William R. Wiley
EMSL In Brief
Environmental Molecular Sciences Laboratory

New Approach Identifies Proteins with Confidence

UStag and high-performance mass spectrometers help drop false discovery rate to near zero

Identifying the thousands of proteins in a cell could become easier, thanks to a new approach developed by scientists at Pacific Northwest National Laboratory and the Department of Energy's Environmental Molecular Sciences Laboratory. This approach accurately identifies proteins and their modifications with high confidence and low ambiguity. The accurate identification provides more reliable proteomics data for those working on applications such as bioremediation or disease identification and treatment.

Yufeng Shen

Nikola Tolić

In conventional approaches, researchers use algorithms (e.g., SEQUEST) to rank the confidence of peptide and protein identifications; however, scoring unavoidably results in incorrect identifications. In the new approach, the team obtained high-quality data from EMSL's LTQ-Orbitrap mass spectrometer. They developed unique sequence tags or UStags method for processing the resultant mass spectra. They examined the top 10 SEQUEST proposed peptide sequences with high mass precision to reject all ambiguous sequences through a residue replacement filter that included all possible amino acid substitutions and potential modifications. The resulting UStags were then used to identify proteins and assign modifications. The team applied this new method to the proteome of *Saccharomyces cerevisiae* and identified a comparable number of peptides as conventional methods but with a 0% false discovery rate. Conventional methods average 5% FDR.

Scientific impact: This research provides a way to generate higher quality data for proteomics research. This research and other discoveries are part of EMSL's work in predicting biological functions from molecular and chemical data.

Societal impact: Unambiguous and highly confident peptide identifications from mass spectra can help better identify rare and important proteins, potentially improving techniques for early disease detection and monitoring changes in environmentally important organisms.

For more information, contact EMSL Communications Manager Mary Ann Showalter (509-371-6017).

Citation: Shen Y, N Tolić, KK Hixson, SO Purvine, L Paša-Tolić, WJ Qian, JN Adkins, RJ Moore, and RD Smith. 2008. "Proteome-wide Identification of Proteins and Their Modifications with Decreased Ambiguities and Improved False Discovery Rates Using Unique Sequence Tags." *Analytical Chemistry* **80**(6):1871-1882. DOI: 10.1021/ac702328x S0003-2700(70)02328-4

Acknowledgments: The National Institutes of Health's National Institute of Allergy and Infectious Disease and National Center for Research Resources along with the EMSL's 2007 Intramural Program funded this research.