

## Supplementary Figure S1

Data for Figures S2 and S3 were generated from whole proteome datasets for *Shewanella oneidensis*, *Escherichia coli* and *Rhodospseudomonas palustris*. The proteome datasets from these organisms were analyzed with Sequest v27 and single copies of the tryptic peptides identifications that passed 98% confidence thresholds were compiled into a curated spectral database. Peak identification and data mining tools were developed using perl scripting language to retrieve peaks as well as matching ions from the curated database. The statistics generated were plotted using R statistical package.

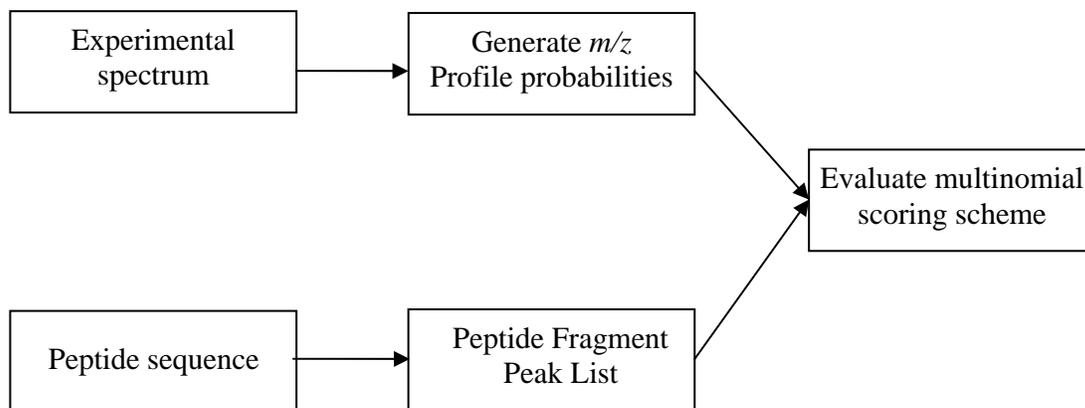


Figure S1: The spectral scoring process is illustrated in this figure. The experimental spectrum is first converted to a probability profile. Fragment peak lists generated from the peptide sequence is matched with the  $m/z$  profile using a multinomial scoring scheme.