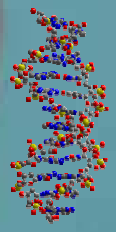
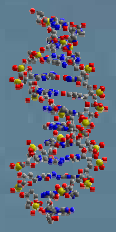
Method to Identify Mass Spectra Containing Post-translational Modifications



Rashida Askia
Fisk University

Research Alliance in Math and Science
Computer Science and Mathematics Division



Mentors:
Mike Leuze, Ph.D
Andrey Gorin, Ph.D

http://www.csm.ornl.gov/Internships/Websites04/r_askia/abstract.html

Abstract

Proteins are complex substances consisting of amino acid residues, and are essential biological compounds. A PTM filter is a procedure that quickly scans mass spectra for possible signs of modified proteins. The recently developed Probability Profile Method (PPM) is used to reliably identify b-ions and y-ions in the mass spectrum. PTM filter performs a *de novo* search for any of these ions that are separated by the mass of modified amino acids, and then reports whether the spectrum is a possible candidate for further PTM analysis. The filter can be adjusted to minimize either "false negatives" or "false positives" among the reported spectra. This filter will allow later PTM searches to be focused only on the mass spectra most likely to contain PTM information, helping to streamline the entire PTM search process.

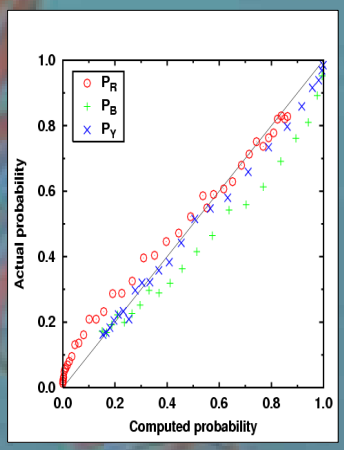
Project Goals/Tasks

Improvements in existing methods that analyze the accuracy of computed probability and predictions of ion peaks are needed. The focus of this project was to carefully develop a post-translational modification filter to identify the peaks and show whether PTMs were present.

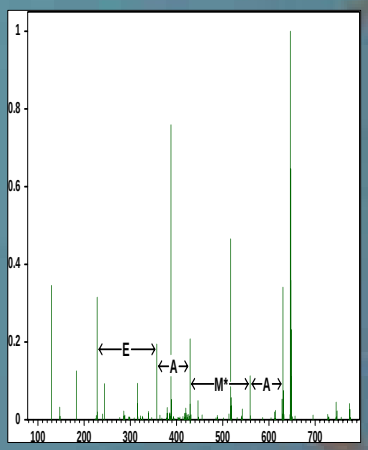
Post-translational Modification

A post-translational modification (PTM) is a change made to a protein after it has been synthesized by a ribosome (a process called translation). Such modifications change the structure and function of a protein, and so their identification and localization within the protein sequence is of value for understanding the biological processes in which they take part.

One of the most common ways to study PTMs employs proteolysis to cut the proteins into peptides, and tandem mass spectrometry to discover the fragments produced by breaking each peptide. Though there are many database search algorithms which are capable of identifying each peptide base sequence, a new approach is being sought to minimize and possibly eliminate error during the PTM process.



Accuracy of PPM predictions



PTM/spectrum graph showing mass vs. intensity

Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) is one of the most important tools used in proteomics for the identifications of proteins. In MS/MS, the mass and abundance of peptide fragments ions are measured and can be used to identify the original peptide sequence. Current methods of correlating MS/MS data to peptides are quite successful. However, there are still several issues to overcome in order to improve the field.

Probability Profile Method

The Probability Profile Method (PPM) is a recently developed system for categorizing ion peaks in mass spectra of peptides. Neutral-loss and complementary peak information is used to assign probabilities for a given peak to belong to several pre-defined categories (e.g. b-ion, y-ion, or other). These computed probabilities have been shown to accurately reflect actual probabilities, and approximately half of b-ion and y-ion peaks have been reliably identified in experiments. The computed probabilities can be used for further analysis of peptide mass spectra.

Results

A tool was developed to test whether its feasible to search a spectra for PTMs. Even though the tool was a success, more data sets are needed in order to evaluate the effectiveness of the PTM tool. This modification procedure is an ongoing effort to improve the way PTMs are identified within the spectrum.

Perl Code

```
 sub ptm_find {
my ($spec) = @_;
my $peaks = $spec->select('$P_b_prob >= 0.2');
my $np = $#$peaks;
my $seq = join '', @{$spec->{seq}};
#print "Seq is '$seq', using $np peaks\n";

for (my $i=0; $i<= $np; $i++)
{
my $mi = $peaks->[$i][0];        # mass of peak $i
my $pbi = $peaks->[$i][10]{B};   # b-prob of peak $i
 for (my $j=$i+1; $j<=$np; $j++)
{
 my $mj = $peaks->[$j][0];
 my $mass = $mj - $mi;

for my $ptm (@$PTM_STEPS)
{
my $diff = abs($ptm->[1] - $mass);
if ($diff <= 0.4) {
my $pbj = $peaks->[$j][10]{B};
my $prob = $pbi * $pbj;
my $tf = 0;   # True or false
 $tf = 1 if index($seq, $ptm->[0]) >= 0;
 #print "Peaks $i and $j: looking for '$ptm->[0]',
$tf\n";
 my $ierr = get_err_grid($diff);
 my $iprob = get_prob_grid_0($prob);
$table->[$ierr][$iprob][$tf]++;
 }
 }
 } # end ptm_find
```


OAK RIDGE NATIONAL LABORATORY
U.S. DEPARTMENT OF ENERGY

Special thanks to Robert M. Day

The Research Alliance in Math and Science program is sponsored by the Mathematical, Information, and Computational Sciences Division, Office of Advanced Scientific Computing Research, U.S. Department of Energy. The work was performed at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. De-AC05-00OR22725. This work has been authored by a contractor of the U.S. Government, accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.