Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 339 (2004) 929–936
Carbohydrate RESEARCH

Review

Web resources for the carbohydrate chemist

Olivier Berteau[a] and Roland Stenutz[b,*]

[a]Department of Chemistry, Swedish University of Agricultural Sciences, PO Box 7015, SE-750 07 Uppsala, Sweden
[b]Arrhenius Laboratory, Department of Organic Chemistry, Stockholm University, SE-106 91 Stockholm, Sweden

Received 13 October 2003; accepted 14 November 2003

Abstract—Bioinformatics has played a pivotal role in advancing genetics and protein sciences. The large amount of information generated by genomics, and now proteomics, has been a driving force. By comparison, glycobiology still generates small amounts of data. The need to organize our knowledge about carbohydrates is however growing constantly and has given rise to an increasing number of public databases and freely available tools. This review gives an overview of the carbohydrate-oriented resources currently available on the Internet. Many of the resources are seldom referred to in the literature and difficult to find, in part because of the constant flux of the net itself, but also because many efforts have been lead by a single individual. As the World Wide Web has matured the number of 'permanent' resources, maintained by organizations rather than individuals, has increased.In this paper, we present some of the more useful and accessible public tools and databases. There are also a few commercial initiatives but these have not been reviewed.

Keywords: Bioinformatics; Database; Glycobiology; Glycomics; Glycosylation

Contents

* Corresponding author. Tel.: +46-08-674-7263; fax: +46-08-154-908; e-mail: roland.stenutz@organ.su.se

0008-6215/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2003.11.008

1. Introduction

The recent progress in the field of genomics has been very impressing. Automated methods for the isolation, amplification and sequence determination of nucleic acids have generated vast amounts of information. But genes alone are not sufficient to describe an organism. For example, human cells have the same genome but express different genes in different organs. Information about which genes are expressed is also required. Hence proteomics came to fill this void. Proteomics tries to describe an organism by the level of protein expression, that is, the active genes. However, the actual activity of proteins is regulated not only on the level of expression but also by various post-translational modifications such as phosphorylation and glycosylation. Whilst much of the concepts of genomics can be transferred to proteomics, it is difficult to apply similar principles to carbohydrates since they are secondary gene products and their structure cannot be easily predicted from the DNA sequence.

The structure of carbohydrates is much more complex than that of nucleic acids and proteins. The latter consist of a reasonably small number of monomers linked in a predetermined sequential fashion. Carbohydrates on the other hand are often branched, can be linked in a number of different ways and the number of possible monomers, not counting the numerous modifications, approaches 100. These residues can be connected in a linear or branched fashion using not only classical glycosidic linkages (acetals), but also by phosphodiesters, either directly or through glycerol or an alditol (teichoic acid-type polymers). Many carbohydrates also show varying degrees of heterogeneity. Despite being macromolecules, glycans require methods more similar to those for small molecules than for proteins and nucleic acids. This means that very little of the methods used for genomics and proteomics research can be adapted for carbohydrates.

Despite these difficulties there is a steadily increasing number of useful Internet resources dedicated to carbohydrates. At present their audience is rather limited but it will no doubt increase as biologist and biochemists turn their attention toward protein glycosylation. This poses new challenges to the carbohydrate community as it will have to communicate its findings to an audience that is accustomed to large, easily searched online data collections.

Two recent reviews of similar scope have recently been published. A paper by Marchal et al.[1] covers the topic from a glycobiologists point of view and a database collection by Baxevanis gives a summary of popular microbiology sites.[2,3]

For clarity, all the web sites referred to in this review are listed in Tables 1–5 and cited with bold numbers in the text.

2. Primary structure

2.1. Nomenclature

The first difficulty one encounters on entering the field of carbohydrate chemistry is the nomenclature. Whereas the nomenclature of proteins and nucleic acids is reasonably easy and consistent, that of carbohydrates is difficult to master even for an expert. As a result, there is no universally accepted description of oligo- and polysaccharides that can serve as a foundation for databases. Some progress was made in the 1996 IUPAC-IUBMB 'Nomenclature of Carbohydrates'[4,5] (**1**), which specified an extended form, a condensed form and a short form. However none of these forms was particularly suited for computer processing because of the use of special symbols and some remaining ambiguities regarding the ordering of branches. This has lead to the development of other structural representations (Fig. 1).

Table 1. Web resources related to the primary structure of carbohydrates

	Name	URL	Retrievable by	Reference
1	IUPAC-IUBMB 'Nomenclature of Carbohydrates'	www.chem.qmw.ac.uk/iupac/		4,5
2	LINUCS	www.dkfz-heidelberg.de/spec/linucs/	Manual input of the structure	6
3	Complex Carbohydrate Structure Database 'CarbBank'	www.boc.chem.uu.nl/sugabase/databases.html	Graphic interface, manual input of the structure, keyword, chemical shifts, references	8
4	SugaBase	www.boc.chem.uu.nl/sugabase/databases.html	Graphical interface, manual input of the structure, keyword, chemical shifts, references	9
5	SweetDB	www.dkfz-heidelberg.de/spec2/sweetdb/	Graphic interface, manual input of the structure, composition, formula, chemical shifts, MS fragments, references	43
6	GlycoBase	ustl.univ-lille1.fr/glycobase/	Manual input of the structure, species	

Table 2. Tools for the analysis of NMR and MS data

	Name	URL	Retrievable by	Reference
7	The CCRC database for xyloglucan oligosaccharide NMR data	www.ccrc.uga.edu/web/specdb/specdbframe.html	Chemical shift, sugar residue, linkage	
8	ECDB—Structure and NMR data of *E. coli* LPS	www.casper.organ.su.se/ECDB/	Sugar residue, chemical shift, manually	
9	GC-EIMS of partially methylated Alditol acetates	www.ccrc.uga.edu/web/specdb/ms/pmaa/pframe.html	Graphic interface	
10	SPECARB—Raman spectra of saccharides	www.models.kvl.dk/users/engelsen/specarb/specarb.html	Manually	
11	CASPER—Simulation of NMR data	www.casper.organ.su.se/casper/	Graphic interface, chemical shifts (Manual/Upload)	11,12
12	GlycoFragments—Simulation of mass spectra	www.dkfz-heidelberg.de/spec/projekte/fragments/	Manual input of the structure	

Table 3. Conformation of carbohydrates and glycoproteins

	Name	URL	Retrievable by	Reference
13	Protein Data Bank	www.rcsb.org/pdb/	PDB ID, keywords	13
14	Pdb2linucs—Annotated PDB	www.dkfz-heidelberg.de/spec/pdb2linucs/	Manually	6
15	HIC-Up—Small molecules from PDB	alpha2.bmc.uu.se/hicup/	PDB ID, keywords	18
16	Image library of biological macromolecules—Small molecules from PDB	www.imb-jena.de/ImgLibPDB/pages/carb.html	Manually	19,20
17	Het-PDB Navi—Small molecules from PDB	daisy.bio.nagoya-u.ac.jp/golab/hetpdbnavi.html	PDB Code, hetero-atom code	
18	3D Monosaccharide database	www.cermav.cnrs.fr/cgi-bin/monos/monos.cgi	Graphic interface	
19	3D Disaccharide database	www.cermav.cnrs.fr/cgi-bin/di/di.cgi	Graphic interface	23
20	GlycoMaps database—Conformational maps of disaccharides	www.dkfz-heidelberg.de/spec/glycomaps/	Graphic interface	
21	SWEET-II—Performs MM calculations	www.dkfz-heidelberg.de/spec/sweet2/	Graphic interface, manual input of the structure	24
22	Glydict—Performs MM calculations	www.dkfz-heidelberg.de/spec/glydict/	Manual input of the structure	25

Table 4. Databases of enzymes and lectins

	Name	URL	Retrievable by	Reference
23	Carbohydrate–Active enzymes	afmb.cnrs-mrs.fr/CAZY/	EC number, family number, clan, species	26–28
24	Bacterial polysaccharide gene database	www.microbio.usyd.edu.au/BPGD/default.htm	EC number, references, gene name, NCBI ID etc.	
25	3D Lectin database	www.cermav.cnrs.fr/lectines/	Graphic interface	
26	A genomics resource for animal lectins	ctld.glycob.ox.ac.uk/	Manually	30,31
27	Thorkild's lectin page	plab.ku.dk/tcbh/lectin-links.htm	Manually	

The German Cancer Research Centre (DKFZ) developed LINUCS[6] (**2**) and uses it for most of its WWW applications. The Consortium for Functional Glycomics (funded by the National Institute of General Medical Sciences, USA) recently decided to use a commercial representation called LinearCode®.[7] Both nomenclatures present a unique description of complex structures in a 'computer-friendly' manner but both have yet to find wider acceptance.

Whilst the problem of storing carbohydrate structures in a computer readable format has been solved the complexity of the glucan structures makes if difficult to design a general, yet intuitive, user interface. An advanced graphical interface was designed for the CCSD (**3**) and SugaBase (**4**, below) but many applications rely on simple pull down menus with limited options. No doubt the shortcomings of the user interfaces is one of the reasons many carbohydrate databases have few users.

Table 5. Databases and prediction tools related to glycoproteins

	Name	URL	Retrievable by	Reference
28	O-GlycBase—Database of glycoproteins	www.cbs.dtu.dk/databases/OGLYCBASE/	Manually	33, 34
29	FindMod—Interpretation of MS	www.expasy.org/tools/findmod/	Manual input, upload, structure, composition	36
30	GlycoMod tool—Interpretation of MS	www.expasy.org/tools/glycomod/	Manual input, structure, composition	37
31	Glypeps—Interpretation of MS	www.dkfz-heidelberg.de/spec/glypeps/	Manual input	38
32	NetOGlyc—Prediction of glycosylation sites	www.cbs.dtu.dk/services/NetOGlyc/	Manual input	39, 40
33	DictyOGlyc—Prediction of glycosylation sites	www.cbs.dtu.dk/services/DictyOGlyc/	Manual input	41
34	NetNGlyc—Prediction of glycosylation sites	www.cbs.dtu.dk/services/NetNGlyc/	Manual input	
35	YinOYang—Prediction of glycosylation sites	www.cbs.dtu.dk/services/YinOYang/	Manual input	
36	DGPI—Prediction of GPI-anchor sites	129.194.185.165/dgpi/index_en.html	Manual input, entry name or access number	
37	Big-PI Predictor—Prediction of GPI-anchor sites	mendel.imp.univie.ac.at/sat/gpi/gpi_server.html	Manual input	42
38	CarbDB	web.mit.edu/glycomics/carb/carbdb.shtml	Not yet available	

IUPAC-IUBMB

Extended form:

α-L-Fuc*p*-(1→4)-β-D-Glc*p*NAc
3
↑
1
β-D-Gal*p*

Condensed form:

Fuc(α1-4)GlcNAc(β1-
|
Gal(β1-3)

Short form:

Fucα4(Galβ3)GlcNAcβ

LINUCS:

[][b-D-Galp]{[(1-3)][b-D-GlcpNAc]{[(4+1)][a-L-Fucp]{}}}

LinearCode™

Ab3(Fa4)GNb

CASPER:

bDGal(1->3)[aLFuc(1->4)]bDGlcNAc

CCSD (CarbBank):

```
a-L-Fucp-(1-4)+
              |
         b-D-GlcpNAc
              |
b-D-Galp-(1-3)+
```

Figure 1. Various structural representations of a simple oligosaccharide (Lewis A, Type 1).

2.2. Structure databases

Structure databases were the first web tools developed for carbohydrates. Indeed, considering the time required to solve the primary structure of an oligo- or polysaccharide, it is of great importance to store the known structures and their analytical data.

The first and by far most ambitious effort in this area was the creation of the Complex Carbohydrate Structure Database (CCSD)[8] better known under the name of the accompanying search engine: CarbBank (**3**). It was an international effort headed by the Complex Carbohydrate Research Center (CCRC, University of Georgia, Athens, GA, USA). In the beginning, the database

was distributed by ftp or on CD but later WWW interfaces were developed. The last release of the CCSD contains ≈50,000 records with data about the primary structure, source, the analytical methods used and literature references. Funding ceased in 1997 but fortunately the database is still available from the Bijvoet Center (University of Utrecht, The Netherlands) and as part of SweetDB (**5**).

A specialized database, GlycoBase, with ≈300 carbohydrate structures from amphibian oviducal mucines and corresponding literature references is available from University of Lille (France) (**6**).

2.3. Analytical data

Whilst databases of known structures are useful, they cannot be used directly to find a compound with unknown structure. No doubt much unnecessary work could be avoided if NMR and MS data for all known structures were stored in a public database. For example, ≈20% of the known *Escherichia coli* O-antigens are identical to structures from other bacteria.

At the Bijvoet Center, a database called SugaBase[9] was created by adding ^{1}H NMR data to 1600 CCSD entries of oligosaccharides. The scope of SugaBase has since been extended to cover ^{13}C NMR data and a wider range of carbohydrates (**4**). Unfortunately SugaBase has met the same fate as the CCSD and is no longer being updated. The interfaces are still available and their content has been included in the recently developed SweetDB (see below) (**5**).

Two other specialized databases containing NMR data are also available. A collection of ^{1}H NMR spectra of fragments from xyloglucans is hosted by the CCRC (**7**) and Stockholm University hosts a database of *E. coli* O-antigens (ECDB) (**8**). The latter database provides the structures of the known O-antigens of *E. coli* with bibliographic data, and NMR chemical shifts, both ^{1}H and ^{13}C.

The only experimental mass spectrometry database available for carbohydrates is a collection of partially methylated alditol acetate maintained by the CCRC (**9**). There is also a small database of ≈25 Raman spectra of various oligo- and polysaccharides (**10**).

2.4. Prediction of analytical data

The structural complexity and diversity of carbohydrates makes the structure elucidation a challenging task despite the progress in spectroscopy. The limited dispersion of the chemical shifts in carbohydrates requires high accuracy in the prediction of NMR data, about 0.2 ppm in ^{13}C—to be compared with 1.3 ppm in a recent tool for the prediction of chemical shifts in proteins.[10] Fortunately, some of the factors that contribute to the poor dispersion of chemical shifts, for example, the absence of a well-defined secondary structure, make it possible to achieve high accuracy using additivity schemes. Computer Aided SPectrum Evaluation of Regular polysaccharides (CASPER) is a WWW tool devoted to the assignment of ^{1}H and ^{13}C NMR spectra (**11**).[11,12] It allows both the simulation of the NMR spectra of known structures and the sequence determination of unknown structures using unassigned NMR spectra and information from chemical analyses (sugar-, configuration- and methylation analysis).

Mass spectrometry is the preferred method in the structure determination of oligosaccharides derived from glycoproteins. The limited number of possible glycosyl residues with identical mass, prior knowledge about the possible structures and the high sensitivity of the method makes it possible to determine the structure of glycopeptides. Also, the small amounts of protein available often precludes the use of NMR spectrometry.

GlycoFragments, hosted by DKFZ (**12**), is a tool for the interpretation of the mass spectra of oligosaccharides and glycopeptides. Prediction of fragmentation patterns is also provided as part of SweetDB.

3. Conformation

3.1. Experimental structures

The more we learn about the diverse biological functions of carbohydrates, the clearer it becomes that it is necessary to go beyond the primary structure and investigate their conformation.

Whilst determining the primary structure of carbohydrates is very difficult, much more so than determining the sequence of nucleic acids or proteins, the determination of the conformation is still more complicated. For this reason experimental is scarce. Although a common way to determine the conformation of a saccharide is by NMR spectroscopy in water, there is no database dedicated to the results from NMR studies. Almost all the available experimental data in databases comes from X-ray diffraction.

In the crystallographic community, there is a long tradition of depositing experimental data in centralized databases. Macromolecular data, for example, structures of proteins and oligonucleotides, is deposited in the Protein Data Bank[13,14] (PDB). The PDB contains more than 700 glycoproteins (10%) and about 200 carbohydrates, most of which interact with proteins (**13**). The vast majority of data comes from crystallography but a few NMR-derived structures, for example, of heparin,[15] can also be found. Conformational data for carbohydrates has been extracted from the PDB[16], but this not simple since the sugar residues often are named incorrectly or have distorted structures. Also, the PDB cannot be searched by carbohydrate structure. A recent

effort, pdb2linucs,[17] tries to eliminate this problem by examining every PDB record for carbohydrate residues and encoding them using LINUCS nomenclature (**14**).

A number of small databases summarize the 'non-protein' data (HETATM) of the PDB in a way that facilitates the identification of small molecules co-crystallized with proteins (**15**–**17**).[18–20]

Small- and medium-sized organic molecules can be found in the Cambridge Structural Database[21] (CSD), which contains ≈4000 entries classified as carbohydrates (≈1.5%). A large number of these are cyclodextrin inclusion complexes and synthetic intermediates and thus of limited biological interest. The CSD is not freely available and does not have a web interface, it is however widely available, at least in the crystallographic community, and certainly the best source of high quality solid-state structures of carbohydrates. A database of papers in which the CSD has been used as the main tool is freely available. Of the ≈1000 entries, 2.5% relate to carbohydrates.[22]

3.2. Computed structures

Because of the scarcity of experimental data on carbohydrates, computational methods have always played a central role in the conformational analysis of carbohydrates, in particular in conjunction with NOE (distance) information from NMR. The results are often presented as the structure of the lowest energy conformation or as a potential energy surface with the glycosidic torsion angles, φ and ψ, as axes.

The calculated structures of about 20 monosaccharides in 90 different conformations and adiabatic (φ,ψ) potential energy surfaces (**18**), and the structures and energies for conformers of ≈20 disaccharides[23] (**19**) can be found on the web pages of Centre de Recherches sur les Macromolécules Végétales (CERMAV, Grenoble, France). A similar, but more extensive database, GlycoMaps, is maintained by DKFZ and contains ≈700 φ,ψ-maps obtained from MD simulations of disaccharides using the MM3 force field (**20**). Larger structures can be analyzed online using SWEET-II[24] (**21**) or off-line using Glydict[25] (**22**) (for N-linked glycans). Because of the large number of possible conformations, an exhaustive analysis is not supported but the structures are easily built and the conformations obtained adequate as a starting points for further analysis. Both tools accept CarbBank-style structures as input.

4. Enzymes

Enzymes having carbohydrates as substrates are collected in a database called Carbohydrate Active Enzymes (CAZy)[26–28] (**23**) at the French Center for National Scientific Research (CNRS). It contains information about glycosidases, glycosyltransferases, lyases and esterases, which can be browsed by organism or CAZy family. The CAZy families are derived from the amino acid sequences and are therefore expected to better reflect the secondary structure and mechanism of the enzyme than the EC numbers, which are based on substrate specificity. This classification is unique to CAZy and has provided new insights into the mechanisms and evolutionary relationships of these enzymes. Links are provided to a number of gene and protein databases.

A more specialized database, the Bacterial Polysaccharide Gene Database,[29] provides data about the genes involved in biosynthesis of bacterial polysaccharides (**24**). Apart from the gene itself, there is information about the associated protein such as the EC number, the reaction catalyzed and its role in the biosynthetic pathway.

5. Lectins

Lectins are carbohydrate binding proteins lacking catalytic activity and which are not antibodies. They are important in a number of biological events involving carbohydrate recognition and we have only recently begun to comprehend their function.

Bibliographic information (Medline) and the three-dimensional structures of ≈200 lectins extracted from the PDB can be browsed on a server at CERMAV (**25**). Oxford University provides 'A genomics resource for Animal Lectins', which contains information not only about animal lectins[30] but also about other proteins having a 'C-type lectin-like domains' (**26**).[31] A comprehensive list of resources related to lectins (Torkild's Lectin Page) is maintained by the Center for Biological Sequence Analysis, Technical University of Denmark (**27**).

6. Glycoproteins

6.1. Experimental data

The area of glycobiology that is developing most rapidly is the study of protein glycosylation. Studies of post-translational modifications, such as glycosylation, is a natural progression of a development that started with the determination of the human genome and has continued with proteomics. The importance of protein glycosylation can be appreciated when one considers that more than half of the known proteins are estimated to be glycoproteins.[32] Despite their importance, there is only one database dedicated to glycoproteins, OGlycBase, (**28**), which contains ≈250 proteins with experimentally verified O- or C-glycosylation site.[33,34] This

database provides the amino acid sequence of the protein and indicates their glycosylation site. More information about glycoproteins can be found in protein databases such as SWISS-PROT.[35]

Several tools for the determination of glycosylation sites from mass spectra are available. FindMod,[36] (**29**) GlycoMod[37] (**30**) and Glypeps[38] (**31**) compare experimentally measured masses to calculated peptide masses and thereby determine glycosylation sites.

6.2. Prediction of glycosylation- and GPI-anchor sites

Only in a small fraction of the expected glycoproteins has the presence of an attached glycan been experimentally verified. Furthermore, it may be desirable to predict glycosylation directly from the amino acid sequence of a protein since this is easy to obtain, for example, from the DNA sequence.

Artificial neural networks can be used to predict glycosylation sites. The Technical University of Denmark hosts NetOGlyc (**32**, for mucin type O-glycosylation),[39,40] DictyOGlyc (**33**, for O-glycosylation sites in the amoeba *Dictyostelium discoideum*),[41] NetNGlyc (**34**, for N-glycosylation in human proteins) and YinOYang (**35**, for *O*-β-D-GlcNAc/phosphorylation sites). The interfaces of the tools are very similar and allow an amino acid sequence to be entered directly or to be uploaded from a file.

Some membrane proteins are anchored on the extracellular side of the plasma membrane by a glycosylated phosphatidylinositol, or GPI anchor. There are two tools that predict if a particular protein is likely to be attached to a GPI anchor. DGPI (**36**, University of Geneva) and big-PI (**37**, Research Institute of Molecular Pathology, Vienna).[42] Both tools accept the sequence of a protein and DGPI also allows the use of a SWISS-PROT access number.

All of these tools use protein sequences. To search by carbohydrate structure one has to use CarbBank, SweetDB or one of the commercial alternatives.

7. Outlook

There is a variety of carbohydrate related resources on the Internet. They are however often difficult to find and seldom cited in the literature. Nearly half of the databases in this review do not have a scientific publication that can be cited. Presumably this makes scientists less inclined to use the tools and, if they do, less likely to refer to them in their work. It is also possible that the tools are simply not found.

A universally accepted structural representation of carbohydrates is still lacking and therefore moving between different database and tools is difficult. SweetDB (**5**),[43] the most comprehensive carbohydrate resource available at present, overcomes this problem and integrates CarbBank and SugaBase with several computational tools. The Consortium for Functional Glycomics is also preparing a database, CarbDB (**38**), linking several resources but it is not yet available.

At present there are few links between carbohydrate and protein databases and the intended audience of most databases is clearly restricted to either carbohydrate chemists or the proteomics community. Since much of the recent interest in carbohydrates originates in the proteomics community it is important to bridge the gap between proteomics and glycobiology.

As research has come to rely more and more on databases and computational tools, there is a growing need for large, multidisciplinary carbohydrate databases. But while it is necessary for scientists to participate in defining the content and the use of databases, it is not obvious that they are most suited to developing and maintaining them. Using and citing existing resources and contributing feedback are simple measures that can be taken to improve on the present situation. One can only wonder if CarbBank and SugaBase would still be receiving funding if they had been cited more frequently.

References

1. Marchal, I.; Golfier, G.; Dugas, O.; Majed, M. *Biochimie* **2003**, *85*, 75–81.
2. Baxevanis, A. D. *Nucleic Acids Res.* **2000**, *28*, 1–7.
3. Baxevanis, A. D. *Nucleic Acids Res.* **2003**, *31*, 1–12.
4. *Pure Appl. Chem.* **1996**, *68*, 1919–2008.
5. McNaught, A. D. *Carbohydr. Res.* **1997**, *297*, 1–90.
6. Bohne-Lang, A.; Lang, E.; Forster, T.; von der Lieth, C. W. *Carbohydr. Res.* **2001**, *336*, 1–11.
7. Banin, E.; Neuberger, Y.; Altshuler, Y.; Halevi, A.; Inbar, O.; Dotan, N.; Dukler, A. *Trends Glycosci. Glycotech.* **2002**, *14*, 127–137.
8. Doubet, S.; Bock, K.; Smith, D.; Darvill, A.; Albersheim, P. *Trends Biochem. Sci.* **1989**, *14*, 475–477.
9. van Kuik, J. A.; Hård, K.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1992**, *235*, 53–68.
10. Meiler, J. *J. Biomol. NMR* **2003**, *26*, 25–37.
11. Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1987**, 168.
12. Stenutz, R.; Jansson, P.-E.; Widmalm, G. *Carbohydr. Res.* **1998**, *306*, 11–17.
13. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissing, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–244.
14. Sussman, J. L.; Lin, D.; Jiang, J.; Manning, N. O.; Prilusky, J.; Ritter, O.; Abola, E. E. *Acta Cryst. D* **1998**, *54*, 1078–1084.
15. Mulloy, B.; Forster, M. J.; Jones, C.; Davies, D. B. *Biochem. J.* **1993**, *293*, 849–850.
16. Petrescu, A. J.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R. *Glycobiology* **1999**, *9*, 343–352.
17. Lütteke, T.; Frank, M.; von der Lieth, C.-W. *Carbohydr. Res.*, doi:10.1016/j.carres.2003.09.038.
18. Kleywegt, G. J.; Jones, T. A. *Acta Cryst. D* **1998**, *54*, 1119–1131.
19. Sühnel, J. *Comput. Appl. Biosci.* **1996**, *12*, 227–229.

20. Reichert, J.; Sühnel, J. *Nucleic Acids Res.* **2002**, *30*, 253–254.
21. Allen, F. H.; Kennard, O. *Chem. Des. Autom. News* **1993**, *8*, 31–37.
22. Allen, F. H.; Motherwell, W. D. S. *Acta Cryst. B* **2002**, *58*, 407–422.
23. Mazeau, K. P. S. *Carbohydr. Res.* **1998**, *311*, 203–217.
24. Bohne, A.; Lang, E.; von der Lieth, C. W. *J. Mol. Model.* **1998**, *4*, 33–43.
25. Frank, M.; Bohne-Lang, A.; Wetter, T.; von der Lieth, C. W. *In Silico Biol.* **2002**, *2*, 427–439.
26. Coutinho, P. M.; Henrissat, B. Carbohydrate-Active Enzymes: An Integrated Database Approach. In *Recent Advances in Carbohydrate Bioengineering*; Gilbert, H. J., Davies, G., Henrissat, B., Svensson, B., Eds.; The Royal Society of Chemistry: Cambridge, 1999; pp 3–12.
27. Coutinho, P. M.; Henrissat, B. The Modular Structure of Cellulases and other Carbohydrate-Active Enzymes: An Integrated Database Approach. In *Genetics, Biochemistry and Ecology of Cellulose Degradation*; Ohimiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karite, S., Kirumura, T., Eds.; Uni Publishers: Tokyo, 1999; pp 15–23.
28. Davies, G. J.; Henrissat, B. *Biochem. Soc. Trans.* **2002**, *30*, 291–297.
29. Reeves, P. R.; Hobbs, M.; Valvano, M. A.; Skurnik, M.; Whitfield, C.; Coplin, D.; Kido, N.; Klena, J.; Maskell, D.; Raetz, C. R. H.; Rick, P. D. *Trends Microbiol.* **1996**, *4*, 495–503.
30. Drickamer, K.; Taylor, M. E. *Trends Biochem. Sci.* **1998**, *23*, 321–324.
31. Dodd, R. B.; Drickamer, K. *Glycobiology* **2001**, *11*, 71R–79R.
32. Apweiler, R.; Hermajkob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4–8.
33. Gupta, R.; Birch, H.; Rapacki, K.; Brunak, S.; Hansen, J. E. *Nucleic Acids Res.* **1999**, *27*, 370–372.
34. Hansen, J. E.; Lund, O.; Nielsen, J. O.; Hansen, J. E.; Brunak, S. *Nucleic Acids Res.* **1996**, *24*, 248–252.
35. Bairoch;, A.; Apweiler, R. *Nucleic Acids Res.* **1999**, *27*, 49–54.
36. Wilkins, M. R.; Gasteiger, E.; Gooley, A. A.; Herbert, B. R.; Molloy, M. P.; Binz, P. A.; Ou, K.; Sanchez, J. C.; Bairoch, A.; Williams, K. L.; Hochstrasser, D. F. *J. Mol. Biol.* **1999**, *289*, 645–657.
37. Cooper, C. A.; Gasteiger, E.; Packer, N. H. *Proteomics* **2001**, *1*, 340–349.
38. Lehmann, W. D.; Bohne, A.; von Der Lieth, C. W. *J. Mass. Spectrom.* **2000**, *11*, 1335–1341.
39. Hansen, J. E.; Lund, O.; Tolstrup, N.; Gooley, A. A.; Williams, K. L.; Brunak, S. *Glycoconj. J.* **1998**, *15*, 115–130.
40. Hansen, J. E.; Lund, O.; Engelbrecht, J.; Bohr, H.; Nielsen, J. O.; Hansen, J. E. S.; Brunak, S. *Biochem. J.* **1995**, *308*, 801–813.
41. Gupta, R.; Jung, E.; Gooley, A. A.; Williams, K. L.; Brunak, S.; Hansen, J. *Glycobiology* **1999**, *9*, 1009–1022.
42. Eisenhaber, B.; Bork, P.; Eisenhaber, F. *J. Mol. Biol.* **1999**, *292*, 741–758.
43. Loss, A.; Bunsmann, P.; Bohne, A.; Loss, A.; Schwarzer, E.; Lang, E.; von der Lieth, C. W. *Nucleic Acids Res.* **2002**, *30*, 405–408.