Disulfide Mapping Validation Using Label-Free Differential Analysis

Michael Athanas[1], Scott Peterman[2], David Sarracino2, Bryan Krastins[2], Amol Prakash[2], Taha Rezai[2], Mary F Lopez[2]
[1]VAST Scientific, Cambridge, MA; [2]Thermo Fisher Scientific BRIMS Center, Cambridge, MA

Overview

Purpose: To discover disulfide bonds in a purified protein sample.

Methods: Samples were analyze by comparing a fraction of protein digested under native conditions with a fraction digested under reduced-alkylated conditions using LC-MS and label-free analysis software.

Results: We have demonstrated a workflow for validating and potentially discovering unexpected disulfide bonds using label-free LC-MS methods. The method robustness appears to be sensitive to sample preparation.

Introduction

Disulfide cross-linkages formed by the oxidation of the thiol groups of cysteines play an important role in the folding and stability of many proteins. Knowledge of the disulfide linkages in a protein provides insight into protein structure. Determination of the disulfide bonds in recombinant or synthetic proteins is also important, since formation of the correct disulfide cross linkages are an indication of proper folding and function. Experimentally, protein structure can be determined with time consuming methods such as X-ray crystallography or NMR spectroscopy.

In this presentation, we discuss a method of determining disulfide bonds using LC-MS and label-free differential analysis. The method is based upon accurate mass derived from full-scan data – no MS/MS fragmentation data were used.

When using accurate-mass information for identification assignment, mass degeneracy is often a concern; that is, molecules with the same or similar mass may introduce ambiguity. Because of the mass accuracy of the Orbitrap™-based mass spectrometer and the simple samples (purified proteins) used in this example, mass degeneracy was never a problem.

The full-scan accurate mass approach is preferred because the data are easier to interpret and are independent of the identification efficiency of fragmentation interpretation software such as Mascot™ or SEQUEST®.

We present a work-in-progress wherein the sample we have chosen are pedagogical; however, well studied and understood. Our approach in developing this method is to start with simple proteins where we can validate with previously reported results.

Samples

In this presentation, we emphasize the analysis methods rather than the actual results. We focus upon a simple protein sample of beta lactoglobulin of length 178 and a mass of about 20 kDa. Beta lactoglobulin contains six cysteine and three bonds that have been previously reported:

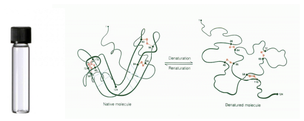
Disulfide bond	82 ↔ 176	Ref.13 Ref.19	
Disulfide bond	122 ↔ 137	Alternate Ref.13	
Disulfide bond	122 ↔ 135	Ref.13 Ref.19	

Uniprot ID P02754

Sample Preparation and Acquisition

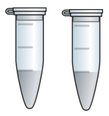
Purified protein samples are weighed and dissolved in a buffer of 8M Urea 150mM Tris-HCl 2.5% n-propanol, pH 8.5 to denature the proteins.

Denatured purified protein



Two equal fractions are constructed. The first fraction is reduced with DTT to break disulfide bonds between cysteines. The sample is then alkylated with iodoacetic acid to modify all cysteines in a way that prevents disulfide bonds from reforming. Enzymatic digest agent is then added to the first fraction for protein digestion into peptides. The second fraction is only treated with the digestion enzyme.

Reduced Alkylated Digest



Digest

The fractions are then divided into four equal volume amounts and placed in a 96-well plate for LC-MS acquisition.

Samples are injected at 150 µL/min onto through a Thermo Scientific Hypersil Gold 5 micron 1 mm x 150 mm column using a Thermo Scientific Surveyor Autosampler with a Thermo Scientific MS pump. Solvents were Fisher Scientific Optima LC-MS grade water with 0.2% formic acid buffer A, acetonitrile 0.2% formic acid buffer B. A gradient from 3% B to 30% B was used to separate the peptides. The ion source was a Thermo Scientific Ion Max source with a standard ESI probe with a 32-gauge metal needle.

The Thermo Scientific Exactive mass spectrometer used for the analyses was tuned on Angiotensin I, teed into a 150-µL/min flow of 15% B. Tune parameters were 20 on sheath, 3 on aux, with a spray voltage of 3.8 kV and a capillary temp of 300 °C. Scan parameters, 60K resolution with 3 microscans and lock mass enabled. The divert valve was used during the first 5 minutes to prevent salts and urea from the digests from contaminating the source.

Database Construction

The entire protein sequence is digested *in silico*. The digested data set is processed to identify all possible peptides with intermolecular and intra-molecular disulfide bonds. Disulfide-linked peptides are expected to possess an increased net positive charge at pH 3.0. For this reason, we consider up to 5 missed cleavages. In addition, we allow for dynamic carboxymethyl modifications of cysteines.

A new database is constructed that considers all possible combinations of intra-disulfide bonds and is combined the single peptides along with the combinations. The combinatorics may be large for a large number of cysteine. The mass of the peptide with the intermolecular disulfide bond is the combined mass of the two peptide chains minus the mass of the two liberated protons.

(1) (R)LSFNPTQLEEQ**C**HI(-) — M1=1658.784
(K)WENGE**C**AQK(K) — M2=1064.447

$M12 = M1 + M2 - 2*M_H$

We also consider a combination of intra-molecular disulfide bonds that may lead to various structural isomers (Figure 2), but isomers are limited to 5 cysteines per peptide with 2 intra-molecular bonds.

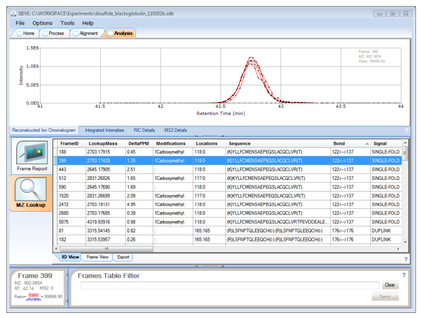
(2) ((K)YLLF**C**MENSAEPEQSLA**C**Q**C**LVR(T) — M = 2703.178 with 1 carboxylmethol modification

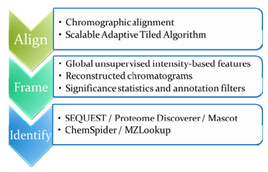
It is not possible with accurate-mass information alone to determine which cysteines are bonded and the correct structural isomer.

The resulting database file is a simple CSV table that can be viewed in Excel and imported readily into the Thermo Scientific SIEVE software. The key column in the database is the zero charge accurate mass value of the peptide / pair.

Data Analysis

The acquired data were processed using SIEVE™ software version 1.3 in a "A vs B" differential analysis experiment with technical replicates. The SIEVE workflow first performs a chromatographic alignment using full scan spectra shape (no peaks). Frames (a.k.a. features) are constructed based upon prominent peaks throughout the full data set. Identification assignment is based upon the accurate mass database described above using the SIEVE *MZLookup* tool. Replicates are required to have a CV<20%.





A disulfide folded peptide displayed in sequence (2) is shown in this view of SIEVE where the red signal is from the digest fraction.

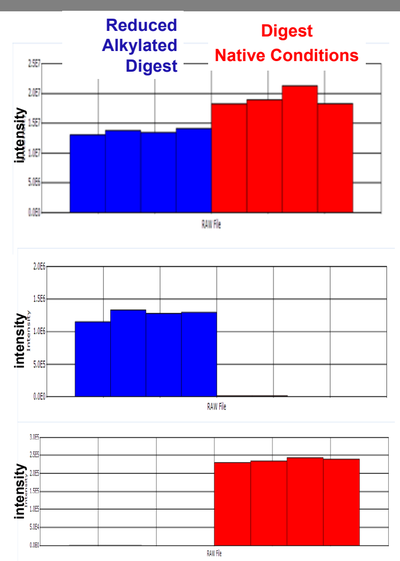
SIEVE

FIGURE 1. SIEVE results demonstrating characteristic intensity signal shapes of four technical replicate injections for A) Non-cysteine containing peptides, B) Reduced cysteine peptides, and C) Native cystine containing peptides.

A. Non-cystiene containing peptide – expect both expression levels to be similar

B. Reduced cysteine peptide – modifications will generally not appear in the natively digested sample

C. Native cystine containing peptide – carboxymethyl modifications will generally not appear in the RAD sample



Results

As can be seen in Figure 2, the label-free differential analysis method has complete coverage. The bond between cysteines 82← →122 is confirmed – a bond not reported in literature.

The bonds between 122← →137, 122← →135 and 135← →137 are generally confirmed but cannot be distinguished since this peptide is not cleaved between cysteines. The bond between 176← →82 is also confirmed.

FIGURE 2. The coverage of reduced cysteine peptides, non-cysteine containing peptides, and peptides containing disulfide bonds for beta lactoglobulin

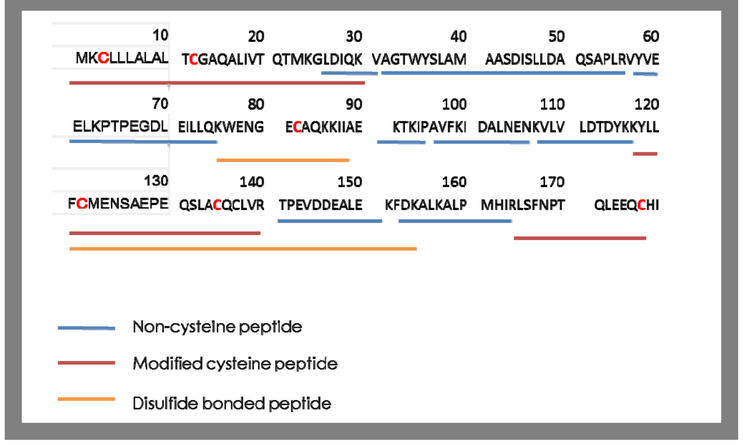
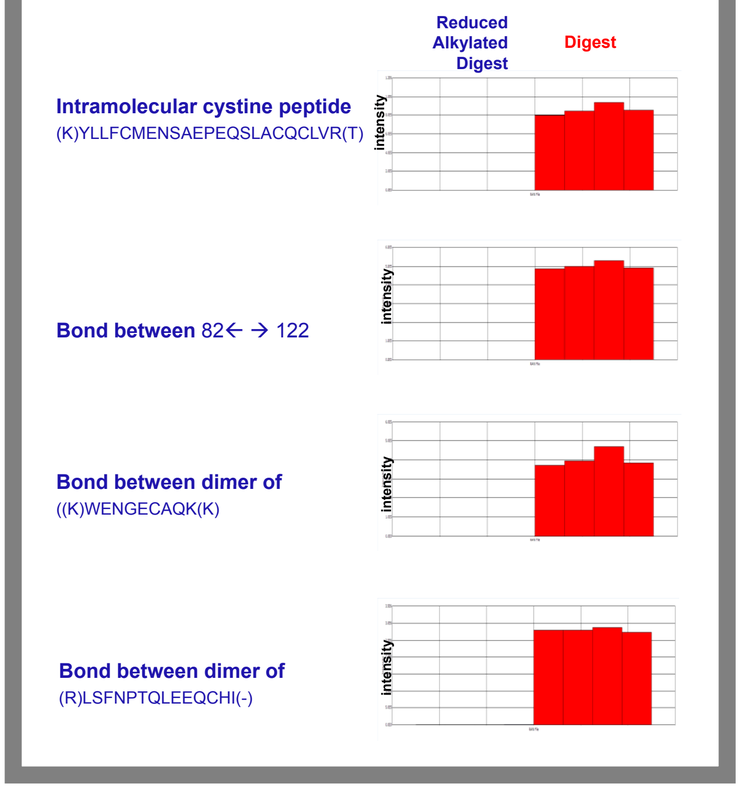


FIGURE 3. Disulfide bond signal intensities from SIEVE differential analysis.



Conclusion

We have demonstrated a workflow for validating and potentially discovering disulfide bonds using label-free LC-MS methods. The method works well for peptides that are well separated by a enzymatic cleavage point; however, it is not possible to distinguish intra-molecular bonds in peptides with three or more cysteines.

The observation of prominent signals for peptide dimers may be an indication of a fault with the sample preparation chemistry. Even though the samples were treated with additional DTT post reduction alkylation, to both quench the alkylation, and return the sampled to a reduced state, non-reduced cystine containing peptides may remain.