Protein structure modeling using backbone chemical shifts

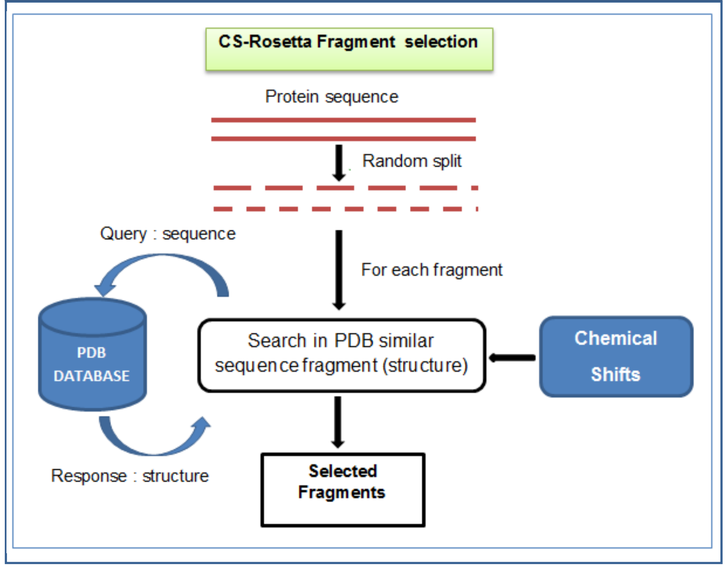
Alexis Marius Wanko[1], Christian Damblon[1]

[1] University of Liege, Department of Chemistry, Liege, 4000, Belgium

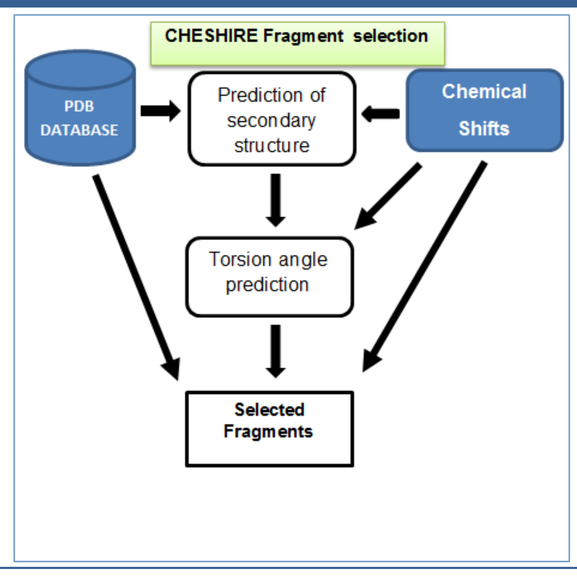
Introduction

The knowledge of the tridimensional structure of a protein is essential to study its interactions and understand its mode of action. In order to determine the three-dimensional structure of proteins, several NMR data can be used (Backbone chemical shifts (CS), NOE distances, residual dipolar coupling, …). Among these parameters, backbone chemical shifts data are NMR parameters that can be rapidly, easily and accurately measured. This parameter is very sensitive to the conformation of amino acids and is used to deduct the secondary structure (TALOS, RCI,...). Therefore, the backbone chemical shifts can be used as constraints on dihedral angles to quickly and easily determine protein structure. Several « de novo » methods like CS-Rosetta, CS23D and CHESHIRE have been recently developed in this purpose.

CS-Rosetta Vs CS23D Vs CHESHIRE



CS-Rosetta Vs CS23D Vs CHESHIRE



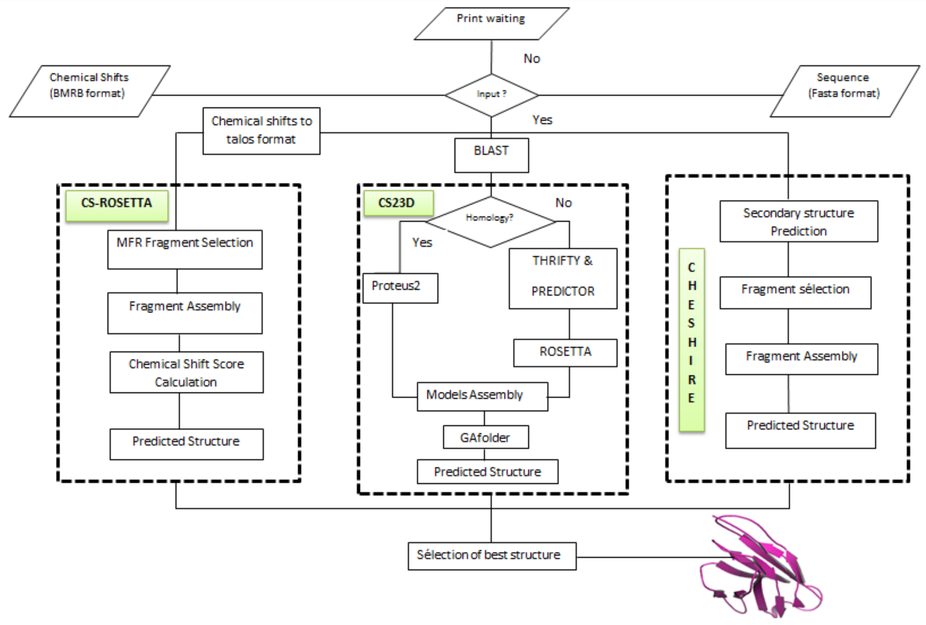
Main purpose

The Purpose of our work is to quickly and easily determine a method to obtain structures of proteins using the backbone chemical shifts.

To achieve our goal, we will:

- Test CS-Rosetta, CS23D and CHESHIRE software using proteins of different sizes for which, the structure (X-ray or NMR structure) and chemical shifts backbone are available
- Develop a platform that can easily compare these three methods based on quality of the structure produced

Method



CS-Rosetta Vs CS23D Vs CHESHIRE

	CS-Rosetta	CS23D	CHESHIRE
Input data	Chemical Shifts (TALOS Format) + Sequence	Chemical Shifts (SHIFTY or BMRB Format) + or without Sequence	Chemical Shifts (Cheshire format) + Sequence
Low resolution Structure generation protocol	Monte Carlo algorithm	GAfolder	Monte Carlo algorithm
Deployment	Platform dependent configuration	web server	Platform dependent configuration
Usage	difficult for non-expert (command line usage)	easy (graphic user interface)	difficult for non-expert (command line usage)
Time calculation	2h - 72h	10 - 15 minutes	~ 1h
Robustness	Tolerate missing backbone CS	Required completeness of backbone CS	Required completeness of backbone CS

Criteria of selection

One of the fundamentals steps of our approach is the selection of the best structure among those obtained. Based on what criteria could we say that a structure is **better** than an other?

Generally, two aspects are taken into consideration during structure validation: agreement of resulting structure with experimental data and geometric validation.

Three criteria are used:

- **Restraint violation**: different measures are available to establish the accordance between structure models and experimental data (number of restraints, RMS violations, …).
- **Accuracy** of the predicted structure measured by positional root mean square deviation (RMSD)
- **Geometric quality evaluation**: different programs are available for NMR-derived structure validation such as PSVS which uses the Assignment Validation Suite (AVS) (Moseley et al. 2004) to identify outliers, SHIFTX (Neal et al. 2003) for back-calculated values and CING which uses VASCO routine (Rieping and Vranken 2010) for referencing correction

Conclusions

The combination of the three protocols should allow us to determine the structure of a large number of proteins. Additionally, this platform is designed to reduce as much as possible user intervention during the generation process of the structure.

Usage of only chemical shifts and sequence as input data should allow us to use current, recent arsenal of structure generation and refinement tools available to generate 3D structure

Unlike the previous softwares, our platform does not generate a set of structures, but aim to predict the most probable structure which fit with experimental data.

Despite the fact that selection process of the best structure may be the source of a bias, we will devoted special attention on the validation criteria of the quality of a structure:

- validation of restraint violation,
- the accuracy of predicted structures,
- geometric quality of predicted structures

to minimize errors when selecting the best structure.

Contact

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

References

1. Lisa R. *Structure of the bamc two-domain protein obtained by rosetta with a limited nmr data set*. 2011.
2. James M. Thompsona, Nikolaos G. Sgourakisa, Gaohua Liub, Paolo Rossib, Yuefeng Tangb, Jeffrey L. Millsc, Thomas Szyperskic, Gaetano T. Montelioneb, and David Bakera. Accurate protein structure modeling using sparse NMR data and homologous structure information. PNAS 2012
3. David S. Wishart, David Arndt, Mark Berjanskii, Peter Tang, Jianjun Zhou1 and Guohui Lin1 CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Research, 30 May 2008, Vol. 36.
4. Berjanskii,M.V., Neal,S. and Wishart,D.S. (2006) *PREDITOR:a web server for predicting protein torsion angle restraints*. Nucleic Acids Res.
5. Chris A.E.M. Spronka, Sander B. Nabuursa, Elmar Kriegera, Gert Vrienda, Geerten W. Vuisterb. Validation of protein structures derived by NMR spectroscopy. Nuclear Magnetic Resonance Spectroscopy 45 (2004) 315–337.
6. Mark Berjanskii, Peter Tang, Jack Liang, Joseph A. Cruz, Jianjun Zhou,Edward Bassett, Cam MacDonell, Paul Lu, Guohui Lin, and David S. Wishart. *GeNMR: a web server for rapid NMR-based protein structure determination*. Nucleic Acids Res, 2009.
7. Benjamin Bardiaux. Structure calculation of proteins from solution and solid-state NMR data : Application to monomers and symmetric aggregates. PhD thesis. January, 2009
8. Geerten W. Vuister , Rasmus H. Fogh, Pieter M. S. Hendrickx, Jurgen F. Doreleijers and Aleksandras Gutmanas. An overview of tools for the validation of protein NMR structures. J Biomol NMR (2014) 58:259–285
9. G. Fuentes, A.D.J. van Dijk, and A.M.J.J. Bonvin Nuclear Magnetic Resonance-Based Modeling and Refinement of Protein Three-Dimensional Structures and Their Complexes. Methods in Molecular Biology, vol. 443, Molecular Modelling of Proteins