

The samples were dissolved in 25 μL of 100 mM ammonium bicarbonate buffer and vortexed well. A total of 5 μL from each sample was used for intact protein analysis. The remaining 20 μL of each sample was used for enzymatic digestions.

Enzymatic digestion:

Each sample was split equally (10 μL /tube) into 2 tubes. One sample was digested with trypsin and the other tube was digested with LysC. The digestion protocol was the same for both enzymes and involved first diluting the sample 1:1 with 2,2,2-trifluoroethanol, then reducing with DTT at 90°C for 20 min. Alkylation was done using iodoacetamide (IAM) for 1 hour at room temperature followed by a quenching step with DTT to remove excess IAM. The enzyme samples were then diluted 10-fold with 25 mM ammonium bicarbonate and enzyme was added (1:50 enzyme:substrate ratio). Following an overnight incubation at 37 °C, the digestions were quenched using 1 μL of TFA.

Intact protein analysis:

For intact protein analysis, the samples were analyzed by both nanoESI LC/MS and a Bioanalyzer using a protein sizing assay (Agilent Protein 230 assay). The Bioanalyzer results indicated 1 protein in sample A and 3 proteins in sample B, with one of the proteins in B being the same mass as the sample A protein. The LC/MS analysis was done on an Agilent 6520 Q-TOF using the HPLC-Chip/MS interface with a C8 HPLC-Chip. The mass spectrometer was operated in MS only mode and data was collected from 400 to 3200 m/z. Sample A showed one major peak and sample B showed 3 major peaks. Deconvolution of the mass spectra for the major peaks showed that the major peak in sample A was also in sample B (33162.83 Da). The two other major peaks in sample B were at 12522.36 Da and 24581.23 Da. It should be noted that the deconvolution algorithm found more than a dozen minor variants of the 3 major proteins, but only the major proteins were reported as per the instructions.

Database searching workflow:

- 1) Perform database search in Spectrum Mill using standard Q-TOF search parameters using the appropriate enzymatic specificity (either trypsin or LysC) against SwissProt.
- 2) Create saved results database containing the AGER protein
- 3) Search against saved results database using “No enzyme” specificity to determine the potential C-terminal peptides.
- 4) Search against saved results using 600 ppm fragment tolerance to account for larger peptide fragments.
- 5) Search against saved results using “Unassigned single mass gap” to find potential unlocalized modifications.
- 6) Search against fixed N-terminal GSHM modification using No enzyme specificity to determine the N-terminal peptide.

Total sequence coverage was determined by combining the results for the A and B samples within Spectrum Mill’s Protein/Peptide Summary view.

The fully covered sequence (23-237 from the SwissProt entry Q15109, 75% sequence coverage of the database entry) with GSHM appended to the N-terminus was copied into

the Sequence Editor-Matcher application and it was determined that this theoretical protein mass (33169.13) was 6 Da higher than the measured average mass of 33162.82 Da for the major protein in the A sample and the equivalent protein in the B sample. It was also seen that there were 6 cysteines, so we postulated that there could be 3 disulfide bonds that lowers the mass by 6 Da.

To determine the sequences for the other two proteins in the sample B, we looked for abundant singly LysC or trypsin specific peptides in sample B that were not in sample A. We found the peptides (R)VRVYQIPGKPEIVDSASE(L) in the trypsin data that corresponds to the amino acid locations 115-132. This was taken as one of the possible terminal peptides of the smallest protein. When truncating the sequence to GSHM-(23-132) within the Sequence Editor Matcher, the calculated intact protein average mass of 12524.50 was 2 Da larger than the measured mass of 12522.69. The addition of a disulfide would reduce the mass to 12522.49, which matches closely to the measured mass.

Another abundant no enzyme peptide match that was found is (R)VWEPVPLEEVQLVVE(P), corresponding to the amino acid locations 229-243. Placing the GSHM-23-243 sequence in the Sequence Editor-Matcher, the theoretical average mass is 24585.39 is ~4 Da larger than the measured average mass of 24581.23. With 2 disulfide bonds, these two masses would match.

The SwissProt entry shows that 1-22 is a putative signal peptide, so our results match this idea. It also shows that there are 3 potential disulfide bonds and our data matches this as well. Thus, in summary, we found the following proteins:

Sample A:

GSHM + SwissProt entry AA #'s (23-327) + 3 disulfides: 33163.08 Da, measured = 33162.83

And

Sample B:

GSHM + SwissProt entry AA #'s (23-327) + 3 disulfides: 33163.08 Da, measured = 33162.83

GSHM + SwissProt entry AA #'s (23-243) + 2 disulfides: 24581.36, measured = 24581.23

GSHM + SwissProt entry AA #'s (23-132) + 1 disulfide: 12522.49, measured = 12522.36