Northeast Structural Genomics Consortium
Rapid and automated protein NMR structure analysis of ^{2}H-enriched proteins using CS-Rosetta

Although NMR spectroscopy is well suited for rapid and - in favorable cases - automated structure determination of small (< 150 residues) proteins, solving structures of larger proteins and complexes is considerably more challenging. As the size of the molecule increases, so does the molecule's rotational correlation time and, consequently, the efficiency of ^{1}H-^{1}H relaxation mechanisms. One way to suppress these effects is to incorporate deuterium (^{2}H) into the protein sample, diluting the ^{1}H-^{1}H relaxation networks and increasing ^{13}C and ^{15}N relaxation times, resulting in sharper linewidths for ^{13}C, ^{15}N and remaining ^{1}H nuclei and dramatically improved signal-to-noise ratios. Perdeuteration is generally required for studies of larger proteins, particularly membrane proteins. While deuterium incorporation into protein samples can greatly improve the quality of data collected, the sample preparation can be very challenging. Working together with Profs. M. Inouye and M. Roth, the NESG has developed an approach for efficient production of [^{2}H,^{13}C,^{15}N]-enriched protein samples at a fraction of the cost of standard techniques. The technology is enabled by the co-expression of the protein MazF, which shifts *E. coli* into a dormant state in which it is no longer growing but is still capable of expressing recombinant target proteins. In this state, cells can be condensed as much as 40 fold into ^{2}H-enriched media without impacting protein production levels per cell. Using this condensed single protein production (cSPP) method, we can routinely and efficiently produce proteins in ^{2}H-based, isotope-enriched minimal medium.

During the PSI-1, we published a *fully automatic* strategy for rapid determination of medium-accuracy protein structures using only the sparse constraints that can be obtained using a perdeuterated protein (Zheng et al., 2003). Our strategy of rapid fold determination derives from ideas that were originally introduced for determining NMR structures of larger proteins (Gardner et al, 1997), using [^{2}H,^{13}C,^{15}N]-enriched protein samples with protonated sidechain methyl groups ($^{13}CH_3$). Data collection includes acquiring NMR spectra for determining assignments of backbone and sidechain ^{15}N, H^N resonances, and sidechain $^{13}CH_3$ methyl resonances. Assignments and 3D structures are then determined automatically by the NESG AutoAssign and AutoStructure software packages together with CNS. These structures provide accurate folds, and are good starting points for further refinement to high precision and accuracy using additional NMR data.

In collaboration with Prof. D. Baker (Univ. of Washington), we have recently extended this approach, exploring synergies between the Rosetta structure prediction method and our methods for automated analysis of protein NMR data, with the goal of applying these methods to larger (150 – 400 residue) perdeuterated proteins. The recently introduced CS-Rosetta method (Shen et al., 2008) provides an alternative approach for small protein structure analysis using only backbone and $^{13}C^{\beta}$ chemical shift data. We have applied the fully automated approach outlined above together with CS-Rosetta, using a perdeuterated, ^{15}N,^{13}C, $^{13}CH_3$ I(δ)LV methyl sample of the 85-residue *E. coli* cold shock protein CspA at 0.3 mM protein concentration. The sample was prepared using the cSPP system. Data were collected over a 3 day period at 600 MHz, and analyzed in a fully-automated fashion in an additional 2 day period. The resulting ensemble of 10 structures, shown in Fig. 1, exhibits excellent structure quality scores using several structural quality measures, and is in remarkably good agreement with the X-ray crystal structure of CspA, with backbone rmsd of 0.5 Å and all atom rmsd of 1.2 Å to the crystal structure for well-defined regions of the CS-Rosetta structure (1.1 Å rmsd to crystal structure for core, non-solvent exposed sidechain atoms). In order to further assess the validity of the CS-Rosetta structure, sidechain methyl resonances were assigned, and additional 3D ^{15}N-edited NOESY and ^{13}C-edited NOESY data were acquired. Comparison of the CS-Rosetta structure with these NOESY data reveals that essentially all of the resulting NOE-based constraints are already satisfied in each of the 10 CS-Rosetta structures. *By these criteria, the sparse constraint structure is as accurate as conventional NMR structures determined using complete sidechain resonance assignments*.

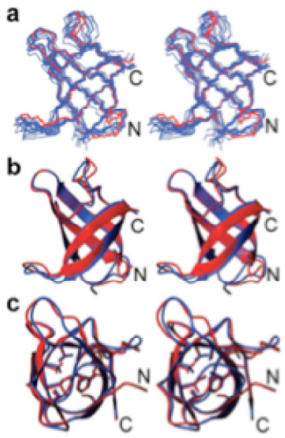

Fig. 1. Fully automated resonance assignments on a perdeuterated protein followed by CS-Rosetta. Stereoview of the superimposition of the CS-Rosetta structure for ^{2}H,^{13}C,^{15}N-enriched CspA (blue) with the 2.0 Å X-ray crystal structure of CspA (red) (pdb ID: 1mjc). (**a**) Backbone line representations of the 10 lowest energy conformers obtained from CS-Rosetta structure compared with X-ray crystal structure. (**b**) Ribbon diagram of the lowest energy conformer versus X-ray crystal structure. (**c**) The packing of the core hydrophobic residues (Y. Tang & G. Montelione, in preparation).

(1) Zheng, et al. Protein Sci, 2003. **12**: 1232-46. (2) Gardner et al., Biochemistry, 1997. **36**: 1389-401. (3) Shen, Y., et al., Proc Natl Acad Sci U S A, 2008. **105**: 4685-90.