Comparison of rapid protein structure determination approaches driven by experimental NMR data

M. Wanko[1], C. Damblon[1], G. Feller[2], A. Tchernychev, V. Volkov

[1] *Structural Biological Chemistry Unit, Department of chemistry, University of Liege, Allée du 6 Août 13, 4000 Liège, Belgium*

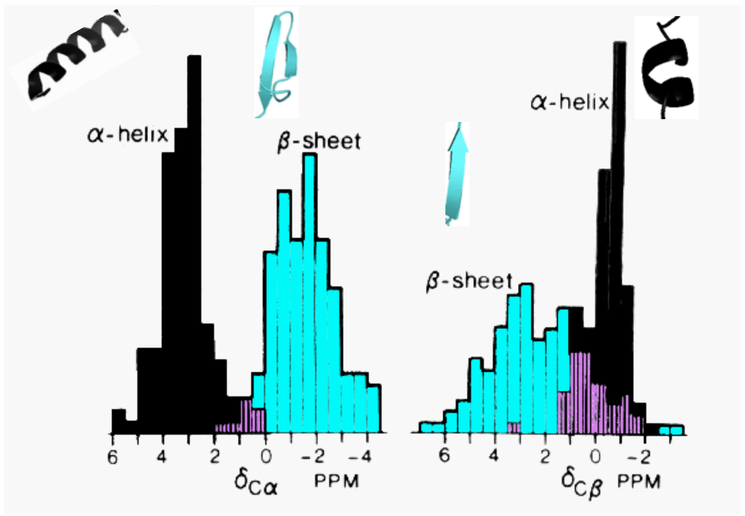
[2] *Laboratory of Biochemistry, Centre for Protein Engineering (CIP), University of Liège, Allée du 6 Août 13, 4000 Liège, Belgium*

e-mail: amwanko@doct.ulg.ac.be

Introduction

Tridimensional structures of proteins are precious sources of information. They allow to understand fundamental biological mechanisms, protein Interaction with other macromolecules. 3D structures are currently determines using NMR and

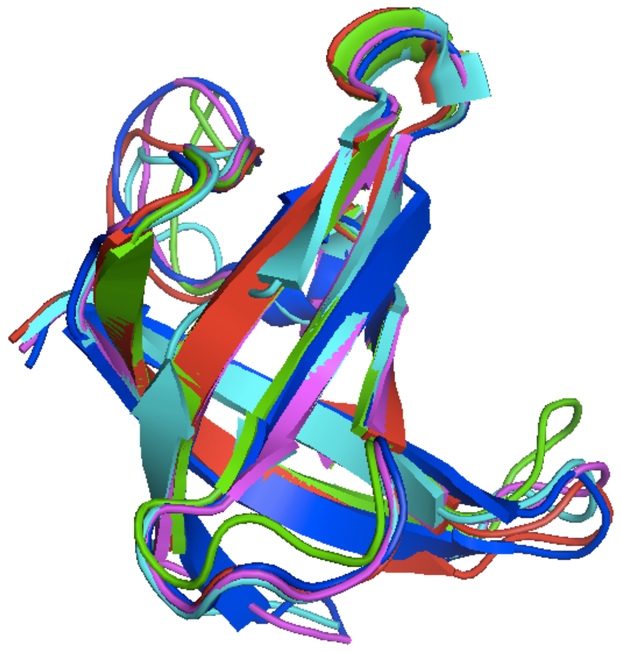
X-Ray Crystallography that faced to some limitations. In order to overcome time consuming limitation of both NMR and Crystallography, modeling approaches driven by minimalist NMR data have been developed. Indeed, it have been shown that NMR backbone chemical shifts are secondary structure dependent.



Backbone chemical shifts is structure dependent

Therefore, different modeling approches driven by only NMR backbone chemical shifts such as CS-Rosetta, RASREC CS-Rosetta, CS-HM-Rosetta CS23D and Cheshire have been developed. To assess whether if these automated methods can indeed produce structures that closely match those manually refined by experts using the same experimental data, these approaches were used to determine 3D structure of a benchmark of proteins.

Application to Cold Shock Protein



Blue: experimental NMR structure
Red: RASREC CS-Rosetta structure
Green: CS-Rosetta structure
Cyan: Modeller structure
Magenta: Calculated structures

Calculated 3D structures driven by NMR backbone chemical are close to NMR experimental structure

It is also true for homology modeling structures.

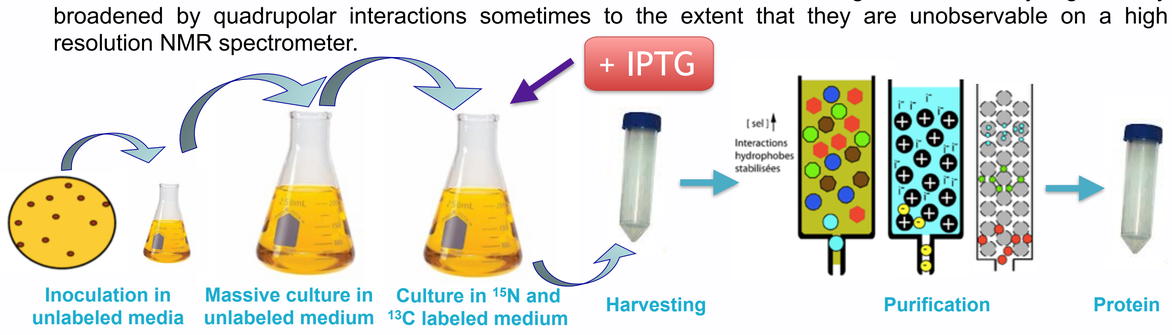
While homology modeling approaches didn't use experimental data, 3D structures determined by calculation approaches under the guidance of NMR data can be used to validate homology structure.

Despite the fact that both homology and calculates approches provides 3D structures close to experimental structure, somme regions were ill-defined

It is because modeling approaches aren't powerful enough or it is due to dynamic?

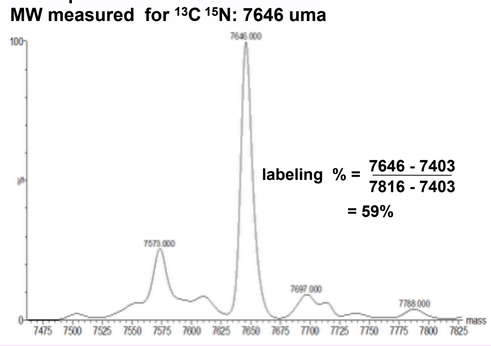
Protein expression and isotopic labeling

During protein expression, E. coli bacteria were transformed and grown in a unlabeled Terrific Broth media After biomass production, cells were starved in M9 medium containing $^{15}NH_4Cl$ and ^{13}C-6 D-Glucose as the only source of carbon and nitrogen. ^{15}N and ^{13}C labeling are needed because ^{12}C which is the most abundant source of carbon, has an even atomic number and even mass so therefore this nuclei is NMR inactive while ^{13}C has an odd mass and is therefore NMR active. ^{14}N signals are usually significantly broadened by quadrupolar interactions sometimes to the extent that they are unobservable on a high resolution NMR spectrometer.

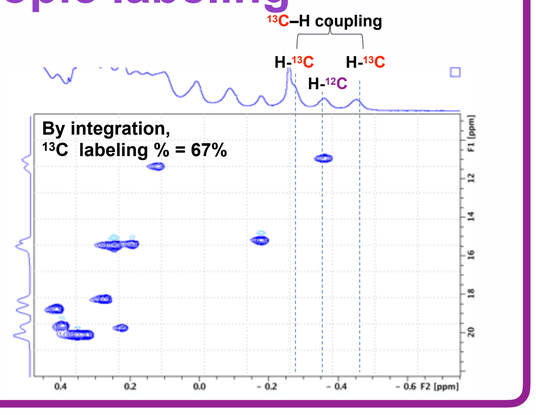


Assessment of isotopic labeling

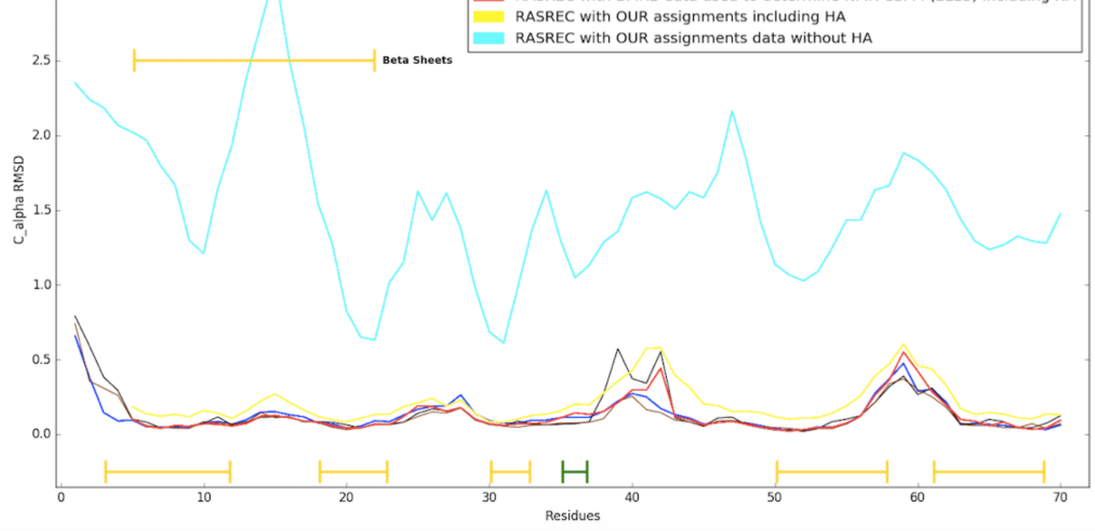
MW expected for ^{12}C ^{14}N: 7403 uma
MW expected for ^{13}C ^{15}N: 7816 uma
MW measured for ^{13}C ^{15}N: 7646 uma



$$\text{labeling \%} = \frac{7646 - 7403}{7816 - 7403} = 59\%$$



Comparison based on Cα-RMSD

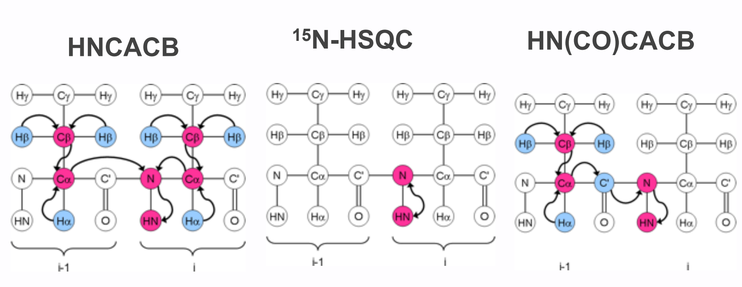


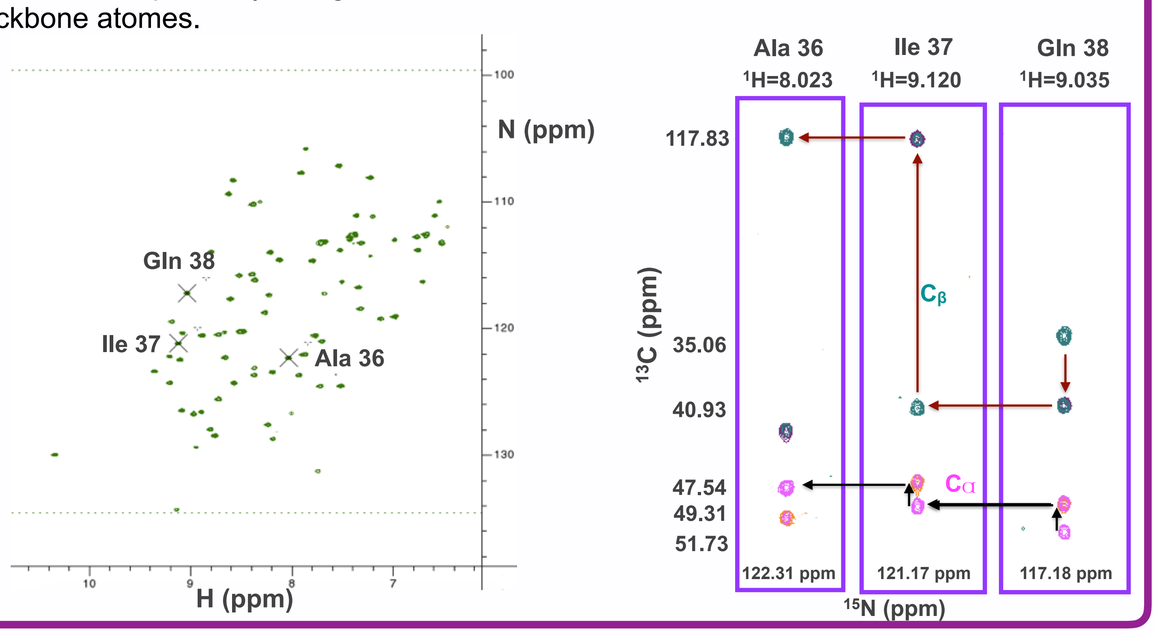
	Well-defined regions ranges	RMSD(Å)	AA/AA$_t$	% of proteins coverage	Gap
CSPM from PDB	4..55, 63..70	0.55	60/70	85 %	1
Rasrec-CS-Rosetta	5..70	1.58	66/70	94 %	0
CS-Rosetta	1..32, 41..66	1.21	58/70	83 %	1
Rasrec-CS-Rosetta sans HA	4..38	1.27	35/70	35 %	0

- AA: Number of amino acids of Well-defined regions
- AA$_t$: Protein's number of amino acids
- % of proteins coverage = (AA÷AA$_t$) * 100

NMR Backbone Chemical shifts assignment

Starting from ^{15}N-HSQC, heteronuclear correlation ^{1}H-^{15}N is used to identifie Cα and Cβ atoms of amino acid in position i and Cα and Cβ atoms of amino acid in position i-1. Resulting chemical shifts can be therefore ordered to sequentially assigned backbone atomes.





Conclusions

Structure calculation driven by NMR backbone chemical is an efficient alternative for 3D structure determination while complete set of expérimental data cannot be obtain for experimental approaches

These approaches are rapidly, provided good results in term of fitness to experimental structure

Perspectives

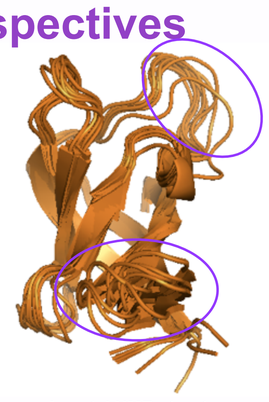
Some regions are ill-defined

Is it due to modeling programs?

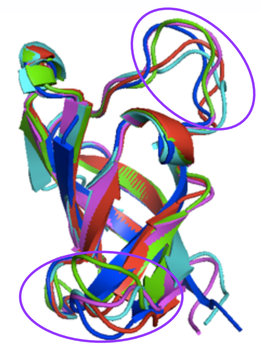
Is it because these regions are dynamic?

Dynamic study using NMR will be used to answer these questions

To improve labeling, starvation time should be increase



20 experimental NMR structures



Blue: experimental NMR structure
Others: Calculated structures

References

A. Rosato, A. Bagaria, D. Baker, B. Bardiaux, A. Cavalli, J. Doreleijers, A. Giachetti, P. Guerry, P. Güntert, T. Herrmann, Y. Huang, H. Jonker, B. Mao, T. Malliavin, G. Montelione, M. Nilges, S. Raman, G. van der Schot, W. Vranken, G. Vuister and A. Bonvin, Nature Methods, 2009, **6**, 625 - 626

A. Rosato, W. Vranken, H. Fogh, J. Ragan, R. Tejero, K. Pederson, H. Lee, H. Prestegard, A. Yee, B. Wu, A. Lemak, S. Houston, H. Arrowsmith, M. Kennedy, B. Acton, R. Xiao, G. Liu, T. Montelione, W. Vuister, J Biomol NMR, 2015, **62**, 413–424

Y. Shen, O. Lange, F. Delaglio, P. Rossi, M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. Singarapu, A. Lemak, A. Ignatchenko, H. Arrowsmith, T. Szyperski, T. Montelione, D. Baker, and A. Bax, PNAS, **105**, 2008, 4685–4690

M. Thompson, G. Sgourakisa, G. Liub, P. Rossib, Y. Tangb, L. Millsc, T. Szyperskic, T. Montelione, and D. Baker, PNAS, 2012, **109**, 9875–9880

Acknowledgements

ULgNMRCENTRE
CIP CENTER FOR PROTEIN ENGINEERING
Université de Liège