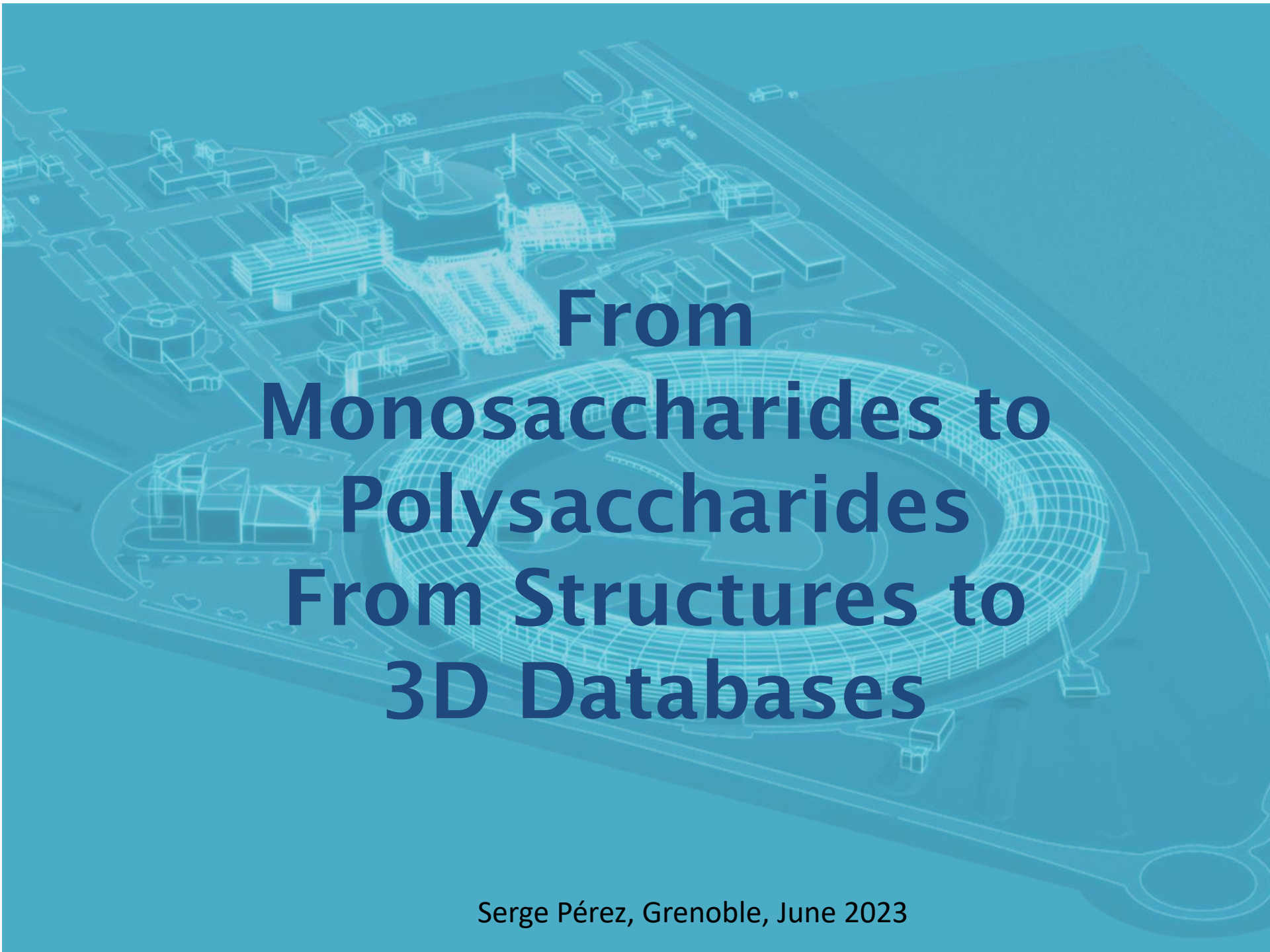


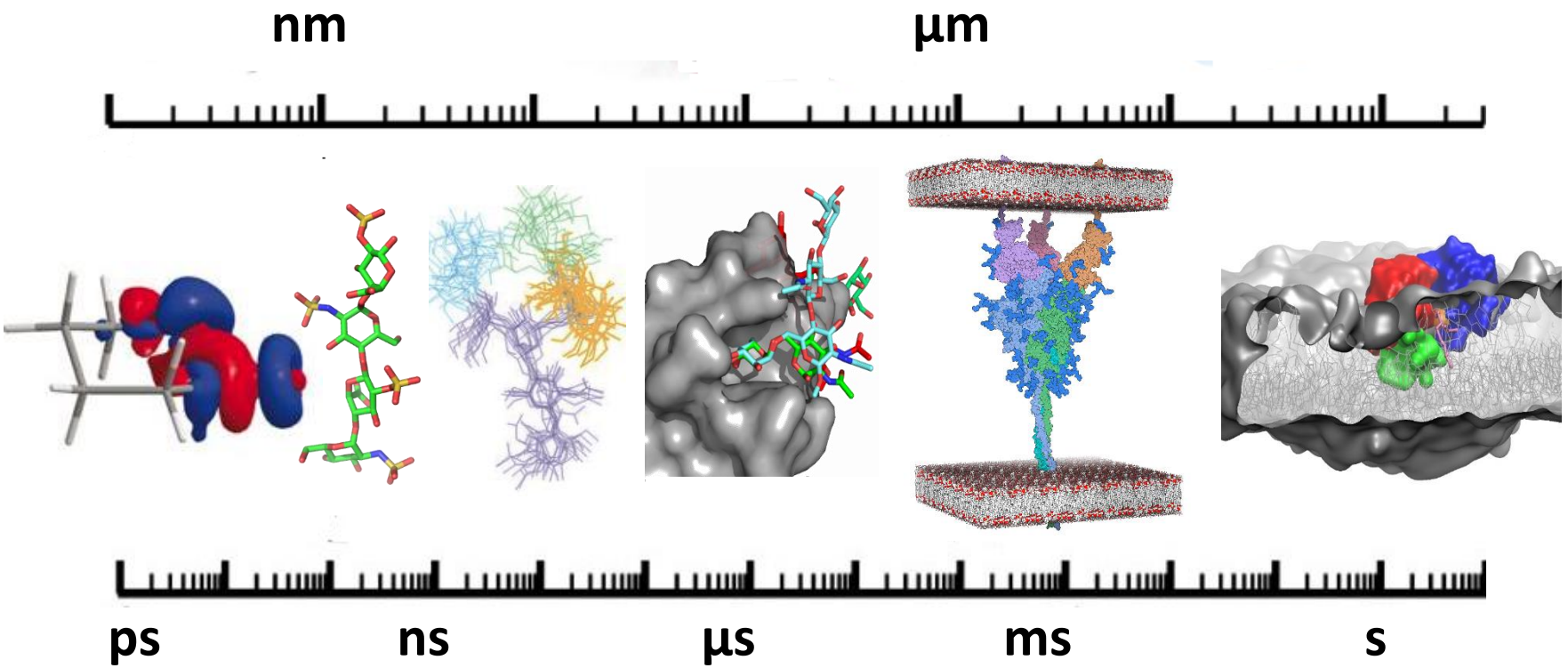
From Monosaccharides to Polysaccharides From Structures to 3D Databases

Serge Pérez, Grenoble, June 2023

Structural Glycobiology

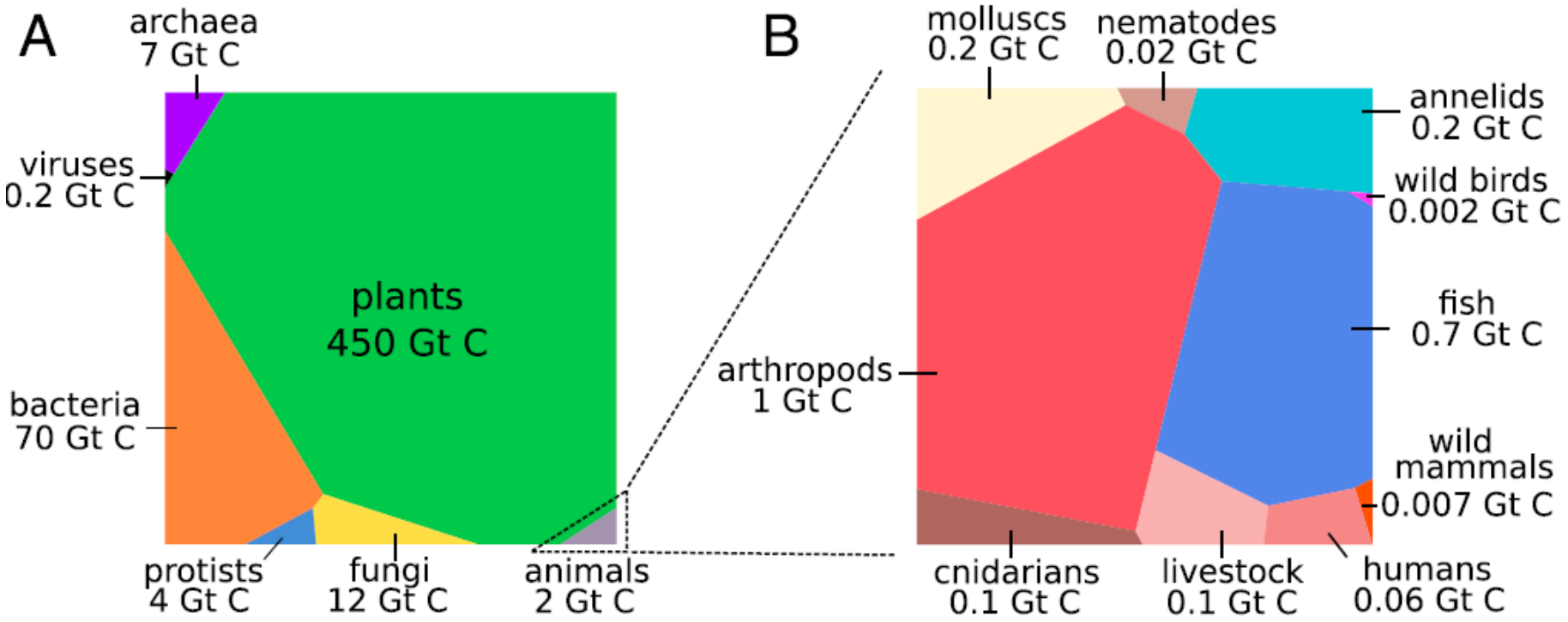
structural glycobiology is the study of how complex glycans are built.

A variety of imaging methods are used, to view molecules in three dimensions to see how they are assembled, how they function, and how they interact.



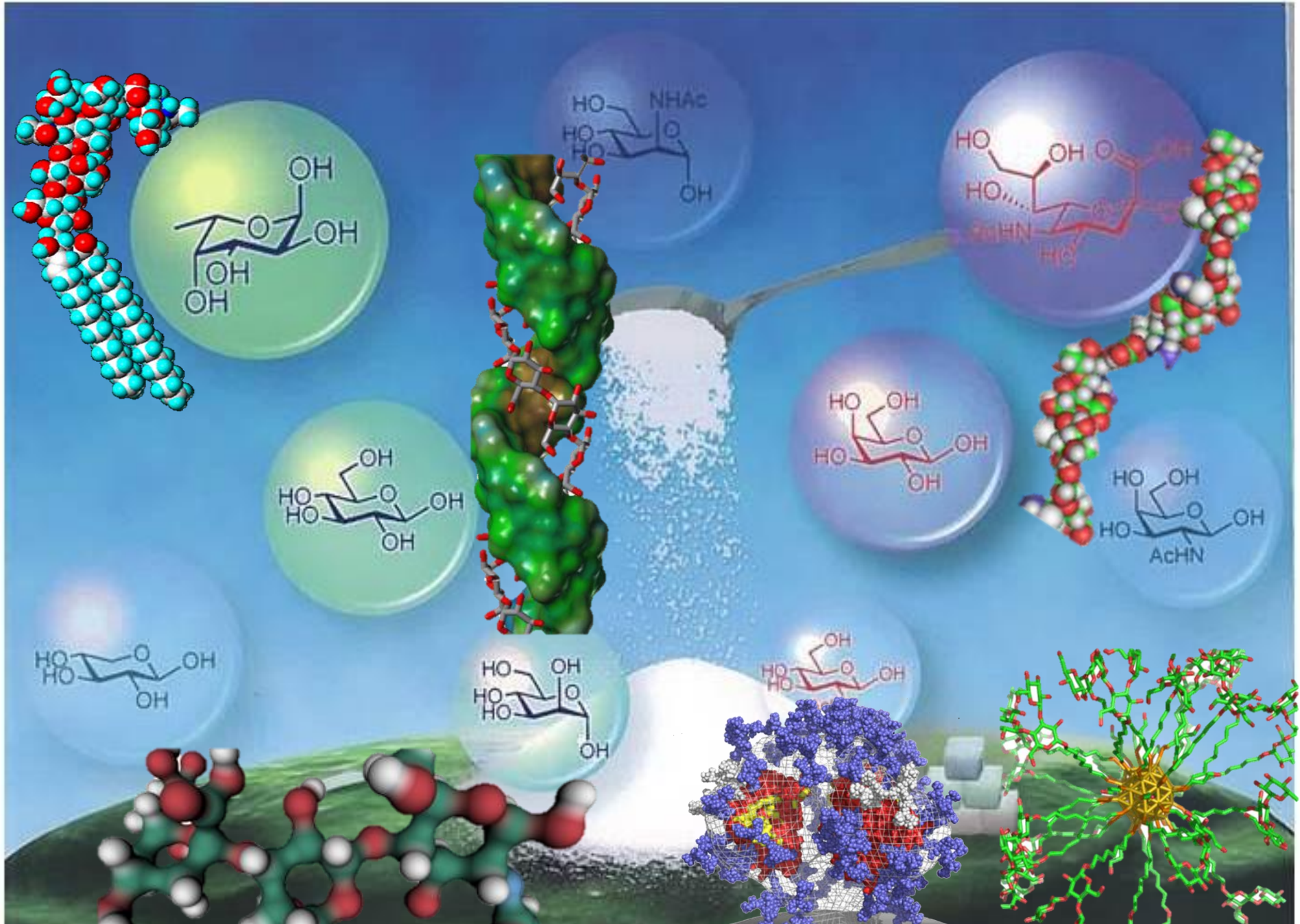
Where are such complex glycans found ?

The Global Biomass Distribution (Gigatons Carbon)

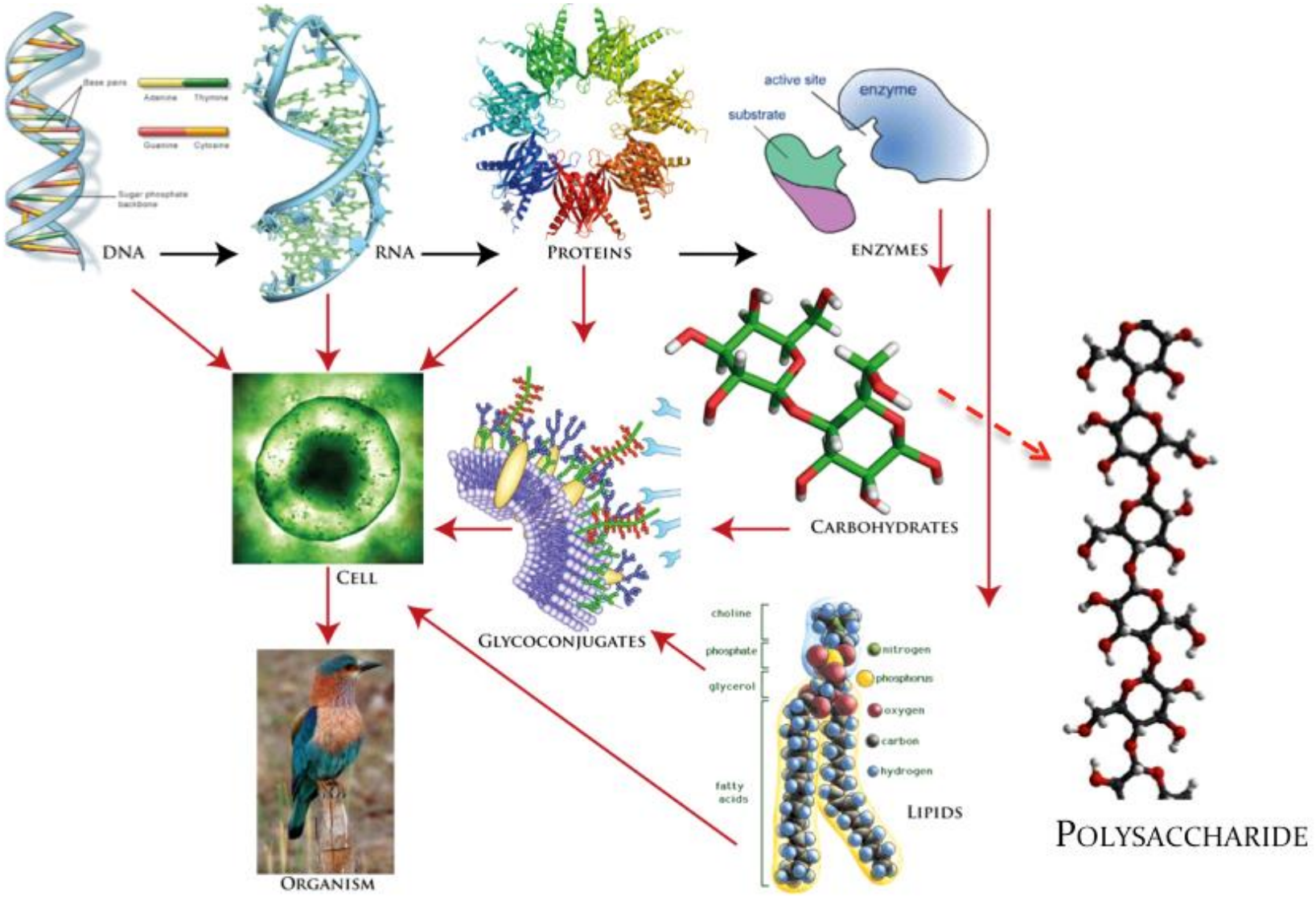


4 Gt C fossil oil extracted / year

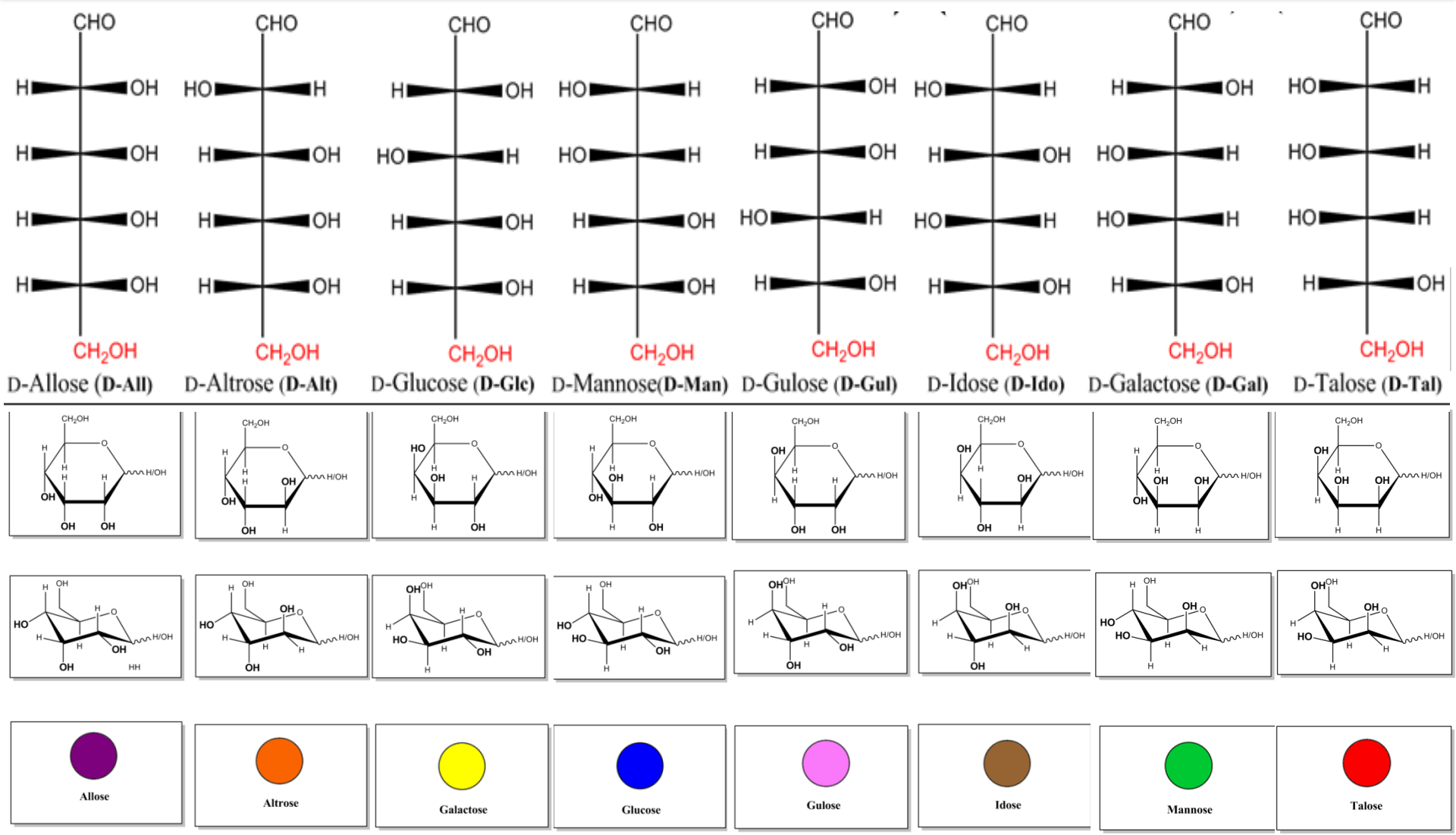
The Biomass distribution on Earth, Y.M. Bar-On, R. Philips & R. Milo, PNAS 2018


OH
O
OH
OH
OH
HO
NHAc
HO
HO
O
OH
HO
OH
O
HO
HN
HO
HO
OH
O
OH
HO
HO
OH
HO
OH
O
OH
HO
HO
OH
HO
OH
O
OH
HO
AcHN
HO
O
OH
HO
HO
OH
HO
OH
O
OH
HO
HO
OH

Carbohydrates in the Scheme of the Central Dogma of Life



Evolution of the Depiction of Monosaccharides



Fischer assigned the dextrorotatory glucose (via glucaric acid) the projection with the OH group at C5 pointing to the right. But the absolute configuration was established in 1951 (Bijvoet) by X-ray crystallography

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

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,

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

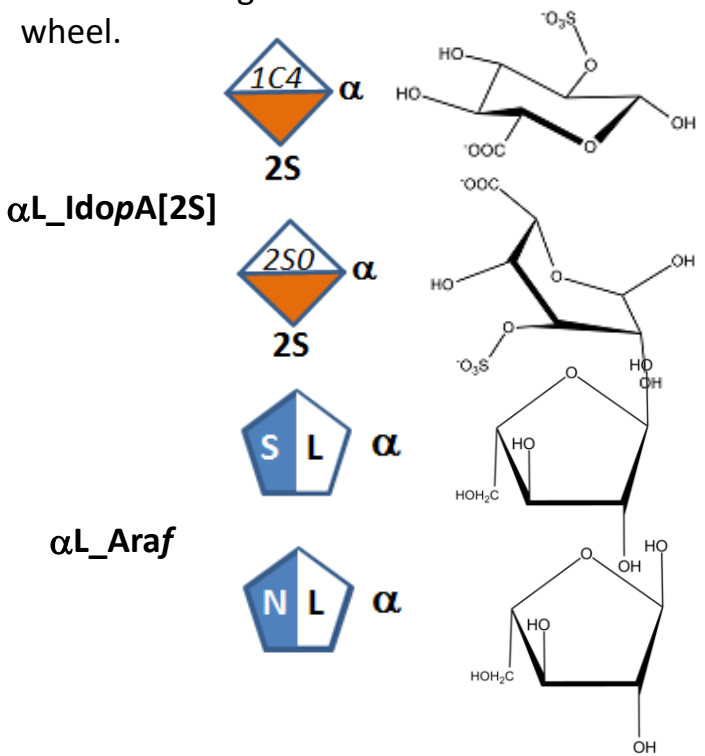
The D-configuration for monosaccharide and the L configuration for Fucose and Idose are implicit and does not appear in the symbol. Otherwise the L configuration, is indicated in the symbol, as in the case of Arabinose or L-Galactose.
For those occurring in the furanose form, a letter *N* or *S* is inserted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

Anomeric Configuration.

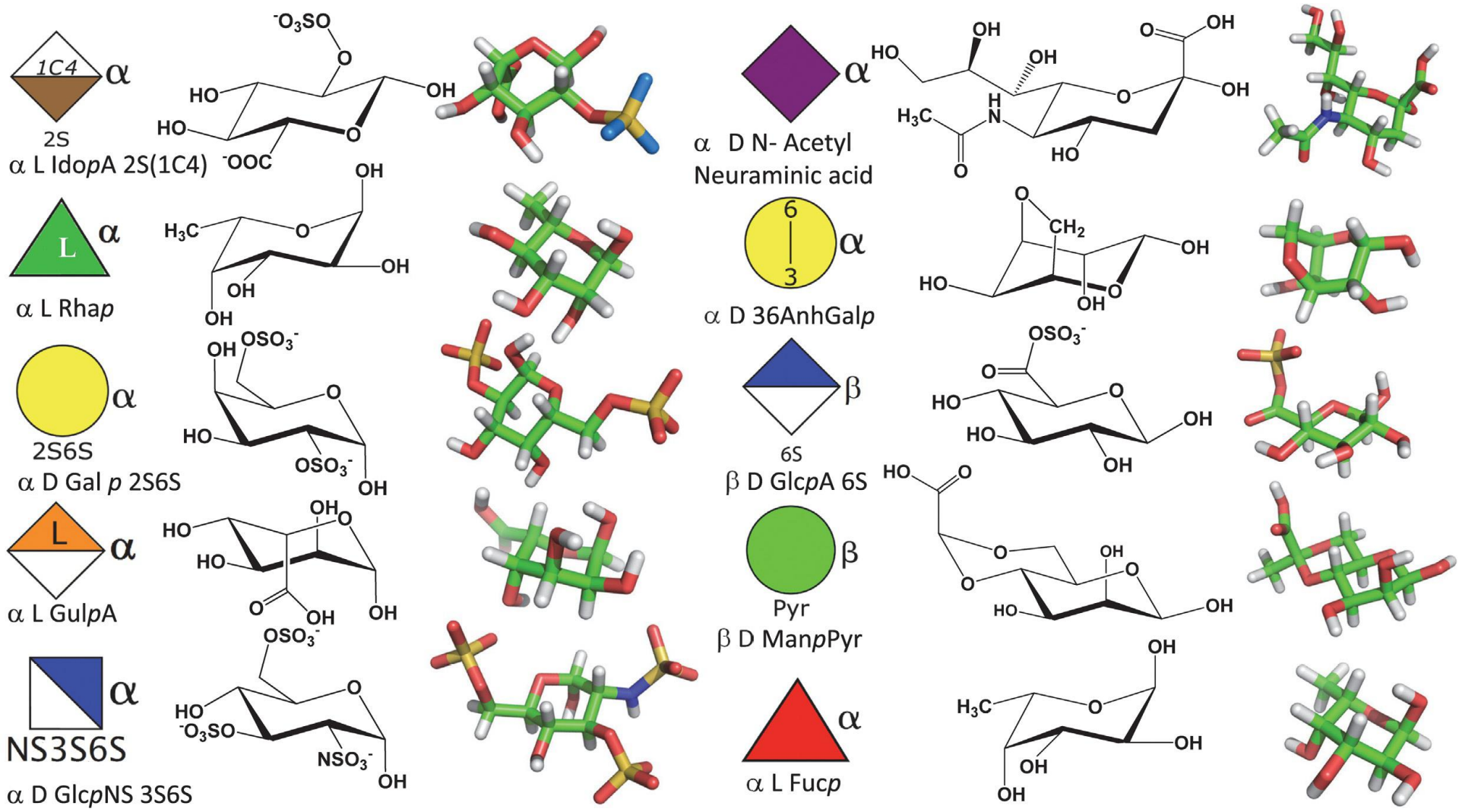
The nature of the glycosidic configuration (α or β) is explicitly set within the symbol.

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.



From Symbol Representation to 3D-Structures

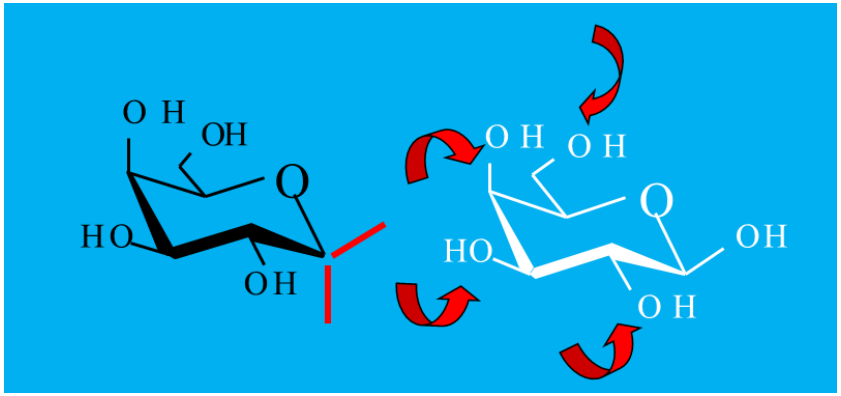


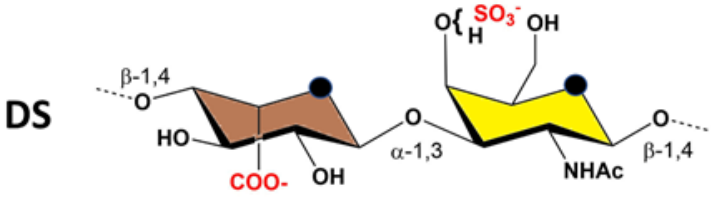
Disaccharides & Higher Oligosaccharides

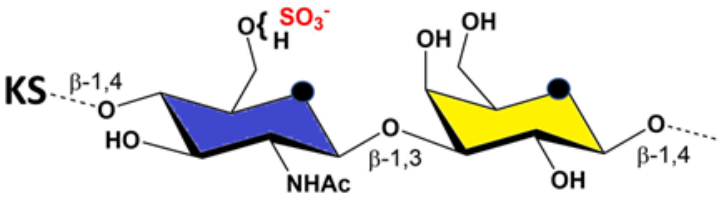
•Have a very high number of monomers (substitution…).

• Have many different ways of connecting monomers.

• Have branching points.







All chemical compounds are described with IUPAC, Simplified Molecular Input Line Entry Specification syntax (SMILES), and InChi encodings that are readable by the vast majority of chemo-informatics tools.

Glycans are encoded in GlycoCT,
WURCS (Web3 Unique Representation of Carbohydrate Structures)
LINUCS (LInear Notation for Unique description of Carbohydrate Sequences).

From Monosaccharides to Polysaccharides Through Crystallography

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

Neutrons are scattered by the atom nucLei.

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

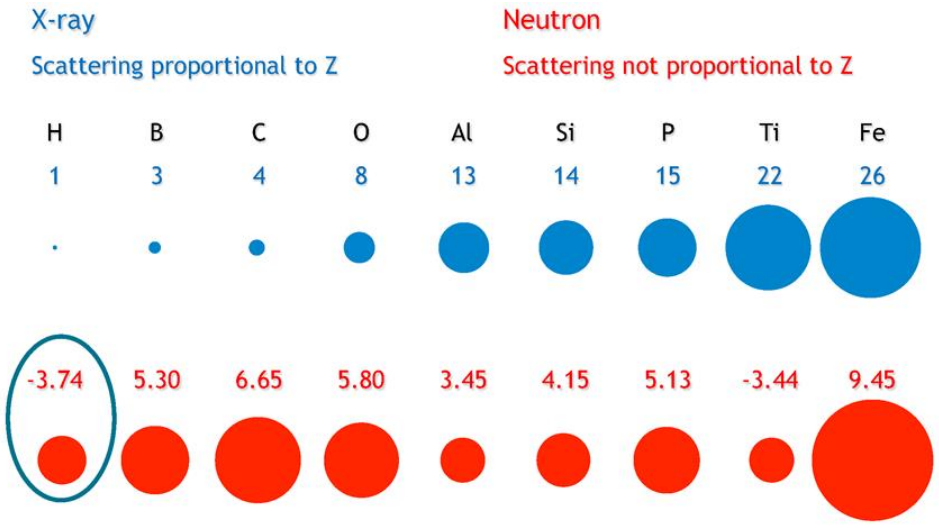
Molecular & Crystal Structures of Carbohydrates

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

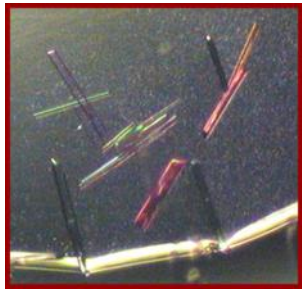


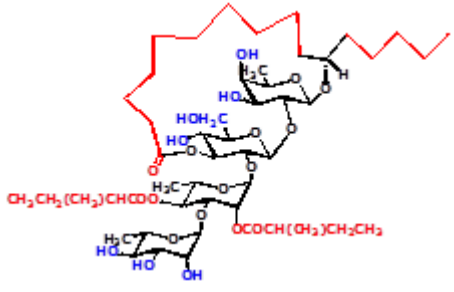
.

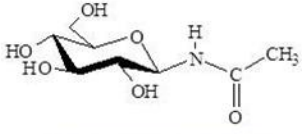
Single crystals usually grown by slow evaporation of saturated solution under well controlled environments

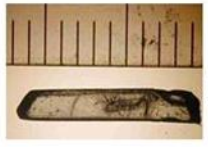
X-ray: Dimensions 0.2 – 0.5 mm / Synchrotron X-ray : 20-30 µm

Neutron: Dimensions over 1.0 mm all dimensions

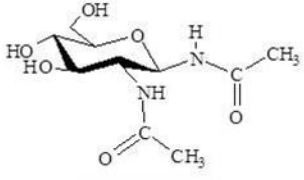




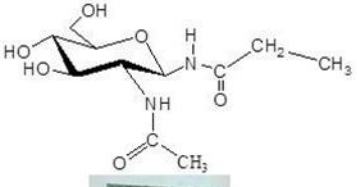




P212121



P21



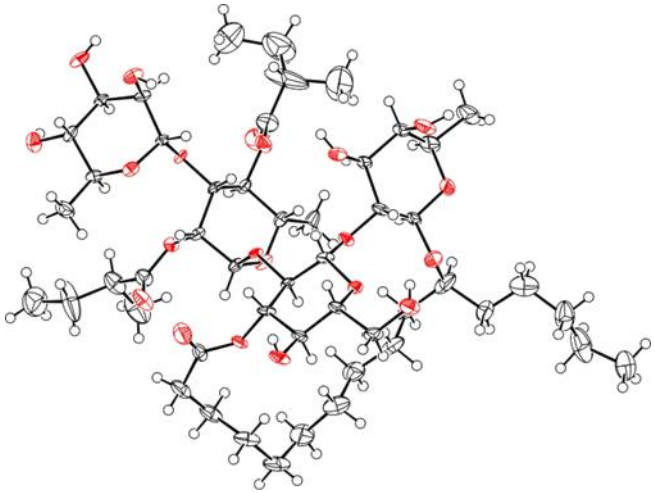
P21

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

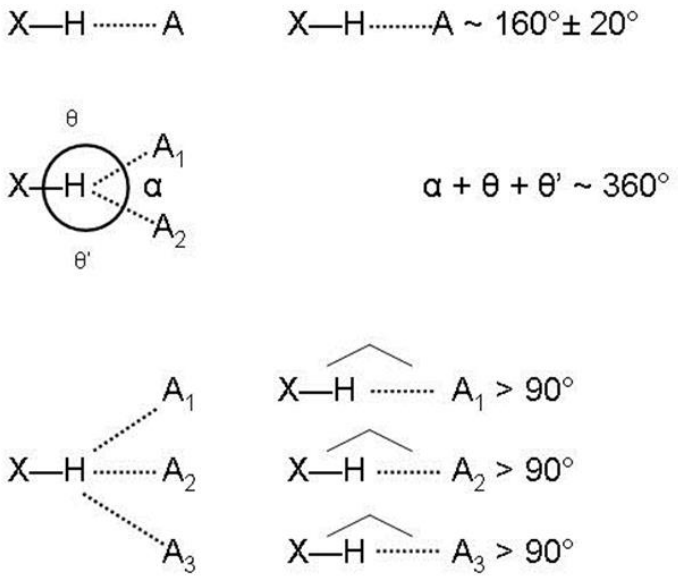
Molecular & Crystal Structures of Carbohydrates

Hydrogen Bonding in Crystalline Oligosaccharides

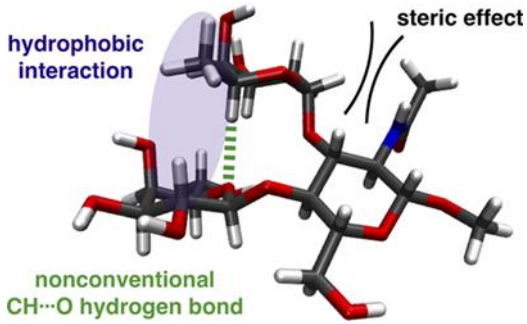
Analysis of high accurate X-ray analysis – Neutron diffraction

dX-dN =(C-H) = -0.096(7)

dX-dN =(O-H) = -0.155(10)



CHO bonds

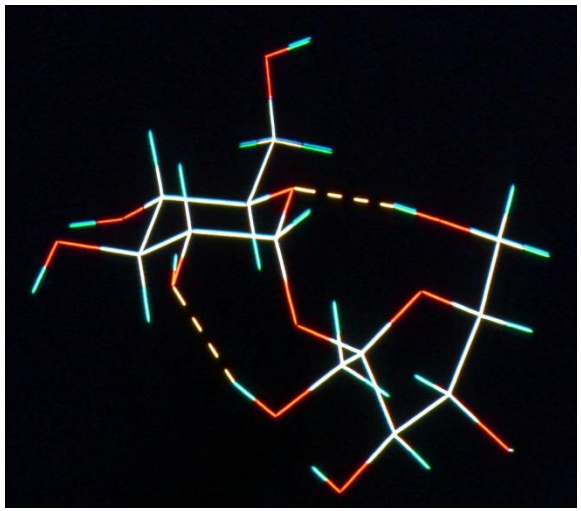


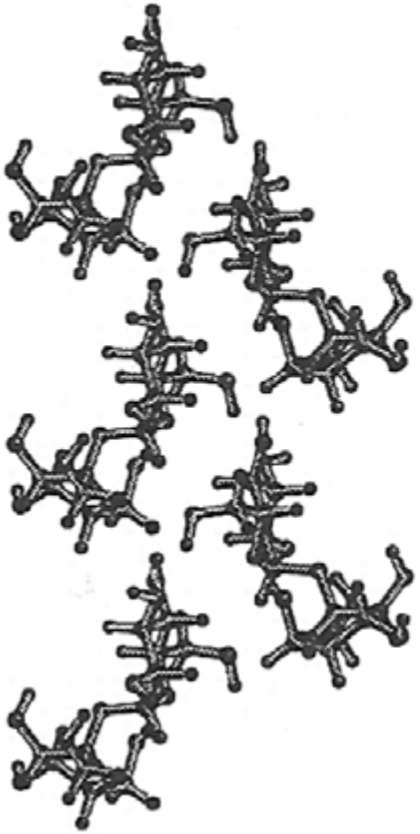
Maximize the Hydrogen Bond interactions throughout the participation of all hydroxyl groups and as many rings oxygen. Two and three-centered bonds

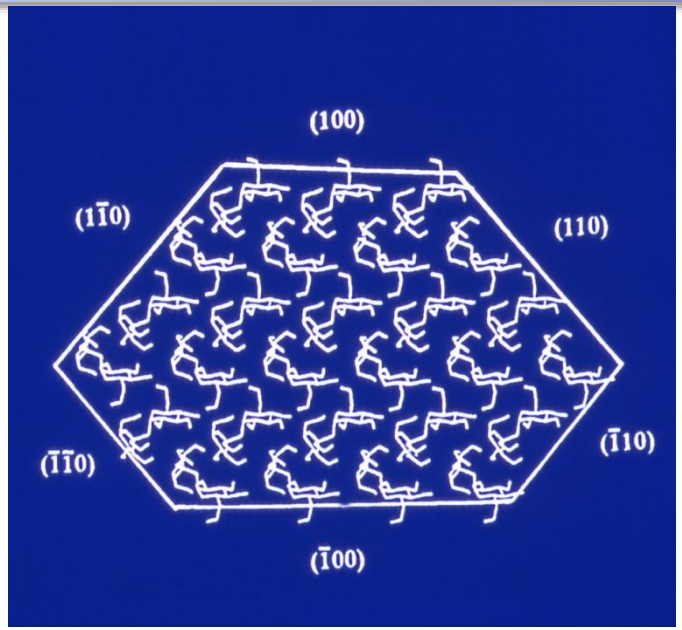
Maximize cooperativity by forming as many finite and infinite chains of hydrogen bonds as possible.

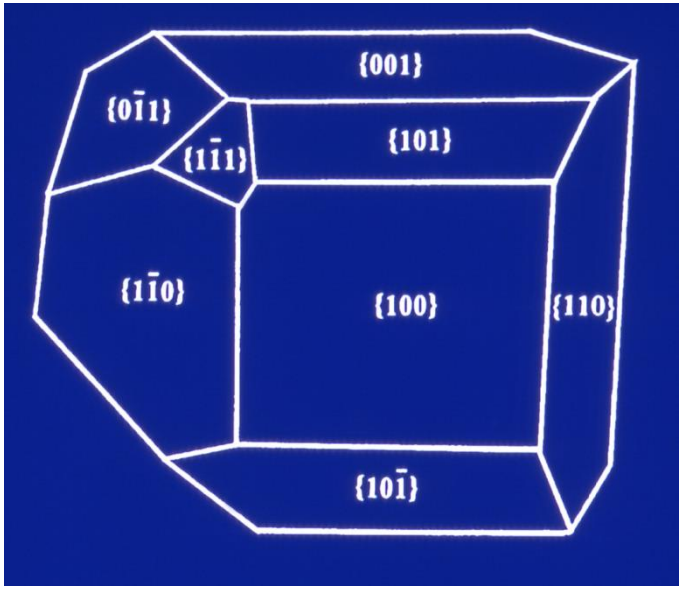
Molecular & Crystal Structures of Carbohydrates

Packing Features









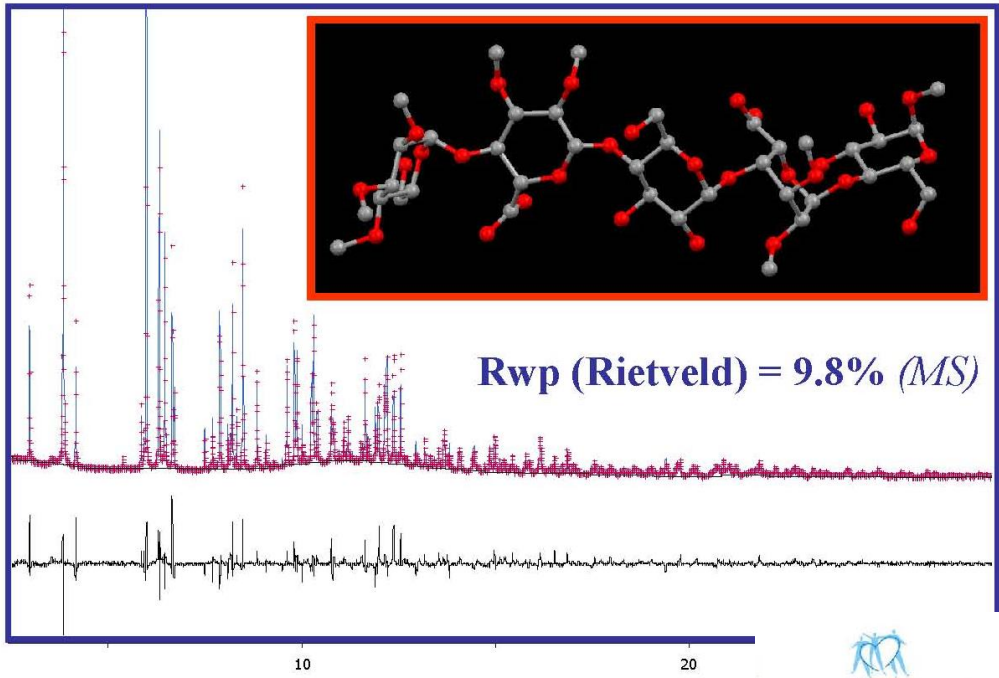
Molecular & Crystal Structures of Carbohydrates

Powder Diffraction

1. Identification of Crystalline Polymorphs

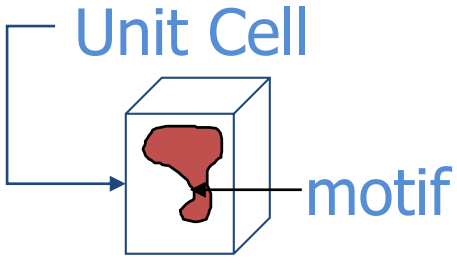
2. Solving Crystal Structures – Rievelt Method + Molecular Modelling

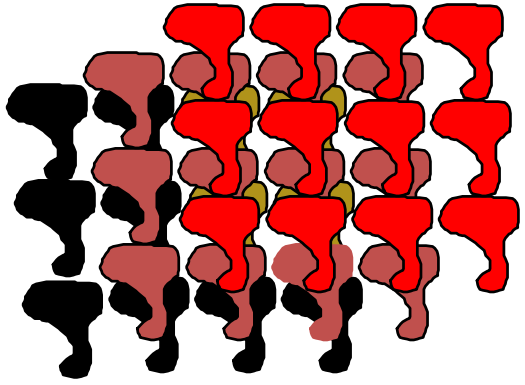
Synthetic Pentasaccharide
ID31@ESRF, λ **=0.8 A**
Monoclinic $P2_1$
a=15.54, b=8.83; c=17.67, β=94.6

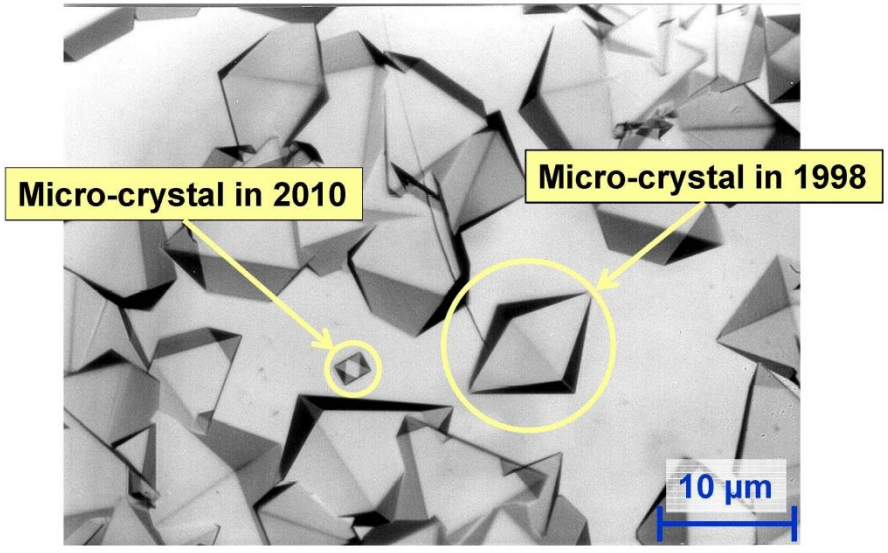


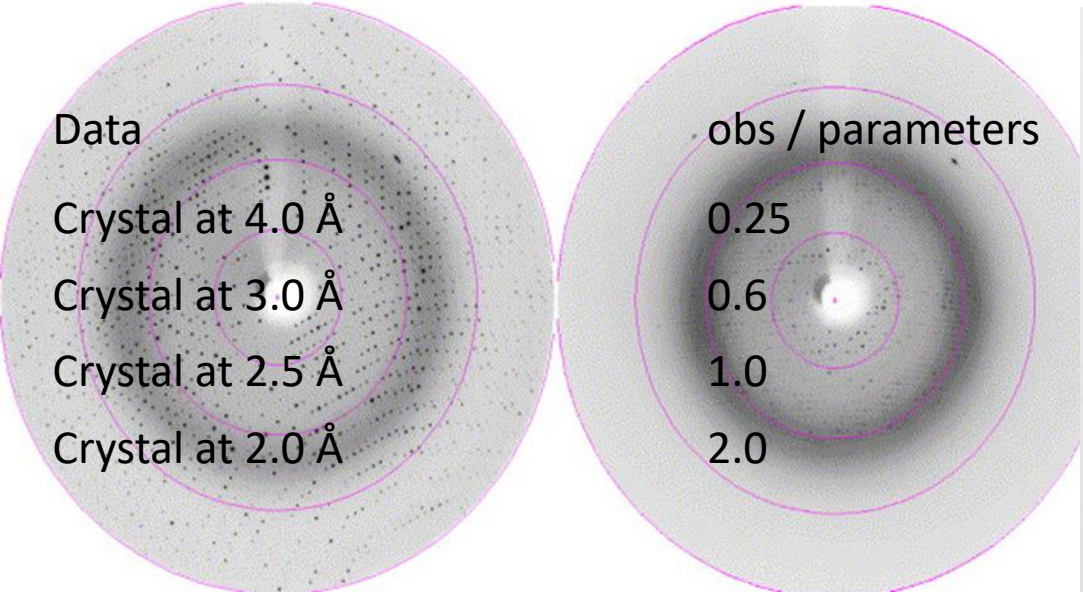
Crystalline Conformations of Oligosaccharides in Proteins

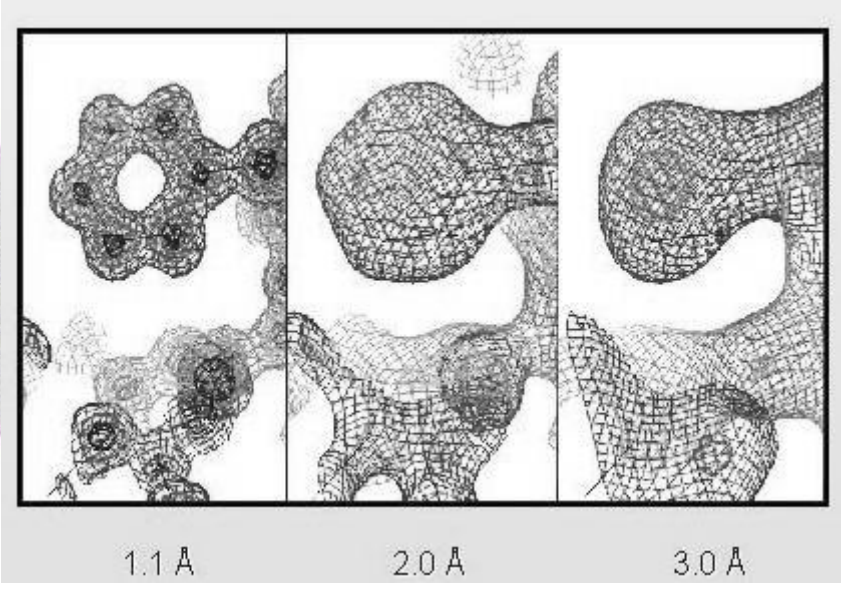
Experimental Conditions and Limitations





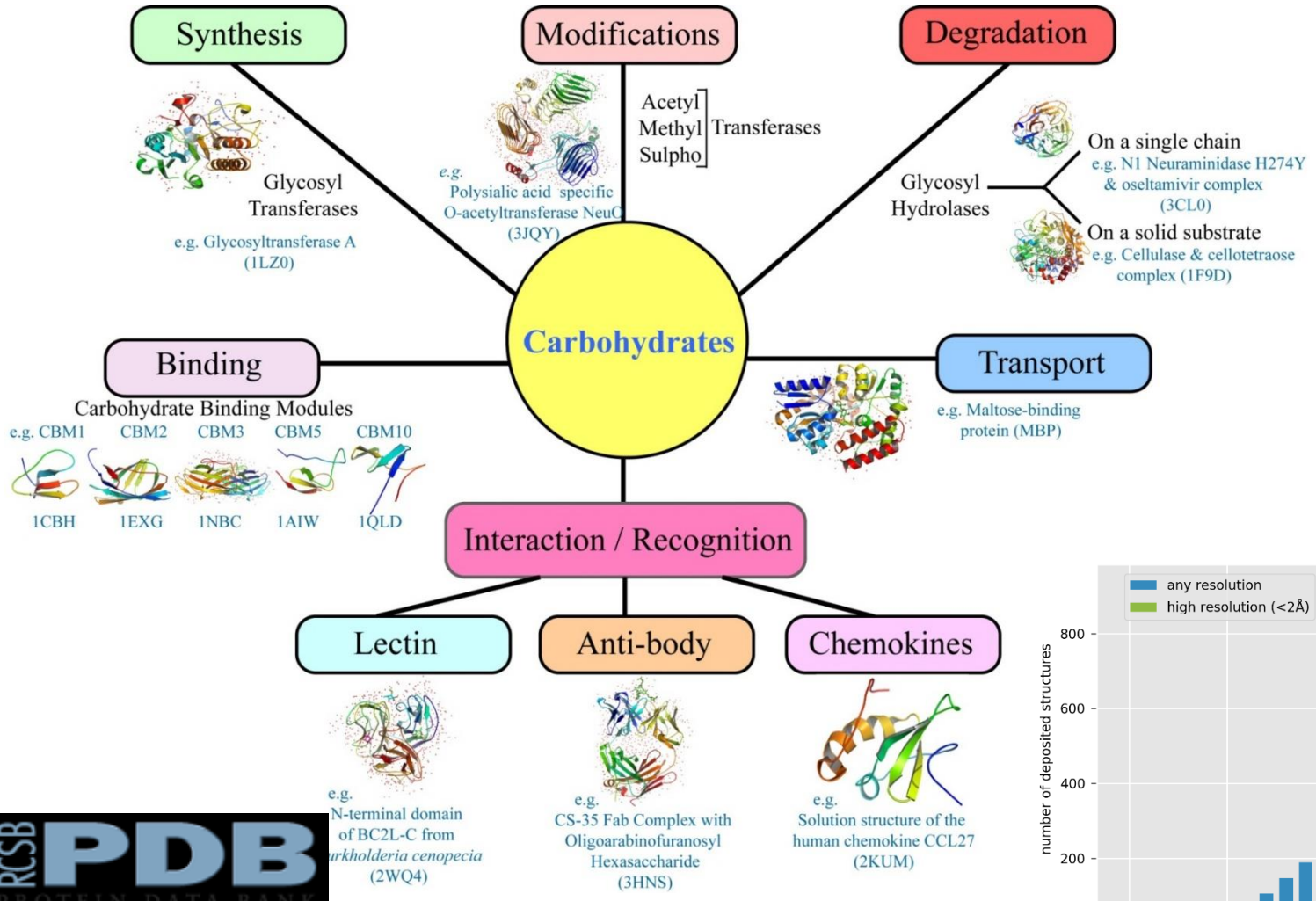


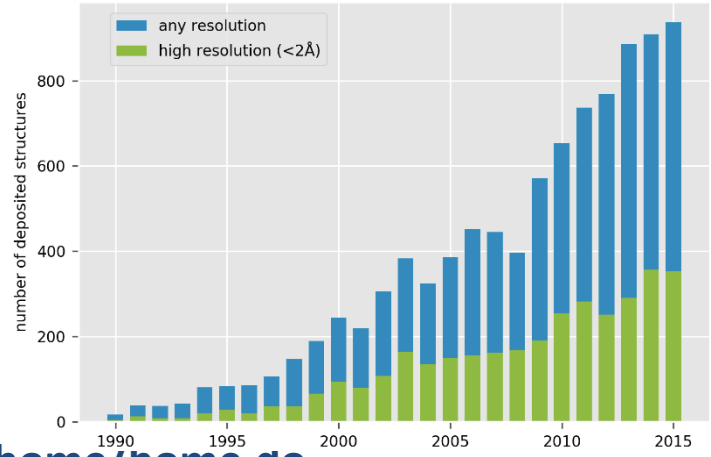




Protein-Carbohydrate Crystal Structures

Protein-Carbohydrate Interactions



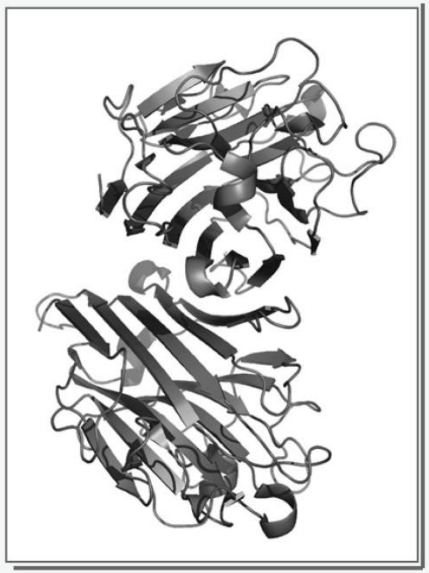


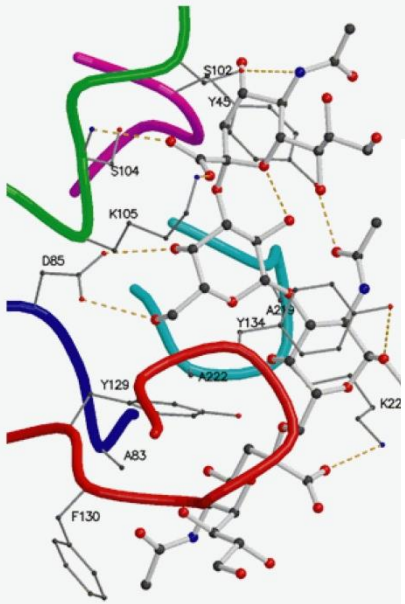
RCSB PDB PROTEIN DATA BANK

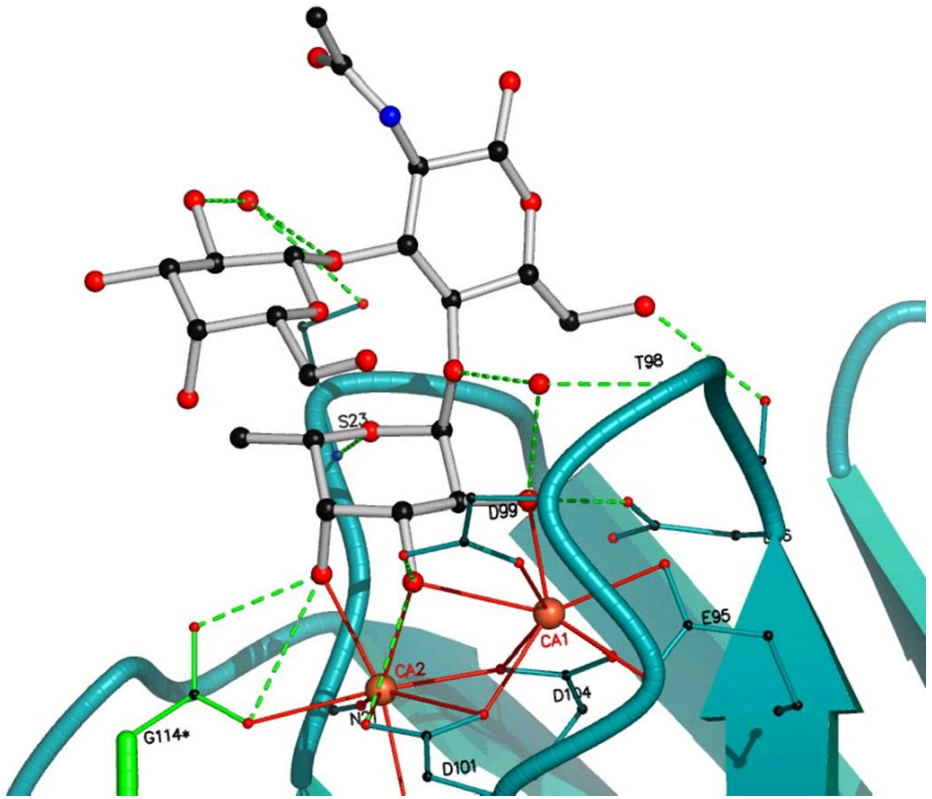
Protein Data Bank : http://www.rcsb.org/pdb/home/home.do

Crystalline Conformations of Oligosaccharides in Proteins

Oligosaccharides –Lectin Complexes

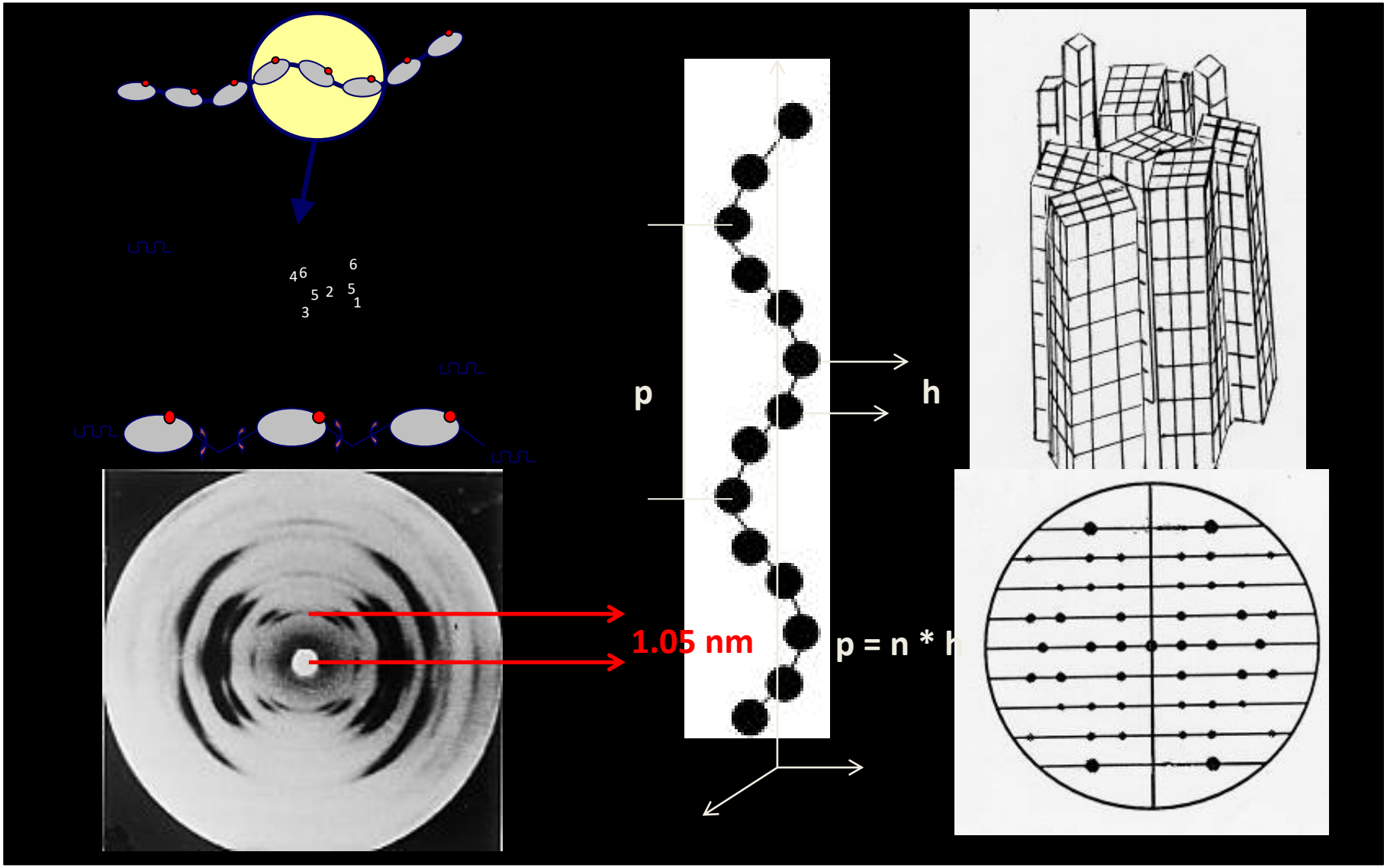






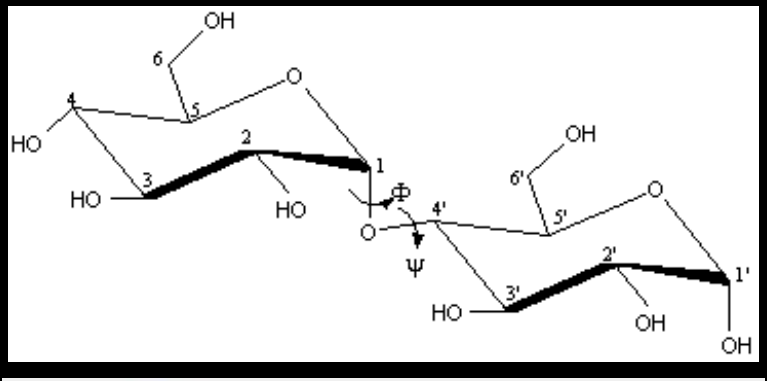
Crystalline Conformations of Polysaccharides

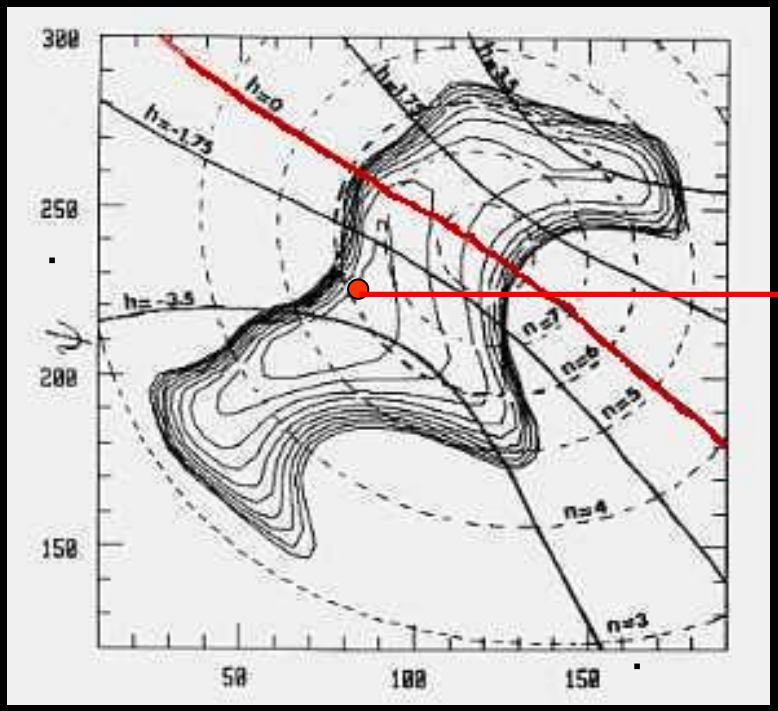
X-Ray Fiber Diffraction of Polysaccharides



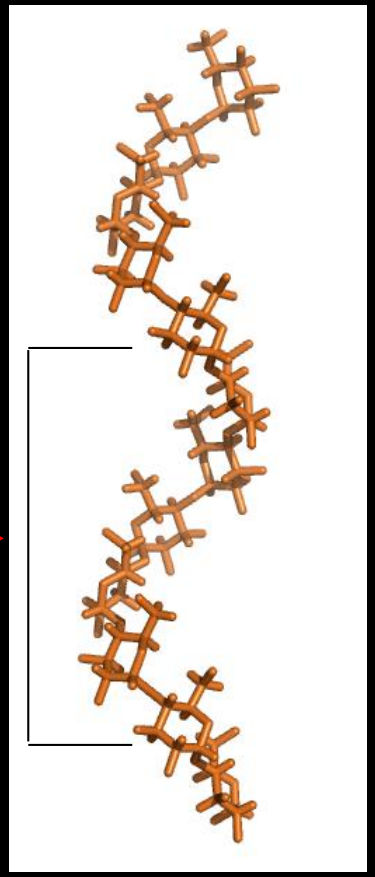
Crystalline Conformations of Polysaccharides

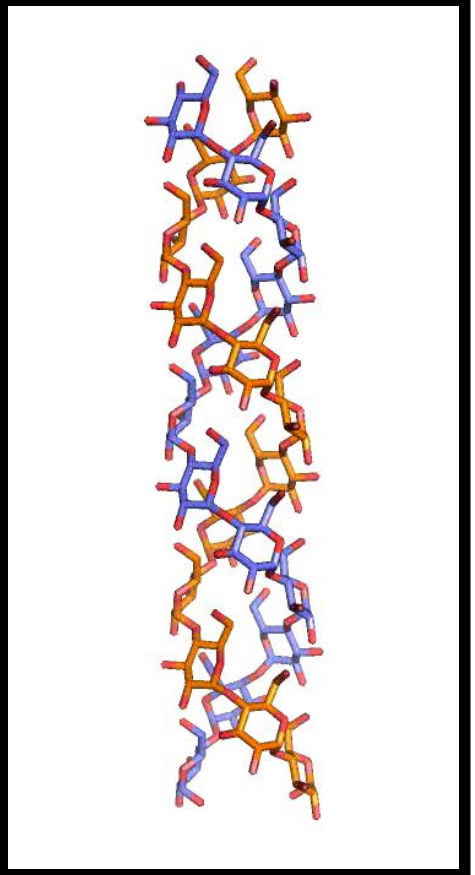
X-Ray Fiber Diffraction of Polysaccharides





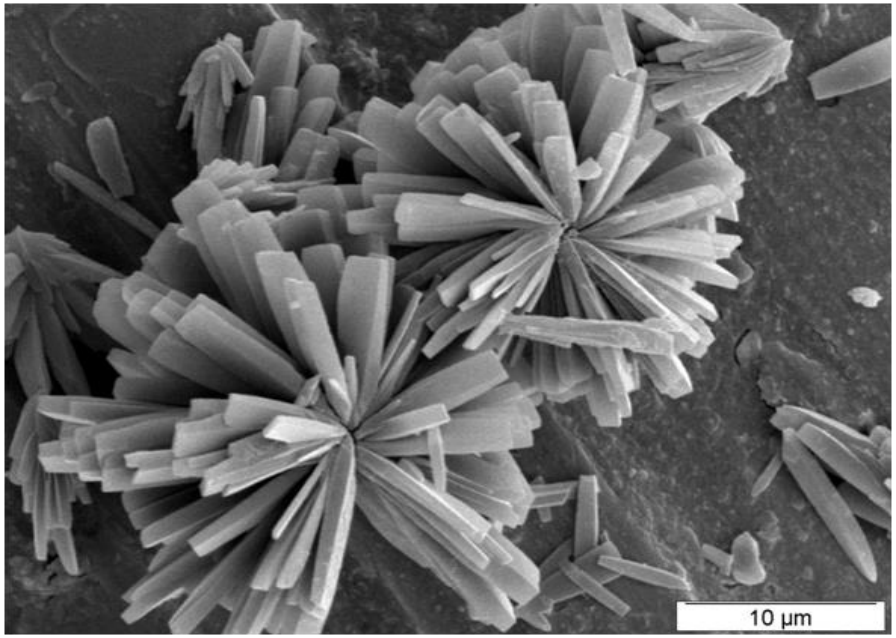
2.1 nm

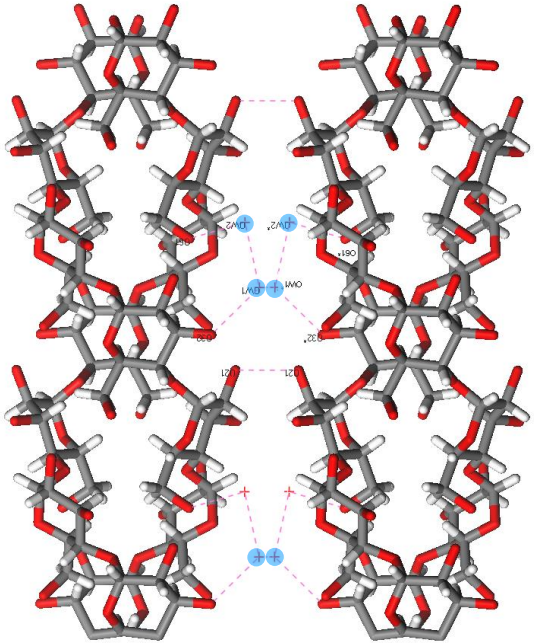


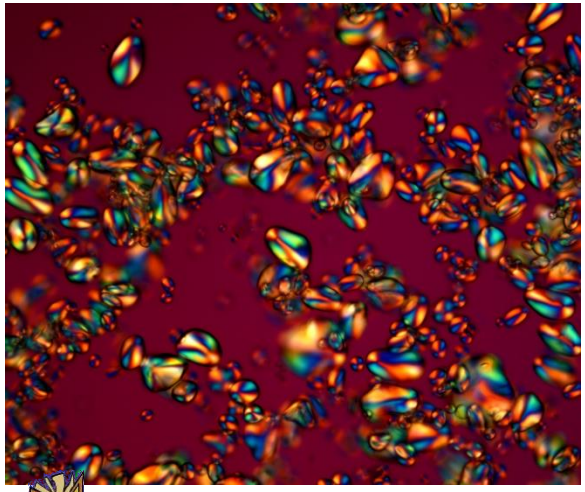


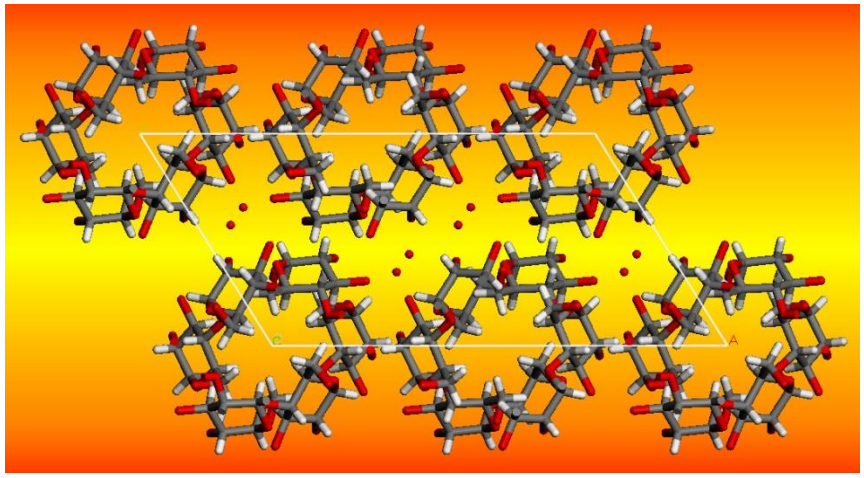
Crystalline Conformations of Polysaccharides

Synchrotron X-Ray Diffraction of Polysaccharides





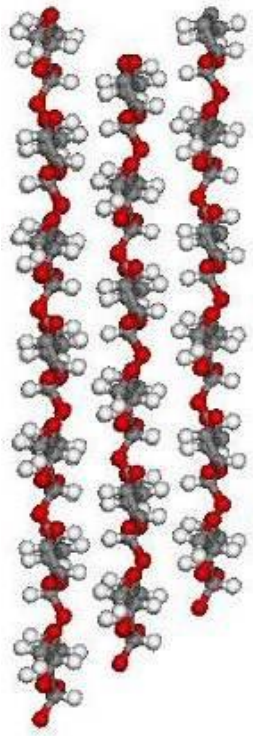


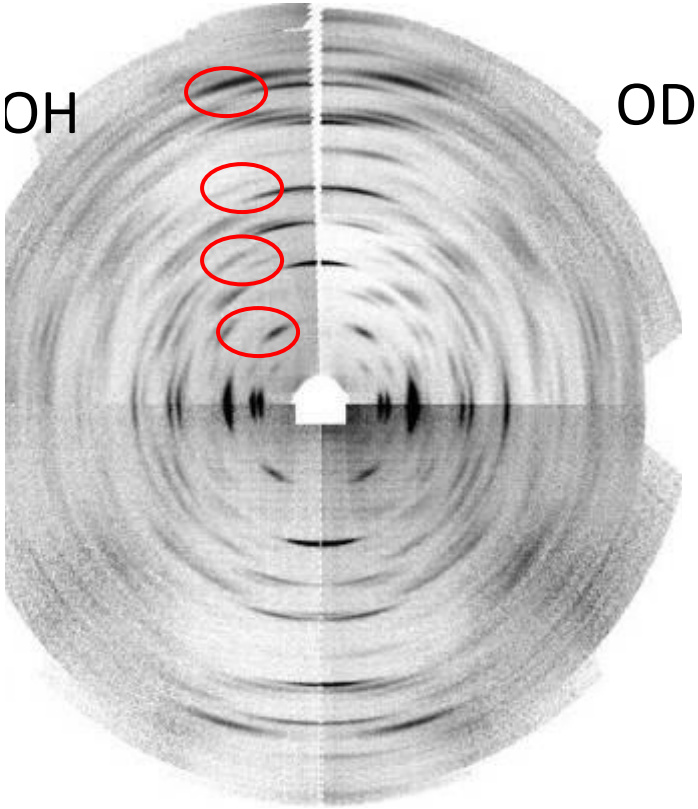


D. Popov, 2009, Macromolecules, 42, 1167-1174

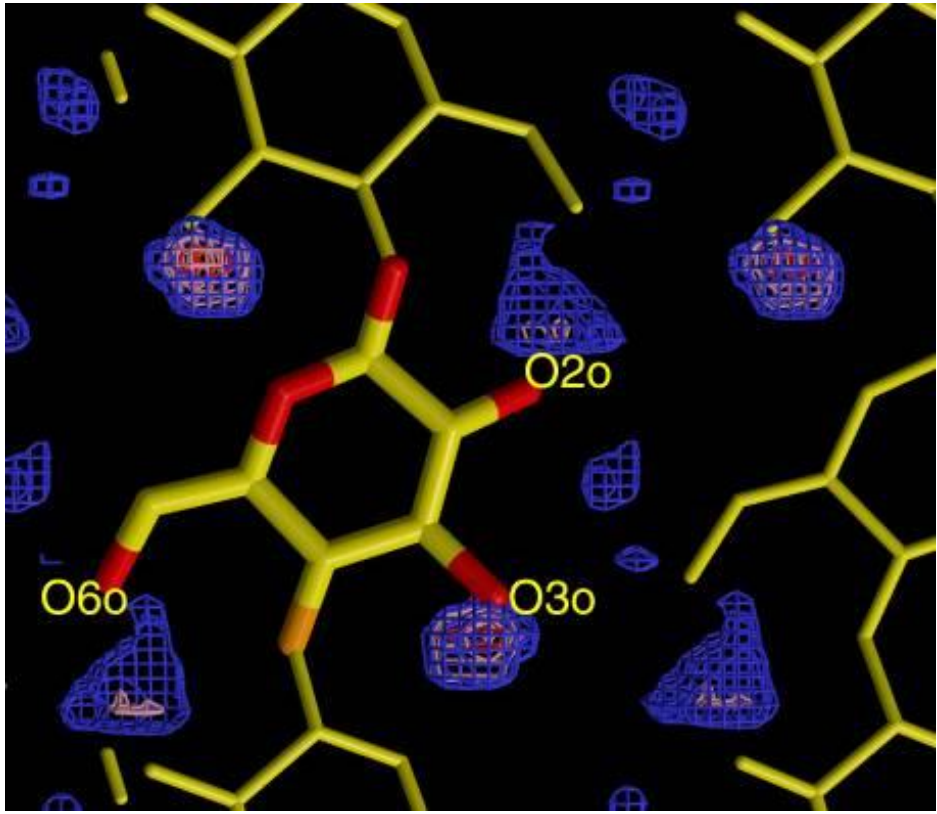
Crystalline Conformations of Polysaccharides

X-Ray Fiber Diffraction using Synchrotron and Neutron Radiations





OD



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

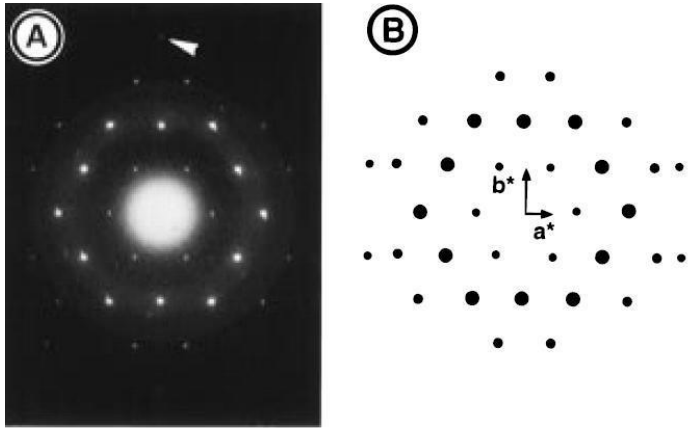
Crystalline Conformations of Polysaccharides

Electron Diffraction of Polysaccharides

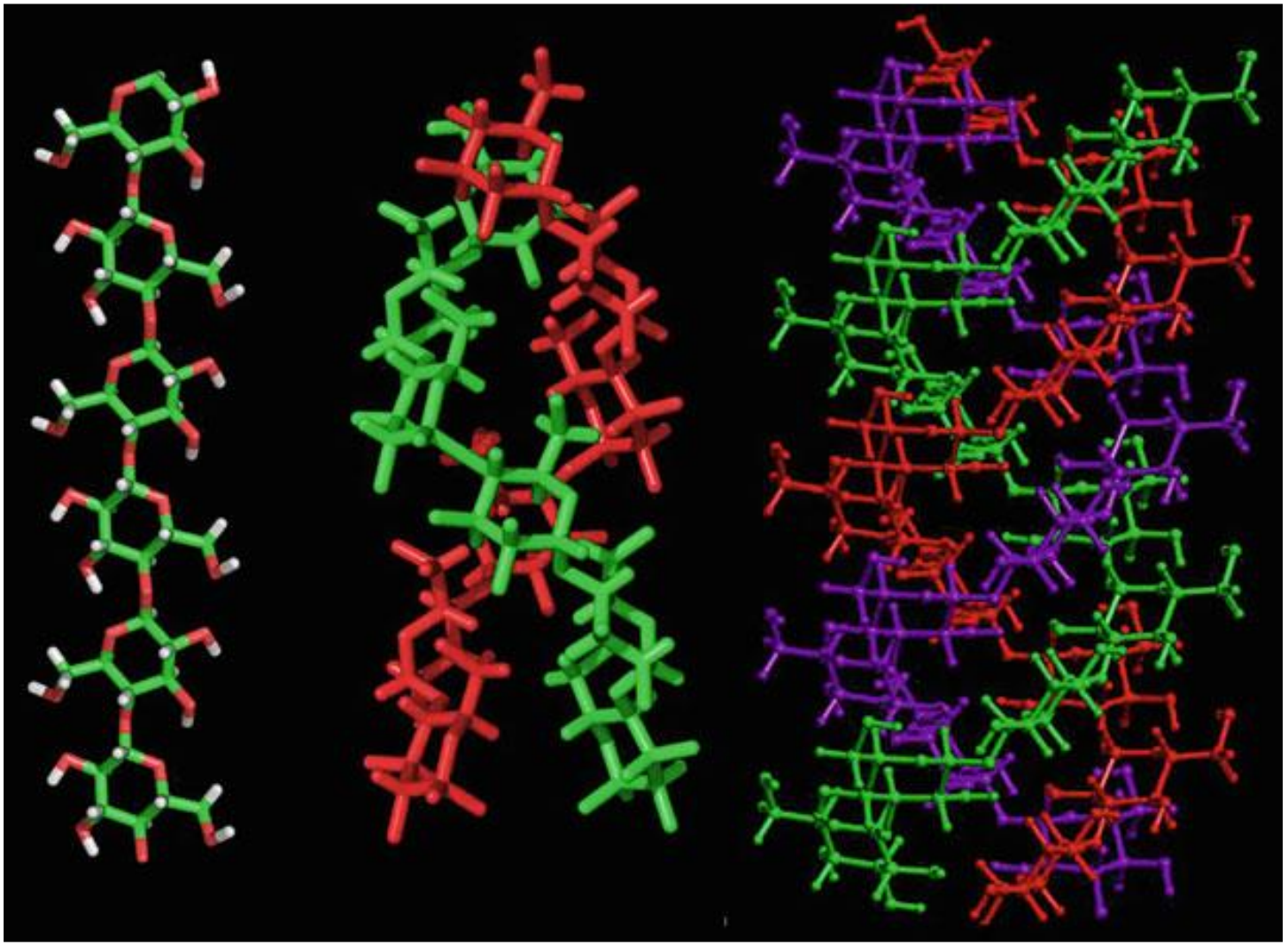
Electrons are charged particles and interact with matter through the Coulomb forces. The incident electrons feel the influence of both the positively charged atomic nuclei and the surrounding electrons.

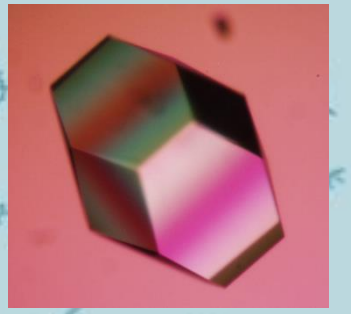
Electron diffraction of solids is usually performed in a **Transmission Electron Microscope** (TEM) where the electrons pass through a thin film of the material to be studied. The resulting diffraction pattern is then observed on a fluorescent screen, recorded on photographic film, on imaging plates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

ED is subjected to several important limitations.
The sample must be electron transparent, i.e. the sample thickness must be of the order of 100 nm or less.
Need careful and time consuming sample preparation.
Many samples are vulnerable to radiation damage caused by the incident electrons.



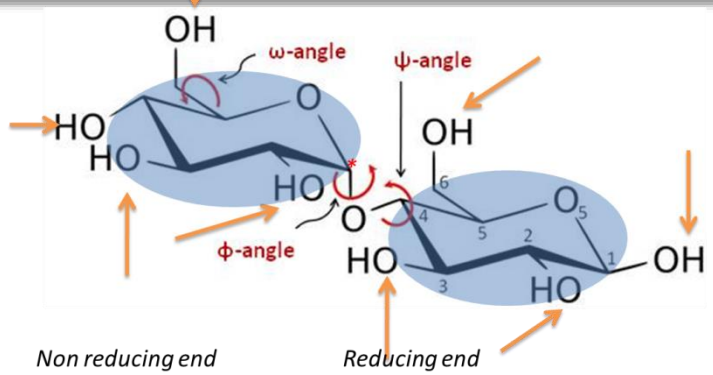
Helical Structures of Polysaccharides

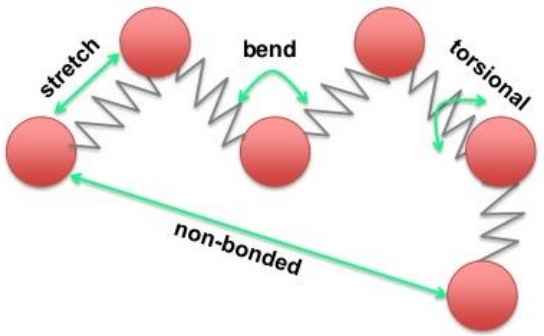


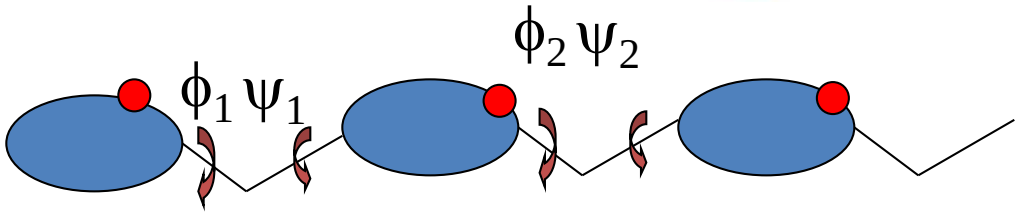


From *in cubo* to *in silico*

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

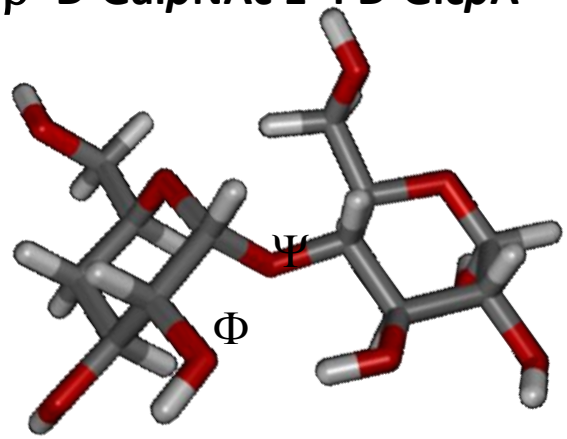
- branched structures

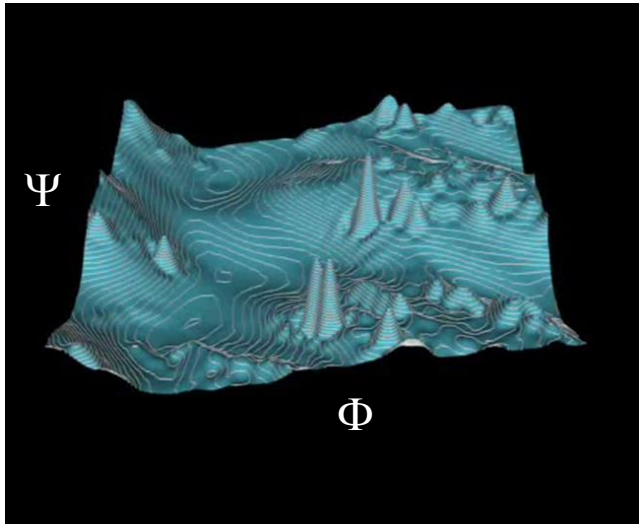
….

But very useful for building starting structures!

Disaccharide: Structural Descriptors

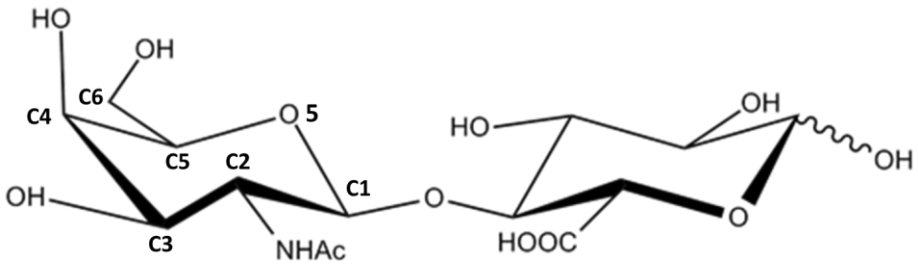
β–D-Gal*p*NAc 1-4 D-Glc*p*A



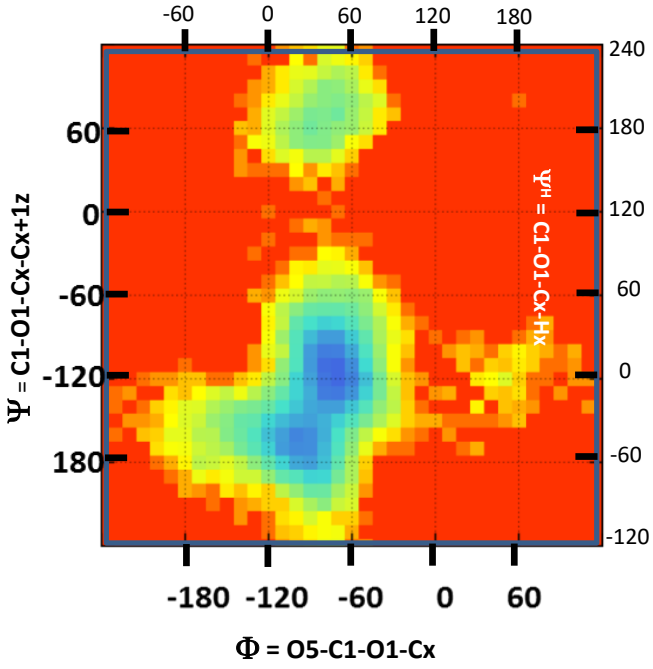


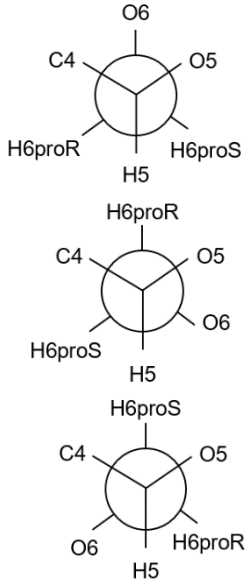
β1-4

Gal*p*NAc β1-4 Glc*p*A



Φ^H = H1-C1-O1-Cx





(Φ, Ψ) = -80; -120
(Φ, Ψ) = -100; -160
(Φ, Ψ) = -90; 70
(Φ, Ψ) = 60; -120

Molecular Mechanics / Dynamics

Initial positions given by the **PDB**
Initial velocities determined based on a Boltzmann distribution of velocities at the target temperature

$$\vec{F} = m\vec{a} = -\frac{dU}{dr}$$

New positions and velocities through integration

MD run → trajectory

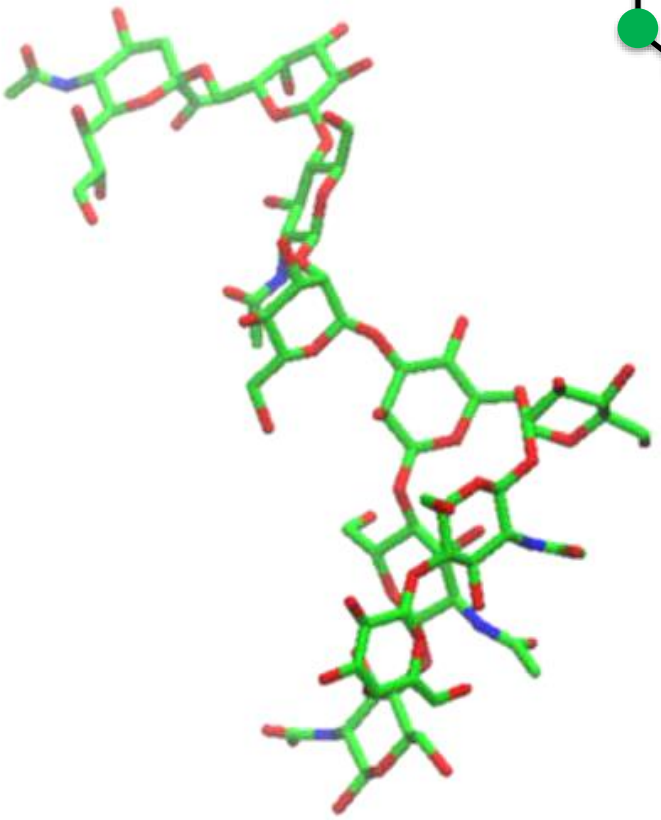
THE FORCE FIELD

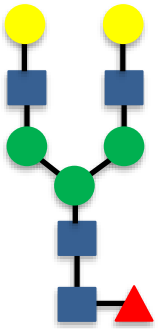
$$v^{\text{Coulomb}}(r) = \frac{Q_1 Q_2}{4\pi\epsilon_0 r},$$

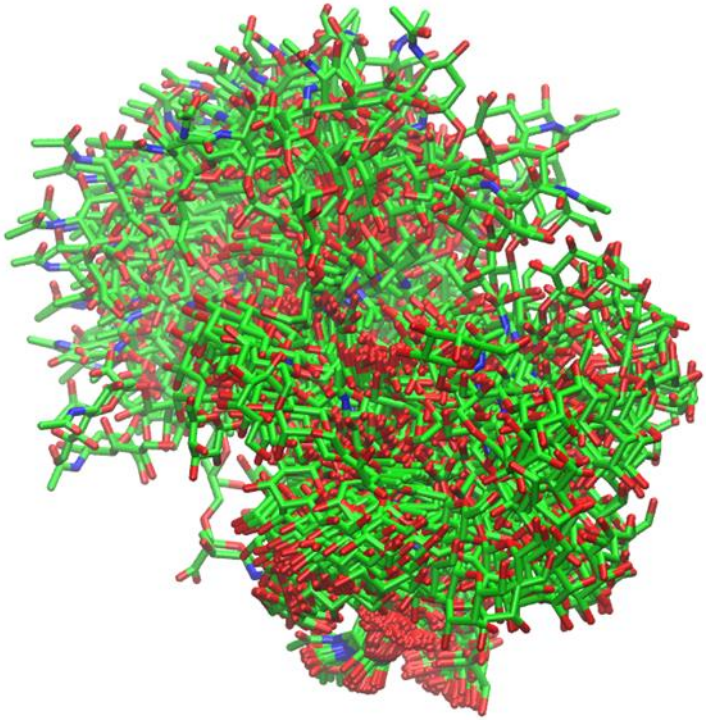
$$v^{\text{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right].$$

$$\begin{aligned} u_{\text{intramolecular}} = & \frac{1}{2}\sum_{\text{bonds}} k_{ij}^{r}\left(r_{ij} - r_{\text{eq}}\right)^2 \\ & + \frac{1}{2}\sum_{\substack{\text{bend}\\ \text{angles}}} k_{ijk}^{\theta}\left(\theta_{ijk} - \theta_{\text{eq}}\right)^2 \\ & + \frac{1}{2}\sum_{\substack{\text{torsion}\\ \text{angles}}}\sum_{m} k_{ijkl}^{\phi,m}\left(1 + \cos(m\phi_{ijkl} - \gamma_m)\right) \end{aligned}$$

Glycans Can be Highly Flexible and Dynamic



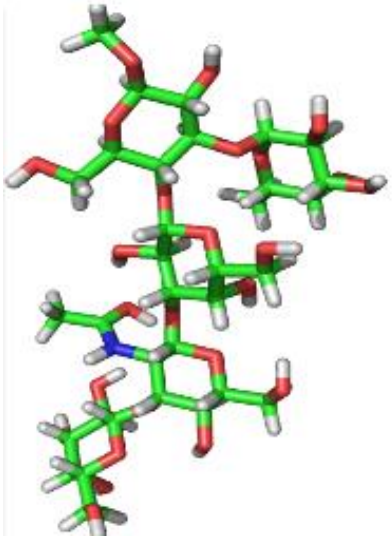


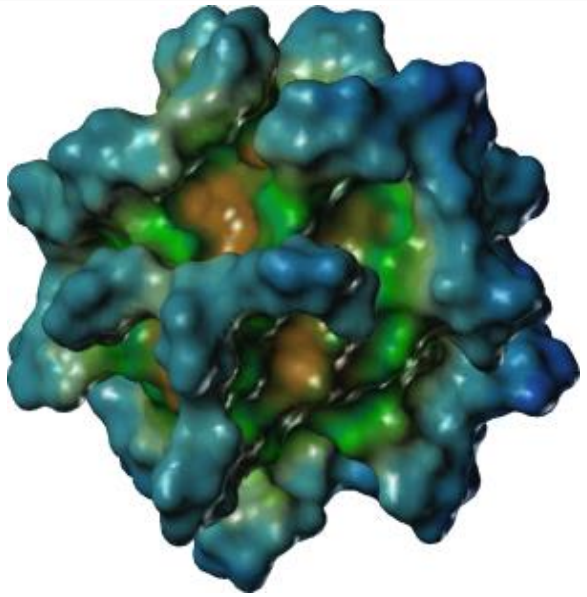


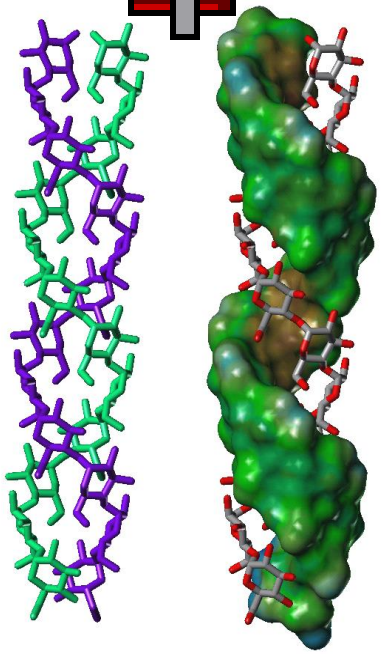
This ensemble of different structures is not necessarily ‘a mess’ not all possible conformations are allowed or equally populated and some of these conformations may actually be functionally important

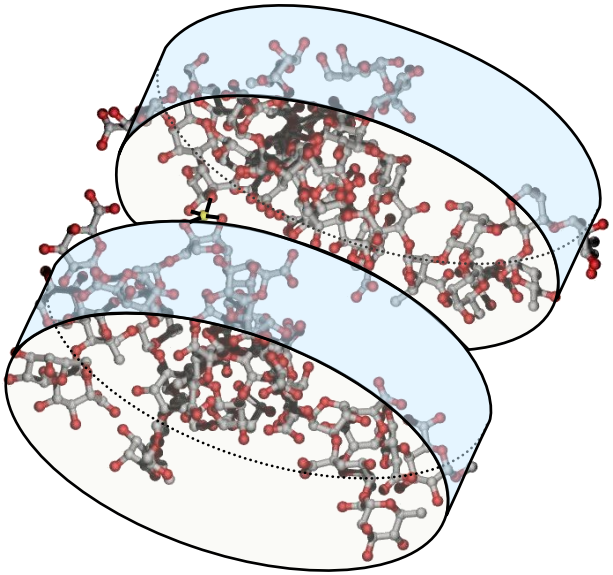
250 ns single trajectory

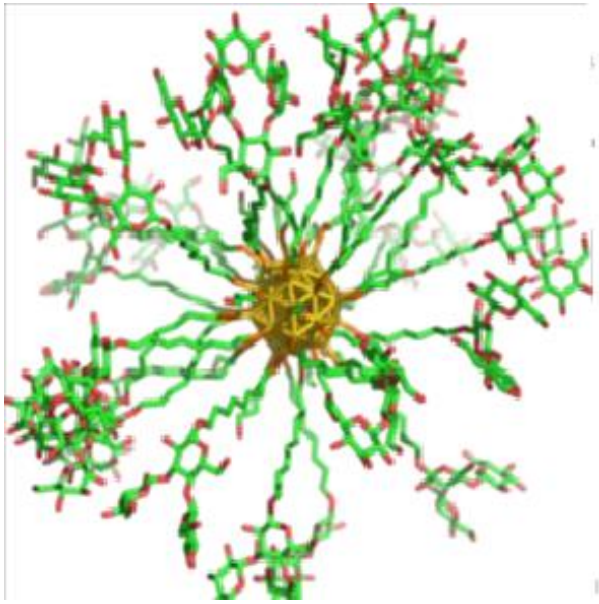
Molecular Modeling at work











Glycan Active Proteins

Synthesis

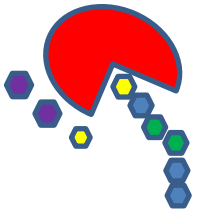
Glycosyl Transferases

Modifications

Acetyl
Methyl
Sulpho
Transferases

Degradation

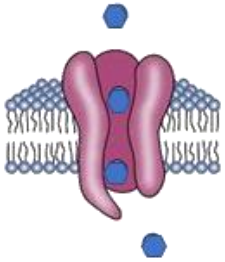
Glycosyl Hydrolases



On a single glycan

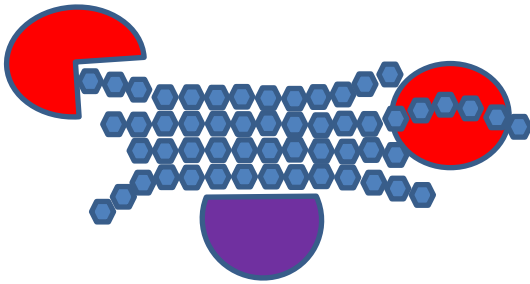
Carbohydrates

Transport



Binding

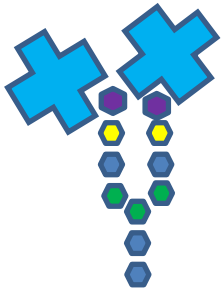
Carbohydrate Binding Modules



On a solid substrate

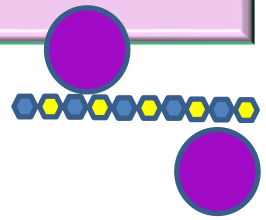
Interaction / Recognition

Antibody

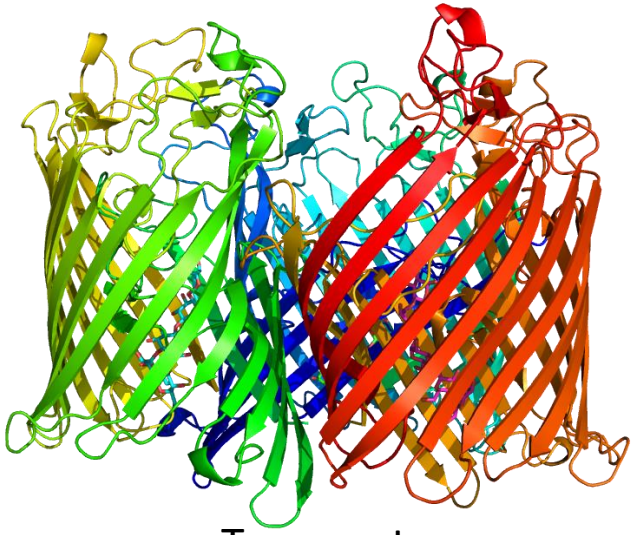


Lectin

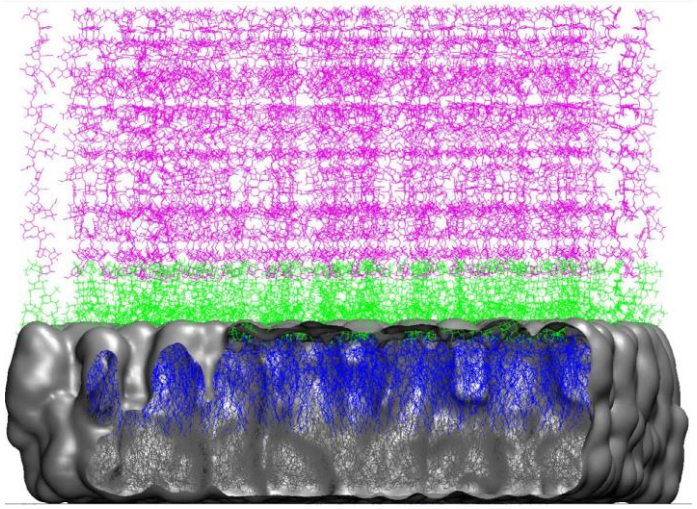
Chemokines



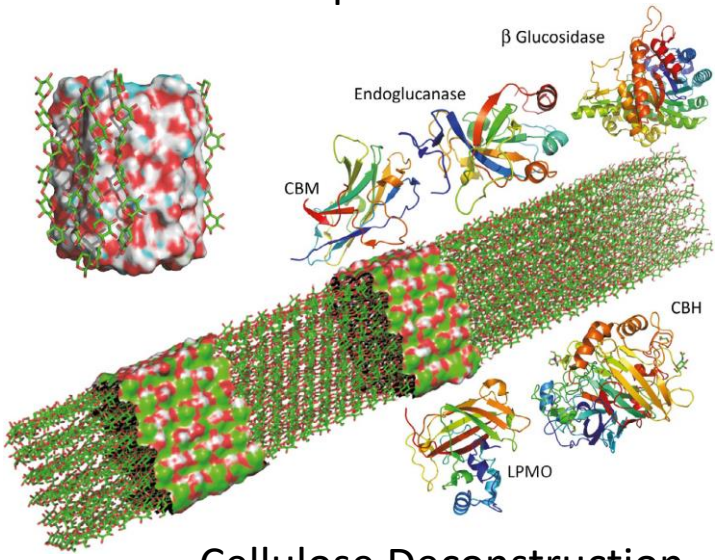
A Wide Range of Applications



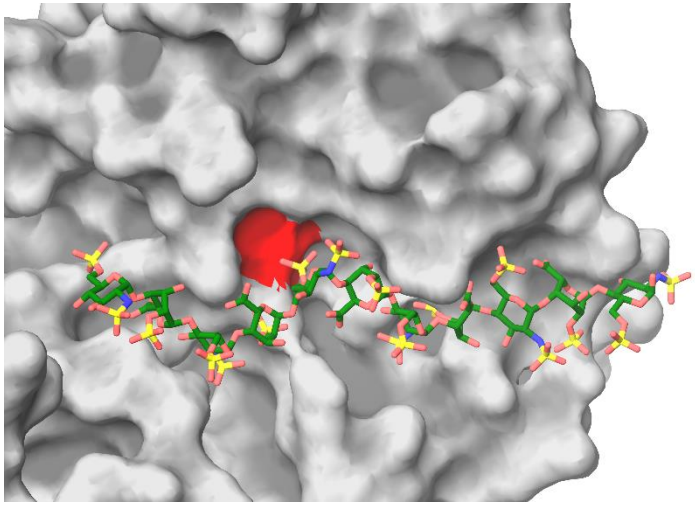
Transport



Lipopolysaccharides

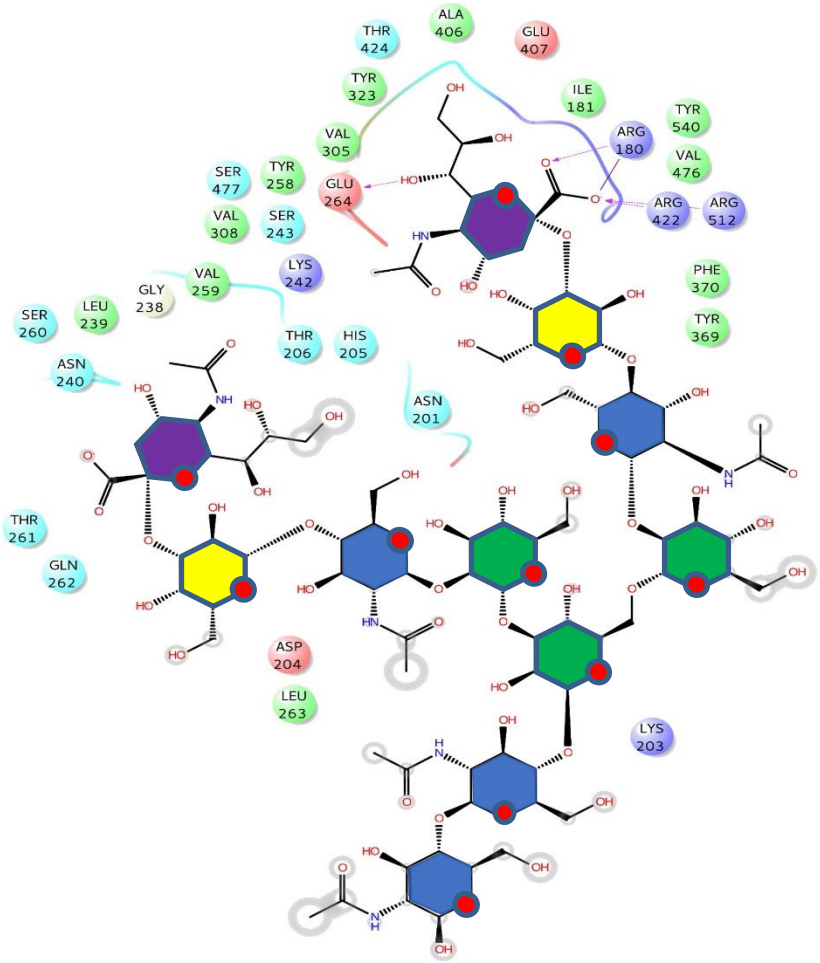


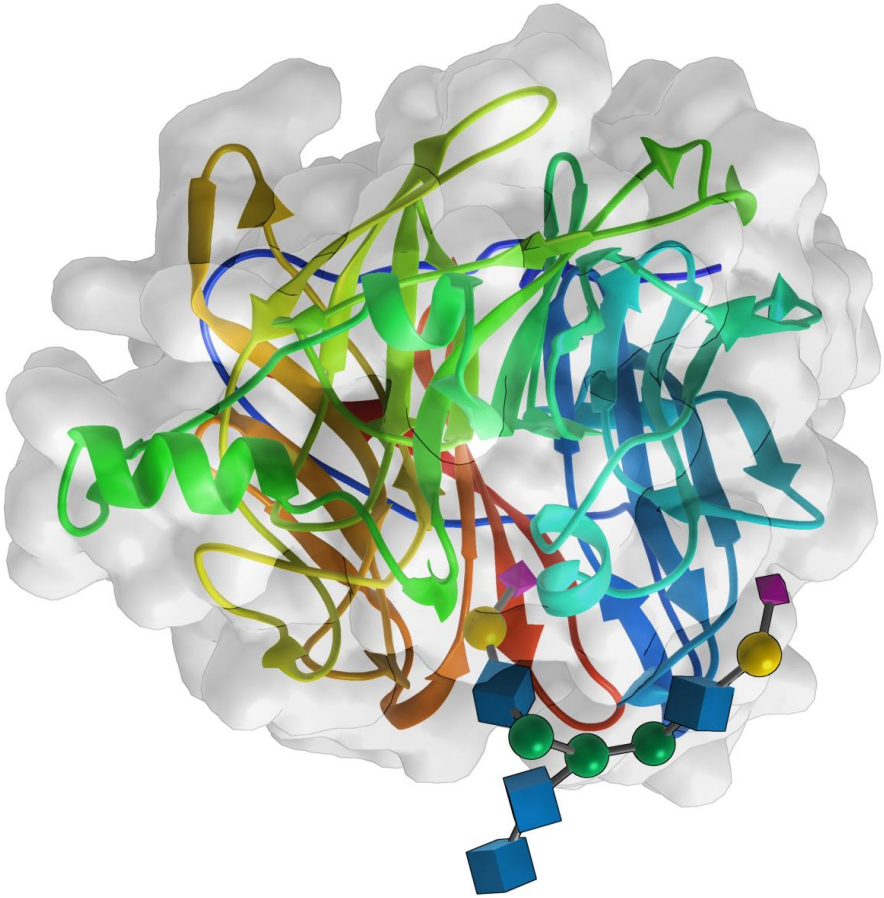
Cellulose Deconstruction



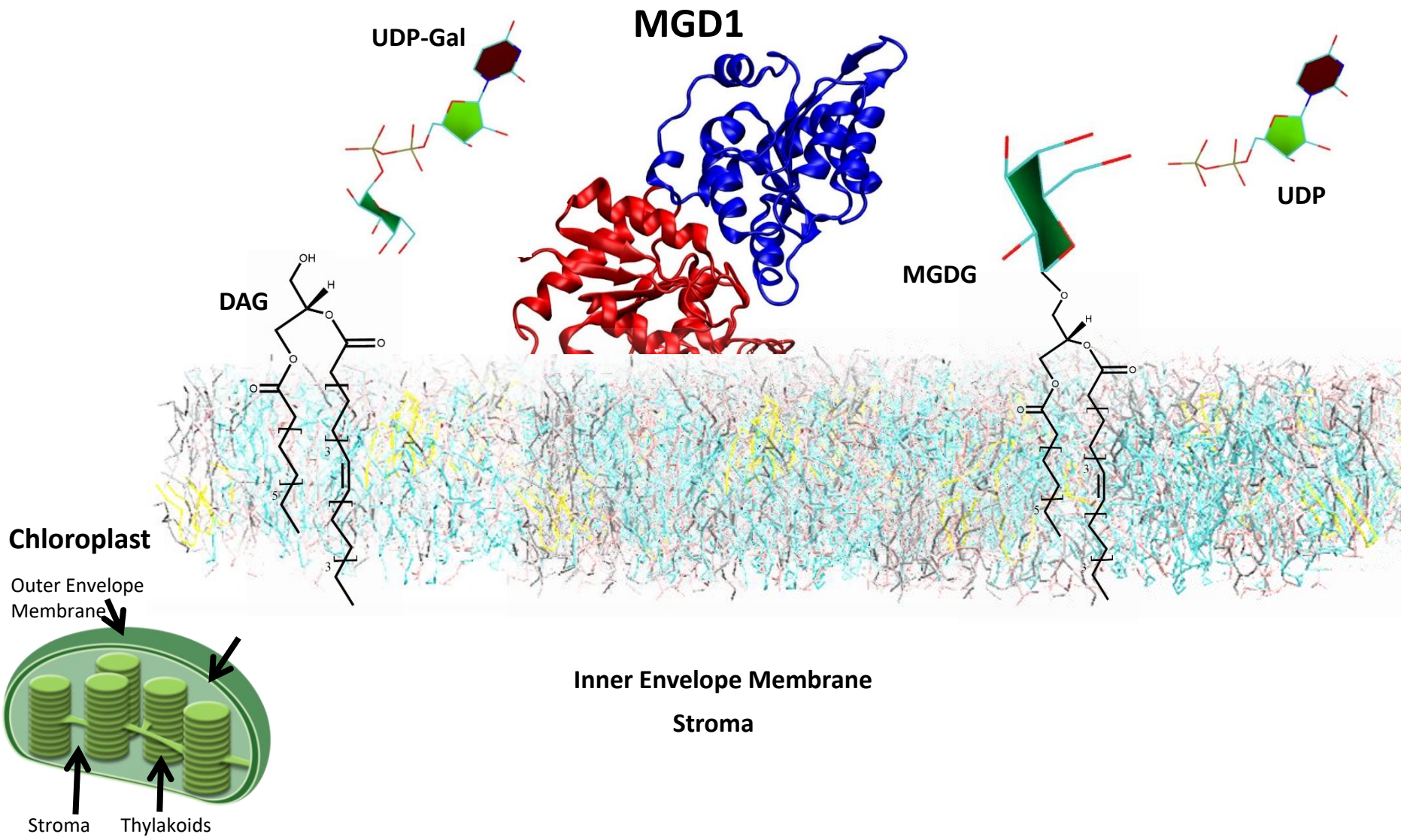
GlycosaminoGlycan Protein Interaction

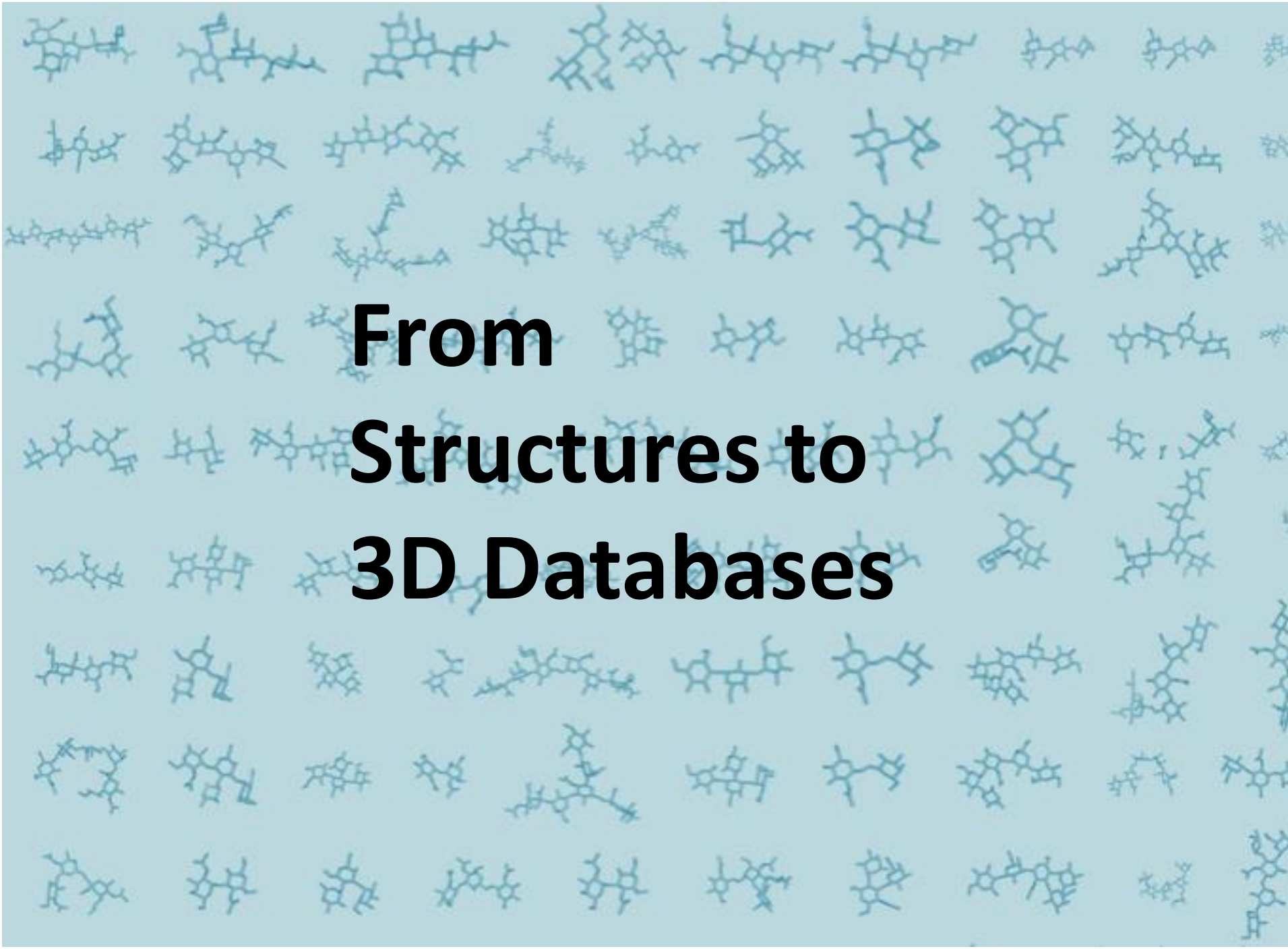
Glycan Receptor Binding by Mumps Virus Hemagglutinin-Neuraminidase





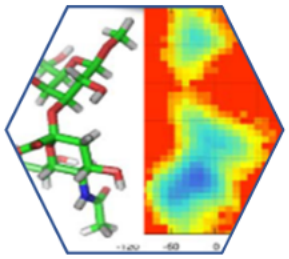
Membrane Assisted Biosynthesis of Glycolipid

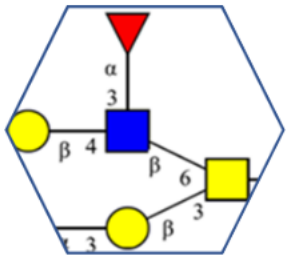



From
Structures to
3D Databases

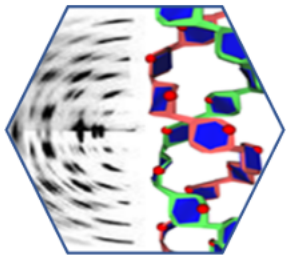
Glyco3D:https://glyco3d.cermav.cnrs.fr/home.php

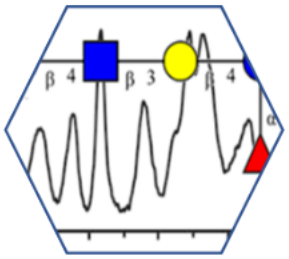
GLYCO3D 2.0


Disac3-DB



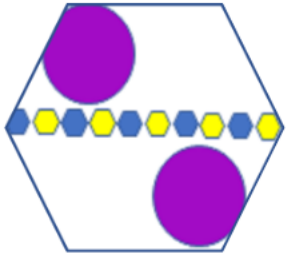
BioOligo-DB


Polysac3-DB



NMR oligo

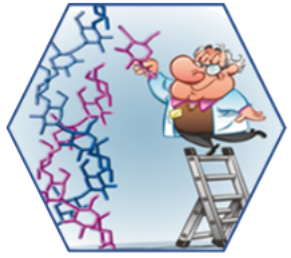

EPS-DB

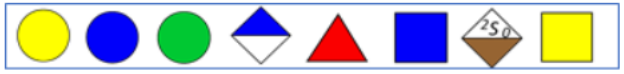

GAG-DB


CBMcarb-DB


Unilectin


mAbscarb-DB


Polys-Glycan Builder

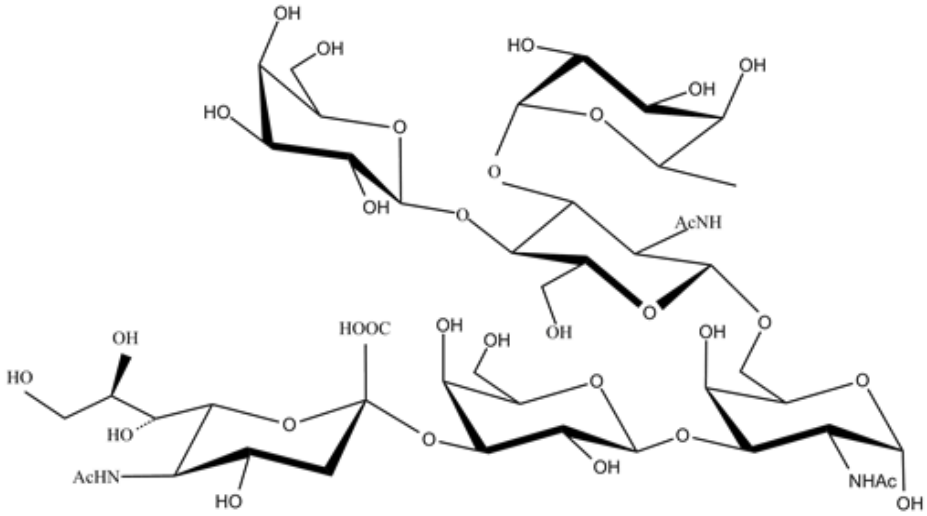

Monosac-DB

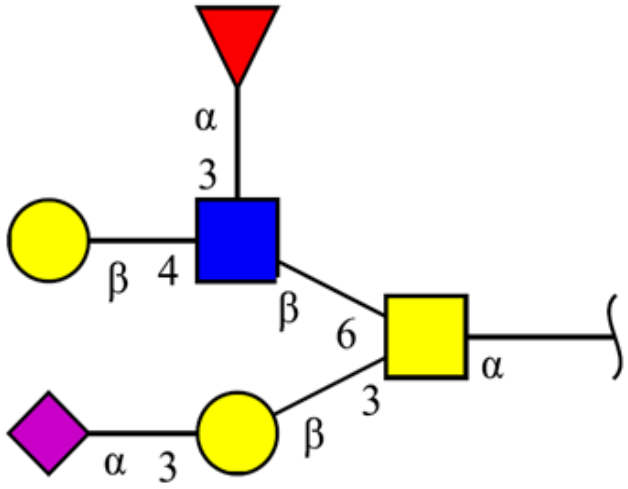

Other tools

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



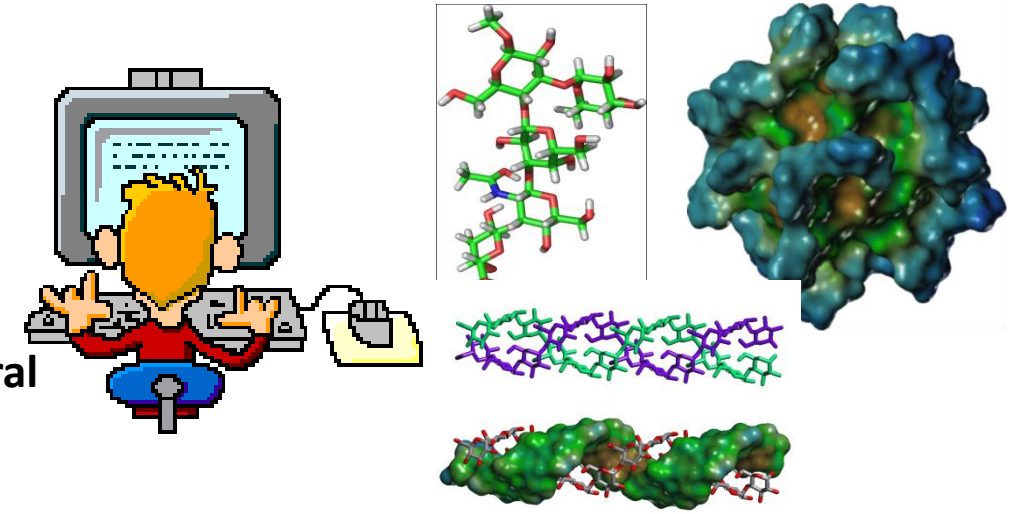


RES
1b:a-dgal-HEX-1:5
2s:n-acetyl
3b:b-dgal-HEX-1:5
4b:a-dgro-dgal-NON-2:6|1:a|2:keto|3:d
5s:n-acetyl
6b:b-dglc-HEX-1:5
7s:n-acetyl
8b:a-lgal-HEX-1:5|6:d
9b:b-dgal-HEX-1:5
LIN
1:1d(2+1)2n
2:1o(3+3)3d
3:3o(3+2)4d
4:4d(5+1)5n
5:1o(6+1)6d
6:6d(2+1)7n
7:6o(3+1)8d
8:6o(4+1)9d

GlycoCT

e-Glycoscience

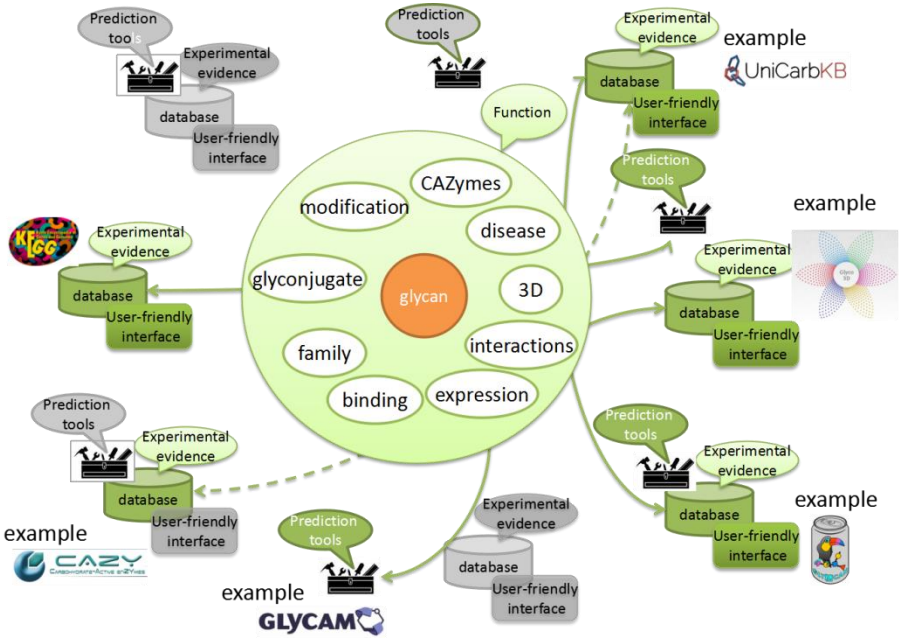
Continued advances in molecular modeling has generated insights for understanding glycan structures and properties. Robust, validated informatics tools are developed in to enable accurate and fast determination of complex **carbohydrate and glycoconjugate structural prediction, computational modeling, and data mining.**



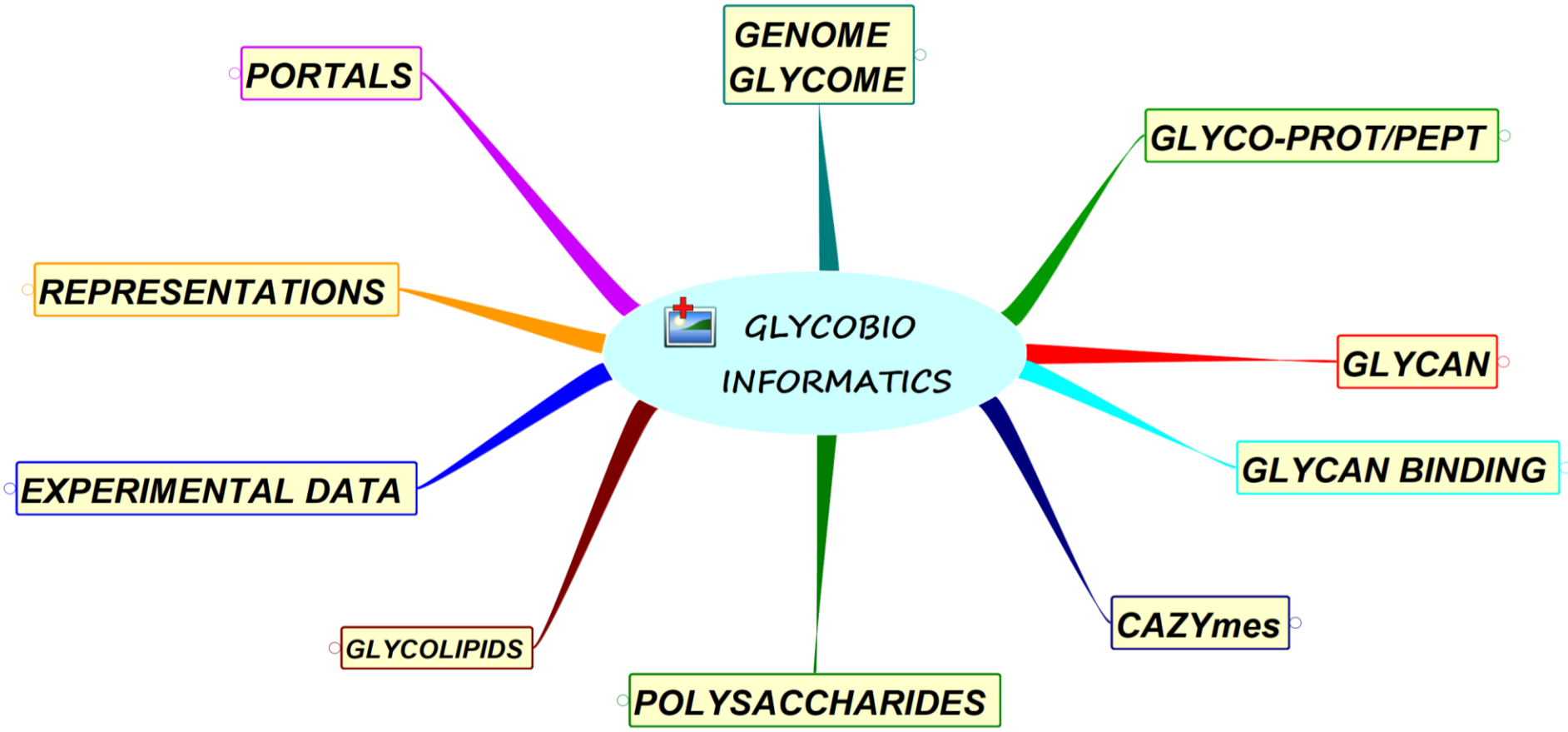
Database have been developedand cover including mammalian, plant and microbial carbohydrates and glycoconjugates.

The carbohydrate **structural database needs to be fully cross-referenced with databases that provide complementary biological information**.

There should be a requirement for deposition of new structures into the database using a reporting standard for minimal information.



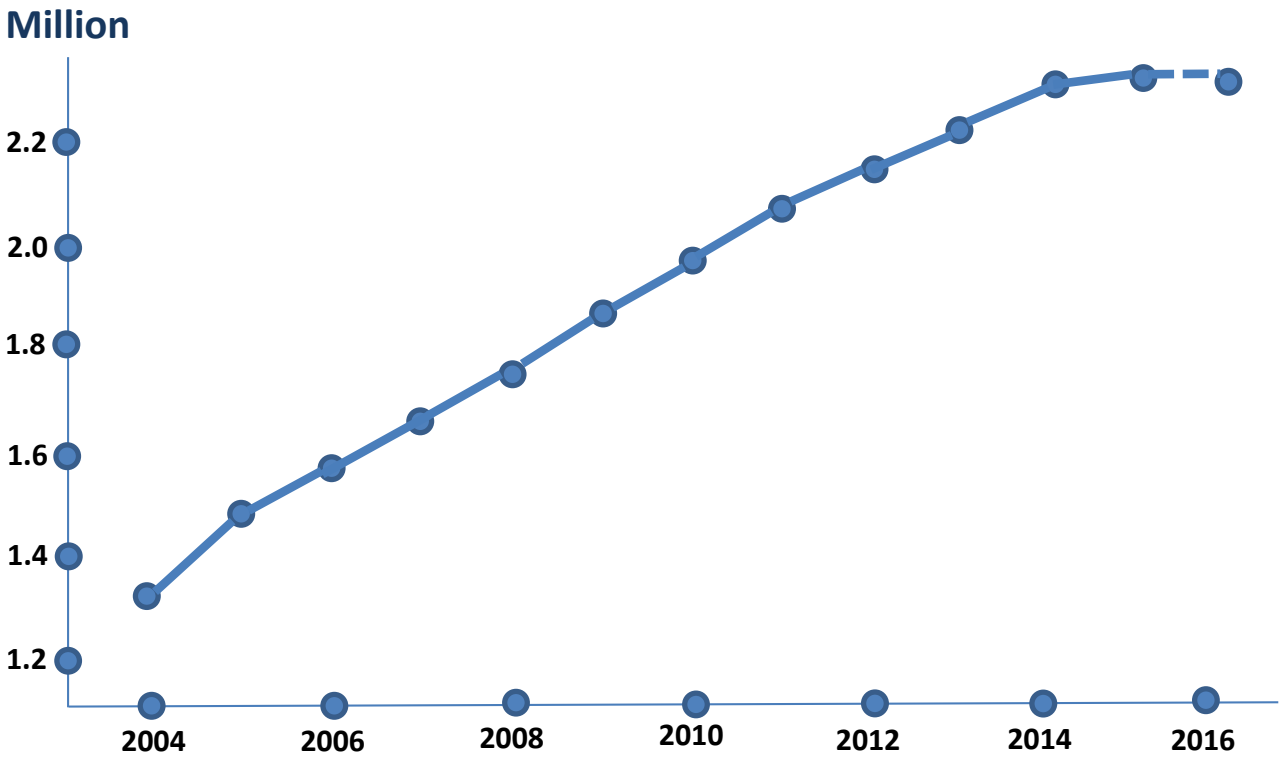
Tools and DataBases



GlyGen: Computational and Informatics Resources for Glycoscience

This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.

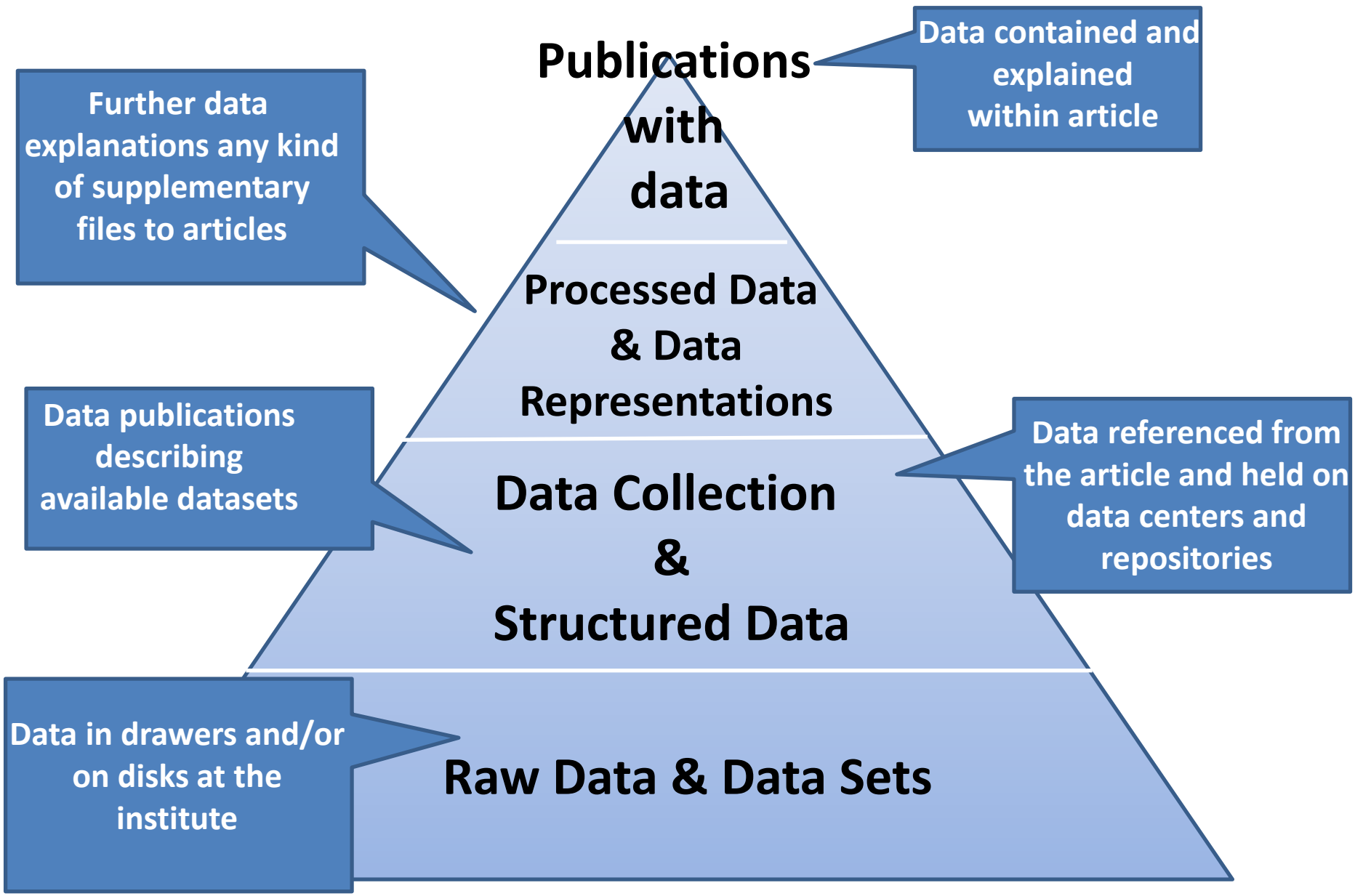
An avalanche of data…



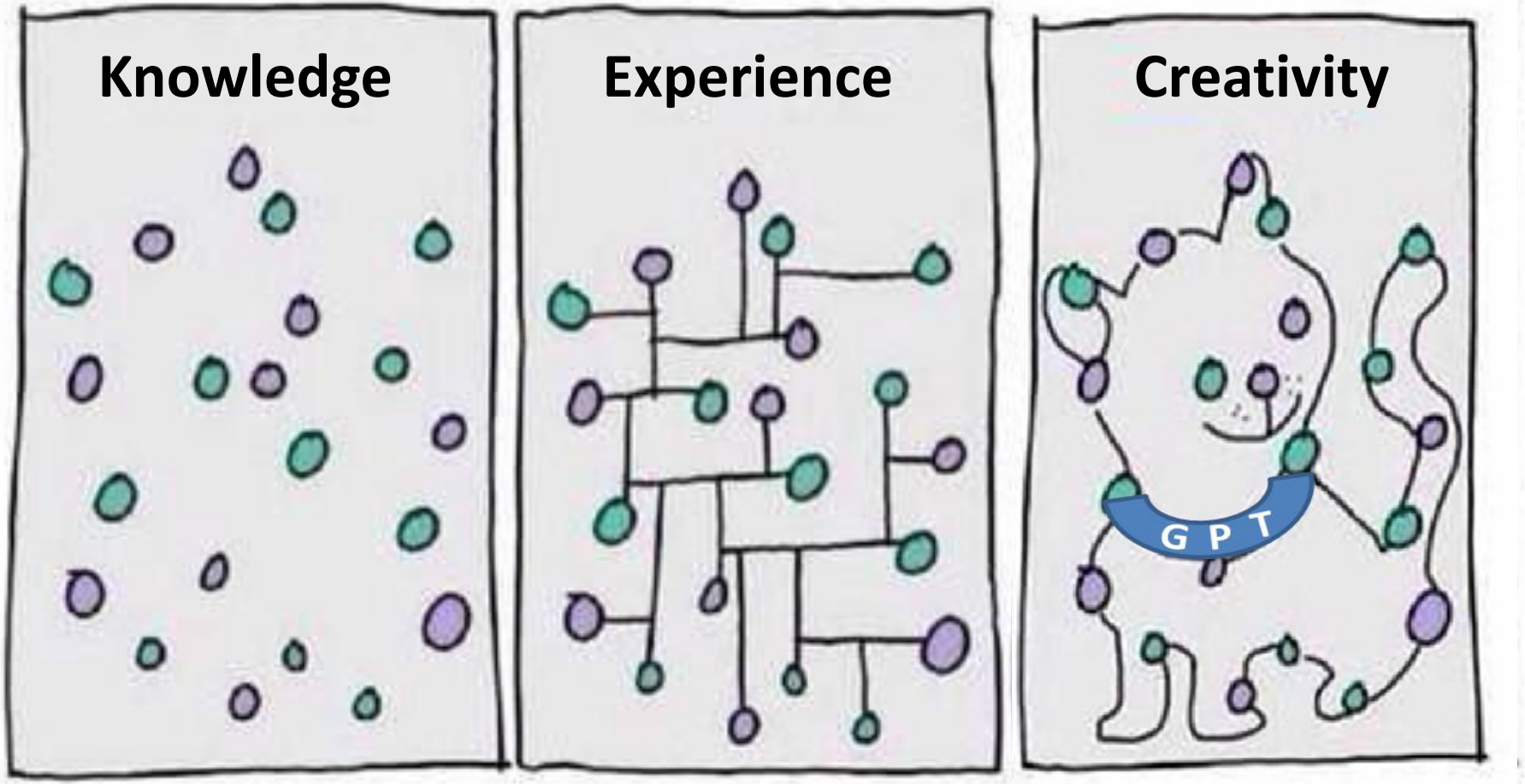
Global scientific output doubles every nine-years

Number of active researchers world-wide 8 Millions

An avalanche of data…



Knowledge, Experience, Creativity



Be FAIR to Glycans…

Update on Standards: Glycan data management and exchange require consolidation and compliance to standards,: Minimum Information Data Required for Glycomics (***MIRAGE***)

FAIR Principles; ***Findability, Accessibility, Interoperability and Reusability.***
Many data are not fully characterized, the lack of information on the metadata (explaining and characterizing the measured or computed data), the ontologies relationships in metadata), and the workflow of different research groups are difficult to adjust. ***Most research data are neither, findable nor interoperable.***

TRUST Principles: ***Transparency, Responsibility, User focus, Sustainability, Technology***

Cross-Referencing: Linking experimental, theoretical, and biological data using **common schemes** and **ontology** will generate a new level of Glycoscience

Data Modeling: Implementing multiscale data (spatial & temporal) faces heterogeneities: simulation steups, force fields, meaning and representation of the produced data
Need for selection and compressions stratefies compatible with the type and amount of data

Big Data and AI Approach : *Standardized, structured & well annotated data required to Deep Learning methods*

It is to realize that
Structural Glycoscience
may be fun and challenging

I Thank you and invite you to visit….

Glycopedia

GlycoPedia

news | e-chapters | resources