

# Disulfide mapping validation using label-free differential analysis

Michael Athanas<sup>1</sup>, Scott Peterman<sup>2</sup>, David Sarracino<sup>2</sup>, Bryan Krastins<sup>2</sup>, Amol Prakash<sup>2</sup>, Taha Rezai<sup>2</sup>, Mary F Lopez<sup>2</sup>

<sup>1</sup>VAST Scientific, Cambridge, MA; <sup>2</sup>Thermo Fisher Scientific BRIMS Center, Cambridge, MA;

## Overview

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

**Methods:** Samples were analyzed by comparing a simple "digest only" fraction with a "reduced-alkylated-digested fraction" using LC-MS and label-free analysis software.

**Results:** We have demonstrated a workflow for discovery and potential verification of disulfide bonds using label-free LC-MS methods.

## 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 for the identification of disulfide bonds using LC-MS and label-free differential analysis (ThermoFisher SIEVE V1.3). 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 instrument (ThermoFisher Exactive) and the simple samples (purified proteins), mass degeneracy is not a problem.

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

We present a preliminary work using proteins that are well studied and understood. Our approach in developing this method was to start with standards that could be validated with previously reported results.

## Samples

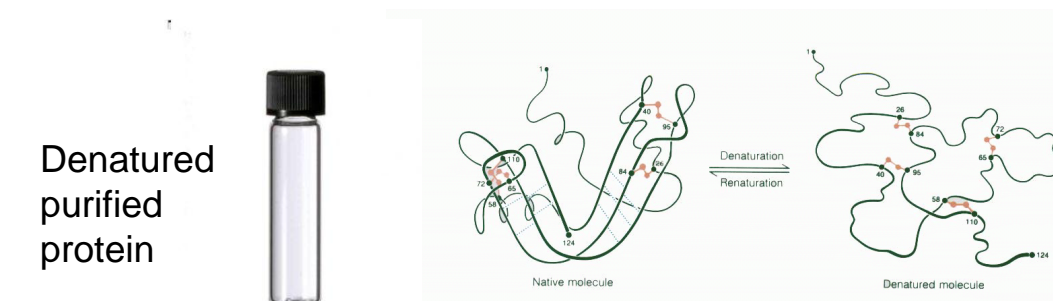
In this presentation, the emphasis is on the methods used in the analysis. We focused on beta lactoglobulin, 178 aa long with a mass of ca 20kDa. Beta lactoglobulin contains six cysteines and three bonds that have been previously reported as shown below:

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

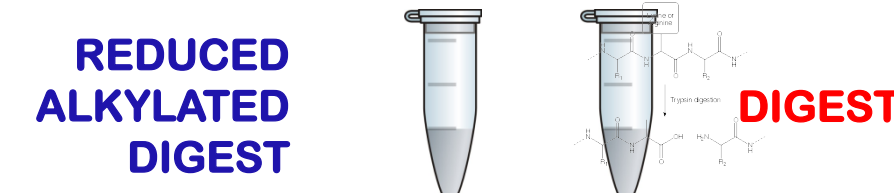
Uniprot ID P02754

## Sample Preparation and Acquisition

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

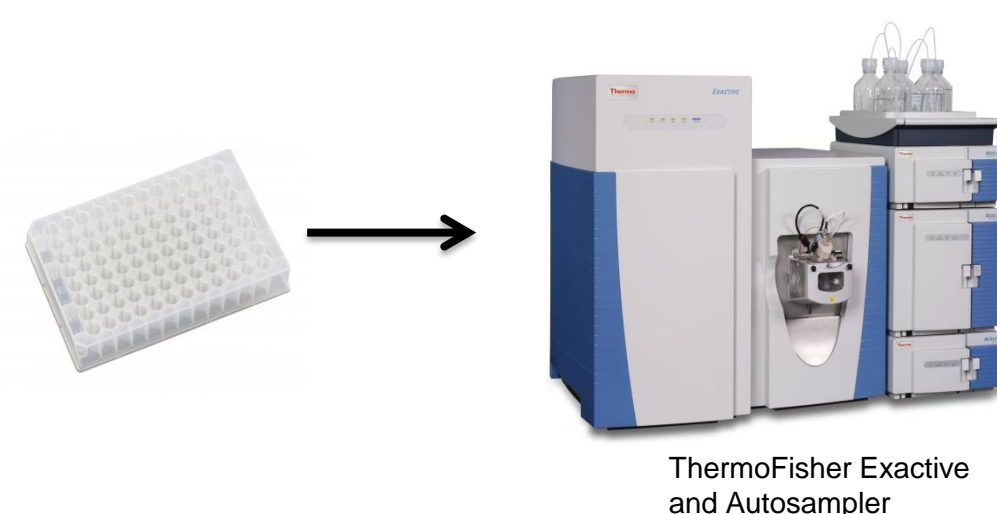


Two equal fractions were constructed. The first fraction was reduced with DTT to break disulfide bonds between cysteines. The sample was then alkylated with iodoacetic acid to modify all cysteines in a way that prevented disulfide bonds from reforming. Trypsin was then added to the first fraction for protein digestion into peptides. The second fraction was only treated with trypsin.



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

Samples were injected at 150uL/min onto through a Hypersil gold 5micron 1mmx150mm column using a Thermo Surveyor Autosampler with a Thermo MS pump. Solvents were Fisher 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 Source was a Ion Max with a standard ESI probe with a 32 gauge metal needle.



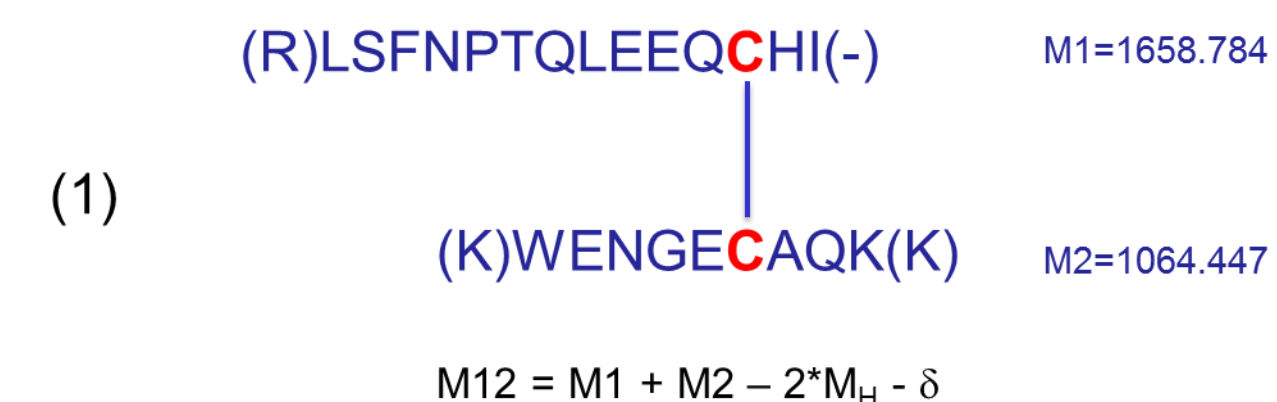
The machine was tuned on Angiotensin I, T'ed into a 150uL min flow os 15% B. Tune parameters were 20 on sheath, 3 on aux, with a spray voltage of 3.8kV and a capillary temp of 300degC. Scan parameters, 60K resolution with 3 microscans and lock mass enabled. The divert valve was used during the first 5 min to prevent salts and urea from the digests from contaminating the source.

## Database Construction

The sequence of the purified protein was digested using an *in silico* digest algorithm. Disulfide-linked peptides were expected to possess an increased net-positive charge at pH 3.0, therefore, we allowed for 5 missed cleavages. In addition, we allowed for dynamic carboxymethyl modifications of cysteines.

1Q	2Q	3Q	4Q	5Q	6Q
MKCLLLALAL	TCGAQALIVT	QTMKGLDIQK	VAGTWYSLAM	AASDISLLDA	QSAPLRVYVE
7Q	8Q	9Q	10Q	11Q	12Q
ELKPTPEGDL	EILLQKWENG	ECAQKKIIAE	KTKIPAVFKI	DALNENKVLV	LDTDYKYYLL
13Q	14Q	15Q	16Q	17Q	
FCMENSAPPE	QSLACQCLVR	TPEVDDEALE	KFDKALKALP	MHIRLSFNPT	QLEEQCHI

The digested data set was processed to identify all possible peptide pairs that were consistent with disulfide bonds. A new database was constructed including the single peptides and peptide combinations. The resulting combinatorics may be very large when a large number of cysteines are included. The mass of the cysteine pair is calculated as: The combined mass of the two peptides, minus the mass of the two liberated protons, minus a systematic shift that may be consistent with a 2-S binding energy (empirically determined as 0.017 Da).



In addition to peptide pairs, we considered the possibility of folded peptides; that is, a peptide with more than two cysteines that had a mass shift consistent with two bonded cysteines.

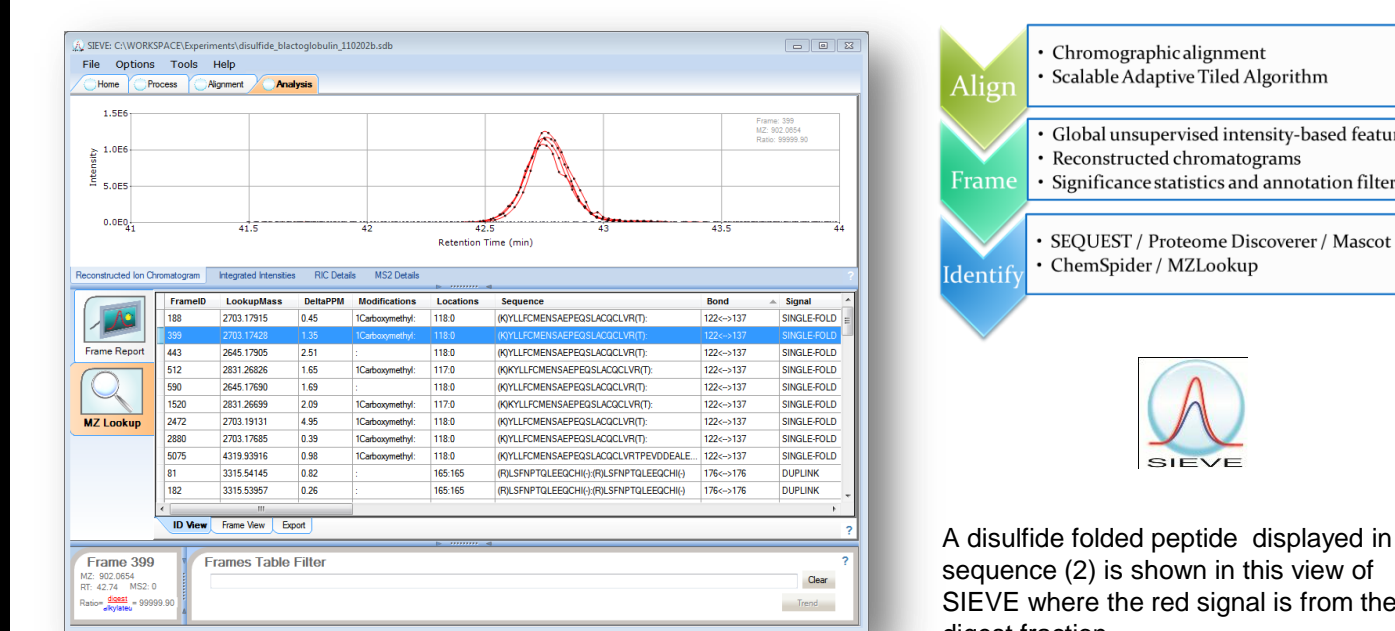


In the case where there are three cysteines, one of the cysteines may be modified. It is not possible to determine which cysteines are bonded with accurate mass information alone.

The resulting database file is a simple CSV table that can be viewed in Excel and imported readily into SIEVE. 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 V1.3 in an "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%.



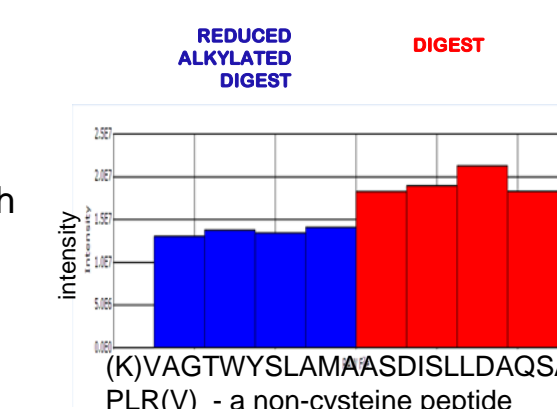
A disulfide folded peptide displayed in sequence (2) is shown in this view of SIEVE where the red signal is from the digest fraction.

Protein / peptide characterization was carried out according to the following steps:

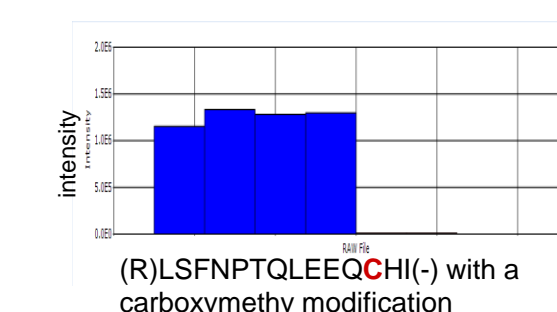
- Find all disulfide bonded peptide pairs or single internally bonded individual peptides. A positive indication is a signal in the digest-only fraction and absence in the reduced alkylated digested (RAD) fraction.
- For each disulfide peptide:
  - Find each single peptide with a carboxymethyl modification(s).
  - Find each single peptide without carboxymethyl modifications.

The last two steps were used to validate the consistency of the disulfide peptide.

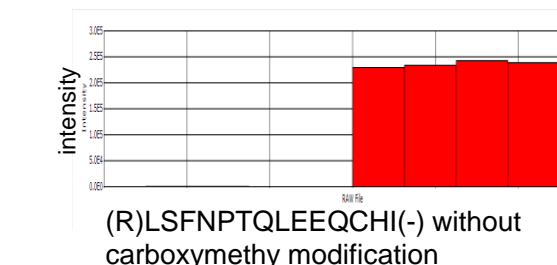
**Non-cysteine peptide** – expect both expression levels to be similar



**Modified cysteine peptide** – modifications will generally not appear in the digested sample

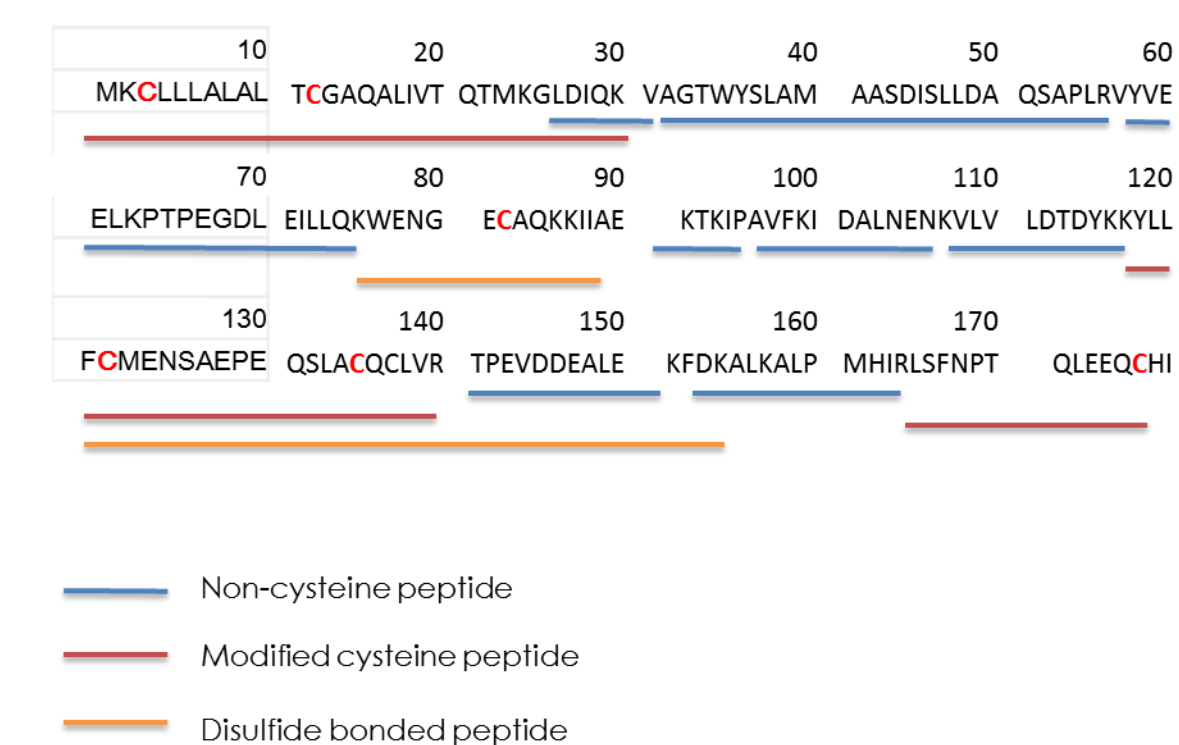


**Unmodified cysteine peptide** – carboxymethyl modifications will generally not appear in the RAD sample

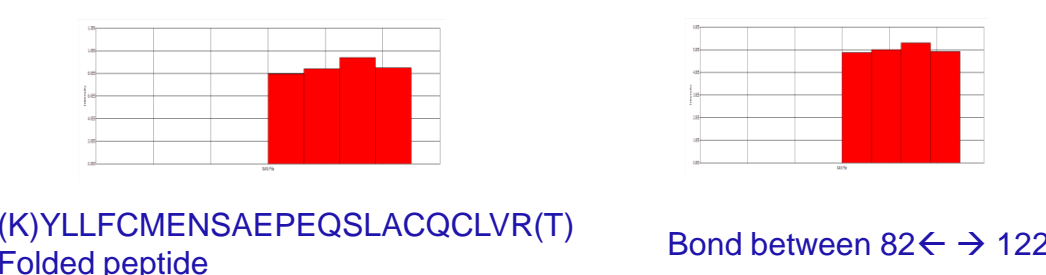


## Results

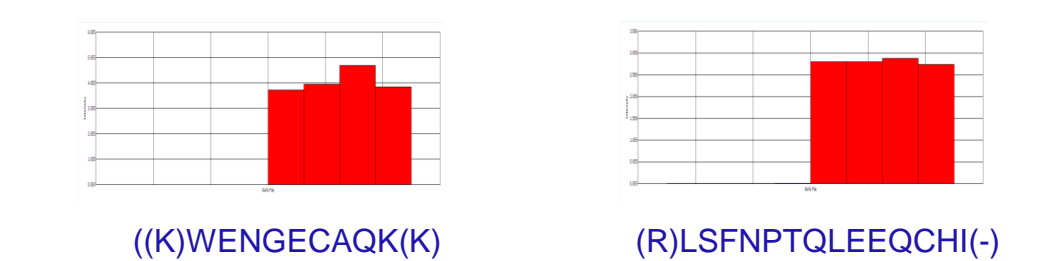
Cysteine peptides, non-cysteine peptides, and disulfide bonded peptides for beta lactoglobulin are shown below.



As is evident from the data, the method provides complete coverage. The bond between cysteines 82 ↔ 122 (not previously reported in the literature) is confirmed. 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.



In addition to intra-cysteine bonds, we also observe clear signals for peptide dimers:



## Conclusion

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

One caveat in this method is the observation of prominent signals for peptide dimers. This may be an indication of insufficient sample preparation chemistry. Although the samples were treated with additional DTT post reduction/alkylation, (in order to quench the alkylation reaction and return the sampled to a reduced state), unalkylated cysteine containing peptides apparently remain.