Poster G-10

Detection of post-translational modifications of proteins by mass shift decomposition in peptide mass fingerprints

Authors:
Axel Rack (*Max Planck Institute for Infection Biology*)
Peter R. Jungblut (*Max Planck Institute for Infection Biology*)
Klaus-Peter Pleissner (*Max Planck Institute for Infection Biology*)

Short Abstract: Exhaustive prediction of post-translational modifications (PTMs) from peptide mass fingerprints is combinatorically demanding. We present an algorithm that identifies PTMs by decomposing mass differences between experimental and theoretical peaks over an alphabet of known PTM induced mass shifts that allows high numbers of PTMs to be searched in parallel.

Long Abstract:

Motivation: Exhaustive identification of post-translational modifications (PTMs) from mass spectrometric data is one of the upcoming milestones in computational mass spectrometry. While MS/MS is becoming the method of choice, peptide mass fingerprinting is still used widely to identify proteins and post-translational modifications (PTMs). Common tools like MOWSE, Mascot and MS-Fit [1, 2, 3] rely on the generation and comparison of all possible theoretical peaks given a set of variable modifications during database search. With increasing numbers of PTMs, the number of peaks that have to be generated and compared increases drastically and computation becomes impractical rapidly. To avoid the combinatorial explosion, however, one can study all observed differences in peak masses between experimental and theoretical spectra, thus transferring the process of PTM prediction to a problem of finding suitable combinations of PTMs that explain observed mass distances between peaks. We know only one existing tool that follows this approach, the FindMod tool [4] that allows parallel searching with 22 modifications and amino acid substitutions. Yet, it restricts predictions to combinations of two of these modifications per peptide.

Problem statement: Given a query M (mass shift between two peaks) and a weighted alphabet Σ of k modifications with weights according to single PTM mass shifts; find all interpretations of M over Σ. An interpretation of M over Σ (referred to as witness) is a vector c that describes the amount of each PTM mass shift needed to add up to the total shift M. A similar but simpler problem that is restricted to a positive query M and positive weights can be computed using a dynamic programming (DP) algorithm as described in [5]. Runtime scales with the number of witnesses and has an additional cost of O(k*M) for preprocessing.

Approach: We improved the existing DP approach to adapt it to the PTM prediction problem by introducing (i) upper bounds for the occurrence of each PTM in a witness and (ii) negative query M and weights. Both changes are inevitable, as we have to consider positive and negative PTM mass shifts and biologically relevant restrictions on PTM occurrences due to AA composition of peptides. Time and memory complexity do not increase by extension (i) whereas a moderate increase appears with respect to (ii).

The algorithm requires an experimental peaklist P, the expected underlying protein sequence S and a set of PTMs with known mass shifts. In the first step, S is theoretically digested to obtain the target peaklist of unmodified peaks. In advance, each mass shift between experimental and target peaks is decomposed over the given set of modifications.
Results: Preliminary tests with artificial data indicate good predictions but still show potential for runtime improvement. However, a complete assessment of algorithmic behaviour and predictive capabilities cannot be made until it is applied to real data.
Discussion: Our algorithm represents, as far as we know, a novel approach to predict PTMs from PMF spectra using number decomposition that allows high numbers of PTMs to be searched in parallel. So far, the only major drawback is the strong dependence of runtime and memory requirements on mass accuracy.

[1] Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr. Biol. 3, 327-332.
[2] http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm
[3] Perkins, D.N.; Pappin, D.J.; Creasy, D.M.; Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis; 20(18), 3551-3567
[4] Wilkins, M.R.; Gasteiger, E.; Gooley, A.; Herbert, B.; Molloy, M.P.; Binz, P.A.; Ou, K.; Sanchez, J.-C.; Bairoch, A.; Williams, K.L; Hochstrasser, D.F. (1999) High-throughput Mass Spectrometric Discovery of Protein Post-translational Modifications. Journal of Molecular Biology, 289, 645-657
[5] Böcker, S.; Lipták, S. (2005) Efficient Mass Decomposition ACM Symposium on Applied Computing; Santa Fe, New Mexico