Automated N-Glycan Composition Analysis with LC-MS/MSMS

H Peltoniemi¹, I Ritamo², J Räbinä² and L Valmu²
Applied Numerics Ltd¹, Nuottapolku 10 A8, 00330 Helsinki and Finnish Red Cross Blood Service², Helsinki, Finland.

MOTIVATION
Compared to proteomic profiling and differentiation the glycan analysis still employs a lot of manual work and can be a burden with increasing number of spectra. The aim of this study is to ease the task by using in house developed glycomic software methods in combination with existing proteomic tools. The resulting workflow (Fig 1) is targeted especially to glycan LC-MS/MSMS analytics and can be run with a minimal amount of human intervention. The method was applied to cord blood derived mononuclear cells. The final goal is to profile and differentiate stem cell surface glycans which are analysed at Finnish Red Cross Blood Service [1,2,3].

WORKFLOW
The workflow (Fig 1) combines LC-MS/MSMS measurements, proteomic software and glyccan specific tools. The glycan methods (steps 4-7, in the workflow) are based on the in house developed R [4] library GlycanID. The library contains functions to basic mass spectrum analysis operations including spectrum matching, statistical scoring and visualization. The aim of the library is to enable fast development of new workflow variants when new requirements appear.

1. LC-MS/MSMS DATA
The start up data is LC-MS 2D map containing chromatographic (time) and mass (m/z) dimensions and embedded tandem MSMS spectra created by fragmentation of glycans.

2. IDENTIFY FEATURES
Potential glycan features are identified with Progenesis LC-MS (Nonlinear Dynamics Ltd) software.

3. EXTRACT MSMS SPECTRA
MS spectra with identified charge states is extracted with Mascot Distiller (Matrix Science Ltd).

4. MATCH COMPOSITIONS (MS)
Searches glycan compositions matching to feature masses. 
- Feature matching is done against:
  A) Theoretical compositions generated with a given set of rules [5], or
  B) Compositions given in a database.
- Several charge carrier ion types and neutral adducts can be used.
- Outliers can be removed by iteratively applying linear fitting and by elimination of compositions with a mass difference greater than two standard deviations.

5. MATCH COMPOSITIONS (MSMS)
Searches glycan compositions matching to precursor mass and MS2 spectrum.
- MSMS precursor compositions are found as in the step 4.
- Fragment matching is done either by:
  A) Generating all the theoretical fragments that any glycan structure with the given composition could produce [5], or
  B) If the precursor composition is given in a database the theoretical or measured spectra in a MSMS fragment database can be used.
- Removal of outlier fragment hits by linear fitting.
- Ranking with statistical score defined by:
  1) the probability that a random set of fragments would have as many or more shared peaks with measured spectrum as the ranked composition [5] and
  2) the probability that by randomly selecting the observed number of shared peaks the same or higher amount of intensity can be covered.
- Optional filtering:
  1) Spectrum is classified as a glycan if any mass difference between two peaks matches a given list of masses, typically composed by one or two monosaccharide masses.
  2) A given monosaccharide can exist in a composition only if the MSMS spectrum contains at least one of the given marker ions. This feature is developed especially for differentiation of Neu5Ac and Neu5Gc residues.

6. COMBINE MS AND MSMS
Combines the results from the MS and MSMS matching so that the MSMS identification is transformed to the MS features if the mass and retention time differences are less than the given tolerances.

7. DECONVOLUTE
Inverts feature-glycans to glycan-features map.
- Calculates the total intensity and score for each glycan by summing the measured feature intensities and MSMS scores with different charge states and charge carrier types.
- Groups the glycans so that the ones matching to the same set of features are given in the same group.
- For each group one glycan is marked as unique if there is only one glycan that spans all the group features and has the highest score. Otherwise the group is marked to be contradictory.

8. PROFILE
The glycan profile is created from the deconvoluted data and the possible contradictory groups are manually resolved based on biological information available.

REFERENCES
1) A. Hakkinen et al., Glycemics of bone marrow-derived mesenchymal stem cells can be used to evaluate their cellular differentiation stage, Glycoconjug. J. 2009 26:367-384.

ACKNOWLEDGMENTS
Glykos Finland Ltd is acknowledged for glycan sample preparation and comparative MALDI-TOF analysis.