ProLuCID: using probability and statistical scores to improve sensitivity and specificity of CID and ETD database search results

Tao Xu, John Venable, Sung Kyu Park, Daniel Cociorva, Bingwen Lu, Lujian Liao, Johannes Hewel, Catherine C L Wong, Xuemei Han, James Wohlschlegel, John R. Yates III
Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA 92037

Overview

ProLuCID uses the XCorr function with a modified XCorr scoring algorithm that incorporates a probability model to improve the specificity and sensitivity of protein identification. The XCorr function is modified to include a probability model that considers the number of possible peptide sequences that can match a given spectrum. This allows for a more accurate assessment of the probability that a peptide sequence is correct, and helps to reduce the number of false positive identifications.

Methods

ProLuCID combines the strengths of the descriptive and probability-based models to provide enhanced computational efficiency, sensitivity and specificity. The XCorr function is modified to include a probability model that considers the number of possible peptide sequences that can match a given spectrum. This allows for a more accurate assessment of the probability that a peptide sequence is correct, and helps to reduce the number of false positive identifications.

Results

Figure 1: A histogram showing the distribution of the XCorr values for forward and reverse protein searches.

Figure 2: A histogram showing the distribution of the XCorr values for forward and reverse protein searches.

Figure 3: A histogram showing the distribution of the XCorr values for forward and reverse protein searches.

Figure 4: A histogram showing the distribution of the XCorr values for forward and reverse protein searches.

Figure 5: A histogram showing the distribution of the XCorr values for forward and reverse protein searches.

Figure 6: A histogram showing the distribution of the XCorr values for forward and reverse protein searches.

Figure 7: A histogram showing the distribution of the XCorr values for forward and reverse protein searches.

Figure 8: A histogram showing the distribution of the XCorr values for forward and reverse protein searches.

Table 1: Number of Proteins identified by SEQUEST and ProLuCID in a saliva sample

Table 2: Number of Proteins identified by SEQUEST and ProLuCID in a saliva sample

Table 3: Number of Proteins identified by SEQUEST and ProLuCID in a saliva sample

Table 4: Number of Proteins identified by SEQUEST and ProLuCID in a saliva sample

Acknowledgments

This work was supported in part by grants from the United States National Institutes of Health (GM070525-05 and GM061025-02).