

Glykoinformatika

Databáze a nástroje

Aplikovaná bioinformatika, Jaro 2014

SACHARIDY – KÓDY A GRAFICKÁ ZNÁZORNĚNÍ

- Základní jednotky sacharidů (glykanů) jsou **monosacharidy**.
- NA/proteiny – základní jednotky (nukleotidy, aminokyseliny) jsou jasně definovány a jejich počet není velký.
- Monosacharidů je příliš **mnoho**, nelze (obecně) použít jednopísmenný kód.

A Novel Linear Code[®] Nomenclature for Complex Carbohydrates

Banin, Ehud; Neuberger, Yael; Altshuler, Yaniv; Halevi, Asaf; Inbar, Ori; Dotan Nir; and Avinoam Dukler^{*}

Glycominds Ltd, 1 Yodfat St., Alon Bldg., Global Park, Lod, 71291, Israel

FAX: 972-8-9181081, E-mail: dukler@glycominds.com

Trivial Name	Monosaccharide / Core ¹	Linear Code
D-Glcp	D-Glucose	G
D-Galp	D-Galactose	A
D-GlcpNAc	<i>N</i> -Acetylglucosamine	GN
D-GalpNAc	<i>N</i> -Acetylgalactosamine	AN
D-Manp	D-Mannose	M
D-Neup5Ac	<i>N</i> -Acetylneuraminic acid	NN
D-Neup	Neuraminic acid	N
KDN ²	2-Keto-3-deoxynanonic acid	K
Kdo	3-deoxy-D-manno-2 Octulopyranosylono	W
D-GalpA	D-Galacturonic acid	L
D-Idop	D-Ioduronic acid	I
L-Rhap	L-Rhamnose	H
L-Fucp	L-Fucose	F
D-Xylp	D-Xylose	X
D-Ribp	D-Ribose	B
L-Araf	L-Arabinofuranose	R
D-GlcpA	D-Glucuronic acid	U
D-Allp	D-Allose	O
D-Apip	D-Apiose	P
D-Fruf	D-Fructofuranose	E

The Linear Code assigns a single letter code to the most common structures of monosaccharides found in vertebrates

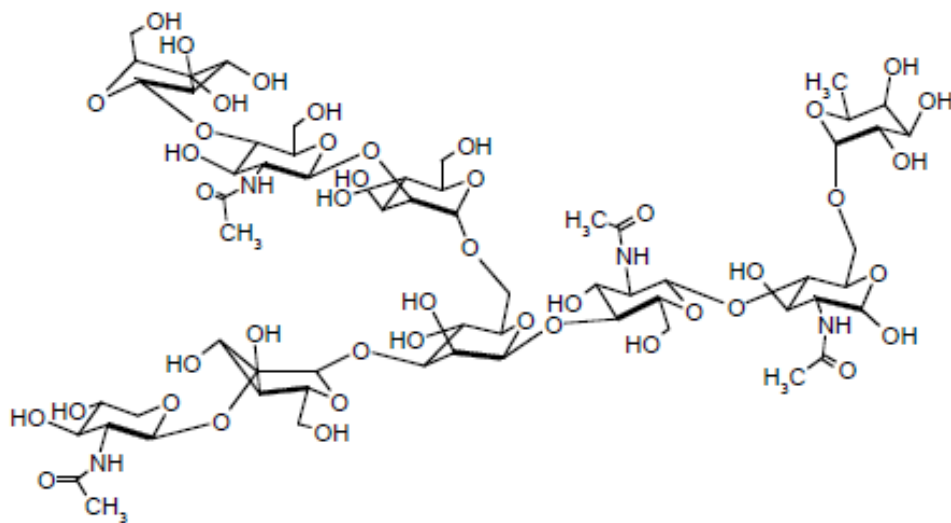
- Jen pro omezenou skupinu sacharidů.
- Problém: **modifikace, vazby, větvení.**
- Ani „lineární“ kód není jednopísmenný.

A Novel Linear Code[®] Nomenclature for Complex Carbohydrates

Banin, Ehud; Neuberger, Yael; Altshuler, Yaniv; Halevi, Asaf; Inbar, Ori; Dotan Nir; and Avinoam Dukler^{*}

Glycominds Ltd, 1 Yodfat St., Alon Bldg., Global Park, Lod, 71291, Israel

FAX: 972-8-9181081, E-mail: dukler@glycominds.com



GNb2Ma3(Ab4GNb2Ma6)Mb4GNb4(Fa6)GNa

- Jen pro omezenou skupinu sacharidů.
- Problém: **modifikace, vazby, větvení.**
- Ani „lineární“ kód není jednopísmenný.

Review

Structural glycobiology: A game of snakes and ladders

Symbol	Core monosaccharides ^a	Symbol	Special cases and common derivatives ^b
	Pentoses		2-<i>N</i>-Acetylhexosamines
A	arabinose	GN	<i>N</i> -acetylglucosamine
R	ribose	LN	<i>N</i> -acetylgalactosamine
X	xylose	MN	<i>N</i> -acetylmannosamine
D	lyxose		Uronic Acids
	Hexoses	GU	glucuronic acid
G	glucose	LU	galacturonic acid
M	mannose	MU	manuronic acid
L	galactose	IU	iduronic acid
I	idose		Sialic Acids (9 carbon)
T	talose	9N	<i>N</i> -acetylneuraminic acid
N	allose	9G	<i>N</i> -glycolylneuraminic acid
E	altrose	9O	3-deoxy- <i>D</i> -manno-oct-2-ulosonic acid (KDO)
K	gulose		Others
	Hexuloses	8O	3-deoxy- <i>D</i> -glycero- <i>D</i> -galacto-non-2-ulosonic acid (KDN)
C	fructose		
P	psicose		
S	sorbose		
J	tagatose		
	6-Deoxyhexoses		
F	Fucose		
Q	quinovose		
H	rhamnose		

Mari L DeMarco and Robert J Woods¹

Complex Carbohydrate Research Center, University of Georgia, Athens, GA,
30602-4712, USA

IUPAC NÁZVOSLOVÍ

UNION OF PURE AND APPLIED CHEMISTRY

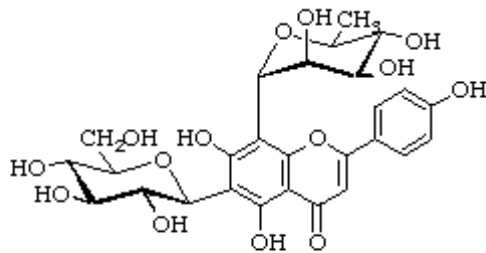
- <http://www.chem.qmul.ac.uk/iupac/2carb/>
- **Systematické názvy/zkrácené názvy (kódy)/triviální názvy**

- **Grafické znázornění (větvení)**

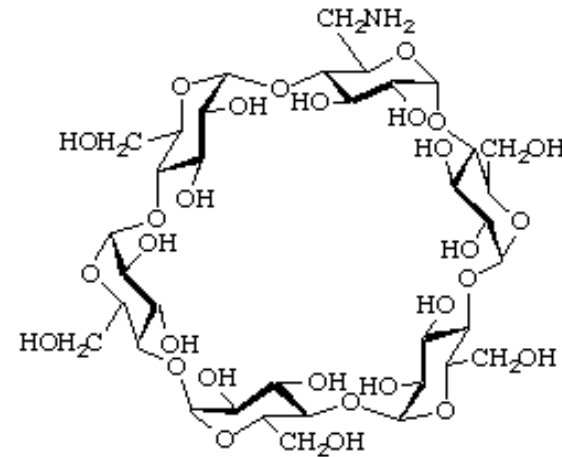
- **Preamble**
 - **2-Carb-0.** Historical
 - 0.1. Early app
 - 0.2. The cont
 - 0.3. Cyclic fo
 - 0.4. Nomencl
 - **2-Carb-1.** Definition
 - 1.1. Carbohy
 - 1.2. Monosac
 - 1.2.1.
 - 1.2.2.
 - 1.3. Dialdose
 - 1.4. Diketose
 - 1.5. Ketoald
 - 1.6. Deoxy st
 - 1.7. Amino st
 - 1.8. Alditols
 - 1.9. Aldonic a
 - 1.10. Ketoalc
 - 1.11. Uronic
 - 1.12. Aldaric
 - 1.13. Glycosi
 - 1.14. Oligosa
 - 1.15. Polysac
 - 1.16. Conven
 - **2-Carb-2.** Parent m
 - 2.1. Choice o
 - 2.2. Numberi
 - 2.2.1.
 - **2-Carb-3.** The Fischer pr
 - 2.3.3. Choi
 - **2-Carb-4.** Configuration:
 - 4.1. Use of D and
 - 4.2. The configurat
 - 4.3. Configuration:
 - 4.4. Racemates an
 - 4.5. Optical rotatic
 - **2-Carb-5.** Cyclic forms a
 - 5.1. Ring size
 - 5.2. The Fischer p
 - 5.3. Modified Fisc
 - 5.4. The Haworth
 - 5.5. Unconvention
 - 5.6. The Mills dep
 - 5.7. Depiction of c
 - 5.8. Conformation
 - **2-Carb-6.** Anomeric form
 - 6.1. The anomeric
 - 6.2. The anomeric
 - 6.3. Mixtures of ar
 - 6.4. Use of α and
 - **2-Carb-7.** Conformation
 - 7.1. The conforma
 - 7.2. Notation of ri
 - 7.3. Notation of va
 - 7.4. Enantiomers
 - **2-Carb-8.** Aldoses
 - 8.1. Trivial names
 - 8.2. Systematic
 - 8.3. Multiple configurat
 - 8.4. Multiple sets of chir:
 - 8.5. Anomeric configurat
 - **2-Carb-9.** Dialdoses
 - 10.1. Classification
 - 10.2. Trivial names
 - 10.3. Systematic names
 - 10.4. Configurational pre
 - **2-Carb-10.** Ketoses
 - 11.1. Systematic names
 - 11.2. Multiple sets of chi
 - **2-Carb-11.** Diketoses
 - 12.1. Systematic names
 - 12.2. Dehydro names
 - **2-Carb-12.** Ketoaldoses (aldol
 - 13.1. Trivial names
 - 13.2. Names derived fro
 - 13.3. Systematic names
 - 13.4. Deoxy alditols
 - **2-Carb-13.** Deoxy sugars
 - 14.1. General principles
 - 14.2. Trivial names
 - 14.3. Systematic names
 - **2-Carb-14.** Amino sugars
 - 16.1. Replacement of hy
 - 16.2. Replacement of OI
 - 16.3. Unequal substitutio
 - **2-Carb-15.** Thio sugars and ot
 - **2-Carb-16.** Other substituted n
 - 17.1. Trivial names
 - 17.2. Systematic names
 - 17.3. Derivatives
 - **2-Carb-17.** Unsaturated monosaccha
 - 17.1. General principles
 - 17.2. Double bonds
 - 17.3. Triple bonds and cumulat
 - **2-Carb-18.** Branched-chain sugars.
 - 18.1. Trivial names
 - 18.2. Systematic names
 - 18.3. Choice of parent
 - 18.4. Naming the branches
 - 18.5. Numbering
 - 18.6. Terminal substitution
 - **2-Carb-19.** Alditols
 - 19.1. Naming
 - 19.2. *meso* Forms
 - **2-Carb-20.** Aldonic acids
 - 20.1. Naming
 - 20.2. Derivatives
 - **2-Carb-21.** Ketoaldonic acids
 - 21.1. Naming
 - 21.3. Derivatives
 - **2-Carb-22.** Uronic acids
 - 22.1. Naming and numbering
 - 22.2. Derivatives
 - **2-Carb-23.** Aldaric acids
 - 23.1. Naming
 - 23.2. *meso* Forms
 - 23.3. Trivial names
 - 23.4. Derivatives
 - **2-Carb-24.** O-Substitution
 - 24.1. Acyl (alkyl) names
 - 24.2. Phosphorus esters
 - 24.2.1. Phosphates
 - 24.2.2. Phosphonates
 - 24.2.3. Phosphinates
 - 24.3. Sulfates
 - **2-Carb-25.** N-Substitution
 - **2-Carb-26.** Intramolecular anhydrides
 - **2-Carb-27.** Intermolecular anhydrides
 - **2-Carb-28.** Cyclic acetals
 - **2-Carb-29.** Hemiacetals and hemithioacetals
 - **2-Carb-30.** Acetals, ketals and their thio analogues
 - **2-Carb-31.** Names for monosaccharide residues
 - 31.1. Glycosyl residues
 - 31.2. Monosaccharides as substituent groups
 - 31.3. Bivalent and trivalent residues
 - **2-Carb-32.** Radicals, cations and anions
 - 32.1. Radicals
 - 32.2. Cations
 - 32.3. Anions
 - 32.4. Radical ions
 - **2-Carb-33.** Glycosides and glycosyl compounds
 - 33.1. Definitions
 - 33.2. Glycosides
 - 33.3. Thioglycosides.
 - 33.4. Selenoglycosides
 - 33.5. Glycosyl halides
 - 33.6. N-Glycosyl compounds (glycosylamines)
 - 33.7. C-Glycosyl compounds
- ligosaccharides
- ligosaccharides without a free hemiacetal group
- ligosaccharides with a free hemiacetal group
- branched oligosaccharides
- cyclic oligosaccharides
- [7.4.1.](#) Semisystematic names
- [7.4.2.](#) Systematic names
- ligosaccharide analogues
- use of symbols for defining oligosaccharide structures
- general considerations
- representations of sugar chains
- the extended form
- the condensed form
- the short form
- oligosaccharides
- names for homopolysaccharides
- designation of configuration of residues
- designation of linkage
- naming of newly discovered polysaccharides
- uronic acid derivatives.
- amino sugar derivatives
- oligosaccharides composed of more than one kind of re
- substituted residues
- glycoproteins, proteoglycans and peptidoglycans
- Official Names for Carbohydrates and Derivatives with the
- [1st monosaccharides](#)
- [and mon trival names](#)
- [and names formed by modification](#) of non-standard mon
- ...ucose-based Terms

IUPAC NÁZVOSLOVÍ

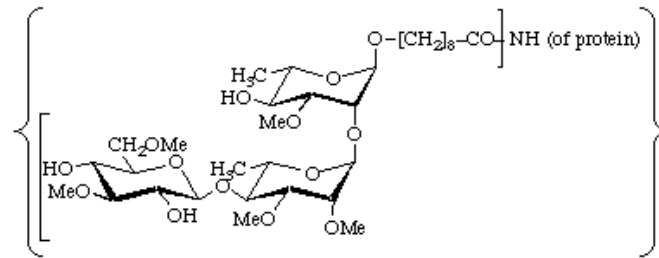
UNION OF PURE AND APPLIED CHEMISTRY



6-β-D-Glucopyranosyl-4',5,7-trihydroxy-8-α-L-rhamnopyranosylflavone;
trivial name violanthin



6¹-Amino-6¹-deoxycyclomaltohexaose

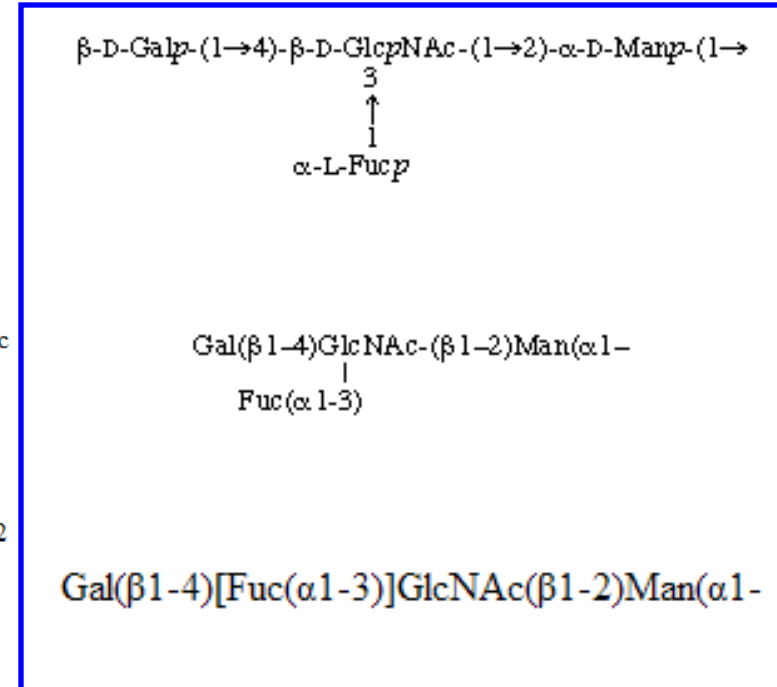


Poly-[(3,6-Di-O-methyl-β-D-glucopyranosyl)-(1 → 4)-(2,3-di-O-methyl-α-L-rhamnopyranosyl)-(1 → 2)-(3-O-methyl-α-L-rhamnopyranosyloxy)-(1-O → 9)-nonanoyl-(1 → N)]-protein
{[β-D-Glcp3,6Me2-(1 → 4)-α-L-Rhap2,3Me2-(1 → 2)-α-L-Rhap3Me-(1-O → 9)-nonanoyl-(1 → N)]_n-protein}

IUPAC NÁZVOSLOVÍ

UNION OF PURE AND APPLIED CHEMISTRY

Abequose	Abe	Iduronic acid	IdoA
Allose	All	Lyxose	Lyx
Altrose	Alt	Mannose	Man
Apiose	Api	Muramic acid	Mur
Arabinose	Ara	Neuraminic acid	Neu
Arabinitol	Ara-ol	<i>N</i> -Acetylneuraminic acid	Neu5Ac
2-Deoxyribose	dRib	<i>N</i> -Acetyl-2-deoxyneur-2-enaminic acid	Neu2en5Ac
Fructose	Fru	<i>N</i> -Glycoloyneuraminic acid	Neu5Gc
Fucose	Fuc	3-Deoxy-D- <i>manno</i> -oct-2-ulosonic acid	Kdo
Fucitol	Fuc-ol	Rhamnose	Rha
Galactose	Gal	3,4-Di- <i>O</i> -methylrhamnose	Rha3,4Me2
Galactosamine	GalN	Psicose	Psi
<i>N</i> -Acetylgalactosamine	GalNAc	Quinovose	Qui
β -D-Galactopyranose 4-sulfate	β -D-Galp4S	Ribose	Rib
Glucose	Glc	Ribose 5-phosphate	Rib5P
Glucosamine	GlcN	Ribulose	Ribulo (or Rul)
2,3-Diamino-2,3-dideoxy-D-glucose	GlcN3N	Sorbose	Sor
Glcitol	Glc-ol	Tagatose	Tag
<i>N</i> -Acetylglucosamine	GlcNAc	Talose	Tal
Glucuronic acid	GlcA	Xylose	Xyl
Ethyl glucopyranuronate	GlcpA6Et	Xylulose	Xyhulo (or Xul)
Gulose	Gul	2- <i>C</i> -Methylxylose	Xyl2CMe
Idose	Ido		



- **Tři IUPAC způsoby jak pomocí kódů znázornit oligosacharid.**
- **Existuje jich více!**

SACHARIDY – KÓDY A GRAFICKÁ ZNÁZORNĚNÍ

Linear	IUPAC	$\{\alpha\text{-D-GalpNAc-(1}\rightarrow\text{3)-}[\alpha\text{-L-Fucp-(1}\rightarrow\text{2)]-}\beta\text{-D-Galp-(1}\rightarrow\text{4)-}\beta\text{-D-GlcpNAc-(1}\rightarrow\text{3)-}\beta\text{-D-Galp-(1}\rightarrow\text{4)-}\beta\text{-D-Glcp}\}$																								
	LINUCS	[[[b-D-Glcp]{{{(4+1)[[b-D-Galp]{{{(3+1)[[b-D-GlcpNAc]{{{(4+1)[[b-D-Galp]{{{(2+1)[[a-L-Fucp]}}{(3+1)[[a-D-GalpNAc]}}}}}}}}}}]]]]]]																								
	LinearCode	ANa3 (Fa2) Ab4 GNb3 Ab4 Gb4 (spaces added for clarity)																								
	GLYCAM	0LN (0fA) ZLB 4Gn 3LB 4GB (with LinearCode precedence rules for branching)																								
Graphical	CFG																									
	Oxford																									
	GLYCAM/Oxford		<table border="1"> <thead> <tr> <th></th> <th>CFG</th> <th>Oxford</th> <th>GLYCAM</th> </tr> </thead> <tbody> <tr> <td>D-Galp</td> <td>●</td> <td>◇</td> <td>L</td> </tr> <tr> <td>D-GalpNAc</td> <td>■</td> <td>◆</td> <td>LN</td> </tr> <tr> <td>D-Glcp</td> <td>●</td> <td>□</td> <td>G</td> </tr> <tr> <td>D-GlcpNAc</td> <td>■</td> <td>■</td> <td>GN</td> </tr> <tr> <td>L-Fucp</td> <td>▲</td> <td>◊</td> <td>f</td> </tr> </tbody> </table> <p>Oxford-type linkage:</p> <p>..... α-linkage</p> <p>— β-linkage</p>		CFG	Oxford	GLYCAM	D-Galp	●	◇	L	D-GalpNAc	■	◆	LN	D-Glcp	●	□	G	D-GlcpNAc	■	■	GN	L-Fucp	▲	◊
	CFG	Oxford	GLYCAM																							
D-Galp	●	◇	L																							
D-GalpNAc	■	◆	LN																							
D-Glcp	●	□	G																							
D-GlcpNAc	■	■	GN																							
L-Fucp	▲	◊	f																							
PDB	GLYCAM	0LN—ZLB—4Gn—3LB—4GB 0fA	(shown in CFG graphical format for illustrative purposes only, actual connection information defined by PDB coordinates)																							

KÓDY A GRAFICKÁ ZNÁZORNĚNÍ – existuje mnoho způsobů a všechny jsou používány!



The Symbolic Representation Of Monosaccharides In The Age Of Glycobiology

Serge Perez

Abstract

Nomenclature

At the Instigation of Glycobiology

Beyond the Symbols

120 Monosaccharides

References

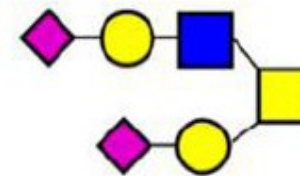
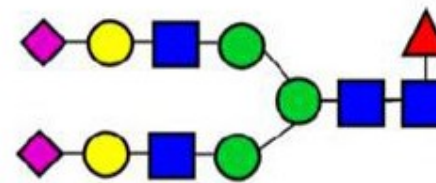
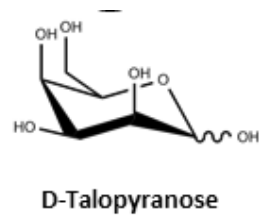
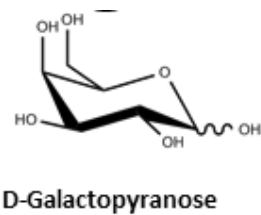
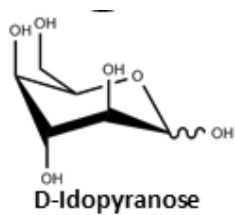
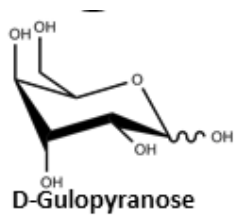
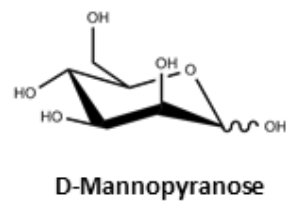
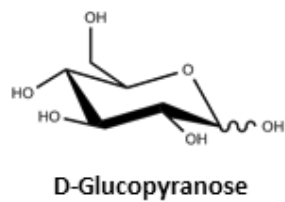
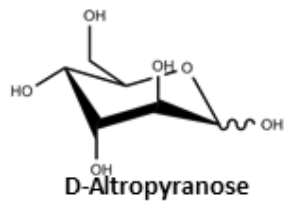
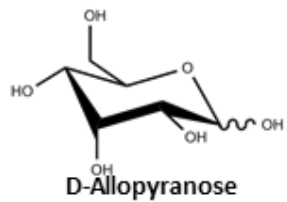
Abstract

This chapter offers a general background for students embarking in glycoscience to grasp an essential component of the field, i.e. the alphabet of the building blocks that constitute the many naturally occurring glycans and complex carbohydrates. There is a need to conform to the recommendations of nomenclatures of carbohydrates whilst the constraints required by the developing field of glycobiology in terms of visualization and encoding. The present chapter offers a unified presentation of the nomenclature and symbols that form part of the “language” used to communicate more effectively. It covers about 120 monosaccharides which have been identified as the building blocks of the vast majority of glycans. As a picture speaks a thousand words, it is hoped that this language could be understood as quickly and effectively as possible and used by those working in

[Click here for printable version](#)

<http://www.glycopedia.eu/Abstract>





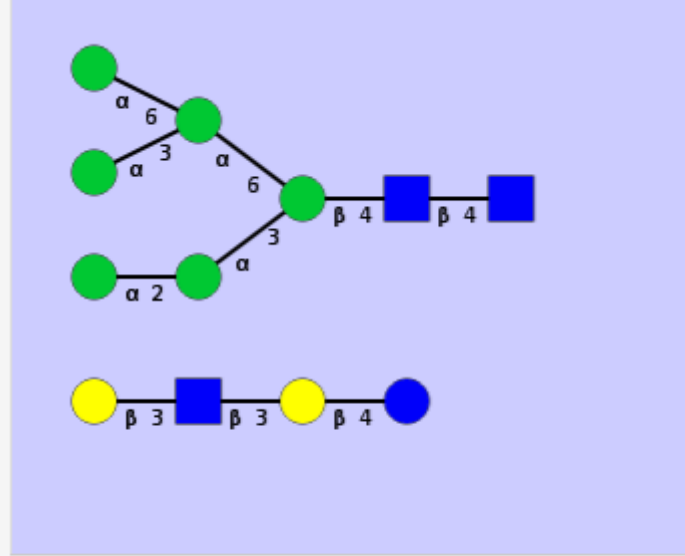
GLYCAN BUILDER

File View Structure

Delete Copy Paste Orientation Select all Select none Bracket Repeat Move CW Move CCW

1st Linkage 2nd Linkage

Star, Red Triangle, Yellow Circle, Blue Circle, Green Circle, Yellow Square, Blue Square, Purple Diamond, Light Blue Diamond



File View Structure

Export Image export Import Import from string Restart

BMP EPS JPG PDF PNG PS SVG

Orientation Select all Select none Bracket Repeat Move

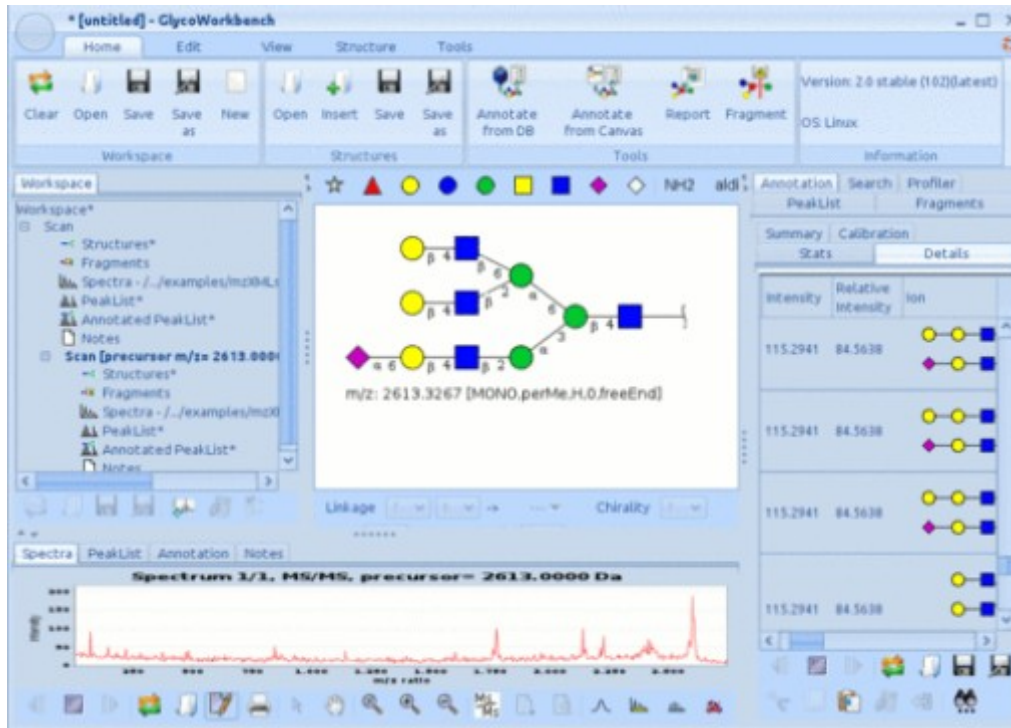
Search

<http://www.unicarbkb.org/builder>



<http://code.google.com/p/glycoworkbench/>

Semi-Automatic Interpretation and Annotation of Mass Spectra of Glycans



GlycoWorkBench

developed by Alessio Ceroni, Kai Maass and David Damerell

GlycoWorkbench is a suite of software tools designed for rapid drawing of glycan structures and for assisting the process of structure determination from mass spectrometry data. The graphical interface of GlycoWorkbench provides an environment in which structure models can be rapidly assembled, their mass computed, their fragments automatically matched with MSn data and the results compared to assess the best candidate. GlycoWorkbench can greatly reduce the time needed for the interpretation and annotation of mass spectra of glycans. GlycoWorkbench can also be used as JAVA program library to calculate the mass of glycans, generate images for glycan structure.

GLYCAM BIOMOLECULE BUILDER

http://glycam.ccruc.uga.edu/ccrc/biombuilder/biomb_index.jsp

Carbohydrate 3D Structure Predictor

This tool allows you to generate 3D structures for linear and branched oligosaccharides.

Please choose linkage configuration (α or β)

Configuration

α β

Isomer

L D

Ring Type

f p

Monosaccharides

Man	Gal	Glc	Ido	All	Alt	Gul	Tal
Xyl	Lyx	Rib	Ara	Fru	Psi	Sor	Tag
		Fuc	Rha	Qui	GalNAc	GlcNAc	ManNAc
	GalA	GlcA	IdoA	Neu5Ac	KDN	KDO	Neu5Gc

Sequence Termination (Aglycon)

-OH -OME -OtBu

1C4 4C1

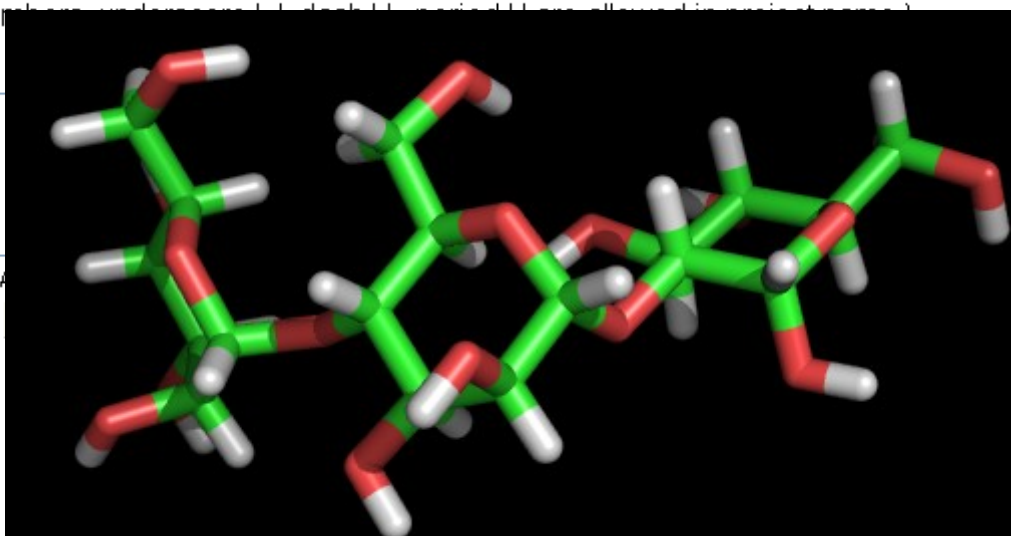
Linkages

1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9
2-1	2-2	2-3	2-4	2-5	2-6	2-7	2-8	2-9

Project Name : (Only letters, numbers and underscores are allowed in project name)

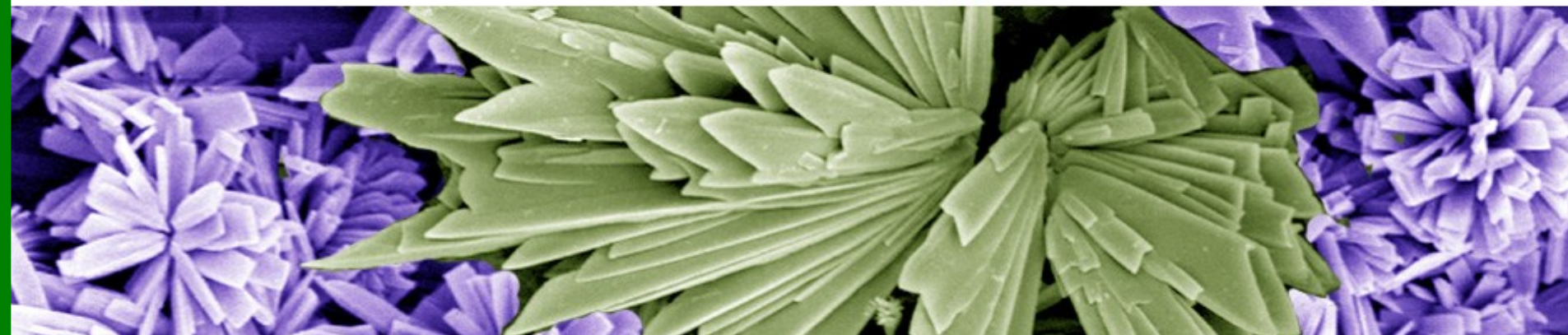
Add Branches [HELP - How to build a Carbohydrate](#)

Please be patient when building. A



DATABÁZE

- Databáze obsahující informace o **proteinech** (součást glykoproteinů, lektiny).
- Vlastní databáze **glykanů** (struktury).
- Databáze **enzymů** a **drah** účastnících se syntéz a odbourávání glykanů (sacharidů).
- Informace o **interakcích** protein-sacharid
- „**Glykocentra**“ – sdružené databáze, více databází, analytické nástroje



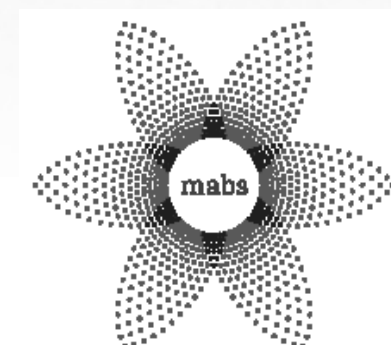
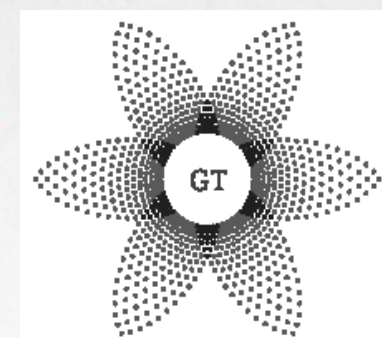
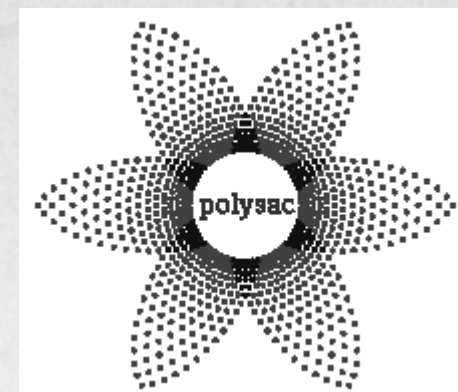
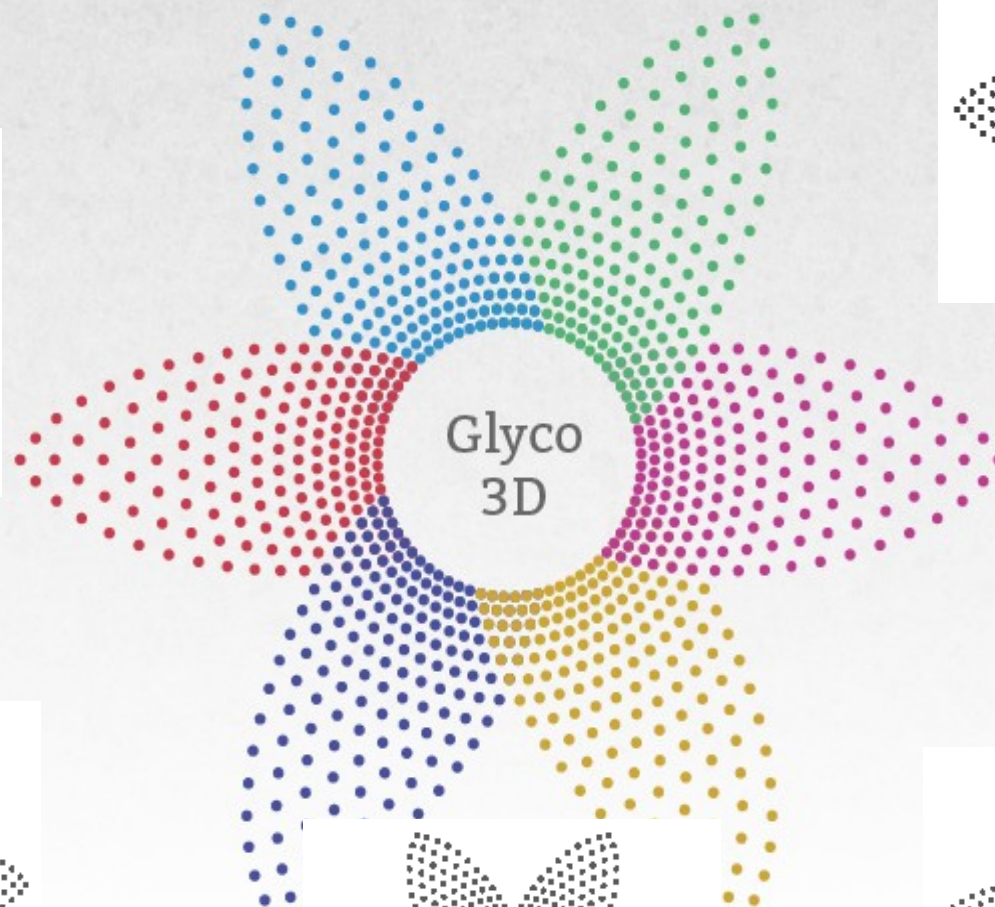
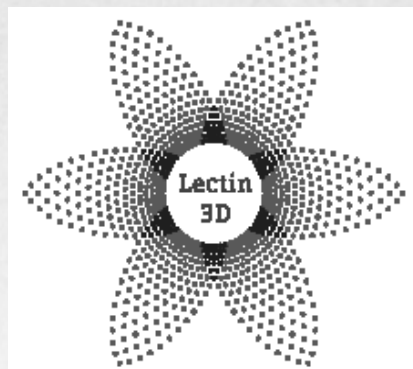
<http://www.cermav.cnrs.fr/en>

DATABÁZE Glyco3D

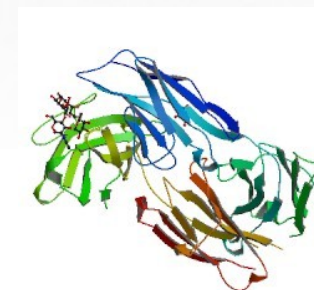
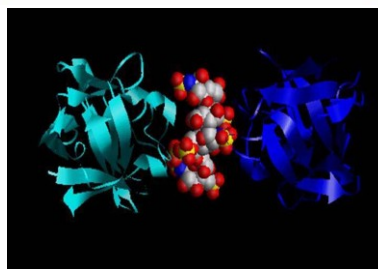
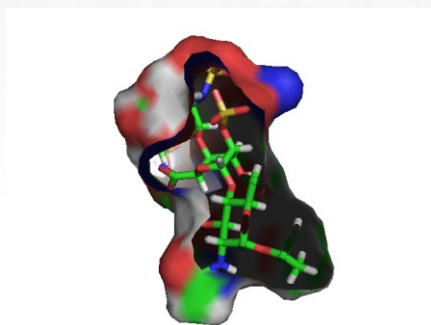
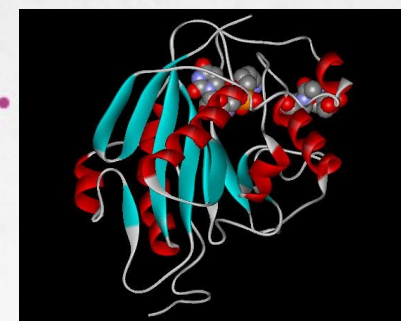
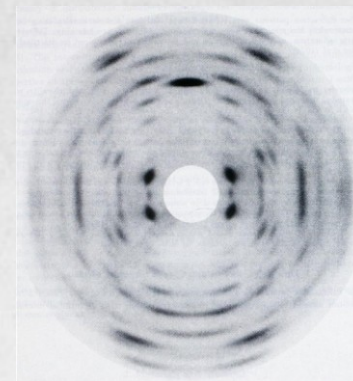
Glyco3D features a family of databases covering the 3D features of monos, di, oligo, polysaccharides, glycosyltransferases, lectins, monoclonal antibodies and glycosaminoglycan-binding proteins. Other databases are also made available that completes the picture of glycan 3D structure decoding. that are made freely available to the scientific community. A search engine has been developed that scans the full content of all the data bases for queries related to sequential information of the carbohydrates or other related descriptors. This database ensemble offers a unique opportunity to characterize the 3D features that a given oligosaccharide can assume in different environments.

<http://glyco3d.cermav.cnrs.fr/>

<http://glyco3d.cermav.cnrs.fr/>



<http://glyco3d.cermav.cnrs.fr/>



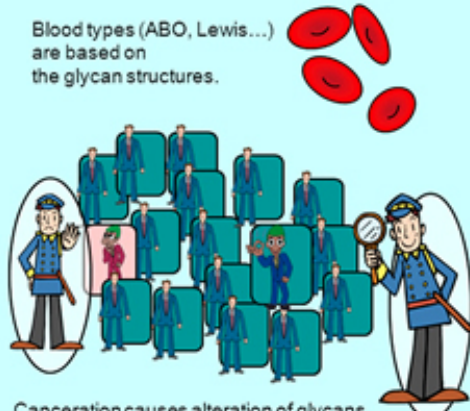
What is Glycan?



Functions and Roles of Glycans

Discrimination of cell variety

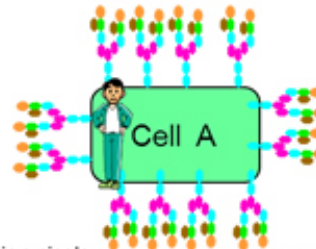
Blood types (ABO, Lewis...) are based on the glycan structures.



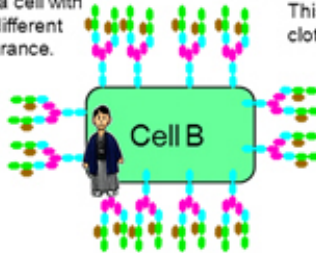
Canceration causes alteration of glycans. Cell differentiation accompanies alteration of glycans.

Applicable to development of cancer vaccine and regenerative medicine

Glycans are binding to proteins and lipids and mostly present on the cell surface.



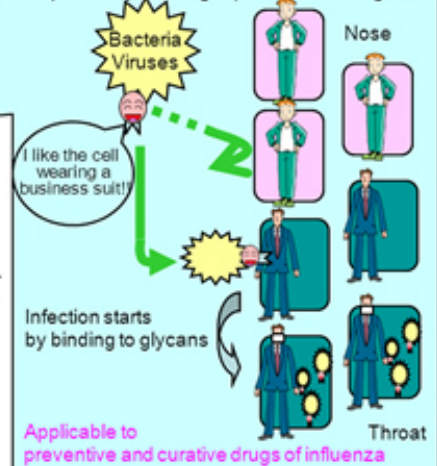
Difference in a single constituting monosaccharide results in a cell with totally different appearance.



Glycans with different structures bind to different species, individuals, and organs. This resembles unique clothing of cells.

Door to the infection with pathogenic microorganism

Bacteria and viruses bind to specific glycan structures on cell surfaces, which causes specificity among species and organs.



I like the cell wearing a business suit!

Infection starts by binding to glycans

Applicable to preventive and curative drugs of influenza

Quality control of glycoproteins



Glycoproteins with disrupted or lost glycans by aging or damage are proteolyzed and metabolized.

Applicable to elongation of drug

Protection of cells and proteins

Protect cells and proteins from enzyme digestion and heat or other degradation causes

Applicable to the development of an antifreeze organ preservation solution for transplantation that does not harm

Pharmaceutical production

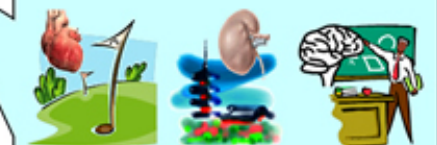
"Bare" proteins and proteins "wearing fur" are rejected as xenobiotics in the human body cannot work as a medicine.



Production by bacteria
Production by animal cells

Applicable to

Drug delivery system



Drugs are delivered to appropriate organs by selecting specific glycans.

Applicable to the medicine for lysosomal storage diseases



Drugs will be efficiently activated by

DATABÁZE

LEKTINY



LfDB

Lectin Frontier DataBase

About Lectin Frontier DataBase (LfDB)

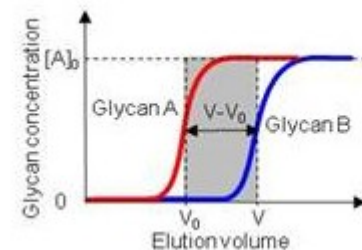
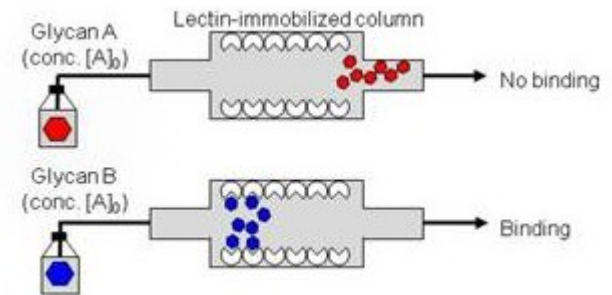
Significance of LfDB













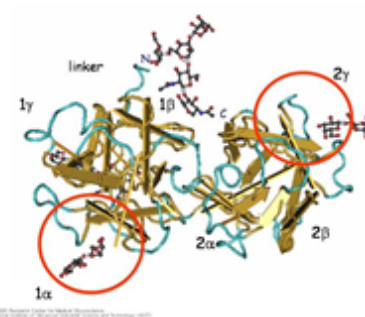
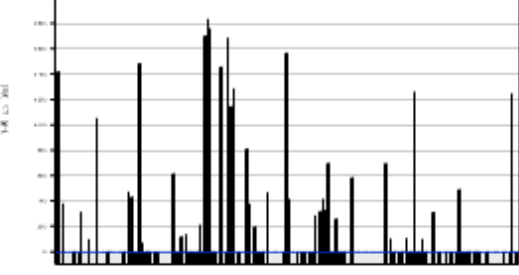


Fig. 1. A photograph of the automated FAC-FD system

of sugar-protein interactions, LfDB would be a valuable resource in the studies of glycan-related biology.

Lectin Frontier DataBase (LfDB) provides quantitative interaction data in terms of the affinity constants (K_a) of a series of lectins toward a panel of pyridylaminated (PA) glycans obtained by an automated frontal affinity chromatography with fluorescence detection (FAC-FD) system (Fig. 1). Since the data are accurate and reliable, providing the absolute values



Lectin ID	LPR-000008	Molecular Structure Jmol	
Lectin Name	RCAI, RCA120		
Lectin Family	R-type Lectin		
Monosaccharide Specificity	Gal		
Species (Scientific Name)	Caster Bean, Castor oil bean (<i>Ricinus communis</i>)		
Japanese Name [Family]	トウゴマ, ヒマ [トウダイグサ科]		
Kingdom	Plant		
Organ	seed		
Specific organ			
Number of CRD	2		
3D-fold	β-trefoil		
Accession (External Link)	GenBank: X52908, X03179, X02388 PDB: 2AAI Pfam: PF00652, PF00161		
Sequence ID		Interaction Graph [Viewer] [GlycanList] 	
Sequence	<pre> 1 MKPGGNTIVI WMYAVATWLC FGSTSGWSFT LEDNNIFPKQ YPIINFTTAG ATVQSYTNFI 61 RAVRGRLTIG ADVRHEIPVL PNRVGLPINQ RFILVELSNH AELSVTLALD VINAYVVGYS 121 AGNSAYFFHP DNQEDAEAIT HLFTDVQNRV TFAFGGNYDR LEQLAGNLRE NIELNGNPLE 181 EAISALYYYS TGGTQLPTLA RSFIICIQMI SEAARFQYIE GEMRTRIRYN RRSAPDPSVI 241 TLENSWGRLS TAIQESNQGA FASPIQLQRR NGSKFSVYDV SILIPIIHALM VYRCAPPSS 301 QFSLLRPVV PNFNADVCMC PEPIVRIVGR NGLCVDVRDG RFHNGNAIQL WPCKSNTDAN 361 QLWTLKRDNT IRSNGKCLTT YGSPGVYVM IYDCNTAATD ATRWQIWDNG TIINPRSSLV 421 LAATSGNSGT TLTVQINIYA VSQGWLPTNN TQPFVTTIVG LYGLCLQANS GQVWIEDCSS 481 EKAEQQWALY ADGSIRPQQN RDNCLTSDSN IRETVVKILS CGPASSGQRW MFKNDGTILN 541 LYSGLVLDVR ASDPSLKQII LYPLHGDPNQ IWLPLF </pre>		



Lectin ID	LPR-000008
Lectin Name	RCA I, RCA120
Lectin Family	R-type Lectin
Monosaccharide Specificity	Gal
Species (Scientific Name)	Caster Bean, Castor oil bean (<i>Ricinus communis</i>)
Japanese Name [Family]	トウゴマ, ヒマ [トウダイグサ科]
Kingdom	Plant
Organ	seed

Molecular Structure Jmol

3D structure

3D structure

3D structure

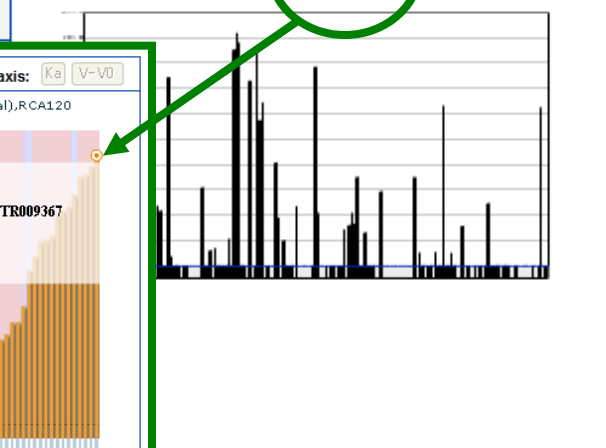
3D structure

Binding Site

Binding Site

Binding Site

Interaction Graph [Viewer] [GlycanList]



Style: X-axis sort: Y-axis:

RCA I (RCA120_Final), RCA120

Family: Galactosylated
oligo_id: ONG-00004d
JCGGDB ID: JCGG-STR009367

JCGGDB ID: JCGG-STR009367 Family: Galactosylated

One Parameter

Legend: Glc (blue circle), Man (green circle), Gal (yellow circle), GlcNAc (blue square), GalNAc (yellow square), Fuc (red triangle), Xyl (white star), Neu5Ac (purple diamond), Neu5Gc (white diamond), GlcA (yellow circle), IdoA (blue circle), GlcN (green circle), dJUA (white star)

```

TAG ATVQSYTNFI
ALD VINAYVVGYS
LRE NIELNGNPLE
RYN RRSAPDPSVI
ALM VYRCAPPSS
IQL WPCKSNTDAN
DNG TIINPRSSLV
ANS GQWIEDCSS
QRW MFKNDGTILN

```

RCMG

What is Glycan?

HOPI

JCGGDB

Japan Consortium for Glycobiology and Glycotechnology DataBase

DATABÁZE

„GLYKOGENY“

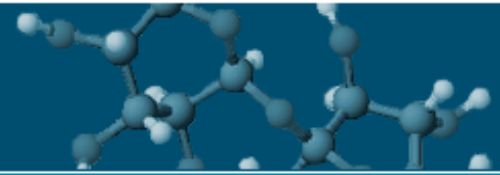


About GlycoGene DataBase

Significance of GGDB

Glycogene includes genes associated with glycan synthesis such as glycosyltransferase, sugar nucleotide synthases, sugar-nucleotide transporters, sulfotransferases, etc. At present, over 180 human glycogenes were identified, cloned and characterized. In "Construction of GlycoGene Library Project" (April, 2001 - March, 2004), we collected and compiled the data on such glycogenes as GlycoGene Database (GGDB), which is the first database to store information on substrate specificity. GGDB provides necessary information for the analysis of glycogenes.

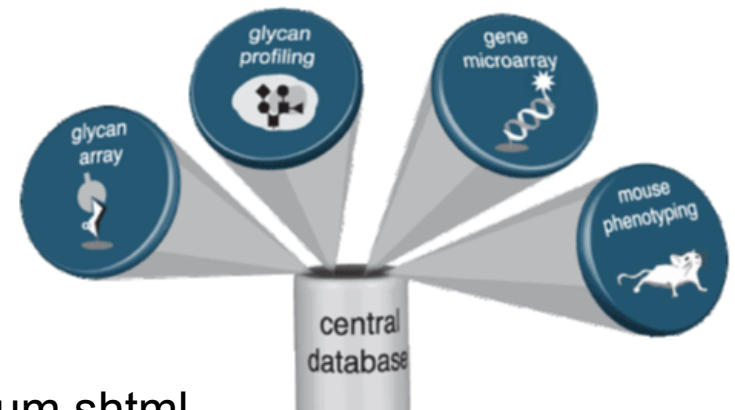
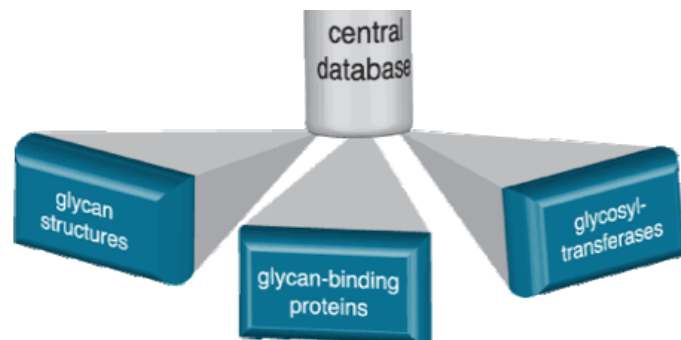
<http://jcggdb.jp/rcmg/ggdb/>



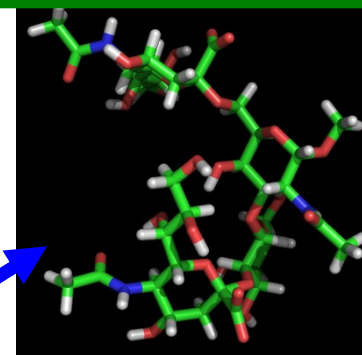
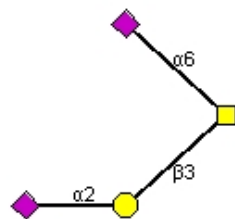
What is the CFG?

The Consortium for Functional Glycomics (CFG) is a large research initiative funded by [NIGMS](#) to define the paradigms by which protein-carbohydrate interactions mediate cell communication. The CFG works with the scientific community to create unique [resources and services](#) that Participating Investigators can use in their own research. These resources and services, which are provided free of charge by the CFG's [Scientific Cores](#), include glycan array screening, gene microarray screening, mouse phenotyping, glycan profiling, a reagent bank, and data analysis tools. Resulting data sets are integrated and made accessible to the community via the CFG's [central database](#). The CFG also has publicly accessible [specialty databases](#) that offer detailed information on [glycan-binding proteins](#), [glycan structures](#), and [glycosyltransferases](#). The number of investigators using CFG resources in their work continues to grow, as reflected by the rate of [new publications involving CFG resources](#). Our vision is to harness the combined research power of our Participating Investigators and our Scientific Cores to achieve our goals.

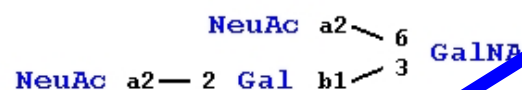
CONSORTIUM FOR FUNCTIONAL GLYCOMICS



Cartoon Representation



IUPAC 2D Representation



IUPAC Code

NeuAc α 2-2 *Gal* β 1-3(*NeuAc* α 2-6) *GalNAc*

Linear Code

NNa2Ab3(NNa6)AN

Link to 3D Model (GLYCAM - Web)

[Link to GLYCAM for NNa2Ab3\(NNa6\)AN](#)

Sub Structure Search Interface

[Load this structure for sub structure search](#)

General Information

Glycan Family:	O-linked
Sub. Family:	Core 1
Last Updated:	05/18/2004
Oligosaccharide Molecular Wt.:	965.8682
Calculated Oligosaccharide Molecular Wt.:	965.853
Permethylated MW.:	1240
Composition:	Hex ₁ NeuAc ₂ HexNAc ₁
Status:	Public

References

Zamze S, Harvey DJ, Pesheva P, Mattu TS, Schachner M, Dwek RA, Wing DR Glycobiology 1999;[823-831]{[PubMed](#)}

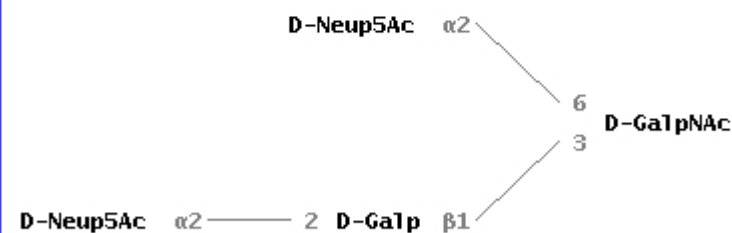
Biological Sources

Taxonomy Name	Organ	Tissue Type	Cell Type
Mus musculus(Mouse)	Brain	-	-

Customize the below sub-structure by clicking on monosaccharides or linkages that you wish to modify.

- Clicking on a monosaccharide allows you to add extensions, replace it with a different monosaccharide, or add modifiers (phosphate groups, sulfate groups, etc.) Extensions can only be added towards the non-reducing end.
- Clicking on a monosaccharide and selecting "Trim" removes it, along with all associated monosaccharides toward the non-reducing end.
- Clicking on a linkage allows you to replace it with other linkages, including "wild cards." Click "Run Query" to search for glycans that contain the finished structure.

Build from other templates



Match the motif at non-reducing end only:

Run Query



UniCarbKB

<http://www.unicarbkb.org/>

UniCarbKB is an initiative that aims to promote the creation of an online information storage and search platform for glycomics and glycobiology research. The knowledgebase will offer a freely accessible and information-rich resource supported by querying interfaces, annotation technologies and the adoption of common standards to integrate structural, experimental and functional data.

GlycoSuiteDB - a database of curated glycan structural data (previously commercial)

GLYCAM - an environment for 3-D glycan structures

EUROCarbDB - an experimental and structural database

GlycoBase - an experimental LC database of derivatized glycans

GlycoMod - a tool for calculating composition of glycans from MS data

RINGS - a web resource providing algorithmic and data mining tools

SugarBind - a database of pathogen-glycan interactions



UniCarbKB

<http://www.unicarbkb.org/>

UniCarbKB is an initiative that aims to promote the creation of an online information storage and search platform for glycomics and glycobiology research. The knowledgebase will offer a freely accessible and information-rich resource supported by querying interfaces, annotation technologies and the adoption of common standards to integrate structural, experimental and functional data.

Published online 13 November 2013

*Nucleic Acids Research, 2014, Vol. 42, Database issue D215–D221
doi:10.1093/nar/gkt1128*

UniCarbKB: building a knowledge platform for glycoproteomics

Matthew P. Campbell¹, Robyn Peterson¹, Julien Mariethoz², Elisabeth Gasteiger³, Yukie Akune⁴, Kiyoko F. Aoki-Kinoshita⁴, Frederique Lisacek^{2,5} and Nicolle H. Packer^{1,*}

¹Biomolecular Frontiers Research Centre, Macquarie University, North Ryde, NSW 2109, Australia, ²Proteome Informatics Group, Swiss Institute of Bioinformatics, Geneva, Switzerland, ³Swiss-Prot Group, Swiss Institute of Bioinformatics, Geneva, Switzerland, ⁴Department of Bioinformatics, Faculty of Engineering, Soka University, 1-236 Tangi-machi, Hachioji, Tokyo, Japan and ⁵Section of Biology, Faculty of Sciences, University of Geneva, Switzerland

Welcome to the Carbohydrate-Active enZymes Database

The **CAZY** database describes the families of structurally-related catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds.

<http://csdb.glycoscience.ru/bacterial/>



Bacterial Carbohydrate Structure DataBase

This database is aimed at provision of structural, bibliographic, taxonomic and related information on bacterial carbohydrate structures. Two key points of this service are:

- **covering** - is above 90% in the scope of bacterial carbohydrates. This means the negative search answer remains valuable scientific information.
- **consistence** - we manually check the data, and aim at high quality error-free content.

PROGLYCPROT

A Repository of Experimentally Characterized
Glycoproteins of Prokaryotes

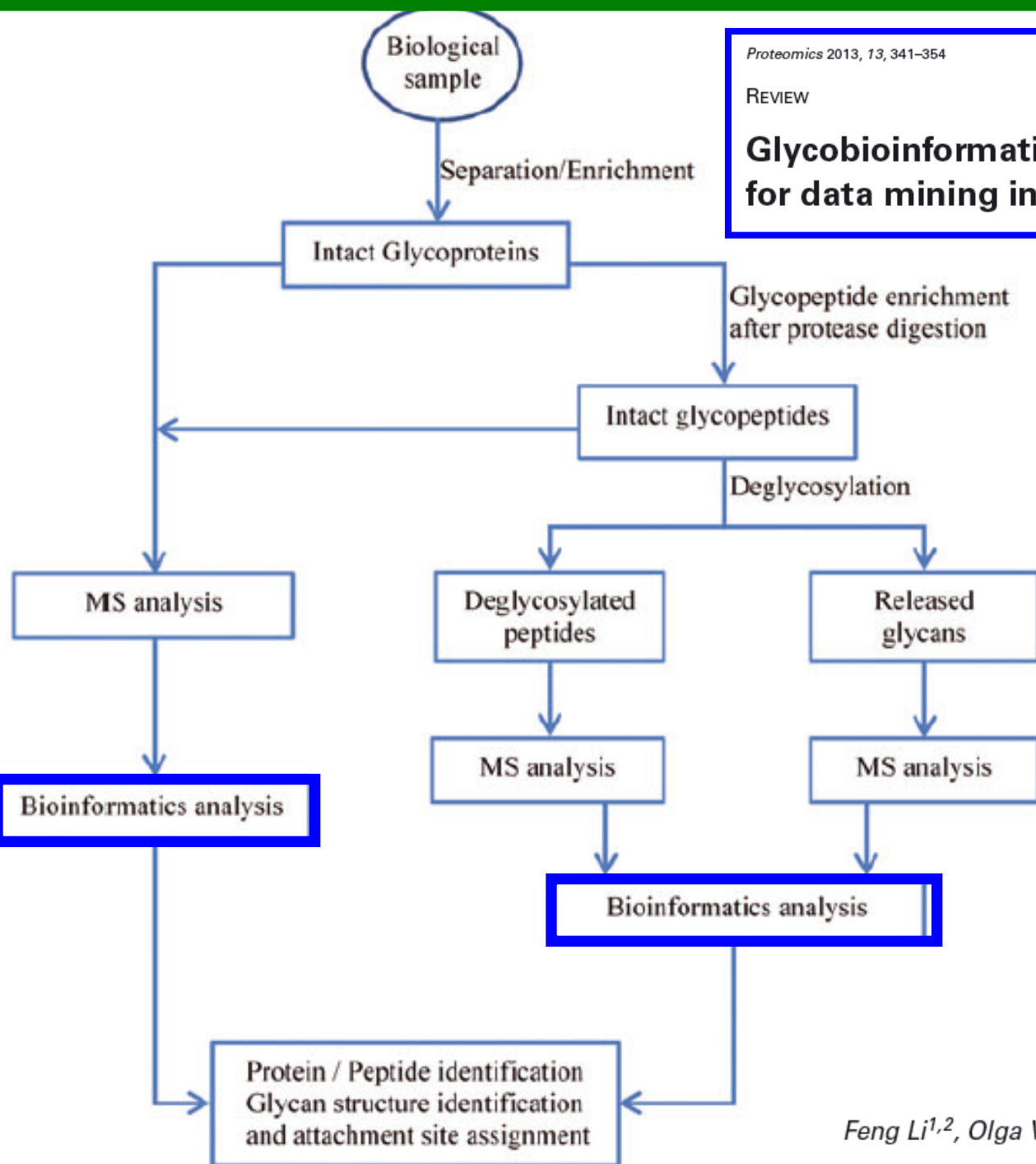
ProGlycProt (**Pro**karyotic **Glyc**oproteins) is a manually curated, comprehensive repository of experimentally characterized bacterial glycoproteins and archaeal glycoproteins, generated from an exhaustive literature search. This is the focused beginning of an effort to provide concise relevant information derived from rapidly expanding literature on prokaryotic glycoproteins, their glycosylating enzyme(s), glycosylation linked genes, and genomic context thereof, in a cross-referenced manner.

<http://www.proglycprot.org/>

SOFTWAREVÉ NÁSTROJE PRO ANALÝZU GLYKANŮ

- Charakterizace glykanů: hmotnostní spektroskopie (**MS**), vysokoúčinná kapalinová chromatografie (**HPLC**), nukleární magnetická rezonance (**NMR**).
- Velká část softwarových nástrojů je zaměřena na zpracování a interpretaci experimentálních dat, tj. analýza glykanů je bez využití bioinformatiky velmi obtížná (prakticky nemožná)...

Glycobioinformatics: Current strategies and tools for data mining in MS-based glycoproteomics



Chcete vědět víc?

C7250 Charakterizace proteinů hmotnostní spektrometrií

Glycoprotein fractions



Glycoconj J (2013) 30:89–117
DOI 10.1007/s10719-012-9444-8

Analytical glycobiology at high sensitivity: current approaches and directions

Milos V. Novotny • William R. Alley Jr. •
Benjamin F. Mann

NÁSTROJE PRO INTERPRETACI MS DAT

GlycoMod Tool

<http://web.expasy.org/glycomod/>

GlycoMod is a tool that can predict the possible oligosaccharide structures that occur on proteins from their experimentally determined masses. The program can be used for free or derivatized oligosaccharides and for glycopeptides [Documentation / Mass values / Reference / Disclaimer].

Note: You can use **GlycanMass** to calculate the mass of an oligosaccharide structure from its oligosaccharide composition.

Glyco-Peakfinder

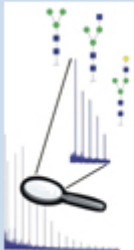
mass

residue

ion/charge

modifications

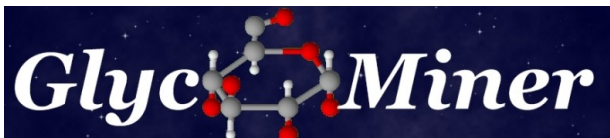
<http://www.glyco-peakfinder.org/>



Introduction

The "Glyco-Peakfinder" is a tool for fast annotation of glycan MS spectra. MS-profiles, MSⁿ spectra with different types of ions (glycosidic cleavages and/or cross-ring cleavages) can be calculated in parallel. The option of detecting differently- and/or multiply-charged ions in one calculation cycle provides a fast and complete annotation of the whole spectrum. All the additional options of "Glyco-Peakfinder" (e.g. calculation of modifications either at the reducing end or within the sequence) increase the field of application from native glycans to a variable set of glycoconjugates. The results from "Glyco-Peakfinder" can be used for advanced database searches in GLYCOSCIENCES.de.

NÁSTROJE PRO INTERPRETACI MS DAT



<http://www.chemres.hu/ms/glycominer/>

GlycoMiner, has been developed to automatically identify MS/MS spectra obtained in LC-MS runs which correspond to N-glycopeptides. The program complements conventional proteomics analysis, and can be used in a high-throughput environment. It interprets the spectra and determines the structure of the corresponding glycopeptides. GlycoMiner runs under Windows, can process spectra obtained on various instruments, if monoisotopic, singly charged spectra are supplied. The algorithm works similar to a human expert; evaluates the low mass oxonium ions; oligosaccharide losses from the molecular ion and identifies the mass of the peptide residue.

GlycoWorkBench

developed by Alessio Ceroni, Kai Maass
and David Damerell

GlycoWorkbench is a suite of software tools designed for rapid drawing of glycan structures and for assisting the process of structure determination from mass spectrometry data. The graphical interface of GlycoWorkbench provides an environment in which structure models can be rapidly assembled, their mass computed, their fragments automatically matched with MSⁿ data and the results compared to assess the best candidate. GlycoWorkbench can greatly reduce the time needed for the interpretation and annotation of mass spectra of glycans. GlycoWorkbench can also be used as JAVA program library to calculate the mass of glycans, generate images for glycan structure.

<http://code.google.com/p/glycoworkbench/>

3D STRUKTURA SACHARIDŮ

- RTG krystalografie – struktury **glykoproteinů** a sacharidů v komplexu s proteiny (**lektiny, enzymy, protilátky**).
Problém: **velká flexibilita sacharidů** (ve struktuře je viditelná jen část glykanu, změny přirozených konformací).
Problém: Kvalita 3D struktur sacharidů v PDB může být nízká...
- Usnadnění vyhledávání „glykostruktur“ v PDB:
Glycoconjugate Data Bank (<http://www.glycostructures.jp/>)
Glycosciences.de (<http://www.glycosciences.de/database/index.php>)

Glycoconjugate Data Bank

4 analogous glycans found

Glycan Structure	# of glycans
Neu5Ac2 a2 8 Neu5Ac1	4
Neu5Ac3 a2 8 Neu5Ac2 a2 8 Neu5Ac1	6
Neu5Ac1 a2 8 Neu5Ac2 a2 3 Gal3 a1 4 Glc4	1
Neu5Ac1 a2 8 Neu5Ac2 a2 8 Neu5Ac3 a2 8 Neu5Ac4	1
Neu5Ac1 a2 8 Neu5Ac2 a2 3 Gal3 b1 4 Glc5 NGA4* b1 4 X	1

Neu5Ac5 a2 8 Neu5Ac1
Neu5Ac5 a2 3 Gal3
Neu5Ac1

glycan id	pdb title
1v0f_04	Endosialidase of bacteriophage k1f in complex with oligomeric alpha-2,8-sialic acid
3gvk_06	Crystal structure of endo-neuraminidase nf mutant
3gvk_09	Crystal structure of endo-neuraminidase nf mutant
3hmy_01	Crystal structure of hcr/t complexed with gt2

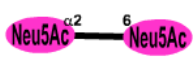
Glycan Structure Search

Create Sugar

Glc Gal
Man Fuc
GlcNAc GalNAc
Neu5Ac Xyl
GlcA GlcN

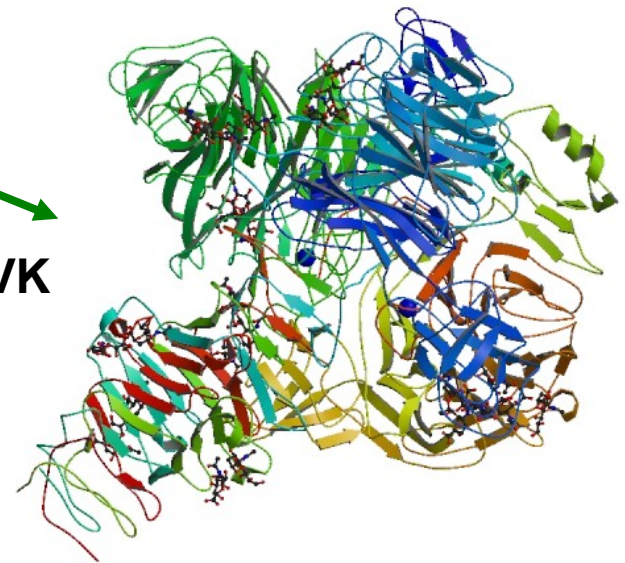
Create Bond

Remove
Clear all
Search



Glycan Structure Search

3GVK



Database / Search / PDB data

Search for carbohydrate containing PDB entries by criteria like species or the compound / classification terms. You can choose predefined, frequent terms from the pull-down-menus or enter your own queries manually. For selection from the pull-down-menus, java script must be activated in your browser to copy the selected value to the text field below.

The wildcards * (matches anything) and ? (matches any single character) can be used in "Source" and "Compound / Classification" queries.

To search for PDB entries by carbohydrate (sub-)structure, use the *structure search* in the **beginner mode** or the **advanced mode**.

Substructure / Search / Beginner

Click [here](#) to reset input.

a-D-Galp

1-3

a-D-Glcp

1-3

a-D-GalpNAc

1-1

with 3D-Co-ordinates (Sweet2) | with NMR data | max # residues

with PDB entries | min. resolution | all chains

species human

(use NCBI-Taxon-ID e.g. 9606 or names e.g. hum

Search Glycosciences

Search in BCSDb

Search now

3D STRUKTURA SACHARIDŮ

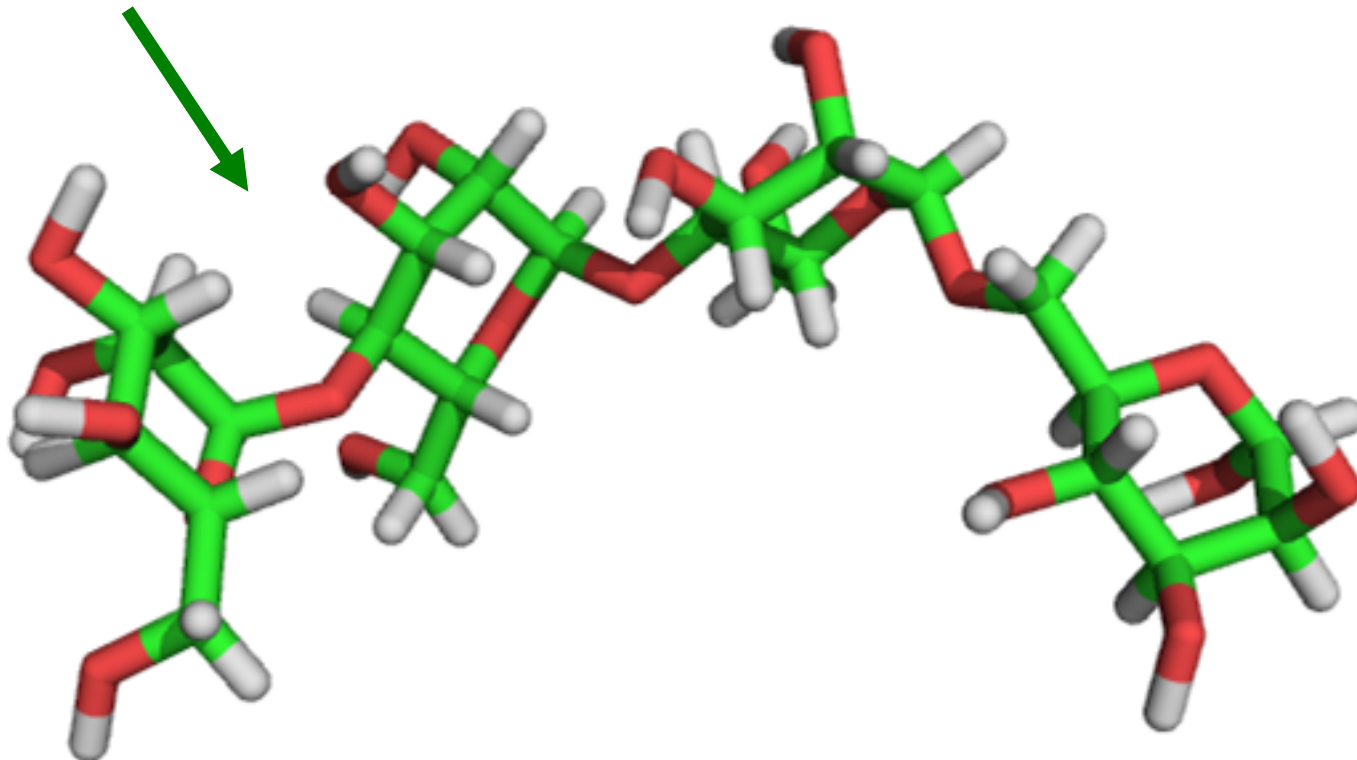
- Určení struktury **komplexních** sacharidů je obecně problémem.
RTG – velká flexibilita, problémy s krystalizací.
NMR – tradiční metoda pro určení struktury oligosacharidů (práce v roztoku), problémy s přiřazením signálů a vyhodnocením dat (malé rozdíly mezi jednotlivými jádry).
- Molekulové modelování sacharidů je často **nezbytnou** součástí interpretace experimentálních dat.

Sweet is a program for constructing 3D models of saccharides from their sequences using standard nomenclature.



<http://www.glycosciences.de/modeling/sweet2/doc/index.php>

Remember - not all constructions are reasonable.



PREDIKCE GLYKOSYLACE

DictyOGlyc 1.1

The DictyOGlyc server produces neural network predictions for GlcNAc O-glycosylation sites in *Dictyostelium discoideum* proteins.

NetCGlyc 1.0 Server

NetCGlyc 1.0 produces neural network predictions of C-mannosylation sites in mammalian proteins.

NetGlycate 1.0 Server

NetGlycate 1.0 server predicts glycation of ϵ amino groups of lysines in mammalian proteins.

NetNGlyc 1.0 Server

The NetNglyc server predicts N-Glycosylation sites in human proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequons.

NetOGlyc 4.0 Server

The NetOglyc server produces neural network predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins.

CHCETE VĚDĚT VÍC?

Cell. Mol. Life Sci. (2010) 67:2749–2772
DOI 10.1007/s00018-010-0352-4

Cellular and Molecular Life Sciences

REVIEW

Bioinformatics and molecular modeling in glycobiology

Martin Frank · Siegfried Schloissnig

Received: 23 December 2009 / Revised: 8 March 2010
© The Author(s) 2010. This article is published online

Abstract The field of glycobiology is concerned with the study of the structures, properties, and biological functions of the family of biomolecules called carbohydrates. Bioinformatics for glycobiology is a particularly active field, because carbohydrates exhibit a high diversity and their chains are often branched. In recent years, improvements in experimental analytical techniques and the recent years have led to a tremendous amount of carbohydrate structure data generated. Consequently, the availability of databases and tools to retrieve and analyze these data in an efficient manner has become of fundamental importance to progress in glycobiology.

Bioinformatics for glycobiology

2751

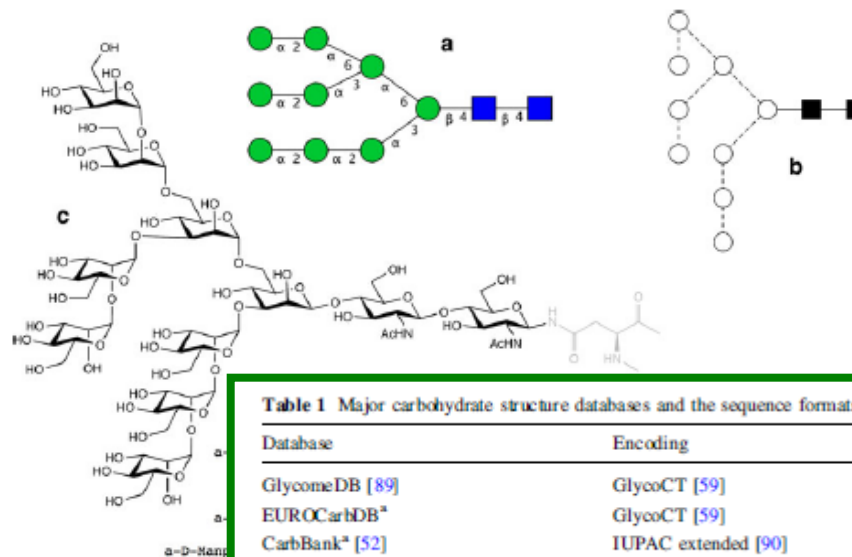


Fig. 2 Different graphical representations of a branched oligosaccharide. **a** Schematic representation, **b** Extended IUPAC

Therefore, derivatization is for most cases rather than an exception. The different forms of some monosaccharides occur naturally in their L-form. Additionally, each of the

Table 1 Major carbohydrate structure databases and the sequence formats used

Database	Encoding	URL
GlycomeDB [89]	GlycoCT [59]	http://www.glycome-db.org/
EUROCarbDB ^a	GlycoCT [59]	http://www.ebi.ac.uk/eurocarb/
CarbBank ^a [52]	IUPAC extended [90]	http://www.boc.chem.au.nl/sugar/
KEGG ^a [83]	KCF [91]	http://www.genome.jp/kegg/glyco/
GLYCOSCIENCES.de ^a [82]	LINUCS [92]	http://www.glycosciences.de/
CFG ^a [84]	Glycominds Linear Code [®] [47]	http://www.functionalglycomics.org/
BCSDB ^a [93]	BCSDB linear code	http://www.glyco.ac.ru/bcsdb3/
GlycoSuiteDB [87]	IUPAC condensed [94]	http://glycosuite.db.expasy.org/
GlycoBase (Dublin) ^a [86]	Motif based	http://glycobase.ucd.ie/
GlycoBase (Lille) ^a [95]	Linkage path	http://glycobase.univ-lille1.fr/ba