

# Supplementary Materials

## Semi-supervised Model-based Validation of Peptide Identifications in Mass Spectrometry-based Proteomics

August 17, 2007

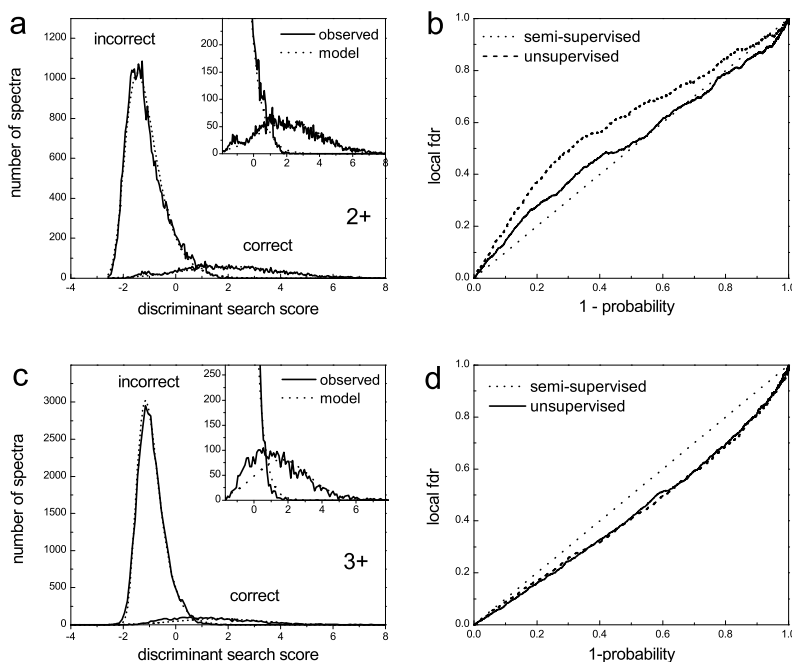


Figure 1: MASCOT search of doubly and triply charged MS/MS spectra in Control Mix II data set. (a) The histogram of TANDEM discriminant search score  $S$  plotted separately for incorrect and correct peptide assignments to doubly charged (a) and triply charged (c) spectra (solid line). Also shown are the distributions learned by the semi-supervised model (dashes). Inset shows a zoom-in picture of the distributions. (b) Actual **fdr** plotted against the complement of posterior probability ( $1 - p$ ) computed using the unsupervised EM (dashes) and the semi-supervised EM (solid) algorithms. Ideal model is represented by a 45 degree line (dots).