

The Microbial Interactome Database: An Online System for Identifying Interactions Between Proteins of Microbial Species

G.B. Hurst,¹ D.A. Pelletier,¹ D.D. Schmoyer,¹ M.B. Shah,¹ W.H. McDonald,¹ N.E. Baldwin,¹ N.F. Samatova,¹ A. Gorin,¹ B.S. Hooker,² V. Kery,² W.R. Cannon,² D.L. Auberry,² K.J. Auberry,² K.D. Victry,² R. Saripalli,² H.S. Wiley,² S.J. Kennel,¹ M.V. Buchanan¹

Center for Molecular and Cellular Systems

¹Oak Ridge National Laboratory and ²Pacific Northwest National Laboratory

The Center for Molecular and Cellular Systems (CMCS) is generating large amounts of protein-protein interaction data from microbial species through various “pipeline” protocols, as described in other abstracts. To make these data available to the scientific community, we are implementing a Microbial Interactome Database (MID) that will allow users to interact with the data at a variety of levels.

The primary data source for the MID includes results from mass spectrometric identification of proteins interacting with affinity-tagged “target” or “bait” proteins, resulting from our endogenous and exogenous pipelines. We are currently acquiring these data for *R. palustris* and *S. oneidensis*. As of December 2005, we have performed affinity tagging, isolation, and mass spectrometric analysis for over 210 distinct target proteins (“baits”) in *R. palustris*, and 53 proteins in *S. oneidensis*, with multiple replicate analyses in many cases. While rigorous and automated methods for distinguishing authentic interactors from non-specific interactions and background proteins are under development (as described in other abstracts from the CMCS), an empirical method based on filtering at several stages (individual mass spectra, comparison of estimates of interactor protein quantities with an appropriate average value over all comparable samples, frequency of observation, etc.) is currently in place. Interactors identified by this approach include some false positives (artifactual interactors), but also homologs to known complexes and potential novel interactions; experimental, literature-based, or informatic confirmation of these high-throughput results are described in other abstracts. This screen of four interaction subnetworks (DNA-directed RNA polymerase, DNA polymerase, ATP synthase and the degradosome) provides results that are consistent with homologous systems studied by other methods. For example, with α , β , and β' subunits as baits, we redundantly identify the core subunits of the RNA polymerase enzyme (α , β , β' and ω) as well as several sigma factors present only under certain conditions. For DNA polymerase, on the other hand, we have to date identified none of the expected interactors from experiments using β , τ/γ , δ , and χ subunits as baits in *R. palustris*; this result helps define the sensitivity of our current protocol, and suggests that larger culture sizes, more efficient isolation, or more sensitive detection will be required to study subnetworks representing complexes present at the level of ~10 copies per cell.

An internal project database tracks and summarizes data from all experiments, and provides access to CMCS investigators through a web interface. Publicly accessible views of selected pages of this website show a summary of the status of targeted proteins

for the endogenous approach (http://maple.lsd.ornl.gov/cgi-bin/gtl_demo/public_target_status.cgi) and for the exogenous approach (http://maple.lsd.ornl.gov/cgi-bin/gtl_demo/public_ex_target_status.cgi), and example interactor identifications for a selected subnetwork (http://maple.lsd.ornl.gov/gtl_demo/index.html).

We are currently designing a web-based resource to provide access to our results for the scientific community. Important components of this resource include presentation of our complete set of high-throughput “pipeline” results, comparison of putative interactions suggested by pipeline results with results from other databases of interacting protein and literature studies, and incorporation of results from confirmatory experiments such as live-cell imaging and surface plasmon resonance. Critical to the success of this resource will be an interactive filtering scheme to accommodate researchers with diverse scientific requirements, who will require different levels of stringency in the reliability of interactions reported, as well as different tools for extracting subsets of the data. Graphical representations (such as Cytoscape) of protein interaction subnetworks, incorporating various levels of experimental and comparative data, will be included.

Research sponsored by the Genomics:GTL program, Office of Biological and Environmental Research, U.S. Department of Energy, under contract No. DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.

This manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a paid-up, nonexclusive, irrevocable, worldwide license to publish or reproduce the published form of this contribution, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, or allow others to do so, for U.S. Government purposes.