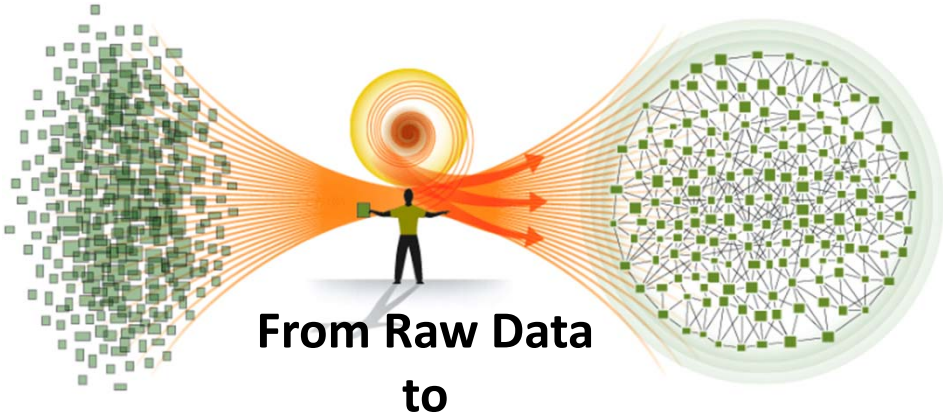
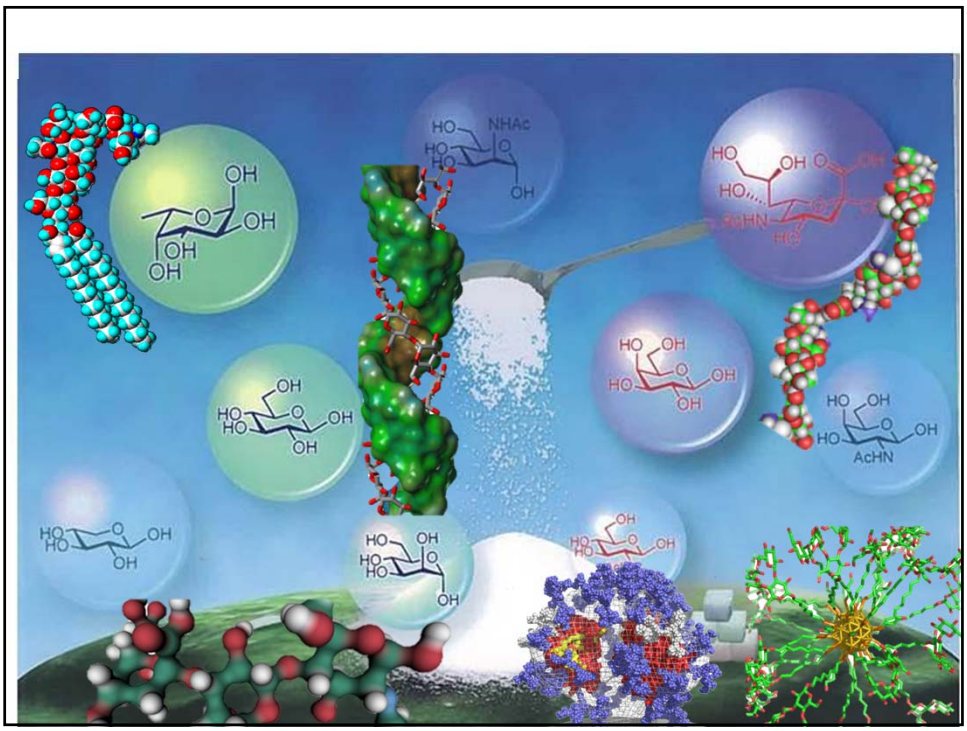
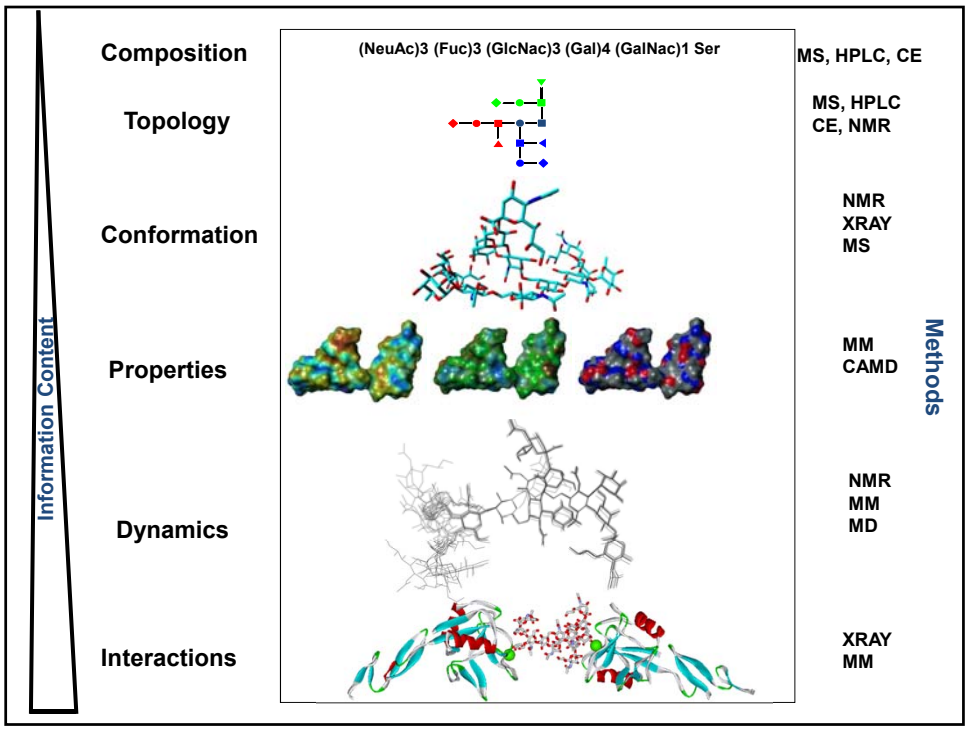
Integrating Knowledge



From Raw Data to Databases & Semantic Web

Serge Perez, Firenze, Ottobre 2016

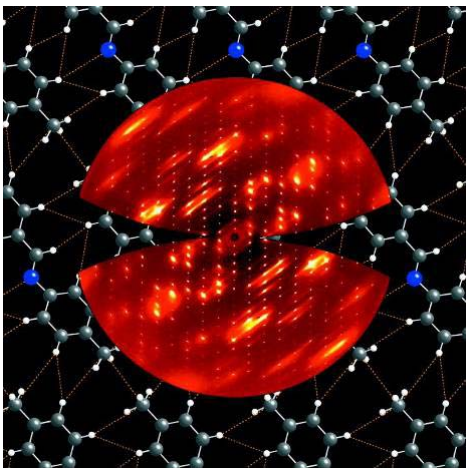


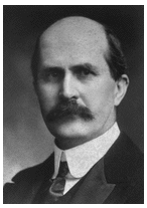


Crystal Structures

From Monosaccharides To Polysaccharides

International Year of Crystallography



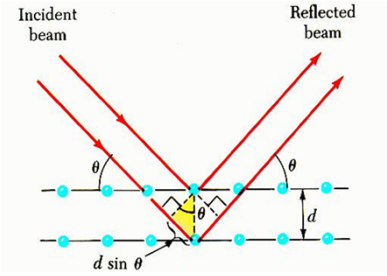


W. L. Bragg

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.

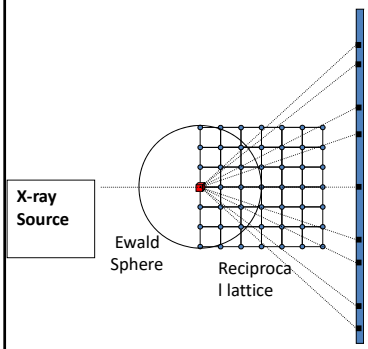
It also commemorated the 400th anniversary of Kepler's observation in 1611 of the symmetrical form of ice crystals, which began the wider study of the role of symmetry in matter.

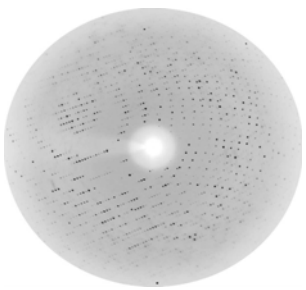


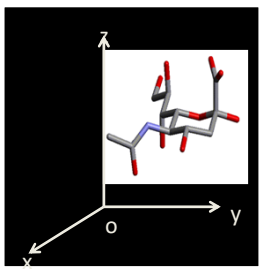
X-ray interact with the spatial distribution of Valence electrons.

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

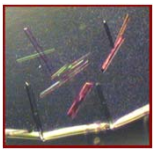
Molecular & Crystal Structures of Carbohydrates

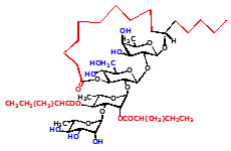
Experimental Conditions and Limitations (X and N)

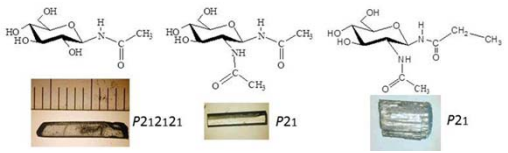
X-ray and Neutron have wavelengths in the same order as the interatomic distances (Angstrom).
Electron are the scattering elements of the incident X-ray.
Nuclei are the scattering elements of the incident Neutron radiation.

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.







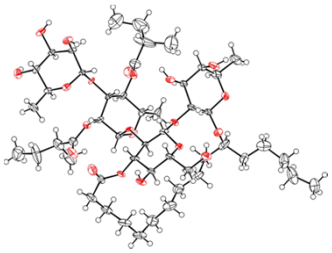
Molecular & Crystal Structures of Carbohydrates

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, α, β, γ); 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

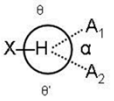
Hydrogen Bonding in Crystalline Oligosaccharides

Analysis of high accurate X-ray analysis – Neutron diffraction

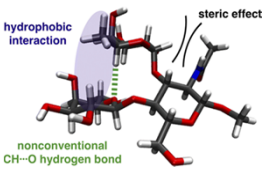
dX-dN =(C-H) = -0.096(7)
dX-dN =(O-H) = -0.155(10)

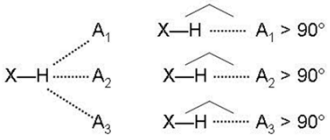
X—H ····· A X—H·····A ~ 160°± 20°

CHO bonds



α + θ + θ' ~ 360°

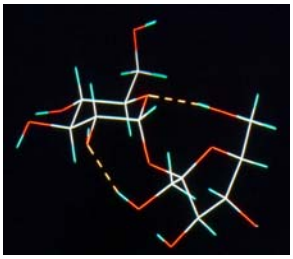


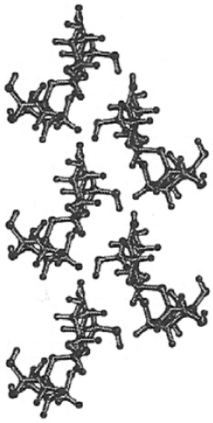


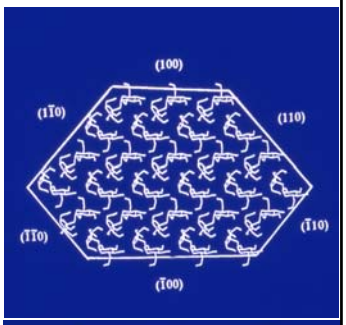
Maximize the Hydrogen Bond interactions throughout the participation of all hydroxyl groups and as many rings oxygen. Two and three-centered bonds.
Maximize cooperativity by forming as many finite and infinite chains of hydrogen bonds as possible.

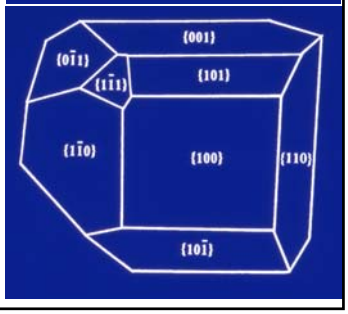
Molecular & Crystal Structures of Carbohydrates

Packing Features







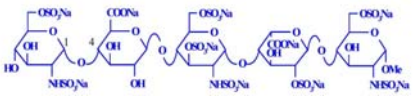


Molecular & Crystal Structures of Carbohydrates

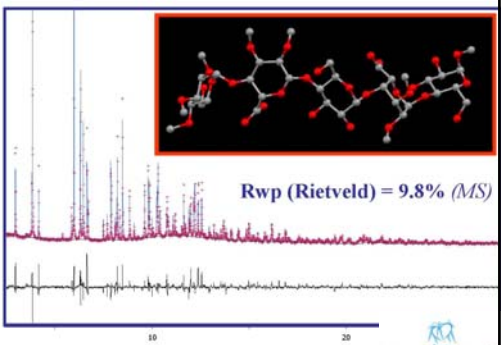
Powder Diffraction

1. Identification of Crystalline Polymorphs

2. Solving Crystal Structures – Rievelt Method + Molecular Modelling

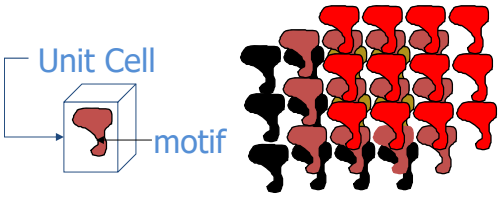


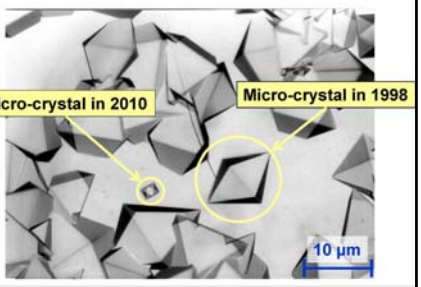
Synthetic Pentasaccharide
ID31@ESRF, λ =0.8 A
Monoclinic $P2_1$
a=15.54, b=8.83; c=17.67, β=94.6

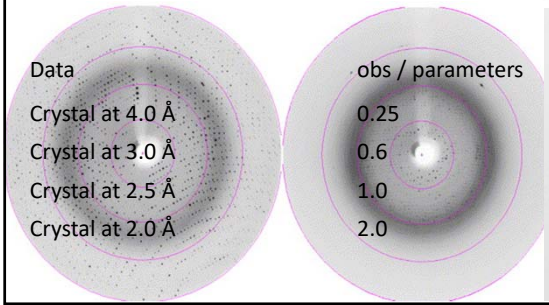


Crystalline Conformations of Oligosaccharides in Proteins

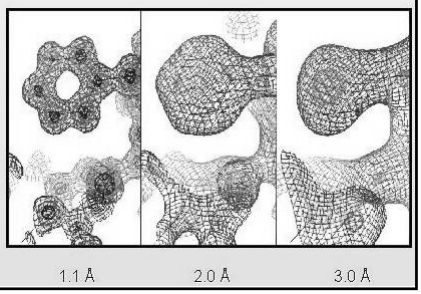
Experimental Conditions and Limitations

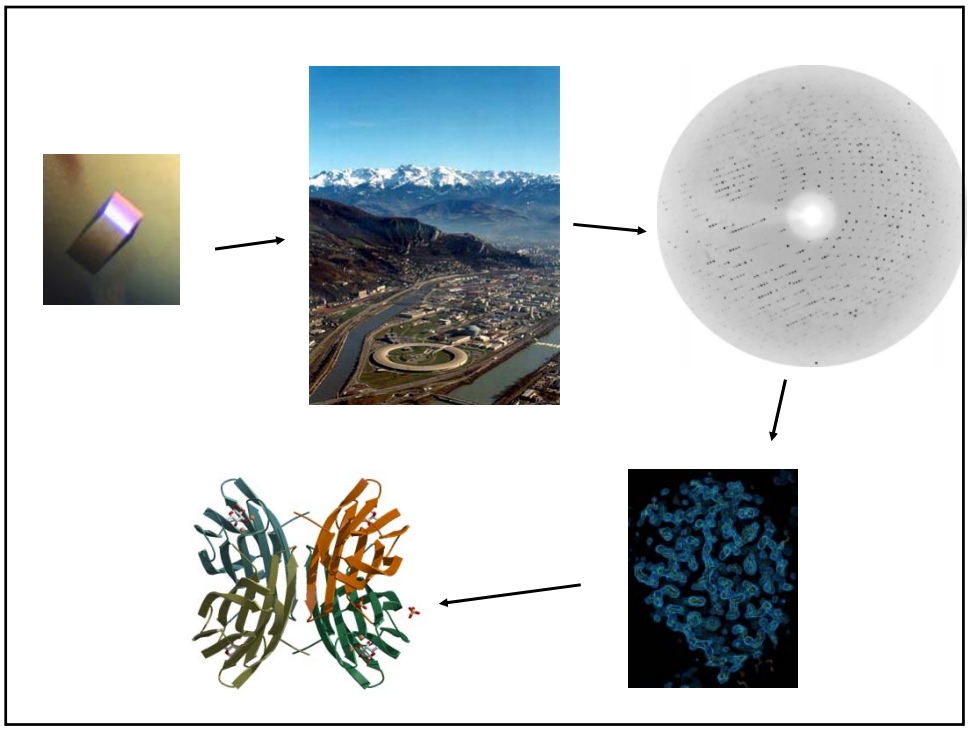


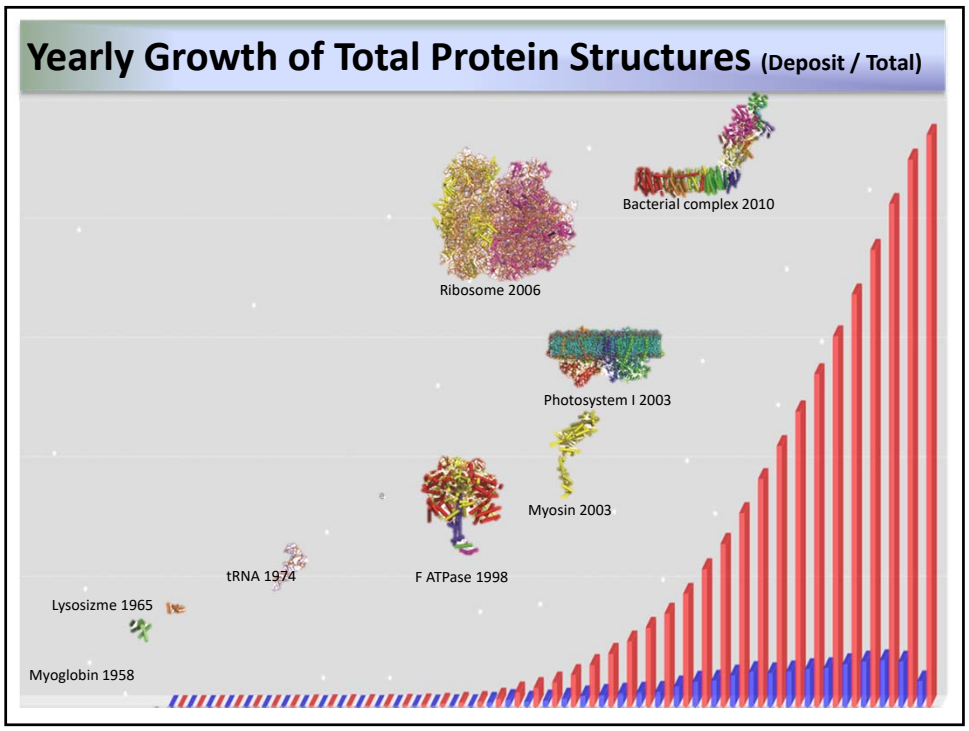




Data	obs / parameters
Crystal at 4.0 Å	0.25
Crystal at 3.0 Å	0.6
Crystal at 2.5 Å	1.0
Crystal at 2.0 Å	2.0

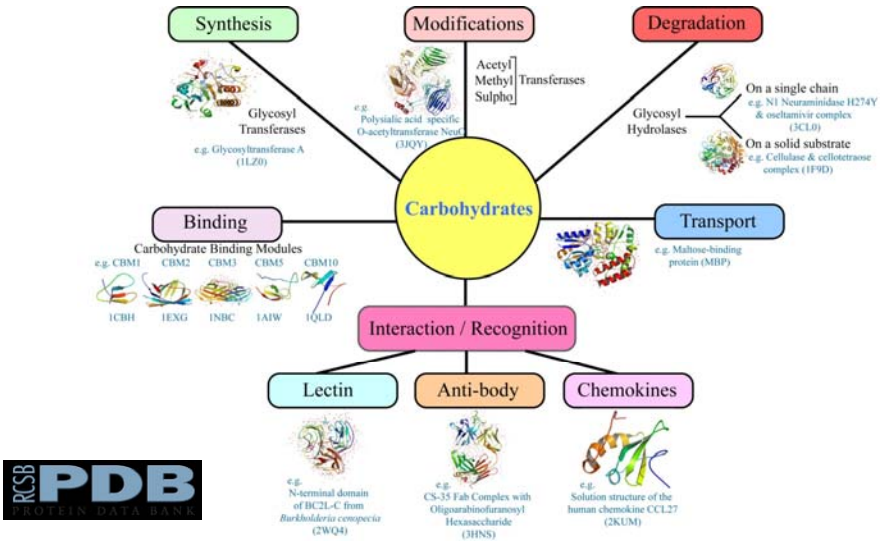





Yearly Growth of Total Protein Structures (Deposit / Total)
Bacterial complex 2010
Ribosome 2006
Photosystem I 2003
Myosin 2003
F ATPase 1998
tRNA 1974
Lysosizme 1965
Myoglobin 1958

Crystalline Conformations of Oligosaccharides in Proteins

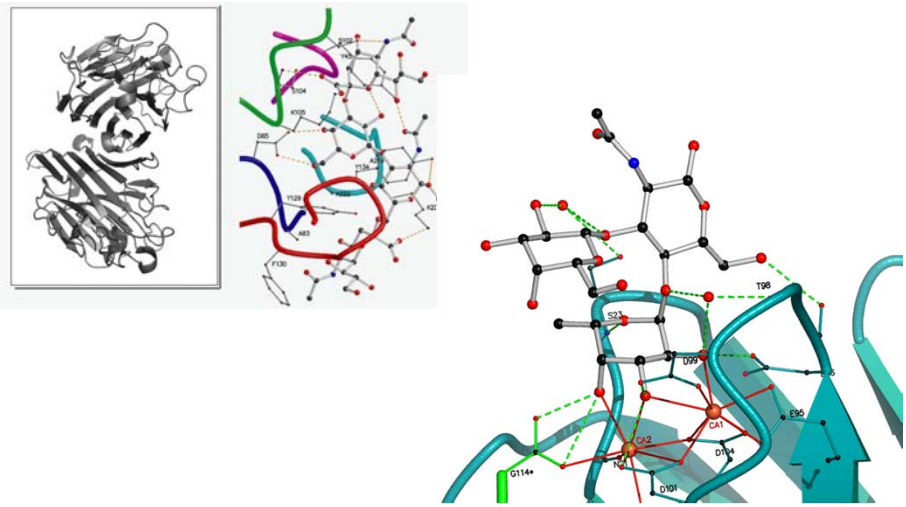
Protein-Carbohydrate Interactions



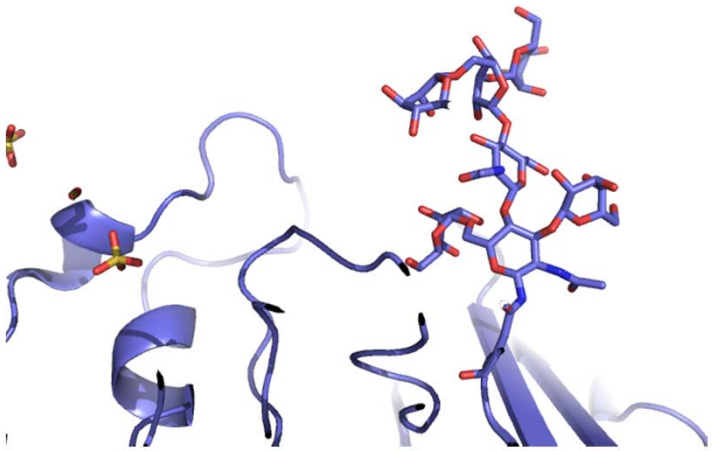
Protein Data Bank : http://www.rcsb.org/pdb/home/home.do

Crystalline Conformations of Oligosaccharides in Proteins

Oligosaccharides –Lectin Complexes



Crystalline Conformations of Oligosaccharides in Proteins

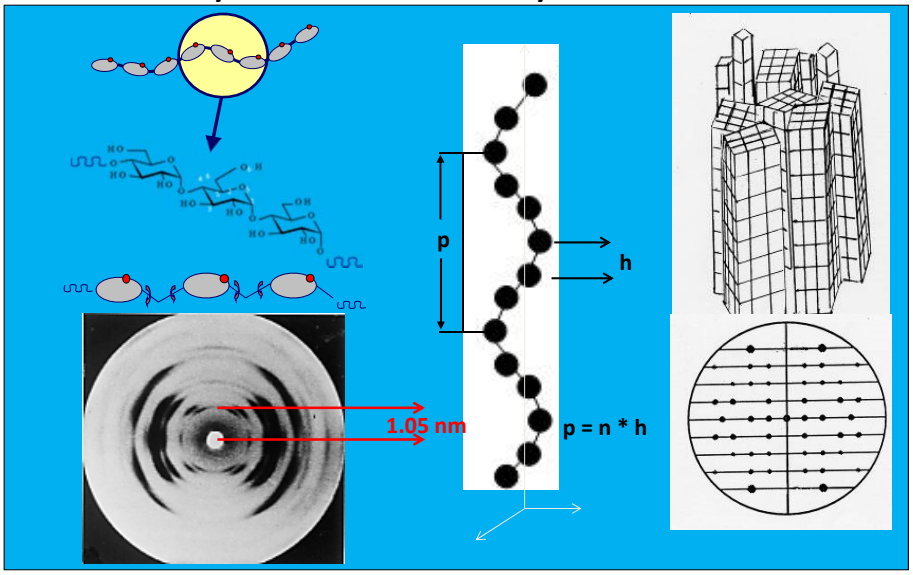


N-linked glycan:

NeuAcα2-6Galβ1-4GlcNAcβ1-2Manα1
6
Manβ1-4GlcNAcβ1-4GlcNAc-Asn
3
NeuAcα2-6Galβ1-4GlcNAcβ1-2Manα1
Fucα1
6

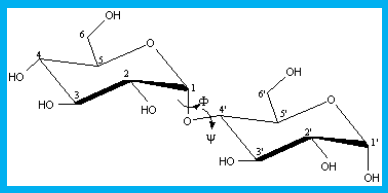
Crystalline Conformations of Polysaccharides

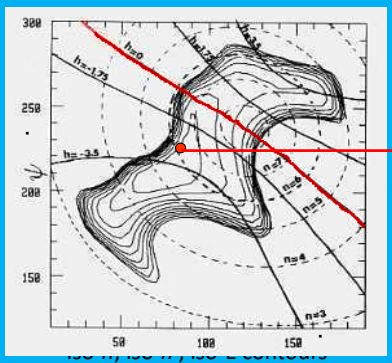
X-Ray Fiber Diffraction of Polysaccharides



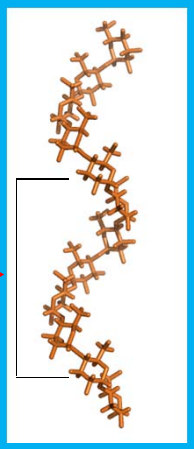
Crystalline Conformations of Polysaccharides

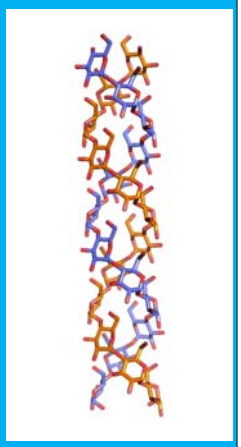
X-Ray Fiber Diffraction of Polysaccharides





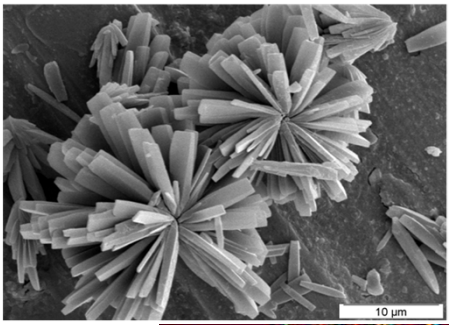
2.1 nm

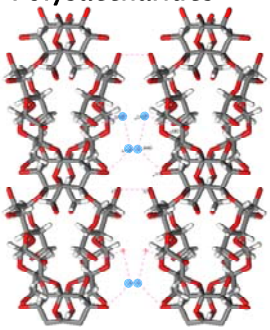


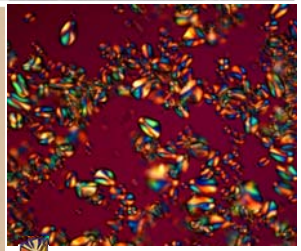


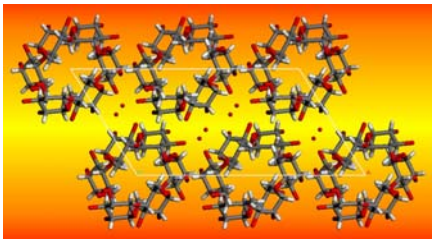
Crystalline Conformations of Polysaccharides

Synchrotron X-Ray Diffraction of Polysaccharides





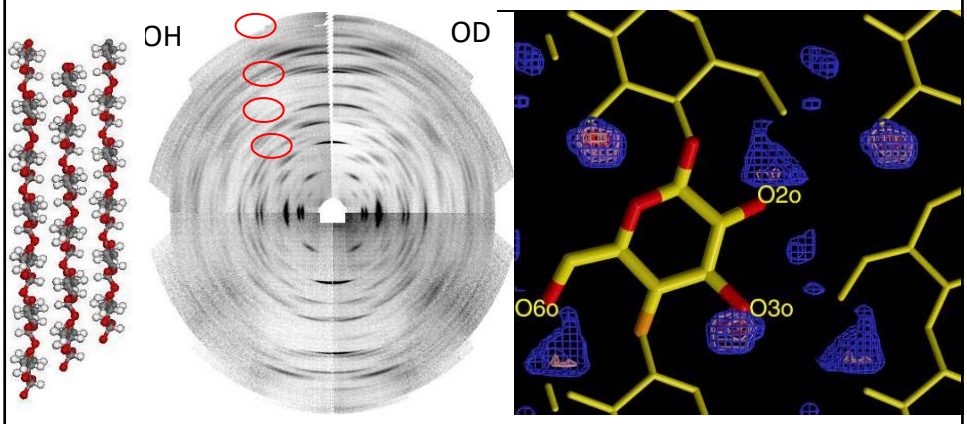




D. Popov, 2009, Macromolecules, 42, 1167-1174

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

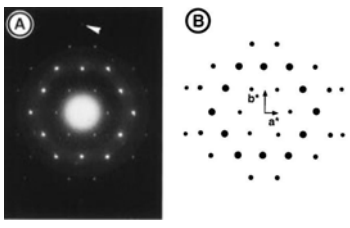
Electrons are charged particles and interact with matter through the Coulomb forces. The incident electrons feel the influence of both the positively charged atomic nuclei and the surrounding electrons.

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.

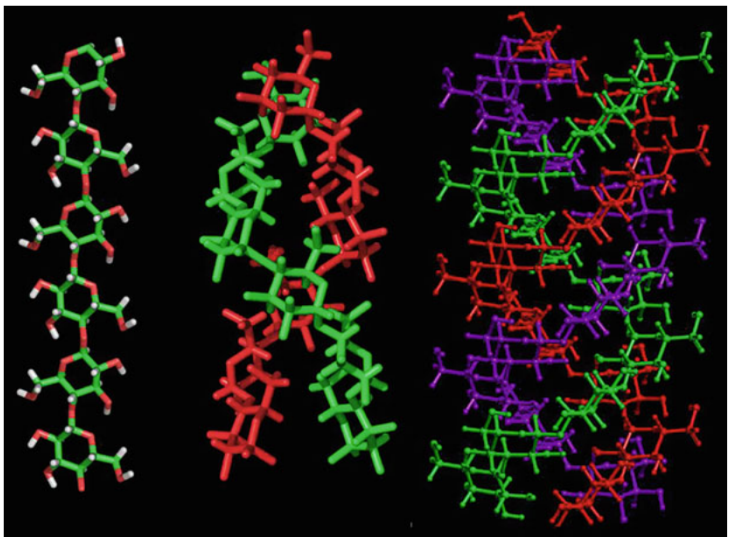
Electron diffraction in TEM is subject to several important limitations.

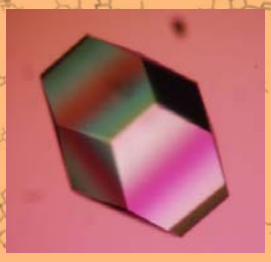
The sample to be studied must be electron transparent, meaning the sample thickness must be of the order of 100 nm or less.

- 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*

Computational GlycoScience

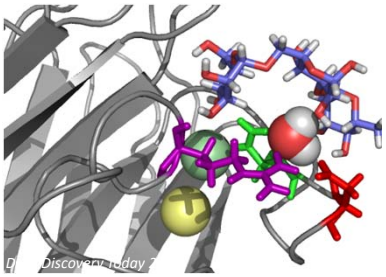
-Principles of Molecular Mechanics

-Carbohydrates
Structure Building, Force Fields

-Receptors
Homology modelling, Force Fields

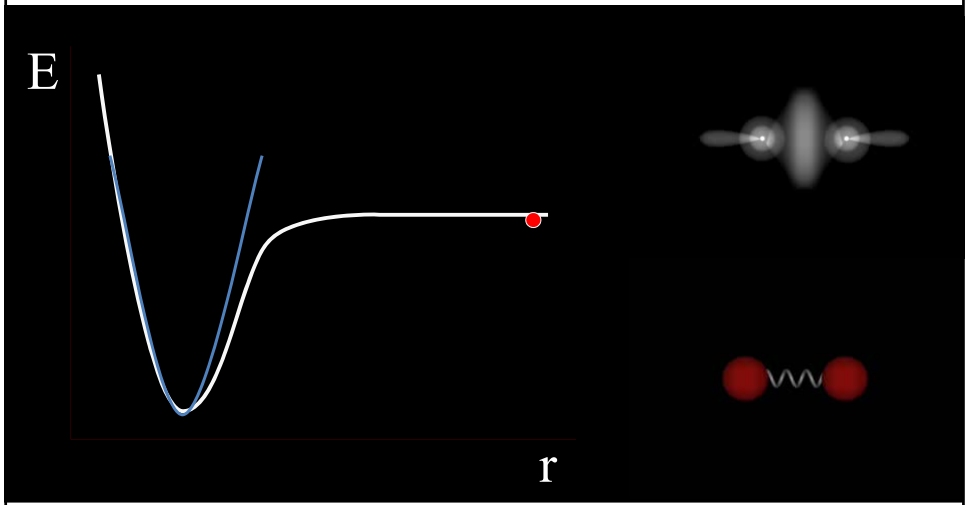
-Carbohydrate-Receptor Complexes
Molecular Docking, Molecular Dynamics

-Case study

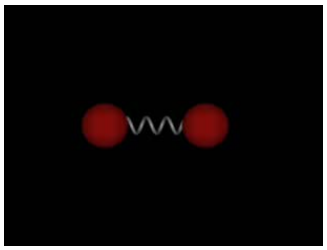


27

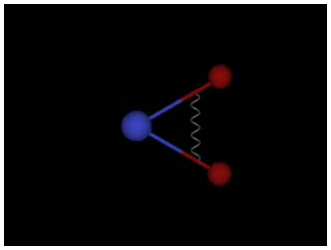
Molecular Mechanics: How does it work ?



Calculation of Energy.

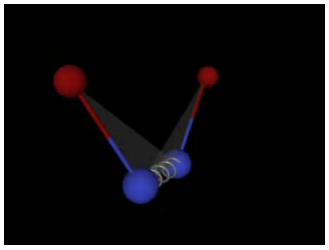


Binding

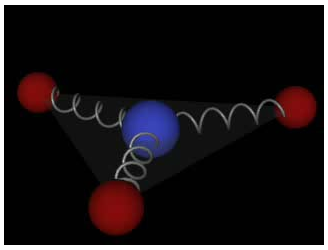


Bending

Terms…



Torsion



Out of Plane

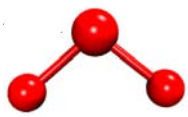
Bonding Interactions

-Molecular mechanics

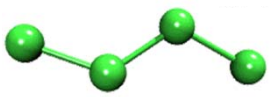
$E_{improper}$ is the energy required to deform a planar group of atoms from its equilibrium
(Hookean potential)

$E = E_{str} + E_{bend} + E_{tor} + E_{improper} + E_{vdW} + E_{qq}$

E_{str} represents the energy required to stretch or compress a bond from its equilibrium
(Hookean potential)



E_{bend} represents the energy required to bend a bond from its equilibrium
(Hookean potential)

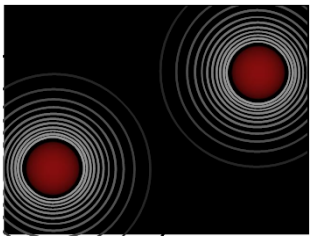


E_{tor} is the energy of torsion needed to rotate about bonds.
(Simple periodic function)

30

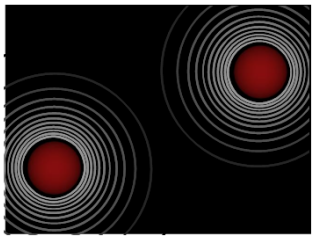
Non Bonding Interactions.

δ+



Electrostatic

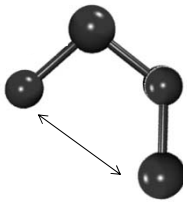
δ–



Van der Waals

Carbohydrates

-Molecular mechanics



$$E = E_{str} + E_{bend} + E_{tor} + E_{improper} + E_{vdW} + E_{qq}$$

E_{vdW} represents the sum of the long-range attractive or repulsive (due to an overlap of electron atomic orbitals) forces between atoms. The Lennard Jones potential is a mathematical simple van der Waals model (the so-called 6-12 equation)

E_{qq} represents the electrostatic energy between atoms calculated through the Coulomb potential function.
The relative dielectric constant ε is used for mimicking solvent effects.

32

Holtje H. et al., *Molecular Modeling* (2008), 2, 15-28 WILEY-VCH

Force-Fields.

$$E_{bond} = \frac{k_b}{2}(l - l_0)^2 \longrightarrow k_b,\ l_0$$

$$E_{angl} = \frac{k_a}{2}(\theta - \theta_0)^2 \longrightarrow k_a,\ \theta_0$$

$$E_{torsion} = \sum_i^N \frac{V_i}{2}[1 + \cos(\omega i - \gamma)] \longrightarrow V_i,\ \gamma$$

$$E_{oop} = \frac{k_p}{2}\delta^2 \longrightarrow k_p$$

$$E_{elect} = \sum_{i=1}^N \sum_{j=i}^N \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \longrightarrow q_i,\ q_j$$

$$E_{VdW} = \sum_{i=1}^N \sum_{j=i}^N 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \longrightarrow \varepsilon,\ \sigma$$

?

$$E = E_{bond} + E_{angl} + E_{torsion} + E_{oop} + E_{elect} + E_{VdW} + \ldots$$

Parametrization.

Experimental Data
- Cristallography
- Spectroscopies

ab initio Calculations

$k_b,\ l_0$

$k_a,\ \theta_0$

$V_i,\ \gamma$

k_p

$q_i,\ q_j$

$\varepsilon,\ \sigma$

Portability !!!

Force-fields.

$$E_{bond} = \frac{k_b}{2}(l - l_0)^2 \longrightarrow k_b,\ l_0$$

$$E_{angl} = \frac{k_a}{2}(\theta - \theta_0)^2 \longrightarrow k_a,\ \theta_0$$

$$E_{torsion} = \sum_i^N \frac{V_i}{2}[1 + \cos(\omega_i - \gamma)] \longrightarrow V_i,\ \gamma$$

$$E_{oop} = \frac{k_p}{2}\delta^2 \longrightarrow k_p$$

$$E_{elect} = \sum_{i=1}^N \sum_{j=i}^N \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \longrightarrow q_i,\ q_j$$

$$E_{VdW} = \sum_{i=1}^N \sum_{j=i}^N 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \longrightarrow \varepsilon,\ \sigma$$

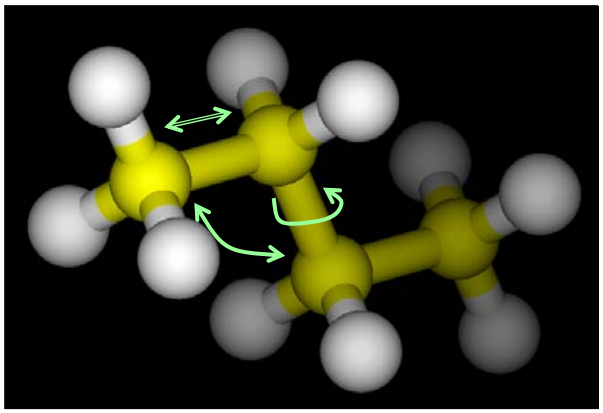
MM2
MM3
MM4
AMBER
GROMOS
CHARMM
SCF
TRIPOS…

$$E = E_{bond} + E_{angl} + E_{torsion} + E_{oop} + E_{elect} + E_{VdW} + \ldots$$

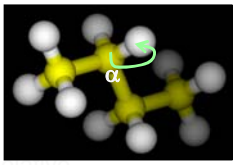
Potential Energy.

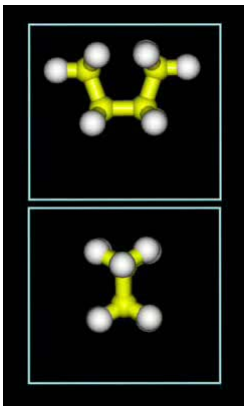
$$E = f(\underbrace{x_1, y_1, z_1}_{\text{Atom 1}}, \underbrace{x_2, y_2, z_2}_{\text{Atom 2}}, \underbrace{x_3, y_3, z_3}_{\text{Atom 3}}, \ldots \underbrace{x_N, y_N, z_N}_{\text{Atom N}})$$

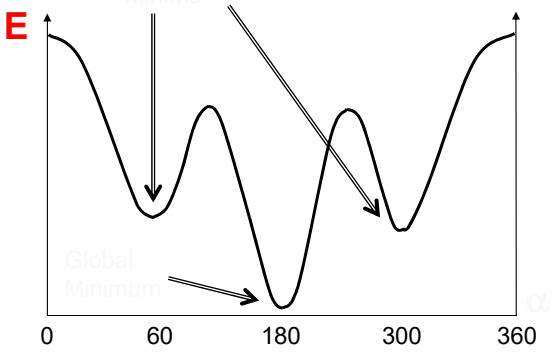
Internal coordinates (3N-6)



Potential Energy.



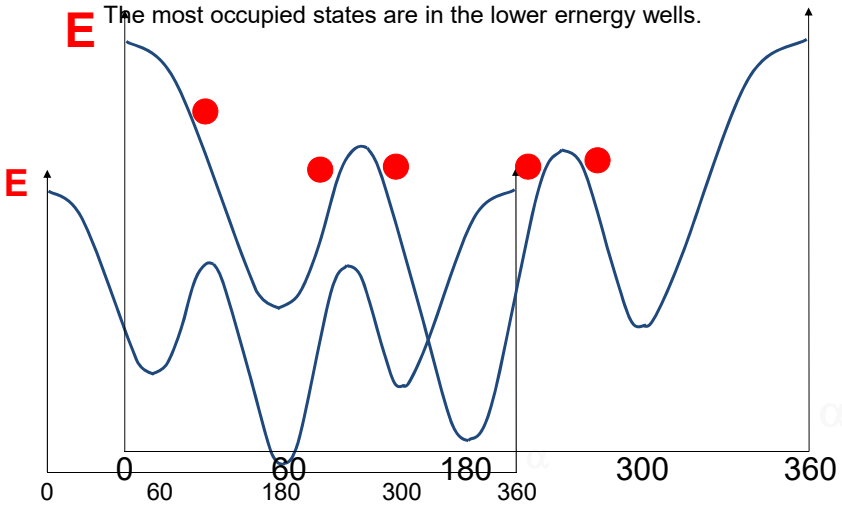




Potential Energy. The minima.

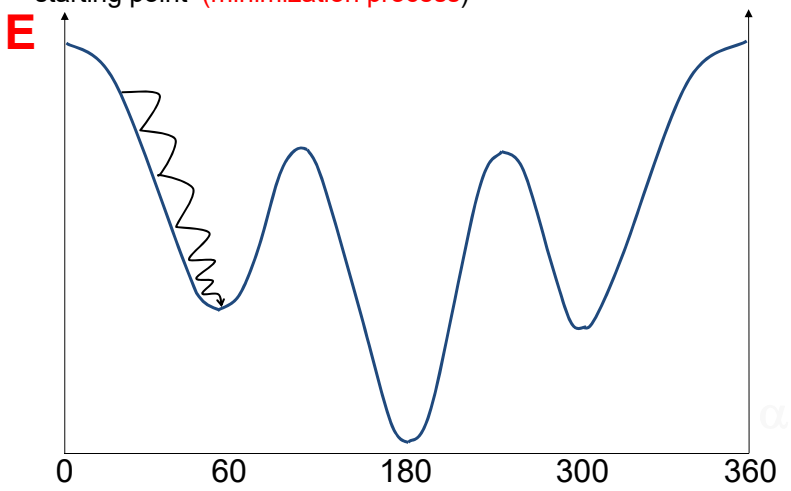
The molecules have a tendancy to remain in the lowest energy areas.

The most occupied states are in the lower ernergy wells.

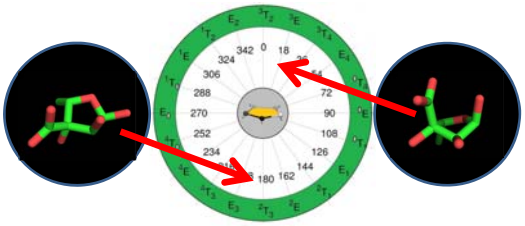


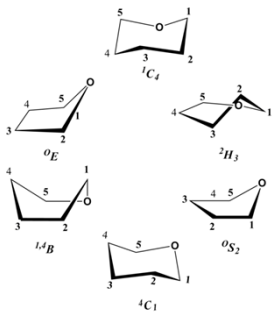
Exploring the Conformational Space…

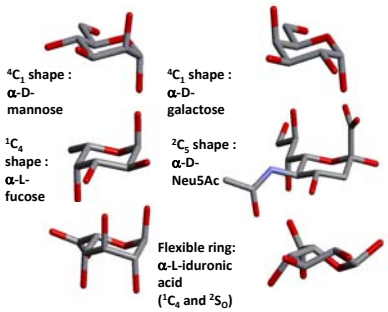
Convergence towards the closest energy minimum from the starting point (minimization process)



Carbohydrates





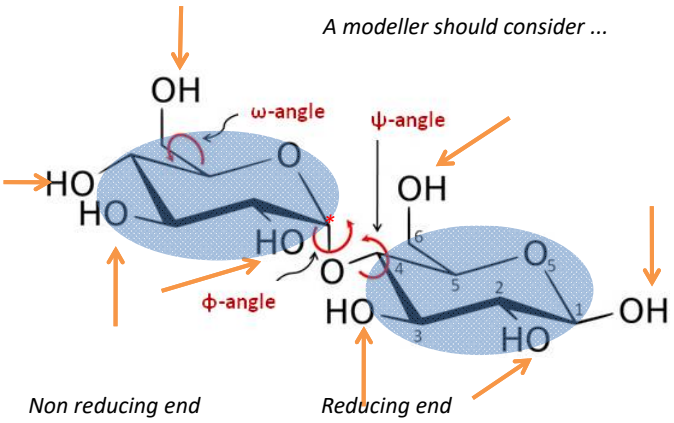


More monosaccharides on http://glycopedia.eu & http://glyco3d.cermav.cnrs.fr

40

Carbohydrates

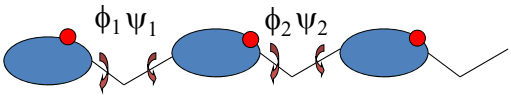
- « A group of organic compounds that contain C, H, O »



More monosaccharides on http://glycopedia.eu

41

Exploring the Conformational Space of Flexible Oligosaccharides



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

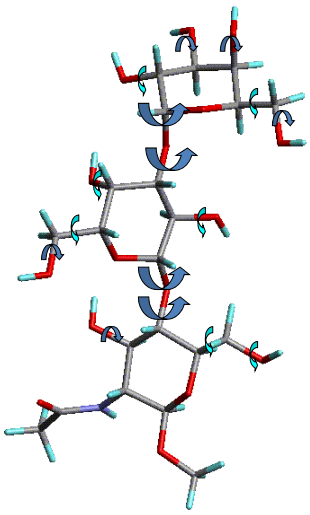
Combine the lowest energy minima of each disaccharide map

- long range interactions
- branched structures
....

But very useful for building starting structures!

Exploring the Conformational Space of Flexible Oligosaccharides

Systematic search of all possible conformations ?



For a trisaccharide:

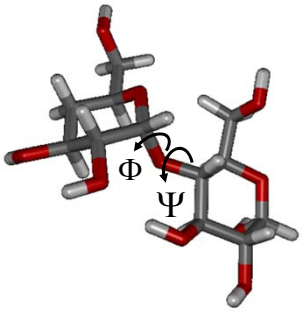
4 torsions to be searched
with 20° steps
18^4 conformations

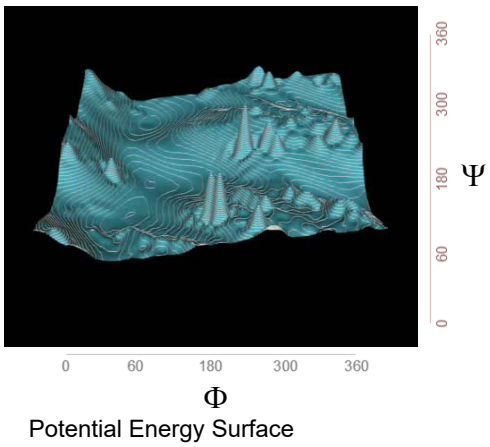
12 pendant groups
Staggered orientations:
3^{12} combinations

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

αGal(1-3)βGal(1-4)βGlcNAc

Exploring the Hyper Space...

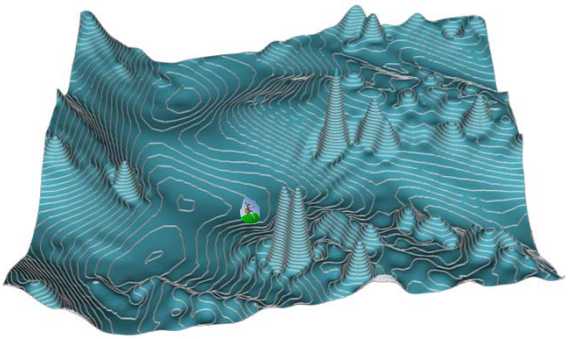




Potential Energy Surface

Exploring the Hyper Space...

Exploring the Hyper Space...



The explorer ....

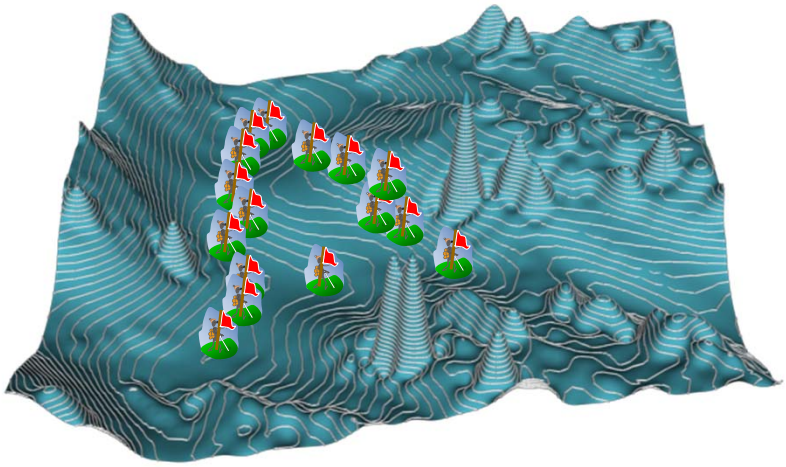
Does not “see” the location of other minima.

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

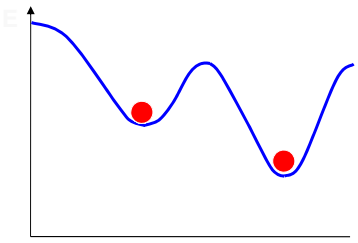
Monte Carlo Calculations...

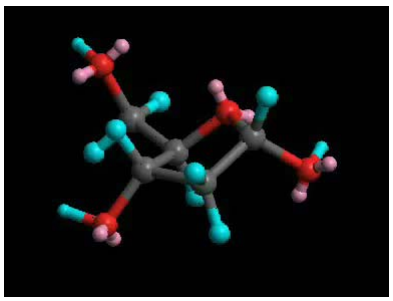
are based on small **random** variations of atomic coordinates.

Uses an algorithm that favors low energy conformations.

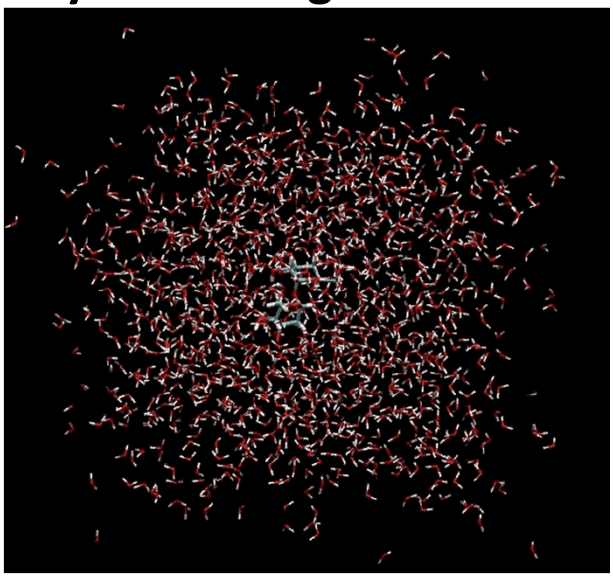


Molecular Dynamics…





Molecular Dynamics of Hydrated Oligosaccharide

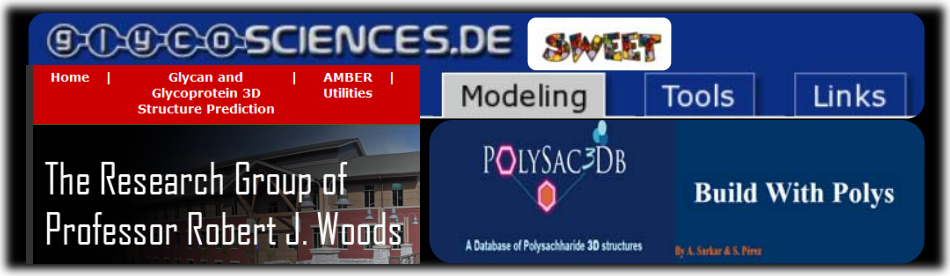


Carbohydrates

-Carbohydrate builders: the starting point for simulations

- ✓ Molecular modelling programs: Sybyl, MOE (not free for academia)

- ✓ On-line servers (free for academia)



Carbohydrates

-Carbohydrate builders: the starting point for simulations

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

1C4 4C1

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

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

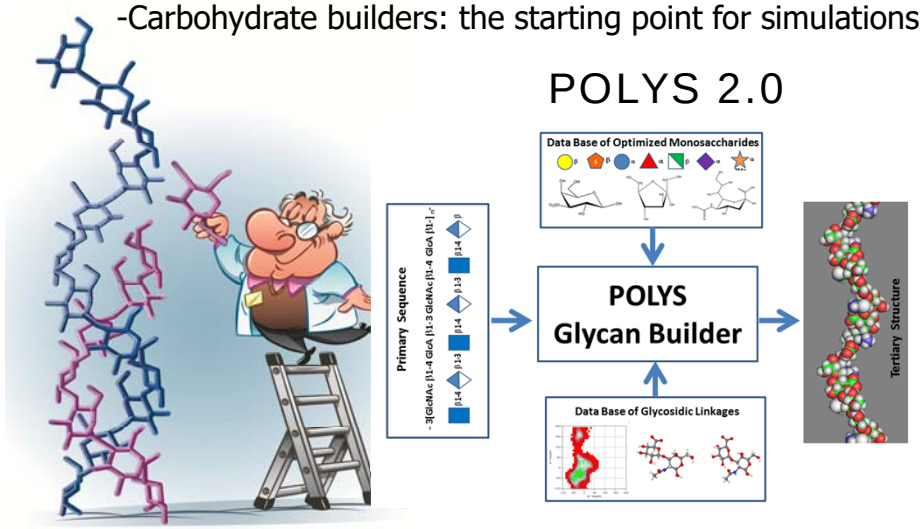
Project Name : glycam (Only letters, numbers, underscore '_', dash '-' , period '.' are allowed in project name)

☐ Add Branches HELP - How to build a Carbohydrate

http://glycam.ccrc.uga.edu/

Carbohydrates

-Carbohydrate builders: the starting point for simulations

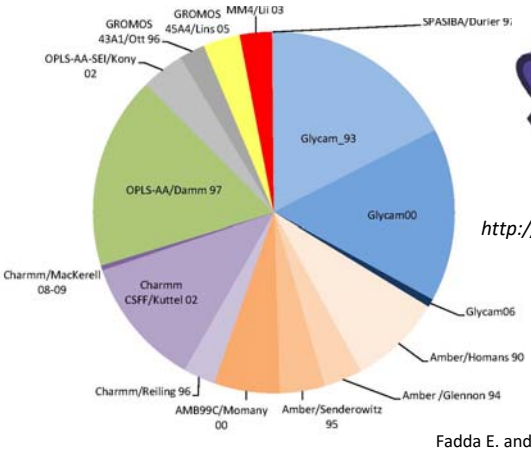


S.Engelsen, P.I Hansen & S. Perez (2013) An Open Source Software Package for Building 3-D Structures of Polysaccharides, *Biopolymers. (2014)*

Carbohydrates

-Force fields

$$V(r^n) = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos\left(n\phi - \gamma\right)\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_r R_{ij}}\right]$$

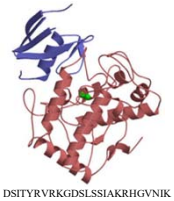


http://glycam.ccrc.uga.edu/

Fadda E. and Woods R.J., *Drug. Disc. Today* (2010), 15, 596-609

Receptors

-How to build an homology model



Target identification

✓UniProt database

Template Selection

✓BLAST-P

Target-Template Alignement

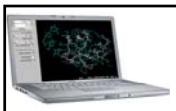
✓Modeller

Model Construction and Refinement

✓Procheck, Whatif

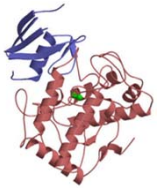
Holtje H. et al., *Molecular Modeling* (2008), 4, 111-178

Nurisso A. et al., *Homology Modeling: Methods and Protocols* (2012), 857, 137-173



Receptors

-How to build an homology model



Target identification

✓UniProt database

Template Selection

✓BLAST-P

Target-Template Alignement

✓Modeller

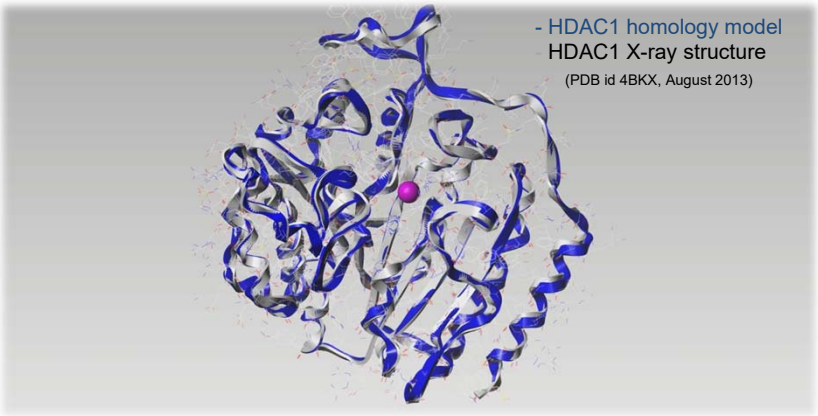
Model Construction and Refinement

✓Procheck, Whatif

Holtje H. et al., *Molecular Modeling* (2008), 4, 111-178
Nurisso A. et al., *Homology Modeling: Methods and Protocols* (2012), 857, 137-173

55

Receptors



RMSD (backbone atoms): 0.5 Å !

Holtje H. et al., *Molecular Modeling* (2008), 4, 111-178
Nurisso A. et al., *Homology Modeling: Methods and Protocols* (2012), 857, 137-173

56

Receptors

-Homology model builders

✓ Modeller (free for academia), Sybyl, MOE (not free for academia)

Modeller
Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints

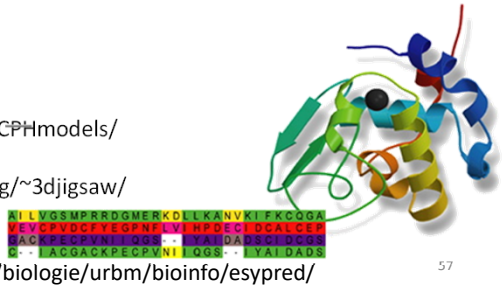
SYBYL

Tripos
A Certara Company

CHEMICAL COMPUTING GROUP
MOE
Molecular Operating Environment

✓ On-line servers (free for academia)

SWISS MODEL
http://swissmodel.expasy.org/
CPH MODEL
http://www.cbs.dtu.dk/services/CPHmodels/
3D JIGSAW
http://bmm.cancerresearchuk.org/~3djigsaw/
EASYPRED3D



http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/

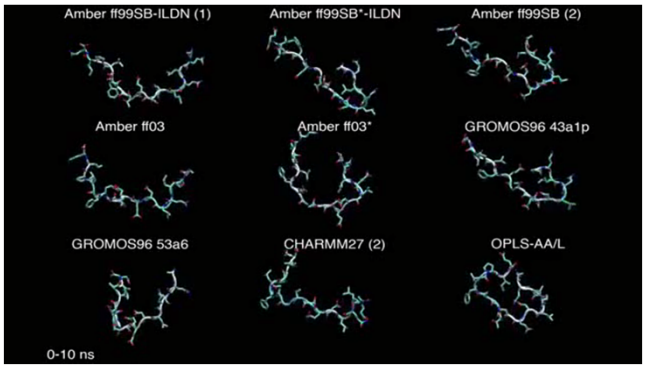
57

Receptors

-Force fields

$$V(r^n) = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2$$

$$+ \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos\left(n\phi - \gamma\right)\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_r R_{ij}}\right]$$



Cino E.A et al., *J.Chem.Theory and Comput.* (2012), 8, 2725-2740
Ponder J.W. and Case D., *Advances in Protein Chemistry* (2003), 66, 27-85

58

Carbohydrate-Receptor Complexes

'*What does glycan X look like* ?'

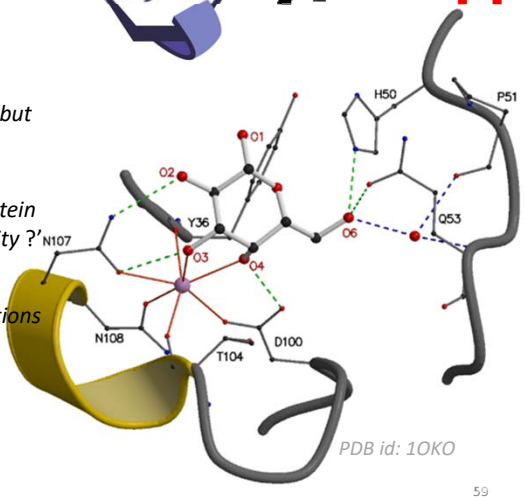
'*How does glycan X bind to protein Y* ?'

'*How does protein Y recognize glycan X but not glycan Z* ?'

'*What mutation can be made in the protein or the glycan to alter specificity or affinity* ?'

'*How reliable are the theoretical predictions* ?'

GLYCAM + AmberFF



PDB id: 1OKO

Woods R. and Tessier M.B., *Curr. Op. Struct. Biol.* (2010), 20(5), 575-583

59

Carbohydrate-Receptor Complexes

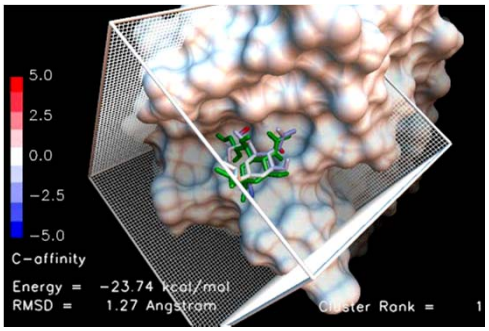
-Molecular Docking

✓ Search algorithm:

moves the ligand (flexible) into the protein pocket (rigid) and generates different ligand conformations.

✓ Scoring function:

evaluates the quality of interactions.



- Quick structural estimation of interactions
- No full flexibility
- No solvent effects
- Qualitative energy of binding

Holtje H. et al., *Molecular Modeling* (2008), 5, 181-207 WILEY-VCH
Sotriffer et al., *Virtual Screening* (2011), 857:137-73 WILEY-VCH

60

Carbohydrate-Receptor Complexes

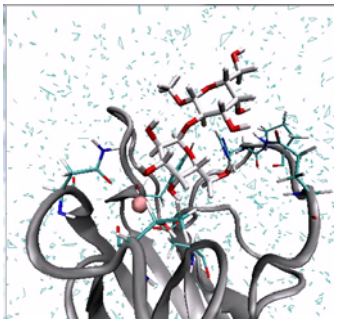
-Molecular Dynamics

✓MDs give dynamic information about protein structures over the time

✓MDs are important because biological systems are flexible and exposed to solvent effects

✓Possibility to study time-dependant phenomena, such as molecular vibrations or diffusion

✓Possibility to study temperature-dependant phenomena, such as free energies of binding



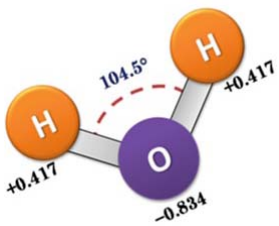
Case D. et al., AMBER 12, University of California 2012
Nurisso A. et al., *Homology Modeling: Methods and Protocols* (2012), 857, 137-173

61

Carbohydrate-Receptor Complexes

-Molecular Dynamics

✓A realistic biological system is always expected to be located in a solvated environment. Systems are embedded in box of explicit solvent molecules



✓Several water models have been developed, but one of the simplest and most widely used is the TIP3P model

Case D. et al., AMBER 12, University of California 2012
Nurisso A. et al., *Homology Modeling: Methods and Protocols* (2012), 857, 137-173

62

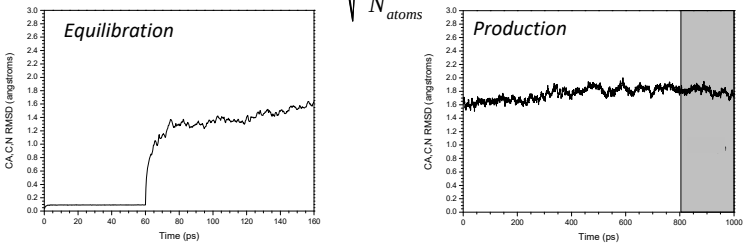
Carbohydrate-Receptor Complexes

-Molecular Dynamics

✓Generation of representative time-dependent molecular conformations (trajectories)

✓Properties calculations as a function of time. E.g. The root-mean-square deviation (RMSD), the measure of the average distance between the atoms of superimposed proteins

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} d_i^2}{N_{atoms}}}$$



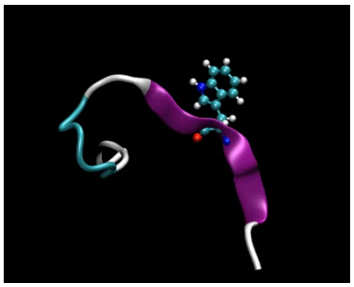
63

Nurisso A. et al., *Homology Modeling: Methods and Protocols* (2012), 857, 137-173

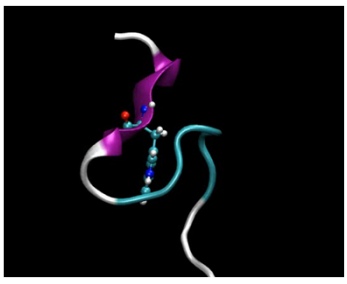
Carbohydrate-Receptor Complexes

-Molecular Dynamics

✓GPUs are processors for accelerating Molecular Dynamics calculations



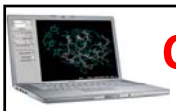
4core CPU
80 ns/day (304 atoms)



4core CPU + GPU
367 ns/day (304 atoms)

Le Grand S. et al., *Comp. Phys.Comm.*, (2013), 184, 374-380
Salomon-Ferrer R. et al., *J. Chem. Theory Comput.*, (2013), 9 (9), 3878–3888
CECAM workshop, Lausanne 2012

64



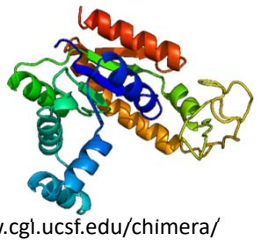
Carbohydrate-Receptor Complexes

-Molecular docking programs

AUTODOCK http://autodock.scripps.edu/ DOCK http://dock.compbio.ucsf.edu/
GOLD http://www.ccdc.cam.ac.uk/Solutions/GoldSuite/Pages/GOLD.aspx
FLAP http://www.moldiscovery.com/soft_flap.php

-Molecular Dynamics programs

DESMOND http://www.schrodinger.com/products/14/3/
NAMD http://www.ks.uiuc.edu/Research/namd/
AMBER http://ambermd.org/
GROMACS http://www.gromacs.org/
CHARMM http://www.charmm.org/



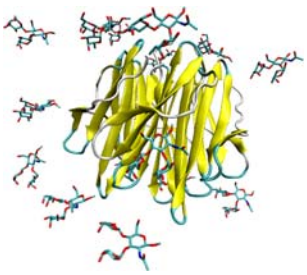
-Visualization programs

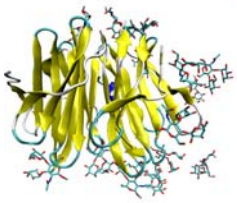
PYMOL http://www.pymol.org/ CHIMERA http://www.cgl.ucsf.edu/chimera/
VMD http://www.ks.uiuc.edu/Research/vmd/
SweetUnityMol://www.glycopedia.eu

65

A Case Study: Lectins Binding Sugars

-

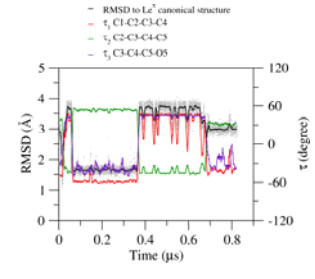


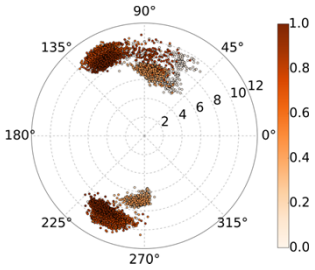


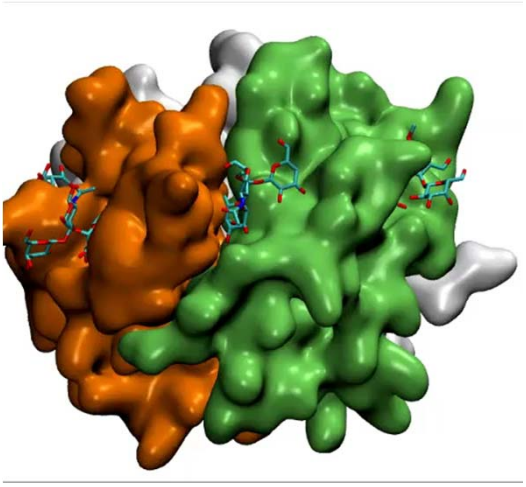
Multiple µs MD simulations of RSL in 32 Lewis x molecules

66

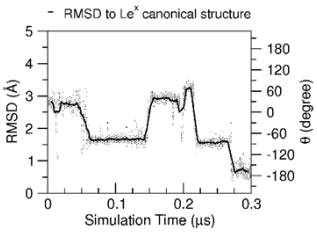
Molecular Dynamics of Lewis x in RSL

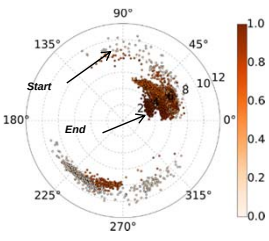


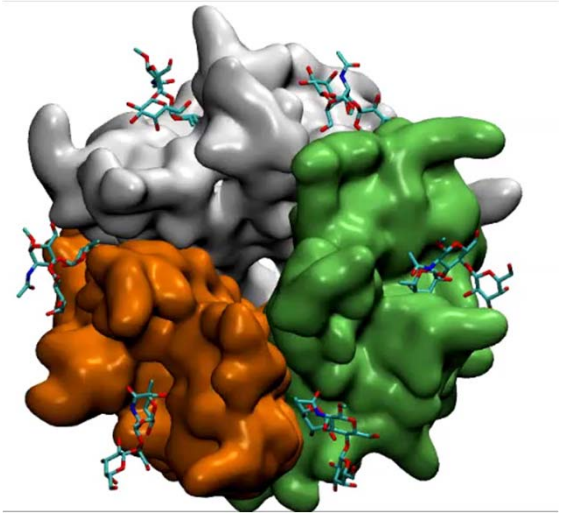




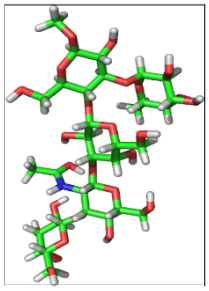
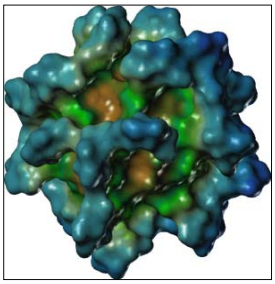
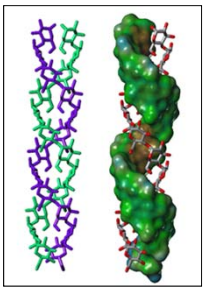
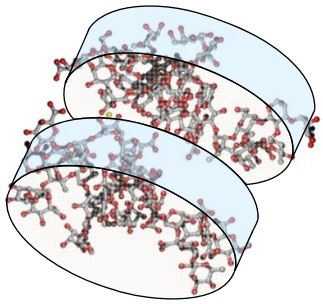
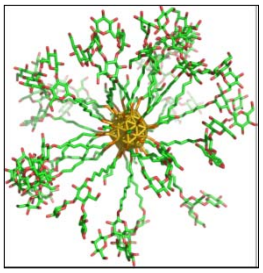
Forcing Lewis x out of Binding Site







A Wealth of Applications

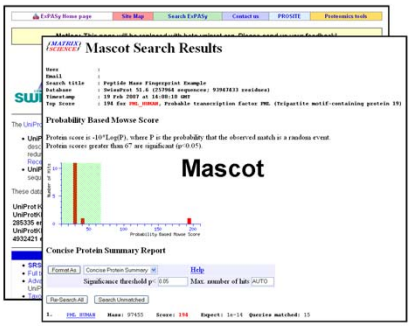
69

From Structures to 3D Databases

Glycoinformatics

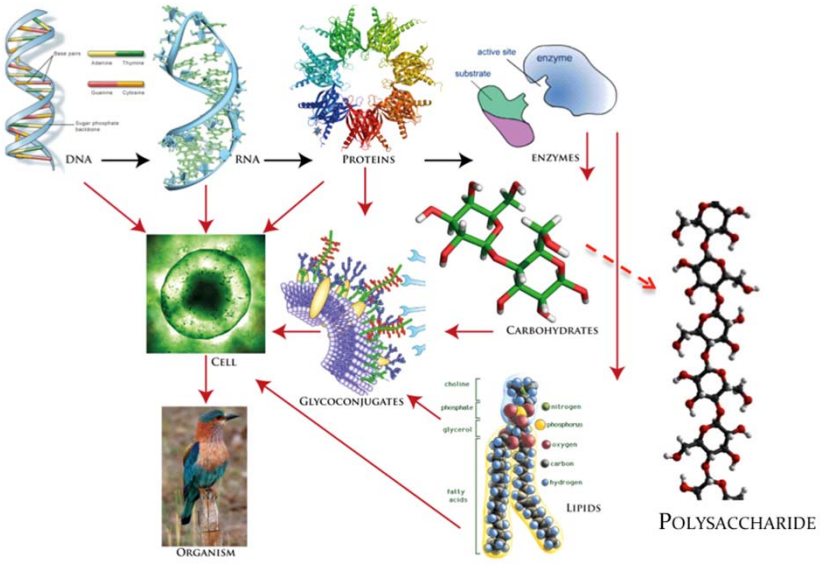
Genomics

Proteomics



Glycomics

Carbohydrates in the Scheme of the Central Dogma of Life



Challenges for Glycoinformatics

Structures as Primary Access Key

Bioinformatic

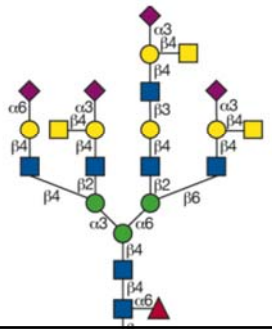
Sequences of residues

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

Glycoinformatic

Topology of Residues

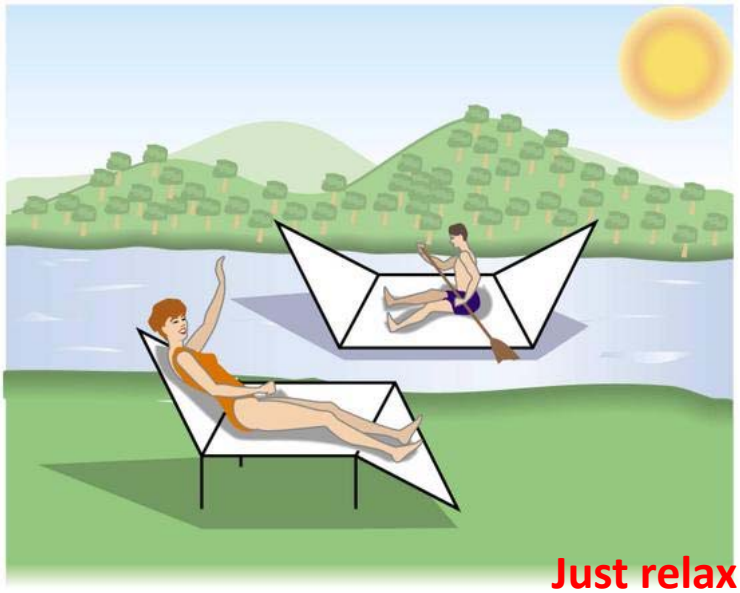


Symbol Nomenclature for Graphical Representation of Glycans (2015), *Glycobiology*, **25,** 1323-1324

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 …….

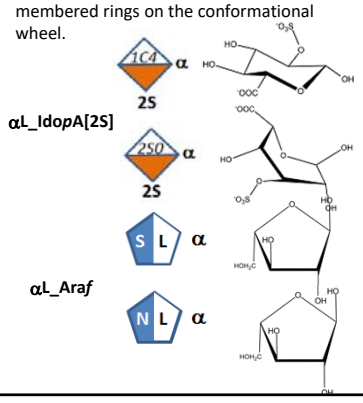
Extending the Symbolic Representation of Monosaccharides

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

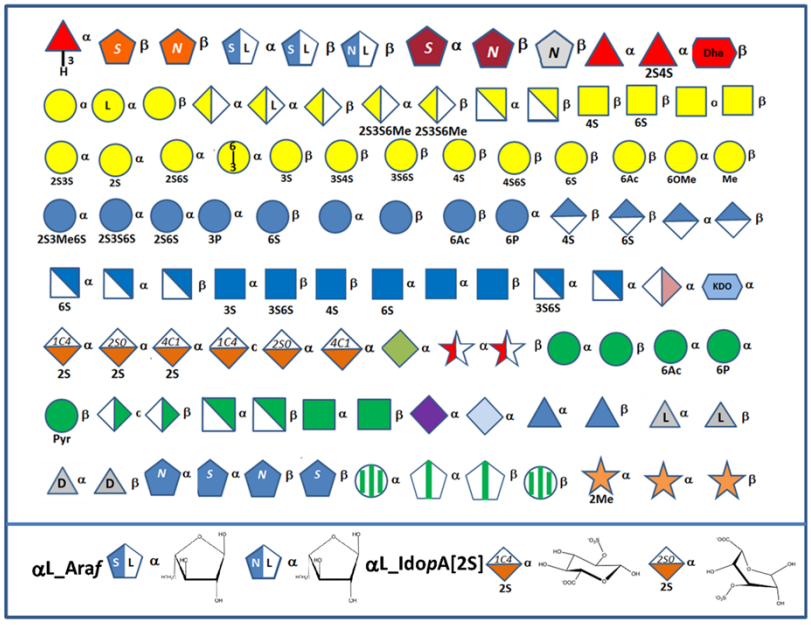
Absolute Configuration: D or L
The D-configuration for monosaccharide and the L configuration for Fucose and Idose are implicit and does not appear in the symbol. Otherwise the L configuration, is indicated in the symbol, as in the case of Arabinose or L-Galactose.
For those occurring in the furanose 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.



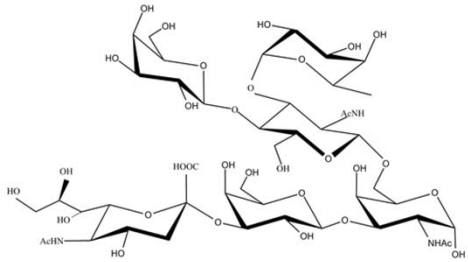
More than 150 Monosaccharides

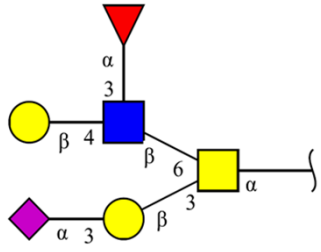


Encoding of Glycan Structures

Lewis X and Sialyl Acid on Core 2

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



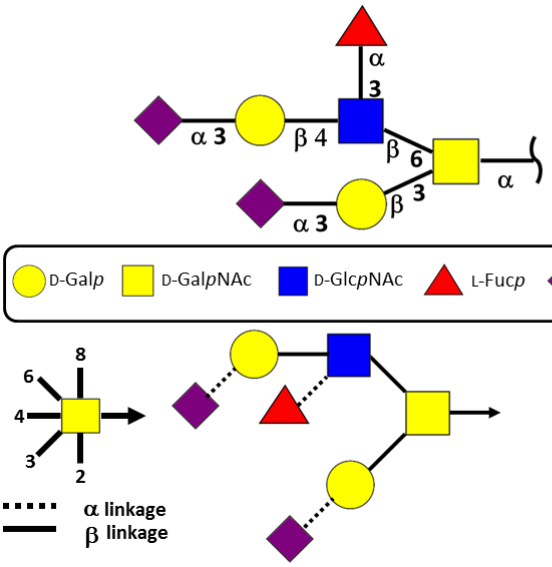


RES
1b:a-dgal-HEX-1:5
2s:n-acetyl
3b:b-dgal-HEX-1:5
4b:a-dgro-dgal-NON-2:6|1:a|2:keto|3:d
5s:n-acetyl
6b:b-dglc-HEX-1:5
7s:n-acetyl
8b:a-lgal-HEX-1:5|6:d
9b:b-dgal-HEX-1:5
LIN
1:1d(2+1)2n
2:1o(3+3)3d
3:3o(3+2)4d
4:4d(5+1)5n
5:1o(6+1)6d
6:6d(2+1)7n
7:6o(3+1)8d
8:6o(4+1)9d

GlycoCT

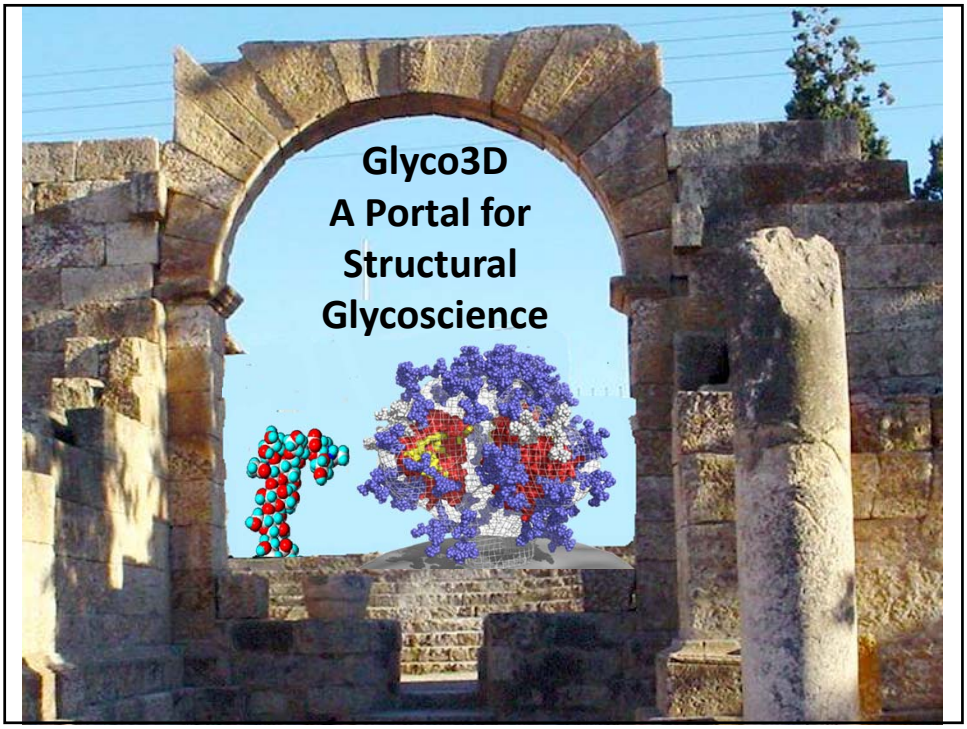
Needs for Standardizations

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

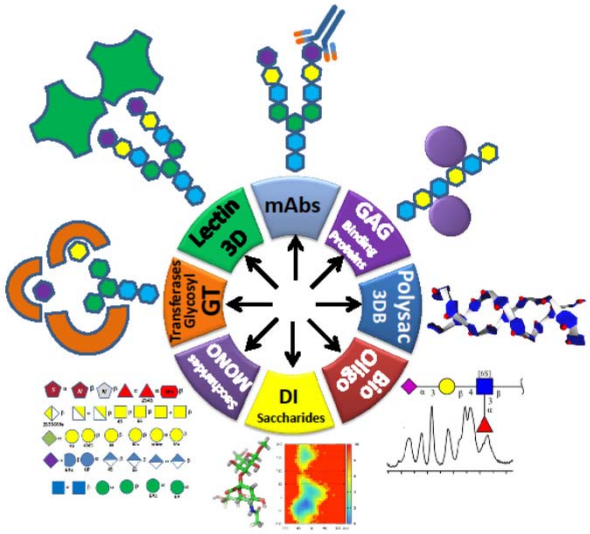


Major Glycan Structure Formats

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



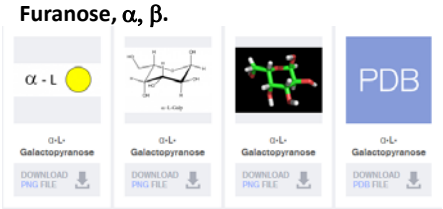
Glyco3D



Monosaccharides

Glycopedia

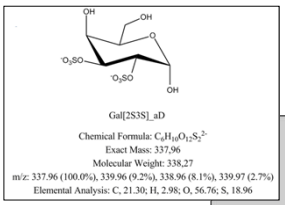
The Templates: (128 entries)
Hexoses, pentoses, ketoses, D, L, pyranose
Furanose, α, β.

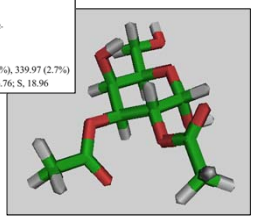


The Bioactive units: (150 entries)
Components of oligo, polysaccharides
glycans, conjugates.

Glyco3D

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

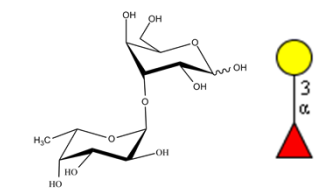


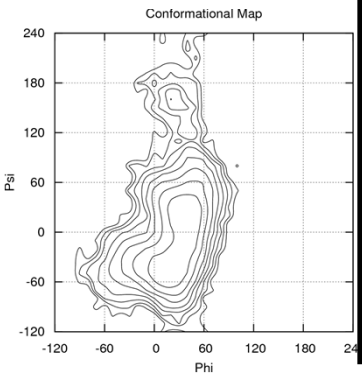


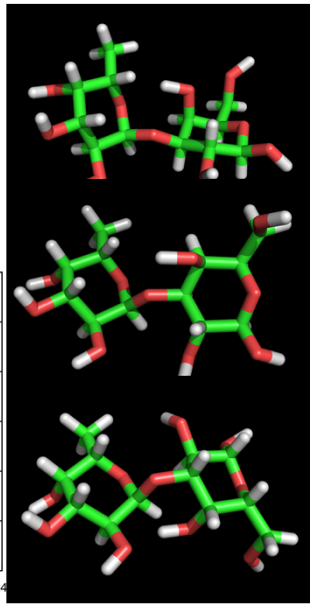
Disaccharides

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

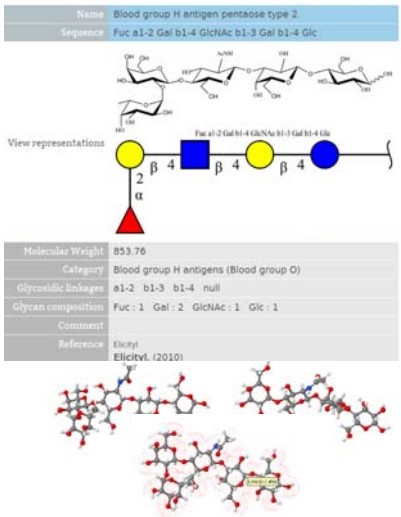


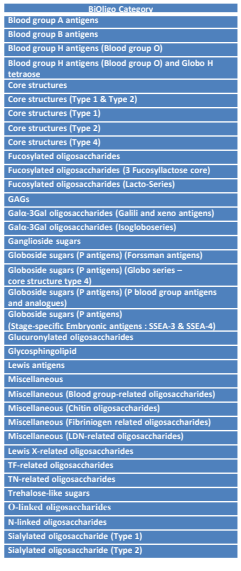




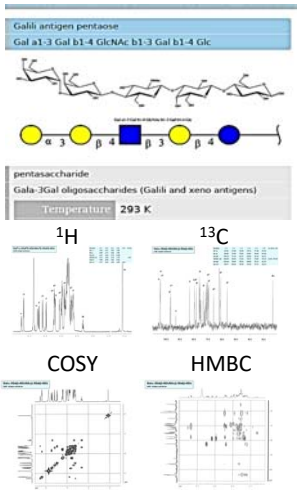
Bio-Oligosaccharides : 3D / NMR

Source: (Literature)
Tri- to octa-sacchartide
Content:
Total : 260 entries





Source: Glycan (bacterial fermentation)
Content:
Total : 150 entries

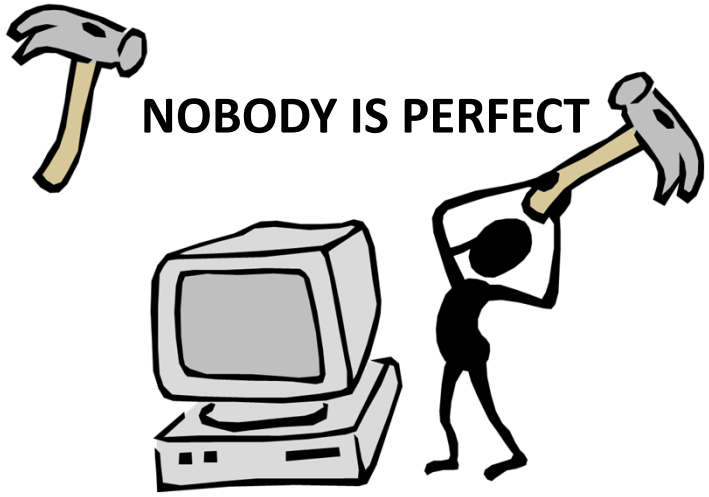


A. Sarkar, S. Drouillard, A. Rivet & S. Perez (2015) Databases of Conformations and NMR Structures of Glycan Determinants

Curations

Annotations

Curations



Annotations

Lectins

Source: X-ray - PDB

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

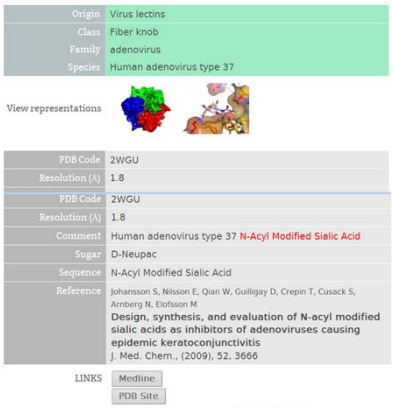
Content:
Total : 1186
Complexed sugar: 748
Free Lectins: 438
Origin : 6
Classes: 56

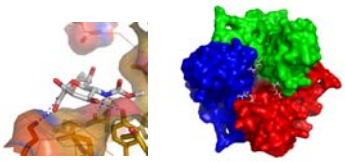
Search:
Species
Family
Sugar
PDB

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

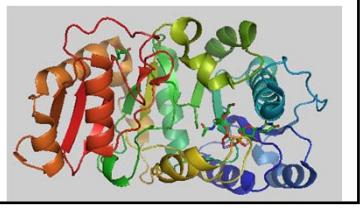
Classification of the GTs
based on their origin:
Animal, archea, bacteria, plant, virus, yeast & fungi

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)

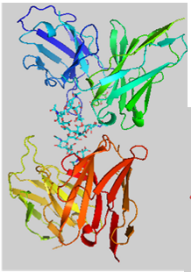
Display & Download
3D Structure (Jmol Applet)
Download PDB File
Still Image
Download Image

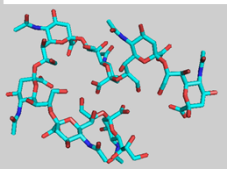


Monoclonal Antibodies / GAG Binding Proteins

Source: X-ray - PDB

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



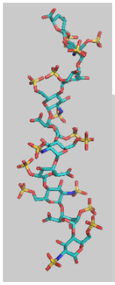


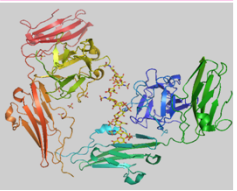
Molecule Information
Class
Family
Species
View representation
PDB Code
Resolution
Sequence
Reference
Links (Medline, PDB, Swiss Prot.)

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

Source: X-ray - PDB

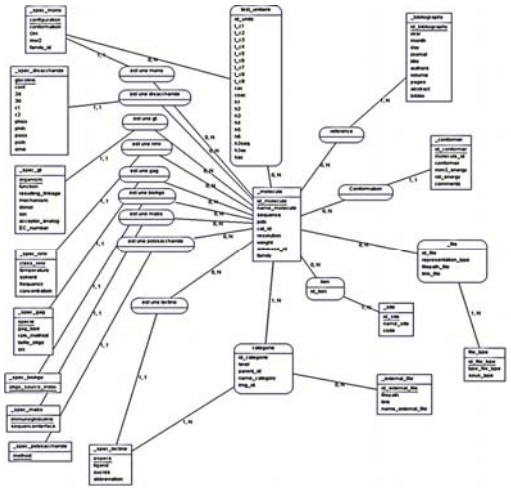
Content:
Total : 46
Classification of GAG binding proteins
Chemokine
Complement proteins
ECM proteins
Enzymes
Growth factors
Lectins
Toxins
Virus
Search:
Family, PDB
Structure of GAG





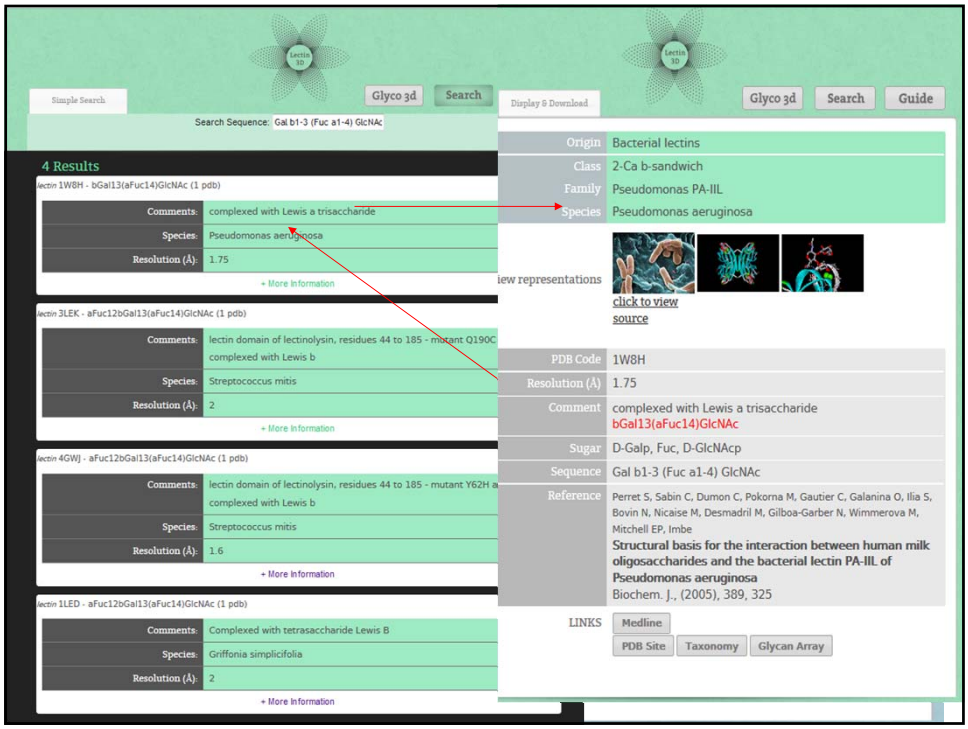
Informatics Implementation

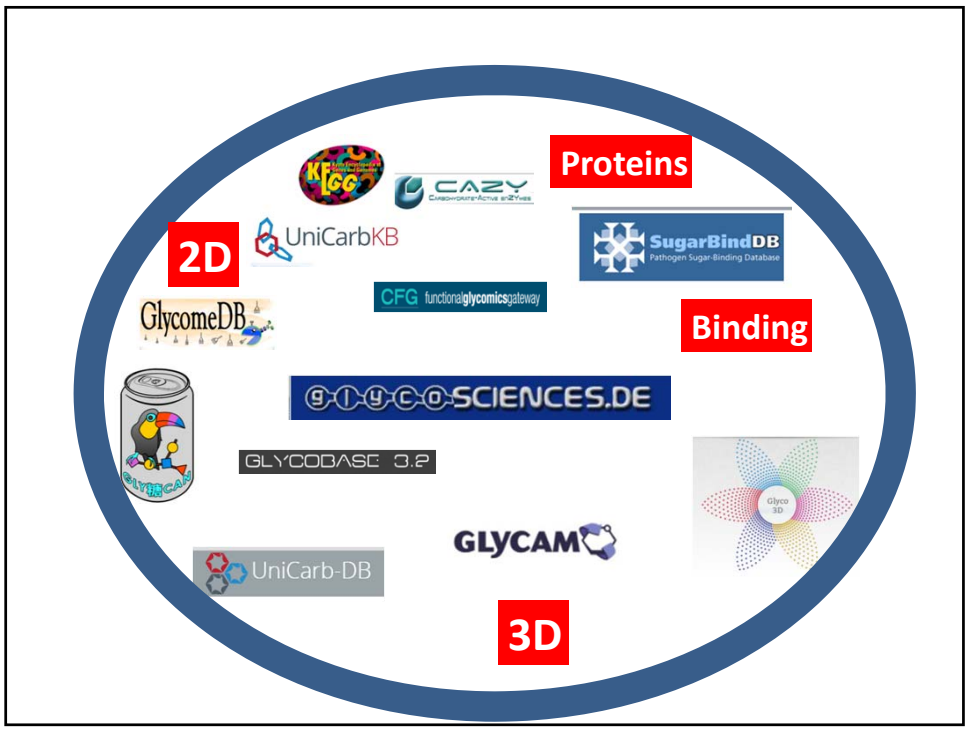
Relational Data Base: Language : PHP 5.4, DBase : MySql 5.5.24

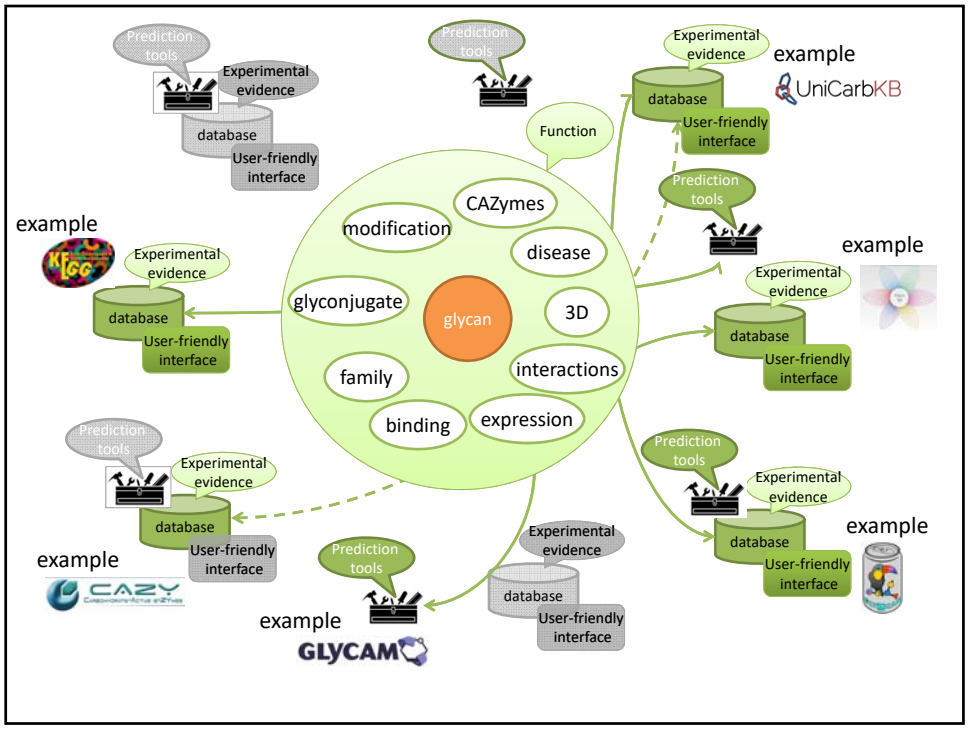


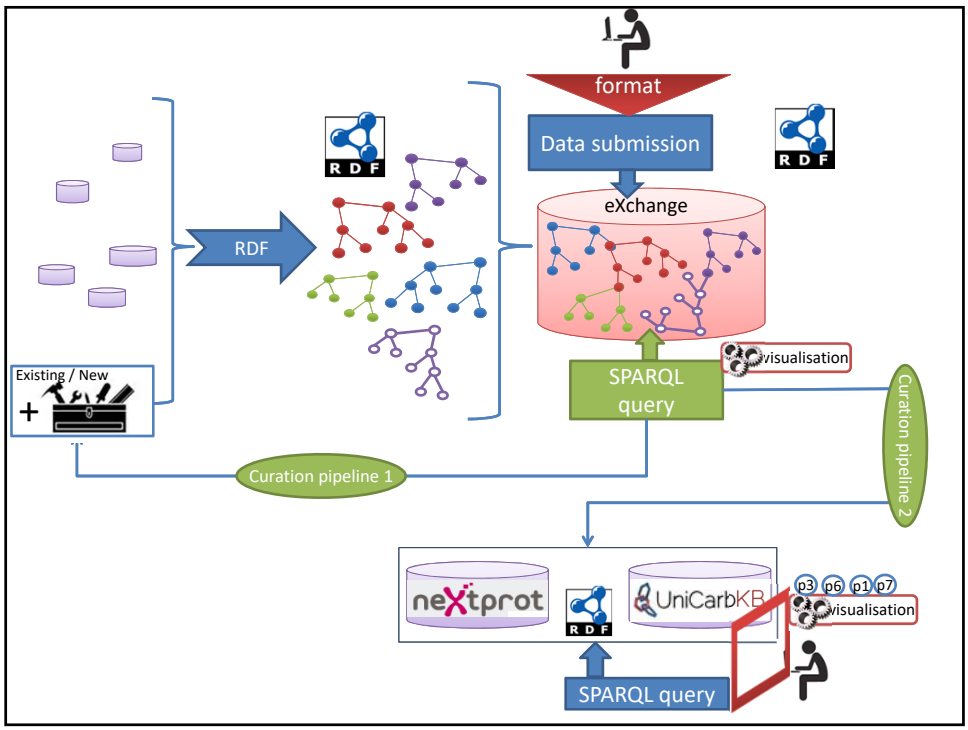
Development Environment

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


Glyco 3d
Search
Guide
Simple Search
Search Sequence: Gal b1-3 (Fuc a1-4) GlcNAc
4 Results
lectin 1W8H - bGal13(aFuc14)GlcNAc (1 pdb)
Comments: complexed with Lewis a trisaccharide
Species: Pseudomonas aeruginosa
Resolution (Å): 1.75
+ More Information
lectin 3LEK - aFuc12bGal13(aFuc14)GlcNAc (1 pdb)
Comments: lectin domain of lectinolysin, residues 44 to 185 - mutant Q190C complexed with Lewis b
Species: Streptococcus mitis
Resolution (Å): 2
+ More Information
lectin 4GWJ - aFuc12bGal13(aFuc14)GlcNAc (1 pdb)
Comments: lectin domain of lectinolysin, residues 44 to 185 - mutant Y62H complexed with Lewis b
Species: Streptococcus mitis
Resolution (Å): 1.6
+ More Information
lectin 1LED - aFuc12bGal13(aFuc14)GlcNAc (1 pdb)
Comments: Complexed with tetrasaccharide Lewis B
Species: Griffonia simplicifolia
Resolution (Å): 2
+ More Information
Origin: Bacterial lectins
Class: 2-Ca b-sandwich
Family: Pseudomonas PA-IIL
Species: Pseudomonas aeruginosa
click to view source
PDB Code: 1W8H
Resolution (Å): 1.75
Comment: complexed with Lewis a trisaccharide bGal13(aFuc14)GlcNAc
Sugar: D-Galp, Fuc, D-GlcNAcp
Sequence: Gal b1-3 (Fuc a1-4) GlcNAc
Reference: Perret S, Sabin C, Dumon C, Pokorna M, Gautier C, Galanina O, Ilia S, Bovin N, Nicaise M, Desmadril M, Gilboa-Garber N, Wimmerova M, Mitchell EP, Imbe
Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of Pseudomonas aeruginosa
Biochem. J., (2005), 389, 325
LINKS
Medline
PDB Site
Taxonomy
Glycan Array


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

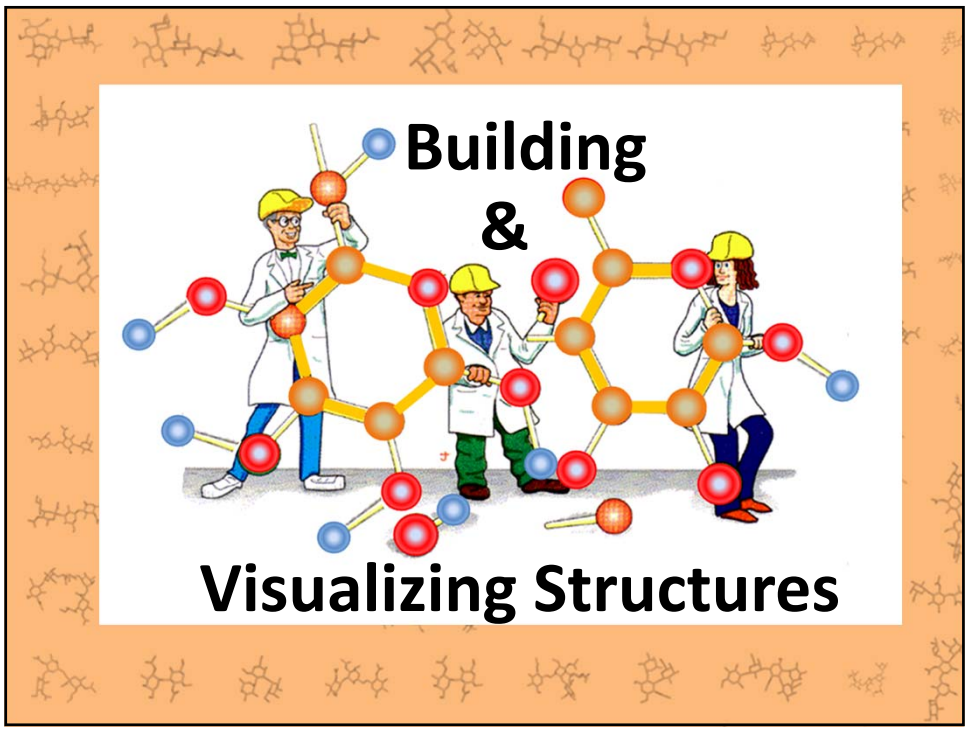

Prediction tools
Experimental evidence
database
User-friendly interface
example
UniCarbKB
Function
CAZymes
modification
disease
glyconjugate
glycan
3D
family
interactions
binding
expression
KEGG
CAZY
GLYCAM


format
Data submission
RDF
eXchange
Existing / New
RDF
visualisation
SPARQL query
Curation pipeline 1
Curation pipeline 2
neXtprot
UniCarbKB
p3 p6 p1 p7
visualisation
SPARQL query

The Semantic Web

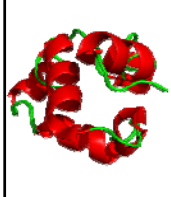
The semantic web is an extension of the web which promotes common data formats and exchange protocols –**Ressource Description Network (RDF)** to provide a common framework that allows data to be shared and reused across applications and community boundarie.

An **ontology** is a formal naming and definition of the types, properties, and interrelationships of the entities that relly or fundamentally exist for a particular domain. Ontologies are created to limit complexity and to organize information.



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, …)

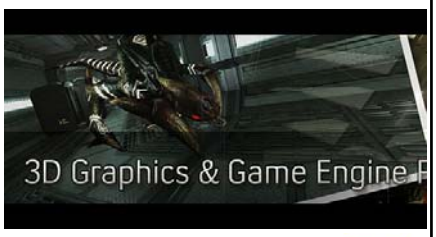
Compatible with accepted pictorial representations used in carbohydrate chemistry, biochemistry and glycobiology and structural biology format (pdb).

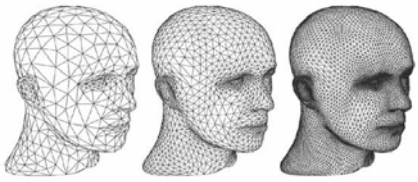
Production of publication-quality figures.

Open Access / No steep learning curve

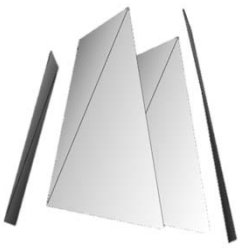
Multiple platforms i.e.Windows, MacOS and Linux operating systems, web pages, …..

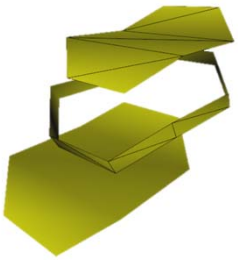
From Game Engine to Macromolecular Graphics





Tessalation

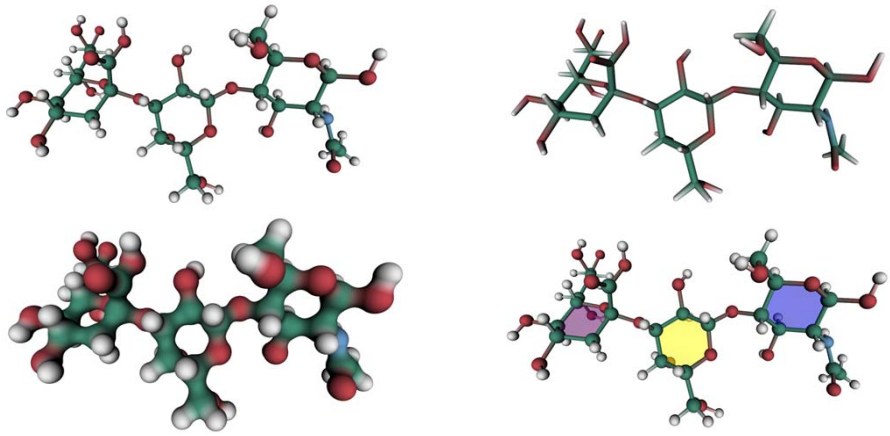




Unity3D provides an optimized set of graphical primitives for rendering.
We use triangulated spheres, triangulated cubes and lines. - mesh

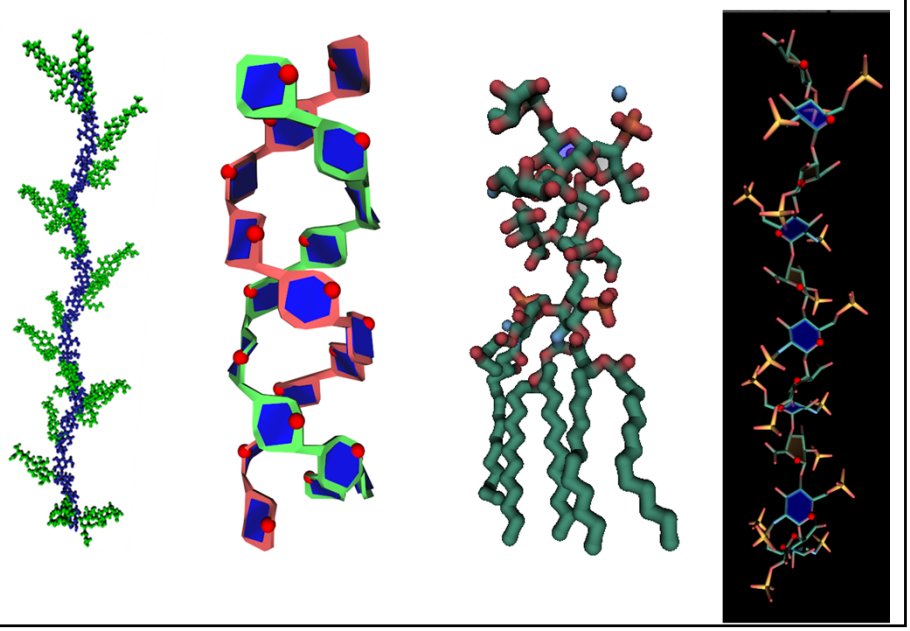
Sweet Unity Mol: *Ring Blending*

Neu5Ac α2-3 Gal β1-4 GlcNAc

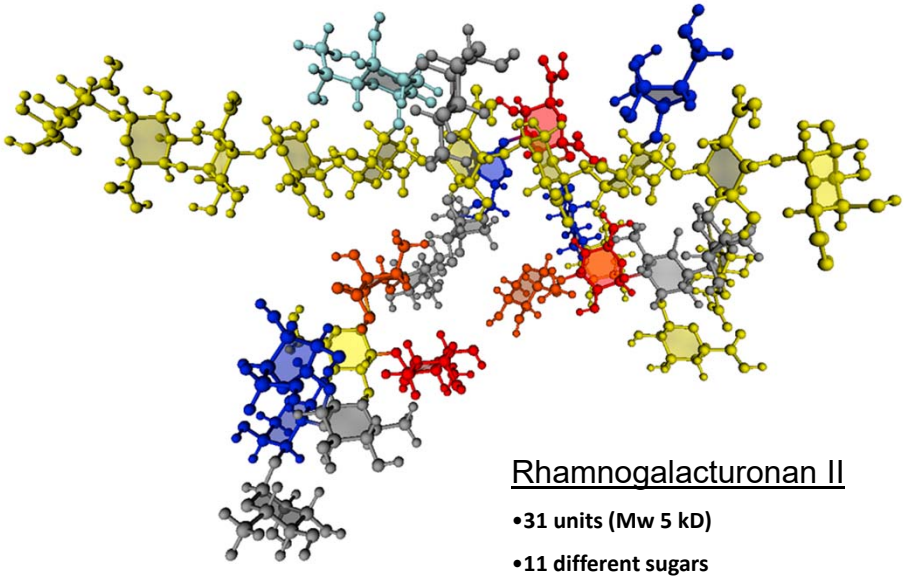


HyperBalls: is composed of spheres depicting atoms linked by hyperboloid primitives rather than simple cylinders

Polysaccharides



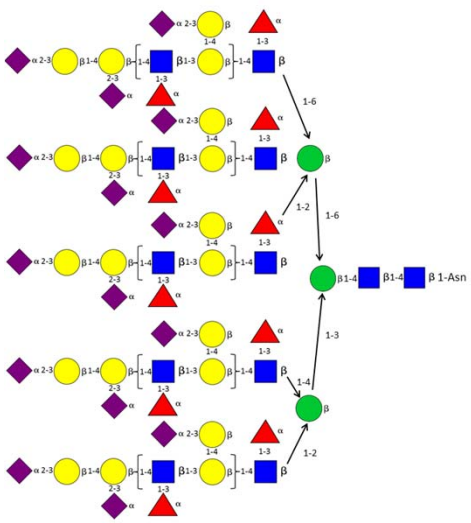
Complex Oligosaccharides & Glycans



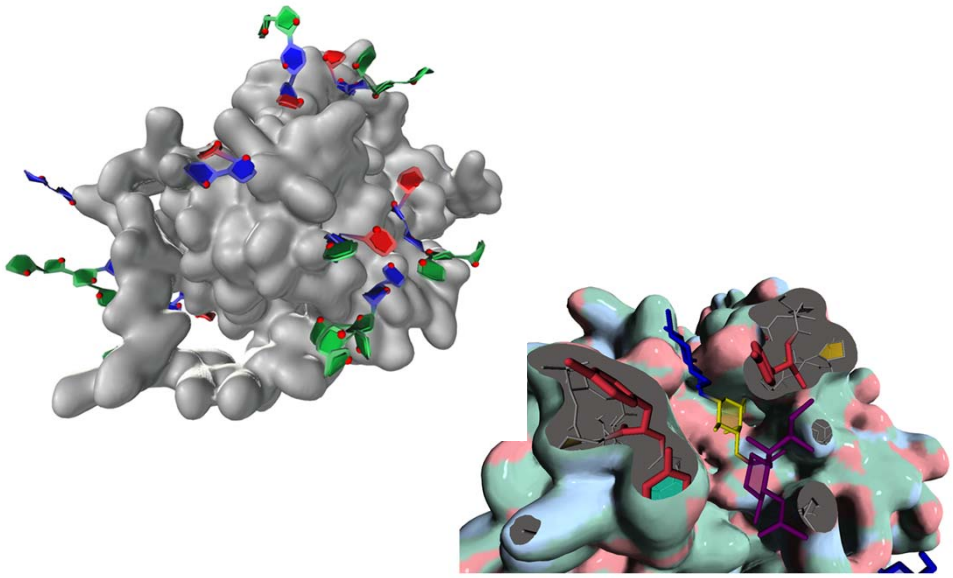
Rhamnogalacturonan II

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

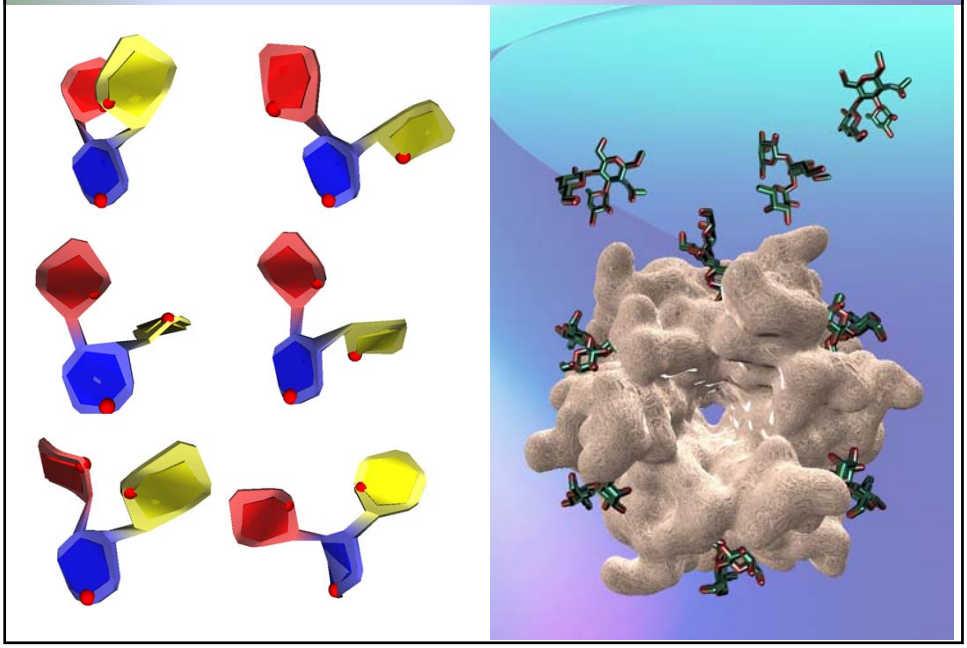
Complex Oligosaccharides & Glycans



Glycoproteins – Protein Carbohydrate Interactions

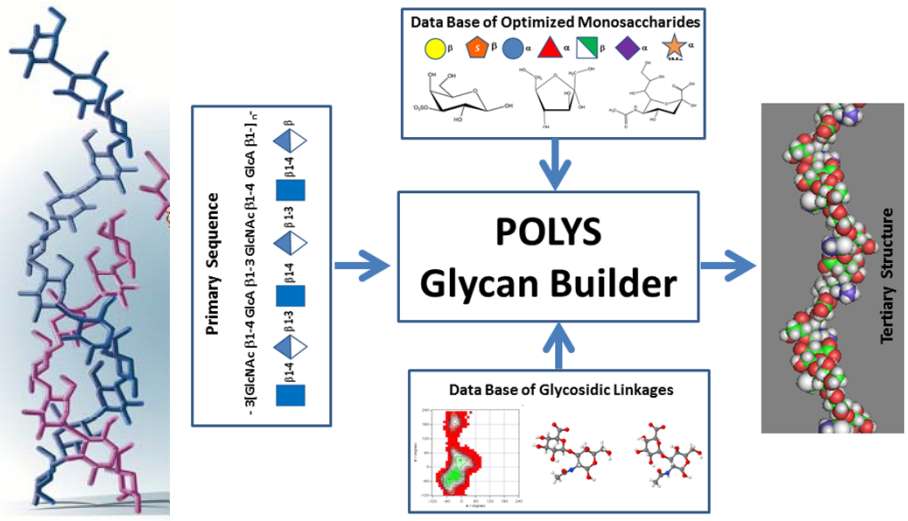


The Hidden Conformations of LewisX

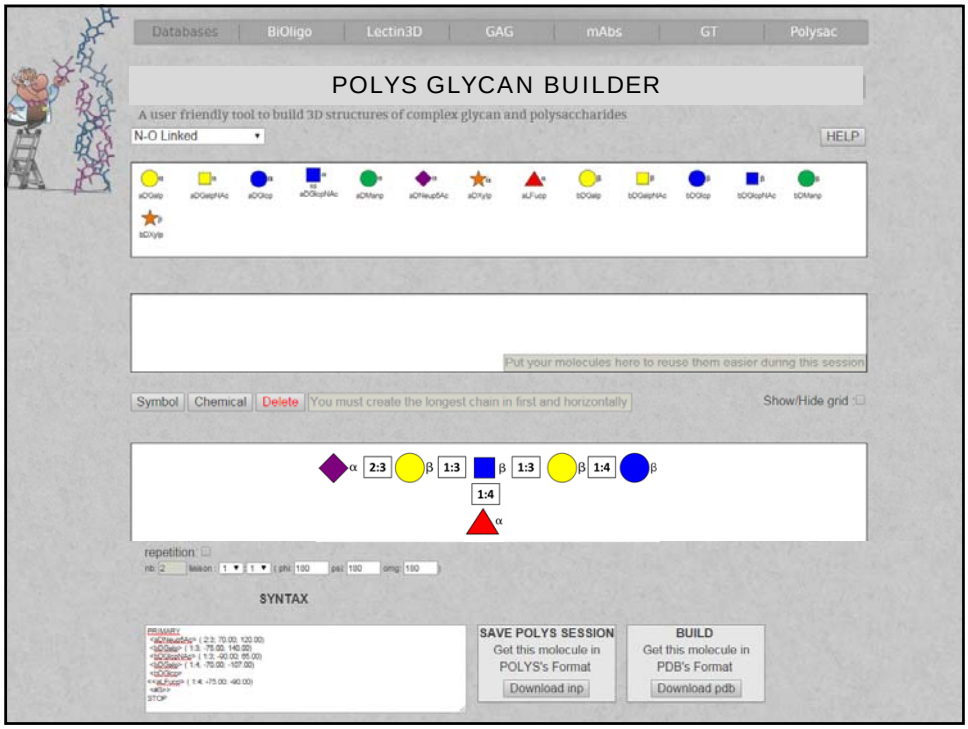


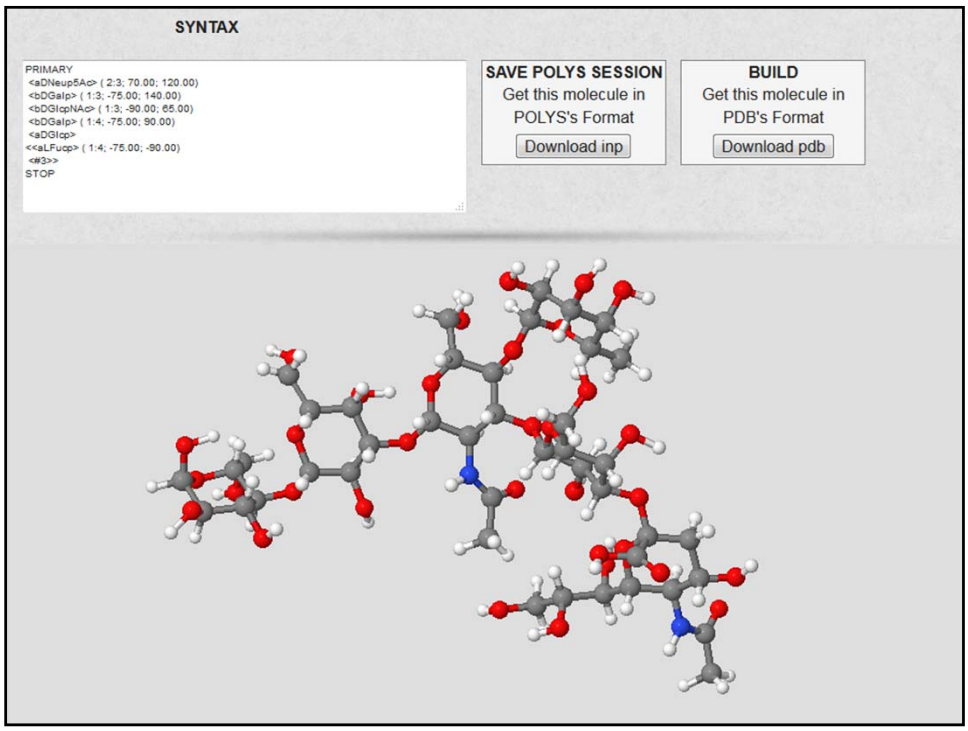
We Can Build Your Glycans

We Can Build Your Glycans: POLYS 2.0



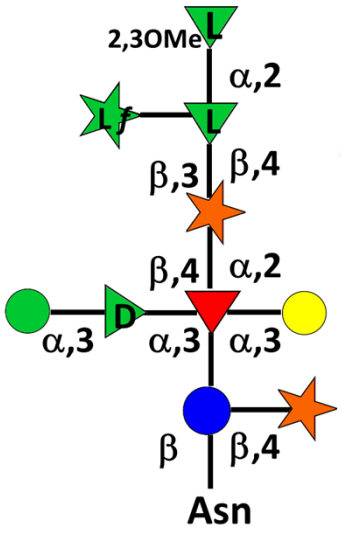
S.Engelsen, P.I Hansen & S. Perez (2013) An Open Source Software Package for Building 3-D Structures of Polysaccharides, *Biopolymers.*

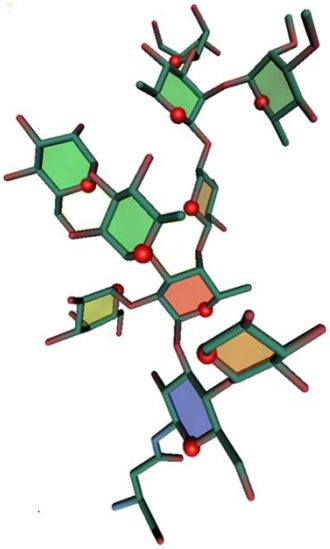

Databases
BiOligo
Lectin3D
GAG
mAbs
GT
Polysac
POLYS GLYCAN BUILDER
A user friendly tool to build 3D structures of complex glycan and polysaccharides
N-O Linked
HELP
Put your molecules here to reuse them easier during this session
Symbol
Chemical
Delete
You must create the longest chain in first and 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

From 1D to 3D Representations

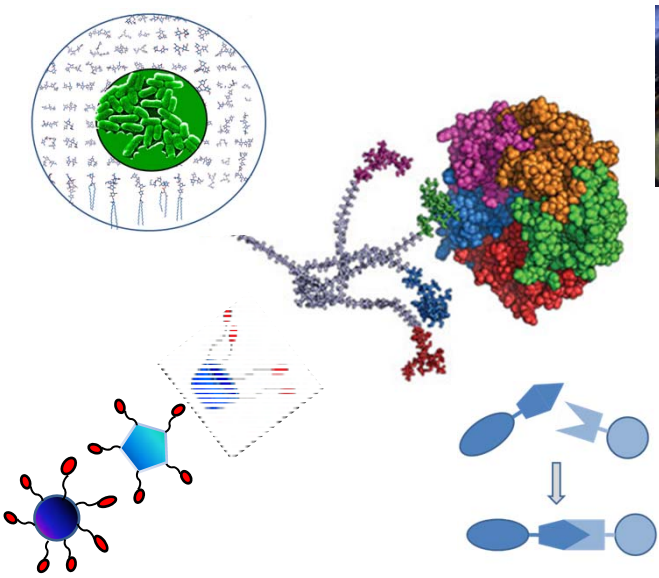
D-Man*p*
D-Glc*p*
D-Gal*p*
L-Fuc*p*
D-Xyl*p*
L-Rha*p*
D-Rha*p*
L-Ara*f*





3 Dimensional Thinking in Glycosciences

An iconic case…



Multivalency at work on Cholera Toxin.
Five **GM1**os moieties linked to a **Calix[5]arene s**caffold. Picomolar inhibition potency (**IC50 = 450 pM**) for Cholera Toxin B. Multivalency effect, with a relative inhibitory potency of **100 000** compared to a monovalent GM1os derivative

J. Garcia-Hartjes, S. Bernardi, C.A. G. M. Weijers, T. Wennekes, M. Gilbert, F. Sansone,A. Casnati & H. Zuilhof*, *Org. Biomol. Chem.*, 2013,