Orthoganol Informatic Pipelines for Analysis of Chemically Crosslinked Protein-Protein Complexes

Alexandre Panchaud[1], Pragya Singh[1], Jimmy Eng[2], Scott A. Shaffer[1], David R. Goodlett[1*]

[1]Department of Medicinal Chemistry, University of Washington,

[2]Genome Science, University of Washington,

To facilitate characterization of protein-protein interactions using mass spectrometry, two orthoganol informatics pipelines and a new data acquisition strategy have been designed, tested and shown to accurately detect known sites of prote-protein interaction. First a high-charge-state-driven data acquisition scheme was utilized to enrich datasets for crosslinked peptides and a pre-screening database search done to remove linear peptides. This reduced tandem MS data set is screened by two informatics pipelines one of which, Popitam (Singh et al. Anal chem. 2008), uses a sequence database containing the proteins known to be in the protein complex, and the other, xComb (Panchaud et al. in review 2010), uses a peptide database consisting of a select set of all theoretically possible peptide-peptide crosslinked speices in linearized and concatenated form.

The Popitam approach considers each peptide to be “post-translationally” modified with an unknown mass at an unknown amino acid that is in fact the second peptide in the crosslinked pair. False positives are reduced and database selectivity increased by acquiring precursors and fragments at high mass accuracy. The xComb approach creates a library of a rational subset of all theoretically feasible cross-linked peptides (for a given protein set) as FASTA files. This library is then used as a database that is examined by any standard proteomic search engine to match tandem mass spectral datasets to identify cross-linked peptides. By searching against a peptide library of linearized, cross-linked peptides, rather than a linearized protein library, search times are decreased and the process is decoupled from any specific search engine. A further benefit of decoupling from the search engine is that protein cross-linking studies may be conducted with readily available informatics tools for which scoring routines already exist within the proteomic community. We will present data from protein complexes and viral structures.

Sponsor: NWRCE