Web resources for the Carbohydrate Chemist

Roland Stenutz
Stockholm University
roland@organ.su.se

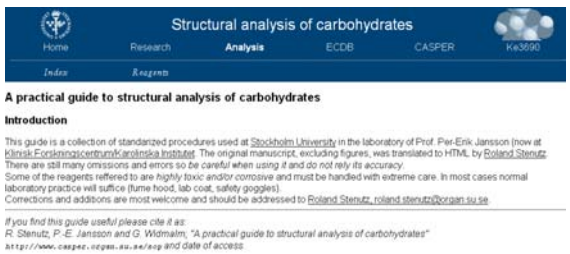
Content

- **Technical considerations**
- Resources by topic

Resources with static content

GlycoWord – Encyclopedia
www.glycoforum.gr.jp/science/word/wordE.html

IUPAC-IUBMB Nomenclature
http://www.chem.qmw.ac.uk/iupac/



Structural analysis – Laboratory manual
www.casper.organ.su.se/sop/

Resources with static content

Content is indexed by search engines, e.g. Google (www.google.com) and Altavista (www.altavista.com).

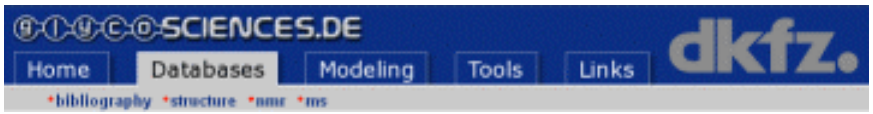
Full text searches are best.

Searches for compounds can be very difficult!

Some tricks have to be used to get useful results.

Compare nr of hits searching for "glucose" (3×10^6), "glucopyranose" (8×10^3) or "50-99-7" (2×10^3).

Databases



glycoSCIENCES.DE
www.glycosciences.de

PDB – Crystal structures
www.rcsb.org/pdb/

CAZy – Enzymes
afmb.cnrs-mrs.fr/CAZY/acc.html

Databases

Large collection (100-100000) of related data.

Searches can be complex, e.g. for (sub)structures.

There are implied by the context; i.e. 5.15 might be a NMR chemical shift but not a price.

There is very little garbage and redundancy in the databases.

Can be difficult to find using search engines since they have little text that can be indexed.

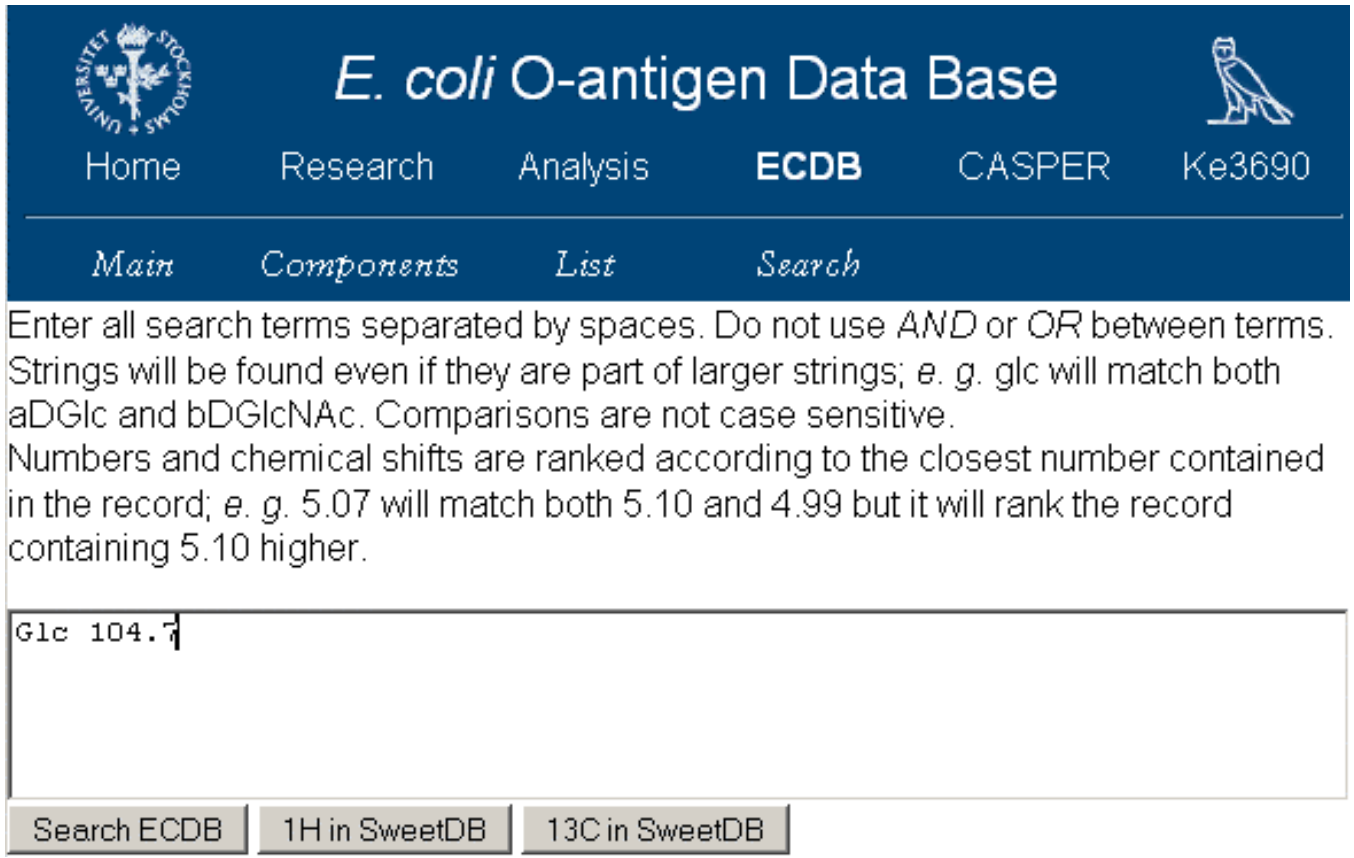
Applications

Most applications can be thought of as databases with an unlimited number of records.

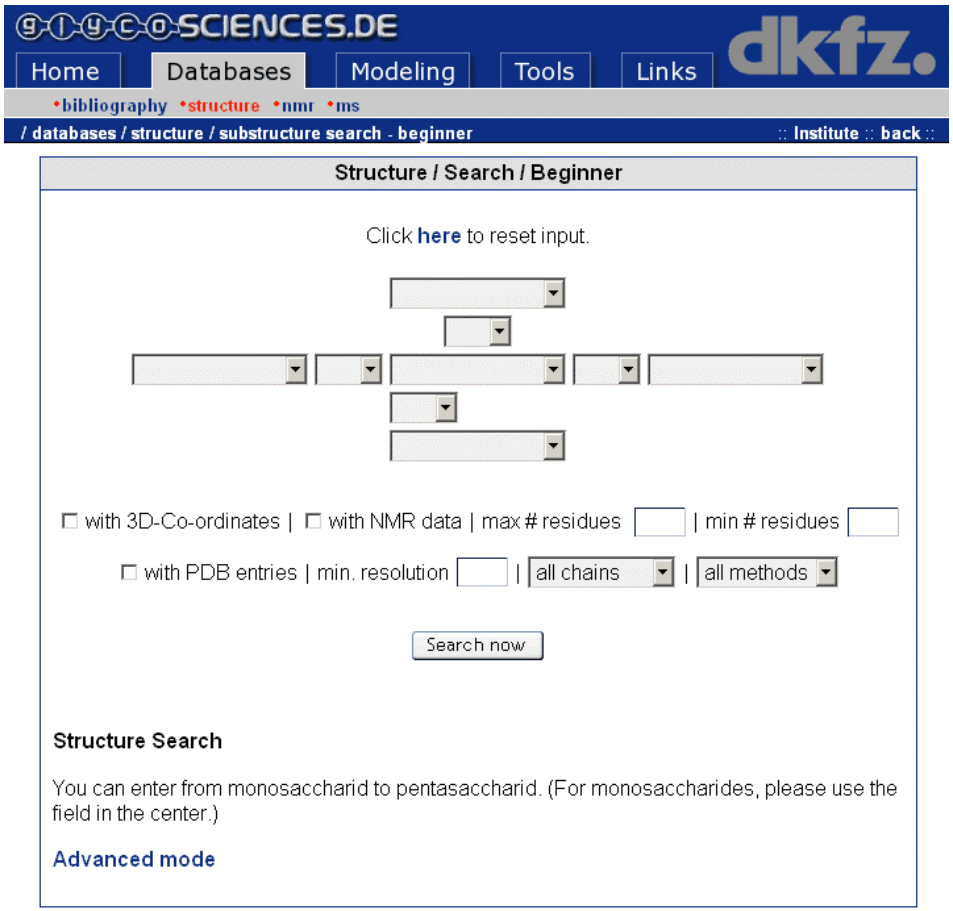
They require relatively complex interfaces since a request for data must contain all the information necessary to generate the data.

They create content “on–the–fly”.

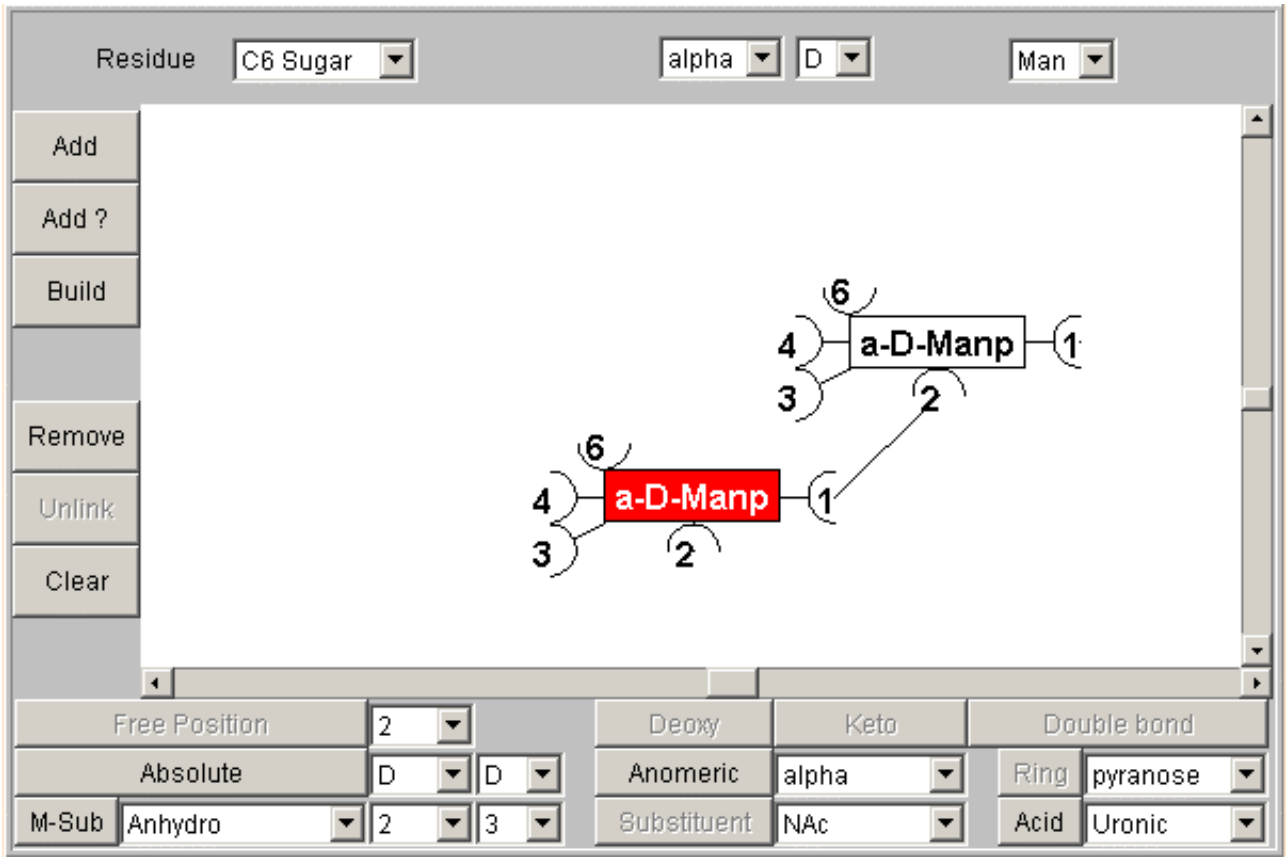
Interfaces - trivial



Interfaces


SCIENCES.DE
dkfz.
Home
Databases
Modeling
Tools
Links
•bibliography •structure •nmr •ms
/ databases / structure / substructure search - beginner
:: Institute :: back ::
Structure / Search / Beginner
Click here to reset input.
with 3D-Co-ordinates | with NMR data | max # residues | min # residues
with PDB entries | min. resolution | all chains | all methods
Search now
Structure Search
You can enter from monosaccharid to pentasaccharid. (For monosaccharides, please use the field in the center.)
Advanced mode

Interfaces - complex



Structured answers – easy to transfer

***** Hit 2 *****
CC: CCSD:3436
AU: Adeyeye A; Jansson PE; Lindberg B; Abaas S; Svenson SB
TI: Structural studies of the Escherichia coli O-149 O-antigen polysaccharide
CT: Carbohydr Res (1988) 176: 231-236
FC: 014fe513
AM: 1H-NMR
BS: (GS) Escherichia coli, (GT) O149
SB: Jansson PE DA: 01-08-1990
MT: LPS
AN: O-antigen
SI: CBank:6914
---
structure: Repeat-4)-b-D-GlcpNAc-(1-3)-b-D-GlcpNAc4,6Py-(1-3)-b-L-Rhap-(1-
==========end of record

Content

- Technical considerations
- **Resources by topic**

Structure

→4)-α-D-Rha-(1→4)-α-L-GalNAcA-(1→3)-β-D-BacNAc-(1→

Complex Carbohydrate Structure Database (CarbBank)
www.boc.chem.uu.nl/sugabase/databases.html
also → glycosciences.de

ECDB – *E. coli* O-antigen structures and NMR
www.casper.organ.su.se/ECDB/

GlycoBase of USTL – oligosaccharides from amphibians
ustl.univ-lille1.fr/glycobase/

→4)-α-D-Rha-(1→4)-α-L-GalNAcA-(1→3)-β-D-BacNAc-(1→

Conformation

PDB – Protein Data Bank, "Brookhaven DB"
Protein structures, incl. glycoproteins

www.rcsb.org/pdb

GlycoMaps Database, SWEET-II etc…
Conformational databases and applications for oligosaccharides

www.glycosciences.de

Disaccharide Database
Conformational maps for some disaccharides

www.cermav.cnrs.fr/databank/disaccharides

Spectroscopy

SugaBase – NMR database, mainly ^{1}H, often incomplete.
www.boc.chem.uu.nl/sugabase/databases.html

CASPER – NMR from structure & structure from NMR.
www.casper.organ.su.se/casper

GlycoFragments – MS fragmentation from structure.
www.glycosciences.de

Enzymes, Lectins and Glycoproteins

CAZy – Carbohydrate Active Enzymes

afmb.cnrs-mrs.fr/CAZY/

3D Lectin Database

www.cermav.cnrs.fr/lectines/

BPGD - Bacterial Polysaccharide Gene Database

www.microbio.usyd.edu.au/BPGD/default.htm

Using Internet resources

The question must be chosen with care!

Ask the same question in different ways.

Ask different search engines/data bases the same question and compare the results.

Always – verify the results!

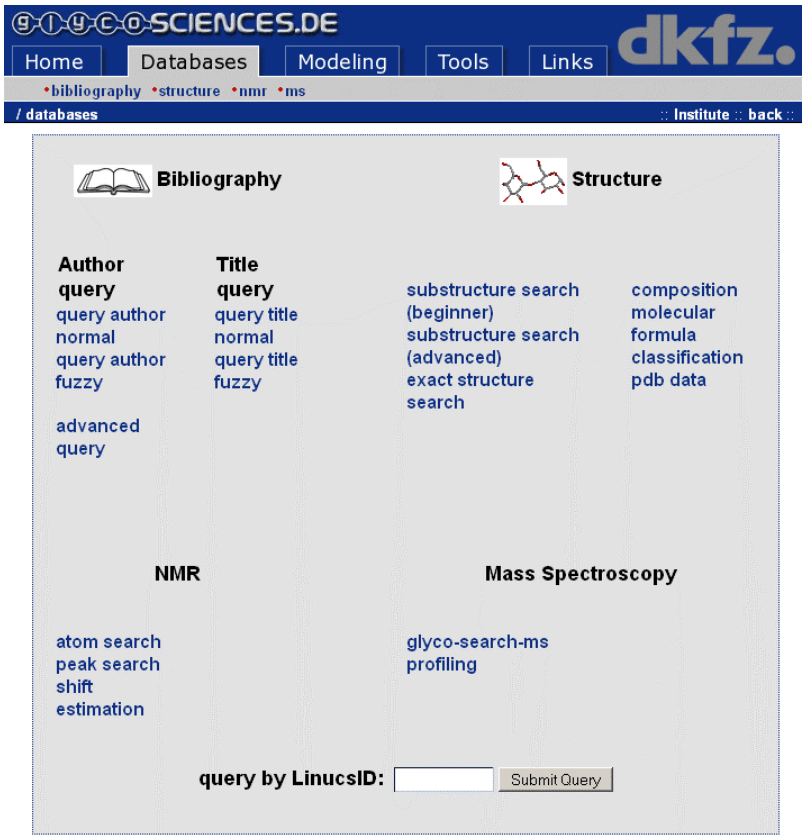
Searching

Even if you know exactly what information you want it can be very difficult to find it.

Information is spread-out in different locations and the question may need paraphrasing.

It is very difficult to get a complete answer – but often you get a hint about how to proceed.
e.g. you might not find the data sought but a reference to a paper that contains the data…

Portals
One interface – several DB:s



glycoSCIENCES.DE

Searchable by
structure/substructure
bibliographic information
NMR
MS

Contains
CarbBank, Sugabase,
+ applications (3D-structure)

Future directions

Consortium for Functional Glycomics
Carbohydrate-protein interactions.
Glycosylation disorders in knock-out mice.

web.mit.edu/glycomics/consortium/

Japanese Consortium for Glycobiology & Glycotechnology
Everything – and then some…

www.jcgg.jp

EuroCarbDB
Structure (primary & 3D) & spectroscopy (NMR, MS)

Russian initiative
CarbBank/NMR (structure & NMR)

Future directions

Cross-linking between resources – makes it easy to find related information.

Portals – one interface to different resources.

Better interfaces – current interfaces are often too complex.

XML – allows data to be transferred directly to local applications.

Conclusion

There is a wide range of carbohydrate related resources available on the WWW.

Many provide useful information but all are rather limited in scope.

There are problems transfering data between databases. The interfaces are difficult to use. Manuals or instructions are often missing.