GlycoFragment and GlycoSearchMS: web tools to support the interpretation of mass spectra of complex carbohydrates

Klaus Karl Lohmann* and Claus-W. von der Lieth

German Cancer Research Center, Central Spectroscopic Department B090, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

ABSTRACT

In recent years, mass spectrometry has become the method of choice for high-sensitivity glycan identification. Currently, only a few tools assisting mass spectra interpretation are available. The web application GlycoFragment (www.dkfz.de/spec/projekte/fragments/) calculates all theoretically possible fragments of complex carbohydrates and aims to support the interpretation of mass spectra. GlycoSearchMS (www.dkfz.de/spec/glycosciences.de/sweetdb/ms/) compares each peak of a measured mass spectrum with the calculated fragments of all structures contained in the SweetDB database. The best-matching spectra and the associated structures are displayed in order of decreasing similarity. Since both algorithms work very efficiently, they are well suited to be used for automatic identification of series of mass spectra of complex carbohydrates.

INTRODUCTION

The term 'glycomics' describes the scientific attempt (1) to identify and study all the carbohydrate molecules—the glycome—produced by an organism such as human or mouse. Rapid and sensitive high-throughput analytical methods employing mass spectrometry (MS) and high-performance liquid chromatography (HPLC) techniques are currently applied to provide information on the glycan repertoire of cells, tissues and organs (2,3). One of the aims of the emerging glycomics projects is to create a cell-by-cell catalogue of glycosyltransferase expression and detected glycan structures. Glycan profiling of normal and diseased forms of a glycoprotein has provided new insights for future research in rheumatoid arthritis, prion disease and congenital disorders of glycosylation (4–8). In all these diseases, differences in glycosylation indicate that there are cellular or genetic changes that affect the activity of specific glycosyltransferases.

In recent years, MS (3,9–12) has become the method of choice for high-sensitivity glycan identification and characterization. One strategy in functional glycomics is to compare the glycan repertoires found in normal and diseased or treated tissue. To analyse glycosylation pattern, a typical protocol works as follows: after the extraction of proteins and their tryptic digestion, an enzymatic or chemical treatment cleaves off the attached sugar moieties. The permethylated *N*-glycans are subsequently analysed using MS and the assignments of peaks are based on compositions, taking into account biosynthetic considerations. Alternative sequences are possible. In such cases electrospray tandem MS/MS may be used to differentiate between alternative possibilities. Currently, the assignment of MS peaks is mainly done manually, and only a few applications are available to support this process. GlycanMass (13) is a simple web tool which allows calculation of the mass of an oligosaccharide composition. GlycoMod (14) is designed to find all possible compositions of a glycan structure from its experimentally determined molpeak.

The aim of this work is to provide efficient web tools to support the interpretation of mass spectra. GlycoFragment enables easy generation of all theoretically possible fragments of a defined glycan structure and thus supports the assignment of all peaks contained in a mass spectrum. GlycoSearchMS takes a mass spectrum, searches through a database of theoretically calculated spectra and identifies the best matching spectra.

GlycoFragment

GlycoFragment (15) generates all theoretically possible A-, B-, C-, X-, Y- and Z- fragments of oligosaccharides according to the definitions of Domon and Costello (16).

*To whom correspondence should be addressed. Tel: +49 6221 424670; Fax: +49 6221 424554; Email: k.lohmann@dkfz.de
Correspondence may also be addressed to Claus-W. von der Lieth. Email: w.vonderlieth@dkfz.de

Figure 1. Input of an *N*-glycan structure using the extended alphanumeric description for oligosaccharides as proposed in the IUPAC nomenclature. The user has the option to indicate the ion used or other adducts. Four different ways to display the resulting fragment can be activated.

Input of carbohydrate structures

The extended alphanumeric description for oligosaccharides (see the IUPAC Nomenclature of Carbohydrates—Recommendations) is used to input carbohydrate structures (see Figure 1). The non-reducing termini are displayed at the left side; the reducing end is at the right. Each three-letter code for a monosaccharide unit is preceded by the anomeric descriptor (α, β) and the configuration symbol (D, L). The ring type or size is indicated by attaching f or p (furanose or pyranose) to the saccharide code. The locants of linkages are given in parentheses between the monomer symbols. Several forms of derivatization and substitution of the reducing end are implemented.

Since mass spectra provide only information about masses of oligosaccharide fragments, it is not possible to distinguish between isomeric residues having the same mass. Often only the composition of a monosaccharide is given in the literature to characterize complex carbohydrates. Generalized abbreviations are used for monosaccharides exhibiting the same mass. The code 'hex' is used for all hexoses (galactose, glucose, mannose, etc.); 'hexnac' for all hexoseamines with an acetyl group linked to the C-2. Carbohydrates are not ionized as efficiently as proteins that can be protonated. Therefore, chemical derivatization techniques are frequently used to improve the sensitivity for carbohydrates. Reducing sugar chains contain a single reactive carbonyl group that can be substituted separately from the many hydroxyl groups. The carbohydrate may be linked to a reagent that already contains a functional group that can be easily protonated. Reductive amination with an aromatic amine is one of the commonly applied derivatization techniques. Therefore, 2-aminopyridine (2AP), 4-aminobenzoic-2-(dimethylamino)ethylester (4ABDEEAE), 4-aminobenzoic acid ethyl ester (4ABEE) and 4-Trimethylaminoaniline (4TMAPA) are added as templates. For the quick input of derivatives formed by reductive amination, GlycoFragment accepts codes as hex-2AP, hexnac-2AP, hex-4ABDEEAE, hexnac-4ABDEEAE etc. (see e.g. 'derivatives' on the web page).

Generation of fragments

The algorithm calculates the mass of all possible fragments of an oligosaccharide in three steps. First, a complete topological description of all atoms and their connectivity is generated. This is done using the program Sweet-II (17,18), which can interpret the extended nomenclature of complex carbohydrates. Sweet-II selects corresponding templates from a database of predefined monosaccharide units and connects the respective sugar units according to the specified linkage information. In a second step the molecular structure is analysed and bonds, which have to be broken to produce a specific fragment, are assigned. In the third step, each possible fragment is generated by virtual bond cleavage and a subsequent summation of atomic masses of the atoms belonging to a certain fragment. The monoisotopic or average masses are calculated with a resolution of 1 mDa.

Output

Several output options are provided. The default is a list where only fragments resulting from glycosidic cleavages (B- and Y-type fragments) are displayed, ordered by increasing mass (see Figure 2). Output of the less frequently occurring C- and Z-fragments as well as all possible ring fragmentations can produce rather long lists of ions. Therefore, a listing of these fragments is available only on demand. Linkage path information allows unambiguous definition of residues in different antennas. A more interactive way to find masses which occur due to a specific fragmentation is provided by the output option 'view as structure' (see Figure 3).

GlycoSearchMS

The GlycoSearchMS tool compares a measured mass spectrum with a library of calculated spectra. The current version contains all theoretically possible spectra of about 5000 *N*- and 1200 *O*-glycans, which were extracted from SweetDB (19). The GlycoFragment algorithm was used to calculate the

Mass in amu	Ion
[M-H]- 1233.4264 Mass: 1234.4334 [M+H]+: 1235.4412 [M+Na+]: 1257.4224 [M+K+]: 1273.5314	
185.0426	$B_{4,4,3}$ + Na+
	$B_{4,4,6,3}$ + Na+
	$B_{4,4,6,6}$ + Na+
244.0797	Y_4 + Na+
447.1591	$Y_{4,4}$ + Na+
509.1482	$B_{4,4,6}$ + Na+
771.2647	$Y_{4,4,6}$ + Na+
833.2539	$B_{4,4}$ + Na+
1036.3332	B_4 + Na+
1095.3703	$Y_{4,4,3}$ + Na+
	$Y_{4,4,6,3}$ + Na+
	$Y_{4,4,6,6}$ + Na+

Figure 2. Fragments resulting from glycosidic cleavages (B- and Y-type fragments) for the *N*-glycan of Figure 1 are displayed, ordered by increasing mass. The linkage path information—a list of attachment positions on the reducing side of the glycosidic linkage—allows an unambiguous assignment of residues in different antennas.

Figure 3. Output option 'view as structure'. When moving the cursor over a glycosidic linkage, a small box appears on the screen displaying the associated B- and C-ion (blue background) on the non-reducing side. The corresponding Y- and Z-ion (red background) are shown on the reducing side. Moving the cursor over the sugar will display all possible A- and X-ions, indicating the cleaved bonds by their number.

masses (of all A-, B-, C-, X-, Y- and Z-fragments) and to assign the peaks according to the definitions of Domon and Costello. The user has the opportunity (i) to select the derivatization of the reducing sugar, (ii) to tell if the experimental spectrum derives from permethylated or peracetylated glycans and (iii) to tell which ESI ion has been used.

The search algorithm compares each peak of the input spectrum with the calculated fragments of all structures contained
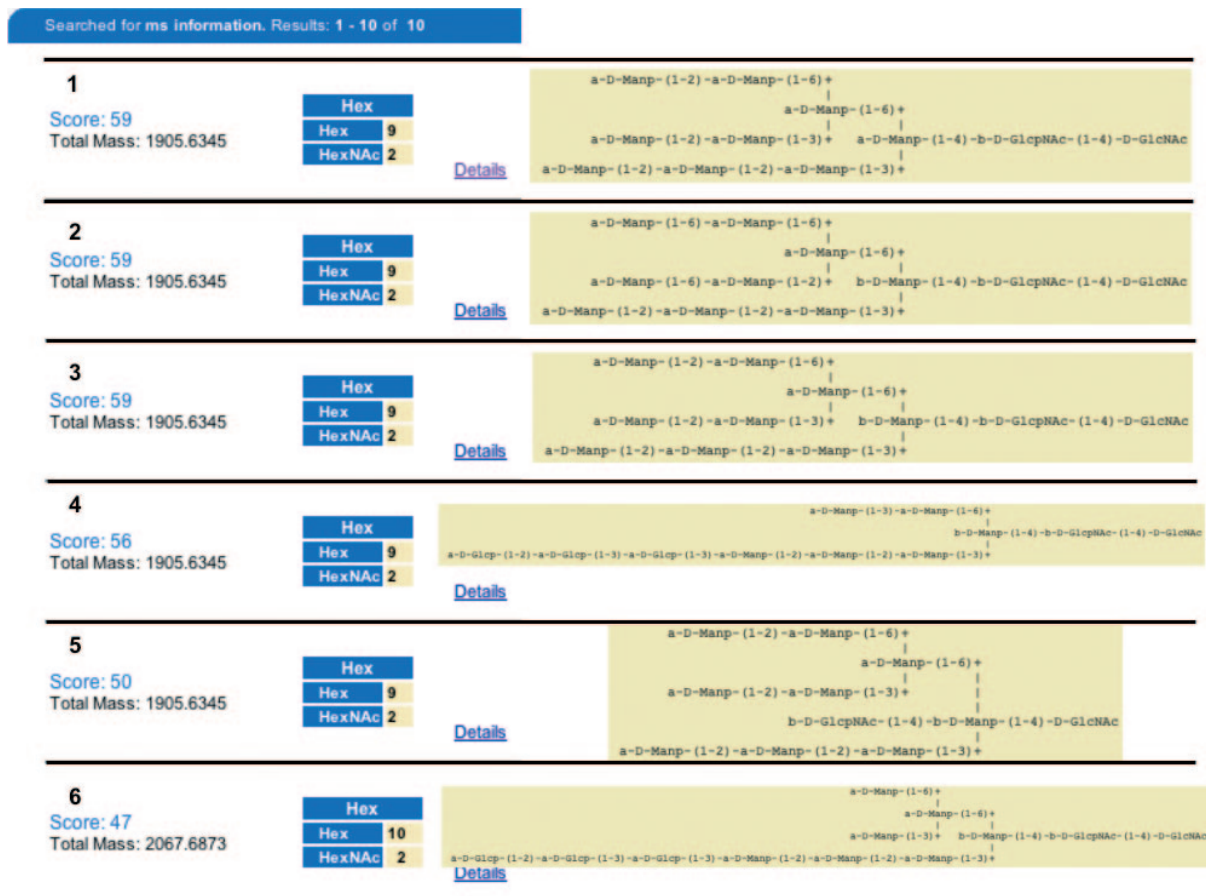
Figure 4. Overview of the results of a GlycoSearchMS run. The composition of the best-matching spectra and their structures are displayed in descending order of MS_{score}.

in the database. The number of matched peaks within a certain tolerance is used to compute the MS_{score} by which the best-matching spectra are ranked (see Figures 4 and 5).

$$MS_{score} = \frac{\sum_1^n [1 - (|P_s - P_r|/Err)]}{n_{input}} \times 100,$$

where n is the number of input peaks, P_s is the mass-to-charge ratio (m/z) input peak, P_r is the m/z reference peak in the library and Err is the tolerance (in mDa).

Output

Several ways to analyse in detail the results of the search are provided. A more general overview lists only the MS_{scores} the glycan composition and the structure of the best-matching spectra (see Figure 4). Upon activating the detail button for each retrieved entry, the entire structure is displayed and a list of where experimental and library masses are matching is displayed. By moving the cursor over the matched mass, the origin of this fragment is indicated, displaying ion type and linkage information (see Figure 5).

The algorithm has been intensively tested for N- and O-glycans and glycolipids. The retrieved results depend on the comprehensiveness of the database searched. It turned out that fragments originating from breaking two bonds within

a sugar ring (A- and X-ions) are highly sensitive markers to differentiate between the various branching pattern of N-glycans found in nature. Ions resulting from inner-ring fragmentations of permethylated glycans are indicative for assigning the exact linkage type.

The current implementation of the GlycoSearchMS interface allows an interactive interpretation of one spectrum at a time.

DISCUSSION

Progressing glycomics projects will dramatically accelerate the understanding of the roles of carbohydrates in cell communication and hopefully lead to novel therapeutic approaches to the treatment of human disease. To screen the glycan content of various tissues using high-throughput techniques and HPLC and/or MS to detect glycans, automatic procedures for a reliable identification of N- and O-glycans, are required. However, high-throughput techniques will soon overwhelm the current capacity of methods if no automation is incorporated into glycomics. The implementation and testing of suitable algorithms is currently the most active field in the development of applying bioinformatics to glycobiology. Since the GlycoFragment and GlycoSearchMS algorithms work very efficiently, such methods are well suited to use for automatic
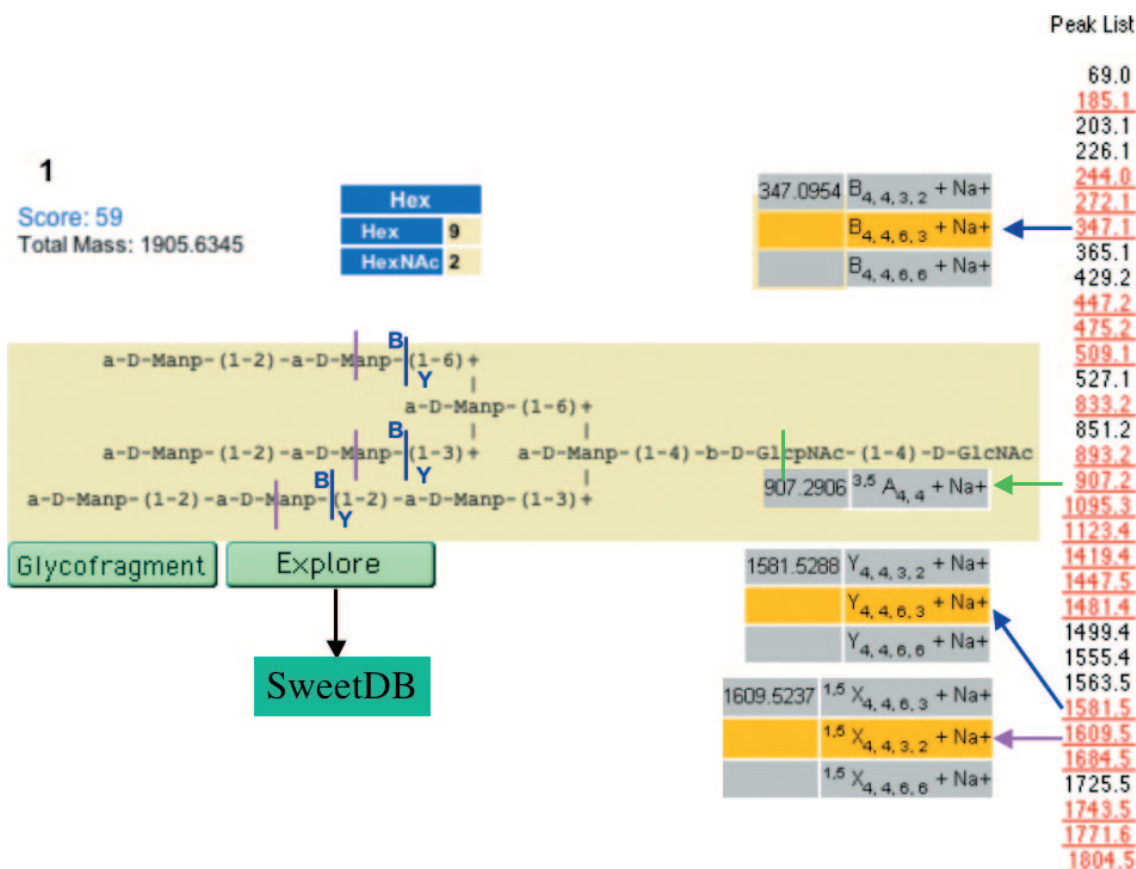
Figure 5. Detailed information on matches between input and library peaks of a GlycoSearchMS run. Masses where experimental and library masses are matching are given in red in the peak list. When moving the cursor over a matched mass the type of ion and the linkage path information assign its structural origin. The position of fragmentation is indicated using different colours.

identification of series of mass spectra of complex carbohydrates.

REFERENCES

1. Feizi,T. and Mulloy,B. (2003) Carbohydrates and glycoconjugates. Glycomics: the new era of carbohydrate biology. *Curr. Opin. Struct. Biol.*, **13**, 602–604.
2. Rudd,P.M., Colominas,C., Royle,L. Murphy,N., Hart,E., Merry,A.H., Hebestreit,H.F. and Dwek,R.A. (2001) A high-performance liquid chromatography based strategy for rapid, sensitive sequencing of *N*-linked oligosaccharide modifications to proteins in sodium dodecyl sulphate polyacrylamide electrophoresis gel bands. *Proteomics*, **1**, 285–294.
3. Dell,A. and Morris,H.R. (2001) Glycoprotein structure determination by mass spectrometry. *Science*, **291**, 2351–2356.
4. Rudd,P.M., Elliott,T., Cresswell,P., Wilson,I.A. and Dwek,R.A. (2001) Glycosylation and the immune system. *Science*, **291**, 2370–2376.
5. Peracaula,R., Tabares,G., Royle,L. Harvey,D.J., Dwek,R.A., Rudd,P.M. and de Llorens,R. (2003) Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology*, **13**, 457–470.
6. Peracaula,R., Royle,L., Tabares,G. Mallorqui-Fernandez,G., Barrabes,S., Harvey,D.J., Dwek,R.A., Rudd,P.M. and de Llorens,R. (2002) Glycosylation of human pancreatic ribonuclease: differences between normal and tumour states. *Glycobiology*, **13**, 227–244.
7. Rudd,P.M., Merry,A.H., Wormald,M.R. and Dwek,R.A. (2002) Glycosylation and prion protein. *Curr. Opin. Struct. Biol.*, **12**, 578–586.
8. Butler,M., Quelhas,D., Critchley,A.J. Carchon,H., Hebestreit,H.F., Hibbert,R.G., Vilarinto,L., Teles,E., Matlhijs,G. Scholler,E. *et al.* (2003) Detailed glycan analysis of serum glycoproteins of patients with congenital disorders of glycosylation indicates the specific defective glycan processing step and provides an insight into pathogenesis. *Glycobiology*, **13**, 601–622.
9. Harvey,D.J. (2001) Identification of protein-bound carbohydrates by mass spectrometry. *Proteomics*, **1**, 311–328.
10. Kuster,B., Krogh,T.N., Mortz,E. and Harvey,D.J. (2001) Glycosylation analysis of gel-separated proteins. *Proteomics*, **1**, 350–61.
11. Sagi,D., Conradt,H.S. Nimtz,M. and Katalinic,J.P. (2002) Sequencing of tri- and tetraantennary *N*-glycans containing sialic acid by negative mode ESI QTOF tandem MS. *Am. Soc. Mass Spectrom.*, **13**, 266–273.
12. Geyer,H., Schmitt,S., Wuhrer,M. and Geyer,R. (1999) Structural analysis of glycoconjugates by on-target enzymatic digestion and MALDI-TOF-MS. *Anal. Chem.*, **71**, 476–482.
13. Appel,R.D., Bairoch,A. and Hochstrasser,D.F. (1994) A new generation of information retrieval tools for biologists: the example of the ExPASy www server. *Trends Biochem. Sci.*, **19**, 258–260.

14. Cooper,C.A., Gasteiger,E. and Packer,N.H. (2001) GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*, **1**, 340–349.

15. Lohmann,K.K. and von der Lieth,C.-W. (2003) GLYCO-FRAGMENT: a web tool to support the interpretation of mass spectra of complex carbohydrates. *Proteomics*, **3**, 2028–2035.

16. Domon,B. and Costello,C.E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate*, **5**, 397–409.

17. Bohne,A., Lang,E. and von der Lieth,C. (1998) W3-SWEET: carbohydrate modeling by internet. *J. Mol. Model.*, **4**, 33–43.

18. Bohne,A., Lang,E. and von der Lieth,C.W. (1999) SWEET—www-based rapid 3D construction of oligo- and polysaccharides. *Bioinformatics*, **15**, 767–768.

19. Loss,A., Bunsmann,P., Bohne,A. Loss,A., Schwarzer,E., Lang,E. and von der Lieth,C.W. (2002) SWEET-DB: an attempt to create annotated data collections for carbohydrates. *Nucleic Acids Res.*, **30**, 405–408.