Supplemental Information

Protein identification and validation using SEQUEST.

Protein identification was carried out as explained in the materials and methods section. The statistical analysis performed was based in the two-variable Gaussian method[1] with minor modifications. In this method the Xcorr of each peptide was normalized as a function of the peptide length, as used in PeptideProphet[2]. For a regular search two Gaussian clusters will be observed in a histogram of corrected Xcorr and ΔCn scores. It is assumed that for each score a higher value means a higher correlation between a theoretical tandem mass spectrum of a peptide sequence and an experimental one. According to the method's statistical model, clusters of lower scores correspond to the random peptide sequence assignments (false positive matches) and the second, higher score distribution to positive matches. The mean and the standard deviations from the score distributions were determined by least square fitting using the following Gaussian Mixture Model (GMM) using Microsoft Excel spreadsheets.

$$g(XCc) = \frac{1}{\sqrt{2\pi\sigma_{X-}}}\exp\left(-\frac{XCc-\mu_{X-}}{2\sigma_{X-}^2}\right)*(1-\alpha)+\frac{1}{\sqrt{2\pi\sigma_{X+}}}\exp\left(-\frac{XCc-\mu_{X+}}{2\sigma_{X+}^2}\right)*\alpha \quad (1)$$

$$XCc = \frac{\ln(XCorr)}{\ln(peptide-length)} \quad (2)$$

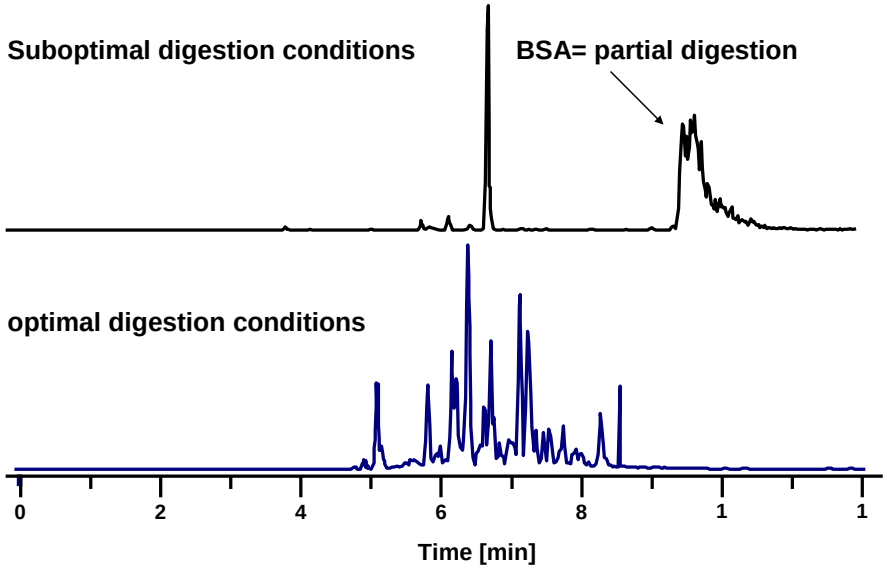
Were α is the weight of the true positives in the overall score distribution.

$$g(DCc) = \frac{1}{\sqrt{2\pi\sigma_{D-}}}\exp\left(-\frac{DCc-\mu_{D-}}{2\sigma_{D-}^2}\right)*(1-\alpha)+\frac{1}{\sqrt{2\pi\sigma_{+}}}\exp\left(-\frac{DCc-\mu_{D+}}{2\sigma_{D+}^2}\right)*\alpha \quad (3)$$

$$DCc = \sqrt{\Delta Cn} \quad (4)$$

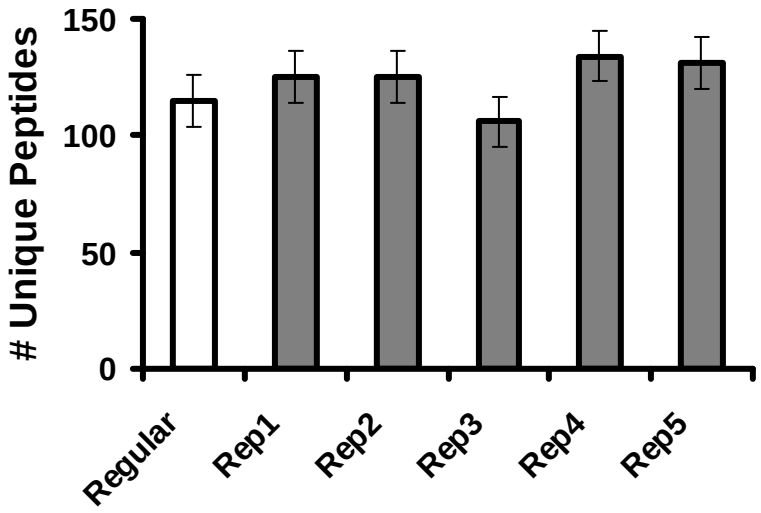
Scoring of these distributions has been shown to be independent of charge state, but for this analysis we considered them as dependent to get a better estimation of the overall score distribution for each charge state. The false discovery rate was estimated as explained before[1] using only the parameters of the random match cluster for easy of calculation.

Evidence of incomplete digestion through the LC-MS chromatogram



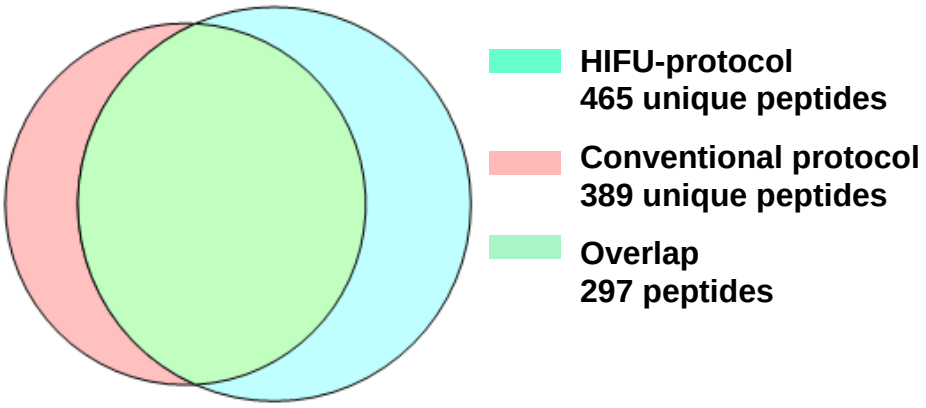
Supplementary Figure 1. Base-peak chromatograms corresponding to the LC-MS/MS analysis of two BSA digestions under suboptimal conditions (pH<4) and using optimal conditions (pH≈ 8.2).

Reproducibilityof the digestion process using four standard proteins



Supplementary Figure 2. Histograms showing the number of total unique peptides identified in each run for 4 different proteins during five technical replicates after 3 min reduction and alkylation time and 1 min digestion time using ultrasound irradiation (grey bars) and compared to the regular protocol (white bar), no replicates were done as mentioned in the text.

Application of HIFU for the in-solution digestion of a global proteome



Supplementary Figure 3. Venn diagram showing the peptide overlap obtained after LC-MS/MS analysis of mouse plasma digestions using either HIFU assisted digestion or conventional 37 ºC overnight digestion.

References

1. Lopez-Ferrer, D.; Martinez-Bartolome, S.; Villar, M.; Campillos, M.; Martin-Maroto, F.; Vazquez, J., Statistical model for large-scale peptide identification in databases from tandem mass spectra using SEQUEST. *Analytical Chemistry* 2004, 76, (23), 6853-60.

2. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R., A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Analytical Chemistry* 2003, 75, (17), 4646-4658.