Genomes to Life Center for Molecular and Cellular Systems
Annual Report
FY 2003

Compiled by:

Michelle Buchanan, Director
Oak Ridge National Laboratory

Frank Larimer, Deputy Director
Oak Ridge National Laboratory

Steve Wiley, Deputy Director
Pacific Northwest National Laboratory

October 2003

This research was sponsored by the Office of Biological and Environmental Research, U.S. Department of Energy, under Contract DE-AC05-00OR22725 with Oak Ridge National Laboratory (ORNL), managed and operated by UT-Battelle, LLC and under Contract DE-AC06-76RL01830 with Pacific Northwest National Laboratory (PNNL), managed and operated by Battelle

Table of Contents

Executive Summary 1

1.0 Protein Complex Identification 3

2.0 Research and Development and Support Activities 34

3.0 FY 2004 Plans and Milestones 85

Appendix – List of Project Participants/Contributors 87

Executive Summary

The first year of the Genomes to Life (GTL) Center for Molecular and Cellular Systems focused on establishing a pipeline for the efficient analysis of protein complexes, identifying "bottlenecks" in the analysis processes, and developing approaches that could eliminate these bottlenecks. Oak Ridge National Laboratory (ORNL) and Pacific Northwest National Laboratory (PNNL) staff worked closely together to develop a comprehensive program that would lead to an integrated process for protein complex analysis. Different protocols were evaluated at the two Laboratories and compared to optimize the analytical processes. Regular teleconferences were established to exchange information between the two Laboratories. To test the established protocols two organisms were used *Rhodopseudomonas Palustris* (*R. palustris*) at ORNL and *Shewanella oneidensis* (Shewanella) at PNNL. Particular emphasis was given in this first year in establishing a common laboratory information management system (LIMS), that would allow data to flow across the two Labs.

The following is a compilation of reports from various activities conducted during this first year of the project. This report is divided in to several sections, including summaries of core activities focused on isolation of specific complexes from *R. palustris* and Shewanella, research activities which assessed different complex isolation approaches, LIMS and bioinformatics development, automation, and initial imaging studies. Overall, the various activities conducted in the first full year of this project may be summarized as follows:

- A broad host range plasmid system for expression of affinity tagged protein targets was designed, constructed, and tested. Test genes have been cloned into entry and expression vectors. Protein expression has been verified in both *E. coli* and *R. palustris.*
- Growth conditions have been established for both organisms. For example, *R. palustris* (strain CGA10 from Harwood's lab) has been grown under four different culture conditions ranging from autotrophic to heterotrophic modes.
- A method of tandem affinity purification, TAP, has been optimized and found to minimize artifacts compared to single tags pulldowns.
- Pulldown experiments have been conducted on complexes in both organisms, including GroESL, nitrogenase, ATP synthase, CO_2 fixation, uptake hydrogenase, ribosome, photosynthesis reaction center, signal recognition partical, and Clp protease in *R. palustris* and RNA polymerase, the degradosome, oxidoreductase, tyrosine phosphatatase and methionine sulfoxide reductase in Shewanella.
- Several exogenous pulldown schemes for rapid isolation of complexes from Shewanella have been examined and optimized, including capture approaches employing novel solid phase capture technique that could be incorporated in a high throughput, automated process.
- Affinity isolated proteins from *R. palustris* and Shewanella complexes have been analyzed by "bottom-up approaches. In the "bottom-up" approach, tryptic digests of proteins to yield peptides, which were analyzed by mass spectrometry, and the resulting peptides were identified using automated searches of data bases. In addition, a "top-down" approach was employed to analyze intact proteins from *R. palustris* complexes. In this approach, the proteins were analyzed by high performance FTICR, allowing important modifications in the proteins isolated from the complexes to be identified.
- A commercial LIMS system (Nautilus) is being configured for the current pipeline processes, working towards having the system functional and up-to-date to enable accurate tracking of the

production activities. Staff at ORNL have taken the lead on this task and are coordinating with staff at PNNL to ensure a fully integrated system across the Center.

- PNNL-developed PRISM system for MS data analysis is being adapted for use at ORNL.
- Evaluated several approaches for chemical cross-linking of protein complexes using model systems
- Evaluated approaches for reproducible and automatable tryptic digestion of proteins.
- Evaluated several approaches for automation of the reagent production, complex pulldown and analysis pipelines and equipment ordered.
- Developed approaches for isolating and imaging live bacteria using AFM; manuscript published.

Based on the progress during this first year, the research teams met in Oak Ridge in late October 2003 to discuss these results and formulate an integrated experimental plan for FY04. In this experimental plan ORNL will produce clones for use in both endogenous pulldowns at ORNL and exogenous pulldowns at PNNL. PNNL will also use the clones to produce antibodies that will be shipped to ORNL for endogenous pulldowns. This overall scheme for an integrated high throughput pipeline for the analysis of protein complexes as part of this Center is summarized in **Figure ES1**, with activities at ORNL highlighted in orange and those at PNNL, in blue. To start the new project period, 192 new *R. palustris* target genes were selected for use in development of protocols for high throughput/automated cloning into affinity tag expression plasmid system. Additional clones for Shewanella are being prepared for shipment to PNNL. PNNL will use these Shewanella clones (1) to produce proteins in E. coli for exogeneous pulldowns, and will also (2) produce scFv against *R. palustris* proteins for ORNL to use in pulldowns. Note that this scheme employs two approaches for pulldowns—one using exogenous pulldowns that will be high throughput but less selective and therefore more subject to artifacts and a second that uses slower, but more selective endogenous pulldowns that can confirm complexes tentatively identified with the exogenous approach. This first year has shown that a multiple pulldown approach will provide the highest quality data. In addition to these activities, effort will also be focused on automating the pipeline and completing the LIMS systems across the two Laboratories for sample tracking. Data will be collected and analyzed using a common system and databases will be established.

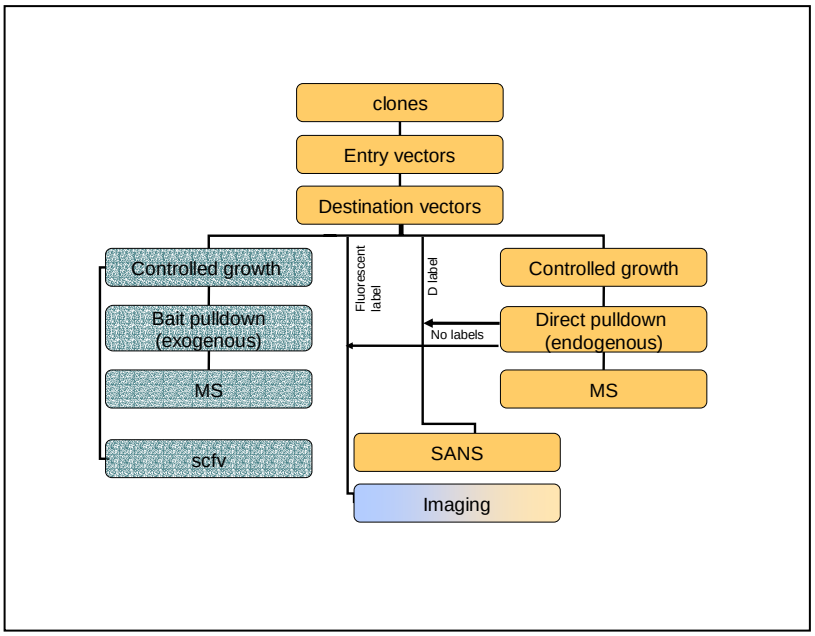

Figure ES1. Integrated high throughput pipeline