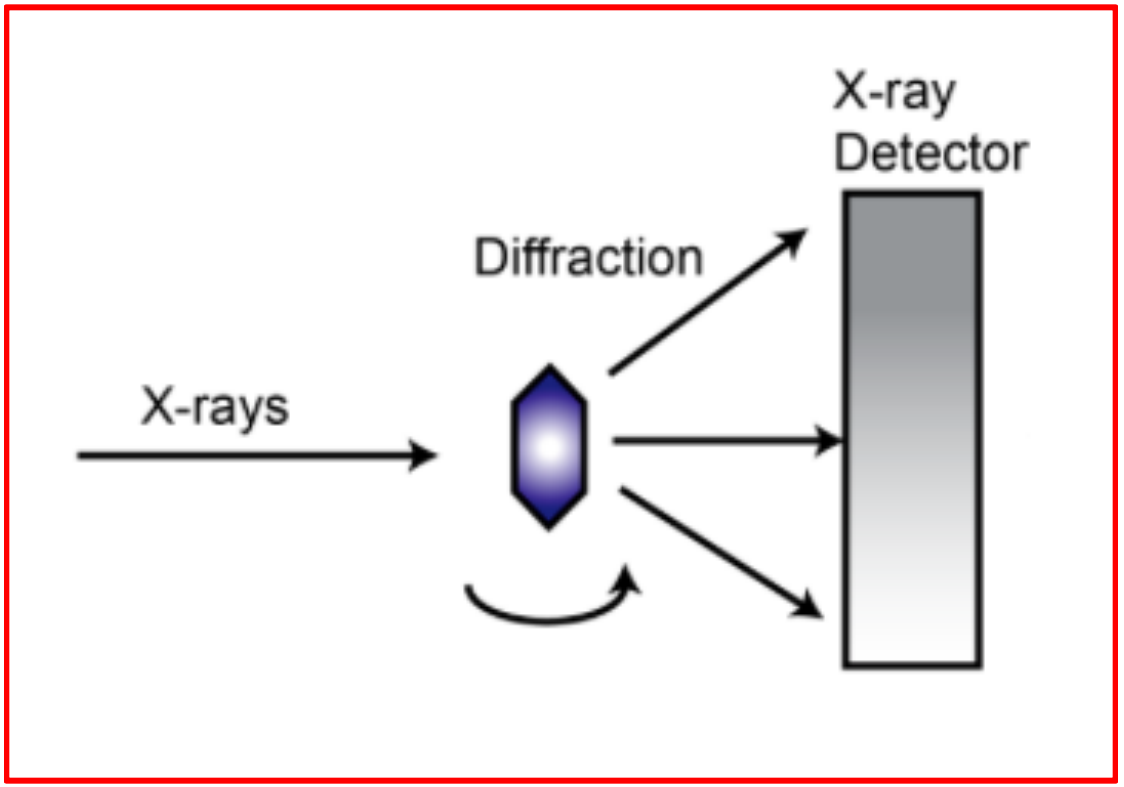
Rapid protein structure determination using experimental NMR data

Marius Wanko
University of Liege
amwanko@doct.ulg.ac.be

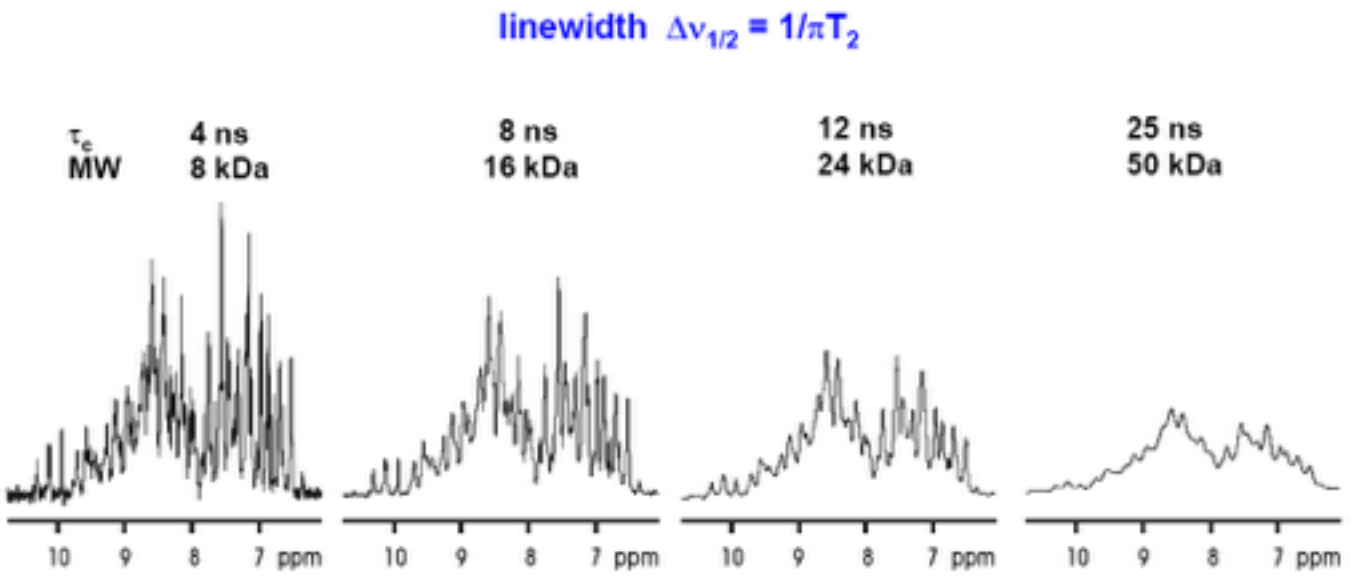
Protein 3D structure determination using experimental techniques

Liquid NMR

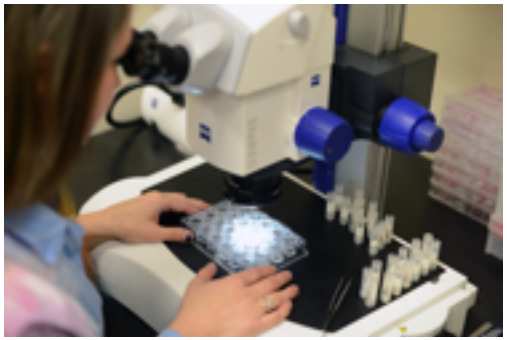


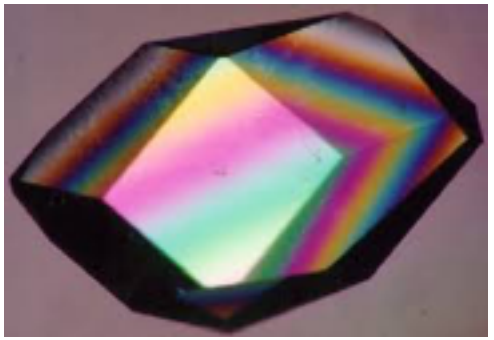
X-Ray Crystallography

- NMR spectra broadening increases with protein size

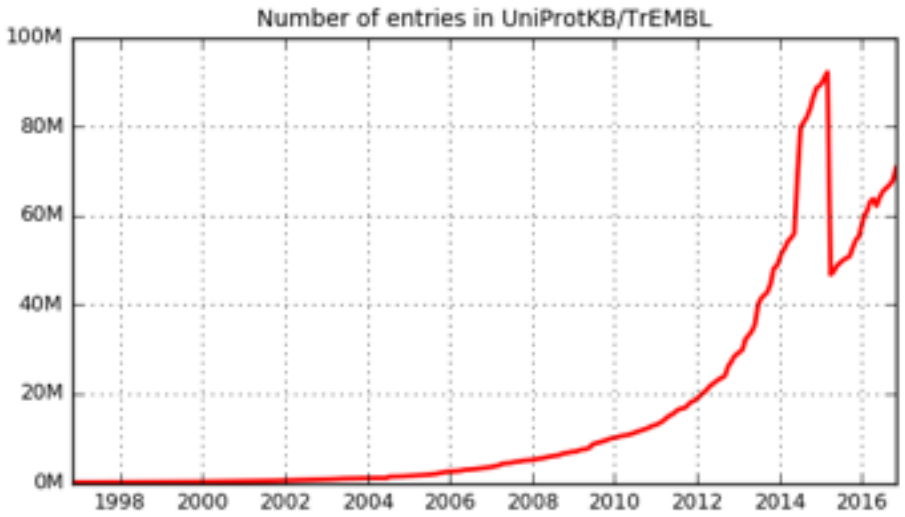


- X-Ray crystallography is crystal dependent



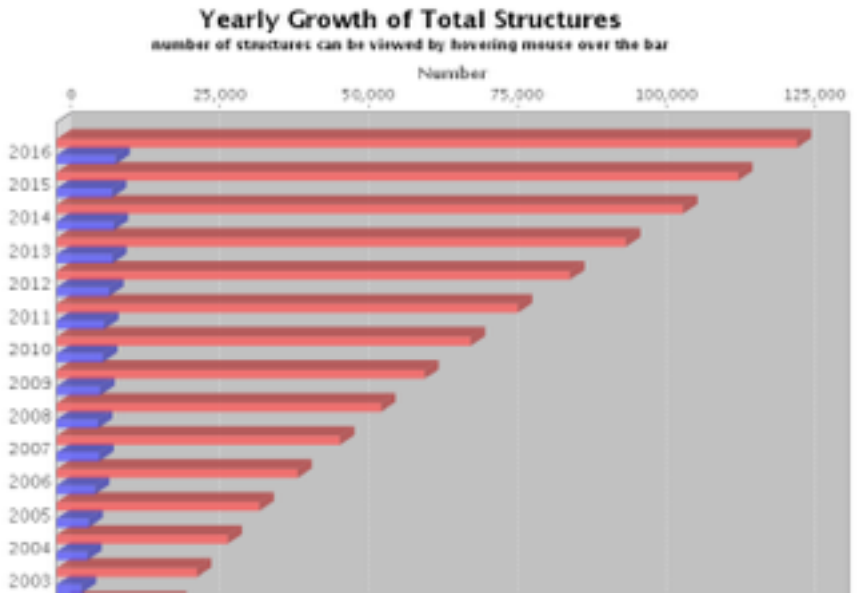


November 2016: 60 000 000 Sequences



UniProtKB/TrEMBL statistics

November 2016: 124 430 Structures



PDB statistics

Less than 1% of protein sequences have 3D structure

Why this gap is so important?

- Experimental approaches are time consuming
- They don't provide complete sets of data

In silico Modeling helps to fill up the gap

Calculation Techniques

1. Homology modeling
2. Threading modeling

} Template-based modeling

3. Ab Initio modeling → Energy-based modeling

↳ Faster and cheaper than X-ray and NMR

How do we validate the 3D structures ?

Guiding and validating structure calculation using experimental data

NMR experiment-driven modeling

4. NMR experiment-driven modeling

i. Rapid data acquisition (1 week)

(Serrano et al. J Biomol NMR 2012)

ii. Experimental NMR data provides structural information

(Mielke et al. Nucl Magn Reson Spectrosc 2009)

iii. Sparse NMR data can guides structure calculation

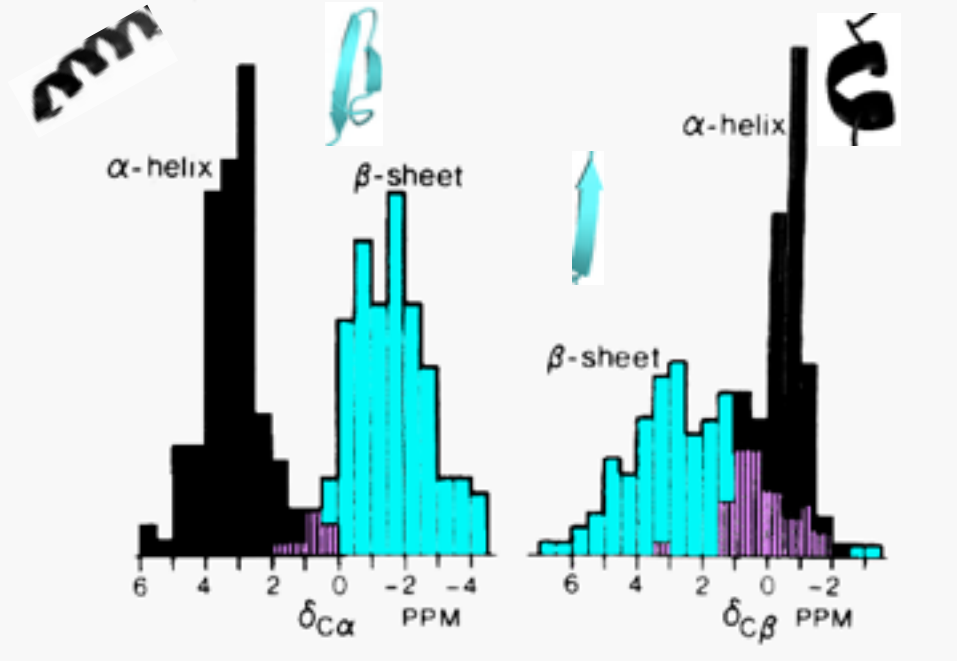
(Shen et al. PNAS 2008)

Which types of NMR data do we need?

- NMR backbone chemical shifts

Backbone chemical shifts-driven modeling

- Backbone chemical shifts are secondary structure dependent



- CS-ROSETTA structure calculation
  - Fragment-based modeling

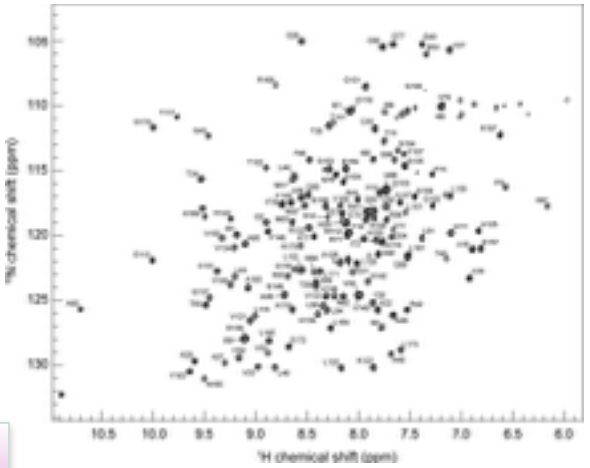
ROSETTA fragments selection

Amino acids sequence

Random split into fragments of 3 or 9 residues

For each fragment

Experimental assigned chemical shifts



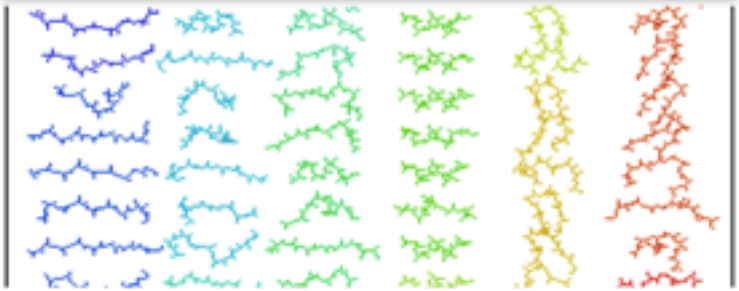
Query: sequence

Structural Database

Searching best matched fragments

Response: structures

3 and 9 amino acids selected fragments

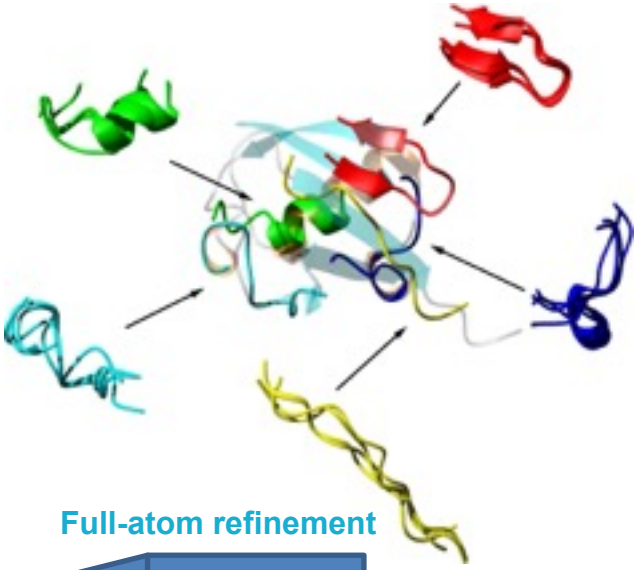


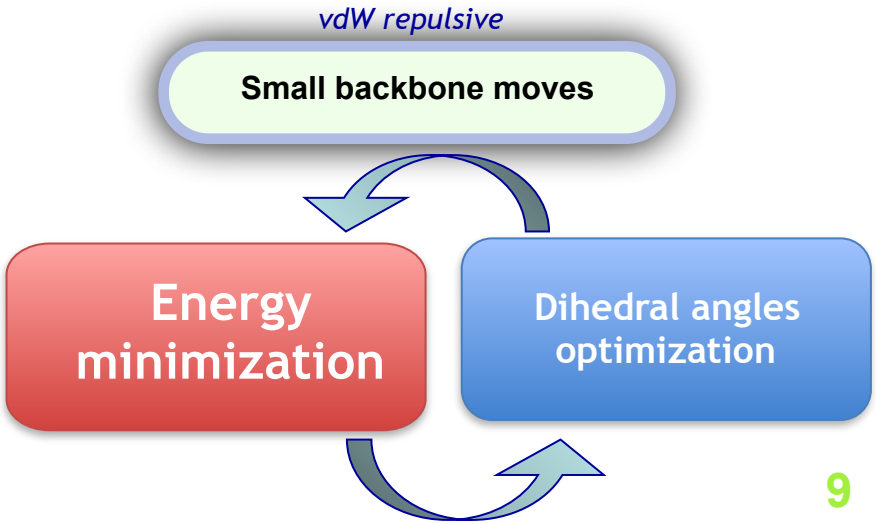
Calculation of Chemical Shifts using SPARTA

CS-ROSETTA structure calculation

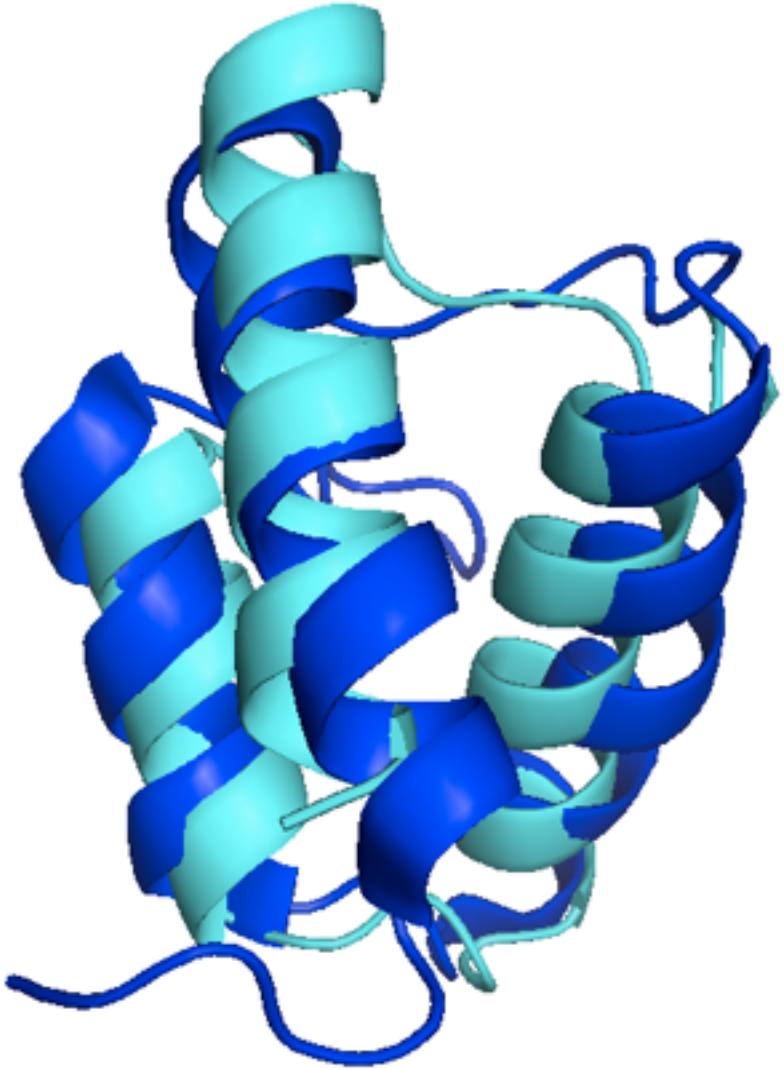
Where do the chemical shifts are used?

- During fragments assembly
- During energy re-scoration
- During refinement

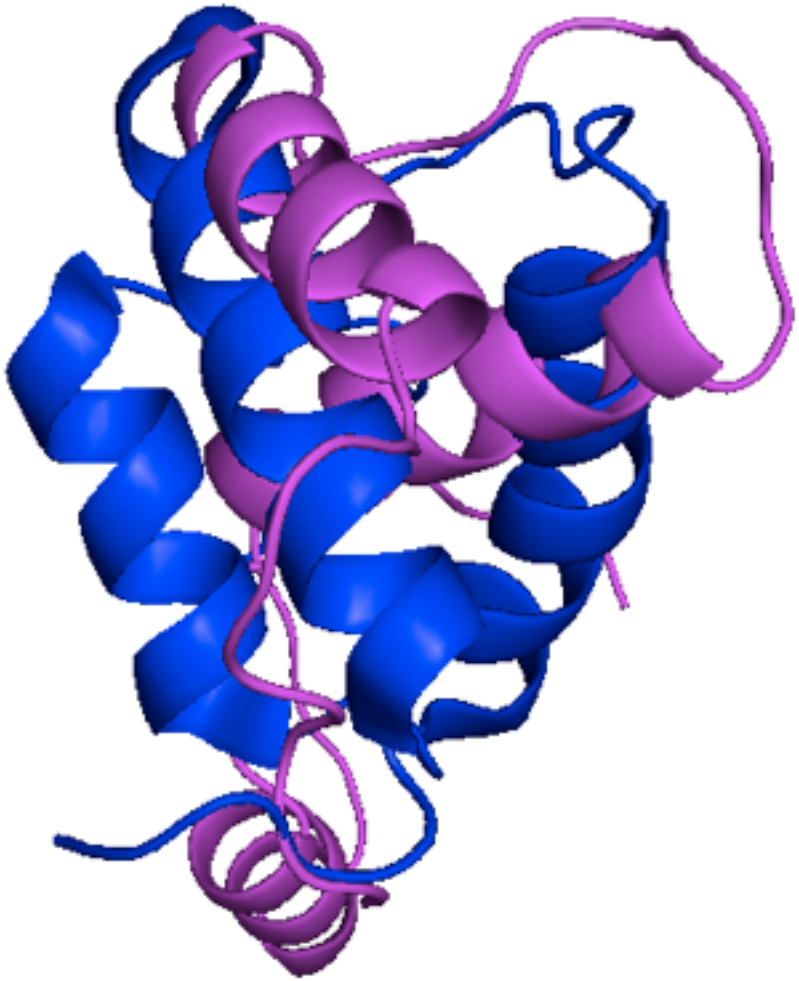




CS-ROSETTA versus Homology modeling



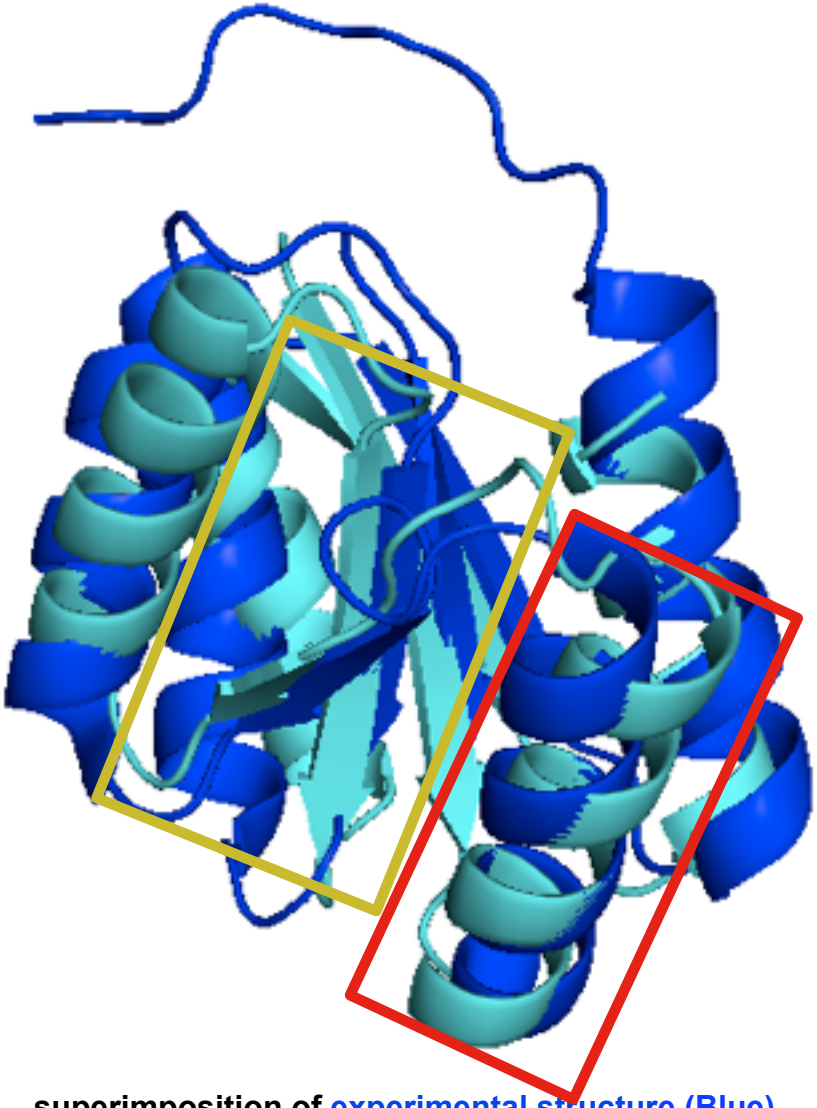
superimposition of experimental structure (Blue)
and
CS-ROSETTA structure (cyan)



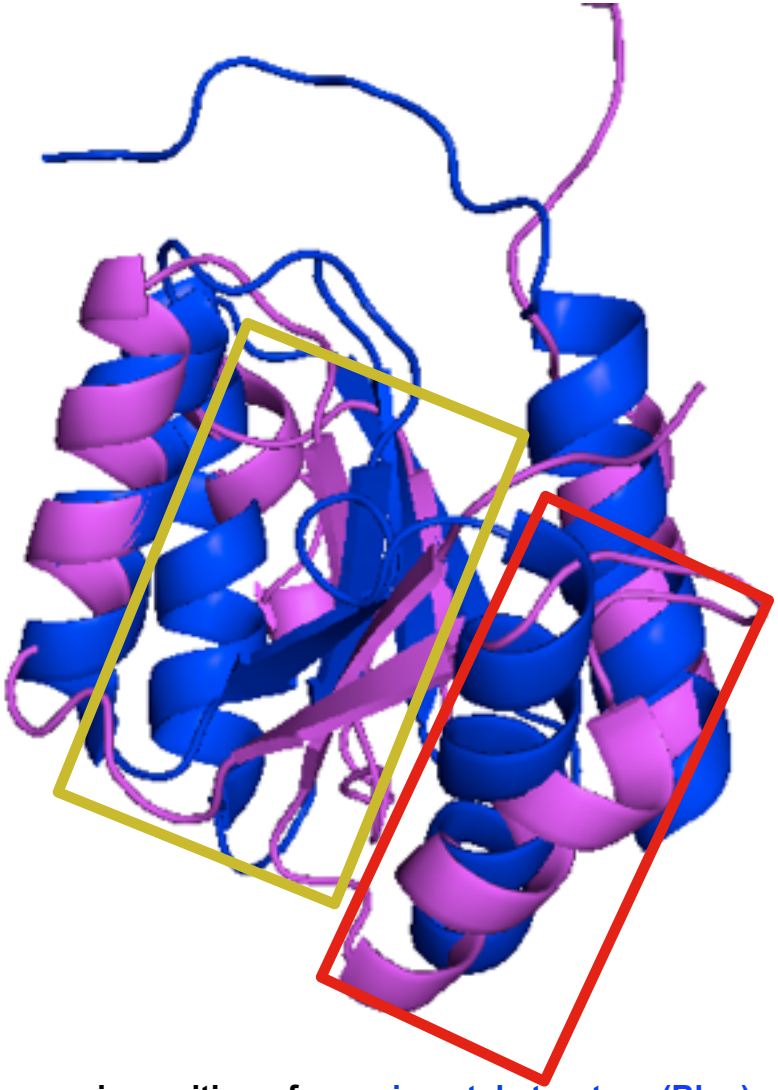
superimposition of experimental structure (Blue)
and
Modeller structure (magenta)

Protein PDB code: 2KTA. Length: 74 amino acids

CS-ROSETTA versus Homology modeling



superimposition of experimental structure (Blue)
and
CS-ROSETTA structure (cyan)



superimposition of experimental structure (Blue)
and
Modeller structure (magenta)

Protein PDB code: 2LND. Length: 112 amino acids

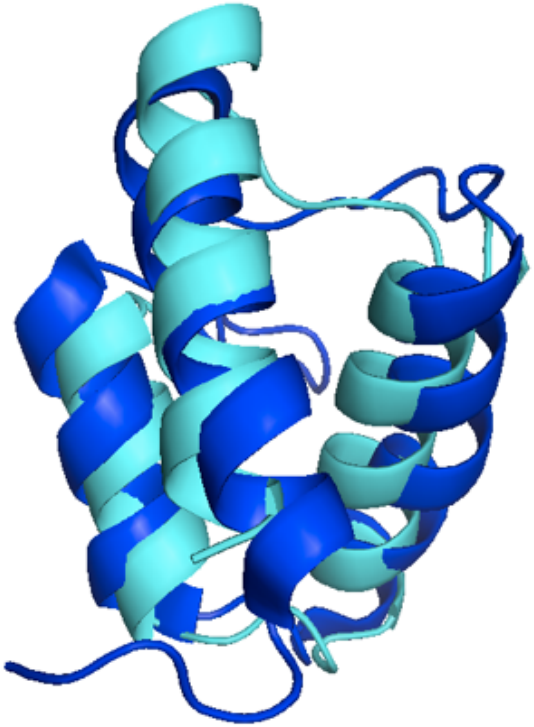
CS-HM-ROSETTA Structure Calculation

Do homologous structures play a central role?

- CS-HM-ROSETTA combines homology with Chemical shifts

- These restraints are used during fragment assembly and refinement

CS-ROSETTA versus CS-HM-ROSETTA



superimposition of experimental structure (Blue)
and
CS-ROSETTA structure (cyan)

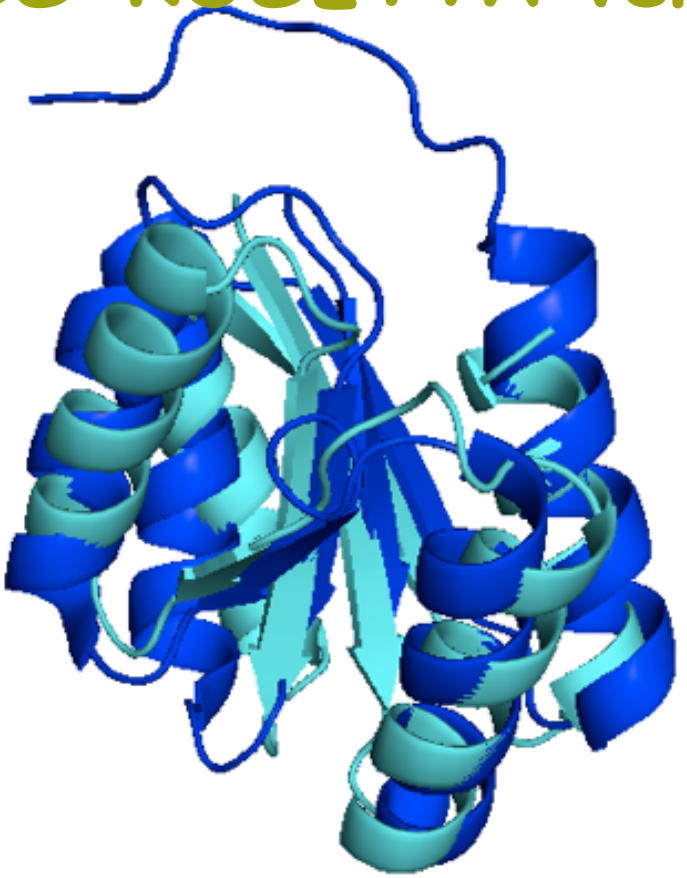
superimposition of experimental structure (Blue)
and
CS-HM-ROSETTA structure (green)

	CS-ROSETTA	CS-HM-ROSETTA
Clashscore	3.26	1.25
Cα RMSD	2.61 A	1.53 A

Clashscore is the number of serious clashes per 1 000 atoms

Protein PDB code: 2KTA. Length: 74 amino acids

CS-ROSETTA versus CS-HM-ROSETTA



superimposition of experimental structure (Blue)
and
CS-ROSETTA structure (cyan)

superimposition of experimental structure (Blue)
and
CS-HM-ROSETTA structure (green)

	CS-ROSETTA	CS-HM-ROSETTA
Clashscore	3.14	1.26
C_α RMSD	2.16 A	1.04 A

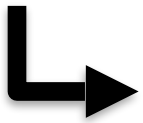
Protein PDB code: 2LND. Length: 112 amino acids

NMR-experiment-driven modeling

- Rosetta-based modeling required assigned chemical shifts as input
- Fully automated approaches have been developed recently
  - J-UNIO (Serrano et al. J Biomol NMR 2012)
  - FLYA (Güntert. Eur Biophys J 2009)
  - PONDEROSA (Lee et al. J Biomol NMR 2014)
- Most of automated approaches were tested during CASD-NMR 2013

NMR experiment-driven modeling

- CASD-NMR: Critical Assessment of Automated Structure Determination of Proteins from NMR Data
  - Assess the performance of different NMR-driven structure determination methods
- CASP: Critical Assessment of Automated Structure Prediction
  - Assess performance of protein structure prediction methods from sequence
- Most of structure calculation methods obtained pretty good structures during CASD-NMR 2013



Recommendation: Usage of more than one structure calculation methods and combine results

NMR experiment-driven modeling

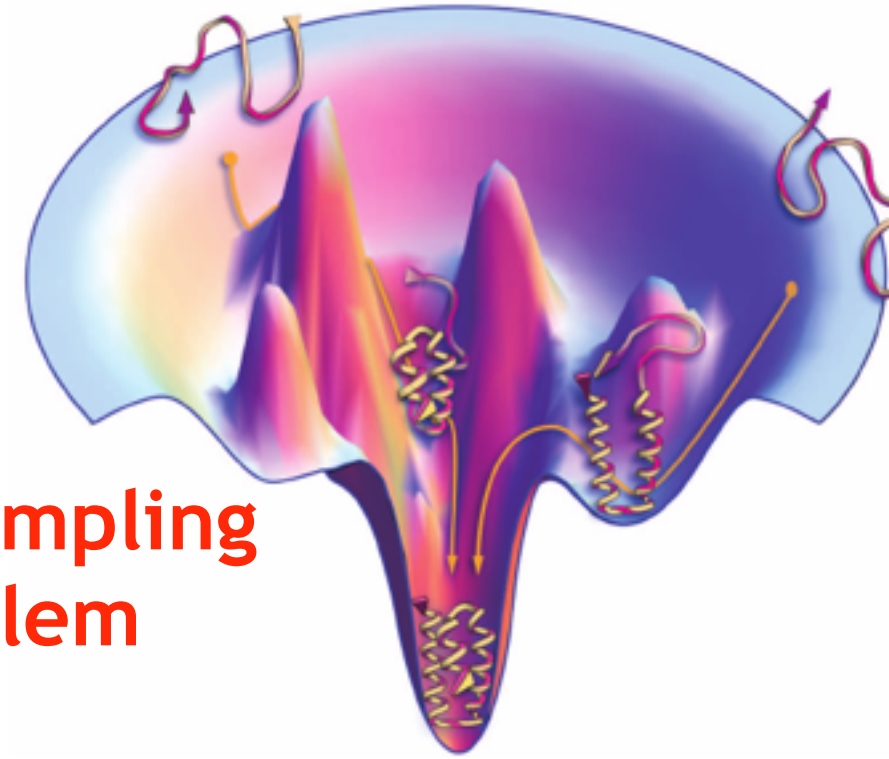
- **Our Aims:**

1. Tested as many as structure calculation approaches driven by incomplete sets NMR data

2. Development of a platform easy to use by non-specialist that allows structure calculation by different methods

Is a challenge for number of pathways to native structure increases with protein size?

- Protein size increase with pathways to native structure



→ **Conformational Sampling becomes a problem**

- **Sampling improvement is as important as incorporation of additional experimental data**

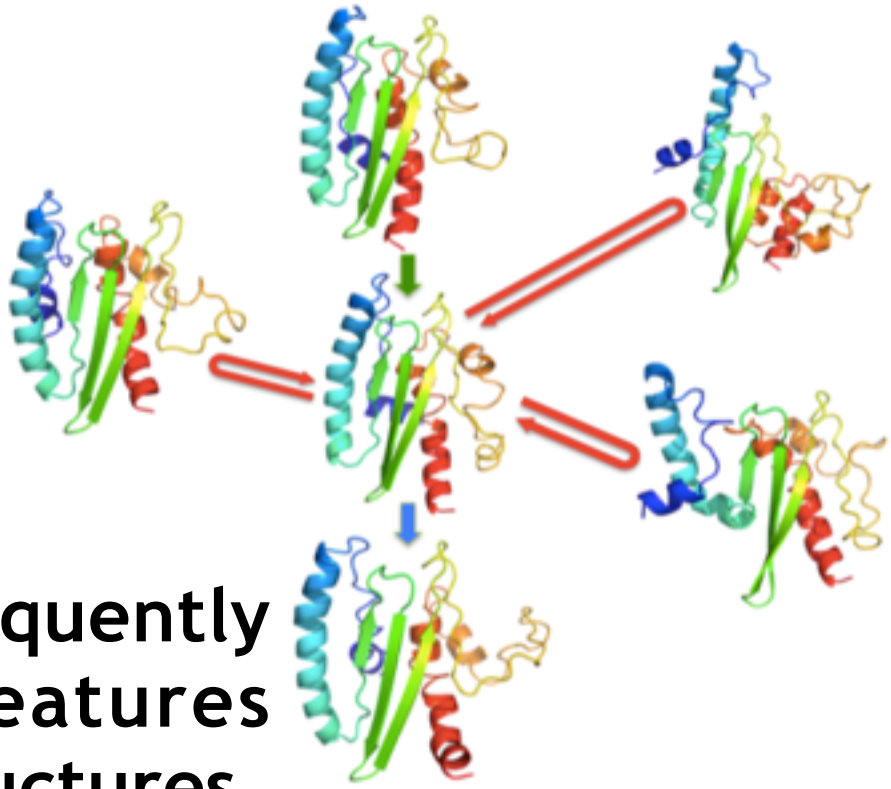
(Lange and Baker . Proteins 2012)

RASREC sampling methodology

- RASREC: Resolution Adapted Structural RECombination

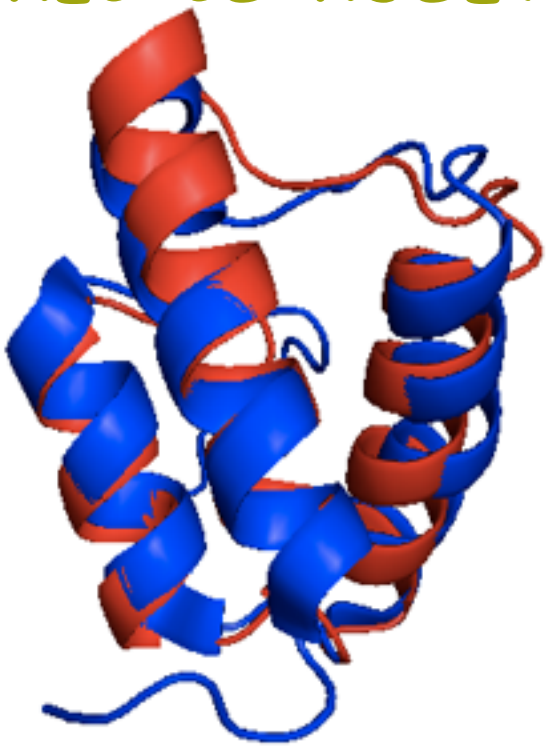
- Structural features occur frequently

- RASREC recombines frequently occurring structural features found in intermediate structures



CS-HM-ROSETTA versus RASREC-CS-ROSETTA

superimposition of experimental structure (Blue) and CS-HM-ROSETTA structure (green)

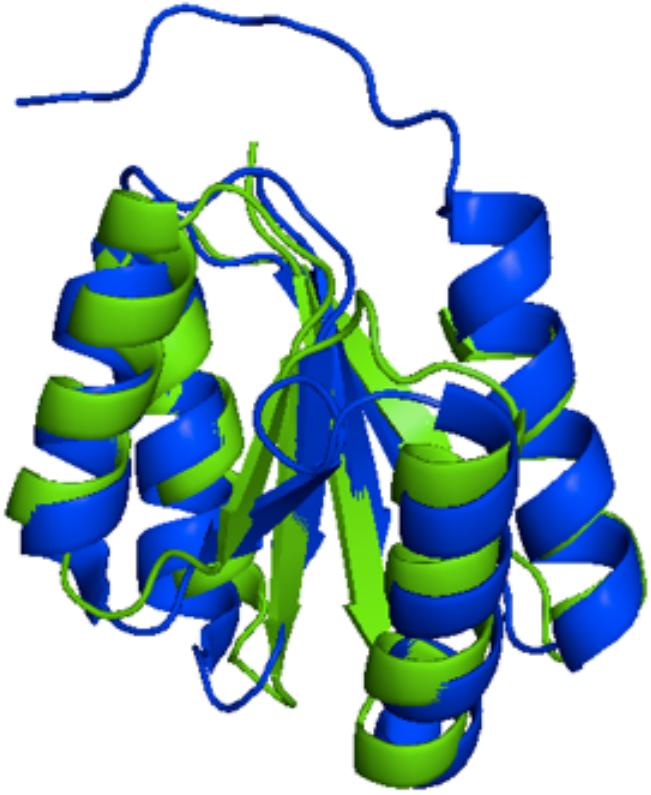


superimposition of experimental structure (Blue) and RASREC-CS-ROSETTA structure (red)

	CS-HM-ROSETTA	RASREC CS-ROSETTA
Clashscore	1.25	1.20
RMSD	1.53	0.64

Protein PDB code: 2KTA. Length: 74 amino acids

CS-HM-ROSETTA versus RASREC-CS-ROSETTA



superimposition of experimental structure (Blue)
and
CS-HM-ROSETTA structure (green)

superimposition of experimental structure (Blue)
and
RASREC-CS-ROSETTA structure (red)

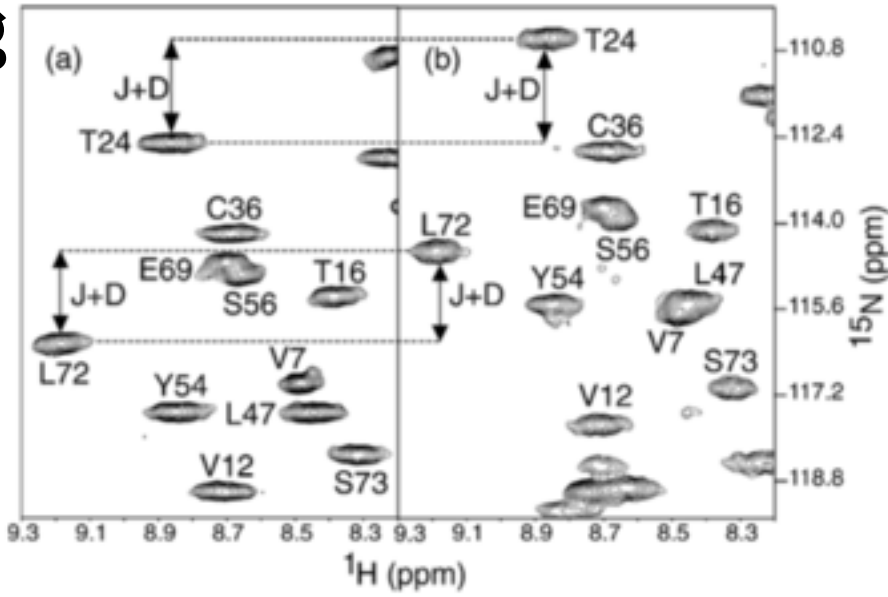
	CS-HM-ROSETTA	RASREC CS-ROSETTA
Clashscore	1.26	2.26
RMSD	1.04	1.51

Protein PDB code: 2LND. Length: 112 amino acids

Perspectives

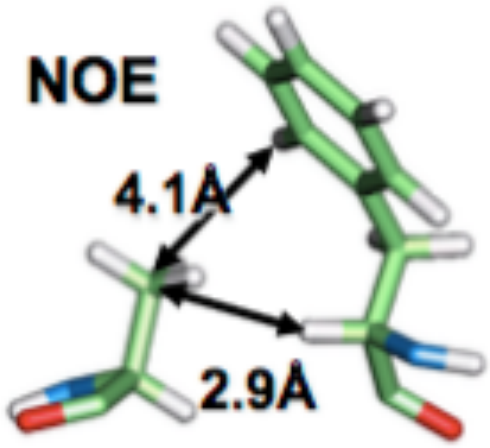
1. Residual Dipolar Coupling of backbone

- Easy to obtain
- Provide orientation information



2. Introduction of incomplete NOEs distances

- NOEs data seem very important
- 12 NOEs are necessary per atom



Can sparse NOEs distances be used?

Perspectives

- Backbone CS
- RDCs
- NOEs

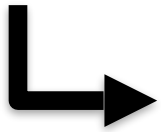
1. One u-^{15}N, u-^{13}C labeled sample
2. Data acquisition take atoms 2 weeks
   - Including assignment

- Towards automatic programs
  - J-UNIO
    - Automatic backbone chemical shifts assignment
    - Automatic side chain assignment
    - Automatic NOEs assignment by CYANA

Perspectives

- CYANA requires at least 90% of assignment
  - Backbone + side chains

- In most cases, J-UNIO doesn't reach this level and need human intervention



Combination of J-UNIO with RASREC-AutoNOE-ROSETTA

✤ AutoNOE-ROSETTA

- Incomplete NOEs
- Automatic NOEs assignment

Summary

Incomplete sets of NMR data

structure calculation methods

Rapid protein structure determination

Development of a platform that combines structure calculation methods

Acknowledgment

Christian Damblon

Andre Joachim

Gorges Feller

ULgNMRCENTRE

Université de Liège

Young Belgian Magnetic Resonance Scientist

YBMRS 2016 15th edition

YBMRS 2016

5th and 6th December 2016

THANK FOR YOUR ATTENTION