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

H Peltoniemi[1], I Ritamo[2], J Räbinä[2] and L Valmu[2]

Applied Numerics Ltd[1] , Nuottapolku 10 A8, 00330 Helsinki and Finnish Red Cross Blood Service[2], 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 glycan 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

MSMS 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.

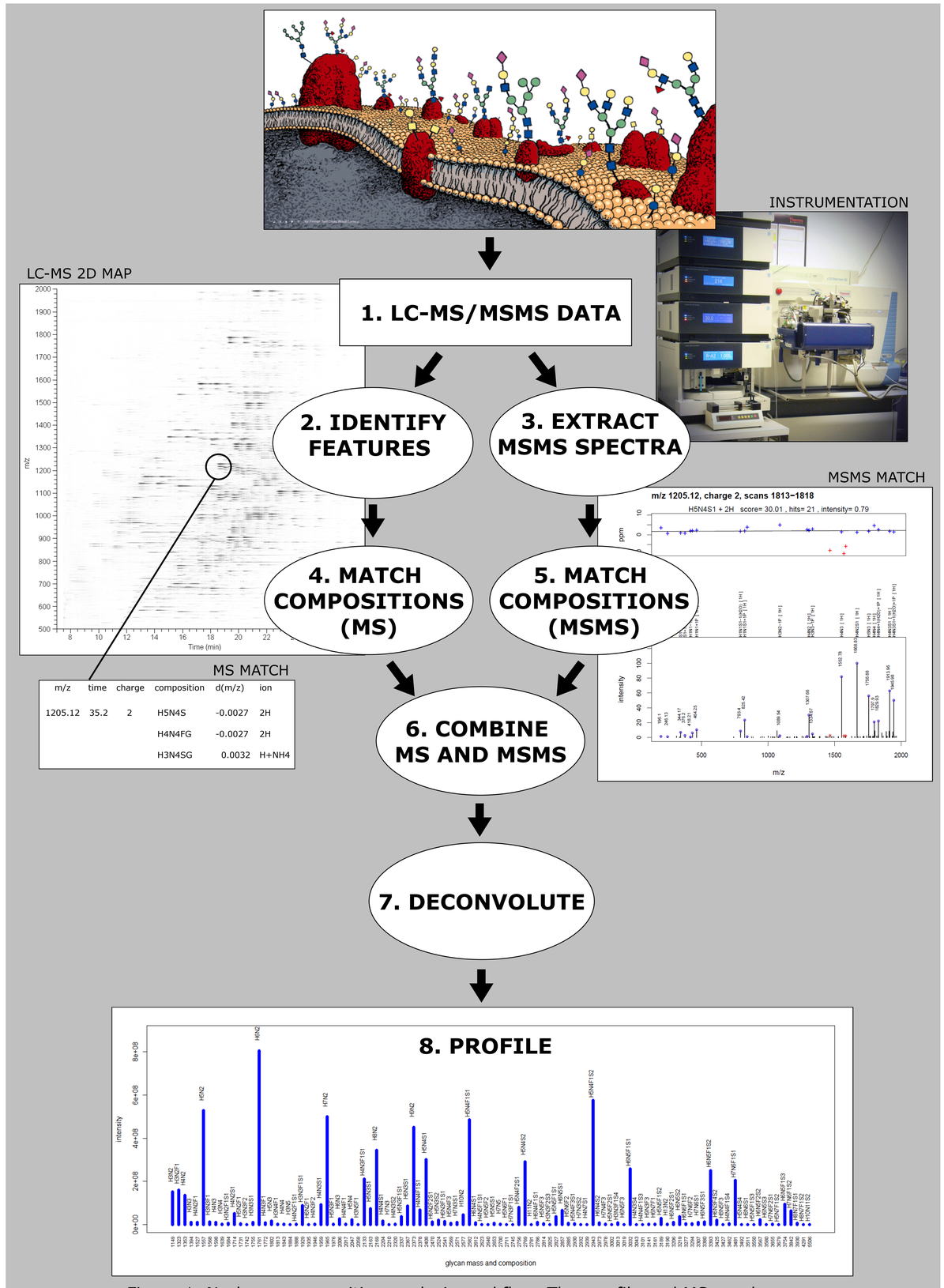

Figure 1. N-glycan composition analysis workflow. The profile and MS graphs are result of the analysis of cord blood derived mononuclear cells sample.

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.

MONONUCLEAR CELLS

The workflow was applied to cord blood derived mononuclear cells sample previously analysed by MALDI-TOF [3].

SAMPLE PREPARATION AND MS

The N-glycans were prepared and permethylated as described in [3,6] and kindly provided by Glykos Finland Ltd. Permethylated N-glycans were loaded to reversed phase (RP) precolumn (Atlantis dC18, Waters) and separated in analytical RP column (PepMap 100, Dionex Corporation). Ultimate 3000 LC instrument (Dionex Corporation) was operated in nano scale with flow rate of 0.3µl/min. Eluted glycans were introduced to LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific Inc.) via ESI Chip interface (Advion BioSciences Inc.) in positive-ion mode.

WORKFLOW RUN

The composition matching was done against auto generated glycan compositions (Hex 3-15, HexNAc 2-15, Fuc 0-6 and Neu5Ac 0-6) and with the assumption of the N-glycan core. Mass tolerances were set to 0.01 Da or 10 ppm. Also the differentiation of Neu5Ac from Neu5Gc was tested.
The workflow started from total of 2444 MS features and 449 MSMS precursors and ended to about 100 glycan compositions classified as credible ones. The differentiation of Neu5Ac and Neu5Gc was shown to be possible with MSMS data, especially with the help of low mass marker ions.

COMPARISON WITH MALDI RESULTS

The calculated glycan compositions (Fig 1) fit very well to the already published glycan structure data of cord blood derived mononuclear cells performed by MALDI-TOF mass spectrometer [3]. High intensity mannose structures are again visualized as well as sialylated complex N-glycans with mainly two or three antennaes. Some differences can be observed, for example in the sialylation degree of the glycans, but these are likely due to the different ionization methodology used in these studies.

DISCUSSION

The number of features detected after LC-MS glycan analysis is far greater than in other analytical methods typically performed in one dimension (Fig 2). In LC-MS analysis the complexity is increased both by electrospray ionization (ESI), which produces multiple charged ions, as well as by second dimension, introduced by chromatographic retention time. Additional complexity is still involved in the number of different metal adducts detected. Therefore, in order to utilize the possibilities that LC-MS analysis introduces to glycan structure determination, a competent data handling tool is essential in order to simplify otherwise extremely laborious interpretation of the data.

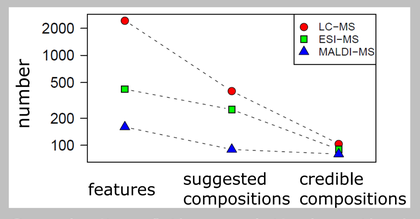

Figure 2. Complexity of the analytical methods based on mononuclear cell data. The graph is approximative because of different setups between the methods.

REFERENCES

1) A. Heiskanen *et al*., Glycomics of bone marrow-derived mesenchymal stem cells can be used to evaluate their cellular differentiation stage, *Glycoconj. J.* 2009 **26**:367-384.
2) T. Satomaa *et al*., The N-glycome of human embryonic stem cells, *BMC Cell Biol*. 2009, in press.
3) H. Hemmoranta *et al*., N-glycan structures and associated gene expression reflect the characteristic N-glycosylation pattern of human hematopoietic stem and progenitor cells, *Exp. Hematol.* 2007 **35**:1279-1292.
4) The R Project for Statistical Computing, www.r-project.org.
5) S. Joenväärä *et al*., N-Glycoproteomics – an automated workflow approach., *Glycobiology* 2008 **18**(4):339-349.
6) P. Kang *et al*., Solid-phase permethylation of glycans for mass spectrometric analysis., *Rapid Commun. Mass Spectrom.* 2005 **19**:3421-3428.

ACKNOWLEDGEMENTS

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