Proposal ID #: 19837

Title: Progress Report for Protein Interactions and Interfaces Science Theme Proposal, year 1

Date: March 15, 2007

Investigative Team:

John R. Cort (Washington State University / PNNL), Theresa A. Ramelot (Miami University), Michael A. Kennedy (Miami University), Adelinda Yee (University of Toronto), Bin Wu (University of Toronto), and Cheryl H. Arrowsmith (University of Toronto), James M. Aramini (Rutgers University), Gaetano T. Montelione (Rutgers University)

The Northeast Structural Genomics Consortium (NESGC, www.nesg.org) used EMSL's High Field NMR facility in 2006-2007 to study the structure and dynamics of proteins and protein-ligand complexes within the EMSL Science Theme Proposal entitled "Protein Interactions and Interfaces". The NESGC, supported by the NIH Protein Structure Initiative (PSI), is engaged in high-throughput NMR and X-ray structure determination of proteins that are not similar in amino acid sequence to proteins of known structure. This coarse survey of structurally uncharacterized proteins is known as structural genomics. Concisely stated, the goal of the PSI is

> ***to make the 3D atomic resolution structures of most proteins easily available from their corresponding DNA sequences.***

To be sure, the long-range value of this endeavor lies in the area of functional genomics—assigning biochemical and cellular functions to genes and proteins. We must remember that ALL proteins function at the most basic level by binding to other molecules! In this sense, by determining the structure of an uncharacterized protein, we learn the characteristics of its surfaces; these characteristics determine to what other molecules it will bind and with what affinity. Such information will only become more valuable in the future as knowledge of each protein's "interactome" grows, both through experimentation and, increasingly, through de-novo prediction of protein-protein and protein-ligand interactions. To be responsive to the original EMSL Science Theme "Biological Interactions and Interfaces", we prioritized protein targets for NMR data collection and structure determination at EMSL that are either dimeric or are known to bind ligands, metals, DNA, and so on. As solution state NMR is an excellent tool for the study of protein dynamics as well as structure, our structural genomics efforts will merge easily into the new version of this EMSL Science Theme, "Biological Interactions and Dynamics". In fact, we routinely characterize the dynamical behavior of proteins during the process of structure determination. Dynamics and protein ligand interactions are connected: internal regions of greater mobility in globular proteins are often ligand binding sites or enzyme active sites, and changes between bound and unbound states of a protein are often found to be correlated with changes in these dynamics.

Summary of Activity during the first 10 months (May ----mid March) of the Science Theme Proposal:

EMSL NMR instruments are used extensively for data collection for protein structure determination. The most heavily-used instruments are the 600 (glacier) and 750 (rainier). Also used less frequently are the cryoprobe equipped 600 (baker) and 800 (denali), and the ultrahighfield 900 (everest). Also used was the ICP-MS instrumentation for metal determination. A novel metal binding function was discovered in one protein and identity of the metal was confirmed ICP-MS. Also, extensive use was made of the EMSL linux cluster "king" for structure determination and data processing. Specialized NMR experiments called isotope edited-filtered NOESYs used for determining protein dimer and ligand-bound structures have been implemented and optimized as part of this work, and these experiments are now a capability that is utilized by other EMSL users.

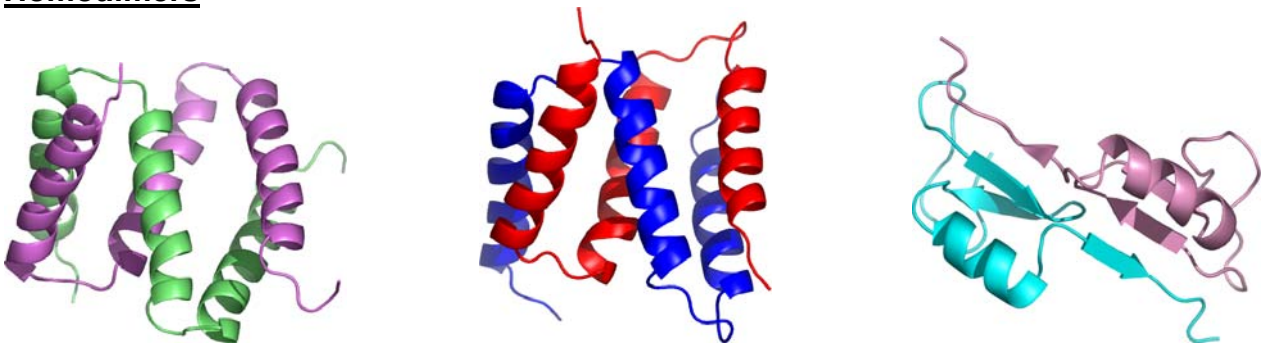
- Proteins for which complete data sets were collected for protein structure determination: **9**
- Proteins for which partial data sets were collected to supplement data collected elsewhere in the NESGC (typically high field NOESY data): **5**
- Proteins subjected to specialized NMR experiments to characterize ligand binding, dimer interfaces, or functional activity: **5**
- Proteins submitted for ICP-MS analysis of bound metals: **5**
- Structures deposited to the Protein Data Bank determined with NMR data collected at EMSL: **12**

EMSL Resources Requested:

For the second year of this science theme proposal, we request the same amount of time as in the previous year: 60 weeks of NMR time: 60% at 600 MHz, 40% at 750 MHz (generally 800 & 900 MHz is not needed). Some need for the cryoprobe-equipped 600 MHz instrument (baker) is anticipated (6 weeks). We also request access to computing resources (king) and ICP-MS instrumentation for metal analysis.

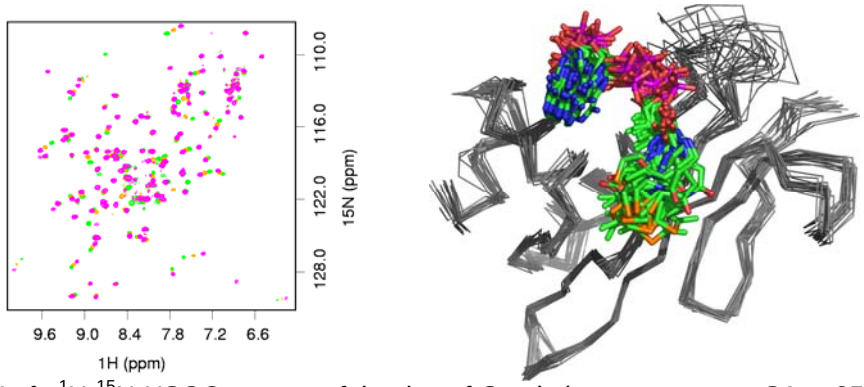
Highlights of systems studied during the first 10 months of the science theme proposal

Homodimers

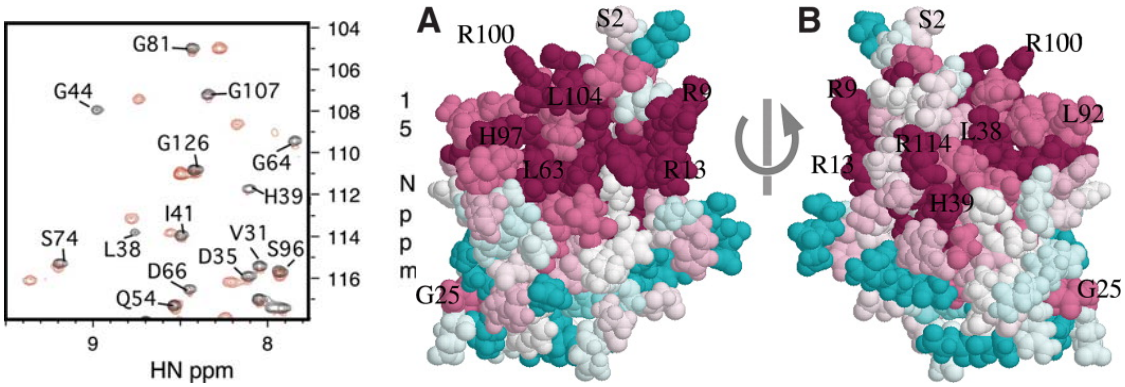


Left: *Escherichia coli* YejL (ER309), Middle: *Vibrio parahemolyticus* YejL homolog (VpR61), unknown function, Right: *Bacillus subtilis* YqaI (SR450, pdb), unknown function. Individual monomer subunits colored separately.

Ligand Binding

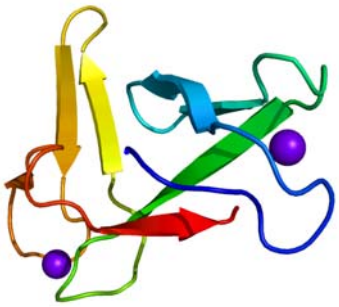


Left, ^{1}H-^{15}N-HSQC spectra of titration of *Staphylococcus aureus* $SA_{COL}2532$ with coenzyme A (green: 0X, orange: 1X, magenta: 10X excess ligand) demonstrating fast-exchange binding kinetics ($k_{ex} >= 10^{5}$ sec^{-1}). Right, structural ensemble with bound ligand determined experimentally using the ^{13}C-edited-filtered NOESY experiment. Structures were calculated using the HADDOCK package running on the 32-node Linux cluster "king".

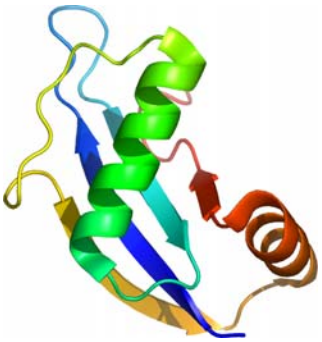


Left, ^{1}H-^{15}N-HSQC spectrum of *Pseudomonas aeruginosa* PA4608 (PaT7, 1YWU.pdb) with 0.7 equivalents of cyclic-diGMP bound, demonstrating slow-exchange kinetics (k_{ex} <= 1 sec^{-1}), in constrast to the previous example. Right (A&B), mapping of conserved residues onto the surface of. Darker, warmer colors represent greater conservation in the sequence family; clustering of conserved residues in proteins often indicates a binding site.

Metal Binding Sites

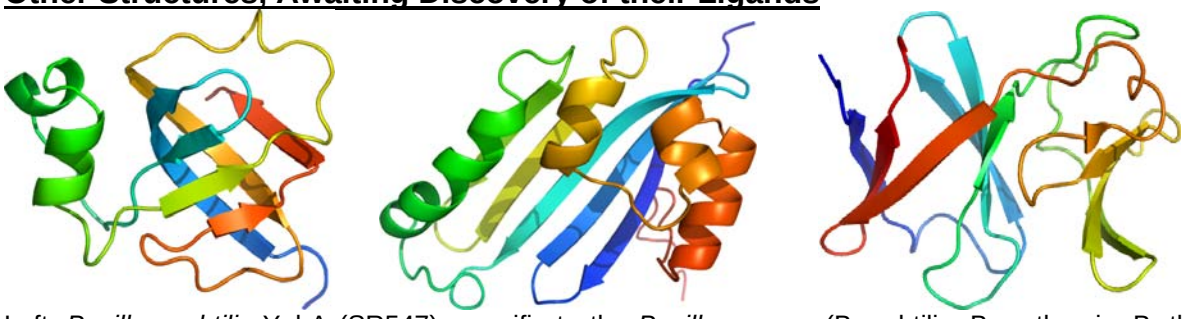


Novel Fold Architecture



Left, *E. coli* YfgJ (ER317, 2JNE.pdb), showing dual Zinc-binding sites suggested initially by clustering of cysteine residues in the NMR structure, and confirmed by ICP-MS analysis. Right, protein NE1680 (NeT5, 2HFQ.pdb) from *Nitrosomonas europea*, an ammonia oxidizing bacterium important for biogeochemical cycling of nitrogen. NE1680 was found to have a novel fold architecture that had not been previously observed.

Other Structures, Awaiting Discovery of their Ligands



Left, *Bacillus subtilis* YobA (SR547), specific to the *Bacillus* genus (B. subtilis, B. anthracis, B. thurigiensis, B. cereus). OB fold architecture suggests oligonucleotide or oligosaccharide binding function. Middle, *E. coli* YehR (ER538, 2JOE.pdb), unknown function. Right, *E. coli* NirD (ET100, 2JO6.pdb) from nitrite reductase operon.

Other Proteins—complete data sets

Rhodopseudomonas palustris RPA2110 (RpT6), a domain-swapped dimer; *Helicobacter pylori* HP1203/NusL (PT1);

Other Proteins--partial data sets (high field NOESY experiments)

Salmonella typhimurium proteins STM0327(StR65, 2JN8.pdb), STM2801/YgaC(StR72, 2G7J.pdb), PLST013/PefL (StR82); *E. coli* YqcC (ER225, 2HGK.pdb), probably interacts with the pseudouridine synthase TruC or with that enzyme's RNA substrate.

Other Proteins—specialized data collection and analysis for functional studies

Allochromatium vinosum DsrC (1YX3.pdb): Cys mutants & binding studies with DsrEFH.
Aquifex aeolicus 5,10-methenyltetrahydrofolate synthase (1SOU.pdb): mutants of active site residues.
Staphylococcus aureus $SA_{COL}2532$: titration with CoA ligand and spectroscopy of the complex (2H5M.pdb).
Pseudomonas aeruginosa PA4608: characterization of ligand binding.

Publications:

The solution NMR structure of Escherichia coli YtfP expands the structural coverage of the UPF0131 protein domain family. J.M. Aramini, Y.J. Huang, G.V.T. Swapna, J.R. Cort, P.K. Rajan, R. Xiao, R. Shastry, T.B. Acton, J. Liu, B. Rost, M.A. Kennedy, and G.T. Montelione. Proteins: Structure, Function, and Bioinformatics, in press (2007).

NMR structure and binding studies confirm that PA4608 from *Pseudomonas aeruginosa* is a PilZ domain and a cyclic-di-GMP binding protein. T.A. Ramelot, A. Yee, **J.R. Cort**, A. Semesi, C.H. Arrowsmith, and M.A. Kennedy. Proteins: Structure, Function, and Bioinformatics, 66, 266-271 (2007).

Structure of an Acetyl-CoA Binding Protein from Staphylococcus aureus Representing a Novel Family of GCN5-related N-Acetyltransferase-like Proteins with a Conserved Cysteine in the Binding Site. John R. Cort, Theresa A. Ramelot, Diana Murray, Thomas B. Acton, Li-Chung Ma, Rong Xiao, Gaetano T. Montelione, Michael A. Kennedy. Journal of Structural and Functional Genomics, submitted.