The MAGGIE Project: Production and Isolation of Tagged Native/Recombinant Multiprotein Complexes and Modified Proteins from Hyperthermophilic *Sulfolobus solfataricus*

Denise Munoz[1], Jill Fuss[1], Kenneth Stedman[2],Daojing Wang[1], Michael W. W. Adams[3], Gary Siuzdak[4] Nitin S. Baliga[5], Steven R. Holbrook[1], John A. Tainer[1,6], and Steven M. Yannone[1]

[1] Dept. of Molecular Biology, Lawrence Berkeley National Lab
[2] Center for Life in Extreme Environments, Portland State University
[3] Department of Biochemistry and Molecular Biology, University of Georgia
[4] Center for Mass Spectrometry, The Scripps Research Institute
[5] Institute for Systems Biology
[6] Department of Biochemistry and Molecular Biology, The Scripps Research Institute

The development of effective discovery based science technologies benefits from a means for validating novel discoveries. As part of the MAGGIE project, we are developing recombination-based viral shuttle vectors for high throughput recombinant protein expression. These tools will be applied to express and isolate poly-histidine tagged Sulfolobus proteins from the native organism. Metal affinity purification of tagged Sulfolobus proteins from Sulfolobus will allow mild purification conditions preserving weak interacting protein partners. Additionally, room temperature protein purification will take advantage of the hyperthermophilic nature of Sulfolobus to “thermally trap” protein complexes assembled in Sulfolobus, which grow optimally at 80 degrees C. By isolating Sulfolobus proteins from a native background we will facilitate native cofactor assembly and preserve normally occurring post-translational modifications of targeted proteins and protein complex assemblies. We will interface with other MAGGIE investigators to identify protein complex component proteins with advanced MS/MS technologies, structurally characterize stable complexes with small angle x-ray scattering in solution, and validate novel protein complexes identified in other MAGGIE organisms. Ultimately, we aim to identify metabolic modules and design functional gene islands suitable to transfer metabolic processes between microbes to address specific DOE missions and develop generally applicable molecular and biophysical technologies for GTL.

Development and Application of Phylogenetically-Constrained Metabolic Activity 'Signature' Probes

Fred Brockman[1]* (fred.brockman@pnl.gov), David Culley[1], Mustafa Syed[2], Banu Gopalan[3], Tanuja Bompada[2], Margaret Romine[1], and Natalia Maltsev[2]

*Presenting author

[1]Microbiology Group, Pacific Northwest National Laboratory, Richland, WA; [2]Bioinformatics Group, Argonne National Laboratory, Argonne, IL; [3]Computational Biology and Bioinformatics, Pacific Northwest National Laboratory, Richland, WA

Microbial communities in Nature carry out complex processes, often involving many different trophic groups that utilize different electron donors and electron acceptors and have very different anabolic and catabolic capabilities. To better understand how the material and energy flow in these complex processes, technologies are needed to interrogate, at fine resolution, the spatial organization of gene expression in microbial communities.

Bulk extraction of RNA and analysis on microarrays largely destroys the spatial and functional linkages that are the key to understanding how communities function and interact. In cell (in situ) RT-PCR has been successfully applied to bacteria in a few cases, but is an enzymatically difficult method that is prone to false positives and false negatives and not amenable to the high-throughput processes the GTL program is seeking.

Therefore, the goals of the project are to:

- Develop mRNA-targeted, non-PCR-based, fluorescence in situ hybridization (mRNA-FISH) using near-infrared oligonucleotide probes coupled to advanced microscopy able to detect a very small number of target molecules in prokaryotic cells.
- Develop grid-based computational tools to simultaneously utilize all community metagenome datasets to design a suite of phylogenetically-constrained, mRNA-targeted metabolic function "signature" probes. Such probes will better enable the development of hypotheses regarding the functional processes and metabolic linkages occurring in multi-species communities.

Single microbial cells are being hybridized with mRNA-targeted oligonucleotides labeled with near-infrared fluors, near-infrared quantum dots, and conjugates for enzyme-labeled fluorescence (ELF). The advantage of these read-outs over typical fluorescent read-outs is decreased background signal (noise). The higher signal to noise ratio, combined with the advanced microscopes present in the Environmental Molecular Sciences Laboratory at PNNL, enables detection of a lower number of hybridization events per cell. Initially, we are working with the easily permeable *E. coli* model system to better understand the

parameters which are most problematic to obtaining hybridization signal. Immobilization of mRNA inside of the cell appears to be the parameter that most limits mRNA-targeted FISH in *E coli*. (We anticipate that in other bacteria and archaea, membrane permeabilization will also be a major limiting factor. We are currently pursuing methods of immobilizing mRNA inside the cell prior to and coincident with cell permeabilization. These experiments are using *E. coli*, *Shewanella oneidensis*, and *Desulfovibrio vulgaris* as the model bacteria.

Past work in this project analyzed the metagenome of the living microbial community present in low biomass (~10,000 cells per gram) subsurface sediments (60 to 150 feet deep) beneath a leaking high-level radioactive waste tank at the DOE Hanford Site. Low-coverage shotgun sequencing was performed by the DOE Production Genomics Facility. 95% of all protein hits were to the Actinobacteria (high GC Gram positive) phyla and gamma division of the Proteobacteria phyla. The e-score distribution for Proteobacteria hits was relatively high (median of 80%); in contrast, the e-score distribution for Actinobacteria hits was quite low (median of 45%). Surprisingly, this e-score distribution was consistently observed across all role categories. Data from amplified and cloned 16S rDNA, from the identical template DNA used for the shotgun sequencing, showed a relatively diverse Actinobacteria community containing 10 genera from the Micrococcineae, Propionibacterineae, and Steptomycineae suborders. This represents the most comprehensive metagenome analysis to date for the Actinobacteria, and indicates these subsurface Actinobacteria are quite novel at the genomic level compared to other studied Actinobacteria.

To demonstrate the phylogenetically-constrained, mRNA-targeted metabolic function "signature" probe approach, these metagenome data are being used to develop mRNA-targeted probes for key anabolic and catabolic genes that differentiate the Micrococcineae, Propionibacterineae, and Steptomycineae – both between these three suborders, and between the novel Hanford genes and previously known genes within each order. To accomplish this task, grid-based computational tools that utilize all existing prokaryotic metagenome sequence are being developed and applied at ANL.

Network Inference of Prokaryote Biology

Sergej V. Aksenov, Dmitriy Leyfer, Robert Miller and Bruce W. Church
Gene Network Sciences, 31 Dutch Mill Rd, Ithaca NY 14850

Most prokaryotes of interest to DOE are poorly understood. Even when full genomic sequences are available, the functions of only a small number of gene products are clear. The critical question is how to best infer the most probable network architectures in cells that are poorly characterized. The project goal is to create a computational network inference framework that uses machine-learning network inference methodologies to integrate large-scale dynamical and probabilistic simulation, high-throughput bioinformatics-derived evidence, together with high-throughput experimental data to automate the discovery of biology. For this process to be efficient the level of resolution of the modeling paradigm must properly match both the constraining experimental data and the bioinformatics. To integrate the diversity of high-throughput experimental data and bioinformatics, we have developed the GNS Network Inference Engine (NIEngine), a parallel software platform for both probabilistic and mechanistic network inference. Probabilistic models provide a course-grained modeling framework appropriate for static gene and protein expression experiments and genome-level bioinformatics. The output from these coarse-grained models in turn provides the feedstock for network inference of high-resolution mechanistic models that are constrained by more detailed time-dependent experimental data. NIEngine produces ensembles of networks which span the diversity of networks consistent constraint data and prediction and prediction uncertainty are computed through analysis of the network ensemble both topologically and through forward simulation. We will present examples using *E. coli* reference models employing the MCM prokaryote framework developed by Michael Shuler where we are currently benchmarking the methodology in preparation for application to *Shewanella oneidensis*.

Highly Integrated Microscopies of Molecular Machines (M^3): Structural Dynamics of Gene Regulation in Bacteria

Carlos J. Bustamante, Jan T. Liphardt, Tracy Nixon*, Eva Nogales, and Haw Yang

*Physical Biosciences Division, Lawrence Berkeley Natl. Lab, University of California, Berkeley, * Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park*

Microbial molecular machines such as the NtrC response regulator are complex and dynamic, sometimes changing their composition and structure on the millisecond or second timescales. The full characterization of microbial molecular machines requires an integrated approach composed of complementary imaging and other characterization technologies that span a broad range of spatial and temporal resolutions.

We are using a combination of cryo-EM, optical tweezers, atomic force microscopy, and single-molecule fluorescence to obtain the composition of molecular machines, the structures of various conformational states, and the machines' dynamics and mechanochemistry. Cryo-EM is used for structural characterization, Optical Tweezers are used to investigate the mechano-chemistry and dynamics of various nucleic-acid binding molecular motors, AFM is used to image protein-DNA interactions, and single-molecule fluorescence is used to obtain dynamic information in real time of these machines. The complementary capabilities of these methods makes their combined use well-suited to address the structural and dynamical complexity of multi-molecular assemblies. Progress has been made in several subprojects:

1. **Pausing behavior of single *E. coli* RNA polymerases**. Using optical tweezers, we have carried out single-molecule studies of the pausing behavior of single *E. coli* RNA polymerases. We found that enzyme processivity is highly temperature dependent, and that temperature differentially various elongation behaviors. The results constrain models of the mechanochemistry of microbial polymerases.
2. **Structure of NtrC-DNA complexes**. By bending promoter DNA regions, the nitrogen response regulator NtrC contacts the sigma^{54}-containing RNA polymerase (RNAP-σ^{54}) and activates it by catalyzing the formation of the open promoter complex. In collaboration with Professor Tracy Nixon, we are investigating this process by the direct visualization of NtrC-DNA complexes using AFM.
3. **Structural studies of NtrC**. Using electron microscopy and single particle image reconstruction we have generated density maps at about 25 Å resolution of the full length NtrC protein in its oligomerized state. Furthermore, by docking the atomic structures of the different domains a pseudo-atomic model of the full length, activated NtrC has been generated.
4. **Structural dynamics of the central domain of NtrC1 ATPase from *Aquifex Aeolicus.*** Single molecule fluorescence experiments were used to study the nucleotide-dependent conformational distribution of single monomers of the central domain of NtrC1 ATPase from *A. Aeolicus*.
5. **Sequences that direct the movement of FtsK**. FtsK is a membrane-bound and septum-localized *E. coli* translocase that coordinates cell division with chromosome segregation. Optical tweezers studies have been used to identify the sequences that direct the movement of the *E. coli* FtsK translocase, an ATP-driven machine required for DNA replication, recombination, and transfer within and between cells.
6. **Characterization of DNA gyrase**. Magnetic tweezers and optical tweezers have been used to perform a mechanochemical analysis of DNA gyrase, a molecular machine that uses the energy of ATP hydrolysis to introduce essential negative supercoils into DNA.
7. **User-friendly, automated, scalable optical tweezers**. A user-friendly "mini" optical tweezers sytem has been developed to permit other GTL investigators to perform tweezers studies in their labs. These tweezers are also highly automated and fully computer controlled, to enable their use in a high thruput context.
8. **Single-molecule sensor of mechanical force with optical readout**. A single-molecule sensor of mechanical force with optical readout has been developed, which will permit the real-time measurement of stresses and strains inside molecular machines and their substrates, such highly bent promoter regions. We are now applying these nanoscale strain gauges to study the bacterial histone analog, H-NS.
9. **New approaches for Cryo-EM of molecular machines**. We have developed a novel approach, the Orthogonal Tilt Reconstruction method (OTR), that eliminates the missing cone problem present in the best of the traditional methods. Combined with 3D Multivariate Statistical Analysis (MSA), the method identifies conformationally distinct shapes, categorizes them separately, and carries out a multi-reference alignment on the full data set.