Advanced Technologies for Identifying Protein-Protein Interactions*

R. Hettich[1], G. Hurst[1], W. McDonald[1], H. Connelly[1], D. Pelletier[1], C. Pan[1], N. Samatova[1], G. Kora[1], V. Kertesz[1], S. Gaucher[2], T. Iqbal[2], M. Hadi[2], M. Young[2] , G. Orr[3], M. Romine[3], D. Panther[3], S. Reed[3], D. Hu[3], E. Livesay[3], B. Hooker[3], S. Wiley[3], S. Kennel[1], and M. Buchanan[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]Sandia National Laboratory, Livermore, CA; [3]Pacific Northwest National Laboratory, Richland, WA

The major focus of the **Center for Molecular and Cellular Systems (CMCS**) is the establishment of a *high-throughput identification pipeline for measuring protein-protein interactions*. In the development and operation of this pipeline, several needs have been uncovered that, if addressed, would provide dramatic improvements in overall performance. In particular, improvements are needed in measurement sensitivity, enhanced accuracy, speed of identification, more comprehensive characterizations, and information management tools. The impetus for development of advanced technologies is to alleviate current bottlenecks of the pipeline as well as continue to pursue state-of-the-art technologies for enhanced measurement throughput and data quality. Recent research has focused on key developments in techniques for more comprehensively examining protein interactions, improving LC-MS throughput, applying automated detection of FRET using flow cytometry, and developing computational tools for data mining.

For more extensive characterization of the molecular details of protein interactions, MS methodologies have been developed to extract information from *intact proteins* by either direct top-down MS measurements or by chemical cross-linking of proteins with MS detection of their proteolytic peptides. As a demonstration of the top-down MS approach, tandem affinity purifications were conducted for the Gln family of proteins from *R. palustris*. These proteins are key to sensing internal cellular ammonium levels and transducing the signal. Under ammonium-rich conditions, the proteins GlnK1, GlnK2, and GlnB are unmodified; however, under ammonium-starvation conditions, these proteins are modified by uridylylation and activate the AmtB ammonium transporter and the glutamine synthetase enzyme that are important for nitrogen fixation. Top-down MS measurements provided not only details about the level of uridylyation, but also verified the presence of multiple isoforms of each protein. Chemical crosslinking, used in conjuction with enzymatic digestion and LC/MS/MS, has potential as an important tool for probing protein-protein interactions. Data interpretation is currently a challenge because the "rules" for crosslinked peptide dissociation have not been well studied. We are thus systematically investigating the dissociation pathways open to crosslinked peptides using a series of defined inter- and intra-molecular crosslinked species. Dissociation pathways unique to crosslinked peptides have been found, and are being incorporated into our MS2Assign software. This will allow for more fully automated analysis of crosslinked peptide sequence, a prerequisite for high throughput experiments.

Because of the complexity of protein-protein interactions in microbial systems, it is essential to continue development of methodologies for high throughput measurements. While LC-MS/MS shows tremendous potential for the identification and characterization of protein-protein interactions, the sequential chromatographic separations are either a current or soon to be rate-limiting step for throughput. In order to address this issue, we have undertaken the development of a protocol for fast LC-MS/MS based on a Q-ToF mass spectrometry platform. This setup includes several HPLC columns connected in parallel for multiplexing the runs and solvent washes. By conducting consecutive runs with 30-min. chromatographic separations, we have been able to examine almost 30 samples in a 24-hour period. The reproducibility and sensitivity are comparable to the longer, more conventional LC-MS/MS methodologies, in which about 12 samples can be analyzed in a 24-hour period. The power of fluorescence resonance energy transfer (FRET) for detecting molecular interactions between proteins can be applied to high throughput screening of protein-protein interactions. Using automated cloning and fluorescence detection approaches, FRET can be used to screen a large number of protein pairs, tagged

with fluorescent proteins or fluorescent antibodies. We have combined Gateway compatible cloning with the application of flow cytometry for high throughput FRET analysis in *Shewanella* to leverage existing molecular reagents. Cyan and Yellow fluorescent proteins were successfully expressed in *Shewanella* and were used as donor and acceptor fluorophores in a FRET pair. In addition, we have used FRET analysis between fluorescent antibodies, directed against *Shewanella* periplasmic proteins. We demonstrate that detection of FRET by flowcytometry enables the rapid screening of a large number of cells.

One of the biggest hurdles in the measurements of protein interactions continues to be limitations in computational approaches for data mining. Recent effort has been given to the development of computational methods for integrating top-down and bottom-up MS analyses ("PTMsearchPlus") and deciphering isotopically labeled samples for protein quantification ("ProRata"). In order to automate the interpretation of the top-down MS data, the algorithm "PTMsearchPlus" has been created and tested with several data sets. This program integrates both top-down and bottom-up MS data to exploit the advantages of each approach. The ability to combine high resolution molecular mass data on the intact proteins with the more extensive sequence coverage of the proteolytic peptides from the bottom-up data provides a powerful platform for not only identifying proteins, but also characterizing the presence of post-translational modifications and isoforms. Protein quantification is very important for not only measuring the amount of protein present, but also the elucidation of the stoichiometry of proteins present in complexes. Stable isotope labeling is one of the most widely employed methods for quantification measurements; however, current computational tools to extract the information are quite remedial. We have developed a new tool, "ProRata," which addresses not only the extraction of isotopic information from the labeled samples, but also provides confidence interval scoring for each protein identification.

* 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.