

## RIPS 2006: TimeLogic Project description

UCLA Faculty: Matteo Pellegrini  
Sponsor Liaison: Roland Luethy, Active Motif Corp.

### Determining the sequence of peptides from tandem massspectra

#### Background

Masspec based proteomics is a modern technology used to identify proteins in complexes, cellular components or bodily fluids. It can not only identify the proteins, but also find posttranslational modifications. For a review of mass spectrometry see [1]. In a typical proteomics experiment a mixture of proteins is subjected to enzymatic proteolysis, most frequently using trypsin for the digestion. The resulting mixture of tryptic peptides is fractionated by HPLC chromatography and the fractions are subjected to tandem masspectroscopy. The first mass separation isolates tryptic peptides according to their mass over charge ratio. Peptides within a mass range can then be selected and fragmented. The resulting fragments are then analyzed with the second masspec step producing the fragmentation spectrum. Fragmentation occurs most frequently by breaking of the peptide bonds, producing a series of fragments with mass differences corresponding to the amino acid masses. However, other bonds break too and not all peptide bonds break with the same probability. There are two types of approaches to derive the peptide sequence from the fragment spectra. One approach relies on using theoretical spectra derived from the protein sequence database and matching these theoretical spectra to the observed ones. Our RIPS project last year used this approach [last years report]. The second approach is the *de novo* interpretation of the fragment spectrum. The advantage of this approach over the database search method is that unanticipated peptides derived from mutations or alternative splice forms can be found. Two tools in this category which are publicly available are Lutefisk [2] and PepNovo [3]. There are also two commercial tools [4, 5]. None of the tools currently available has the capability to consistently produce complete peptide sequences. Frequently only partial sequences or multiple solutions are found. Therefore there is room for improvement and it is worthwhile to investigate alternative approaches.

#### Goals

The main goal is to develop a method for the de-novo interpretation of tandem masspectra, with a scoring function indicating the likelihood of the assignment.

#### Deliverables

- A computer program which takes a collection of MS/MS spectra and outputs ranked candidate peptides and scores for each spectrum.

## References

1. Steen, H. and Mann, M. (2004). The ABC's (AND XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol.* 5(9), 699-711.
2. Taylor, J.A., and Johnson, R.S. (1997). Sequence database searches via *de novo* peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 11, 1067–1075.
3. Frank A, Pevzner P. (2005) PepNovo: *de novo* peptide sequencing via probabilistic network modeling. *Anal Chem.* 77(4):964-73.
4. Dancik, V., Addona, T. A., Clauser, K. R., Vath, J. E., Pevzner, P. A. (1999). *De Novo* Peptide Sequencing via Tandem Mass Spectrometry. *J. Comput. Biol.* 6, 327-42.
5. Ma, B., Zhang, K., Lajoie, G., Doherty-Kirby, A., Hendrie, C., Liang, C., Li, M. (2003) PEAKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry *Rapid Commun. Mass Spectrom.* 17, 2337-2342.