Pakistan J. Zool., vol. 44(1), pp. 53-59, 2012.

Sequence Analysis for the Terminal Sugars of Oligosaccharide Chain

Nasir-ud-Din[1,*], Afshan Kaleem[1], Wajahat Mahmood Qazi[1,2], Zeeshan Iqbal[1], Asma Butt[1], Asma Kanwal[1], Daniel C. Hoessli[1] and Muhammad Iqbal Choudhary[3]

[1]*Institute of Molecular Sciences and Bioinformatics, Lahore, Pakistan*
[2]*Cybernetics Research Lab, Department of Computer Science, GC University, Lahore, Pakistan*
[3]*HEJ Research Institute of Chemistry, Karachi, Pakistan*

Abstract.- Glycan moieties of complex carbohydrates contribute important functional roles of molecules in recognition, memory, half life and molecular transport and targeting in mammalian physiology in normal and pathological conditions. Oligosaccharide chains comprise of sugar residues in specific sequence, particular linkages and defined anomery. These particular features provide diverse function to the glycan moiety. The contribution of component sugar residues, particularly terminal sugar and their vicinal and consequent sugar residues, sequence, linkages and anomery has not been related to its function as yet. In current study, we have developed an algorithm, sequence analysis for the terminal sugar oligosaccharide chain, which is based on terminal carbohydrate moieties, Gal and GalNAc. The developed tool relates the function with terminal sugars up to tetrasaccharide at terminal and vicinal sugar residues, linkages and anomery.

Key words: Galactose/N-acetyl galactose amine, structure-function relationship, terminal sugars, data mining.

INTRODUCTION

The diverse oligosaccharide structures on glycoproteins and –lipids are involved in numerous molecular interactions in physiologic and pathologic situations, and defining these oligosaccharides can be a challenging task. When compared to DNA and protein sequences the carbohydrate structure and function is much more difficult to determine, because of the vast diversity in sequence, linkages and anomery, creating countless possible combinations. The glycan moieties of complex carbohydrates contribute important functional role to these molecules in recognition, memory, half life and molecular transport and targeting in mammalian physiology in normal and pathological conditions (Helenius and Aebi, 2004; Szabo *et al.*, 2009; Willey, 1999; Tanaka *et al.*, 2004; Sunyer *et al.*, 2008). Defining the sugar moieties can eventually suggest how glycoproteins and glycolipids contribute in health and disease states. Oligosaccharide chains comprise on sugar residues in specific sequence, particular linkages and defined anomery. The diversity of sugar residue in an oligosaccharide, different linkages and variable anomery provide diverse function to the glycan moiety. Several glycoconjugate databases are available that provide structural information about the glycan in sequence, anomery and linkages of sugar residues shown in Table I. The functional contribution of sugar residues, particularly terminal sugar and their vicinal and consequent sugar residues, sequence, linkages and anomery has not been correlated yet in any glycan sequence analysis tool.

In current study, a sequence analysis tool for the terminal sugars of oligosaccharide chain has been developed to analyze the relationship of structure of terminal sugars of oligosaccharides, including sequence, linkage and anomery with functions. This relationship is helpful in investigating the function of unknown glycans. The data covering these parameters of diverse sugars components of an oligosaccharide is a complex study and unlike protein has the capacity of performing unlimited functions arising due to different combinations of sequence, linkages and anomery.

Gal and its derivate GalNAc are widely distributed sugars in glycoproteins, -lipids, GPI-anchors, and MUC core structures. The stereochemistry of the hydroxyl groups in the pyranose ring structure of the sugar moieties dictates the function of the oligosaccharide chain. In

* Corresponding author: Institute Tel: +92 4237241653; Fax: +92 4237245753; E-mail: professor_nasir@yahoo.com
0030-9923/2012/0001-0053 $ 8.00/0

Table I.- List of different glycan databases available and their applications.

Databases	Applications	Weblinks
SATTSOC (Sequence Analysis Tool for the Terminal Sugars of Oligosaccharide Chains)	It provides information about the structure of glycans and gives information regarding the relationship between the sequence, linkage and anomery with its functions and also helps in investigation of functions of unknown glycans.	http://www.imsb.edu.pk/Default.aspx
GlycoEpitope Database	It gives information about the epitopic regions of glycan and their antibodies. In some cases the function, subcellular localization and disease involvement of the epitopic region.	http://www.glyco.is.ritsumei.ac.jp/epitope/
KEGG GLYCAN	It provides annotated glycan structures their pathways, enzymes and reactions.	http://www.genome.jp/kegg/glycan/
CCSD (Complex Carbohydrate Structure Database)	It provides literature on the basis of structure similarity of glycan	http://www.boc.chem.uu.nl/sugabase/databases.html
GlycoSuiteDB	It gives information about the type, anomery linkage, configuration, mass composition, subcellular localization, disease involvement and blood group.	http://glycosuitedb.expasy.org/glycosuite/glycodb
Glycosciences.de	It includes glycan structure search by NMR constrains, MS fragments, conformational maps, and the protein carbohydrate complexes from pdb data bank.	http://www.glycosciences.de/
Bacterial Carbohydrate Structure DataBase	It provides annotated glycan structures present in bacteria.	http://www.glyco.ac.ru/bcsdb3/

case of Gal the axial OH group on position 4 represents structural variation from other hexoses, like glucose, and this attribute determines its recognition and binding properties. Work to establish the structure-function relationship of complex, non-linear and highly heterogeneous oligosaccharides need to be done in order to determine the structural variations in the sugar residues into specific functions.

The structure-function relationship is established as a set of association rules (AR) based on tetrasaccharide chain. AR mining is an unsupervised machine learning method used in data mining (Kantardzic, 2002). These rules can identify patterns with the degree of correlation that might not have been known through other machine learning methods (Creighton and Hanash, 2003). AR defines an implication A=>B (Agarwal *et al.*, 1993; Chen *et al.*, 1996), where A and B are frequent item-sets in a dataset. Here B is a set, likely to occur, whenever set A occurs with a particular confidence level. ARs are symmetric in nature as no attribute is given a special treatment. Therefore, attributes on right hand side (RHS) of ARs are not predefined. This method has been proven to be a useful tool for the descriptive analysis of biological sequences (Ahmad *et al.*, 2008a, b; Ahmad *et al.*, 2009; Georgii *et al.*, 2005; Ji and Tan, 2004; Oyama *et al.*, 2002; Qazi and Ahmad, 2010).

MATERIALS AND METHODS

Data for AR mining is extracted from newly designed data, which include the oligosaccharide chain, terminal sugar, its vicinal sugar and two subsequent sugars. The consistency of information

in the newly designed data for tool can be ensured by the removal of redundant data and misinformation in oligosaccharide chain, terminal sugar, etc.

Structure-function relation was established in two steps. First step was to mine ARs, second step was to implement the structure-function relation in the data utilizing the AR as a knowledge base.

ARs are conditional rules defining correlation between tetrasaccharide chain. These ARs are generated on the basis of frequent item sets, which are items that exist in significant portions of data for tool (Pasquier *et al*., 2008). The most appropriate algorithm for AR is Apriori algorithm. Apriori is an efficient algorithm which utilizes large item set to find frequent items in the glycan data using two rule construction values: support value and confidence value (Niimi and Tazaki, 2000). These values perform optimization of data for considering any frequent items as acceptable rule. Support value is defined for the pruning of frequent items; it defines the least percentage of occurrence of frequent items in data. If the confidence of mined rule exceeds from minimum confidence criteria, then the rule is considered in the final mined AR.

In this tool AR mining was performed by implementing a modified version of Apriori algorithm (Agarwal *et al.*, 1993) in C#. The AR is asymmetric in nature, because the member of RHS is the function associated with the oligosaccharide chain, which is predetermined. Therefore, the variation of Apriori algorithm implemented for this study is hereafter called AApriori implemented.

The data for AR mining was prepared by grouping the entries of the glycan data on the bases of the functions of oligosaccharide chains. For each oligosaccharide chain in this data, set of tetra-oligosaccharide from each terminal point from the chain were extracted. This preparation resulted into new dataset. This newly created dataset was presented to AApriori implementation to mine AR at various support levels. The support levels selected were 1, 5, 10 and 15. Indeed the support level 1 is very low, but it reveals all possible AR. The reason for selecting support level 1 was that the data representation for some of oligosaccharide's function was significantly low. At high support level AApriori would not be able to mine any rules for such functions. These mined rules provide all possible correlated knowledge for existing data to entertain any query in its every relevant aspect. These rules define the strongly correlated glycan moieties to express their associated functions. The rules mined are the part of the analysis tool as a knowledge base for the structure-function relationship module (SFRM).

SFRM was implemented using ASP.NET. SFRM module required tetra-oligosaccharide to establish possible structure function relationship for the given chain. It maps the tetra-oligosaccharide with the rules and determines what possible functions are associated with the given chains. These establishments are supported by two statistical quantities i.e. support level of the rule and confidence.

The analysis tool was written in C# Dot Net. The web interface for search and analysis of structure-function relationship and for data management was developed in ASP.net and Microsoft SQL Server 2005, respectively.

This analysis tool facilitates the user to store, retrieve and analyze curated information/data of the oligosaccharide sequences. Analysis of the oligosaccharide chain, up to tetrasaccharide is the most important phase of the analysis-tool. This tool incorporates four search operations:

i) Search by ID
ii) Search by function
iii) Search by terminal moiety and modification
iv) Search by oligosaccharide chain

This tool provides facility to search oligosaccharide chain by DB ID, the format of DB ID is "SAT_[4 digit number]" (example SAT_0003). The search result contain the necessary information about the specific oligosaccharide chain including its IUPAC sequence, structure, name, class, composition, detail function, references and links for these references. This tool also allows user to search oligosaccharide sequences regarding to its specific function by inserting information about terminal sugar, linkage and anomery. Most importantly, this analysis tool allows the user to search oligosaccharide chain in a diverse way. The user can search different oligosaccharide sequences by specifying sugars, linkages, anomery and

modifications up to tetrasaccharides.

The tool allows the user to analyze the structure function relationship of glycans up to 4 sugar moieties from the terminal end.

RESULTS AND DISCUSSION

The glycans attached to proteins or lipids act as receptors for different types of pathogens, and regulate cell adhesion, cell-cell interaction, recognition, cell growth, mobility and motility (Helenius and Aebi, 2004; Szabo *et al.*, 2009; Willey, 1999; Tanaka *et al.*, 2004; Sunyer *et al.*, 2008; Hakomori, 1996). Thus defining the role and diversity of glycans associated with different diseases can be helpful in disease prevention and treatment by vaccine and drug targeting. Aberrant glycosylation is mostly associated with tumor progression and metastasis. Specifically gangliosides and sphingoglycolipids are highly expressed in tumor cells and thus can be used for investigating invasion, cell adhesion, malignancy and metastatic potential associated with different types of tumors (Hakomori, 1996). Sulfated glycoproteins, obtained from the serum of patients of cystic fibrosis (Mawhinney *et al.*, 1987; Lamblin *et al.*, 1991) and chronic bronchitis (Degroote *et al.*, 2003), are not present in normal serum, and can be potential immune targets. Simply by altering the sequence, linkage or anomery of the oligosaccharide the function of the glycoprotein or –lipid changes. An excellent example of this phenomenon is the blood group system, where recognition is solely done through the non-reducing carbohydrate moiety on erythrocytes. By replacing terminal Gal with GalNAc on erythrocytes, the blood group changes from B to A (Cao *et al.*, 2007). Similarly when GalNAc is replaced by GlcNAc of the epitope GalNAc(β1-4)Gal attached to lipid, which is recognized by the bacterium *Streptococcus pneumoniae*, binding is inhibited (Haataja *et al.*, 1993).

In this study the structure-function relationship of oligosaccharides containing terminal Gal and/or GalNAc is investigated. The relationship of sequence, linkage and anomery with function is analyzed up to four subsequent sugar residues from the non-reducing end. In this study 125 functional entries are collected on the basis of terminal Gal/GalNAc moiety, and are sub-divided into fragments on the basis of branches in order to avoid any inconvenience. Each sugar moiety of terminal tetrasaccharide of every fragment is analyzed for its occurrence with a specific function and the confidence level are assigned for individual sugar residues as well as for different sets of combinations. The GalNAc-GD1a is a glycolipid and is known to be associated with brain and liver cancer cells. GalNAc-GD1a is branched structure containing two neuraminic acid groups at 2 and 4 positions. When the analysis tool was applied, the GalNAc-GD1a was first divided into three fragments, where each branch is considered as one template. Then every template is sub-divided as shown in (Fig. 1). For the first fragment 46 rules are obtained, while 50 for second and 90 for third fragment at various support levels. The top seven rules on the basis of confidence level are also given in (Table II). These rules are given accordingly to the confidence for which the given sugar residue/residues in combination are involved in specific function. All three fragments give high confidence levels for the involvement in cancer. The maximum confidence 71% is obtained for the sugar 3 GalNAc(β1-4) and for combine sugars 3^4 GalNAc(β1-4)Gal(β1-4) for being involved in cancer associated antigens. 60% confidence is obtained for the involvement of sugars 2^3 Gal(β1-3)GalNAc(β1-4) and 2^3^4 Gal(β1-3)GalNAc(β1-4)Gal(β1-4) in cancer, while Neu5Ac(α2-3) is assigned with 50% confidence level. The rules are also supported by the fact that different sialylated glycolipids containing Gal(β1-3)GalNAc(β1-4)Gal(β1-4) exact/related sequence are known to be involve as cancer antigens having different sub cellular localizations (Hakomori, 1996; Fredman *et al.*, 1989; Ilyas *et al.*, 1988). Also it is well known that asialo-glycolipids are highly expressed in cancer cells (Hellstrom *et al.*, 1990). By using this tool to investigate the function on the basis of sequence, linkage and anomery, it will provide a clear picture of the possible role of a specific epitope in a given environment, and will guide experimentalist for further investigations.

The developed tool is a novel and useful carbohydrate sequence analysis tool for researchers

Table II.- Association rules for terminal tetrasaccharide for GalNAc-GD1a.

Mode	LHS	RHS	Confidence
Association Rules for first fragment of GalNAc-GD1a			
Sugar3	GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.7142857
Combined-3-4	GalNAc(b1-4)^Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.7142857
Combined-2-3	Gal(b1-3)^GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.6
Combined-2-3-4	Gal(b1-3)^GalNAc(b1-4)^Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.6
Sugar1	Neu5Ac(a2-3)	CANCER ASSOCIATED ANTIGENS	0.5
Sugar4	Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.4651163
Sugar2	Gal(b1-3)	CANCER ASSOCIATED ANTIGENS	0.2258064
Association Rules for second fragment of GalNAc-GD1a			
Sugar3	GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.7142857
Combined-3-4	GalNAc(b1-4)^Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.7142857
Combined-2-3	Gal(b1-3)^GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.6
Combined-2-3-4	Gal(b1-3)^GalNAc(b1-4)^Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.6
Sugar4	Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.4651163
Sugar1	GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.3157895
Sugar1	GAlNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.3157895
Association Rules for third fragment of GalNAc-GD1a			
Sugar1	Neu5Ac(a2-3)	CANCER ASSOCIATED ANTIGENS	0.5
Sugar2	Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.3333333
Sugar3	Glc(b1-1)	CANCER ASSOCIATED ANTIGENS	0.2727273
Combined-2-3	Gal(b1-4)^Glc(b1-1)	CANCER ASSOCIATED ANTIGENS	0.2727273
Combined-2-3-4	Gal(b1-4)^Glc(b1-1)^Cer	CANCER ASSOCIATED ANTIGENS	0.2727273
Combined-3-4	Glc(b1-1)^Cer	CANCER ASSOCIATED ANTIGENS	0.2727273
Sugar4	Cer	CANCER ASSOCIATED ANTIGENS	0.2142857

Means Mode => position of the moieties involved in association rules, LHS => Association Rules, RHS => Function

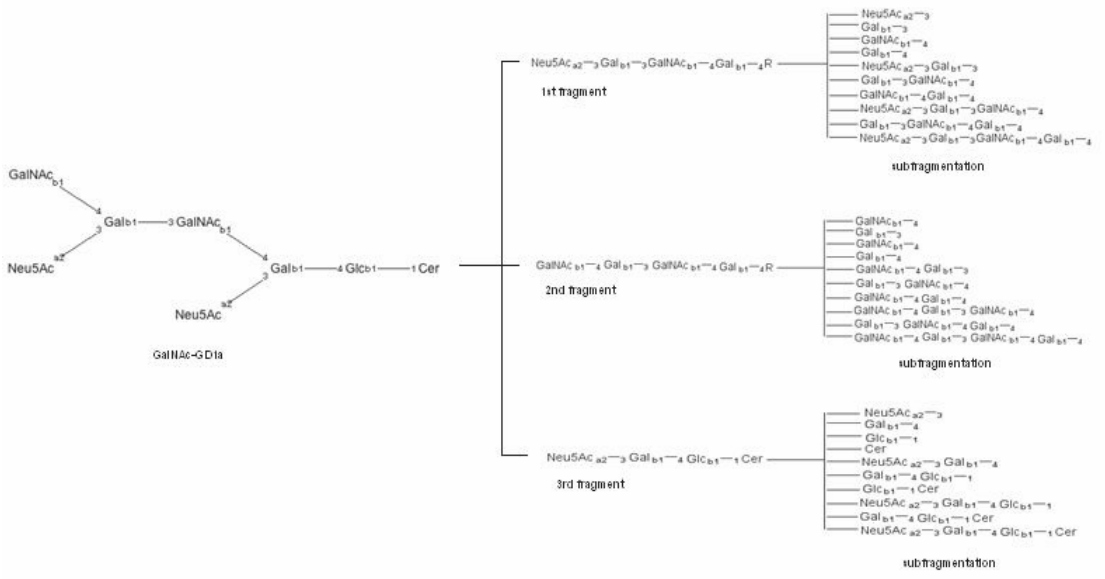


Fig. 1. The fragmentation of GalNAc-GD1a is described. First the branches of glycan structure are separated. The tetrasaccharide of each fragment is shown while all other moieties are set as R. The tetrasaccharide of each fragment is then sub-divided into mono-, di- and trisaccharides using different combinations. These sub-fragments are then set for data mining.

working with complex carbohydrate structures. It is a simple and easy approach for describing the functions of sugars upto tetrasaccharide on the basis of sequence, linkage and anomery. The tool is being extended to define the structure function relationship of complete oligosaccharide chain. Furthermore, a large number of epitopes comprise of branched carbohydrate structures. In the near future this issue will be addressed, and the tool developed will be further extended and define the role of branched sugars as well. The functional information provided through this algorithm is based on the reported experimental data, and all entries are referred. There is no existing glycan tool that describes the structure and function of oligosaccharides, but to date this tool is the only one, which defines the function based on comprehensive structure (sugar component, sequence, linkage and anomery).

ACKNOWLEDGEMENT

Nasir-ud-Din acknowledges partial support from HEC (at H.E.J. Karachi University), Pakistan, EMRO-COMSTECH and Pakistan Academy of Sciences for this work.

REFERENCES

AGARWAL, R., IMIELINSKI, T. AND SWAMI, A., 1993. Mining association rules between sets of items in large databases. *Proc. of the 1993 ACM SIGMOD Conference*, 207-216.

AHMAD, I., HOESSLI, D.C., QAZI, W.M., KHURSHID, A., MEHMOOD, A., WALKER-NASIR, E., AHMAD, M., SHAKOORI, A.R. AND NASIR-UD-DIN., 2008a. MAPRes: An efficient method to analyze protein sequence around post-translational modification sites. *J. Cell. Biochem.*, **104**: 1220-1231.

AHMAD, I., MEHMOOD, A., KHURSHID, A., QAZI, W.M., HOESSLI, D.C., WALKER-NASIR, E., SHAKOORI, A.R. AND NAIR-UD-DIN., 2009. Phosphoproteome sequence analysis and significance: Mining association patterns around phosphorylation sites utilizing MAPRes. *J. Cell. Biochem.*, **108**: 64-74.

AHMAD, I., QAZI, W.M., KHURSHID, A., AHMAD, M., HOESSLI, D.C., KHAWAJA, I., CHOUDHARY, M.I., SHAKOORI, A.R. AND NAIR-UD-DIN., 2008b. MAPRes: Mining association patterns among preferred amino acid residues in the vicinity of amino acids targeted for post-translational modifications. *Proteomics*, **8**: 1954-1958.

CAO, S., LOU, Z., TAN, M., CHEN, Y., LIU, Y., ZHANG, Z., ZHANG, X.C., JIANG, X., LI, X. AND RAO, Z., 2007. Structural basis for the recognition of blood group trisaccharides by norovirus. *J. Virol.*, **81**: 5949–5957.

CHEN, M.S., HAN, J. AND YU, P.S., 1996. Data Mining: An overview from database perspective. *IEEE Transactions on Knowledge and Data Engineering*, **8**: 866-883.

CREIGHTON, C. AND HANASH, S., 2003. Mining gene expression databases for association rules. *Bioinformatics*, **19**: 79-86.

DEGROOTE, S., MAES, E., HUMBERT, P., DELMOTTE, P., LAMBLIN, G. AND ROUSSEL, P., 2003. Sulfated oligosaccharides isolated from the respiratory mucins of a secretor patient suffering from chronic bronchitis. *Biochimie*, **85**: 369-379.

FREDMAN, P., MANSSON, J-E., WIKSTRAND, C-J., VRIONIS, F.D., RYNMARK, B-M., BIGNER, D.D. AND SVENNERHOLM, L., 1989. A new ganglioside of the lactotetraose Series, GalNAc-3'-isoL$_{M1}$ detected in human meconium. *J. Biol. Chem.*, **264**: 12122-12125.

GEORGII, E., RICHTER, L., RUCKERT, U. AND KRAMER, S., 2005. Analyzing micro array data using quantitative association rules. *Bioinformatics*, **21**: ii123-ii129.

HAATAJA, S., TIKKANEN, K., LIUKKONEN, J., FRANQOIS-GERARD, C. AND FINNE, J., 1993. Characterization of a novel bacterial adhesion specificity of *Streptococcus suis* recognizing blood group P receptor oligosaccharides. *J. Biol. Chem.*, **268**: 4311-4317.

HAKOMORI, S-I., 1996. Tumor malignancy defined by aberrant glycosylation and sphingo (glyco)lipid metabolism. *Cancer Res.*, **56**: 5309-5318.

HELENIUS, A. AND AEBI, M., 2004. Roles of N-Linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.*, **73**: 1019-1049.

HELLSTROM, I., BROWN, J.P., HELLSTROM, K.E., HORN, D. AND LINSLEY, P., 1990. Monoclonal antibodies and antigen for human non-small cell lung carcinomas. USA4906562.

ILYAS, A.A., LI, S-C., CHOU, D.H., LI, Y-T., JUNGALWALA, F.B., DALAKAS, M.C. AND UARLES, R.H., 1988. Gangliosides G_{M2}, $IV^4Ga1NAcG_{Mlb}$, and $IV^4GalNAcG_{D1a}$ as antigens for monoclonal immunoglobulin M in neuropathy associated with gammopathy. *J. Biol. Chem.*, **263**: 4369-4373.

JI, L. AND TAN, K.L., 2004. Mining gene expression data for positive and negative co-regulated gene clusters. *Bioinformatics*, **20**: 2711-2718.

KANTARDZIC, M., 2002. Data Mining: Concepts, models, methods, and algorithms. *Wiley-IEEE Press.*, 135-158.

LAMBLIN, G., RAHMOUNE, H., WIERUSZESKI, J.M., LHERMITTE, M., STRECKER, G. AND ROUSSEL, P., 1991. Structure of two sulphated oligosaccharides

from respiratory mucins of a patient suffering from cystic fibrosis. A fast-atom-bombardment M.S. and 1H-N.M.R. spectroscopic study. *Biochem. J*., **275**: 199-206.

NIIMI, A. AND TAZAKI, E., 2000. Rule discovery technique using genetic programming combined with apriori algorithm. *S. Arikawa and S. Morishita (Eds.): DS* 2000, 273-278.

MAWHINNEY, T.P., ADELSTEIN, E., MORRIS, D.A., MAWHINNEY, A.M. AND BARBERO, G.J., 1987. Structure determination of five sulfated oligosaccharides derived from tracheobronchial mucus glycoproteins. *J. Biol. Chem.*, **262**: 2994-3001.

OYAMA, T., KITANO, K., SATOU, K. AND ITO, T., 2002. Extraction of knowledge on protein-protein interaction by association rule discovery. *Bioinformatics*, **18**: 705-714.

PASQUIER, N., PASQUIER, C., BRISSON, L. AND COLLARD, M., 2008. Mining gene expression data using domain knowledge. *Int. J. Software Informatics.*, **2**: 215-231.

QAZI, W.M. AND AHMED, K., 2010. Machine learning in bioinformatics: An approach to protein sequence analysis, VDM Publishers, Germany, ISBN: 978-3639253726

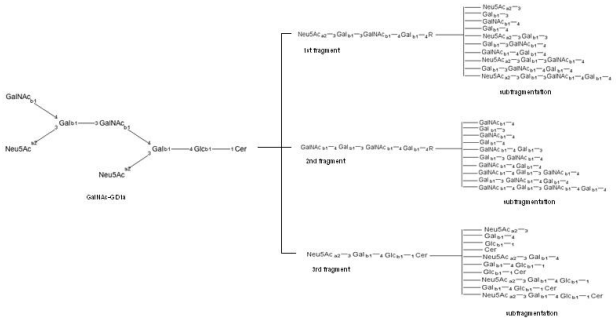
SUNYER, B., DIAO, W. AND LUBEC, G., 2008. The role of post-translational modifications for learning and memory formation. *Electrophoresis*, **29**: 2593-2602.

SZABO, T.G., PALOTAI, R., ANTAL, P., TOKATLY, I., TOTHFALUSI, L., LUND, O., NAGY, G., FALUS, A. AND BUZAS, E.I., 2009. Critical role of glycosylation in determining the length and structure of T cell epitopes. *Immunome Res.*, **5**: Doi: 10.1186/1745-7580-5-4.

TANAKA, K., XU, W., ZHOU, F. AND YOU, G., 2004. Role of glycosylation in the organic anion transporter OAT1. *J. Biol. Chem.*, **279**: 14961-14966.

WILLEY, K.P., 1999**.** An elusive role for glycosylation in the structure and function of reproductive hormones. *Hum. Reprod updates*, **5**: 330-355.

(Received 11 June 2010, revised 17 October 2011)


GalNAc
Gal
GalNAc
Gal
Glc
Cer
Neu5Ac
Neu5Ac
GalNAc-GD1a
1st fragment
2nd fragment
3rd fragment
subfragmentation
subfragmentation
subfragmentation

Association Rules for first fragment of GalNAc-GD1a			
Mode	**LHS**	**RHS**	**Confidence**
Sugar3	GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.7142857
Combined-3-4	GalNAc(b1-4)^Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.7142857
Combined-2-3	Gal(b1-3)^GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.6
Combined-2-3-4	Gal(b1-3)^GalNAc(b1-4)^Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.6
Sugar1	Neu5Ac(a2-3)	CANCER ASSOCIATED ANTIGENS	0.5
Sugar4	Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.4651163
Sugar2	Gal(b1-3)	CANCER ASSOCIATED ANTIGENS	0.2258064
Association Rules for second fragment of GalNAc-GD1a			
Mode	**LHS**	**RHS**	**Confidence**
Sugar3	GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.7142857
Combined-3-4	GalNAc(b1-4)^Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.7142857
Combined-2-3	Gal(b1-3)^GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.6
Combined-2-3-4	Gal(b1-3)^GalNAc(b1-4)^Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.6
Sugar4	Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.4651163
Sugar1	GalNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.3157895
Sugar1	GAlNAc(b1-4)	CANCER ASSOCIATED ANTIGENS	0.3157895
Association Rules for third fragment of GalNAc-GD1a			
Mode	**LHS**	**RHS**	**Confidence**
Sugar1	Neu5Ac(a2-3)	CANCER ASSOCIATED ANTIGENS	0.5
Sugar2	Gal(b1-4)	CANCER ASSOCIATED ANTIGENS	0.3333333
Sugar3	Glc(b1-1)	CANCER ASSOCIATED ANTIGENS	0.2727273
Combined-2-3	Gal(b1-4)^Glc(b1-1)	CANCER ASSOCIATED ANTIGENS	0.2727273
Combined-2-3-4	Gal(b1-4)^Glc(b1-1)^Cer	CANCER ASSOCIATED ANTIGENS	0.2727273
Combined-3-4	Glc(b1-1)^Cer	CANCER ASSOCIATED ANTIGENS	0.2727273
Sugar4	Cer	CANCER ASSOCIATED ANTIGENS	0.2142857

Means Mode => position of the moieties involved in association rules, LHS => Association Rules, RHS => Function