

## Computational challenges in rapid characterization of the protein-protein interactions by Nuclear Magnetic Resonance (NMR)-based methods

Andrey GORIN

*Computer Science and Mathematics Division,  
Oak Ridge National Laboratory,  
Oak Ridge, TN, 37830 [agor@ornl.gov](mailto:agor@ornl.gov)*

Hashim M. AL-HASHIMI

Department of Chemistry, University of Michigan,  
Ann Arbor, MI, 48109, [hashimi@umich.edu](mailto:hashimi@umich.edu)

One of the greatest problems of modern biology is the characterization of potentially insurmountable number of interactions that underlie gene product activity. Assembling a molecular catalogue of proteome function that carries predictive powers requires not only the identification of binding partners, but measurements of binding affinities and specificities along with the structural characterization of protein complexes. Unlike pure structure determination, such functional investigations would ideally be carried out under physiologically relevant solution conditions, severely limiting the number of techniques that can be employed to realize this task.

The limiting factors for using solution NMR in rapid measurements and characterization of protein-protein interactions can be summarized as: (1) assignment procedures needed for characterization of residues at protein interfaces are laborious and time-consuming, and (2) quantitative measurements of binding constants can be limited by overlap of resonances and non-ideal NMR exchange rates. Current approaches for achieving backbone resonance assignments in proteins rely on correlating various nuclei on the backbone ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^1\text{H}$ ) using a variety of triple resonances 3D and 4D experiments (Bax, 1994). Although great progress has been made in automating this assignment approach at the point of data analysis (Montelione et al., 2000), resonance assignments can still, under favorable conditions, consume more than one month of data acquisition and analysis time.

---

The submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

Here we describe our proposed approach for residual dipolar coupling (RDC)-based rapid backbone assignment using Combinatorial Assignment Procedure (CAP), which we previously proposed and successfully implemented for the automatic RDC sequential assignments resonances in RNA (Al-Hashimi et al., 2002). The development of the CAP technology for the proteins poses distinct challenges and difficulties that are related to the size of the systems under study and computational efficiency.

The CAP assignment method works in the following way. RDCs are first measured on a target protein molecule. This leads to the assignment of chemical shift values with a corresponding RDC value. For a given assignment guess, the RDC values will either agree or disagree with the given structure. For the ideal case, the best agreement, measured as the root mean square deviation between measured and ‘best-fit’ calculated RDCs, will occur for the correct set of assignments. All possible RDC assignment permutations can therefore be explored for consistency with a given target structure.

For a proof-of principal we have implemented CAP algorithms for the sequence assignment on a small 76 amino acid protein - ubiquitin. We conducted multiple rounds of the automatic assignments on this protein using both simulated RDC data sets, generated from the crystal structure, and experimental ones, available for two different alignment media from Ad Bax’s laboratory. Several ideas for the acceleration of CAP are under further investigation (one of them may result in 2 order of magnitude acceleration), but even at the current speed many of the proposed methods are feasible for the proteins of 100-200 amino acids.

Finally, we briefly outline methods for accurate measurements of binding constants and for the identification of interacting residues in protein-protein complexes and discuss perspectives and challenges of our “computational NMR methodology” for rapid characterization of the cellular proteomes.

Reference:

1. Al-Hashimi, H.M., Gorin, A., Majumdar, A., Gosser, Y. and Patel, D.J., (2002) *J Mol Biol*, **318**: 637-649.
2. Bax, A., (1994) *Current Opinion in Structural Biology*, **4**: 738-744.
3. Montelione, G.T., Zheng, D., Huang, Y.J., Gunsalus, K.C. and Szyperski, T., (2000) *Nat Struct Biol*, **7** Suppl, 982-985.