The ProbSeq fragmentation model (Skilling, 2000a, 2000b) is used for databank searching and characterized fragmentation data. It is shown that for both databank searching and of the correct sequence for the majority of peptides in the validation set.

Each standard was the product of a tryptic digest with subsequent purification. The standards were obtained from the tryptic peptides used to generate the MS/MS data were obtained from the tryptic digests of known protein standards. The software application, MassLynxTM (Waters, Manchester, UK), was used to automate the collection of both the MS survey and MS/MS data. The MS/MS data was obtained in the data-directed analysis (DDA) mode inside where peptides identified in the MS survey had to pass a set of criteria before being analyzed in the MS/MS mode. Peptides were fragmented if the molecular ion was clearly visible in the MS survey spectra and not within 3 m/z of any other peaks. It was ensured that a wide range of peptide masses were selected for fragmentation and the optimal collision energy for fragmentation, determined from the m/z, was used for the CD experiments.

All the MS/MS data acquired was processed using MasseView. This consisted of the background substraction of noise and deconvolution of the spectra into single monoisotopic peaks using the MaxEnt3 algorithm (Waters, Manchester, UK). A theoretical amino acid sequence was predicted for each MS/MS spectrum, using an *in silico* tryptic digestion of the theoretical protein sequence. The precursor ion masses selected for fragmentation were then filtered to the theoretically defined fragment ions. The resulting spectra were validated against its expected sequence using Biotyper (Machines, Manchester, UK) and Mascot (Matrix Science, London, UK).

### Methods

The ideal result is for the difference in the log likelihood’s to be zero. That is, the correct sequence has the highest likelihood of all those explored. The correct sequence is the top result generated by a search using the tuned ProbSeq model is denoted by .

Finally, it would be advantageous to tune the model using data generated using dietary whey or another type of protein.

**Optimization of a Model of Peptide Fragmentation for MALDI Q-TOF MS/MS Data**

**Conclusions**

**Results**

The tuned values were used to update the ProbSeq model, which was then tested against the validation data set. Testing the ProbSeq model consisted of performing database searches and de novo sequencing for many MS/MS spectra in the validation set. For each search the difference in the log likelihood for the top candidate and the correct sequence was calculated. These values were compared for the ProbSeq model before and after tuning.

This chart illustrates the difference between the log likelihood of the top candidate peptide sequence and correct peptide sequence for de novo sequencing results.


1 Water Corporation MS Technologies Center, Float Road, Wythenshawe, Manchester M23 9LZ, UK
2 University of Manchester Institute of Science and Technology, Physical Methods for Biocatalysis and Post-Genome Science
3 Maximum Entropy Data Consultants Ltd, Tresawsan, Killaha East, Kenmare, Ireland

©2004 Waters Corporation