



computational proteomics

## Laboratory for Computational Proteomics

[www.FenyoLab.org](http://www.FenyoLab.org)

E-mail: [Info@FenyoLab.org](mailto:Info@FenyoLab.org)

Facebook: [NYUMC Computational Proteomics Laboratory](#)

Twitter: [@CompProteomics](#)

## Erratum

### A model of random mass-matching and its use for automated significance testing in mass spectrometric proteome analysis

J. Eriksson and D. Fenyő

On Page 262, the corresponding author should read:

**Correspondence:** Dr. Jan Eriksson, Department of Chemistry, Swedish University of Agricultural Sciences, Box 7015, S-750 07 Uppsala, Sweden

**E-mail:** jan.eriksson@kemi.slu.se

On page 265, equation 1 should read:

$$\rho_i = f_i \frac{k_u}{m_{i+1} - m_i} \quad (1)$$

On page 265, equation 4 should read:

$$f(S) = \left\{ \sum_{k=0}^{k'} p(k) \right\}^H - \left\{ \sum_{k=0}^{k'-1} p(k) \right\}^H \quad (4)$$

On page 264, the legend to Fig. 1 should read:

**Figure 1.** Top left: Mass distribution of the proteins in a genome database (*S. cerevisiae*) compared with the mass distribution of the proteins identified in simulations using random tryptic peptide maps and ranking by the number of matches. Top right: The theoretical number,  $k_u$ , of proteolytic peptides that a protein in a database (*S. cerevisiae*) can yield when at most  $u$  missed cleavage sites are assumed versus protein mass,  $M_p$ . The line represents a least squares fit of a power function to the data. The power equals 1.0. Bottom left: The distribution of  $k_u$  values in the *S. cerevisiae* database for different values of  $u$ . Bottom right: The distribution of  $k_u$  values for proteins identified when using random tryptic peptide maps. The distribution of  $k_u$  values in the whole database is shown for comparison.

On page 264, the legend to Fig. 2 should read:

**Figure 2.** Tryptic peptide mass distribution peaks in two different mass regions (*S. cerevisiae*).

On page 265, the legend to Fig. 3 should read:

**Figure 3.** The frequency of tryptic peptides (within a peptide mass distribution peak) as a function tryptic peptide mass.

On page 267, the legend to Fig. 4 should read:

**Figure 4.** Comparison of simulated and computed (see Section 3.1) frequency functions  $f(S)$  (left panel) of the score (number of matches) and scores required for significance (right panel) for random protein identification in various genomes. The search constraints are: maximum

protein mass,  $M_p < 100$  kDa, maximum number of missed trypsin cleavages,  $u = 2$ , mass accuracy,  $\Delta m = 0.1$  Da, unless stated otherwise in the legend.

On page 268, the legend to Fig. 5 should read:

**Figure 5.** Frequency functions for random protein identification obtained by simulation and model computation for three different cases. Top, (1) all the masses in the maps are between 1396 and 4500 Da; middle, (2) the maps include the entire mass range 800 to 4500 Da; bottom, (3) the maps include masses between 800 and 1396 Da. The model-based computation takes the actual peptide mass distribution as well as all other constraints into account in a direct and rapid way.

On page 268, the legend to Fig. 6 should read:

**Figure 6.** Simulations demonstrating automated model-based significance testing implemented in a protein identification algorithm that ranks the proteins by their respective number of matches. In each map, a fraction of a total of 35 masses originated from a single randomly chosen protein (correlated masses) and the rest of the masses were each from a different protein (noncorrelated). The correlated masses corresponded to a randomly chosen protein sequence-coverage in the range 15–65%. The significance testing efficiently rejects false results. As a lower frequency of false results is tolerated, more true results become nonsignificant.

On page 269, the legend to Fig. 7 should read:

**Figure 7.** A comparison between the use of 2, 4 and 8 mass regions in the model computations of frequency functions for random protein identification.

On page 270, the legend to Fig. 8 should read:

**Figure 8.** The random coverage,  $\delta$ , of a peptide mass distribution peak as a function of peptide mass (see Appendix).

On page 270, the legend to Fig. 9 should read:

**Figure 9.** The mean value of the random coverage,  $\delta$ , in four different mass regions (1: 800–1054 Da, 2: 1055–1395 Da, 3: 1396–2055 Da, 4: 2055–4500 Da) as a function of the mass accuracy. The  $\delta(i, \Delta m)$  functions were derived from the *S. cerevisiae* genome only, but peptide mass distributions are highly conserved between genomes.