO is **more** - Multipurpose data analysis platform*



One of the major strengths and specificities of UNIO is it's flexibility. Modules can be launched separately, out of sequence and with different input data. Therefore the UNIO setup can be tailored for any structure determination project according to a particular problem at hand.

Create your own protocol

CASD-NMR

Critical Assessment of automated Structure Determination of proteins from NMR data

Community-wide initiative for assessing the feasibility of obtaining in an unsupervised manner solution NMR structures of proteins with a quality for direct deposition into the Protein Data Bank.

CASD-NMR 1 (June 2009 - Feb 2010)

10 blind structure determinations of proteins

protein sequence, resonance assignment and **unassigned, refined peak lists**

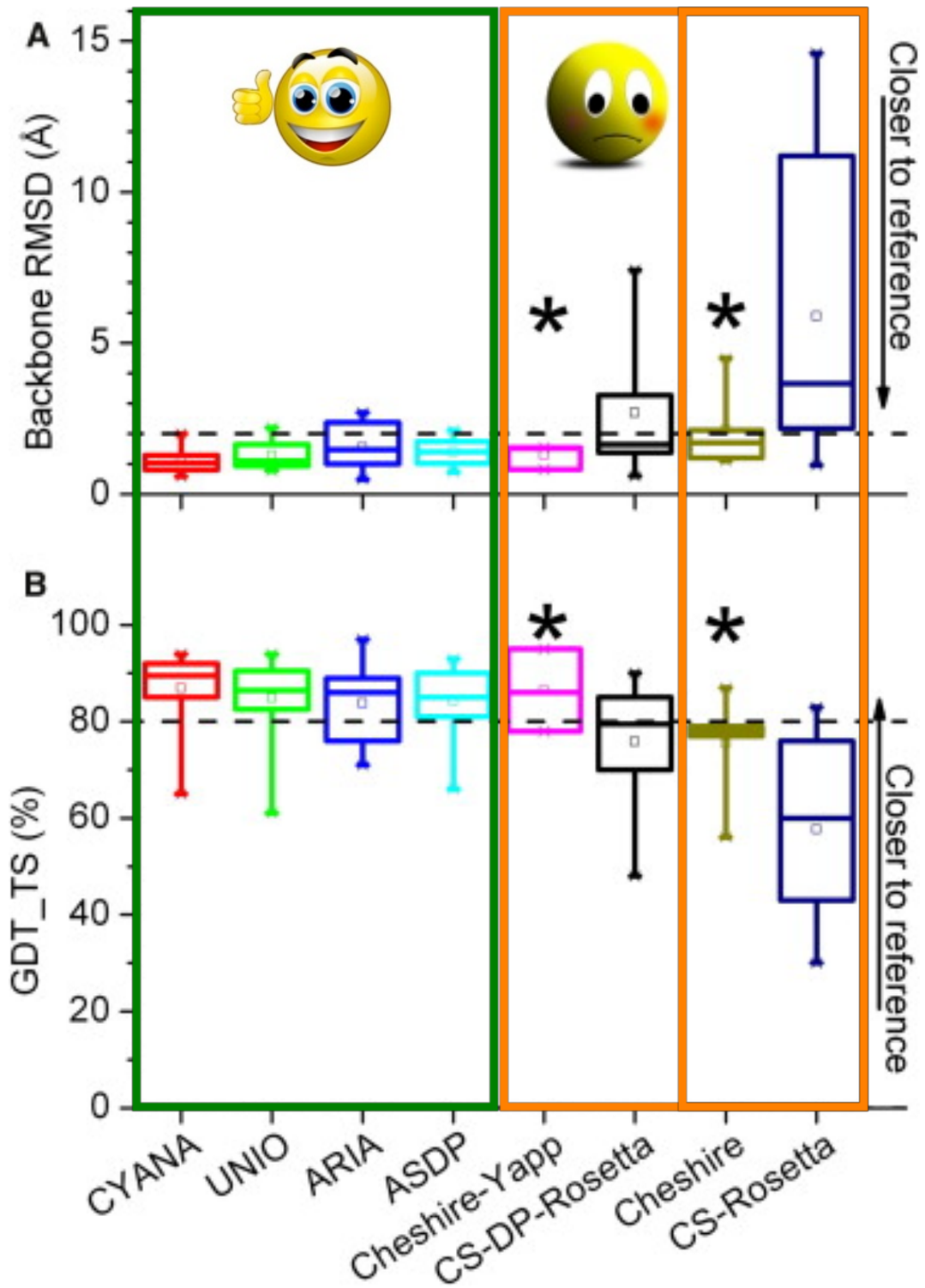
CASD-NMR 2 (March 2011-present)

8 blind structure determinations of proteins (so far)

protein sequence, resonance assignment and **NMR spectra**

Deploying & unifying the computational NMR infrastructure

CASD-NMR 1: Results & Conclusions



Structure **2012** Feb. 20(2): 227-236.

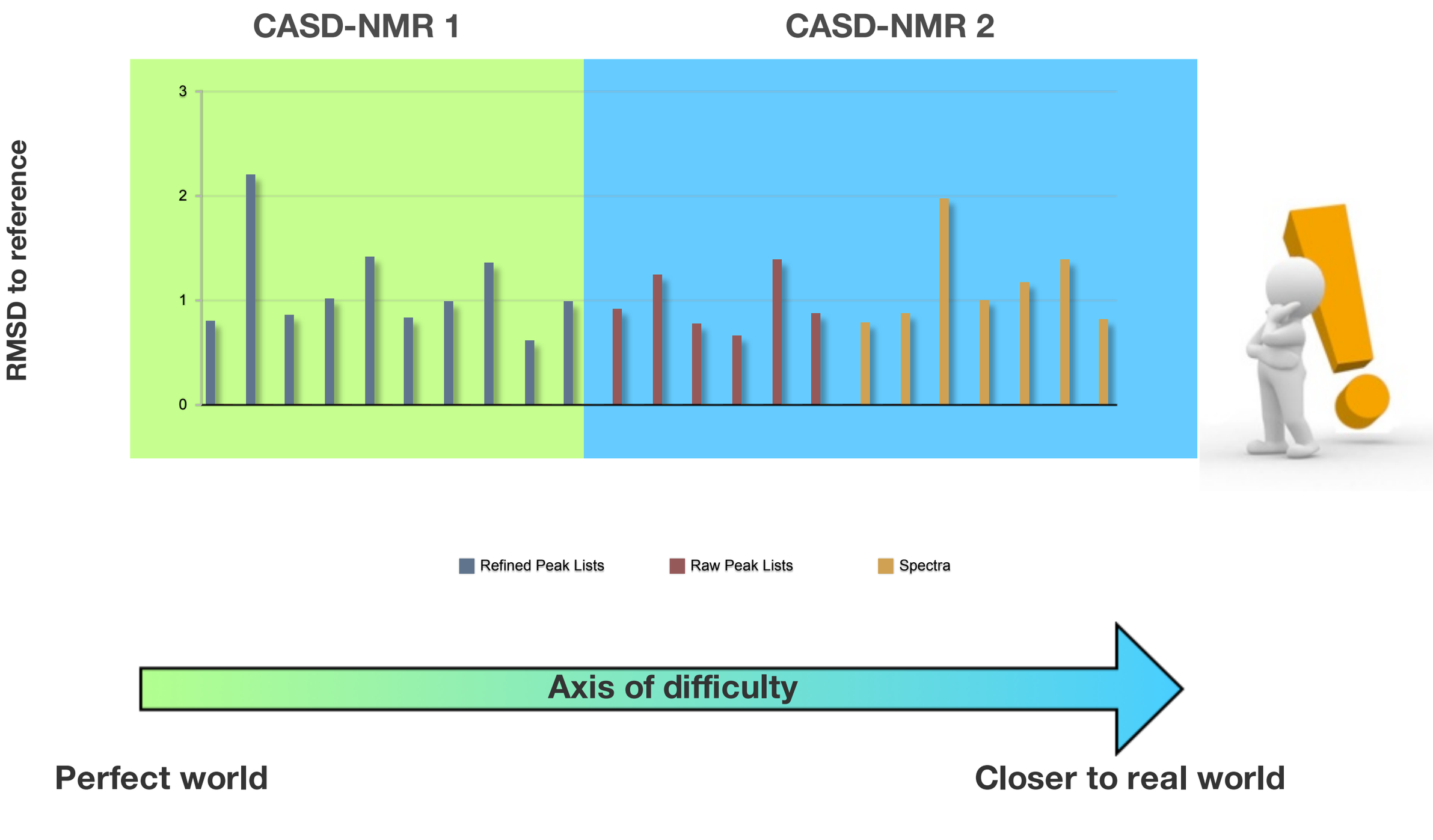
- **Automated NOE assignment yields structures < 2 Å from the reference**

- **Chemical-shift based calculations show high variability in their performance. Filtering with NOESY data might help sometimes.**

"*It can be concluded that* ***NOESY-based methods*** *delivered* ***more consistent and robust performances*** *than CS-based methods, yielding structures* ***on average closer to the reference****. NOESY-filtering as in CS-DP-Rosetta could recover some but not all of the consistency and reliability of the restraint-driven methods (discussed later).* ***Notably,*** *the* ***CS-methods*** *(regardless of whether augmented with NOESY information) are* ***computationally much more demanding*** *than NOESY-based methods.*"

From Structure 2012, 20(2): 227-236.

CASD 1&2-NMR: UNIO'12 performance



Outline - Part 2

UNIO

UNIO'13 - Integration and modeling

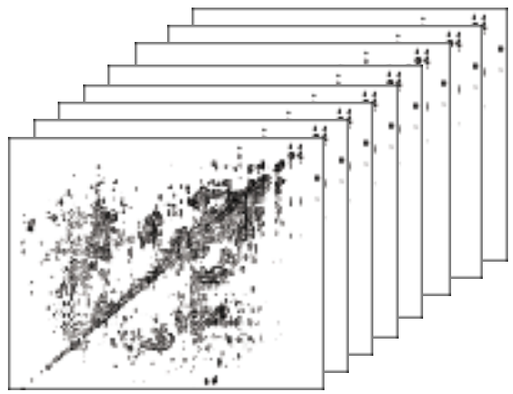
How to use data base information?

Correctness?

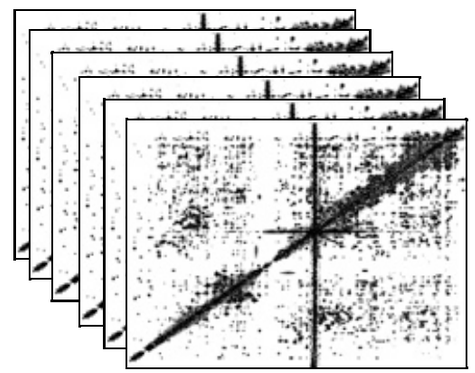
BMRB

PDB

Homology models

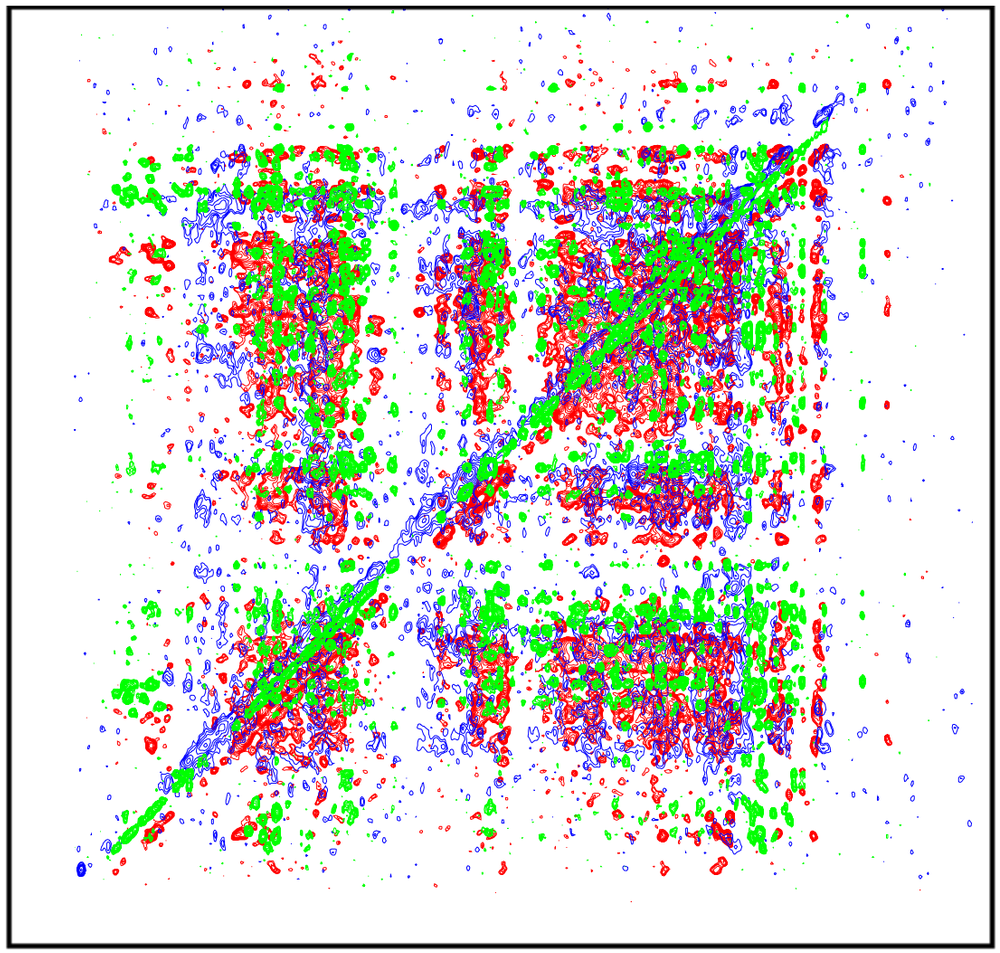


Resonance assignment

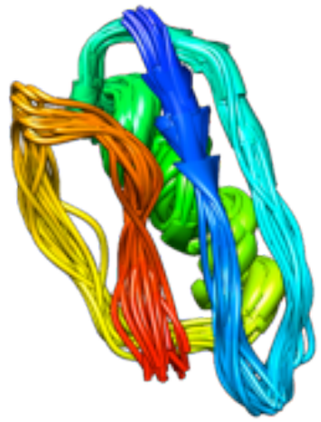


Conformational restraints

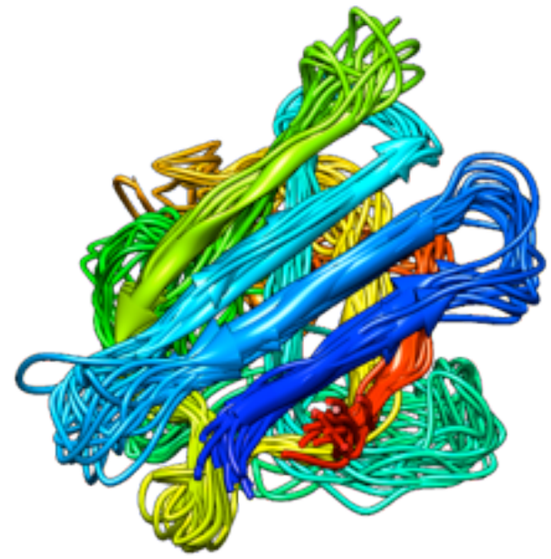
Data base information can be dangerous!



Ubiquitin GB1 SOD



GB1 from ubiquitin spectra
rmsd to reference: 1.7 Å



SOD from GB1 & ubiquitin spectra
rmsd to reference: 2.8 Å

Prerequisite for the application of data base knowledge: Correct chemical shift referencing

UNIO-MATCH

- **LACS: Wang et al. (2005) J. Biomol. NMR 32.**
- **CheckShift: Simon et al. (2007) J. Biomol. NMR 39.**
- **PANAV: Wang et al. (2010) J. Biomol. NMR 47.**
- **VASCO: Rieping et al. (2010) Proteins 78.**
- **RefDB: Zhang et al. (2003) J Biomol NMR 25.**

Potential weaknesses of current procedures:

- ✦ **Only applicable to a subset of backbone atoms.**
- ✦ **Strong assumptions on correct referencing of selected backbone shifts.**
- ✦ **Structure needed as input.**

Not Found

The requested URL [redacted] was not found on this server.

Additionally, a 404 Not Found error was encountered while trying to use an ErrorDocument to handle the request.

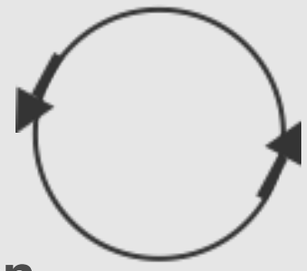
Prerequisite for the application of data base knowledge: Correct chemical shift referencing

- **Data base of average values for secondary chemical shifts of atoms in protein (coil, strand, helix).**

  (e.g.Wang & Jardetzky Protein Sci. 2002)

  - **For each residue, determine the secondary structure elements using the joint probability over CA, CB, CO, N, HN, HA shift values.**

  - **Determine individually for each backbone atom the offset between input chemical shift value and derived secondary distribution function.**



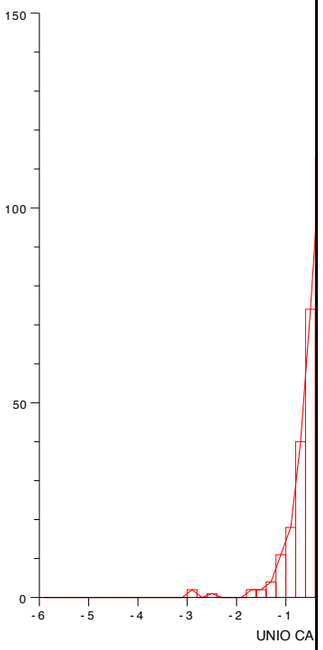
★ **Crucial merit 1: chemical shifts do not need to be assigned, i.e., procedure applicable at the outset of a structure determination process.**

★ **Crucial merit 2: Chemical shift correction and standard deviations is given.**

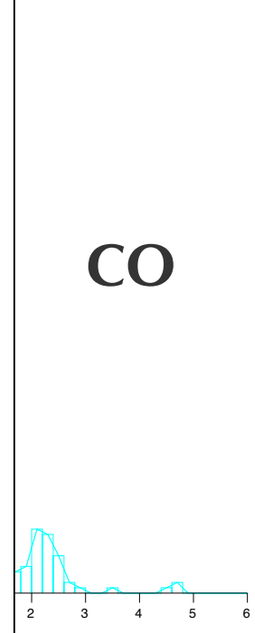
★ **Now: For backbone resonance assignment a safe comparison to chemical shift statistics can be done.**

Guerry & Herrmann (submitted)

Distribution of CA

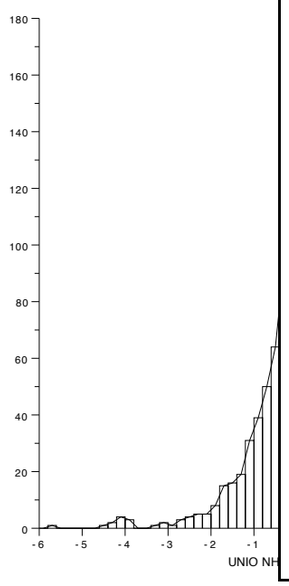


offset in the BMRB

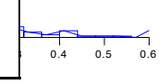


CO

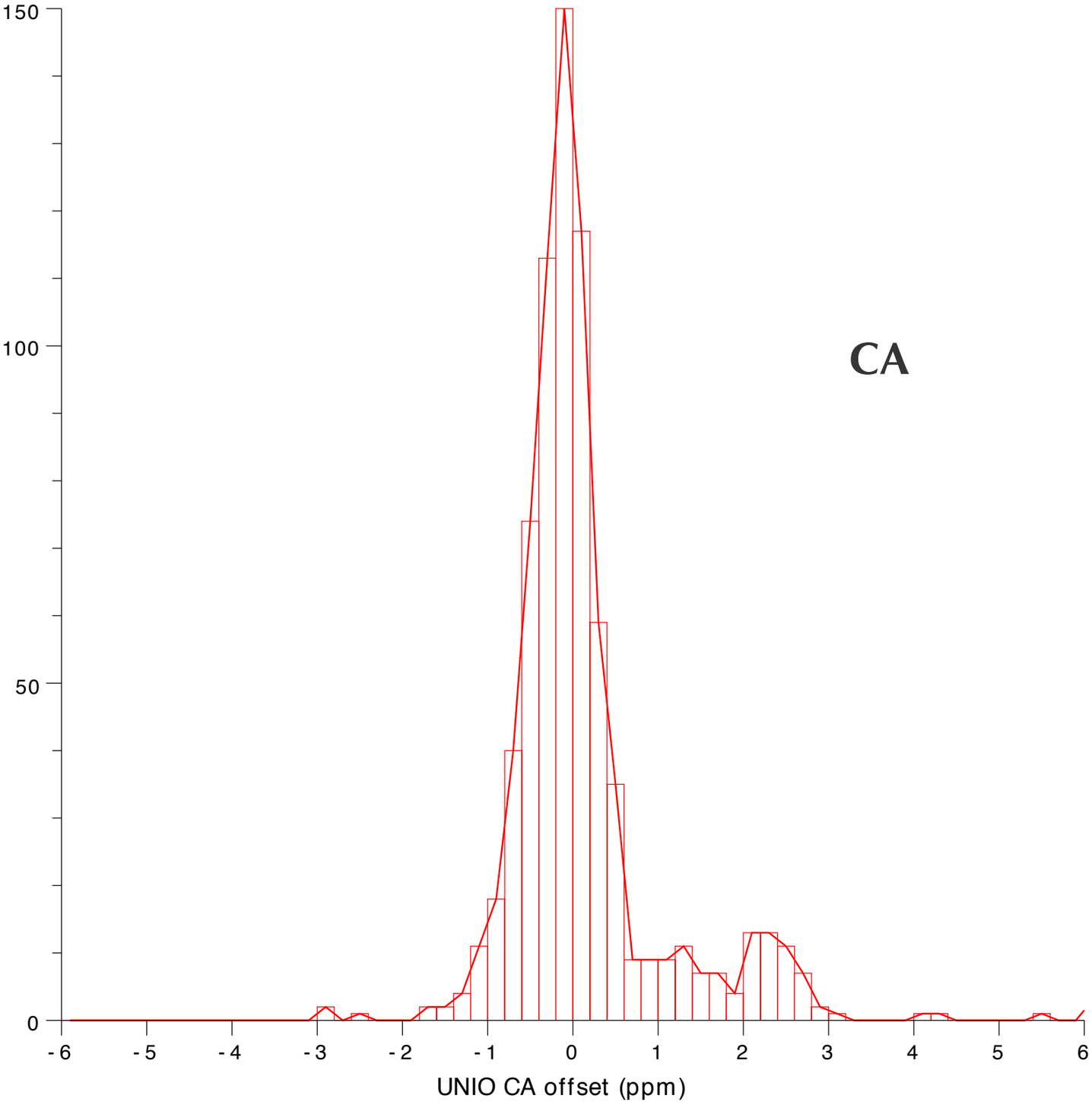
Distribution of NH



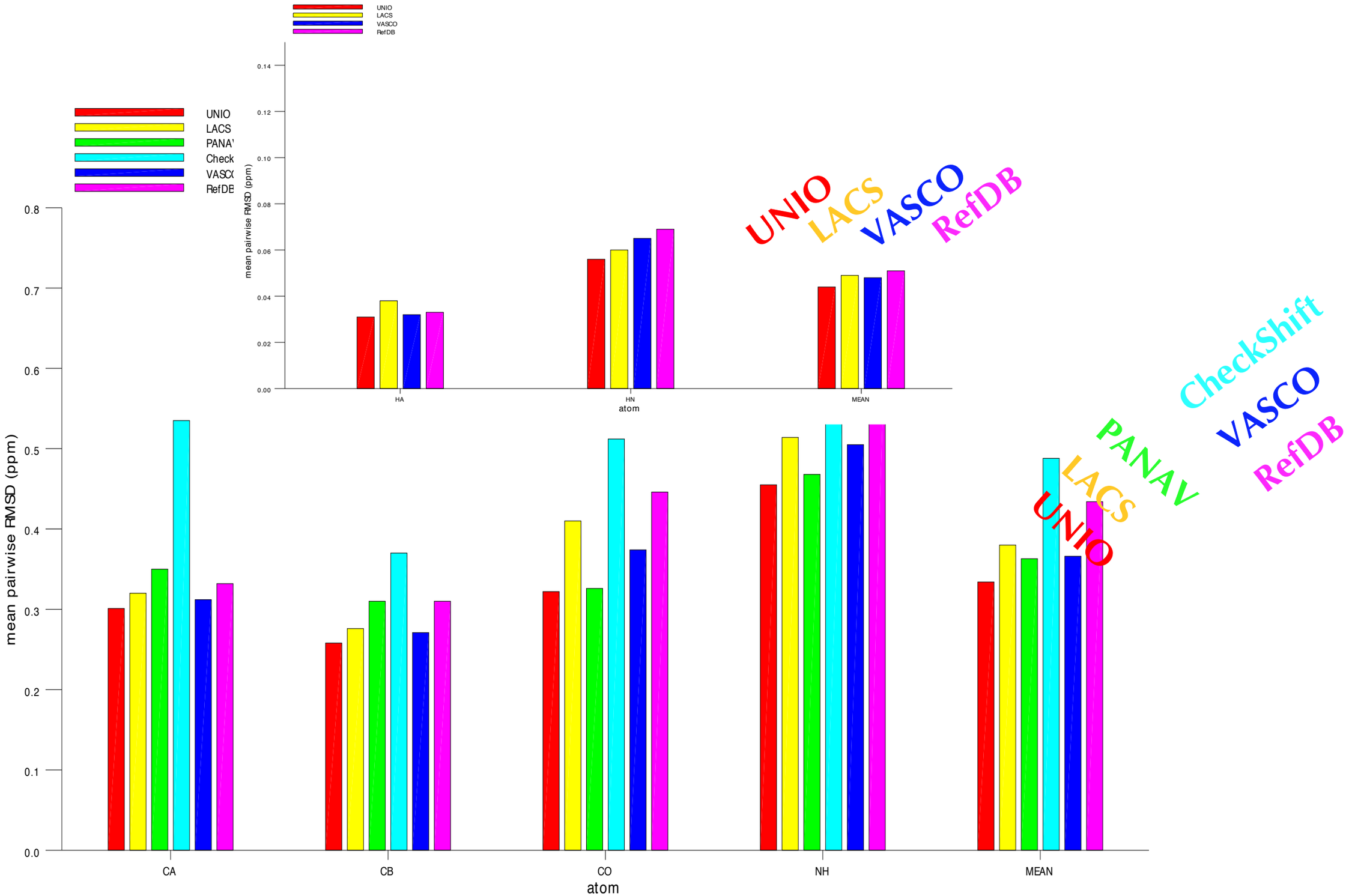
offset in the BMRB



HA



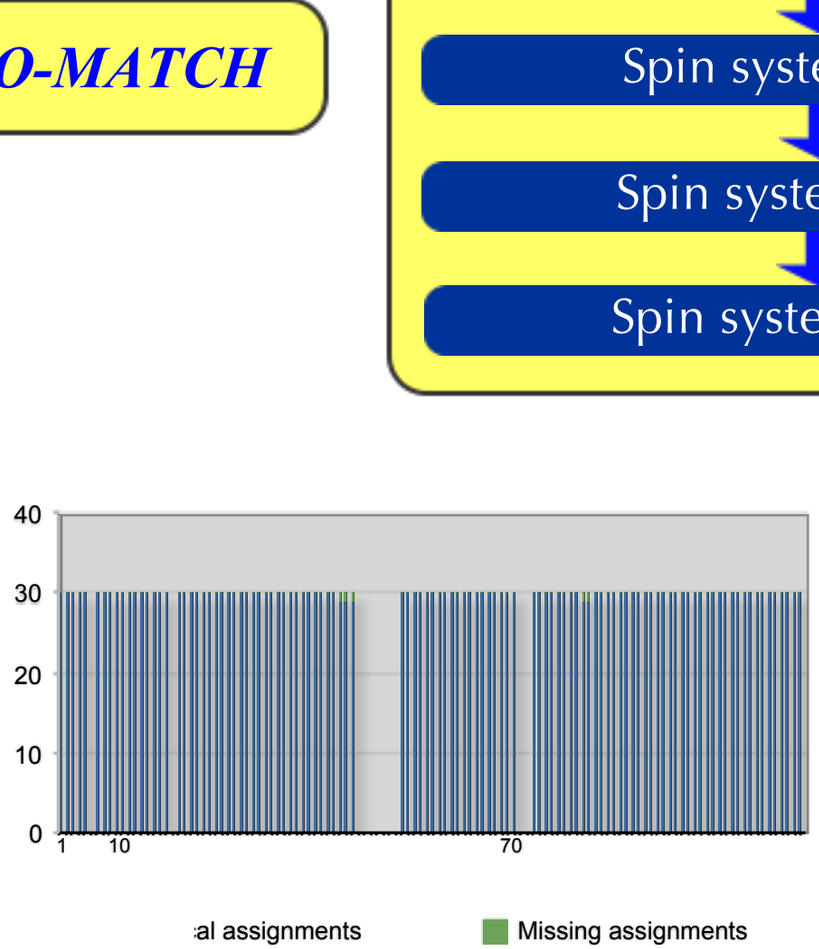
UNIO RefBMRB - Comparison to other programs

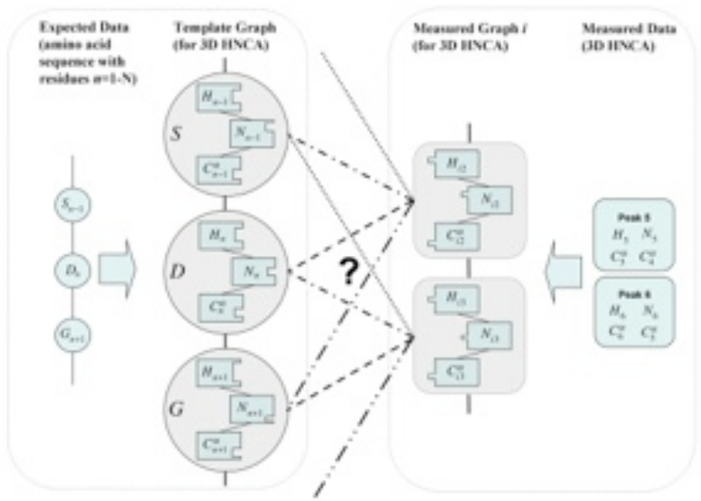


I. Integration & modeling: Backbone resonance assignment

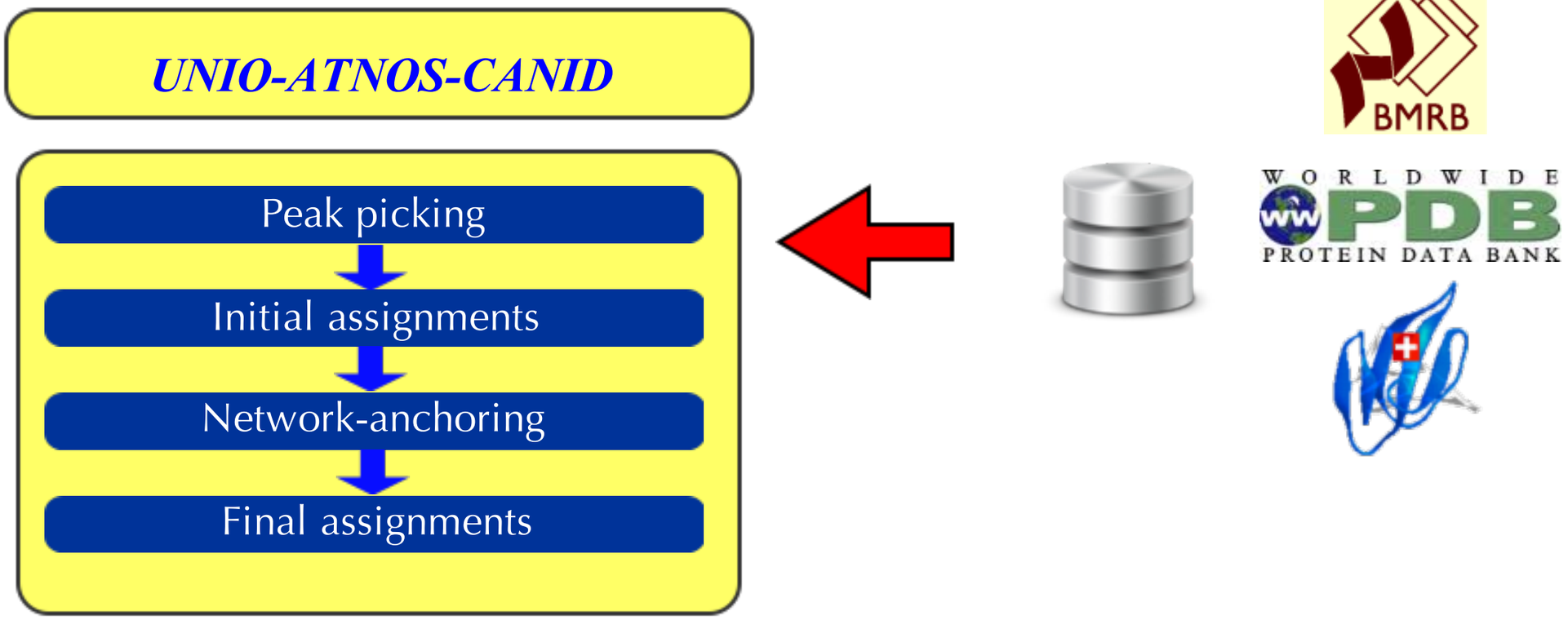
UNIO-MATCH

- Peak picking
- Spin system grouping
- Spin system typing
- Spin system linking
- Spin system placing





2. Integration & modeling: NOE assignment

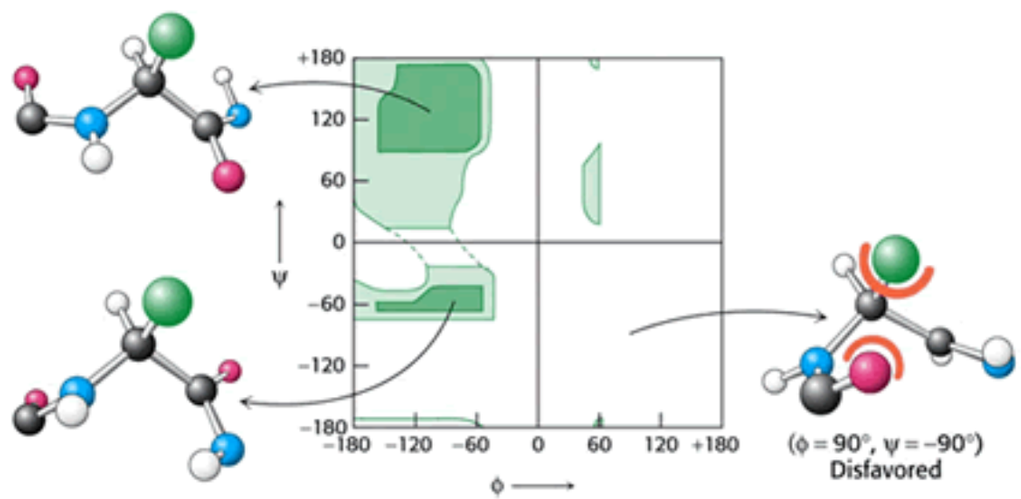


★ **Principal idea 1:** Retain structure-based assignments together with other assignment possibilities in the first round of NOE assignment.

★ **Principal idea 2:** Either structure-based assignments are in agreement with experimental data; if not, algorithm will converge to the ‘real’ structure. Retained structure-based assignments were only an initial disturbance of the procedure.

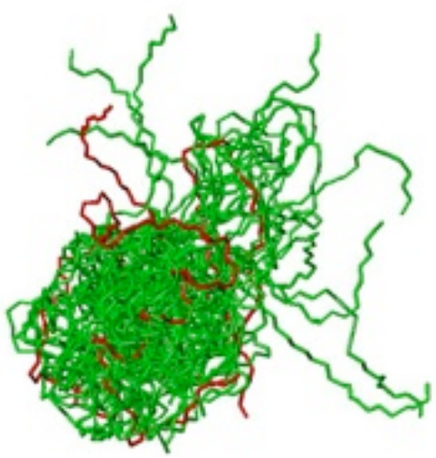
3. Integration & modeling: Distance estimations by exhaustive grid search

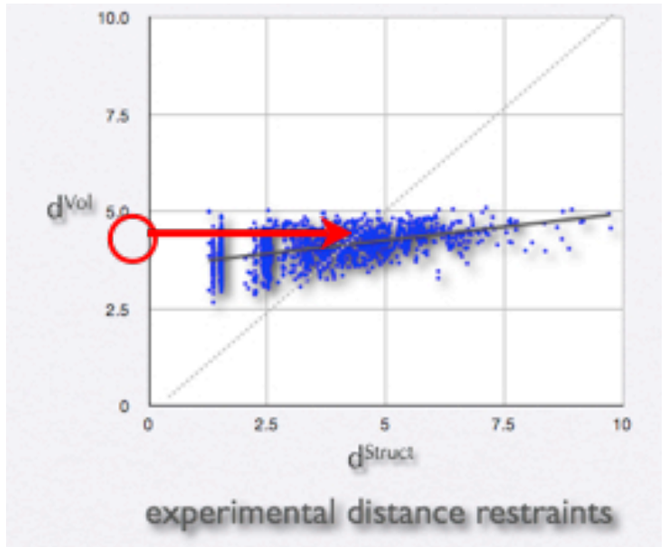
Correct upper bounds for distance restraints.



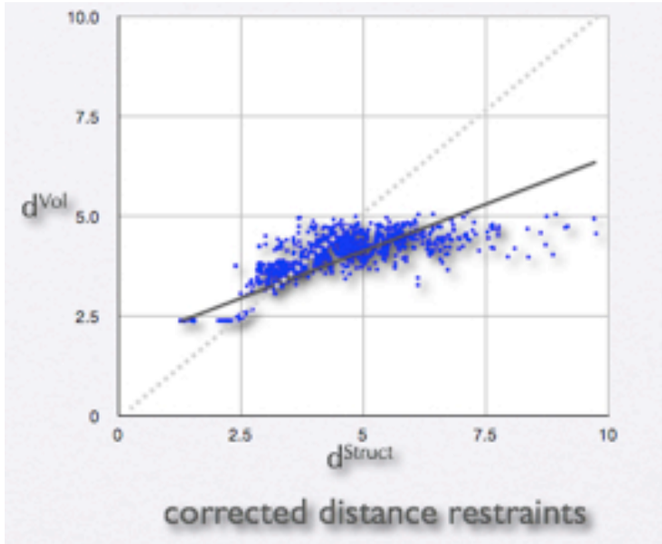
Distance estimation in proteins

- **Exhaustive conformational sampling.**
- Incorporation of a **priori knowledge**.





Without



With

Outline - Part 3

UNIO

UNIO'13 - Integration and modeling

Nanomachineries

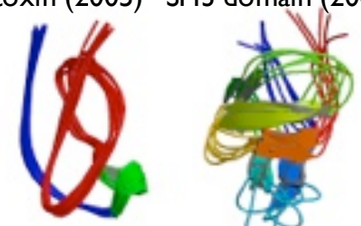
Solid state NMR structure determination

Micro-crystalline proteins

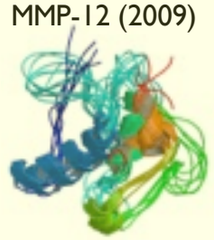
GB1 (2005, 2008, 2009) Ubiquitin (2008)

CrH (2008)

Kaliotoxin (2005) SH3 domain (2002)



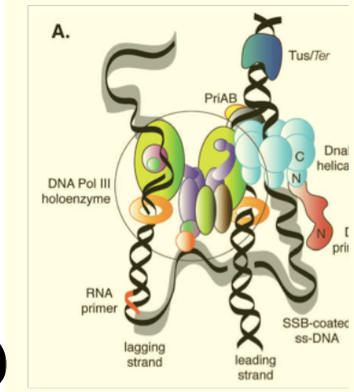
MMP-12 (2009)



Next targets

αβ crystalline (2010)

SOD

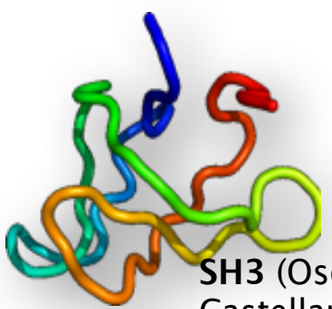


DNA Polymerase III

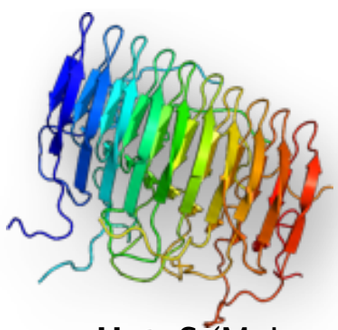
?

4 6 8 10 12 14 16 18 30 1000

Molecular Weight [kDa]



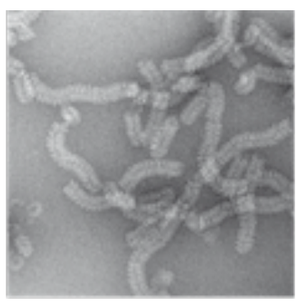
SH3 (Oschkinat group), Castellani et al. Nature 2002

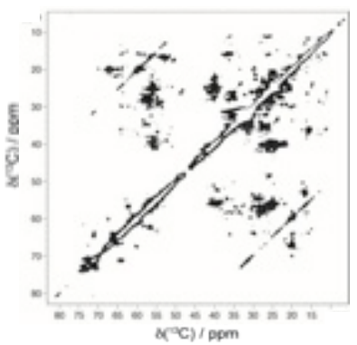


Het-S (Meier group), Wasmer et al. Science 2008

Type-III Secretion system (Lange group) Nature 2012

Virus nucleocapsid:



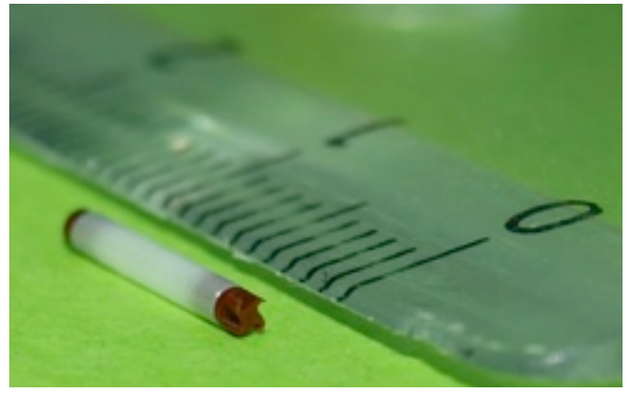


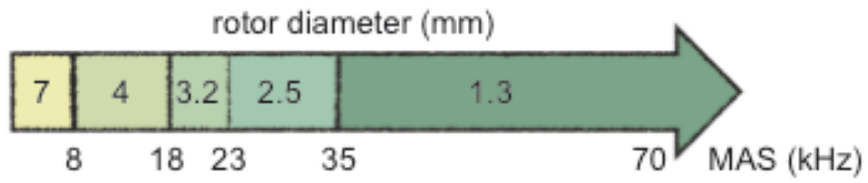
GHz spectra of GigaDalton molecules

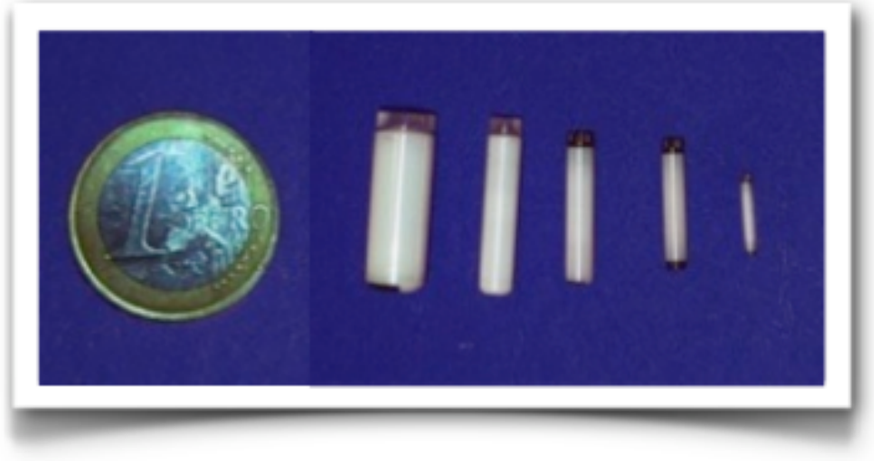
Proton detection, high magnetic field & ultra-fast MAS

- All results were recorded using a 1 GHz instrument and 60 kHz MAS at CRMN, ENS-Lyon

- 100% back-protonation was used in all experiments

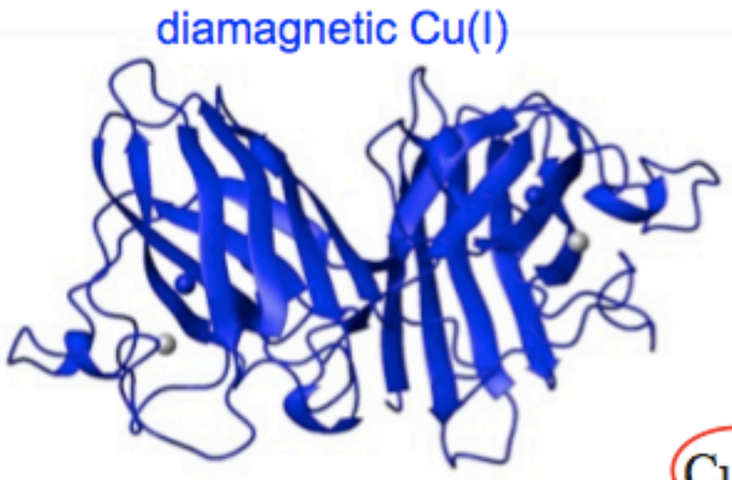


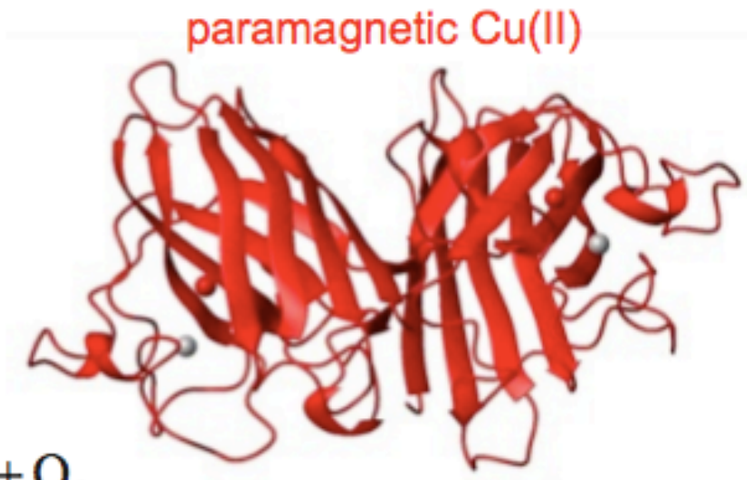




Model system: Human superoxide dismutase (32kDa dimer)

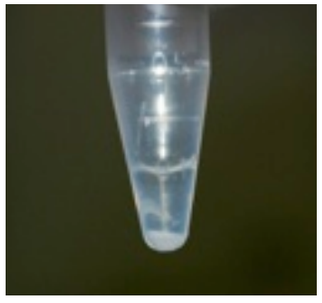
Human CU(II), Zn(II)-superoxide dismutase (32kDa)



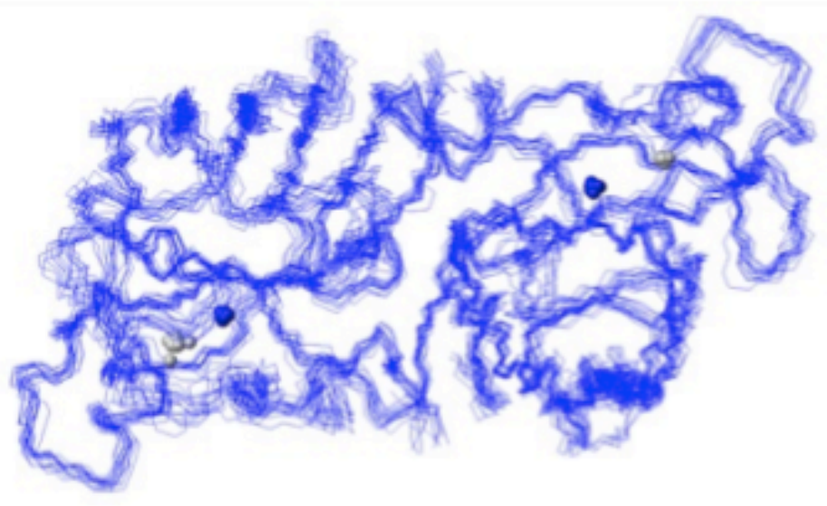


$$Cu^{2+} + O_2^- \rightarrow Cu^+ + O_2$$

$$Cu^+ + O_2^- \rightarrow Cu^{2+} + (O_2^{2-}) \xrightarrow{(2H^+)} Cu^{2+} + H_2O_2$$



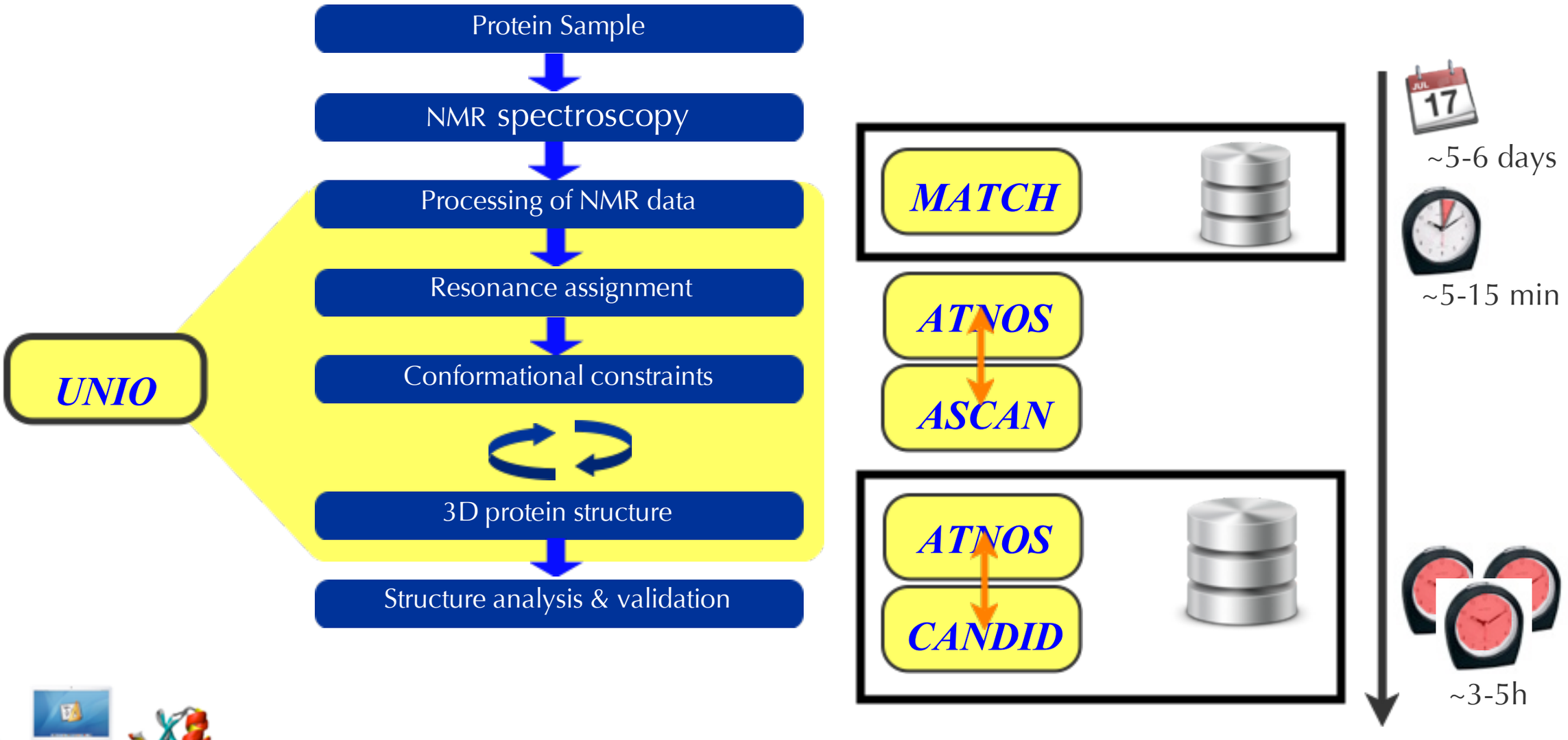
Sample: non-diffracting nanocrystals



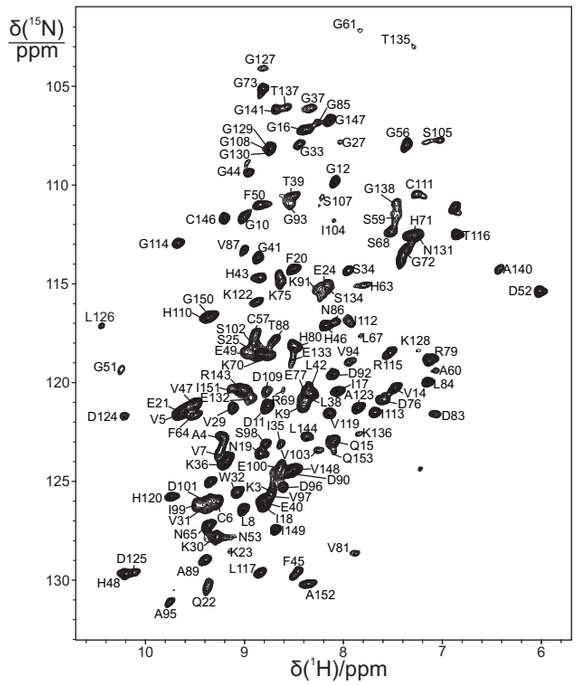
Banci et al., *Eur. J. Biochem.* **2002**

UNIO'13

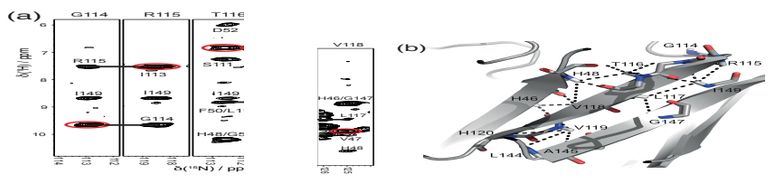
NMR data analysis using comparative modeling

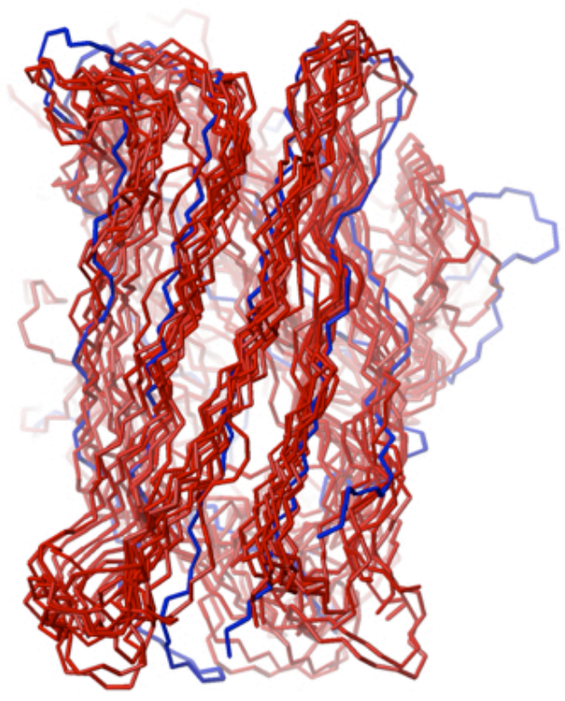


Resonance assignment & distance restraints using proton detection



+



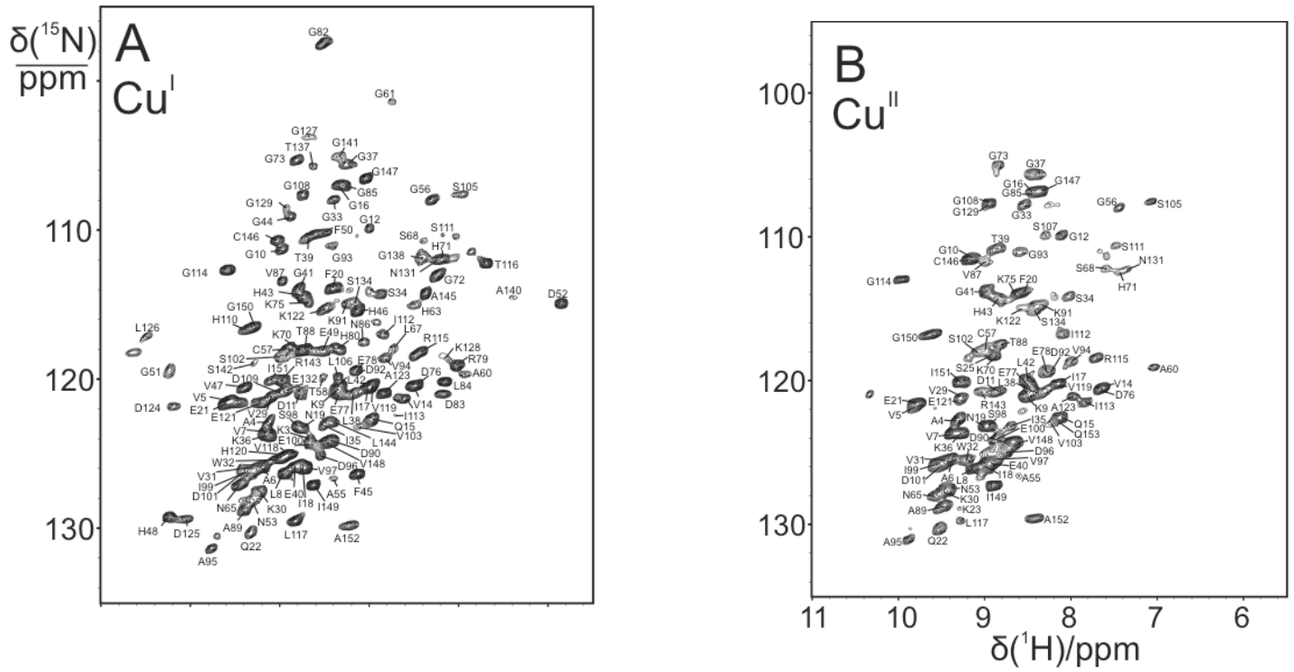


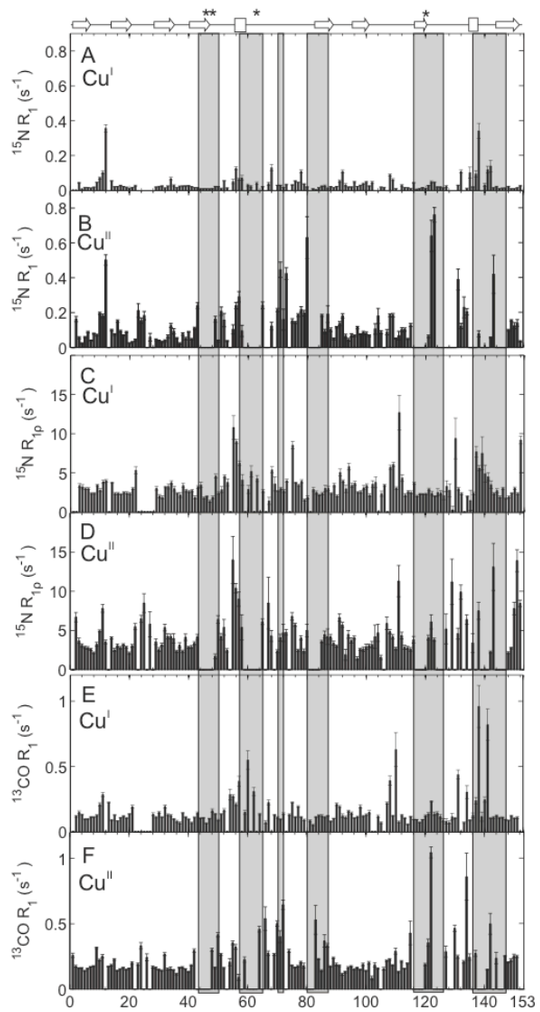
- Using UNIO-MATCH and UNIO-ATNOS/CANDID: 192 1H-1H restraints
- RMSD to mean: 1.64 Å
- RMSD to X-ray: 2.34 Å

Angew. Chem. - Int. Edit. **2011** Oct;, 50(49): 11697-11701

Site-specific backbone dynamics

- H-H spin-diffusion (RDFR) distance restraints -> short-range distance restraints
- Paramagnetic relaxation enhancements (PRE) -> long-range distance restraints
- Gaussian axial fluctuation (GAF) analysis -> order parameters and timescales

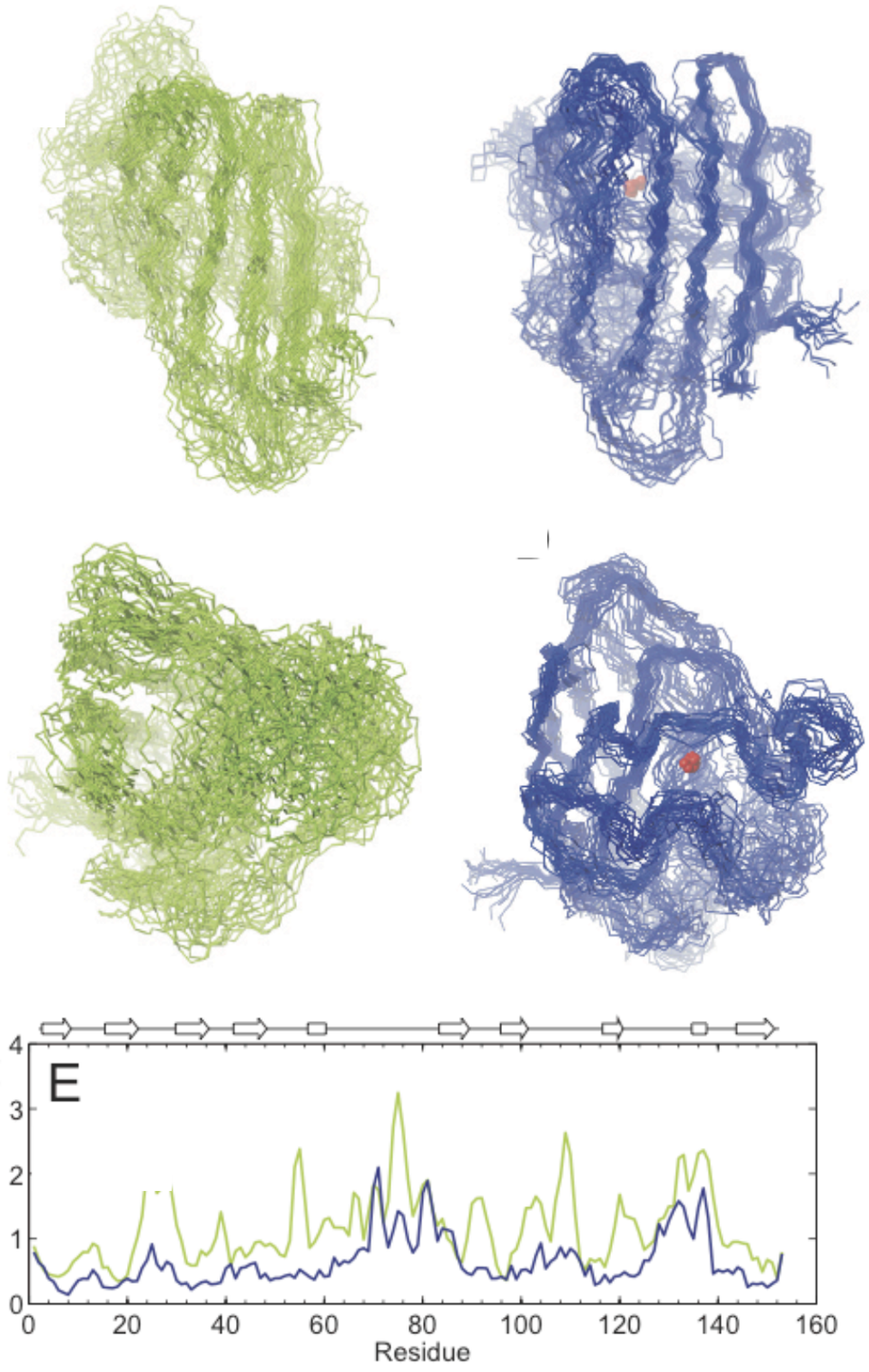




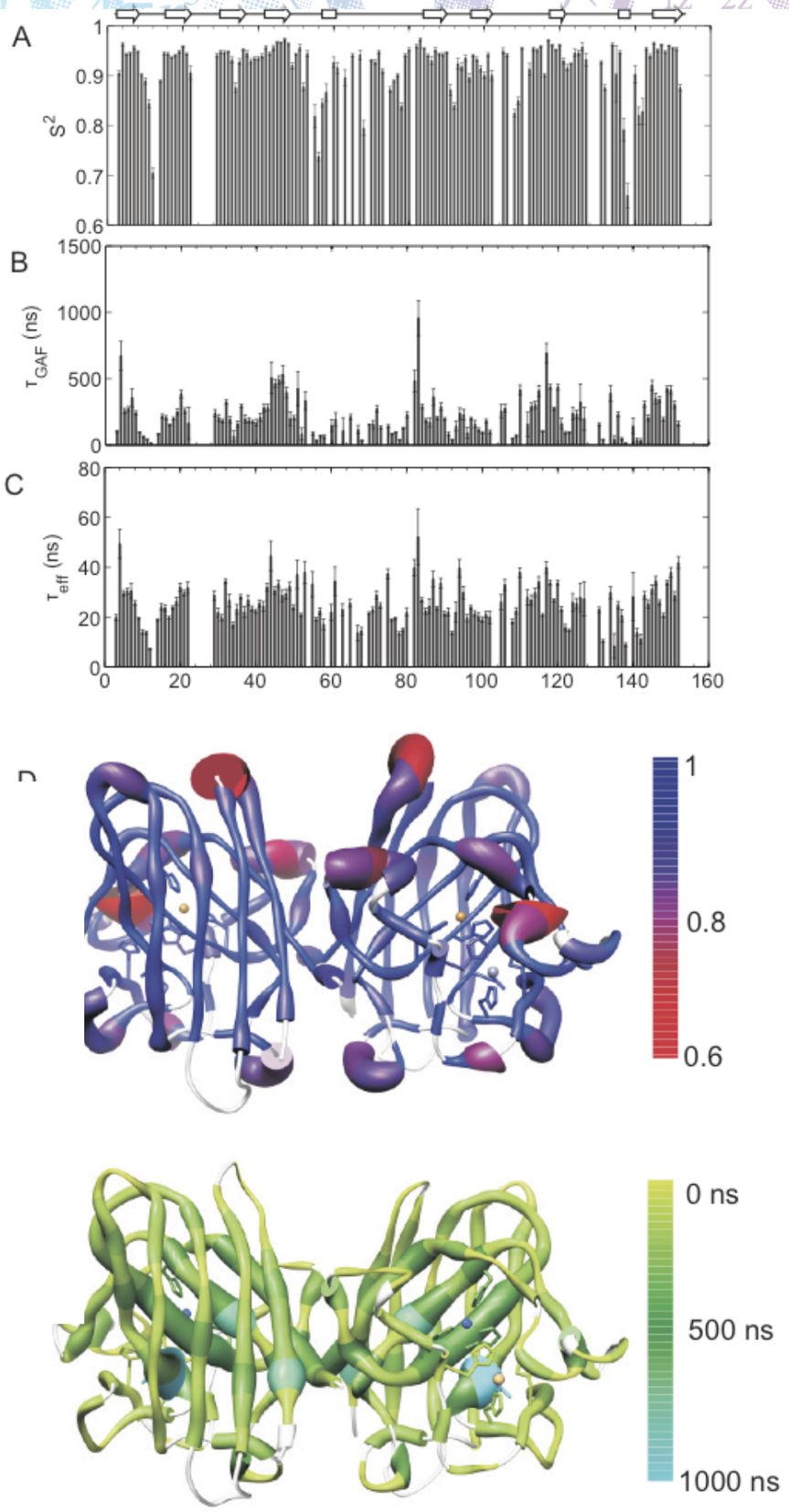
Proc. Natl. Acad. Sci. USA **2012** July, 109(28): 11095-11100.

Refined SOD structure and dynamics using PRE

w/o PRE

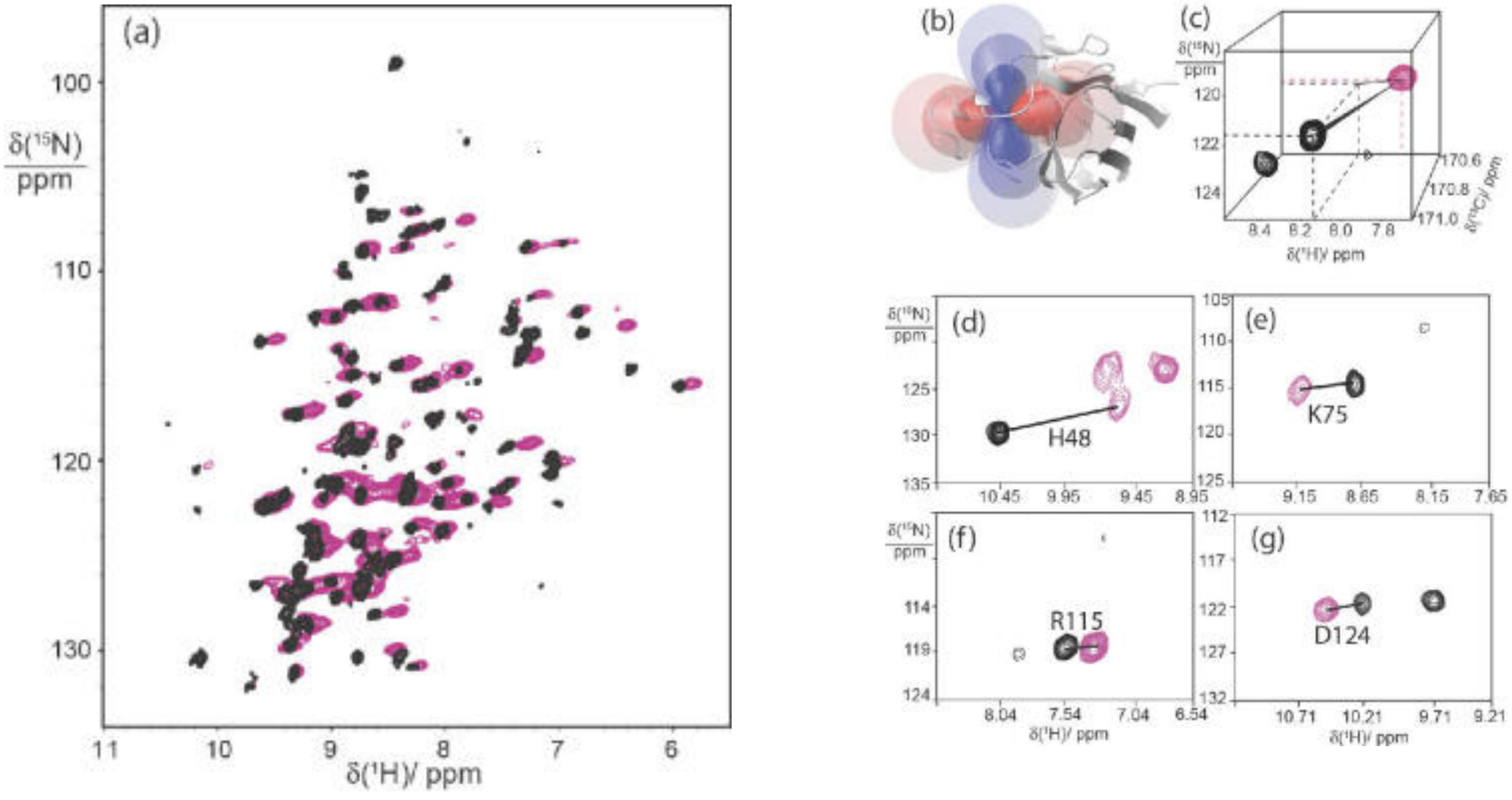


w PRE



Proc. Natl. Acad. Sci. USA **2012** July, 109(28): 11095-11100.

Pseudo contact shifts in SOD



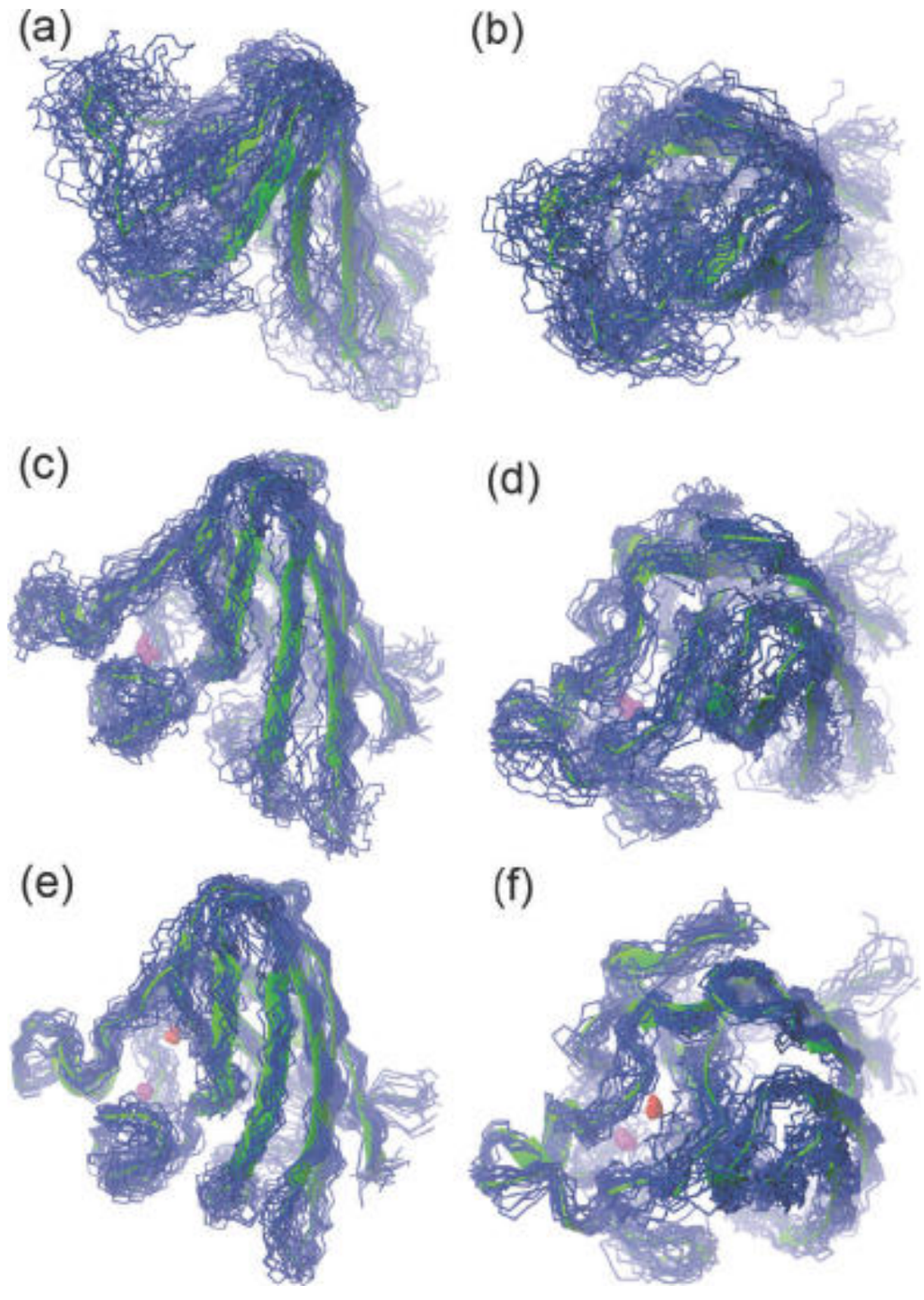
J. Am. Chem. Soc. **2012** Aug., 136(36): 14730-14733.

SOD structure using diamagnetic restraints, PRE and PCS

1H-1H only

1H-1H + PCS

1H-1H + PCS + PRE



J. Am. Chem. Soc. **2012** Aug., 136(36): 14730-14733.

Acknowledgment

UNIO and J-UNIO protocol

Paul Guerry

Solid State NMR

Michael Knight

Guido Pintacuda

CERM, Florence, Italy

Paul Guerry

Jozef Lewandowski

Benedicte Elena

Moreno Lelli

Anne Lesage

Lyndon Emsley