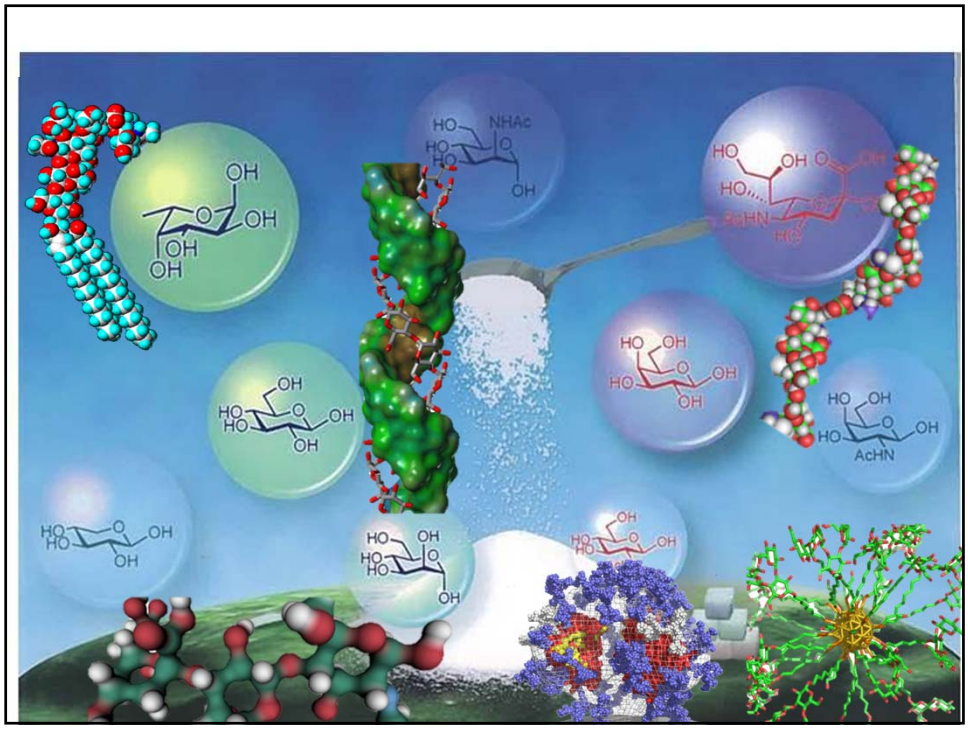
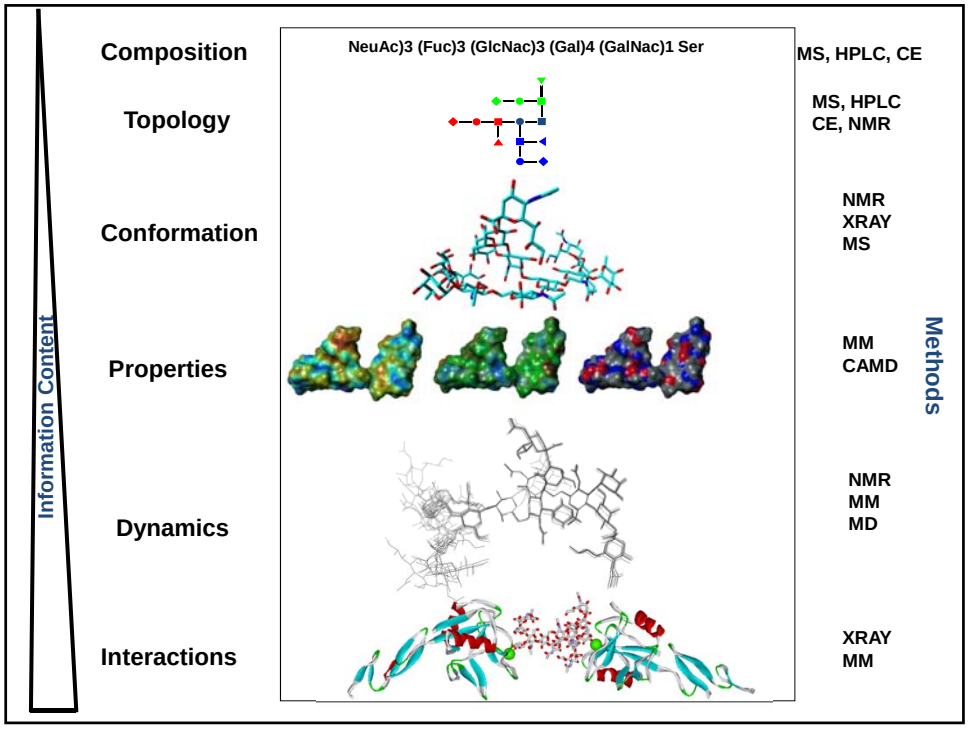
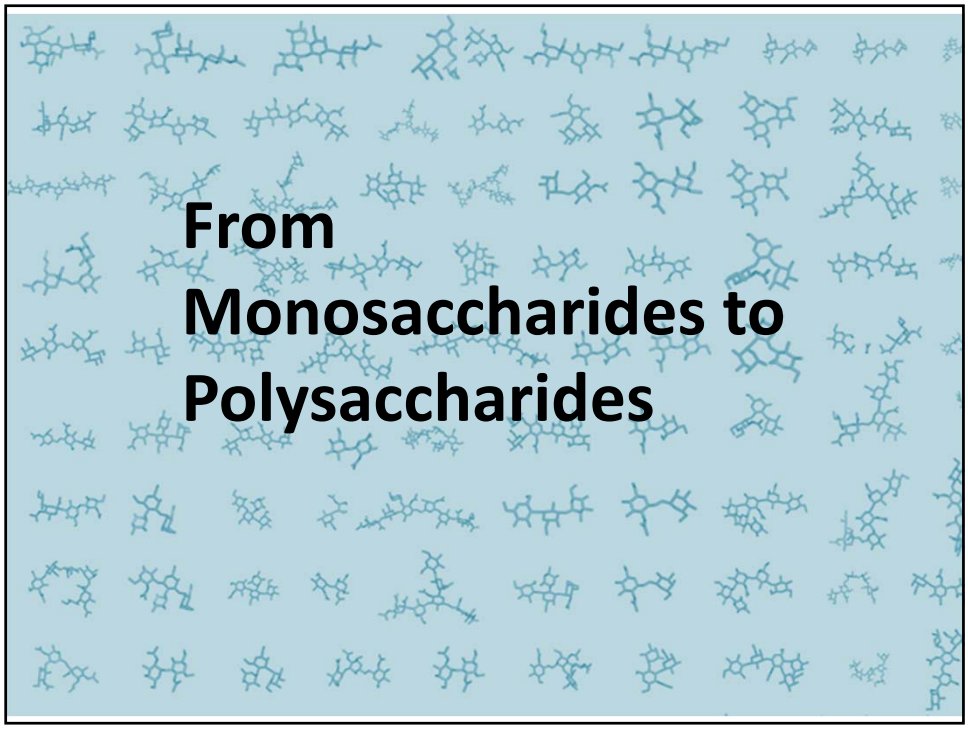
From Monosaccharides to Polysaccharides From Structures to 3D Databases

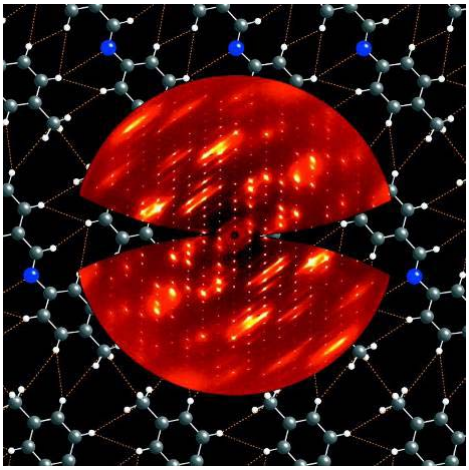
Serge Pérez, Grenoble, June 2016

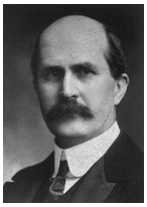



Information Content
Composition
Topology
Conformation
Properties
Dynamics
Interactions
NeuAc)3 (Fuc)3 (GlcNac)3 (Gal)4 (GalNac)1 Ser
MS, HPLC, CE
MS, HPLC
CE, NMR
NMR
XRAY
MS
MM
CAMD
NMR
MM
MD
XRAY
MM
Methods


From
Monosaccharides to
Polysaccharides

International Year of Crystallography



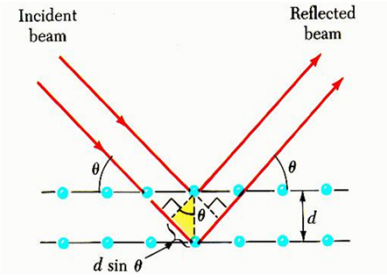


W. L. Bragg

W. L. Bragg

The International Year of Crystallography 2014 (IYCr2014) commemorated the centennial of X-ray diffraction, which allowed the detailed study of crystalline material.

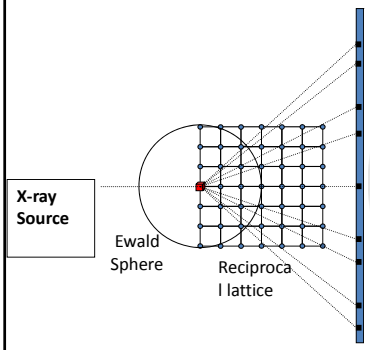
It also commemorated the 400th anniversary of Kepler's observation in 1611 of the symmetrical form of ice crystals, which began the wider study of the role of symmetry in matter.

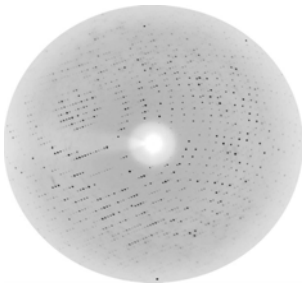


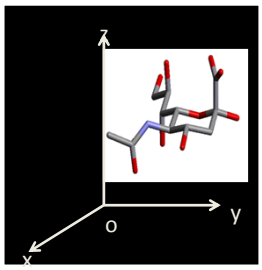
X-ray interact with the spatial distribution of Valence electrons.

Neutrons are scattered by the atom nucei.

Electrons feel the influence of both the positively charged atomic nuclei and the surrounding electrons.







N atoms : 3N observables

Crystallography of Carbohydrates

Molecular & Crystal Structures of Carbohydrates

- •Experimental Conditions and Limitations (X and N)
- •Crystalline Conformations of Oligosaccharides
- •Hydrogen Bonding in Crystalline Oligosaccharides
- •Packing Features
- •Powder Diffraction

Crystalline Conformations of Oligosaccharides in Proteins

- •Experimental Conditions and Limitations
- •Oligosaccharides –Lectin Complexes
- •Glycosaminoglycan-Protein Complexes

Crystalline Conformations of Polysaccharides

- •Experimental Conditions and Limitations
- •X-Ray Fiber Diffraction of Polysaccharides
- •X-Ray Fiber Diffraction using Synchrotron and Neutron Radiations
- •Electron Diffraction of Polysaccharides

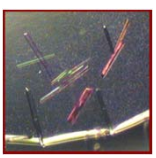
Molecular & Crystal Structures of Carbohydrates

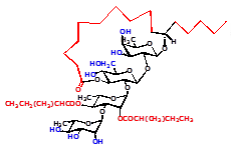
Experimental Conditions and Limitations (X and N)

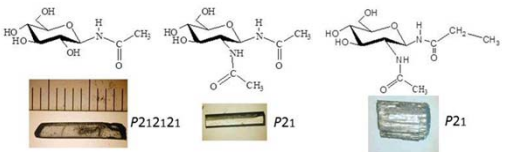
X-ray and Neutron have wavelengths in the same order as the interatomic distances (Angstrom).
Electron are the scattering elements of the incident X-ray
Nuclei are the scattering elements of the incident Neutron radiation

Single crystals usually grown by slow evaporation of saturated solution under well controlled environments.

X-ray: Dimensions 0.2 – 0.5 mm / Synchrotron X-ray : 20-30 µm
Neutron: Dimensions over 1.0 mm all dimensions







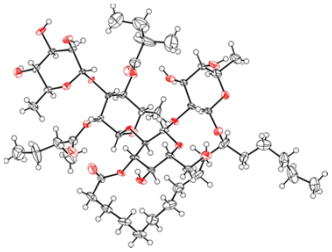
Molecular & Crystal Structures of Carbohydrates

Crystalline Conformations of Oligosaccharides

Cambridge Structural Data Base (CSDB) ⍰ 4000 entries

Unsubsitituted disaccharides ⍰ 60 structures
Unsubsitituted trisaccharides ⍰ 30 structures
Unsubsitituted tetraccharides < 5 structures

Cyclodextrins & cyclic oligoamyloses : > 300 structures



Difficulty to crystallize oligosaccharides having molecular weight 1000 to 5000

Understanding a Structural Report

Unit Cell Parameters (a, b, c, a, b, g); Space Group

Fractional atomic coordinates content of the asymmetric unit: (x/a; y/b; z/c)
Anisotropic Temperature Factors (ORTEP representation ellipsoids)

Bond distances (esds), Bond angles (esds), Torsion angles (esds)
Geometry and conformation of the molecule

Configuration !!!!!

Intra- and Inter molecular hydrogen bonds

Analysis of
Hydration features
Packing features

Emil Fisher

Johannes Bijvoet

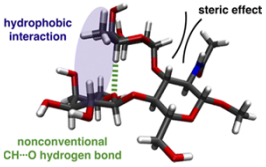
Molecular & Crystal Structures of Carbohydrates

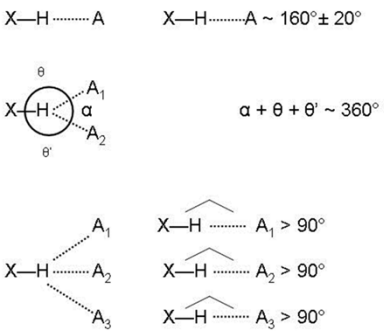
Hydrogen Bonding in Crystalline Oligosaccharides

Analysis of high accurate X-ray analysis – Neutron diffraction

dX-dN =(C-H) = -0.096(7)
dX-dN =(O-H) = -0.155(10)

CHO bonds



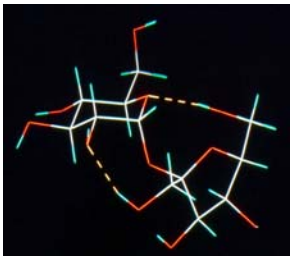


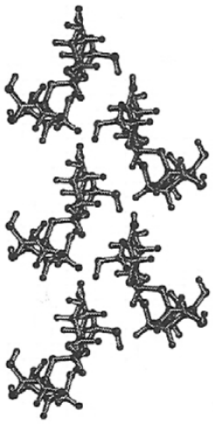
Maximize the Hydrogen Bond interactions throughout the participation of all hydroxyl groups and as many rings oxygen. Two and three-centered bonds

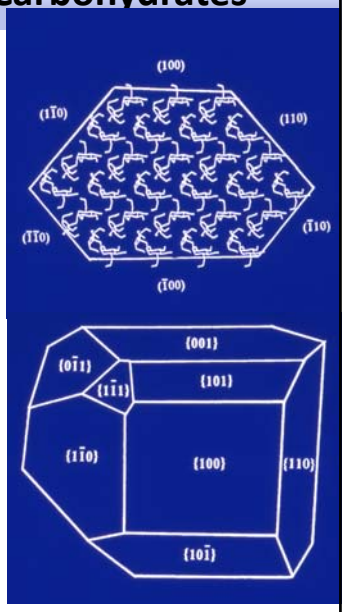
Maximize cooperativity by forming as many finite and infinite chains of hydrogen bonds as possible.

Molecular & Crystal Structures of Carbohydrates

Packing Features





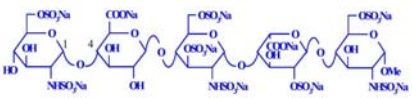


Molecular & Crystal Structures of Carbohydrates

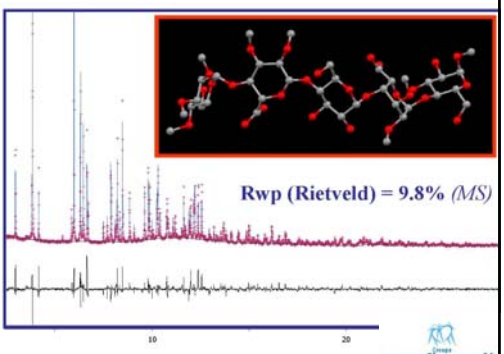
Powder Diffraction

1. Identification of Crystalline Polymorphs

2. Solving Crystal Structures – Rievelt Method + Molecular Modelling



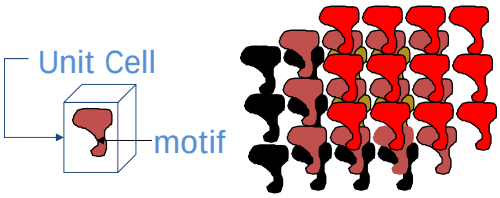
Synthetic Pentasaccharide
ID31@ESRF, **λ =0.8 A**
Monoclinic $P2_1$
a=15.54, b=8.83; c=17.67, β=94.6

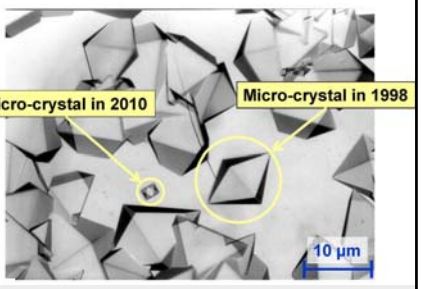


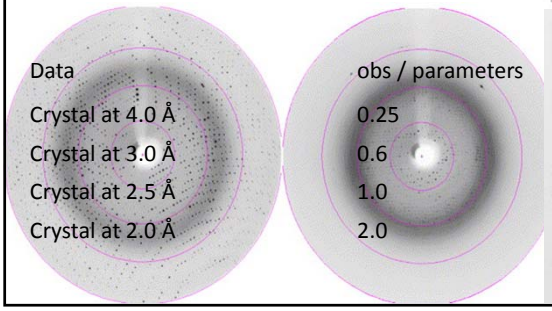
Ph. Ochsenbein, J. Kieffer & M. El Hajji 12th European Powder Diffraction Conference, 2010, Darmstatd

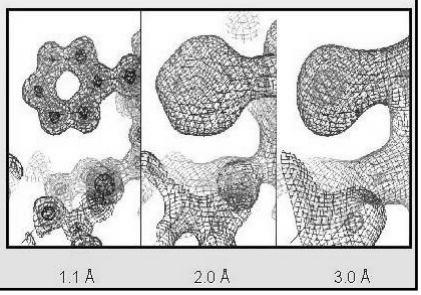
Crystalline Conformations of Oligosaccharides in Proteins

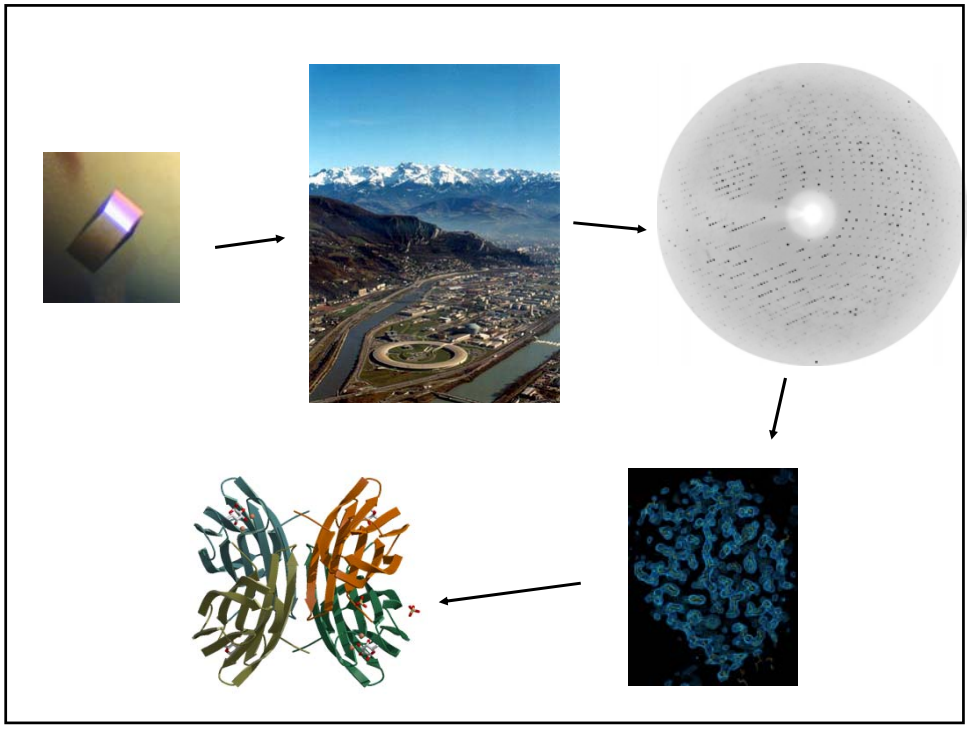
Experimental Conditions and Limitations

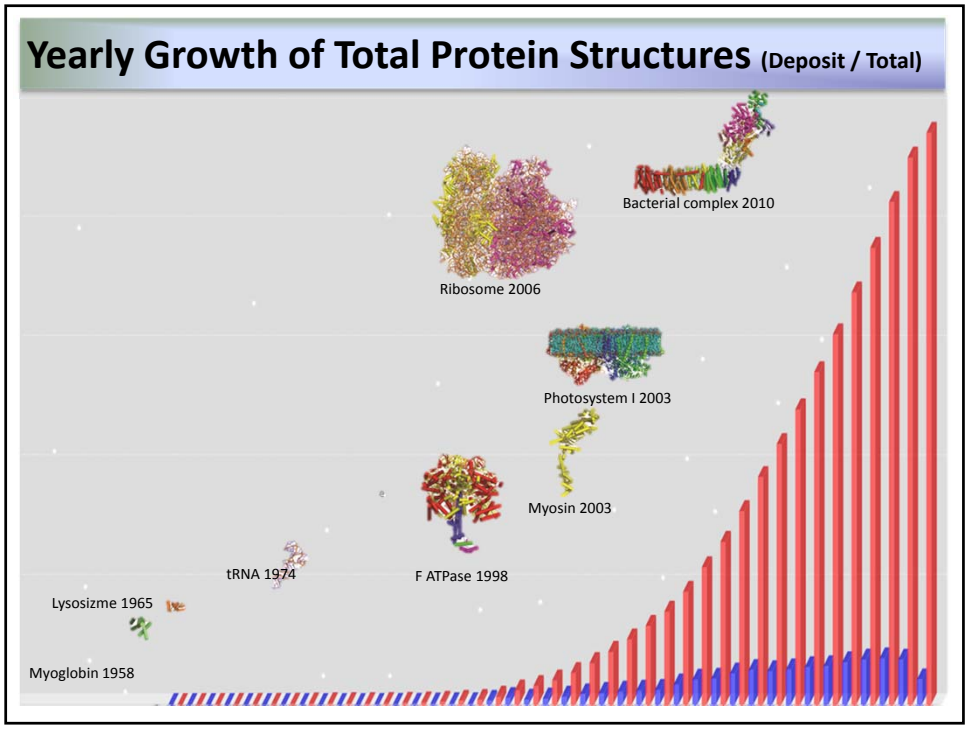






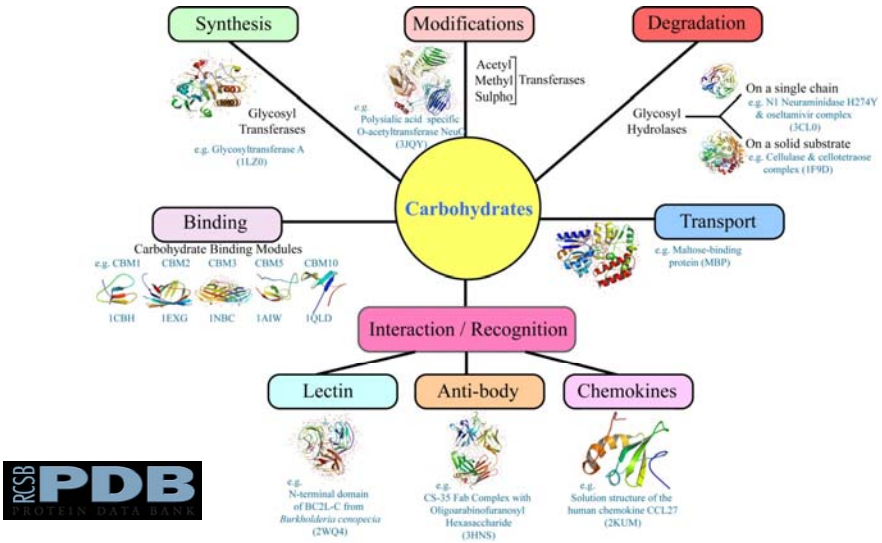





Yearly Growth of Total Protein Structures (Deposit / Total)
Bacterial complex 2010
Ribosome 2006
Photosystem I 2003
Myosin 2003
F ATPase 1998
tRNA 1974
Lysosizme 1965
Myoglobin 1958

Crystalline Conformations of Oligosaccharides in Proteins

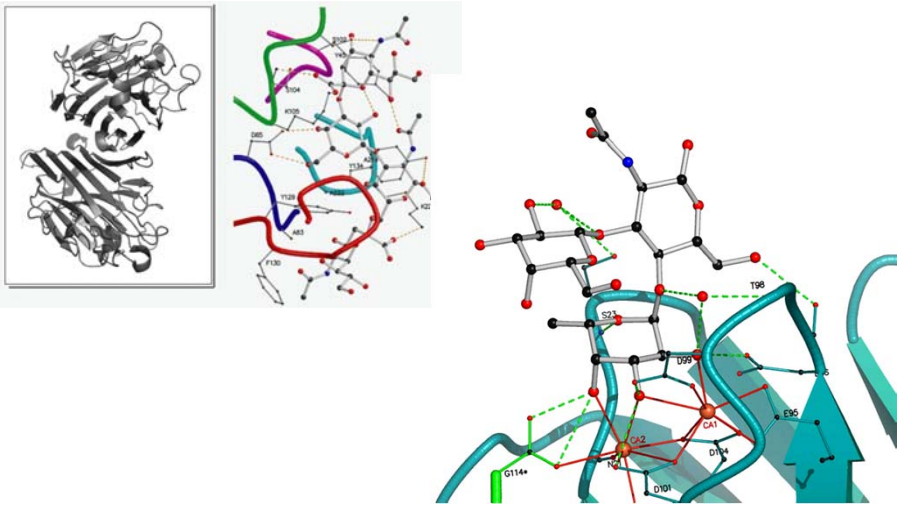
Protein - Carbohydrate Interactions



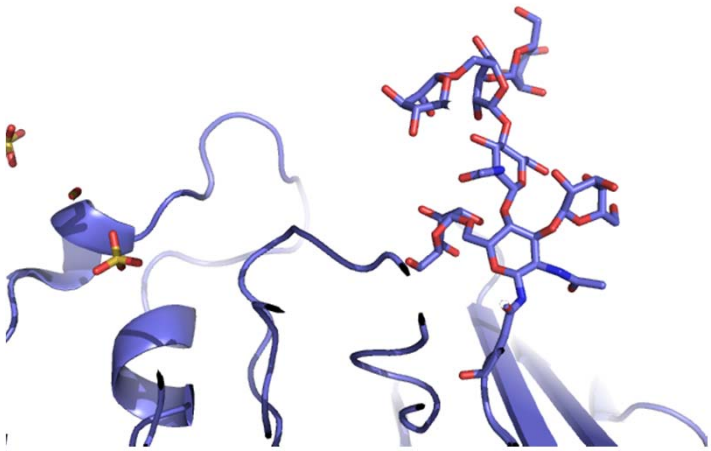
Protein Data Bank : http://www.rcsb.org/pdb/home/home.do

Crystalline Conformations of Oligosaccharides in Proteins

Oligosaccharides –Lectin Complexes



Crystalline Conformations of Oligosaccharides in Proteins



N-linked glycan:

NeuAcα2-6Galβ1-4GlcNAcβ1-2Manα1

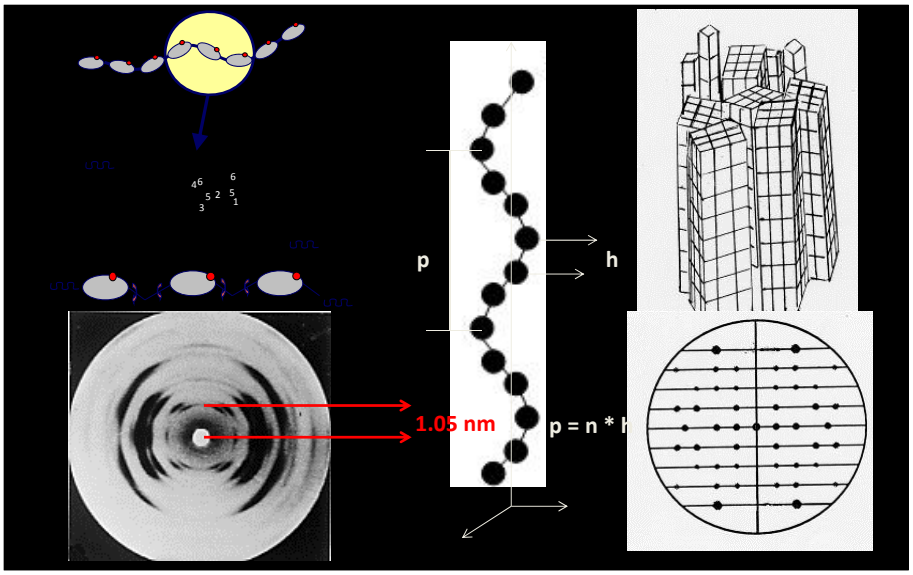
NeuAcα2-6Galβ1-4GlcNAcβ1-2Manα1

Manβ1-4GlcNAcβ1-4GlcNAc-Asn

Fucα1

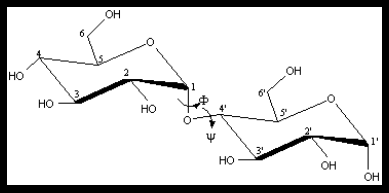
Crystalline Conformations of Polysaccharides

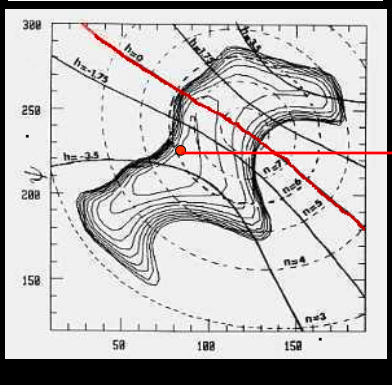
X-Ray Fiber Diffraction of Polysaccharides



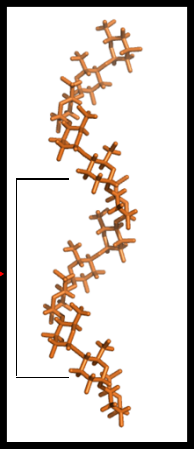
Crystalline Conformations of Polysaccharides

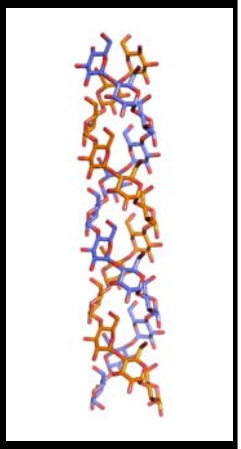
X-Ray Fiber Diffraction of Polysaccharides





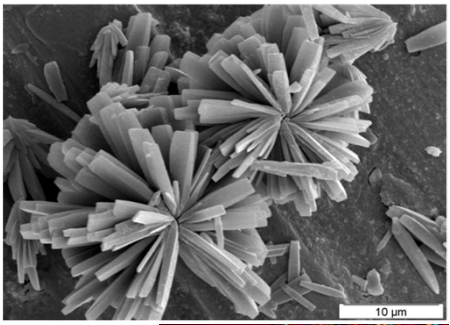
2.1 nm

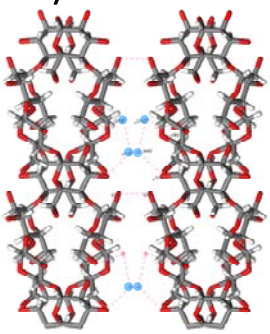


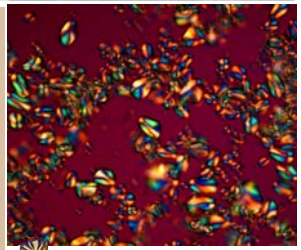


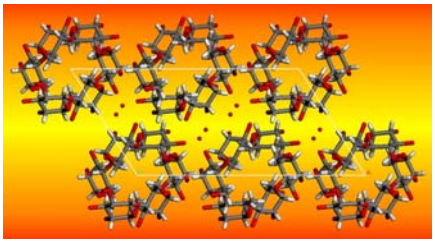
Crystalline Conformations of Polysaccharides

Synchrotron X-Ray Diffraction of Polysaccharides





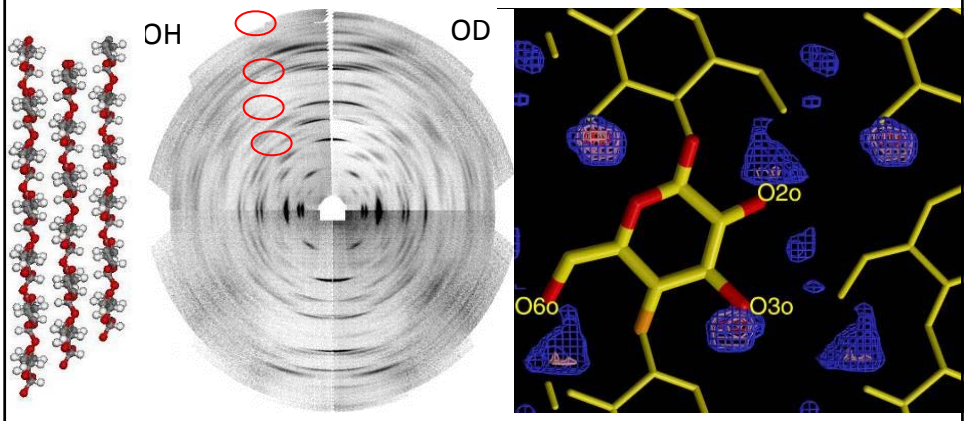




D. Popov, 2009, Macromolecules, 42, 1167-1174

Crystalline Conformations of Polysaccharides

X-Ray Fiber Diffraction using Synchrotron and Neutron Radiations



Y. Nishiyama et al., 2002, J. Am. Chem. Soc., 124, 9074-9082

Crystalline Conformations of Polysaccharides

Electron Diffraction of Polysaccharides

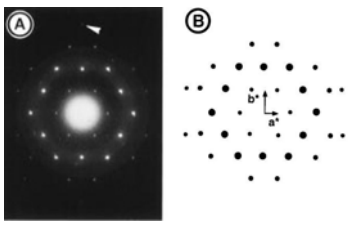
Electrons are charged particles and interact with matter through the Coulomb forces. The incident electrons feel the influence of both the positively charged atomic nuclei and the surrounding electrons.

Electron diffraction of solids is usually performed in a **Transmission Electron Microscope** (TEM) where the electrons pass through a thin film of the material to be studied. The resulting diffraction pattern is then observed on a fluorescent screen, recorded on photographic film, on imaging plates or using a CCD camera.

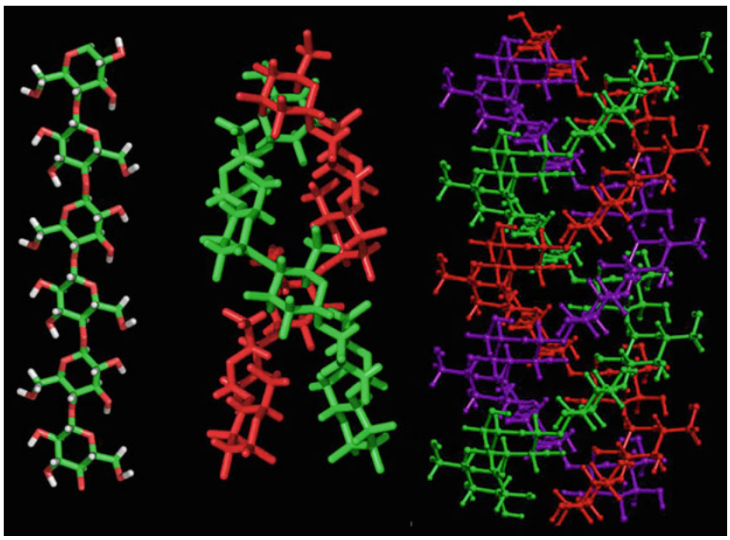
Electron diffraction in TEM is subject to several important limitations.

The sample to be studied must be electron transparent, meaning the sample thickness must be of the order of 100 nm or less.

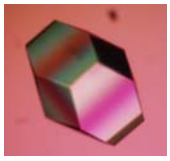
- Careful and time consuming sample preparation are needed.
- Many samples are vulnerable to radiation damage caused by the incident electrons.

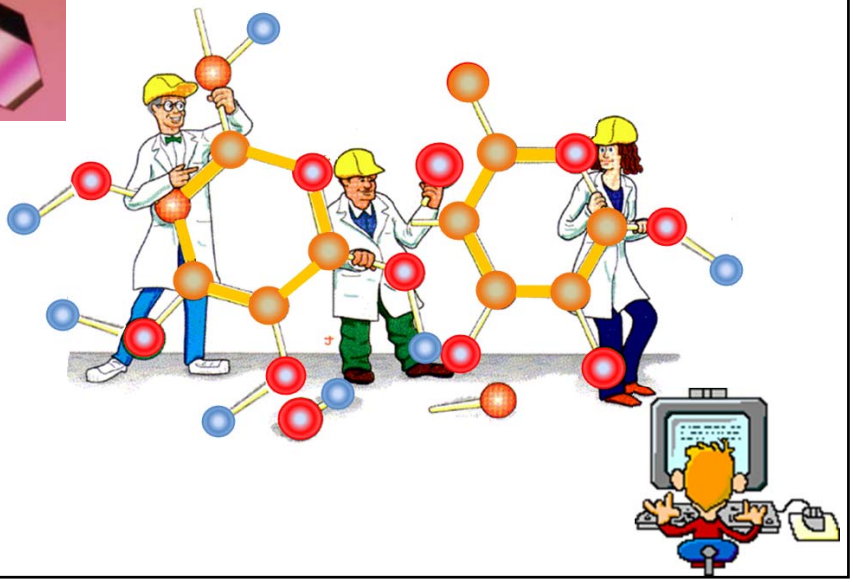


Helical Structures of Polysaccharides

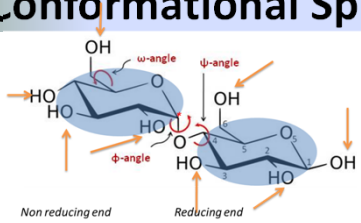


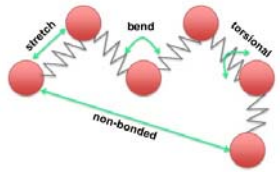
From *in cubo* to *in silico*

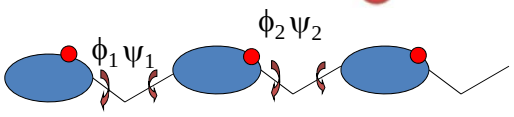




Conformational Space of Oligosaccharides







Combinatorial building

Assumption:
Because of the bulky and (almost) rigid nature of the monosaccharide unit, the conformation of each linkage is independent on the other

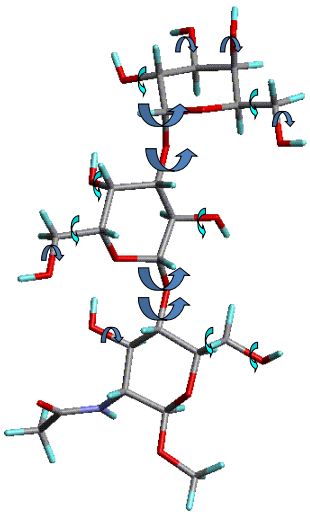
Methods :
Combine the lowest energy minima of each disaccharide map

Not true for
- long range interactions
- branched structures
….

But very useful for building starting structures!

Conformational Space of Flexible Oligosaccharides

Systematic search of all possible conformations ?



For a trisaccharide:

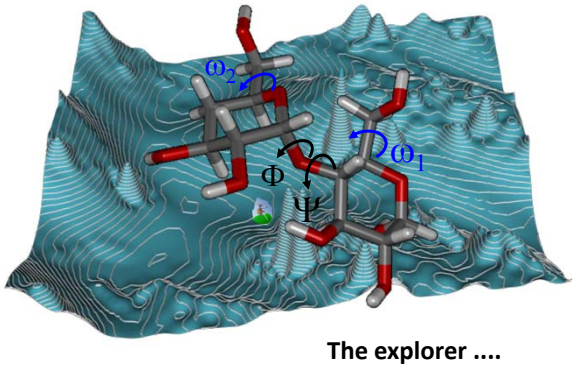
4 torsions to be searched
with 20° steps
18^4 conformations

12 pendant groups
Staggered orientations:
3^{12} combinations

$$> 5\ 10^{10}$$

αGal(1-3)βGal(1-4)βGlcNAc

Conformational Space of Flexible Oligosaccharides

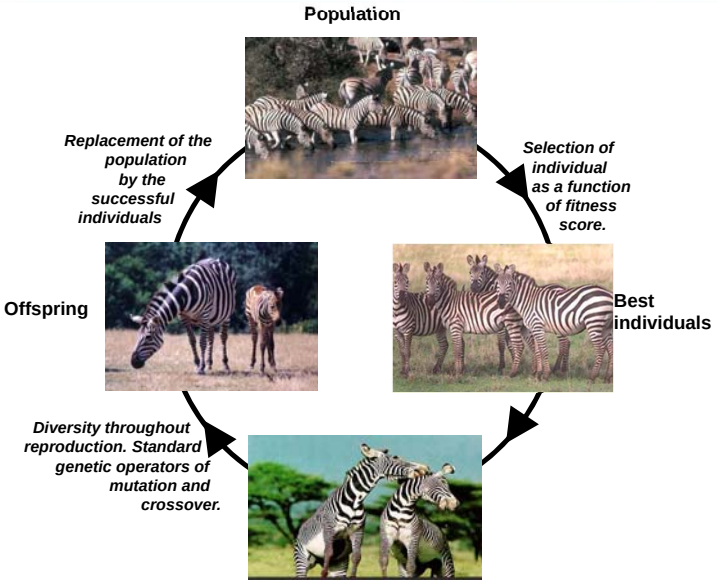


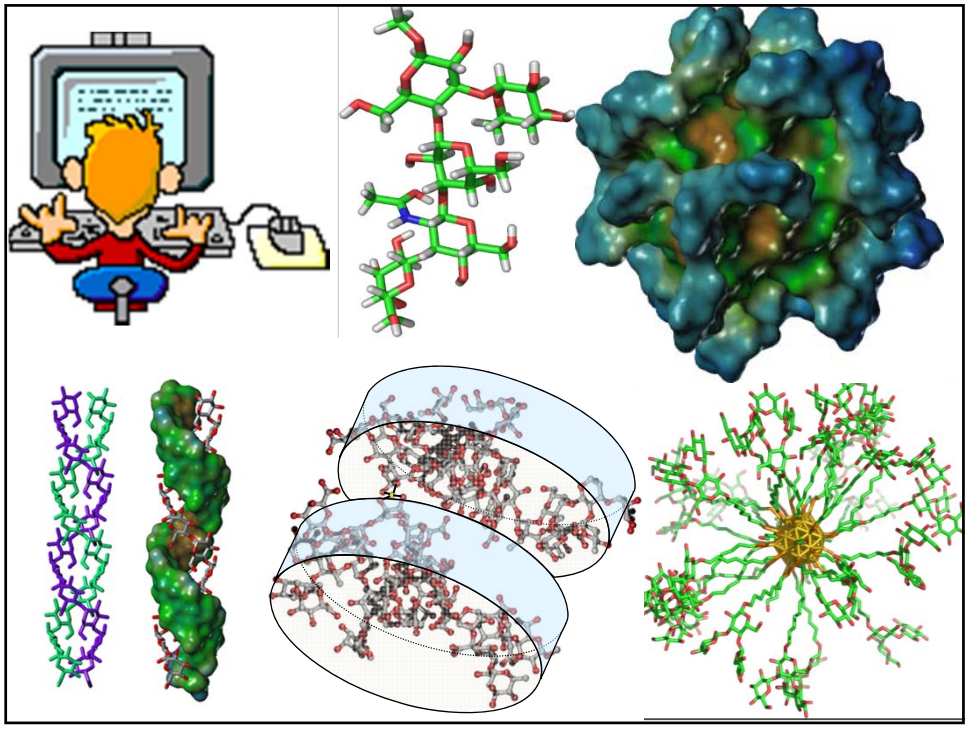
The explorer ....

Does not “see” the location of other minima.

He just knows if he goes “up” or “down”.

Monte Carlo Calculations



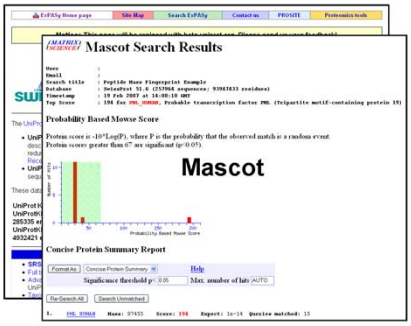


From Structures to 3D Databases

Glycoinformatics

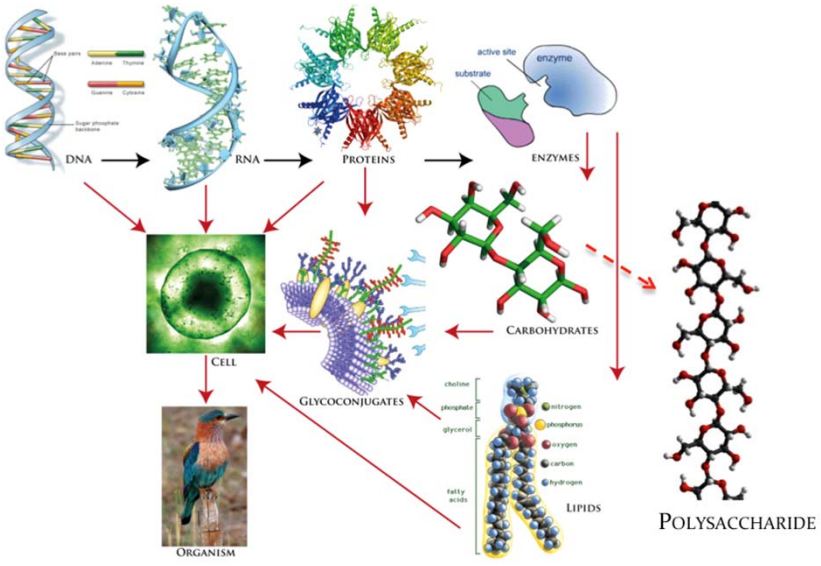
Genomics

Proteomics



Glycomics

Carbohydrates in the Scheme of the Central Dogma of Life



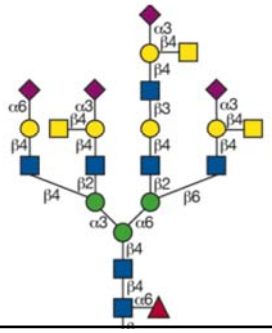
Challenges for Glycoinformatics
Structures as Primary Access Key

Bioinformatic

Galectin-1
source organism="Homo sapiens"
gene gene="LGALS1"
Site /site_type="binding"
/note="Beta-galactoside (Potential)."
1 MACGLVASNL NLKPGECLRV RGEVAPDAKS
31 FVLNLGKDSN NLCLHFNPRF NAHGDANTIV
61 CNSKDGGAWG TEQREAVFPF QPGSVAEVCI
91 TFDQANLTVK LPDGYEFKFP NRLNLEAINY
121 MAADGDFKIK CVAFD

Sequences of residues

Glycoinformatic



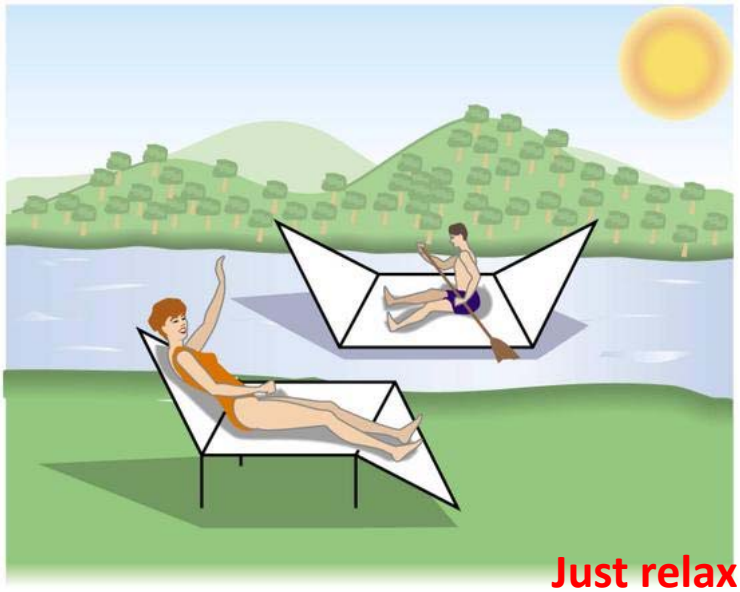
Topology of Residues

Symbol Nomenclature for Graphical Representation of Glycans (2015), *Glycobiology*, **25,** 1323-1324

Hexose	Glc	Man	Gal	Gul	Alt	All	Tal	Ido	
HexNAc	GlcNAc	ManNAc	GalNAc	GulNAc	AltNAc	AllNAc	TalNAc	IdoNAc	
Hexosamine	GlcN	ManN	GalN	GulN	AltN	AllN	TalN	IdoN	
Hexuronate	GlcA	ManA	GalA	GulA	AltA	AltA	TalA	IdoA	
DeoxyHexose	Qui	Rha			6dAltA		6dTal		Fuc
Deoxy HexNAc	QuiNAc	RhaNAc							FucNAc
Dideoxy Hexose	Oli	Tyv		Abe	Par	Dig	Col		
Pentose		Ara	Lyx	Xyl	Rib				
Nonulosonate		Kdn				Neu5Ac	Neu5GC	Neu	
Assigned (I)	Bac	ManHep LD	Kdo	Dha	ManHep DD	MurNAc	MurNGc	Mur	
Assigned (II)	Api	Fru	Tag	Sor	Psi				

A. VARKI, R.D. CUMMINGS, M. AEBI, N.H. PARKER, P.H. SEEBERGER, J.D. ESKO, P. STANLEY, G. HART, A. DARVILL, T. KINOSHITA, J.J. PRESTEGARD, R.L. SCHNAAR, H.H. FREEZE, J.D. MARTH, C.R. BERTOZZI, M.E. ETZLER, M. FRANK, J.F.G. VLIEGENTHART, T. LUTTEKE, S. PEREZ, E. BOLTON, P. RUDD, J. PAULSON, M. KANEHISA, P. TOUKACH, K.F. AOKI-KINOSHITA, A. DELL, H. NARIMATSU, W. YORK, N. TANIGUCHI & S. KORNFELD,

No Comments …….



Just relax …….

37

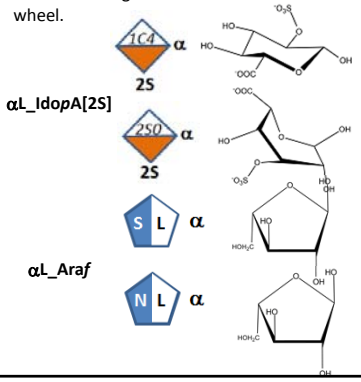
Extending the Symbolic Representation of Monosaccharides

Residue Letter Name: Rib, Ara, Xyl, Lyx, All, Alt, Glc, Man, Gul, Ido, Gal, Tal,….
[O-ester and ethers]: (when present) are shown attached to the symbol with a number, e.g.
6Ac for 6-*O*-acetyl group, 3S for 3-*O*-sulfate group
6P for 6-*O*-phosphate group, 6Me for 6-*O*-Methyl group
36Anh for 3,6-anhydro, Pyr for pyruvate group

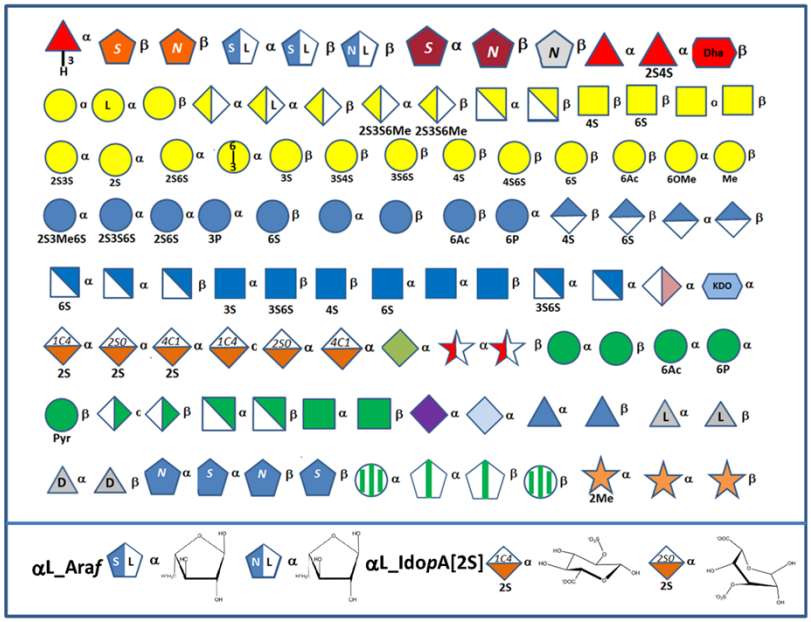
Absolute Configuration: D or L
The D-configuration for monosaccharide and the L configuration for Fucose and Idose are implicit and does not appear in the symbol. Otherwise the L configuration, is indicated in the symbol, as in the case of Arabinose or L-Galactose.
For those occurring in the furanose form, a letter *N* or *S* is inserted in the symbol, indicating the northern *(N)* or Southern (*S*) conformation of the five membered ring.

Anomeric Configuration.
The nature of the glycosidic configuration (α or β) is explicitly set within the symbol.

Ring Conformation.
All pyranoses in the D-configuration are assumed to have $^4C_1$ chair conformation; those in the L configuration are assumed to have $^1C_4$ chair conformation. Otherwise, the ring conformation is indicated in the symbol, as $^2S_0$ in the case of α-L-Idopyranose.
N or *S* indicates the conformation of the five membered rings on the conformational wheel.



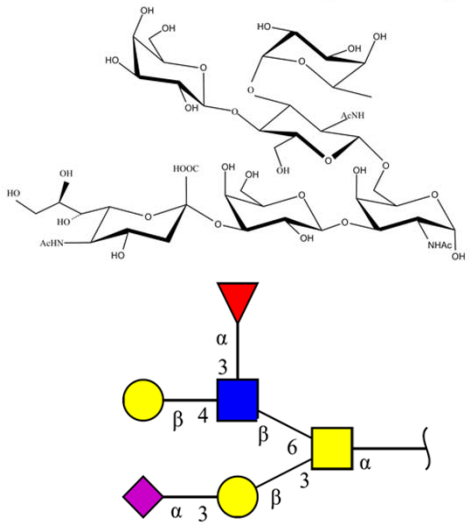
More than 150 Monosaccharides



Encoding of Glycan Structures

Lewis X and Sialyl Acid on Core 2

Neu5Ac a2-3 Gal b1-3 (Gal b1-4 (Fuc a1-3) GlcNAc b1-6) GalNAc

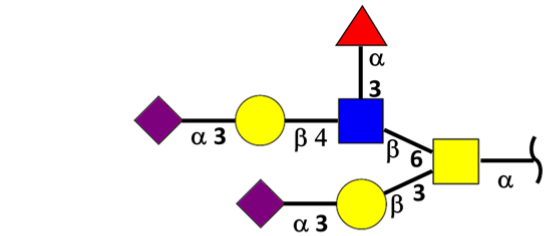


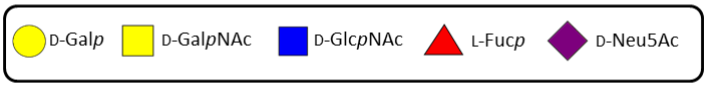
```
RES
1b:a-dgal-HEX-1:5
2s:n-acetyl
3b:b-dgal-HEX-1:5
4b:a-dgro-dgal-NON-2:6|1:a|2:keto|3:d
5s:n-acetyl
6b:b-dglc-HEX-1:5
7s:n-acetyl
8b:a-lgal-HEX-1:5|6:d
9b:b-dgal-HEX-1:5
LIN
1:1d(2+1)2n
2:1o(3+3)3d
3:3o(3+2)4d
4:4d(5+1)5n
5:1o(6+1)6d
6:6d(2+1)7n
7:6o(3+1)8d
8:6o(4+1)9d
```

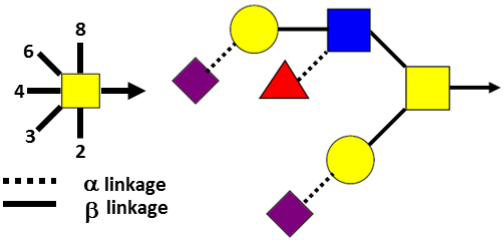
GlycoCT

Needs for Standardizations

Neu5Ac α2-3 Galp β1-3 (Neu5Ac α2-3 Galp β1-4 (Fucp α1-3) GlcpNAc β1-6) GalpNAc



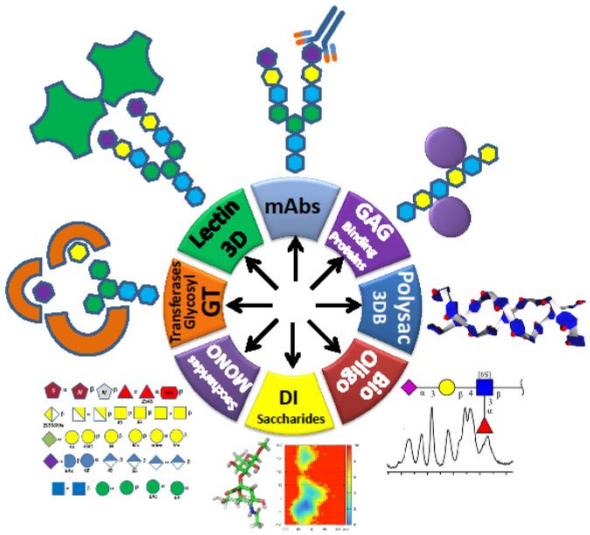




Major Glycan Structure Formats

- **IUAPC**
  (condensed / extended)
- **LINUCS**
- **CarBank**
- **BCSDB linear**
- **KCF**
- **Linear Code**
- **GlycoCT**
- **GLYDE-II**
- **WURCS**

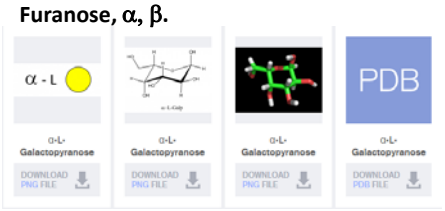
Glyco3D



Monosaccharides

Glycopedia

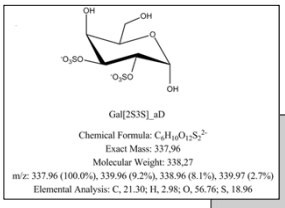
The Templates: (128 entries)
Hexoses, pentoses, ketoses, D, L, pyranose
Furanose, α, β.

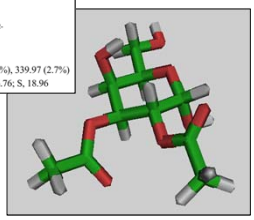


The Bioactive units: (150 entries)
Components of oligo, polysaccharides
glycans, conjugates.

Glyco3D

Molecule Information
Sequence, Family
Configuration/Conformation
Chemical representation
Formula
Exact mass (OH / OMe)
m/z, Elemental analysis

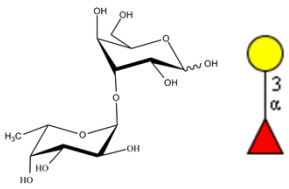




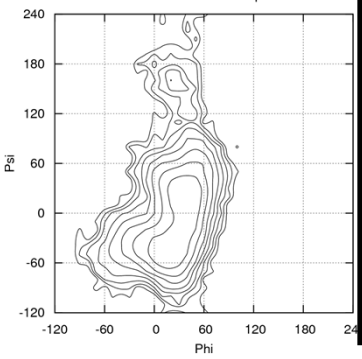
Disaccharides

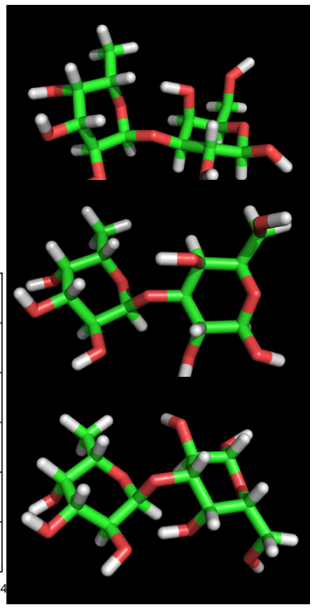
Source: Molecules or
Buidling blocks of
« glycan determinants »
Content: 150 entries
Method:
Molecular Mechanics
(MM3 vacuum)
Search: Sequence, MW.
Molecule Info.
Sequence
Family
Configuration/
Conformation
Chemical representation
Formula, Exact mass, m/z
Elemental analysis
Display & Download
3D Structure (Jmol Applet)
up to 3 low energy conf.
Download PDB Files

Fuc α1-3 Gal



Conformational Map

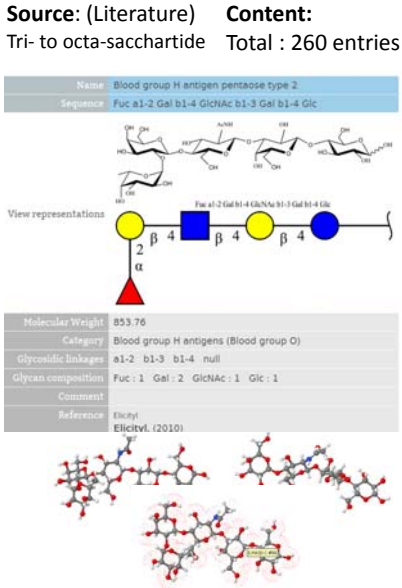


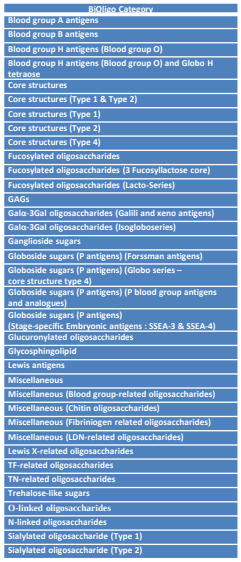


Bio-Oligosaccharides : 3D / NMR

Source: (Literature)
Tri- to octa-sacchartide

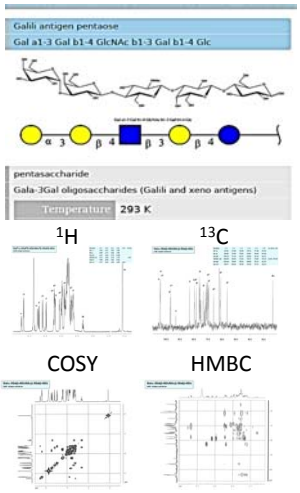
Content:
Total : 260 entries





Source:
Glycan
(bacterial fermentation)

Content:
Total : 150 entries



A. Sarkar, S. Drouillard, A. Rivet & S. Perez (2015) Databases of Conformations and NMR Structures of Glycan Determinants

Curations

Annotations

Lectins

Source: X-ray - PDB

Classification of Lectins
based on their origin:
Algea, Animal, bacteria, fungi & yeast, plant, virus,.

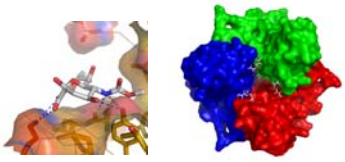
Content:
Total : 1186
Complexed sugar: 748
Free Lectins: 438
Origin : 6
Classes: 56

Search:
Species
Family
Sugar
PDB

Molecule Information
Origin
Class
Family
Species
View representation

PDB Code
Resolution
Comments
Reference
Links (Medline, PDB, Taxonomy)

Display & Download
3D Structure (Jmol Applet)
Download PDB File
Still Image
Download Image



Glycosyl Transferases

Source: X-ray – PDB, NMR

Content:
Total : 375

Classification of the GTs
based on their origin:
Animal, archea, bacteria, plant, virus, yeast & fungi

Sub-classification based
either on the function,
or the fold, i.e. GT-A, GT-B & GT-alike.
GTs are numbered according to the CAZY classification

Search: family
PDB
Authors
Fold
Resulting linkage
Enzyme name
Abbreviation

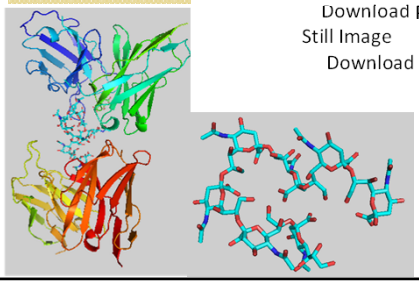
Molecule Information
Enzyme name
Short name
Origin
Organism
Resulting linkage
Fold
Cazy Fmily
Mechanism
PDB Code
Resolution
Complexed with
Comments
Sequence
Reference
Links (Medline, PDB, Swiss Prot, CAZY)

Display & Download
3D Structure (Jmol Applet)
Download PDB File
Still Image
Download Image

Monoclonal Antibodies / GAG Binding Proteins

Source: X-ray - PDB

Content:
Total : 40
Classification of mAbs
Human
Murine
Synthetic
Search:
Family
Sugar
PDB

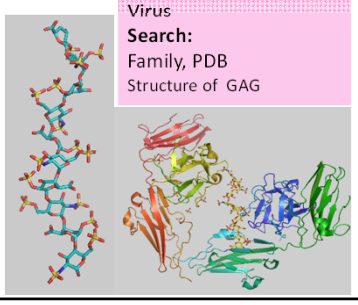


Molecule Information
Class
Family
Species
View representation
PDB Code
Resolution
Sequence
Reference
Links (Medline, PDB, Swiss Prot.)

Display & Download
3D Structure (Jmol Applet)
Download PDB File
Still Image
Download Image

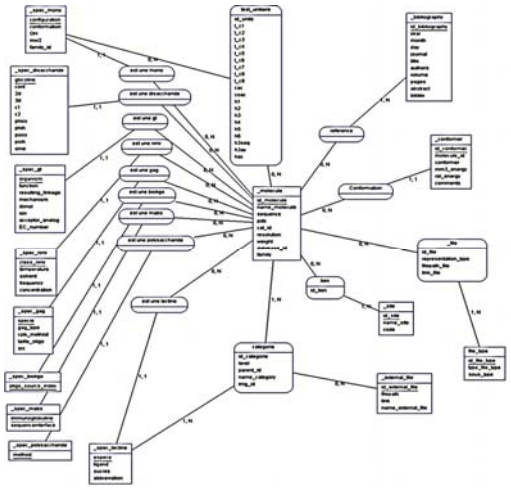
Source: X-ray - PDB

Content:
Total : 46
Classification of GAG binding proteins
Chemokine
Complement proteins
ECM proteins
Enzymes
Growth factors
Lectins
Toxins
Virus
Search:
Family, PDB
Structure of GAG



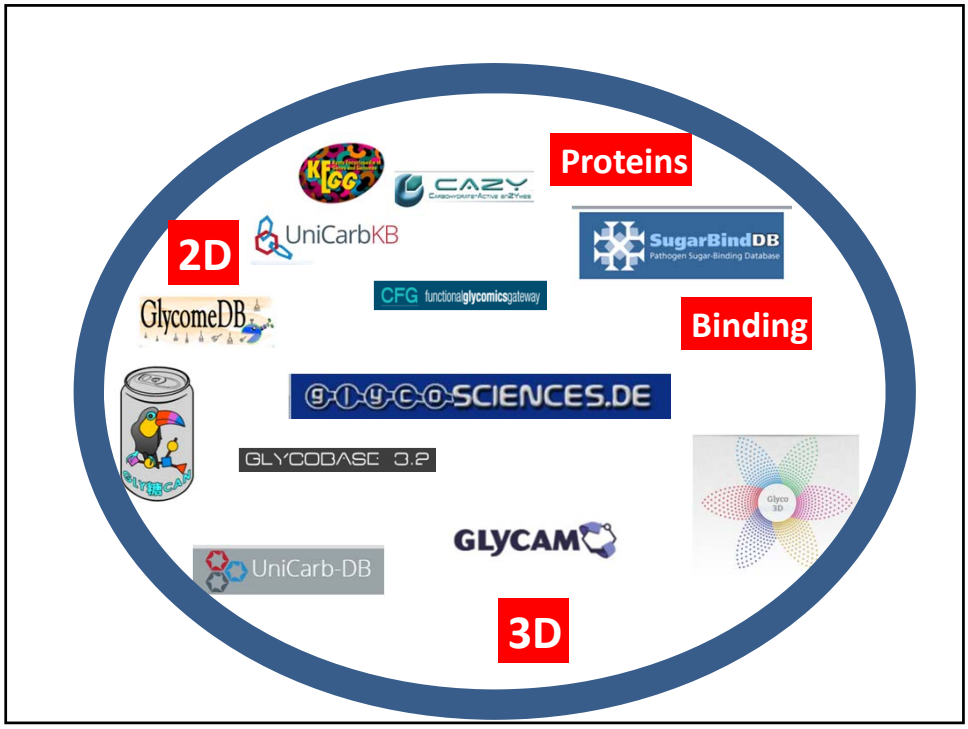
Informatics Implementation

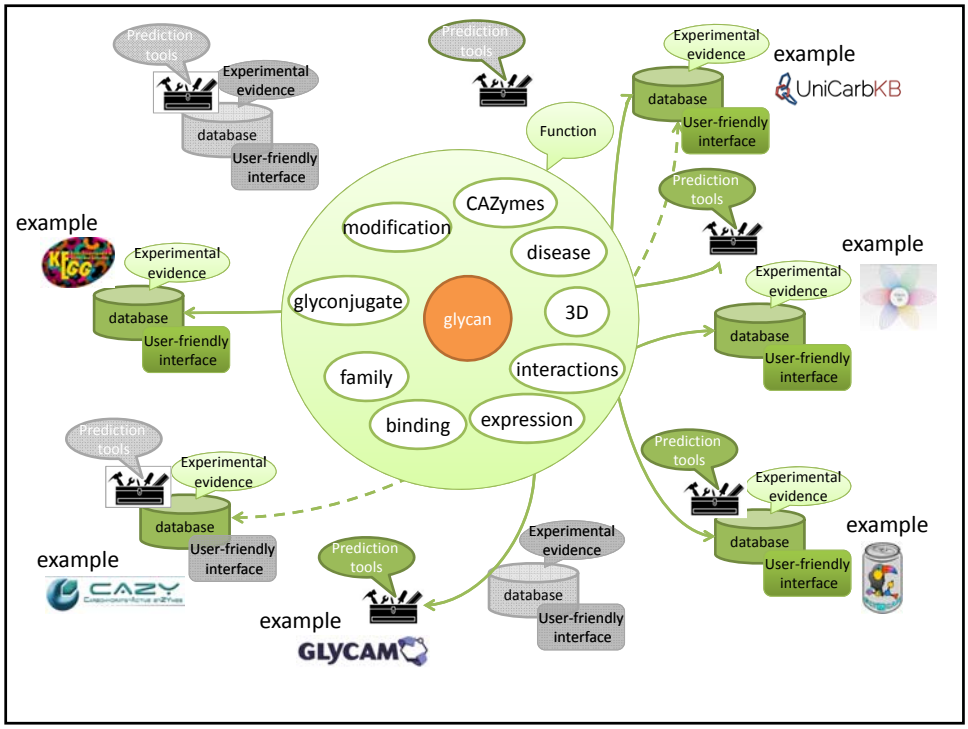
Relational Data Base: Language : PHP 5.4, DBase : MySql 5.5.24



Development Environment

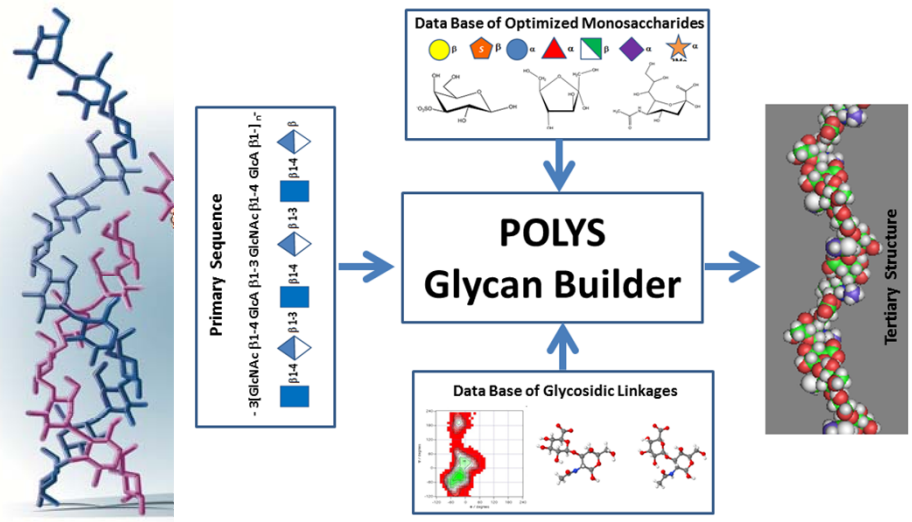
IDE : NetBean 7.3 SERVER : Wamp 2.2 Versionning: TortoiseSVN 1.7.9 , BugReporting: Mantis


Proteins
2D
UniCarbKB
SugarBindDB
Pathogen Sugar-Binding Database
CFG functionalglycomicsgateway
GlycomeDB
Binding
GLYCOSCIENCES.DE
GLYCOBASE 3.2
UniCarb-DB
GLYCAM
3D

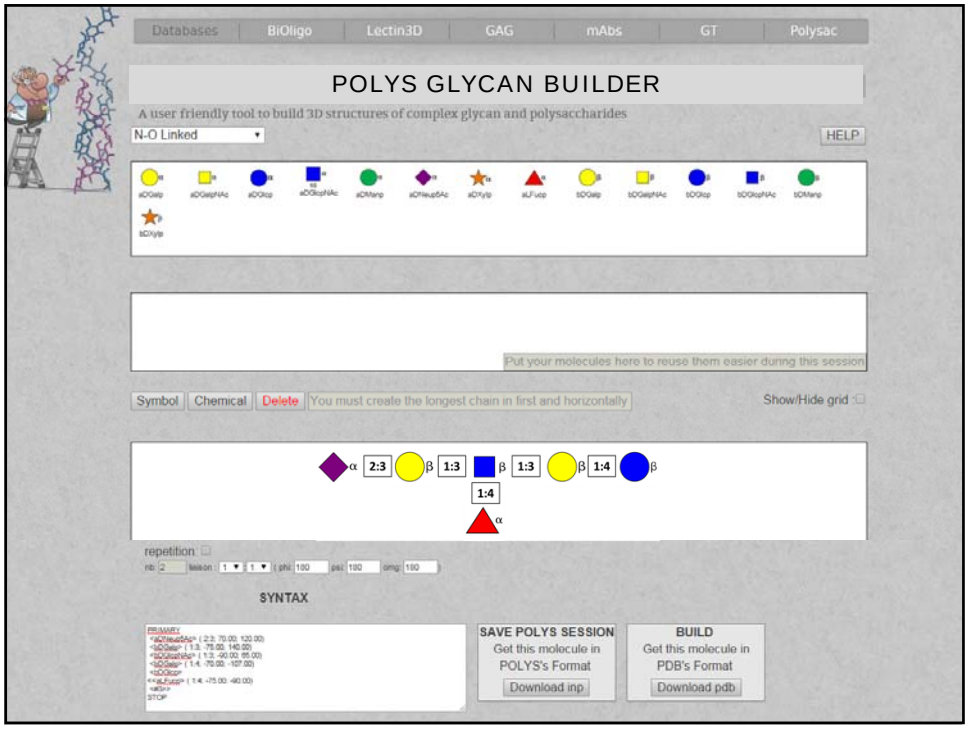

Prediction tools
Experimental evidence
database
User-friendly interface
example
UniCarbKB
Function
CAZymes
modification
disease
glyconjugate
glycan
3D
family
interactions
binding
expression
CAZY
GLYCAM

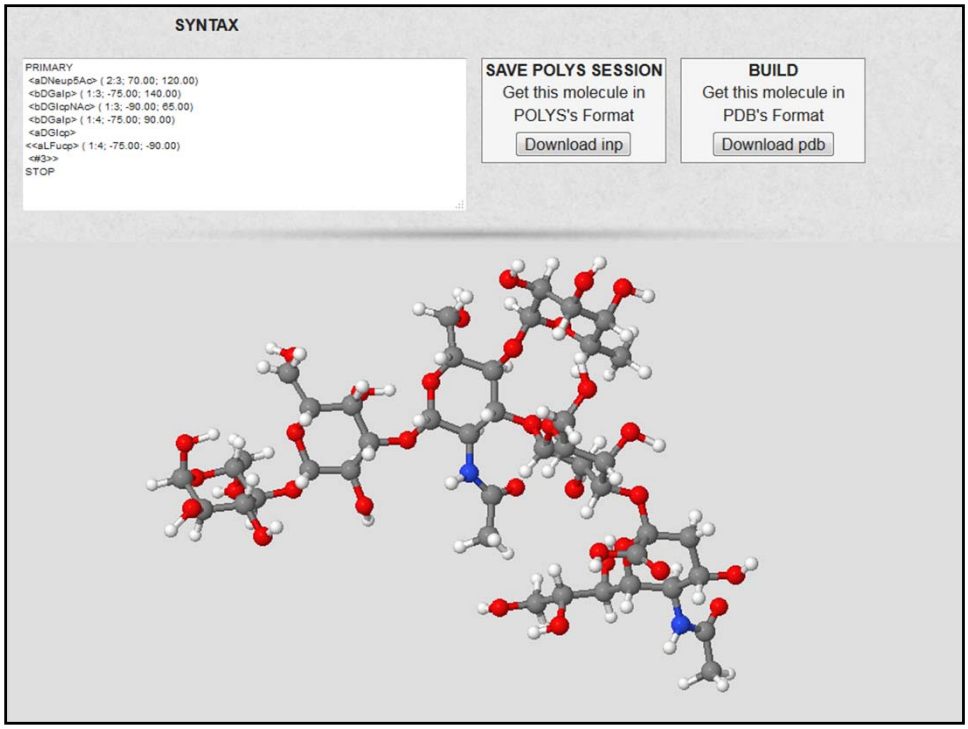
We Can Build Your Glycans

We Can Build Your Glycans: POLYS 2.0



S.Engelsen, P.I Hansen & S. Perez (2013) An Open Source Software Package for Building 3-D Structures of Polysaccharides, *Biopolymers.*


Databases
BiOligo
Lectin3D
GAG
mAbs
GT
Polysac
POLYS GLYCAN BUILDER
A user friendly tool to build 3D structures of complex glycan and polysaccharides
N-O Linked
HELP
Put your molecules here to reuse them easier during this session
Symbol
Chemical
Delete
You must create the longest chain in first and horizontally
Show/Hide grid
repetition
SYNTAX
SAVE POLYS SESSION
Get this molecule in POLYS's Format
Download inp
BUILD
Get this molecule in PDB's Format
Download pdb


SYNTAX
PRIMARY
<aDNeup5Ac> (2:3; 70.00; 120.00)
<bDGalp> (1:3; -75.00; 140.00)
<bDGlcpNAc> (1:3; -90.00; 65.00)
<bDGalp> (1:4; -75.00; 90.00)
<aDGlcp>
<<aLFucp> (1:4; -75.00; -90.00)
<#3>>
STOP
SAVE POLYS SESSION
Get this molecule in POLYS's Format
Download inp
BUILD
Get this molecule in PDB's Format
Download pdb

SWEET UNITY MOL

Biomolecules
Standardized representations

Proteins

Nucleic Acids

Carbohydrates

Identification of monosaccharide types.
Conformations (*C, E, T, B,….).*
Location in single chain / multiple branched. chains.
Depiction of secondary structures.
Constituents of complex assemblies.
(glycoproteins, protein-carbohydrate, …)

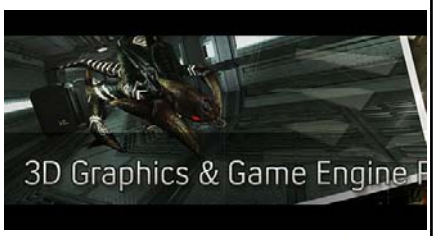
Compatible with accepted pictorial representations used in carbohydrate chemistry, biochemistry and glycobiology and structural biology format (pdb).

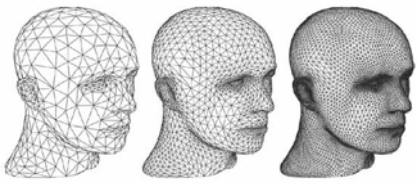
Production of publication-quality figures.

Open Access / No steep learning curve

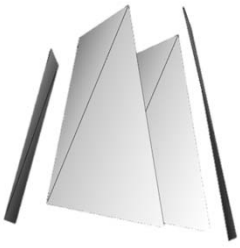
Multiple platforms i.e.Windows, MacOS and Linux operating systems, web pages, …..

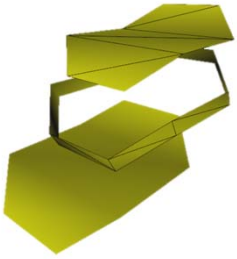
From Game Engine to Macromolecular Graphics





Tessalation

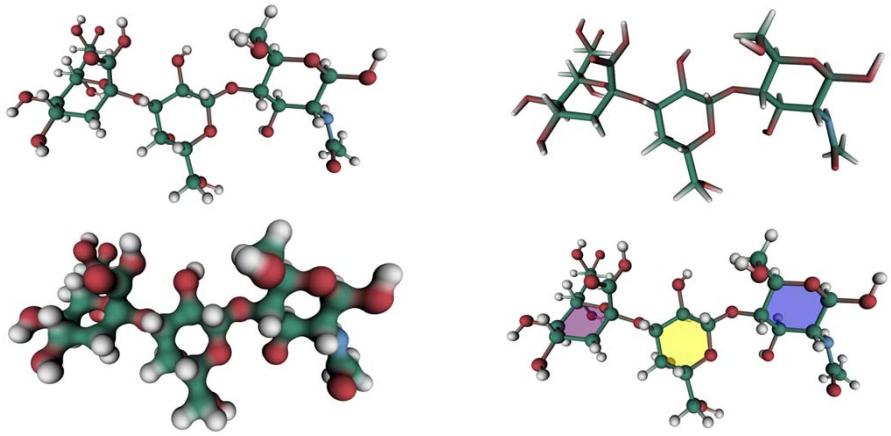




Unity3D provides an optimized set of graphical primitives for rendering.
We use triangulated spheres, triangulated cubes and lines. - mesh

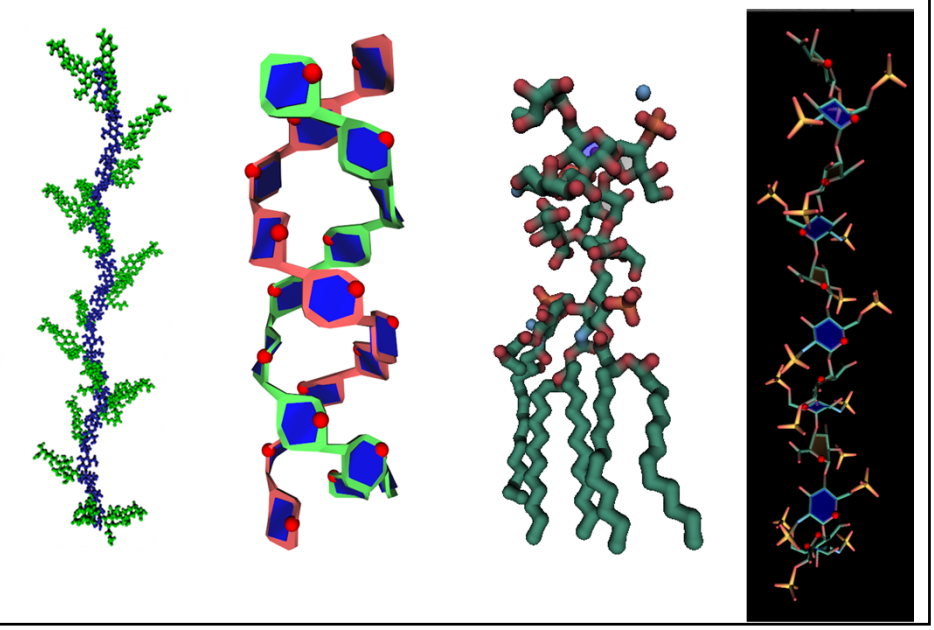
Sweet Unity Mol: *Ring Blending*

Neu5Ac α2-3 Gal β1-4 GlcNAc

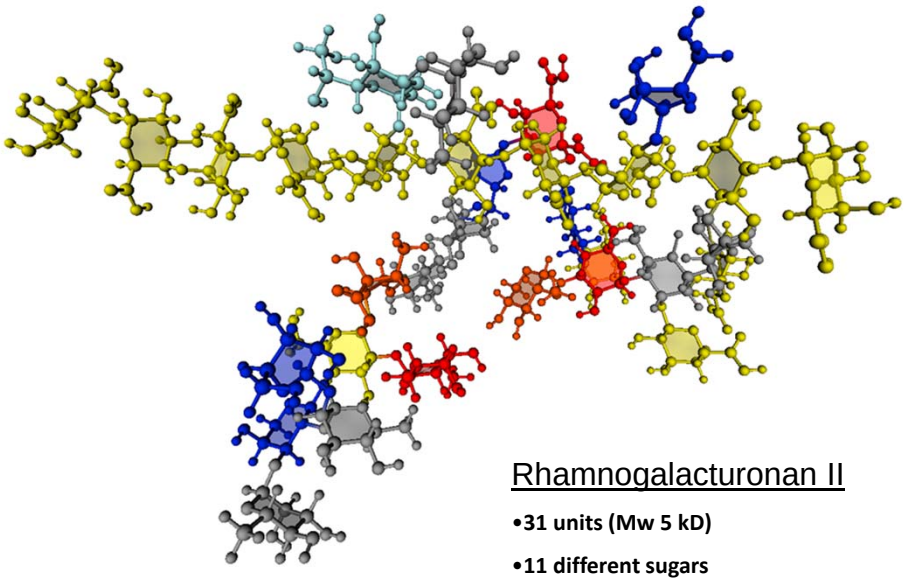


HyperBalls: is composed of spheres depicting atoms linked by hyperboloid primitives rather than simple cylinders

Polysaccharides



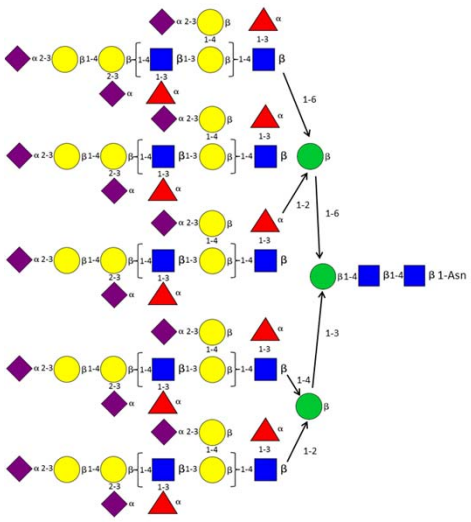
Complex Oligosaccharides & Glycans

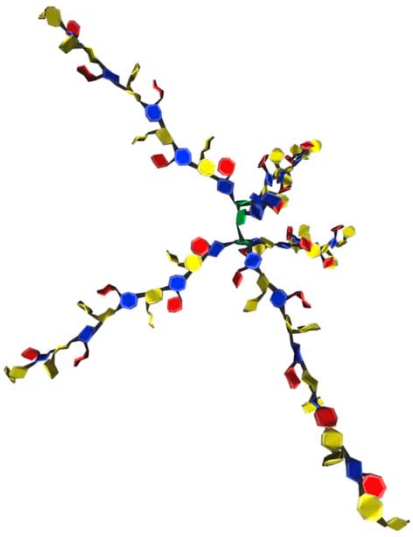


Rhamnogalacturonan II

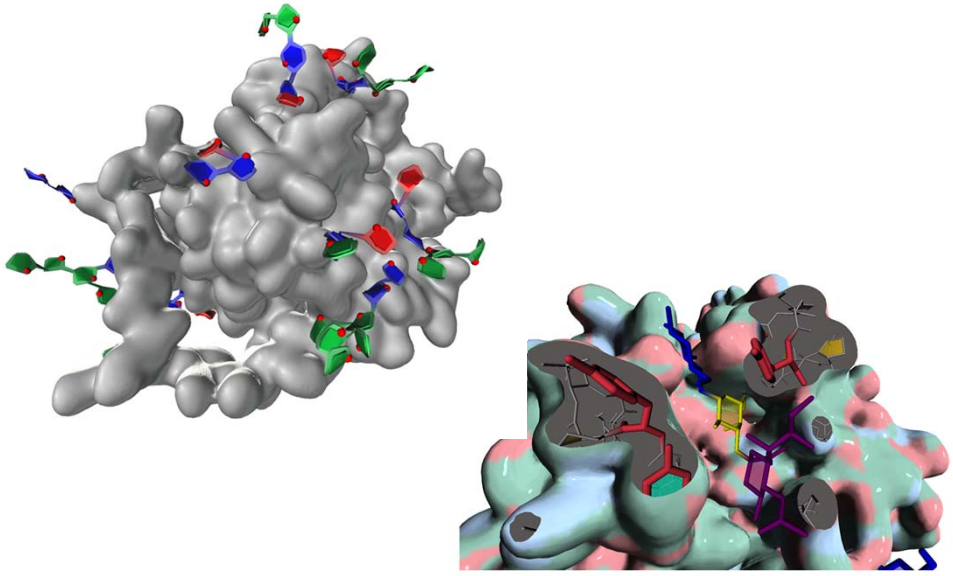
- **31 units (Mw 5 kD)**
- **11 different sugars**

Complex Oligosaccharides & Glycans

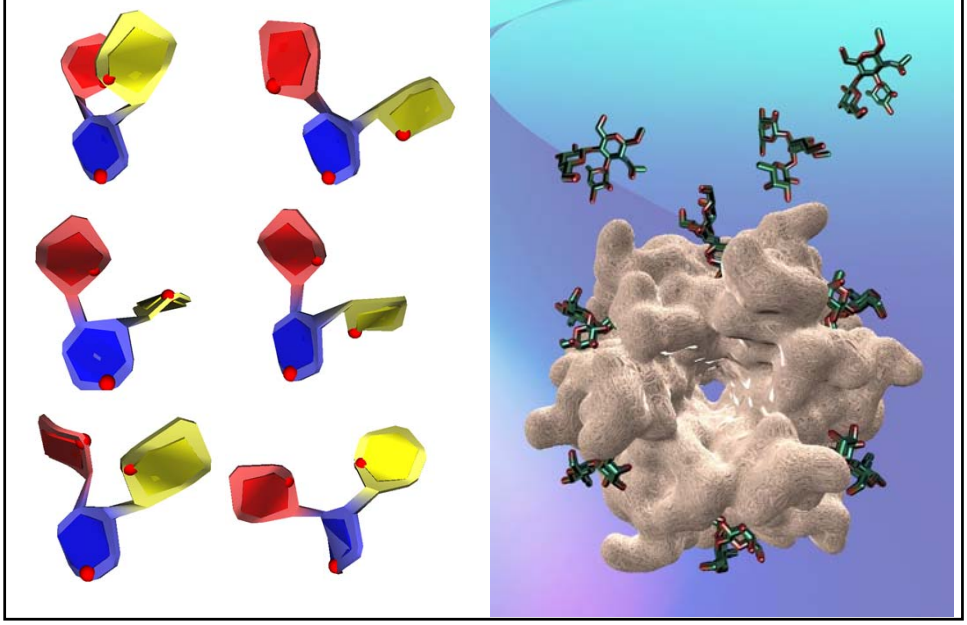




Glycoproteins – Protein Carbohydrate Interactions



The Hidden Conformations of LewisX



From 1D to 3D Representations