

SUPPORTING INFORMATION

A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry

Alexey I. Nesvizhskii^{*†}, Andrew Keller^{*†}, Eugene Kolker[§], and Ruedi Aebersold

Institute for Systems Biology
1441 North 34th Street
Seattle, WA 98103 USA

[†]contributed equally to this work

^{*}corresponding authors: nesvi@systemsbiology.org

akeller@systemsbiology.org

Fax: 206-732-1299

[§]Current address:

BIATECH, North Creek Parkway, Suite 115, Bothell, WA 98011 USA

Table S1. Discriminant functions derived from training dataset of spectra of $[M+H]^+$ precursor ions described in A. Keller *et al. OMICS*, **2002**, 6(2) 207-212. The coefficients weighting each variable are indicated, as well as the correlation between each variable and the discriminant function given by the loading matrix, indicative of the contribution of each variable to discrimination. Contributions can range from none (correlation of 0) to complete (correlation of ± 1).

<i>variable</i>	<i>coefficient</i>	<i>correlation</i>
$Xcorr'$	8.362	0.798
ΔC_n	7.386	0.746
$\ln SpRank$	-0.194	-0.510
dM	-0.314	-0.306
Constant	-0.959	—

$Xcorr'$ is calculated from the SEQUEST $Xcorr$ score in the following way:

$$Xcorr' = \frac{\ln(Xcorr)}{\ln(2L)}$$

where L is the length (number of amino acids) of the assigned peptide.

The discriminant function coefficients listed in Table S1 are used to compute the discriminant score for each peptide assignment to an MS/MS spectrum of $[M+H]^+$ precursor ion as described in A. Keller *et al. Anal. Chem.* **2002**, 74, 5383-5392.

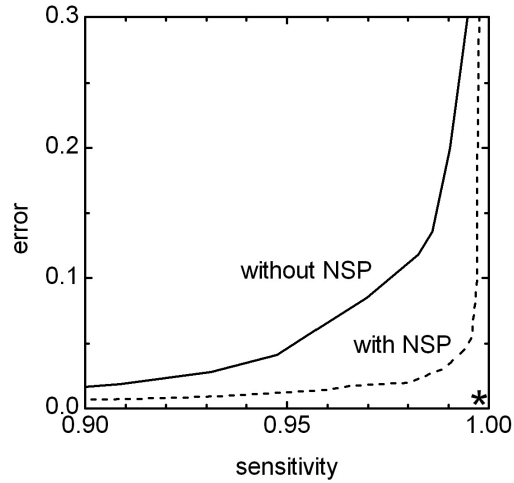


Figure S1. Adjusting peptide probabilities to account for NSP information results in improved discrimination between correct and incorrect peptide assignments. Sensitivity/false positive error rate tradeoff employing thresholds based upon probabilities computed with (dashed line) or without (solid line) adjustment for NSP plotted for the 18prot_Hum dataset. The result of an ideal filter (100% sensitivity and 0% error rate) is indicated by an asterisk.

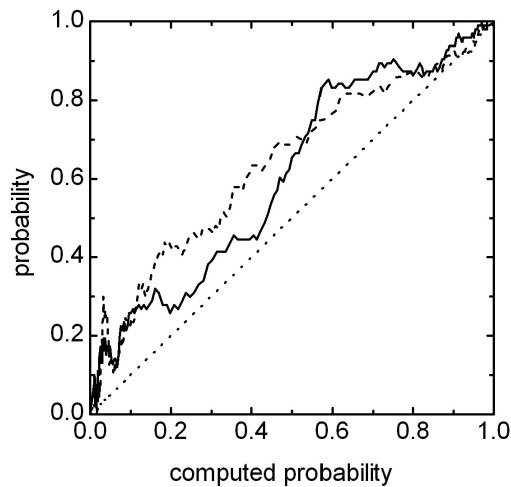


Figure S2. Accuracy of peptide probabilities computed using PeptideProphet. The actual probability (fraction of peptide assignments that are correct) among assignments with indicated computed probabilities derived from the *H. influenzae* sample, Hinf_Hum (solid line) and the *Halobacterium* sample, Halo_Hum (dashed line). The expected probability for an ideal model is indicated by the dotted 45° line.