Protein screening and optimization for NMR-based structural proteomics at the Center for Eukaryotic Structural Genomics (CESG)

Shanteri Singh[1], Robert C. Tyler[1], Claudia C. Cornilescu[1], Jikui Song[1], Qin Zhao[1], Francis C. Peterson[2], Brian F. Volkman[2] Min S. Lee[1], and John L. Markley[1]

[1]University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin, USA 53706-1549, [2]Department of Biochemistry, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin, 53226, http://www.uwstructuralgenomics.org

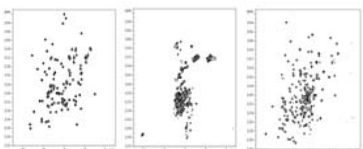
Abstract

We describe CESG's pipeline procedures for evaluating candidates for NMR structure determination and for salvaging targets that fail initial screens. ^{15}N-labeled proteins are evaluated under optimized default solutions conditions as: (HSQC+) suitable, (HSQC+/–) may become suitable, and (HSQC–) unsuitable for NMR study. HSQC+ samples are subjected to 14-day stability tests at room temperature (under conditions that simulate full data collection). Proteins that pass this test are labeled with 13C and 15N, and these proteins enter the NMR structure determination pipelines located at CESG-Madison or CESG-Milwaukee. Salvage pathways for targets that are HSQC+/– include buffer optimization and construct redesign to remove unstructured regions. Proteins that fail the stability test may be rescued by: buffer optimization, addition of proteinase inhibitors, or further purification. Proteins that remain HSQC– or HSQC+/– are transferred to CESG's X-ray crystallization pipeline.
Supported by NIH Grant P50 GM64598.

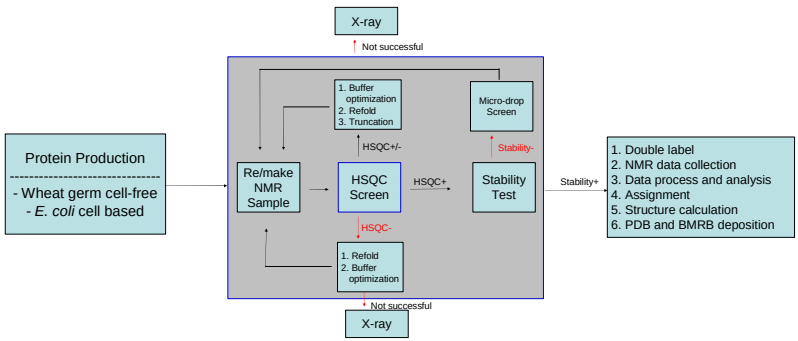
NMR pipeline target evaluation

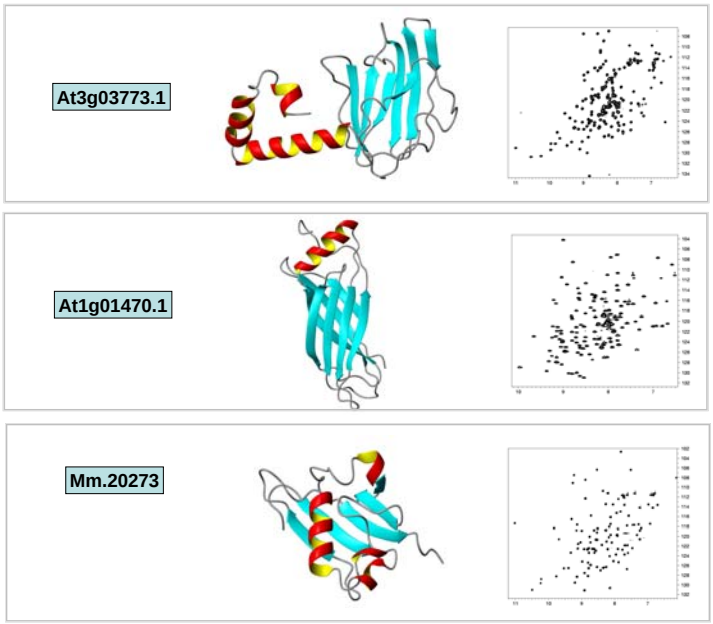
Following production and purification of ^{15}N-labeled target proteins, samples are screened for their suitability for NMR structural analysis. This entails collecting ^{1}H-^{15}N heteronuclear correlation spectroscopy of target proteins and classifying them into HSQC+, HSQC-, or HSQC+/- category. HSQC+ is defined as a target that is suitable for further NMR study. HSQC- is defined as a target that is not suitable for further NMR study. HSQC+/- classify targets that may become suitable for further NMR study. To be classified as HSQC+, the HSQC spectrum of the target protein must display well dispersed peaks characteristic of folded protein. In addition, uniform peak shape and consistent peak count must be satisfied for that particular target protein to be classified as HSQC+. After confirmation and classification into HSQC+ category, the targets of interest are subject to 14-day stability test. Upon satisfactorily concluding this stability test, the target proteins are classified as NMR structure candidates. Proteins are ^{13}C and ^{15}N labelled and enter the NMR structure determination pipeline. Targets that are classified as HSQC+/- enter salvage pathways that include buffer optimization and construct redesign. HSQC+/- classification of target proteins are diverse since the conditions for correct peak count and shape maybe influence by various factors. As an example, a protein aggregate of dimer to multimer may have some characteristics of folded protein but with broad peaks that are very much different from a well folded monomeric protein that has unstructured N- or C-terminal domain. Construct redesign entails removal of unstructured regions based on bioinformatics and observation of peaks in HSQC spectra characteristic of random or unstructured polypeptide chains falling within the narrow amide proton spectral range. Proteins that fail stability test are particularly good targets for buffer optimization. Stability can be achieved by simply changing the salt from K+ to Na+. Addition of additives such as high concentration of DTT can also stabilize Cys containing proteins and/or protease inhibitors can improve stability and convert a target into a NMR structure candidate.

Figure 1. HSQC Spectra of *Arabidopsis* Proteins



NMR Pipeline Target Evaluation Work Flow





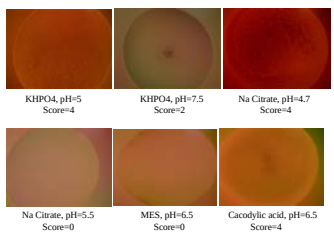
Screening for optimal NMR solution conditions

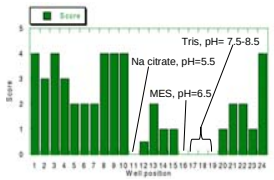
Following production and purification of ^{15}N-labeled target proteins samples are screened for their suitability for NMR structural analysis. A major limitation in the screening process is the difficulty in quickly identifying the optimum solution condition which will maximize solubility and chemical shift dispersion. UW-CESG has adopted a procedure developed by C. Lepre and J. Moore of Vertex Pharmaceticuals (*Journal of Biomolecular NMR*, 12 493-499, 1998) called Micro-Drop analysis, which enables the screening of a large set of solution conditions using very little protein sample. This procedure relies on the use of a 24 well Linbro plate, in which each well is filled with 500 µl of test buffer of varying pH. Next, a siliconized glass cover slip is then spotted with 1µl of a protein solution, typically 10 mg/ml, and mixed with 2 µl of the test buffer. The glass cover slip containing the protein micro-drop is then inverted and sealed onto the corresponding well using petroleum jelly. The sealed micro-drop is then allowed to equilibrate with the test buffer at room temperature. The 24 drops are monitored at regular intervals throughout the equilibration process with a microscope and scored on a scale of 0 to 4 for signs of precipitation. A score of 0 corresponds to no visible precipitation with a score of 4 corresponding to a drop that is completely covered with precipitated protein. Once the optimum buffer is identified the screening procedure can be repeated using known stabilizers such as salts, amino acids, and known cofactors. Using this methodology, testing against 24 different buffer conditions and stabilizers requires approximately 0.5 mg of protein. **Table 1** lists the 24 buffer conditions used in the screening procedure. **Figure 2** display examples of protein micro-drops in a variety of buffers with scoring result. This method was used to stabilize a protein target was that was degrading over time. Stability in this case was achieved by simply switching the salt in the buffer from K+ to Na+.

Table 1. Buffer Conditions

1. $KHPO_4$, pH=5.0	7. $NaHPO_4$, pH=7.5	13. Imidazole, pH=8.0	19. TRIS, pH=8.5
2. $KHPO_4$, pH=6.0	8. Na Acetate, pH=4.5	14. MES, pH=5.8	20. HEPES, pH=7
3. $KHPO_4$, pH=7.0	9. Na Acetate, pH=5.5	15. MES, pH=6.2	21. HEPES, pH=8.0
4. $KHPO_4$, pH=7.5	10. Na Citrate, pH=4.7	16. MES, pH=6.5	22. Bicine, pH=8.5
5. $NaHPO_4$, pH=5.5	11. Na Citrate, pH=5.5	17. TRIS, pH=7.5	23. Bicine, pH=9.0
6. $NaHPO_4$, pH=6.5	12. NH4 Acetate, pH=7.3	18. TRIS, pH=8.0	24. Cacodylic acid, pH=6.5

Figure 2. Micro-drop Examples and Scoring Results





Results from solution optimization screen. Each test buffer listed in **Table 1** is at a concentration of 100 mM, with starting protein solution approximately 10 mg/ml in 10 mM phosphate buffer. Micro-drops and scoring results presented in **Figure 2** were obtained after 48 h at 25 C. The complete scoring results for a target protein are shown in the histogram plot recorded after 48 h. Well position numbers correspond to the buffering condition listed in **Table 1**.

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (P50 GM 64598).