Center for Eukaryotic Structural Genomics
Protein Structure Initiative

National Magnetic Resonance Facility at Madison
NMRFAM

BMRB

WORLDWIDE PDB
PROTEIN DATA BANK

Advances in high-throughput NMR spectroscopy of proteins at CESG, NMRFAM, and BMRB

Bahrami A, Lee W, Dai Z, Tonelli M, Cornilescu G, Chylla RA, Wedell JR, Maziuk D, Ulrich EL, Sedighi M,* Eghbalnia HR,* Livny M,† Markley JL

Center for Eukaryotic Structural Genomics (CESG), National Magnetic Resonance Facility at Madison (NMRFAM), and BioMagResBank (BMRB), Biochemistry Dept. and †Computer Sciences Dept., Univ. of Wisconsin, Madison WI 53706, USA, and *Dept. of Molecular and Cellular Physiology, Univ. of Cincinnati, Cincinnati OH 45267

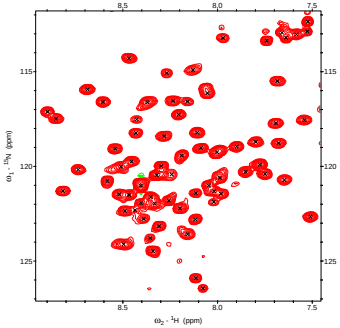
Abstract

Continued development of our probabilistic approach to NMR data collection has led to the combination of adaptive, reduced dimensionality data collection (HIFI-NMR) with automated assignment (PINE). The new approach is called HIFI-PINE. In response to requests from PINE users, we have added additional experiments to those supported. More robust and efficient methods for spectral data collection and signal recognition are being implemented. Data collection and analysis are supported on both Varian and Bruker spectrometers. Intermediate and final results can be analyzed graphically, respectively, with HIFI-Enhancer and PINE-SPARKY. Final results are submitted to the TARA software tool, which generates probabilistic torsion angle restraints for entry into structure determination programs. A high-throughput server (powered by Condor) for CS-Rosetta is now available from the BMRB website.

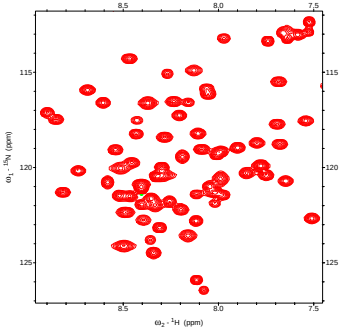
Supported by NIGMS Protein Structure Initiative grants P50 GM074901 and 1U54 GM064598, NCRR grant P41 RR02301 and NLM grant P41 LM05799.

Newton

Newton approaches signal recognition and parameter estimation by model fitting. As an augmentation/replacement for "peak-picking", Newton builds an *n*-D spectrum to match FT-processed data and thus seeks to eliminate manual filtering of peak lists.



DATA: Expanded region of the experimental ^{15}N-HSQC spectrum of a 118-residue protein. The X's denote signals detected by Newton in the model shown below.

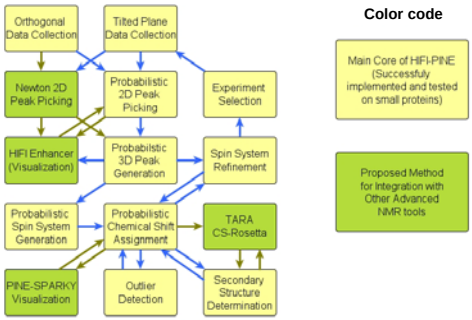


MODEL: Newton analyzed the experimental time-domain data to extract amplitudes, frequencies (equivalent to peak picking), and line widths for detected peaks. The spectrum calculated from the model is shown (note the absence of noise).

HIFI-PINE

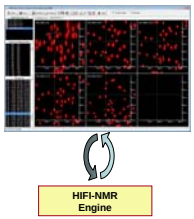
HIFI-NMR (fast data collection) and PINE (automated chemical shift assignment and secondary structure determination) are two of the major methods that have substantially reduced the data collection and analysis time in protein NMR. HIFI-PINE aims to integrate automated and adaptive data collection (HIFI-NMR) with automated data analysis (HIFI-PINE). The iterative approach combined with feedback from the assignment module ensure the acquisition of adequate NMR signal / noise and high quality chemical shift information (peak lists); these, in turn, yield superior backbone and sidechain resonance assignments, and secondary structure prediction in less time. The correlation of associated peaks across spectra, along with the HIFI-Enhancer visualization tool, alleviates the need for post-processing in the fast data collection approach. The main core of HIFI-PINE (yellow boxes) has been successfully tested on small proteins. The final integrated tool (with incorporation of the green boxes) will support fully automated protein structure determination from chemical shifts data alone.

Probabilistic network implemented in HIFI-PINE



HIFI-Enhancer

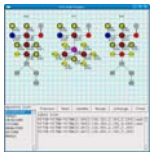
Fully automated structure determination is one of the main goals of structural genomics. However, in practice some proteins are not amenable to fully automated approaches. For these difficult proteins, manual intervention is inevitable. The HIFI-Enhancer interface presents the automated picked peaks in each plane, links all peaks in a series of tilted planes (along with the orthogonal planes) to the spin systems, and enables manual correction of peak picking and spin system assembly.



PINE-SPARKY

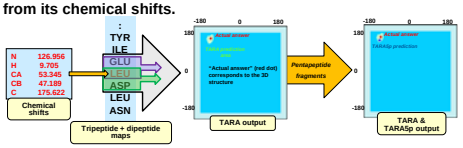
PINE-SPARKY is a powerful tool for visualizing and editing the probabilistic assignments provided by PINE. In PINE-SPARKY, multiple possible assignments for a peak are displayed next to the peak, color coded by their probabilities. The user can then select which alternative is correct and remove the incorrect labels.

"PINE Graph Assigner" is a graphical diagram window in PINE-SPARKY that displays all the atoms and their probabilistic assignments from the selected residue as well as the atoms in the adjacent residues. One can easily assign or unassign one or multiple peaks in the interface.



TARA

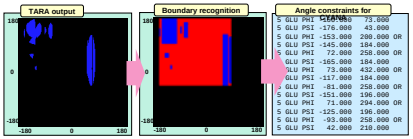
TARA predicts an area in ϕ, ψ (Ramachandran) space for each residue from its chemical shifts.



Then as the final step *TARA5p* provides a more refined angle prediction.

TARA-CYANA

TARA output (including multiple ambiguous torsion angle constraints) can used as input to CYANA for better structure calculation through our TARA-CYANA interface.



Boundary recognition derived from our *combining boxes* algorithm is a critical part of TARA-CYANA.

CS-Rosetta server

The addition of a publicly accessible CS-Rosetta server now allows for the deposition of chemical shift data, which are then processed through the CS-Rosetta software suite using the Condor distributed computing architecture at the BMRB (University of Wisconsin Madison). When decoy generation is complete the depositor is provided with a graph of the RMSD vs. respective energy of the decoys that were generated. This allows the author to determine if the results are meaningful. They are also provided the option to download the rest of the data generated by the CS-Rosetta run. This service is provided in the interest of improving the availability of tools for validating structures and for *ab initio* structure generation.

References

Eghbalnia, H. R., Bahrami, A., Tonelli, M., Hallenga, K. & Markley, J. L. (2005) High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection, *J Am Chem Soc. 127*, 12528-36. NIHMSID: NIHMS6770

Eghbalnia, H. R., Bahrami, A., Wang, L., Assadi, A. & Markley, J. L. (2005) Probabilistic Identification of Spin Systems and their Assignments including Coil-Helix Inference as Output (PISTACHIO), *J Biomol NMR. 32*, 219-33.

Eghbalnia, H. R., Wang, L., Bahrami, A., Assadi, A. & Markley, J. L. (2005) Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements, *J Biomol NMR. 32*, 71-81.

Wang, L., Eghbalnia, H. R., Bahrami, A. & Markley, J. L. (2005) Linear analysis of carbon-13 chemical shift differences and its application to the detection and correction of errors in referencing and spin system identifications, *J Biomol NMR. 32*, 13-22.

Wang, L., Eghbalnia, H. R. & Markley, J. L. (2006) Probabilistic Approach to Determining Unbiased Random-coil Carbon-13 Chemical Shift Values from the Protein Chemical Shift Database, *J Biomol NMR. 35*, 155-165.

Wang, L., Eghbalnia, H. R. & Markley, J. L. (2007) Nearest-neighbor effects on backbone alpha and beta carbon chemical shifts in proteins, *J Biomol NMR. 39*, 247-57.

Bahrami, A., Assadi, A. H., Markley, J. L. & Eghbalnia, H. R. (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy, *PLoS Comput Biol. 5*, e1000307. PMCID: PMC2645676

Lee, W., Westler, W. M., Bahrami, A., Eghbalnia, H. R. & Markley, J. L. (2009) PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy, *Bioinformatics. 25*, 2085-7. PMCID: PMC2645676