An Efficient De Novo Sequencing Algorithm for the Identification of Modified Peptides from MS/MS spectra

Lin He, Xi Han, Bin Ma[*]
David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada
Email: {l22he, x23han, binma}@uwaterloo.ca

Abstract:
The identification of the peptide sequence from its MS/MS spectrum is an important task in proteomics. The identification without the assistant of a protein database is called the de novo sequencing, which is especially useful in the analysis on novel proteins. Most studies on the de novo sequencing problem focus on the peptide without modifications, since the consideration of a large set of modifications would lead to a speed issue and introduce a plenty of false positives.

In this abstract, we propose an efficient dynamic programming algorithm to solve the de novo sequencing problem with modifications. Given k as the maximum number of modifications allowed per peptides, M as the peptide mass, our algorithm maintains a $(k+1)*M$ matrix S, where the i^{th} row, $1 \le i \le k+1$, corresponds to the solution with at most i-1 modifications for each MS/MS spectrum. When the score $S[i, m]$ for a peak of mass m in the i^{th} row is calculated, each possible residue r needs to be evaluated to determine the current residue. If r is a modified amino acid, the $S[i, m]$ will be calculated from $S[i-1, m-m(r)]$; otherwise, it is calculated from $S[i, m-m(r)]$. The optimal score is calculated for each mass m from 1 to M. Trace back from the highest score in $S[1:k+1, M]$, a peptide, modified or not, can be obtained as the best match of the given MS/MS spectrum. The time complexity of our algorithm is $O(knM)$, where n is the number of possible residues.

Keywords: Mass spectrometry; de novo sequencing; modification

[*] To whom correspondence should be addressed. Email: binma@uwaterloo.ca.