Disulfide bonded dipeptide analysis with PEAKS and Q-TOF mass spectrometry

Jiaxi Wang[1], Bin Ma[2], Weiwu Chen[2] *1-McMaster University, 2-BSI*

Bioinformatics Solutions Inc.

Introduction

Proteins and peptides are commonly studied using mass spectrometry; however, the most commonly used tools for MS data analysis are built with the assumption that peptides are linear. Disulphide bonds, creating complexes involving two or more peptides bonded together, cause problems for this kind of analyses. Chemical reduction, using 1,4-dithiothreitol (DTT), can break the disulfide bonds, making the peptides acceptable for standard analysis. But since this makes determination of the disulphide bond location more ambiguous, analysis of intact dipeptides becomes necessary. Also, since chemical reduction can be incomplete, even reduced samples can benefit from this analysis.

Here we present an algorithmic solution for the analysis of MS/MS data of disulfide bonded dipeptides.

Approach

In a traditional MS/MS ions search engine, a peptide library is constructed from the sequences of proteins in a reference database. If a specific protease, such as trypsin, is used, the peptide library can be restricted to those peptides conforming to the digest rules, but such restrictions are not requisite. The peptides in the library are then theoretically fragmented. By comparing the theoretical fragments to an actual MS/MS spectrum, we may be able to conclude a match, thereby identifying the peptide.

The anlysis of dipeptides is confounded by increased complexity of theoretical fragmentation, and the fact that disulphide bonds is unknown. However, with efficient computation, we can employ a philosophy similar to that of a traditional search engine.

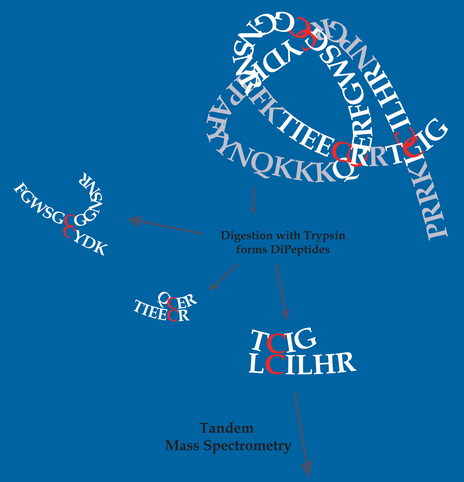
For each protein sequence in a FASTA database, every combination of Cysteine to Cysteine connection is considered, and a library of dipeptides is constructed. For a given MS/MS spectrum, each dipeptide in the library that matches in mass to the experimental peptide is theoretically fragmented. Theoretically derived fragments include traditional y and b ion types, as well as y and b ions that include part of the other strand of the dipeptide.

The candidate's theoretical fragmentation is matched to the experimental spectrum, and a candidate score is produced. To avoid bias towards longer peptides, the candidate score is divided by the number of residues in the dipeptide. The candidate with the highest resulting score is considered the best.

Data

Bovine insulin (Sigma Aldrich) and α-dendrotoxin (Sigma Aldrich) were prepared and digested with chymotrypsin (Yen et al [1]) and trypsin (Belva et al [2]). A chymotryptic and three tryptic disulfide-linked dipeptide were identified and sequenced by Nano-spray Q-TOF Global Ultima (Waters, Micromass).

The MS/MS data were searched using the algorithm. To guard against random matches, the whole NR database was searched, and no conditions were set as to which residues could appear at the N or C termini. That is to say, no enzyme was specified for the search, so chymotryptic or tryptic digest rules where not enforced.



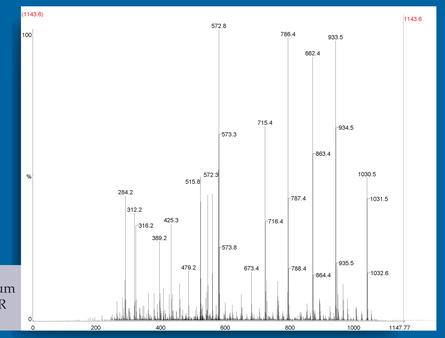


Figure 1: Tandem Mass Spectrum of the DiPeptide TCIG|LCILHR

Candidate Elution

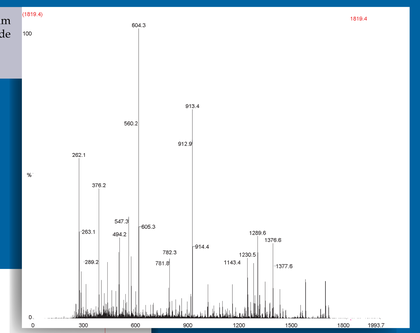
Without prior knowledge of which DiPeptide is represented in a mass spectrum, one must consider all possible dipeptides. Prior knowledge of α-dendrotoxin tells us the three possible DiPeptides, but for most practical applications, the locations of the disulphide bonds -- i.e. which cysteine bonds to which -- is unknown. The algorithm must consider every possible bond location combination.

For this protein there are 15 combinations, (shown to the right) but if the protein is larger, having even 10 cysteine resides, there could be 45 possible DiPeptides. Longer sequences could result in hundreds.

Enzyme specificity presents a further problem, since tightly folded proteins may limit the effectiveness of a digest. Whereas we can easily pick pairs of tryptic or chymotriptic peptides to form our set of DiPeptide candidates, we expand on this set exponentially if we must further consider all cleavage sites.

FGWSGCGGNSNR|CYDK
QCER|TIEECR
TCIG|LCILHR
FGWSGCGGNSNR|TIEECR
QCER|LCILHR
TCIG|CYDK
FGWSGCGGNSNR|LCILHR
QCER|CYDK
TCIG|TIEECR
FGWSGCGGNSNR|QCER
FGWSGCGGNSNR|TCIG
QCER|TCIG
TIEECR|CYDK
TIEECR|LCILHR
CYDK|LCILHR

Figure 2: Tandem Mass spectrum of DiPeptide CYDK|FGWSGCGGNSNR



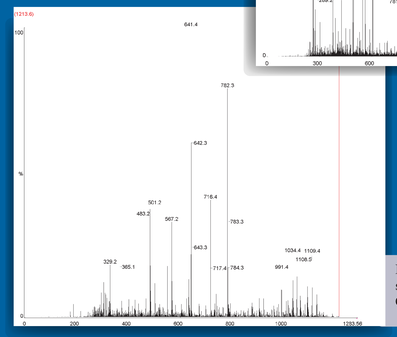


Figure 3: Tandem Mass spectrum of DiPeptide QCER|TIEECR

Results and Conclusions

Of the dipeptides corresponding to the three disulphide bond locations, we were able to identify to correct sequence and correct bond site in two cases and a total of five replicates. In the third case, interference between two precursor ions close in m/z lowered the quality of the MS/MS spectra. As such, the correct answer was not listed first, but had the search been restricted to this protein sequence, the algorithm would have identified it correctly.

The ability of this algorithm to identify the correct dipeptides from a very large number of possiblities means that we draw strong conclusions when identifying dipeptides from the sequences of a few known proteins. The algorithm should be useful for a variety of protein characterization research applications.

References

1. Yen T-Y, Yan H. and Macher B., Characterizing closely spaced, complex disulfide bond patterns in peptides and proteins by liquid chromatography/electrospray ionization tandem mass spectrometry (Journal of Mass Spectrometry, 37, 15-30, 2002).

2. Belva, H., Valois, C., and Lange, C., Determination of disulfide bonds in highly bridged alpha-dendrotoxin by matrix-assisted laser desorption/ionization mass spectrometry (Rapid Communications in Mass Spectrometry, 14, 224-229, 2000).