Introduction

Crucial structural information about protein-protein interactions can be obtained using chemical crosslinking methods in mass spectrometry. In practice, these data are incredibly complex to analyze and quantify, requiring extensive and time-consuming validation to eliminate false positives or low confidence assignments. Here, we present a new rapid approach to analyze crosslink data using Byonic to identify crosslinked peptides and Byologic to facilitate data validation. This approach reduces the validation time for typical protein-protein systems from hours to minutes, and provides portable and robust data output to aid future investigations.

Using this strategy, we have investigated multiple cross linked protein-protein interactions to determine binding interfaces, stoichiometry, and replicability. The Byonic search produces lists of possible peptide linkages containing thousands of potential cross links. Byologic readily condenses these data into related peptide groups, e.g., cross linked peptides with their non-crosslinked homologs, even across replicate experiments or across experimental conditions. This facilitates rapid analysis for cross-link validation at both MS1 and MS2 levels, and easy label-free quantification using XIC analysis. We propose empirical rules, such as peptide length, number of observations, peptide assignment scores, and chromatographic profiles, and demonstrate a method to assign and filter data using ‘comment’ labels to define classes based on analyst confidence. Additionally, data that are not relevant to the current analysis (such as peptides without cross links) can be filtered without loss and investigated at later times based on new developments. Once classifications are made, Byologic produces permanent and flexible report outputs for information transfer between researchers.

Methods

Cross linked peptides are categorized into three groups: true-positives, uncertain, and false-positives, using a set of empirical criteria detailed below. “True positives” succeed on all of these rules, while “False Positives” fail at least three of these rules. This strategy does not guarantee that a given cross link is or is not ‘real,’ but provides an easy transferable, and repeatable framework for analysis. True positives are then used to anchor structural determination, with uncertain assignments filling in based on 3D structural limitations. Thus, only the most reliable mass spectral data is used to complement other structural measurements.

Data analysis workflow:

---

Discussion and Conclusions

Workflow reduces lists of unsorted cross links that number in the 100s to a manageable list of crosslinks of links in the 10s to produce a list of solid leads for structural assignment.

The team has demonstrated one can integrate data from multiple linker types and multiple digest types.

Based on a list of strong leads, analysts can derive initial structure, and can validate either across linkers or with the ‘uncertain’ matches which fit the alignment.

Our integrated software workflow allows exhaustive analysis of these proteins. To date, these analyses have proven to be about an order of magnitude faster for interpretation (ca 1-2 hours as opposed to about ca 2 days with a traditional data analysis approach).

Acknowledgment

Support is gratefully acknowledged from NIH, grant GM100634 and the Stanford Dean of Research.