Mass Spectrometric Identification of Protein Complexes in *Rhodopseudomonas palustris*

Gregory B. Hurst[1], Dale A. Pelletier[1], Stephen J. Kennel[1], Frank W. Larimer[1], W. Hayes McDonald[1], Trish K. Lankford[1], Manesh B. Shah[1], Denise D. Schmoyer[1], Tse-Yuan S. Lu[1], Linda J. Foote[1], Elizabeth T. Owens[1], Cathy K. McKeown[1], Jenny L. Morrell-Falvey[1], Mitchel J. Doktycz[1], Brian S. Hooker[2], Deanna L. Auberry[2], William R. Cannon[2], Kenneth J. Auberry[2], H. Steven Wiley[2], Michelle V. Buchanan[1]

[1]Oak Ridge National Laboratory, Oak Ridge TN [2]Pacific Northwest National Laboratory, Richland WA

OVERVIEW

- The protein-protein interaction network in *Rhodopseudomonas palustris* is elucidated by affinity isolation and mass spectrometry
- Results are available via the Microbial Protein-Protein Interaction Database at mippi.ornl.gov

INTRODUCTION

- Genome-scale elucidation of protein-protein interactions is
  - an emerging component of systems biology
  - a tool for generating hypotheses regarding protein function
- The current focus of the Genomics:GTL Center for Molecular and Cellular Systems (CMCS) is to identify the set of protein complexes used by *Rhodopseudomonas palustris*, and to determine changes in these protein-protein interactions under different metabolic conditions.
- Affinity tags can be introduced into proteins of most gram negative bacteria using a cloning system that combines a ***convenient commercial system*** with a ***broad host range plasmid***.
- *R. palustris* is a ubiquitous, metabolically diverse anoxygenic phototrophic bacterium that
  - fixes nitrogen
  - performs photosynthesis (carbon sequestration)
  - degrades aromatics (bioremediation)
  - produces hydrogen gas (energy production)
- We have identified protein-protein interactions from cultures of *R. palustris* grown under anaerobic photoheterotrophic growth conditions, in the presence or absence of fixed nitrogen. Interactors identified by this approach include homologues of a number of well-characterized protein complexes involved in known metabolic networks as well as numerous potential novel interactions.

EXPERIMENTAL

High-throughput PCR amplification and cloning of selected targets into Gateway Entry vectors

↓

Recombination of entry clones into low copy Gateway compatible broad host range expression plasmid

↓

Expression of "bait" V5/6xHis fusion protein in *R. palustris* and endogenous protein complex formation

↓

Affinity purification (IMAC followed by V5 antibody) of bait protein with interactors

↓

Elution of specific interactors

↓

Trypsin digestion and LC-MS-MS identification of interacting proteins

↓

Statistical and informatics analysis

↓

Post results to internal and public websites

- Primers are designed for each gene of interest, and PCR amplification is performed. A modified commercial system (Gateway, Invitrogen) is used to prepare clones encoding a dual affinity tag (hexahistidine and V5 epitope) at the position corresponding to the C terminus of each protein. A broad host range plasmid allows expression of affinity-tagged proteins in a variety of gram negative bacteria [1]. Proteins are expressed in separate 800 mL *R. palustris* cultures, followed by cell harvest and lysis. A two-stage affinity purification isolates the tagged protein, and other proteins in the complex. Complexes are digested with trypsin, and analyzed by reverse-phase nano-scale HPLC (LCPackings Famos/Switchos/Ultimate system) coupled via electrospray with a quadrupole ion trap mass spectrometer (ThermoFinnigan Deca XP +). Tandem mass spectra are identified using Sequest, and results compiled using DTASelect.
- Further details of methods are available at http://mippi.ornl.gov/methods/index.shtml

RESULTS AND DISCUSSION

- As of late May 2007, progress in analysis of protein-protein interactions in *R. palustris* is as follows:

Distinct Affinity-tagged Gene Products:	
Cloned (entry / expression)	1098 / 879
Cultured and harvested	537
Isolated and analyzed by LC-MS-MS	386
LC-MS-MS Identifications:	
Proteins (distinct / total)	2333 / 73481
Peptides	150288

- Each protein-protein interaction receives a score based on frequencies with which particular "bait" and "prey" proteins are observed across all measurements. [2] Figure 1 illustrates the application of this statistical approach. Details of the statistical method are available at http://www.pnl.gov/statistics/BEPro3/
- The experimental results contain both protein-protein interactions that are
  - consistent with predictions from the literature (blue ovals in Figure 1), [3] and
  - potentially novel (yellow ovals in Figure 1).
- A public website, the Microbial Protein-Protein Interaction Database, provides access to the results:
  - http://mippi.ornl.gov (registration required for full data access).
  - Results are provided in several formats for viewing and downloading (Figure 2a).
  - Statistical scores, links to other data, and additional information for each protein-protein interaction are included (Figure 2b).
  - Biannual updates to the MiPPI database are planned.

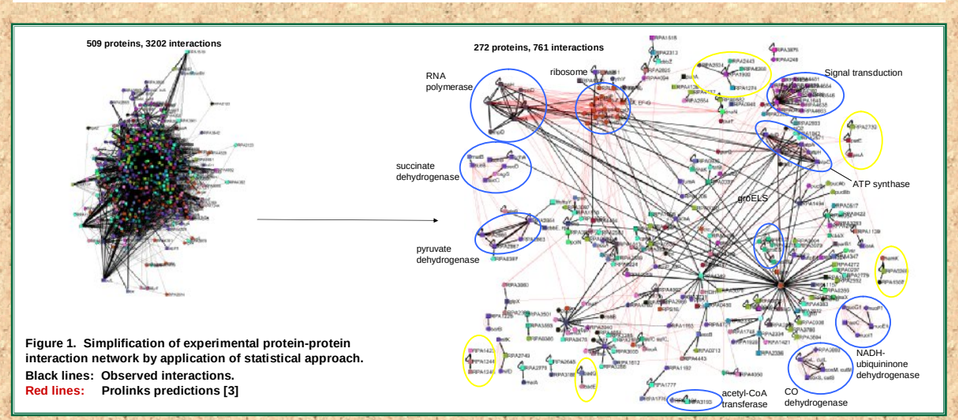
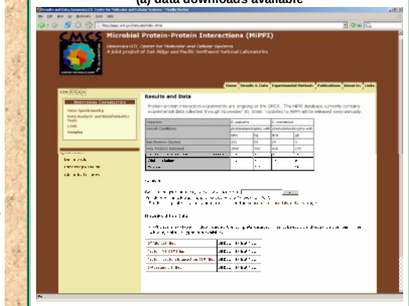


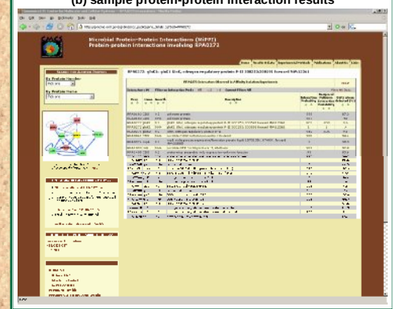
Figure 1. Simplification of experimental protein-protein interaction network by application of statistical approach.
Black lines: Observed interactions.
Red lines: Prolinks predictions [3]

Figure 2. Screen shots of mippi.ornl.gov:

(a) data downloads available



(b) sample protein-protein interaction results



CONCLUSIONS

- To date, we have successfully cloned over 1000 *R. palustris* open reading frames, purified over 300 different affinity-tagged gene products, and identified their protein interaction partners from cultures of *R. palustris* grown under anaerobic photoheterotrophic growth conditions, in the presence or absence of fixed nitrogen. Components of several known complexes have been co-isolated to validate the efficacy of the approach. Abundant complexes found in the cytosol, such as RNA polymerase, provide good coverage of known components of the core enzyme, as well as sub-stoichiometric sigma factors. However, with our approach, it has not been possible to obtain data on low-copy-number complexes, such as DNA polymerase. We have identified interactions among components of other complexes for which homologues are known in other species, including alpha ketoglutarate dehydrogenase, pyruvate dehydrogenase, fixABCX, nitrogenase, NADH:ubiquinone oxidoreductase, succinyl-Co A synthetase, tryptophan synthetase, and GroELS, as well as numerous potential novel interactions.
- Data emerging from the CMCS can confirm known interactions, as well as provide hypotheses for further study of potentially novel protein-protein interactions.
- Current challenges include continuing to ramp throughput and extending the "pipeline" to other organisms of interest.

ACKNOWLEDGMENTS

This research sponsored by the Office of Biological and Environmental Research, U.S. Department of Energy. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725

REFERENCES

1. M.E. Kovach, P.H. Elzer, D.S. Hill, G.T. Robertson, M.A. Farris, R.M. Roop and K.M. Peterson, "4 new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes." *Gene* **1995**, *166*, 175-176.
2. J.L. Sharp, K.K. Anderson, D.S. Daly, D.A. Pelletier, G.B. Hurst, W.R. Cannon, D.L. Auberry, D.D. Schmoyer, W.H. McDonald, A.M. White, B.S. Hooker, K.D. Victry, M.V. Buchanan, V. Kery and H.S. Wiley, "Inferring protein-protein associations with affinity isolation LC-MS/MS assays." *J. Proteome Res.* (submitted).
3. P.M. Bowers, M. Pellegrini, M.J. Thompson, J. Fierro, T.O. Yeates, D. Eisenberg, D. "Prolinks: a database of protein functional linkages derived from coevolution." *Genome Biology* **2004**, *5*, R35.