

Nuclear Magnetic Resonance (NMR) [PRINT](#)

| [BMRB](#) | [MCW NMR](#) | [NMRFAM](#) | [Pine Extranet](#) |

The NMR SPECTROSCOPY Team, headed by John Markley, PhD, is responsible for NMR condition optimization, data collection and processing, and structure determination. This group takes advantage of the facilities available at the [National Magnetic Resonance Facility at Madison \(NMRFAM\)](#) and the [Medical College of Wisconsin \(MCW\)](#) and has solved over 40 NMR structures.

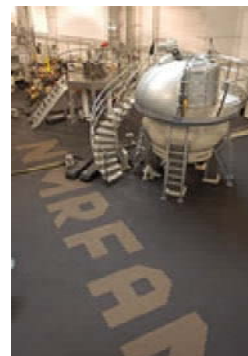
Goals for CESG NMR Team

- Automate screening for optimization of solvent conditions.
- Automate adaptive fast data collection.
- Automate processing and analysis of NMR data.
- Automate probabilistic data analysis.
- Rapid determination of three-dimensional protein structures.

Overview of NMR at CESG

The Teams' streamlined approach employs cryogenic probes and automated analysis to rapidly and efficiently determine three-dimensional protein structures by NMR. For proteins of up to 25 kDa in effective molecular weight that are soluble (> 0.5 mM), folded and stable, we acquire a complete data set consisting of 17 2D and 3D experiments in 10–14 days. Time-domain data are processed with NMR Pipe program, and converted to the XEASY format.

Data analysis is carried out in a semi-automated manner using software from various academic sources. Signals in all 3D experiments are detected automatically and integrated using the SPSCAN program. GARANT or PINE is used for automated assignment of backbone and chemical shifts, and side chain assignments are completed manually in XEASY or CARA. Backbone torsion angle restraints are predicted empirically from chemical shift values using the TALOS package and included in the initial round of structure calculations.

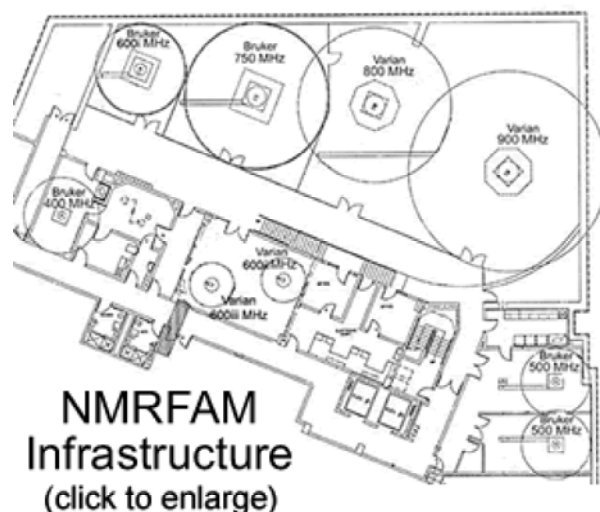


- [Bioinformatics](#)
- [Cell-Free](#)
- [Cloning](#)
- [Crystallography](#)
- [Expression Testing](#)
- [NMR](#)
- [Protein Production](#)
- [Protein Purification](#)
- [Quality Assurance](#)
- [Sesame](#)

Center for Eukaryotic Structural Genomics - NMR

Initial protein structures are generated using an iterative and fully automatic methodology for assignment of NOESY cross peaks provided by the NOEASSIGN module of CYANA. The final stages of structure refinement are accomplished through manual optimization of NOE assignments, NOE intensity-to-distance calibration functions, and backbone torsion angle restraints. Torsion angle dynamics structures that meet a set of objective criteria for agreement with experimental constraints and coordinate precision are subjected to molecular dynamics calculations in explicit solvent using the XPLOR-NIH package before final validation and deposition in the PDB and BMRB databases.

Structures are interrogated using bioinformatic methods that utilize structure and sequence based comparison tools, such as FATCAT, VAST, FFAS03, and Pfam. Bioinformatic analysis can generate a testable functional hypothesis, and experimental validation of our findings led to the identification of a new family of membrane-associated ubiquitin fold proteins, the MUBs. This modular pipeline strategy relies on a predefined framework for data management that enables an efficient workflow even when multiple personnel participate at various stages of the process. The self-contained nature of individual steps allows for substitution of improved software tools as new technology becomes available.



Methods Being Investigated and Publications

+ Structure Determination by NMR Spectroscopy

Here we describe Micelle induced folding of an intrinsically unfolded protein used to determine the structural features of nonphosphorylated TSP9. We have shown that an intrinsically unfolded eukaryotic protein can be folded through interaction of detergent micelles. Thylakoid soluble phosphoprotein of 9 kDa (TSP9) has been identified as a plant-specific protein in the photosynthetic thylakoid membrane. Nonphosphorylated TSP9 is associated with the membrane, whereas, after light-induced phosphorylation, a fraction of the phosphorylated TSP9 is released into the aqueous stroma. By NMR spectroscopy, we have determined the structural features of nonphosphorylated TSP9 both in aqueous solution and in membrane mimetic micelles. The results show that both wild type nonphosphorylated TSP9 and a triple-mutant (T46E + T53E + T60E) mimic of the triphosphorylated form of TSP9 are disordered under aqueous conditions, but adopt an ordered conformation in the presence of detergent micelles. The micelle-induced structural features, which are similar in micelles either of SDS or dodecylphosphocholine (DPC), consist of an N-terminal alpha-helix, which may represent the primary site of interaction between TSP9 and binding partners, and a less structured helical turn near the C-terminus. These structured elements contain mainly hydrophobic residues. NMR relaxation data for nonphosphorylated TSP9 in SDS micelles indicated that the molecule is highly flexible with the highest order in the N-terminal alpha-helix. Intermolecular NOE signals, as well as spin probe-induced broadening of NMR signals, demonstrated that the SDS micelles contact both the structured and a portion of the unstructured regions of TSP9, in particular, those containing the three phosphorylation sites (T46, T53, and T60). This interaction may explain the selective dissociation of phosphorylated TSP9 from the membrane. Our study presents a structural model for the role played by the structured and unstructured regions of TSP9 in its membrane association and biological function.

Song, J., Lee, M.S., Carlberg, I., Vener, A.V., Markley, J.L. (2006) **Micelle-induced folding of spinach thylakoid soluble phosphoprotein of 9 kDa and its functional implications.** *Biochemistry* 45(51):15633-43. [[17176085](#)]

+ Sample Preparation and Labeling

As stimulated by a cooperative agreement with [Cell-Free Sciences](#) (Yokohama, Japan) and [Ehime University](#) (Matsuyama, Japan), CESC's wheat germ cell-free protein production pipeline is quickly becoming the default method for screening targeted ORFs coding for proteins < 20 kDa for protein production and solubility, for preparing [¹⁵N]-labeled proteins to determine their suitability for NMR structural studies, and for preparing [¹³C,¹⁵N]-labeled proteins for structure determinations. Five structures have been solved with proteins made by this approach.

CESC is evaluating the replacement of manual screening by automated screening with a Cell-Free Sciences GeneDecoder1000 robot and is routinely using a Cell-Free Sciences Protemist robot for large-scale protein production.

These two robotic systems are the first of their kind installed outside Japan. CESH has prepared its first protein labeled with 'SAIL' (Stable Isotope Array IsotopeLabeled) amino acids. The SAIL approach, which requires cell-free protein production, promises to speed up the determination of high-quality structures of smaller proteins (those < 20 kDa) and to enable high-throughput determination of structures of protein up to 35 kDa.

+ Protein Refolding / Stabilization

A description of the balanced stabilization-destabilization approach for the prevention of aggregation after protein refolding is as follows. The protein was immobilized on nickel-agarose resin and refolded by stepwise decrement of the denaturant. The elution buffer was 20 mM sodium phosphate, pH 7.0, with 1% glycerol, 0.5 M urea, 300 mM NaCl, and 1 Mimidazole. After the removal of imidazole by ultrafiltration, the His-tag was cleaved with biotinylated thrombin. The protein product was kept in 20 mM sodium phosphate, pH 7.0, with 1% glycerol, 0.5 M urea, and 300 mM NaCl. The protein was found to aggregate extensively over time if any one of the three ingredients (sodium chloride, urea, or glycerol) was omitted. The yield of the protein was around 20 mg/L Luria-Bertani culture medium. The ¹H-¹⁵N HSQC spectrum showed the characteristic signature of a folded protein; thus, the solutes appear to have no deleterious effect on the sample. These solution conditions kept the protein soluble and unaggregated for at least 2 days (enough time for NMR data collection). This approach of balanced stabilization-destabilization may offer a general approach for structural investigations of proteins that tend to aggregate.

Chae, Y.K., Im, H., Zhao, Q., Doelling, J.H., Vierstra, R.D., Markley, J.L. (2004) **Prevention of aggregation after refolding by balanced stabilization-destabilization: production of the *Arabidopsis thaliana* protein At4g21980 (APG8a) for NMR structure determination.** *Protein Prod Purif* 34(2):280-3. [[15003262](#)]

+ HIFI-NMR Data Collection

HIFI-NMR is a rapid method for collecting data from a series of three-dimensional NMR experiments and determining the positions of peaks so as to develop a probabilistic peak list. HIFI-NMR uses the tilted-plane reduced dimensionality approach to data collection developed by E. Kupce and R. Freeman. The pioneering features of HIFI-NMR are that tilted planes are collected adaptively, one at a time, with an on-board algorithm choosing the angle of the next plane to be collected and determining in advance whether or not to collect these data on the basis of an estimation of the impact the new data would have on the evolving model of spectral peak locations in three dimensions. If collection of an additional plane is predicted not to improve the model, the software terminates the current NMR experiment and starts a new one. When data from all NMR experiments have been collected, the software provides peak lists for each experiment (chemical shifts in each dimension and probability measures for peak detection and shift accuracies). One output format option is NMR-STAR for direct deposition into BMRB. The progress of HIFI-NMR can be monitored from the spectrometer console, or data collection can be launched and followed from a remote computer. HIFI-NMR is being exported to other NMR laboratories, but currently is available only for Varian spectrometers. By determining peak positions as part of the data collection of each NMR experiment, the approach is designed to interface directly with PISTACHIO and PECANS.

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(36):12528-36. [[16144400](#)]

+ HIFI-C For Rapid, Reliable Determination of J Couplings

HIFI-C is a novel method for the robust, rapid, and reliable determination of J couplings in multi-dimensional NMR coupling data, including small couplings from larger proteins. HIFI-C is an extension of the adaptive and intelligent data collection approach introduced earlier in HIFI-NMR. HIFI-C collects one or more optimally tilted two-dimensional (2D) planes of a 3D experiment, identifies peaks, and determines couplings with high resolution and precision. The HIFI-C approach offers vital features that advance the goal of rapid and robust collection of NMR coupling data. (1) Tilted plane residual dipolar couplings (RDC) data are collected adaptively in order to offer an intelligent trade off between data collection time and accuracy. (2) Data from independent planes can provide a statistical measure of reliability for each measured coupling. (3) Fast data collection enables measurements in cases where sample stability is a limiting factor (for example in the presence of an orienting medium required for residual dipolar coupling measurements). (4) For samples that are stable, or in experiments involving relatively stronger couplings, robust data collection enables more reliable determinations of couplings in shorter time, particularly for larger biomolecules. The new approach has shown excellent quantitative agreement with values determined independently by the conventional 3D quantitative J NMR method (in cases where sample stability in oriented media permitted these measurements) but with a factor of 2-5 in time savings. The statistical measure of reliability, measuring the quality of each RDC value, offers valuable adjunct information even in cases where modest time savings may be realized.

modest time savings may be realized.

Cornilescu, G., Bahrami, A., Tonelli, M., Markley, J.L., Eghbalnia, H.R. (2007) **HIFI-C: a robust and fast method for determining NMR couplings from adaptive 3D to 2D projections.** *J Biomol NMR* 38(4):341-51. [[17610130](#)]

[+ NMR Data Analysis LACS](#)

NMR Data Analysis LACS provides the means for analyzing NMR data early on, prior to assignment or structure determination, to determine whether the ^{13}C chemical shifts are referenced properly and to identify ^{13}C and ^{13}C peaks with unusual chemical shifts. LACS takes advantage of the finding that, for a correctly referenced protein dataset, linear regression plots of ^{13}C , ^{13}C , or ^1H . (^{13}C - ^{13}C) pass through the origin from two directions, the helix-to-coil and strand-to-coil directions. LACS is available from a webserver at: <http://bija.nmrfam.wisc.edu/MANI-LACS/>. The BMRB uses LACS in screening chemical shift data sets being deposited and notifies depositors of possible problems with chemical shift referencing and the presence of outliers. The approach also has been used to derive unbiased ^{13}C and ^{13}C chemical shift values for residues in random coil.

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. [[16041479](#)]

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(3):155-65. [[16799859](#)]

[+ PINE \(Probabilistic Inference Network of Evidence\)](#)

PINE (Probabilistic Inference Network of Evidence), which is available from a webserver at <http://miranda.nmrfam.wisc.edu/PINE/>, represents the proof of concept and implementation of a multiple step probabilistic data analysis pipeline for protein NMR spectroscopy (A. Bahrami, L. Wang, J. L. Markley, H. R. Eghbalnia, manuscript in preparation). PINE incorporates the capabilities of separate probabilistic tools: PISTACIO (automated backbone and sidechain assignment), PECAN (secondary structure determination), and LACS (referencing offset and outlier detection). The input to PINE is the amino acid sequence and sets of peak lists generated from one or more of the standard types of protein NMR experiments; these can be either probabilistic (e.g., peak lists generated by HIFI-NMR) or traditional peak lists generated by popular NMR data analysis tools. PINE takes into account the interconnectedness of different stages of analysis. PINE begins with a set of local statistical potentials. It then proceeds iteratively until a stationary state for a consistent global similarity measure is achieved. The resulting software enables a seamless and robust integration of multiple steps in the NMR structure determination pipeline. PINE provides as output a probabilistic assignment of backbone and sidechain signals and the secondary structure of the protein. At the same time, it identifies, verifies, and if needed rectifies, problems related to referencing, assignment, or outlying data. PINE can make use of prior information supplied from selective labeling or spin system assignments derived independently by other means. The performance of PINE is much superior to that of the individual tools used sequentially.

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(3):219-33. [[16132822](#)]

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(1):71-81. [[16041485](#)]

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(1):13-22. [[16041479](#)]

[+ NMR Functional Followups](#)

Structures solved by CESG range from proteins of completely unknown function to those with a putatively assigned function based on structure or sequence homology to those with a highly defined function. Functional studies may be carried out by scientists interested in a specific solved protein. NMR has been used to detect and investigate protein:substrate interactions, protein:protein interactions, and complexes with other ligands such as DNA. Additionally, transient secondary and tertiary structures were detected in an apparently disordered protein, and a protein that was disordered in aqueous conditions was found to adopt an ordered conformation in the presence of detergent micelles.

Center for Eukaryotic Structural Genomics - NMR

- John Markley, PhD (PI)
- Arash Bahrami, PhD
- Claudia Cornilescu, PhD
- Sarata Sahu, PhD

NMR at Medical College of Wisconsin (MCW)

- Brian Volkman, PhD (PI)
- Kelly Kjer
- Betsy Lytle, PhD
- Francis Peterson, PhD

NMR studies were carried out at the National Magnetic Resonance Facility at Madison with support from the NIH Biomedical Technology Program (RR02301) and additional equipment funding from the University of Wisconsin, NSF Academic Infrastructure Program (BIR-9214394), NIH Shared Instrumentation Program (RR02781, RR08438), NIH Research Collaborations to Provide 900 MHz NMR Spectroscopy (GM66326), the NSF Biological Instrumentation Program (DMB-8415048), and the U.S. Department of Agriculture.

Center for Eukaryotic Structural Genomics (CESG), Department of Biochemistry
University of Wisconsin-Madison, 445 Henry Mall, Madison, WI 53706 | [Map](#) |
Telephone: 608.263.2183 Fax: 608.890.1942 Email: cesginfo@biochem.wisc.edu

| [Biochemistry](#) | [Private](#) | [PDB](#) | [PSI](#) | [PSI-KB](#) | [PSI-MR](#) | [Sesame](#) | [UW-Madison](#) |
Copyright © 2006 The Board of Regents of the University of Wisconsin System