

Sharing of worldwide distributed carbohydrate-related digital resources: online connection of the *Bacterial Carbohydrate Structure DataBase* and *GLYCOSCIENCES.de*

Philip Toukach, Hiren J Joshi¹, René Ranzinger¹, Yuri Knirel and Claus-W. von der Lieth^{1,*}

N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russia and
¹German Cancer Research Center, Spectroscopic Department (B090), Molecular Modeling, Im Neuenheimer
Feld 280, D-69120 Heidelberg, Germany

Received August 16, 2006; Revised and Accepted October 11, 2006

ABSTRACT

Functional glycomics, the scientific attempt to identify and assign functions to all glycan molecules synthesized by an organism, is an emerging field of science. In recent years, several databases have been started, all aiming to support deciphering the biological function of carbohydrates. However, diverse encoding and storage schemes are in use amongst these databases, significantly hampering the interchange of data. The mutual online access between the Bacterial Carbohydrate Structure DataBase (BCSDB) and the GLYCOSCIENCES.de portal, as a first reported attempt of a structure-based direct interconnection of two glyco-related databases is described. In this approach, users have to learn only one interface, will always have access to the latest data of both services, and will have the results of both searches presented in a consistent way. The establishment of this connection helped to find shortcomings and inconsistencies in the database design and functionality related to underlying data concepts and structural representations. For the maintenance of the databases, duplication of work can be easily avoided, and will hopefully lead to a better worldwide acceptance of both services within the community of glycoscientists. BCSDB is available at <http://www.glyco.ac.ru/bcsdb/> and the GLYCOSCIENCES.de portal at <http://www.glycosciences.de/>

INTRODUCTION

Functional glycomics is an emerging field of science, aiming to create a cell-by-cell catalogue of glycosyltransferase

expression and detected glycan structures in relation to health and diseases (1–3). Analysis of glycans has proven difficult in the past due to their structural complexity. However, modern analytical methods such as mass spectrometry and NMR have afforded the ability to elucidate most structural details at the concentration levels required for glycomics (4,5). Several national and international initiatives aiming to decipher the biological function of carbohydrates have emerged for the recent years (6–8). In a similar fashion to the finished Human Genome Project—which determined the sequences of the chemical base pairs that make up a human DNA—most of these glycomics projects intend to make their data freely accessible under an open access philosophy. Unfortunately, the exchange of data between different glyco-related databases is seriously hampered by the dearth of generally accepted digital exchange formats and standardized structural and biological descriptions (9). Similar to the genomics and proteomics field, a description of glycan structures would be an appropriate way to establish an efficient connection of glyco-related information resources. However, glycan sequences cannot be described by a simple linear one-letter code as each pair of monosaccharides can be linked in several ways and branched structures can be formed. The *GLYCOSCIENCES.de* portal (7) demonstrates that data originating from various resources can be efficiently integrated using a linear notation for unique description of carbohydrate sequences (LINUCS) (10). The extended alphanumeric IUPAC description and glycosidic linking information are applied to build up a hierarchy of the various branches starting from the reducing end of the oligosaccharide chain, which is then converted into a linear representation. However, other larger projects use different ways to encode glycan structures. The commercially available *GlycoSuiteDB* (11) uses the so-called condensed form of the IUPAC description to create a linear representation, where four rules are applied to obtain a unique linear code. The ‘*Glycan Database*’ of the US Consortium for Functional Glycomics (6) uses the so-called

*To whom correspondence should be addressed. Tel: +49 6221 424541; Fax: +49 6221 42454; Email: w.vonderlieth@dkfz.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© 2006 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Linear Code™ (12), using a one or two character-based representation of saccharide units and linkages. The ordering of glycan branches is established using a special lookup table where the hierarchy of monosaccharide structures is defined. The KEGG Carbohydrate Matcher (KCaM) (8,13) uses a connection table based graph representation to encode carbohydrate structures, where monosaccharides are represented by nodes and glycosidic bonds as edges.

GLYCOSCIENCES.de, *GlycoSuite*, *CFG Glycan Database* and *KEGG-Glycan* concentrate on glycan structures found in mammalian species. In contrast, the mission of the Russian Bacterial Carbohydrate Structure DataBase (*BCSDB*) [(14), for URLs see Appendix] is to provide all published glycan structures found in bacteria. Since the monosaccharide namespace as well as the type of linkages found in bacterial polysaccharides differ considerably from those found in mammals, *BCSDB* uses an internal representation of glycans, which diverges from those used to describe structures found in mammals.

Looking at various existing carbohydrate databases accessible through the Internet, it is obvious that diverse ways to encode and store complex carbohydrates are in use. They all seem to work satisfactorily for the purpose they have been designed for. However, users who would like to access all publicly available glyco-related data spread over many databases have not only to cope with varying graphical and non-graphical interfaces to input glycan structures, but also must be aware that the definition of building blocks and topologies may be different. Each database has developed its own set of rules to solve some problematic encoding situations such as treatment of monovalent substituents, phosphates, sulphates, repeat units, unknown linkages and other uncertain structural features of glycan structures.

It is, of course, an attractive vision [expressed during the Joint Meeting of the Japanese and American Consortia for Glycomics (15)] to have a single user interface which will provide access to all relevant world-wide distributed resources without any technical and administrative barrier. A prerequisite for an efficient exchange of data is the agreement to a generally accepted exchange format as well as a common application programming interface. Consequently, several proposals for an XML-based description of glycan structures have already been published (16,17). To avoid any further confusion about XML descriptions of glycans, the seven larger initiatives in this field [*CFG*, *BCSDB*, *GLYCOSCIENCES.de*, *EUROCarbDB*, *KEGG*, *HGPI* and *CCRC* (for abbreviations see Appendix)] agreed to further develop the XML description for the encoding of glycan structures on the basis of the already existing *GLYcan Data Exchange (GLYDE)* (17). The progress discussion is open to all interested scientists and takes place at the forum pages of the *EUROCarbDB* project.

Concerning the technical realization of the online connection between existing databases, it seems that the Simple Object Access Protocol (*SOAP*) is now the broadly accepted procedure for automated communication between web-applications. Being designed to communicate via the Internet, it is well suited to be also used for the exchange of glycan-related data between distributed computers. Taken together, it seems like the field has matured to the point where it is feasible to establish an online connection of distributed

databases, at least between the larger of the established projects.

BACTERIAL CARBOHYDRATE STRUCTURE DATABASE

The *Bacterial Carbohydrate Structure DataBase (BCSDB)* [(14), for URLs see Appendix] is a database containing data on natural carbohydrates with known structure. In addition to the structure and bibliography, each record in the *BCSDB* contains the abstract of the publication, data on the carbohydrate source, methods of structure elucidation, information on the availability of spectral data and assignment of NMR spectra when available, data on conformation, biological activity, chemical and enzymatic synthesis, biosynthesis, genetics and other related data. The search criteria can be fragment(s) of the structure; fragment(s) of the NMR spectrum; and indexed tags, including microorganism, bibliography and keywords.

Currently, the *BCSDB* contains ~8200 records on bacterial carbohydrates, including the corresponding part of *CarbBank* (18) (~3500 records on structures reported before 1995). This coverage is approaching the total number of bacterial carbohydrate structures ever reported. Data from both literature and *CarbBank* have been carefully checked for consistency before the upload, and corrected when necessary. The *BCSDB* interface includes the web-based user part, web-based administrator part and programming gateways for the automated data interchange. The *BCSDB* is available on the Internet for free usage and validated user data submission.

GLYCOSCIENCES.de

The *GLYCOSCIENCES.de* portal (7) is an attempt to link glycan-related data originating from various resources through a unique structural description. The *LINUXS* (LInear Notation for Unique description of Carbohydrate Sequences) (10) notation is used to uniquely encode fully characterized glycans. Currently, the *GLYCOSCIENCES* portal provides access to ~24 000 different entries with nearly 14 000 different carbohydrate moieties. These structures are sourced from a number of sources, including the former *CarbBank* and *SugaBase*-project (19), automatic extraction from the Protein Data Base (*PDB*) (20), and the curation of new entries altogether. The structure-oriented approach to the database allows the data related to a single glycan, but originating from various sources (e.g. experimental NMR spectra, theoretically calculated fragment ions for mass spectra interpretation or experimental or simulated 3D structures) to be easily linked and accessed using a single database query. According to the varying needs of specific research questions, the *GLYCOSCIENCES* portal provides several structure-oriented options to recall glycan-related data. Substructure searches are the most frequently used way to look for glycan structures. The retrieval of glycans matching an exact structure is the most traditional way to access a database. The motif search enables to retrieve all entries, which possess substructures having names such as Lewis^X, blood group H antigen or GM3. All glycan-related scientific data of the *GLYCOSCIENCES.de* portal are freely accessible via the

Retrieval request: of references of author 'Brade' in year 2002 in any journal. Used was the *GLYCOSCIENCES.de* advanced bibliographic search. (Only results are shown)

Two entries were found in *GLYCOSCIENCES.de*

Chemical structure and immunoreactivity of the lipopolysaccharide of the deep rough mutant I-69 Rd-/b+ of Haemophilus influenza Muller-Loennies S; Brade L; Brade H Published 2002 in Eur J Biochem, 269: 1237-42 ReferenceID: 15361 PMID: 11856357
Reference contains 2 structures.

Structural analysis of deacylated lipopolysaccharide of Escherichia coli strains 2513 (R4 core-type) and F653 (R3 core-type). Muller-Loennies S; Lindner B; Brade H Published 2002 in Eur J Biochem, 269: 5982-91 ReferenceID: 15372 PMID: 12444988
Reference contains 1 structure.

Additional four papers are found in *BCSDB*

A novel strategy for the synthesis of neoglycoconjugates from deacylated deep rough lipopolysaccharides Muller-Loennies S, Grimmecke D, Brade L, Lindner B, Kosma P, Brade H. Published 2002 in Journal of Endotoxin Research, 8: 295-305 PMID: 12230919

The structure of the carbohydrate backbone of the lipopolysaccharide from Acinetobacter baumannii strain ATCC 19606 Vinogradov EV, Duus JO, Brade H, Holst O. Published 2002 in European Journal of Biochemistry, 269: 422-430 PMID: 11856300

Monoclonal antibodies against 3-deoxy-a-D-manno-oct-2-ulosonic acid (Kdo) and D-glycero-D-talo-oct-2-ulosonic acid (Ko) Brade L, Gronow S, Wimmer N, Kosma P, Brade H, Published 2002 in Journal Endotoxin Research, 8: 358-364

A single nucleotide exchange in the wzy gene is responsible for the semirough O6 lipopolysaccharide phenotype and serum sensitivity of Escherichia coli strain Nissle 1917 Grozdanov L, Zaehring U, Blum-Oehler G, Brade L, Henne A, Knirel YA, Schombel U, Schulze J, Sonnenborn U, Gottschalk G, Hacker J, Rietschel ET, Dobrindt U. Published 2002 in Journal of Bacteriology, 184: 5912-25

Example 1. Retrieval request of references of author 'Brade' in year 2002 in any journal. Used was the *GLYCOSCIENCES.de* advanced bibliographic search (only results are shown). References from *BCSDB* contain one structure each.

Internet following the open access philosophy: 'free availability and unrestricted use'.

WEB-SERVICES

The SOAP-based web-services are available on the websites of the two projects and are documented in the form of WSDL (for URLs see Appendix) descriptions that provide the possibility of platform-independent formalization of server-side features. WSDL files can be easily integrated into the existing code by using features from various SOAP libraries which allow the transparent work with the SOAP interface under Perl, PHP, Java, etc. Additionally sample PHP clients are available.

DATA TRANSFER FORMATS

GLYcan Data Exchange (GLYDE) version 1.2 (17) was chosen as the structure exchange format. It supports almost all known peculiarities of carbohydrate structures, such as uncertainties in configuration and ring sizes, various combinations of repeating and non-repeating parts, non-carbohydrate linkers, cyclic structures, etc.. GLYDE uses a tree-based approach to structure description. Within this approach the tree root is the reducing and or the rightmost residue in the repeating unit, while all the substituents are the 'children' of the residue they are attached to. Configurations, ring-size and other related information is stored as attributes of the residue. The syntax of GLYDE is XML.

To transfer the bibliographic information two approaches are used: the raw data (as array of strings corresponding to

authors, title terms, journal name, etc.) or PubMed XML. *BCSDB* supports both formats, while *GLYCOSCIENCES.de* currently supports only the former. The former format is simpler in realization but the latter provides more standardization. PubMed XML encodes the bibliographic information using the strictly defined set of rules. More information is available at the NCBI PubMed XML tagged data homepage.

A well-known identifier for an organism is a TaxID provided by NCBI taxonomy database. Both databases provide the search mechanism that uses NCBI TaxID to identify the microorganism. However, the ranking of TaxID is limited to species; thus, no possibility to cross-search for particular strains/serogroups is provided. As this detailed ranking is significant mainly for bacteria, the capability to perform deep species searching is only supported on the *BCSDB* side of the connection. TaxIDs are stored in the *GLYCOSCIENCES.de* database together with structures, while *BCSDB* generates TaxIDs based on genus and species name, making use of an NCBI web service.

EXAMPLES

Three examples are given to demonstrate the established interconnection of both data collections. Example 1, using the bibliographic search of *GLYCOSCIENCES.de*, shows all references found in both resources for author 'Brade' in year 2002. *GLYCOSCIENCES.de* has included only two papers, where NMR spectra are reported. *BCSDB* lists another five papers where the structures of bacterial polysaccharides are described. Example 2 depicts a sub-structure search containing a specified disaccharide fragment

[α -D-Neup5NAc-(2-3)- β -D-Galp] in *GLYCOSCIENCES.de*. The structure input option implemented in *BCSDB* (see Example 2a) is used. The data associated with two entries containing the disaccharide fragment are shown in Example 2b. Example 3 demonstrates a substructure search in *BCSDB* using *GLYCOSCIENCES.de* to input the trisaccharide fragment α -D-Galp-(1-3)- α -D-Manp-(1-4)- α -L-Rhap.

CONCLUSIONS

The capability of web services to make distributed scientific data accessible is clearly demonstrated. To our knowledge,

the implemented mutual online access between *BCSDB* and *GLYCOSCIENCES.de* is the first reported attempt of a structure-based interconnection of two glyco-related databases. For users the advantages are obvious: they can use and have to learn only one interface, always have access to the latest data from both services, and the results of both searches are presented in a consistent way. For the database design and its functionality the establishment of a connection helped to find shortcomings and inconsistencies in both underlying data concepts and structural representations. For the maintenance of the databases, duplication of work can be easily avoided. It can be expected that more frequent use of both services will improve the quality of data.

Substructure search

Expert mode Wizard, 2 residues (A->B) (A)→(B)

Substructure: aDNeup5Ac(2-3)bDGalp

Scope: monomers oligomers any repeating units biological repeating units cyclic

Residue (A): aDNeup5Ac(2-

<div style="display: flex; justify-content: space-between;"> a neuraminic acid (pyranose) <input type="checkbox"/> add substitution </div> <p>aDNeup substitutes C3 of Residue B</p> <p><input type="checkbox"/> is terminal</p>	<input checked="" type="checkbox"/> add substituent Acetylated at 5 <input type="checkbox"/> add substituent <input type="checkbox"/> add substituent <input type="checkbox"/> add substituent
--	---

Residue (B): bDGalp

<div style="display: flex; justify-content: space-between;"> b D galactose (pyranose) <input type="checkbox"/> add substituent </div> <p>bDGalp</p> <p><input type="checkbox"/> has aglycone</p>	<input type="checkbox"/> add substituent <input type="checkbox"/> add substituent <input type="checkbox"/> add substituent <input type="checkbox"/> add substituent
--	--

Search in: all database results of the previous query (*none*)

Search exact structure (not a fragment) **Go!** & display 30 records per page.

Make GLYDE
Search in GlycoSCIENCES
Home
Help

Example 2a. Substructure Search for a -D-Neup5NAc-(2-3)- b -D-Galp in Glycosciences.de using the BCSDB Input wizard.

GlycoSCIENCES data retrieval

GlycoSCIENCES DB returned 1845 structure(s) matching your request
Displaying **30** structures beginning from **0**

[Next 30 structures](#)

Structure (LinusID= 1175). To go to the corresponding Glycosciences DB page [click here](#).

```

a-D-Fucp-(1-3)-b-D-Galp-(1-4)-b-D-GlcpNAc-(1-6)+
      |
      a-D-Neup5Ac-(2-6)+      b-D-Galp-(1-4)-D-Glc
      |
      b-D-GlcpNAc-(1-3)+
      |
      a-D-Neup5Ac-(2-3)-b-D-Galp-(1-3)+
    
```

Molecular weight: 1801
Chemical formula: C₆₈H₁₁₂N₄O₅₁

No NMR data available

References:

Smith DF; Zopf DA; Ginsburg V
Fractionation of sialyl oligosaccharides of human milk by ion-exchange chromatography
Anal Biochem **85** (1978) 602-608
[Pick this reference in PubMed](#)

Structure (LinusID= 285). To go to the corresponding Glycosciences DB page [click here](#).

a-D-Neup5Ac-(2-3)-b-D-Galp-(1-4)-b-D-GlcpNAc-(1-4)-Asn

Molecular weight: 772
Chemical formula: C₂₉H₄₈N₄O₂₀

¹H NMR data:
Recorded in D₂O at 281.0 K
Spectrometer: 500.0 MHz

Residue	Linkage	Atom	δ, ppm	J, from	J, to	J, Hz
b-D-Galp	4,4	H1	4.57	1	2	7.8
b-D-Galp	4,4	H2	3.58	2	3	9.9
b-D-Galp	4,4	H3	4.13	3	4	3.1
b-D-Galp	4,4	H4	3.96	4	5	1.5
b-D-Galp	4,4	H5	3.72	5	6	3.5
b-D-Galp	4,4	H6	3.73	5	6'	8.5
b-D-Galp	4,4	H6'	3.67	6	6'	-11.5
a-D-Neup5Ac	3,4,4	H3A	1.80	3a	3e	-12.8
a-D-Neup5Ac	3,4,4	H3E	2.77	3a	4	11.6
a-D-Neup5Ac	3,4,4	H4	3.68	3e	4	4.6
a-D-Neup5Ac	3,4,4	H5	3.85	4	5	9.6
a-D-Neup5Ac	3,4,4	H6	3.62	5	6	10.5
a-D-Neup5Ac	3,4,4	H7	3.59	6	7	1.8
a-D-Neup5Ac	3,4,4	H8	3.90	7	8	9.2
a-D-Neup5Ac	3,4,4	H9	3.87	8	9	2.3
a-D-Neup5Ac	3,4,4	H9'	3.63	8	9'	6.7
a-D-Neup5Ac	3,4,4	NAC	2.03	9	9'	-12.3

References:

#0:
Vliegenthart JFG; Dorland L; van Halbeek H
High-resolution, 1H-nuclear magnetic resonance spectroscopy as a tool in the structural analysis of carbohydrates related to glycoproteins
Adv Carbohydr Chem Biochem **41** (1983) 209-374

#1:
Hounsell EF
Physicochemical analyses of oligosaccharide determinants of glycoproteins
Adv Carbohydr Chem Biochem **50** (1994) 311-350
[Pick this reference in PubMed](#)

#2:
Stoll MS; Hounsell EF
Selective purification of reduced oligosaccharides using a phenylboronic acid Bond Elut column: potential application in HPLC, mass spectrometry, reductive amination procedures and antigenic/serum analysis
Biomed Chromatogr **2** (1988) 249-253
[Pick this reference in PubMed](#)

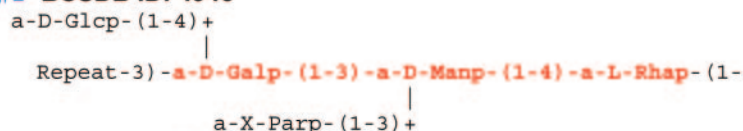
Example 2b. Result querying GLYCOSCIENCES for all structures containing a specified disaccharide fragment α -D-Neup5Ac-(2→3) β -D-Galp. The data associated with two entries are shown.

Input of substructure in *GLYCOSCIENCES.de* using the substructure input matrix (not shown). The substructure Search is performed for *BCSDB* entries.

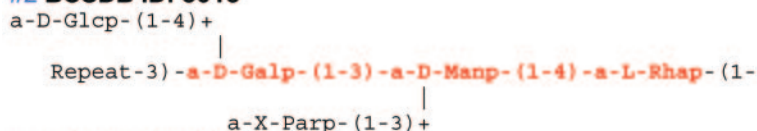
a-D-Galp-(1-3)-a-D-Manp-(1-4)-a-L-Rhap

Results from BCSDB (Search for substructures. Results: 1 - 5 of 5)

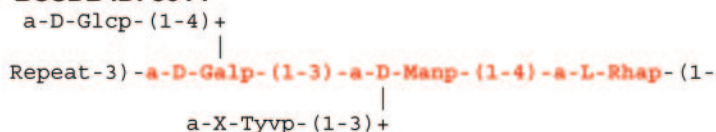
#1 BCSDB ID: 4946



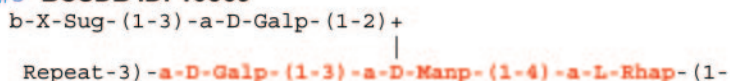
#2 BCSDB ID: 5013



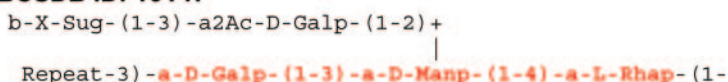
#3 BCSDB ID: 5014



#5 BCSDB ID: 10369

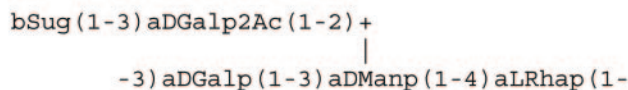


#4 BCSDB ID: 10147



BCSDB entry ID: 10147

Senchenkova SN, Shashkov AS, Knirel YA, Ahmed M, Mavridis A, Rudolph K Structure of the O-polysaccharide of *Erwinia carotovora* ssp. *atroseptica* GSPB 9205, containing a new higher branched monosaccharide



Sug = 3,6,8-trideoxy-4-C-[(R)-1-hydroxyethyl]-D-gulo-octose (erwiniosa)

Erwinia carotovora ssp. *atroseptica* GSPB9205 (NCBI Taxonomy)

Izv. Akad. Nauk Ser. Khim. **0 (5)** (2005) 1-6

Example 3. Querying BCSDB for all structures containing the specified trisaccharide fragment α -D-Galp-(1-3)- α -D-Manp-(1-4)- α -L-Rhap. The *GLYCOSCIENCES.de* substructure input spreadsheet is used. The data associated with BCSDB entry 10147 are additionally shown.

This will hopefully lead to a better worldwide acceptance of both services within the community of glycoscientists.

Since the exchange of data is accomplished through standard, well-documented XML-based descriptions and SOAP protocols; other interested providers of glyco-related databases may easily be linked so that a larger network could grow. It can be envisaged that online connection of thematically related scientific data collections will have a bright

future, and not only in the area of glycosciences. One of the main bottlenecks is currently that broadly accepted standard XML exchange formats are often not yet available. It will definitively be a time-consuming task to come to agreements about such standard descriptions within the various communities. With *GLYDE* 1.2 an XML-based encoding scheme of glycan structures exists, which is sufficiently flexible to link the vast majority of structures contained in *BCSDB*

and *GLYCOSCIENCES.de*. However, GLYDE 1.2 has some shortcomings regarding uncertainties in terminal residues and other fuzzy encodings, which will become more important for glycomics projects. The current focus of discussion is to base a more flexible encoding on the concept of a connection table approach, instead of a tree-like structure as used in GLYDE 1.2. Recently (September 2006, NIH Meeting 'Frontier in Glycomics'), the seven larger projects already mentioned above have agreed to support GLYDE-CT as the main database format for the exchange of glycan structures. The hope is of course that only one format will be used by everyone. A less favourable situation would be that several exchange format exit and parsers must be available for each database.

ACKNOWLEDGEMENTS

P.T. thanks the DKFZ for the fellowship supporting his stay in Heidelberg. The development of *GLYCOSCIENCES.de* at the DKFZ was supported by a Research Grant from the German Research Foundation (DFG BIB 46 HDdkz 01-01) within the digital library program. The development of the BCSDB was supported by the International Science and Technology Center (project 1197p), the Russian Foundation for Basic Research (project 05-07-90099) and Russian President Grant Commetee (project MK-2005.1700.4). Funding to pay the Open Access publication charges for this article was provided by DFG.

Conflict of interest statement. None declared.

REFERENCES

- Lowe, J. and Marth, J. (2003) A genetic approach to Mammalian glycan function. *Annu. Rev. Biochem.*, **72**, 643–691.
- Raman, R., Raguram, S., Venkataraman, G., Paulson, J.C. and Sasisekharan, R. (2005) Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nature Methods*, **1**, 817–824.
- von der Lieth, C.W., Bohne-Lang, A., Lohmann, K.K. and Frank, M. (2004) Bioinformatics for glycomics: status, methods, requirements and perspectives. *Brief Bioinformatics*, **5**, 164–178.
- Harvey, D. (2005) Proteomic analysis of glycosylation: structural determination of N- and O-linked glycans by mass spectrometry. *Expert Rev. Proteomics*, **2**, 87–101.
- Guerardel, Y., Chang, L., Maes, E., Huang, C. and Khoo, K. (2006) Glycomic survey mapping of zebrafish identifies unique sialylation pattern. *Glycobiology*, **16**, 244–257.
- Raman, R., Venkataraman, M., Ramakrishnan, S., Lang, W., Raguram, S. and Sasisekharan, R. (2006) Advancing Glycomics: implementation strategies at the consortium for functional glycomics. *Glycobiology*, **16**, 82R–90R.
- Lutheke, T., Bohne-Lang, A., Loss, A., Goetz, T., Frank, M. and von der Lieth, C.W. (2006) *GLYCOSCIENCES.de*: an Internet portal to support glycomics and glycobiology research. *Glycobiology*, **16**, 71R–81R.
- Hashimoto, K., Goto, S., Kawano, S., Aoki-Kinoshita, K., Ueda, N., Hamajima, M., Kawasaki, T. and Kanehisa, M. (2006) KEGG as a glycome informatics resource. *Glycobiology*, **16**, 63R–70R.
- von der Lieth, C.W. (2004) An endorsement to create open databases for analytical data of complex carbohydrates. *J. Carbohydr. Chem.*, **23**, 277–297.
- Bohne-Lang, A., Lang, E., Forster, T. and von der Lieth, C.W. (2001) LINUCS: linear notation for unique description of carbohydrate sequences. *Carbohydr. Res.*, **336**, 1–11.
- Cooper, C., Joshi, H.J., Harrison, M., Wilkins, M. and Packer, N. (2003) GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res.*, **31**, 511–513.
- Banin, E., Neuberger, Y., Altshuler, Y., Halevi, A., Inbar, O., Dotan, N. and Avinoam, D. (2002) A novel linear code[®] nomenclature for complex carbohydrates. *Trends Glycosci. Glycotechnol.*, **14**, 127–137.
- Aoki, K., Yamaguchi, A., Ueda, N., Akutsu, T., Mamitsuka, H., Goto, S. and Kanehisa, M. (2004) KcAM (KEGG Carbohydrate Matcher): a software tool for analyzing the structures of carbohydrate sugar chains. *Nucleic Acids Res.*, **32**, W267–W272.
- Toukach, F.V., and Knirel, Y.A. (2005) New database of bacterial carbohydrate structures. In *proceedings of the XVIII International Symposium on Glycoconjugates*, Florence, Italy, pp. 216–217.
- Glycomics Consortium for Functional. (2004) Joint meeting of the Japanese and American consortia for glycomics.
- Kikuchi, N., Kameyama, A., Nakaya, S., Ito, H., Sato, T., Shikanai, T., Takahashi, Y. and Narimatsu, H. (2005) The carbohydrate sequence markup language (CabosML): an XML description of carbohydrate structures. *Bioinformatics*, **21**, 1717–1718.
- Sahoo, S., Thomas, C., Sheth, A., Henson, C. and York, W. (2005) GLYDE—an expressive XML standard for the representation of glycan structure. *Carbohydr. Res.*, **340**, 2802–2807.
- Doubet, S., Bock, K., Smith, D., Darvill, A. and Albersheim, P. (1989) The complex carbohydrate structure database. *Trends Biochem. Sci.*, **14**, 475–477.
- van Kuik, J.A., Hard, K. and Vliegthart, J.F. (1992) ¹H NMR database computer program for the analysis of the primary structure of complex carbohydrates. *Carbohydr. Res.*, **235**, 53–68.
- Lutheke, T., Frank, M. and von der Lieth, C.W. (2004) Data mining the protein data bank: automatic detection and assignment of carbohydrate structures. *Carbohydr. Res.*, **339**, 1015–1020.

APPENDIX

Web addresses of tools discussed in the paper are as follows:

Apache Axis: <http://ws.apache.org/axis>
 BCSDB: <http://www.glyco.ac.ru/bcsdb/>
 BCSDB: <http://www.glyco.ac.ru/bcsdb/help/bcsdb.wSDL>
 CCRC (Complex Carbohydrate Research Center). <http://www.ccr.cuga.edu/>
 Codehaus Xfire: <http://xfire.codehaus.org>
 Consortium For Functional Glycomics: <http://www.functionalglycomics.org/>
 EUROCarbDB: <http://www.eurocarbodb.org>
 EUROCarbDB forum: http://www.dkfz.de/spec/EuroCarbDB_forum/
 GLYCOSCIENCES: <http://www.glycosciences.de>
 GLYCOSCIENCES web-services: <http://www.glycosciences.de/soap/soapservice?wsdl>
 Glyde 1.2: <http://www.glyco.ac.ru/bcsdb/help/glyde12.dtd>
 KEGG-Glycam: <http://www.genome.jp/ligand/kcam/>
 HGPI (Human Disease Pglcomics/Proteome Initiative): <http://www.hgpi.jp/>
 Hibernate: <http://www.hibernate.org>
 Monosaccharoide DB: <http://www.dkfz.de/spec/monosaccharide-db/>
 NCBI electronic utilities: http://www.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html
 NCBI PubMed XML tagged data: <http://www.ncbi.nlm.nih.gov/entrez/query/static/publisher.html>
 NCBI Taxonomy: <http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>
 SOAP Lite: <http://www.soaplite.com/>
 Web Service Data Language (WSDL): <http://www.w3.org/TR/wsdl/>