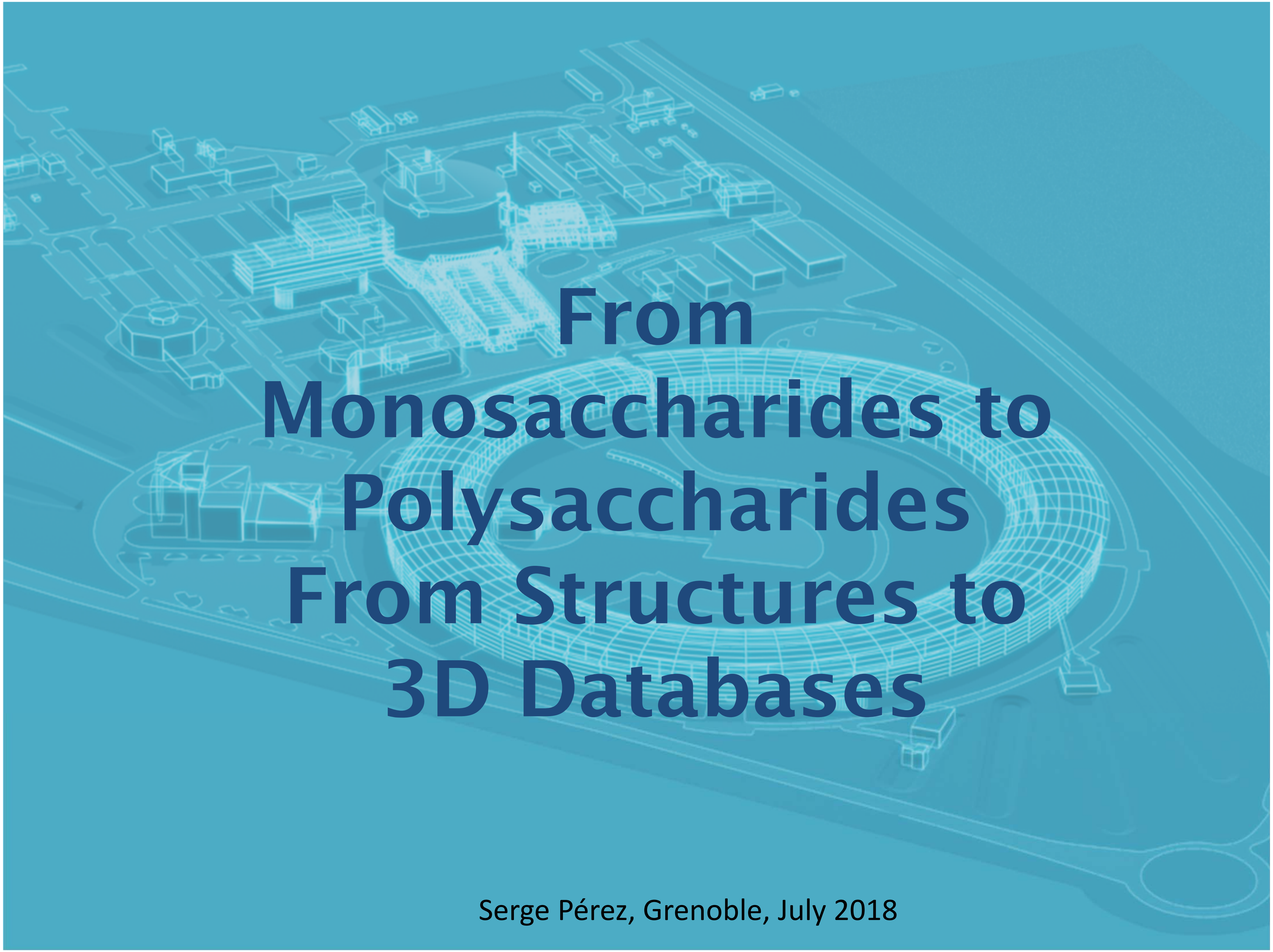
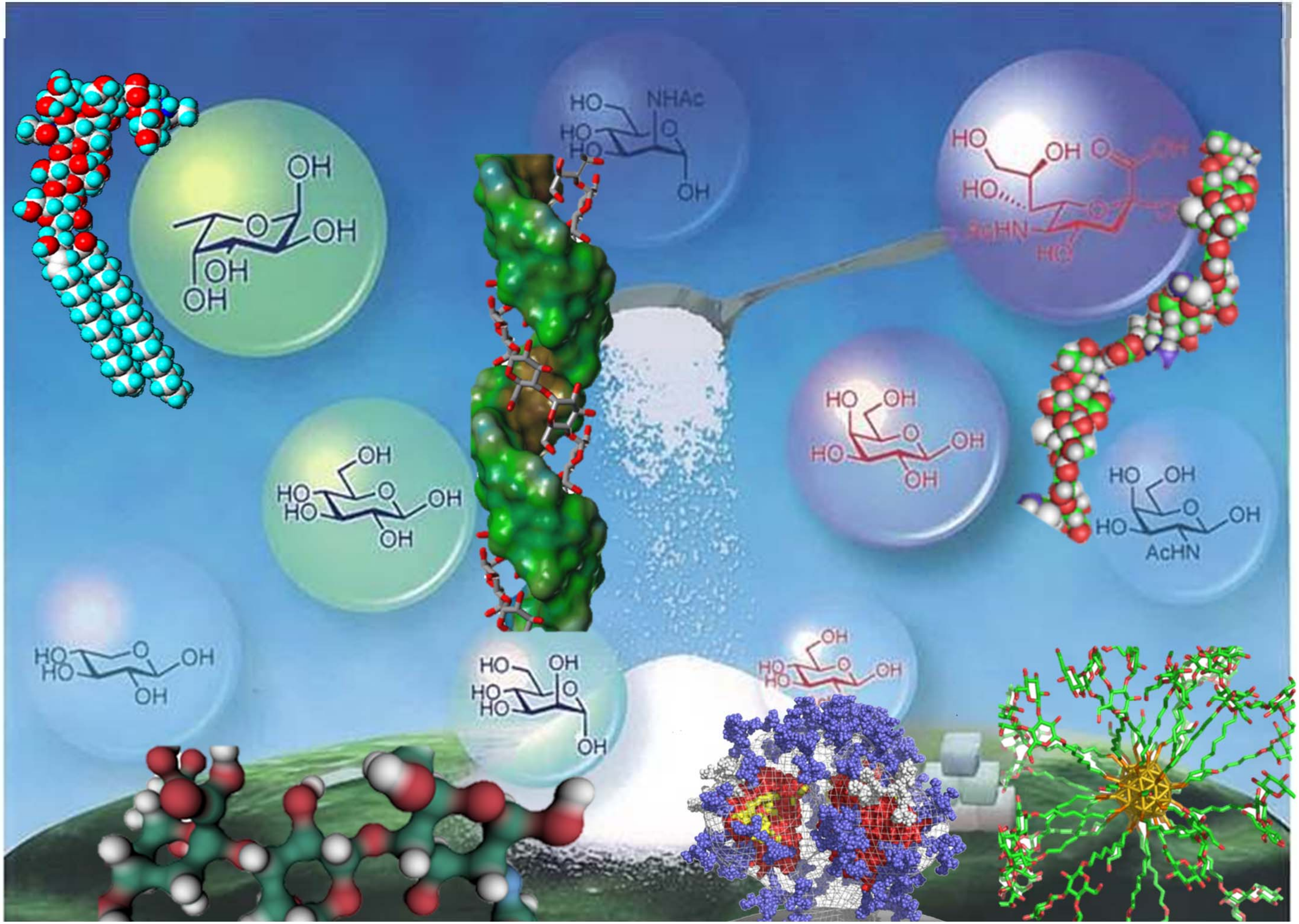
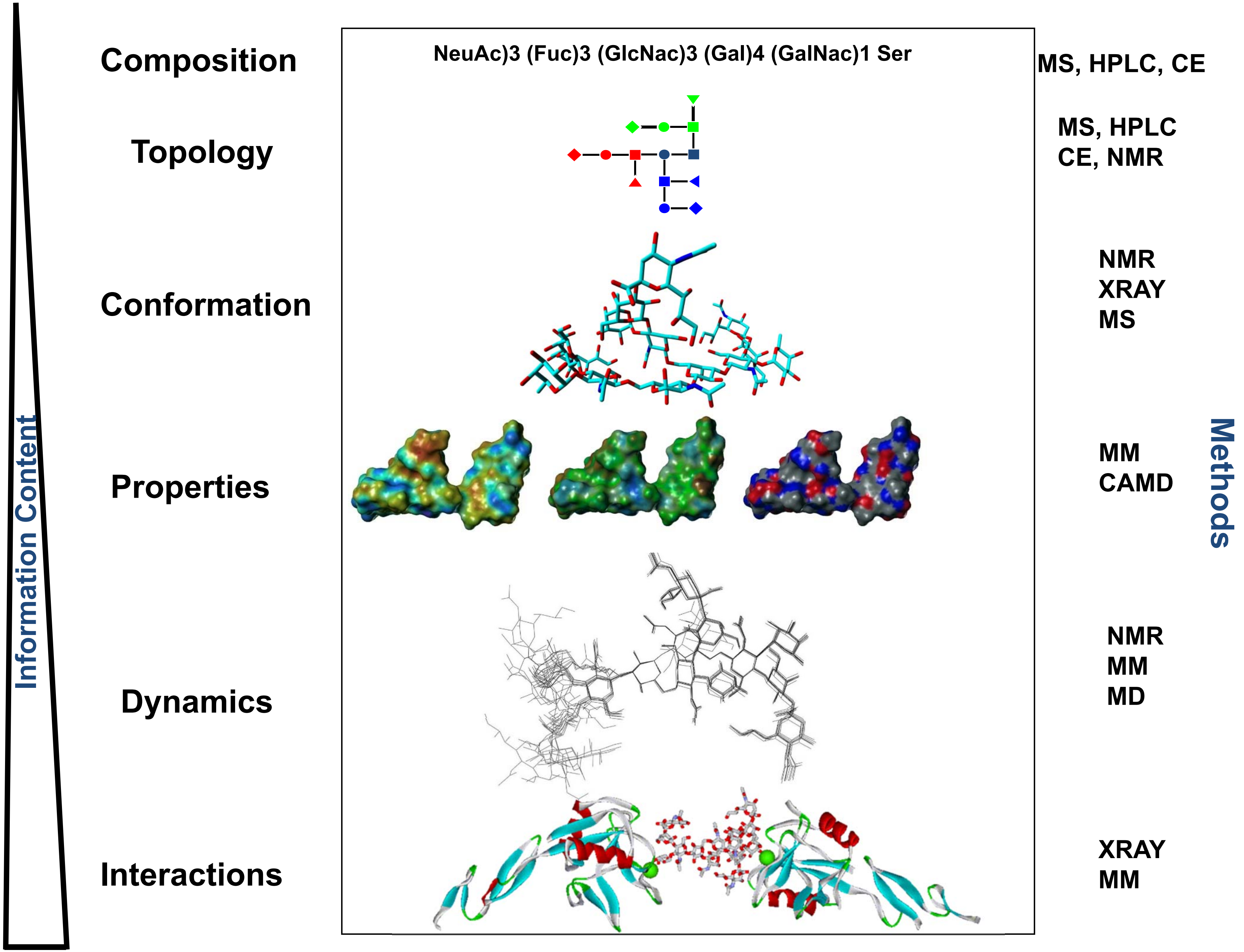


From Monosaccharides to Polysaccharides From Structures to 3D Databases

Serge Pérez, Grenoble, July 2018





Information Content

Composition

NeuAc)3 (Fuc)3 (GlcNac)3 (Gal)4 (GalNac)1 Ser

MS, HPLC, CE

Topology

MS, HPLC
CE, NMR

Conformation

NMR
XRAY
MS

Properties

MM
CAMD

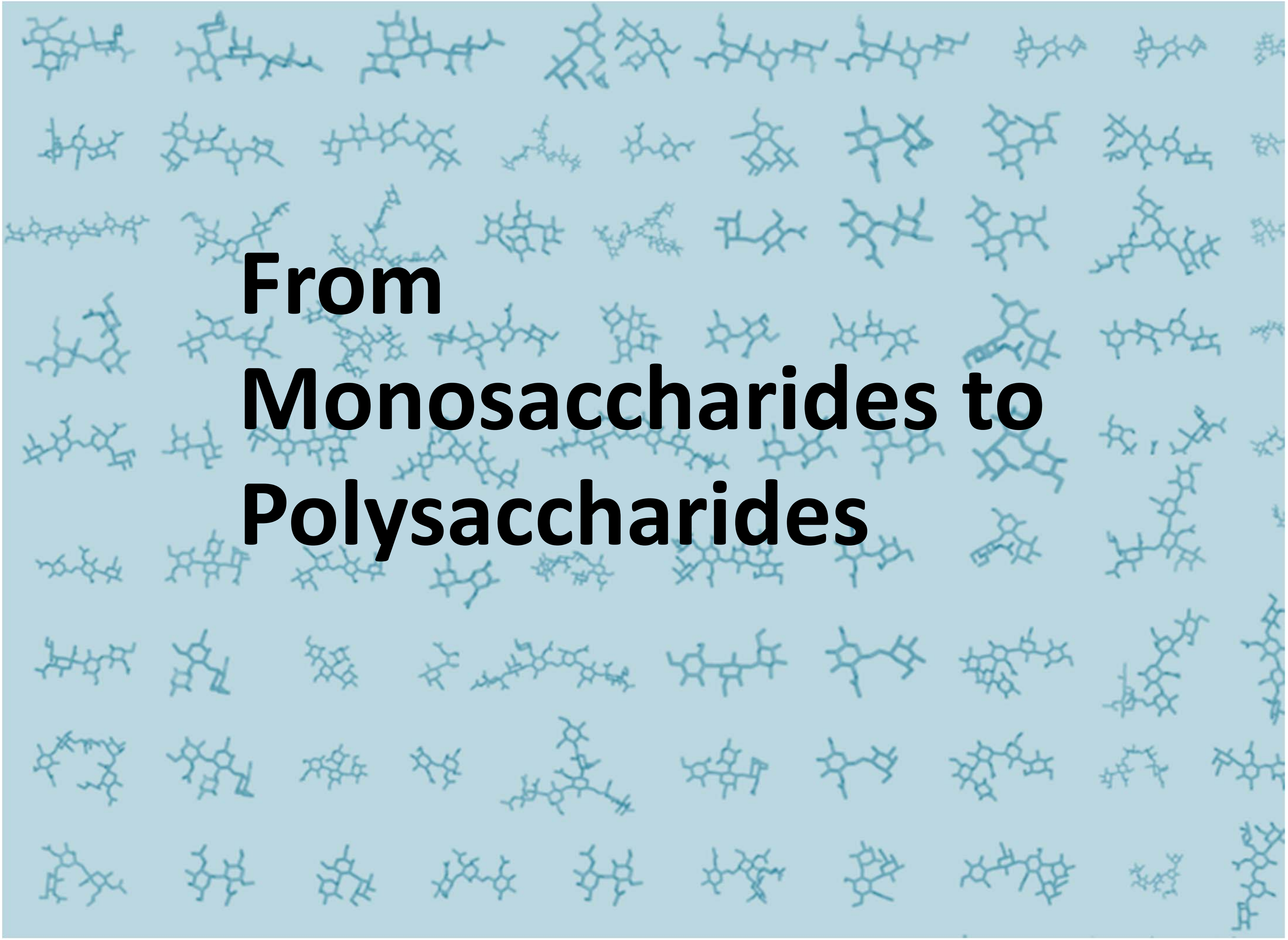
Dynamics

NMR
MM
MD

Interactions

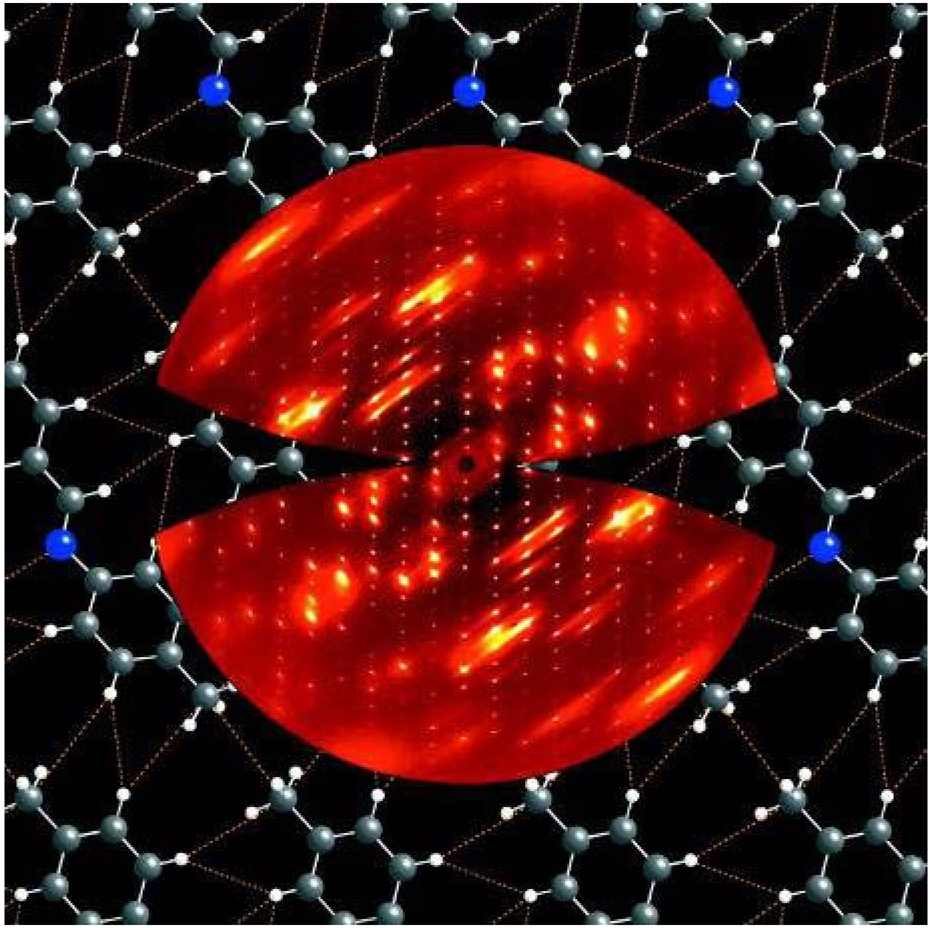
XRAY
MM

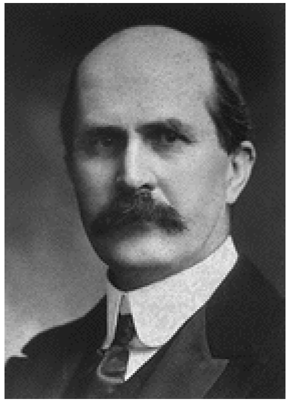
Methods



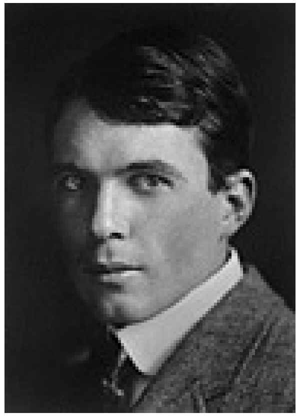
From Monosaccharides to Polysaccharides

International Year of Crystallography





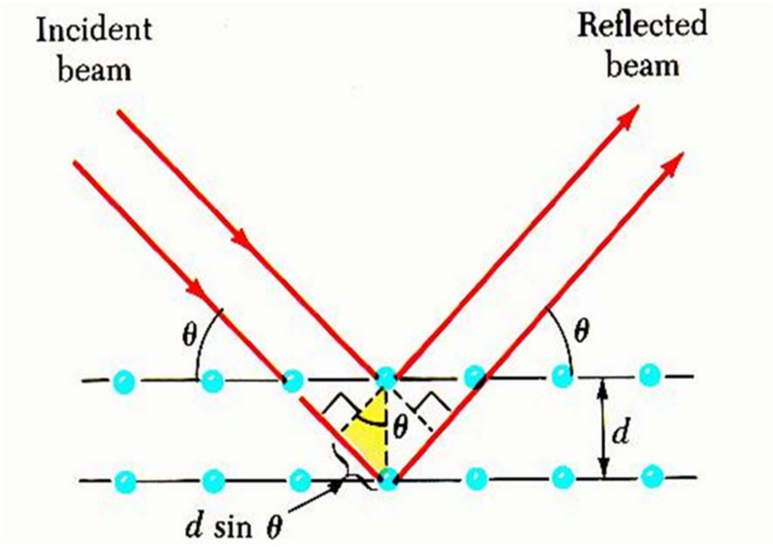
W. L. Bragg



W. L. Bragg

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

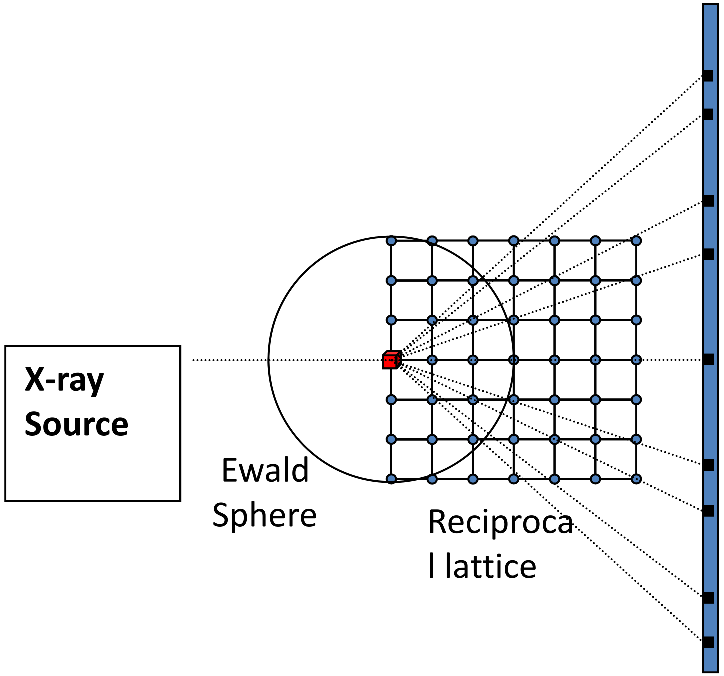
It also commemorated the 400th anniversary of Kepler’s observation in 1611 of the symmetrical form of ice crystals, which began the wider study of the role of symmetry in matter.

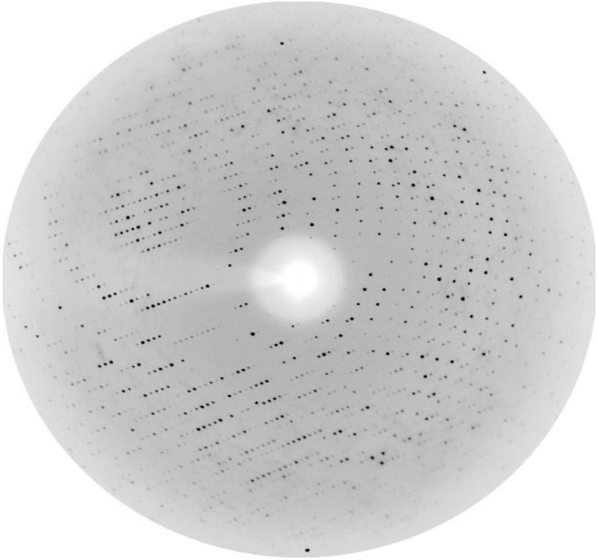


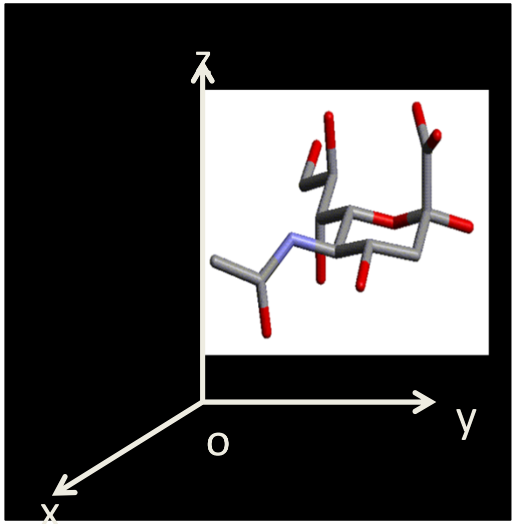
X-ray interact with the spatial distribution of Valence electrons.

Neutrons are scattered by the atom nucei.

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







N atoms : 3N observables

Crystallography of Carbohydrates

Molecular & Crystal Structures of Carbohydrates

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

Crystalline Conformations of Oligosaccharides in Proteins

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

Crystalline Conformations of Polysaccharides

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

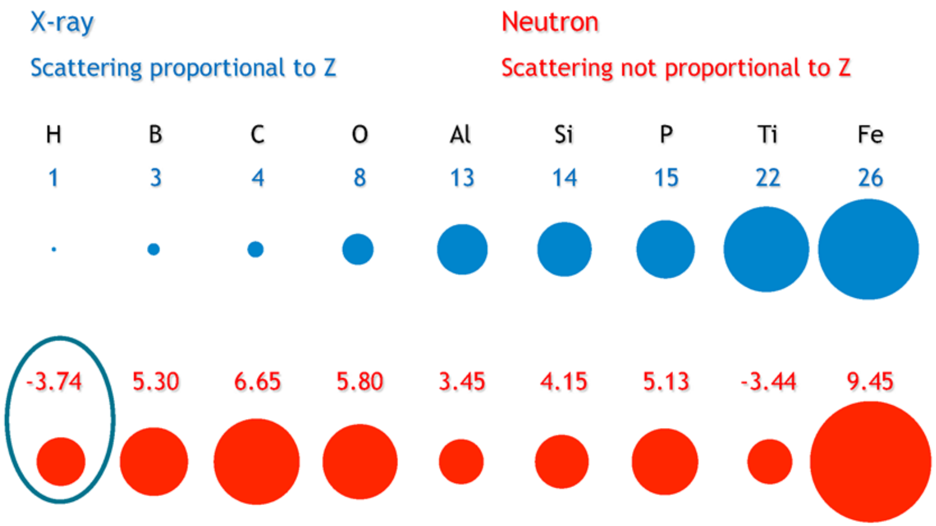
Molecular & Crystal Structures of Carbohydrates

Experimental Conditions and Limitations (X and N)

X-ray and Neutron have wavelengths in the same order as the interatomic distances (Angstrom).

Electron are the scattering elements of the incident X-ray

Nuclei are the scattering elements of the incident Neutron radiation


X-ray

Scattering proportional to Z

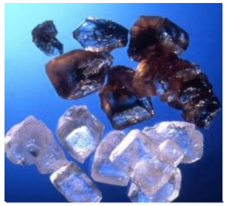
Neutron

Scattering not proportional to Z

H	B	C	O	Al	Si	P	Ti	Fe
1	3	4	8	13	14	15	22	26
-3.74	5.30	6.65	5.80	3.45	4.15	5.13	-3.44	9.45

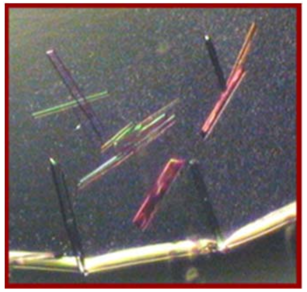
.

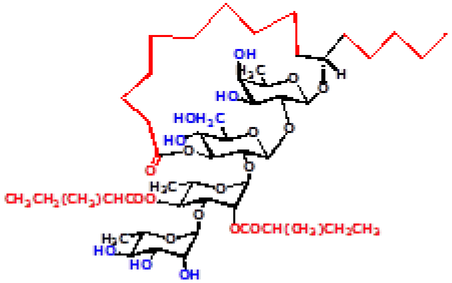
Single crystals usually grown by slow evaporation of saturated solution under well controlled environments

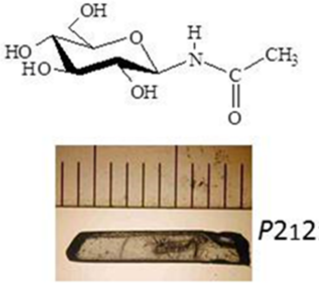


X-ray: Dimensions 0.2 – 0.5 mm / Synchrotron X-ray : 20-30 µm

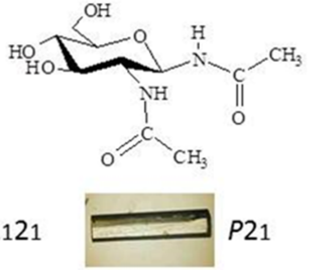
Neutron: Dimensions over 1.0 mm all dimensions



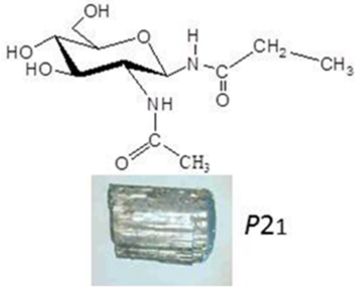




P212121



P21

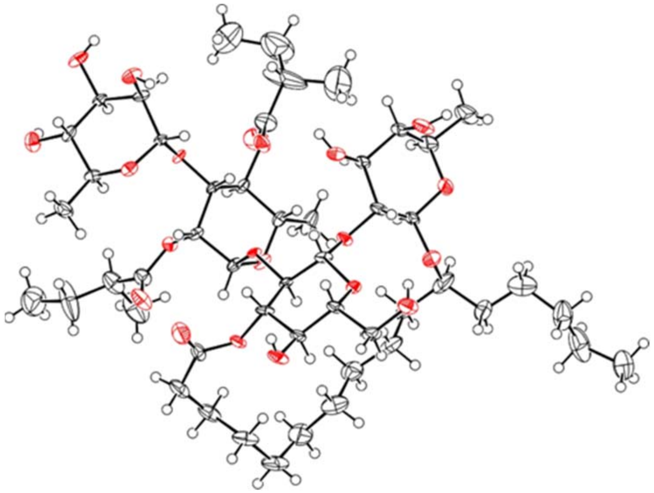


P21

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

Understanding a Structural Report

Unit Cell Parameters (a, b, c, α, β, γ); Space Group

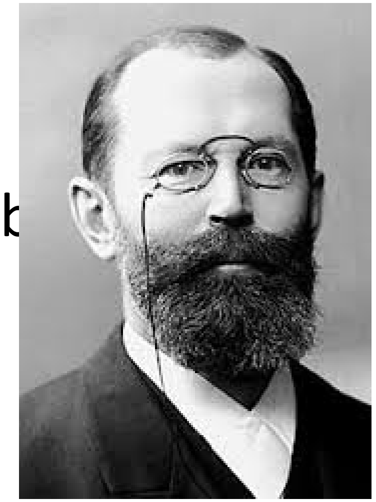
Fractional atomic coordinates content of the asymmetric unit:
(x/a; y/b; z/c)
Anisotropic Temperature Factors (ORTEP representation ellipsoids)

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

Configuration !!!!!

Intra- and Inter molecular hydrogen b

Analysis of : Hydration features
Packing features



Emil Fisher

Johannes Bijvoet

Molecular & Crystal Structures of Carbohydrates

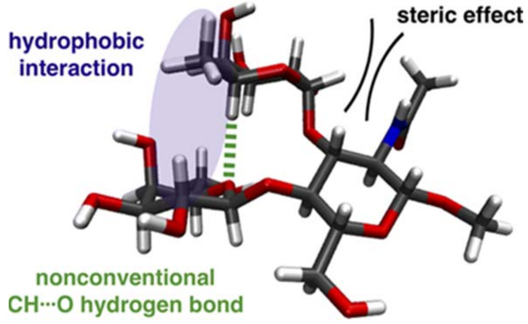
Hydrogen Bonding in Crystalline Oligosaccharides

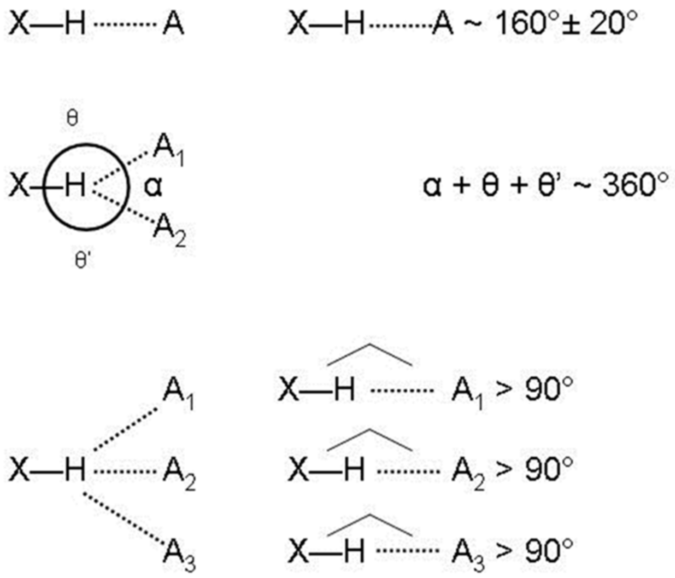
Analysis of high accurate X-ray analysis – Neutron diffraction

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

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

CHO bonds



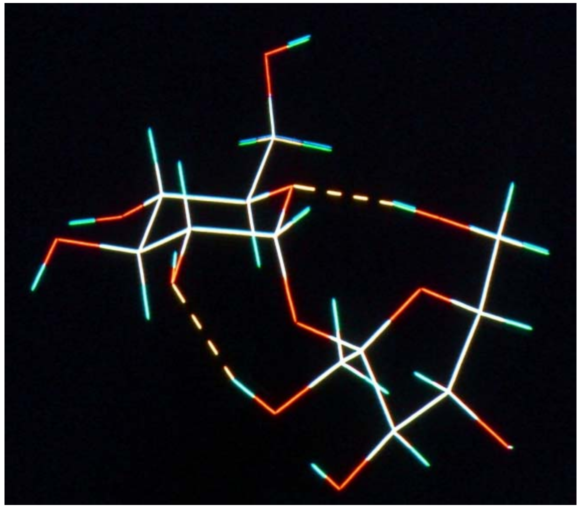


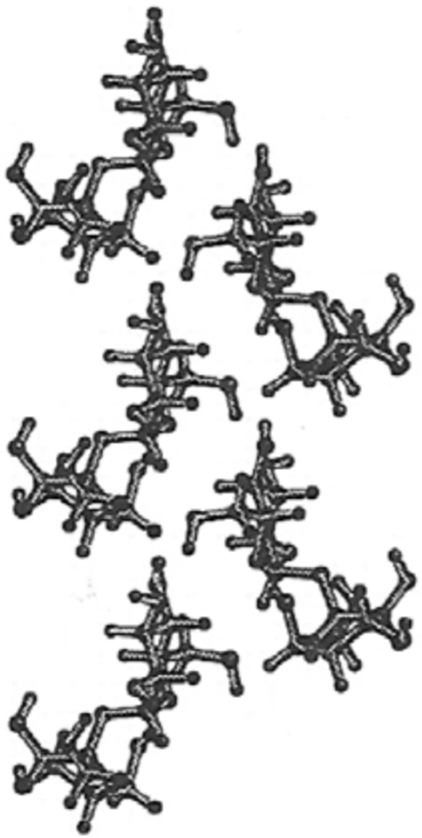
Maximize the Hydrogen Bond interactions throughout the participation of all hydroxyl groups and as many rings oxygen. Two and three-centered bonds

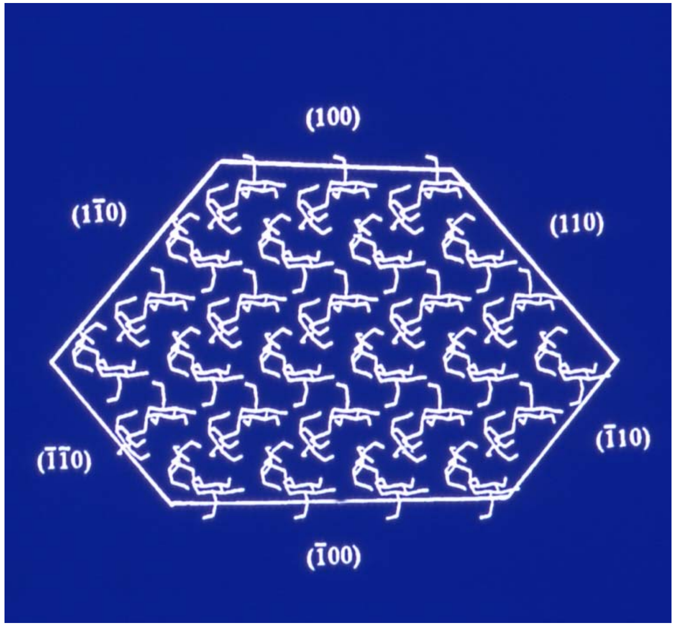
Maximize cooperativity by forming as many finite and infinite chains of hydrogen bonds as possible.

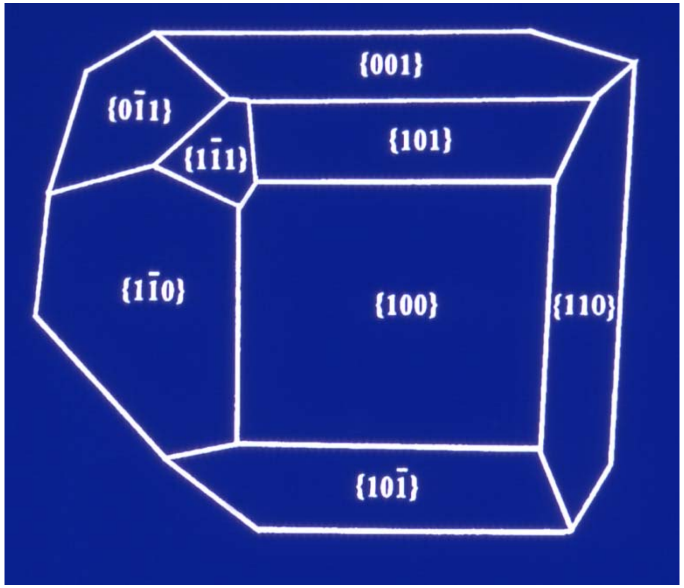
Molecular & Crystal Structures of Carbohydrates

Packing Features









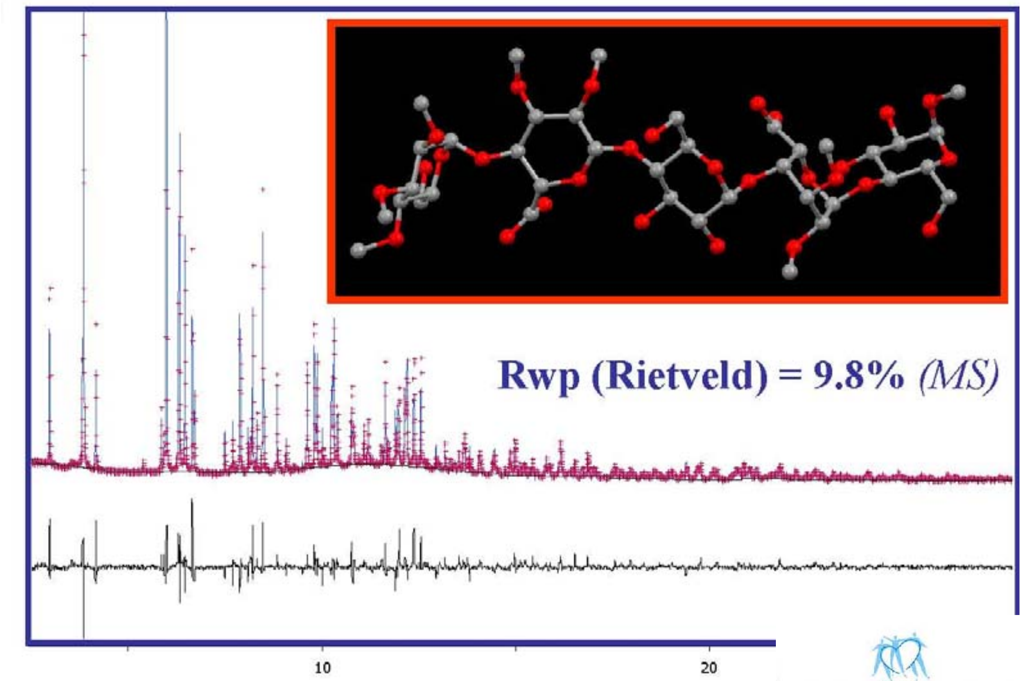
Molecular & Crystal Structures of Carbohydrates

Powder Diffraction

1. Identification of Crystalline Polymorphs

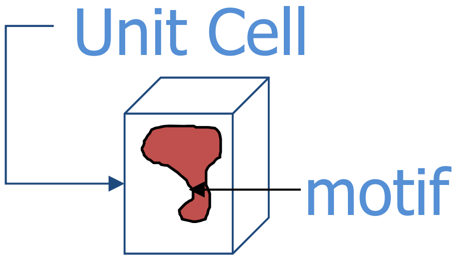
2. Solving Crystal Structures – Rievelt Method + Molecular Modelling

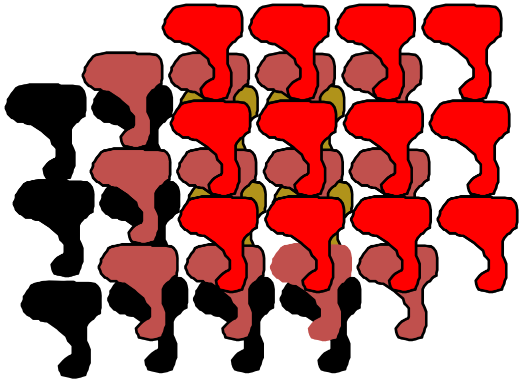
Synthetic Pentasaccharide
ID31@ESRF, λ **=0.8 A**
Monoclinic $P2_1$
a=15.54, b=8.83; c=17.67, β=94.6

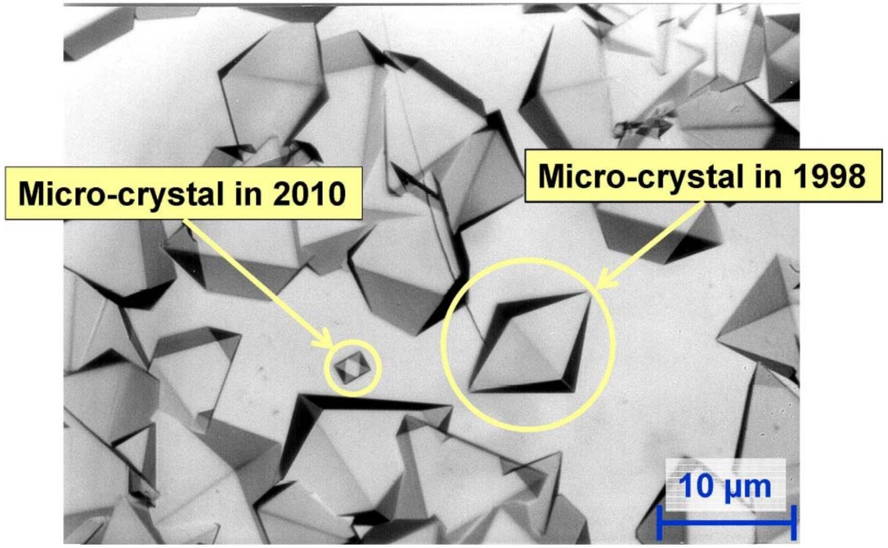


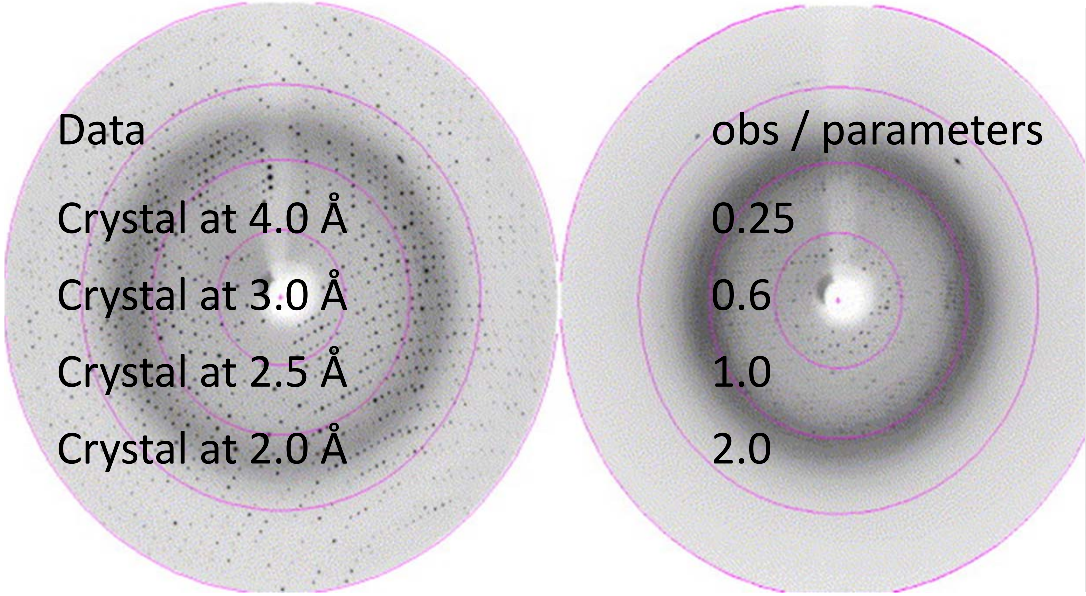
Crystalline Conformations of Oligosaccharides in Proteins

Experimental Conditions and Limitations

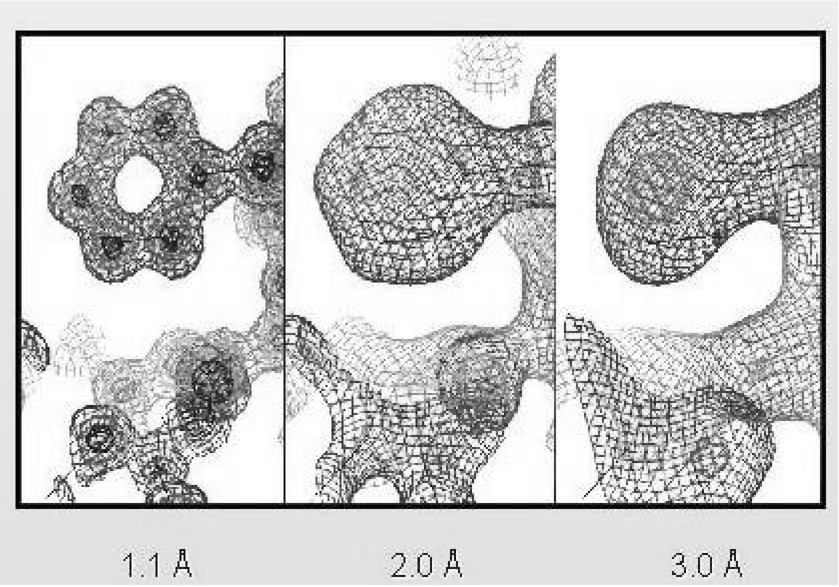


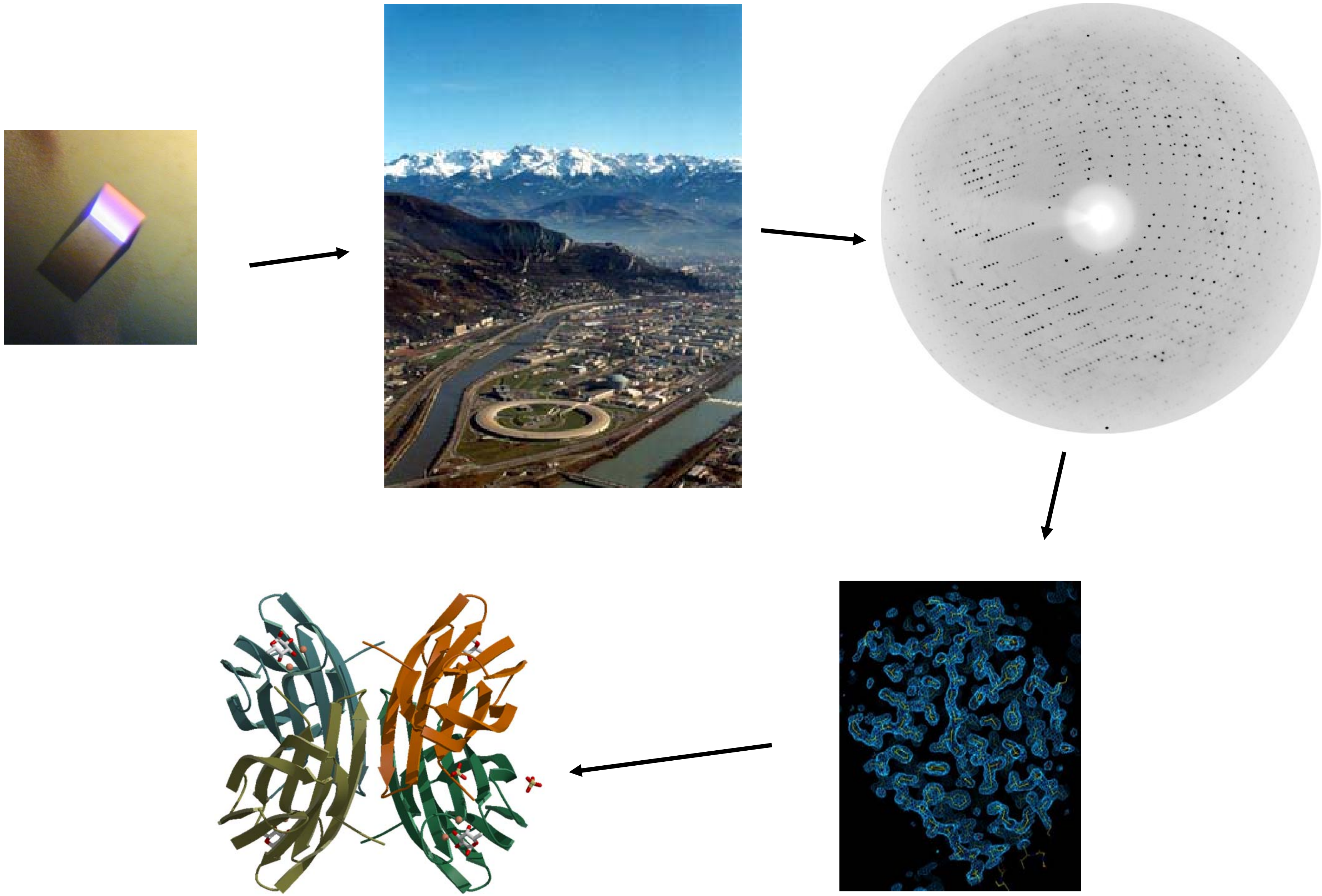




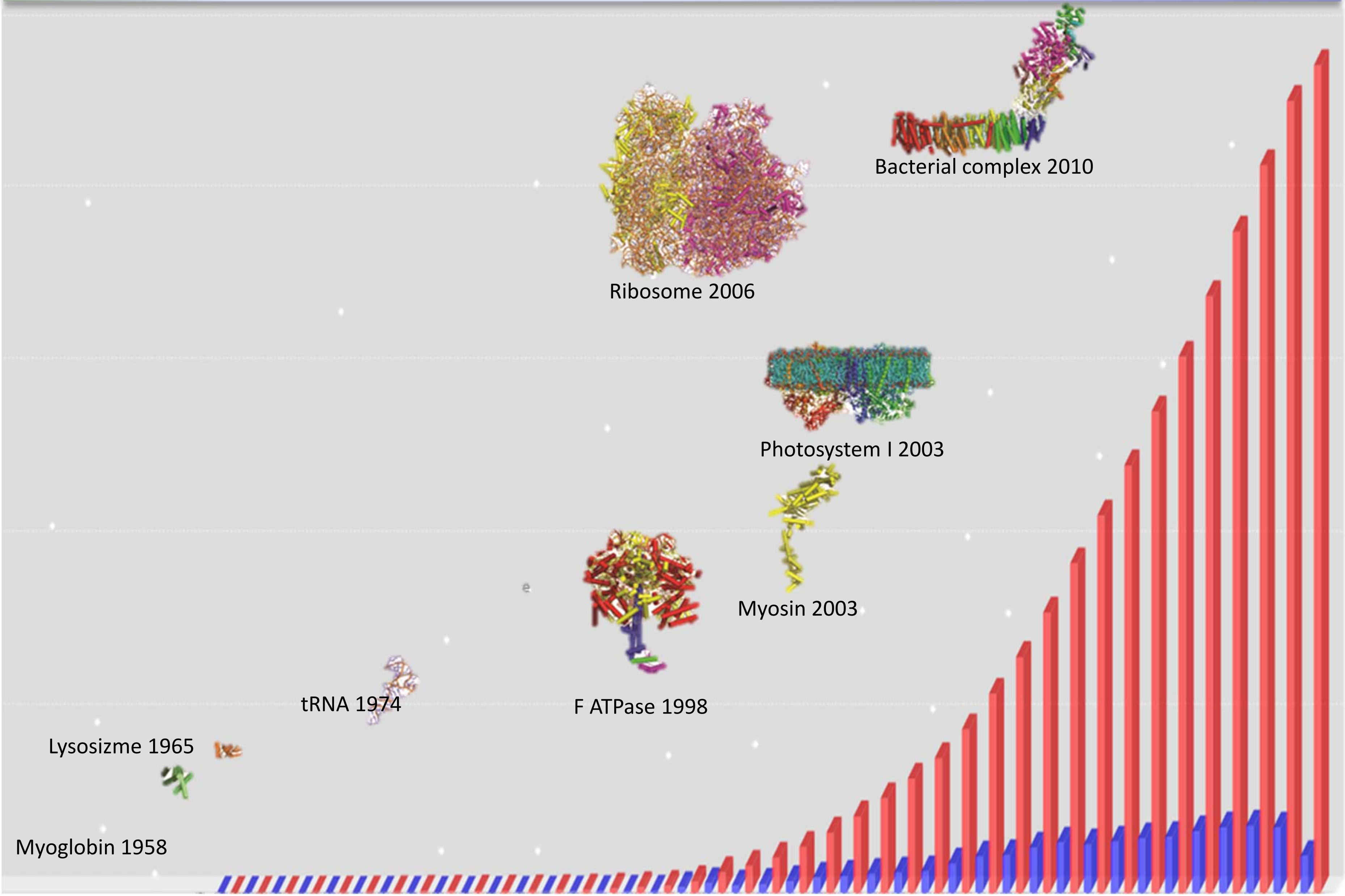


Data	obs / parameters
Crystal at 4.0 Å	0.25
Crystal at 3.0 Å	0.6
Crystal at 2.5 Å	1.0
Crystal at 2.0 Å	2.0



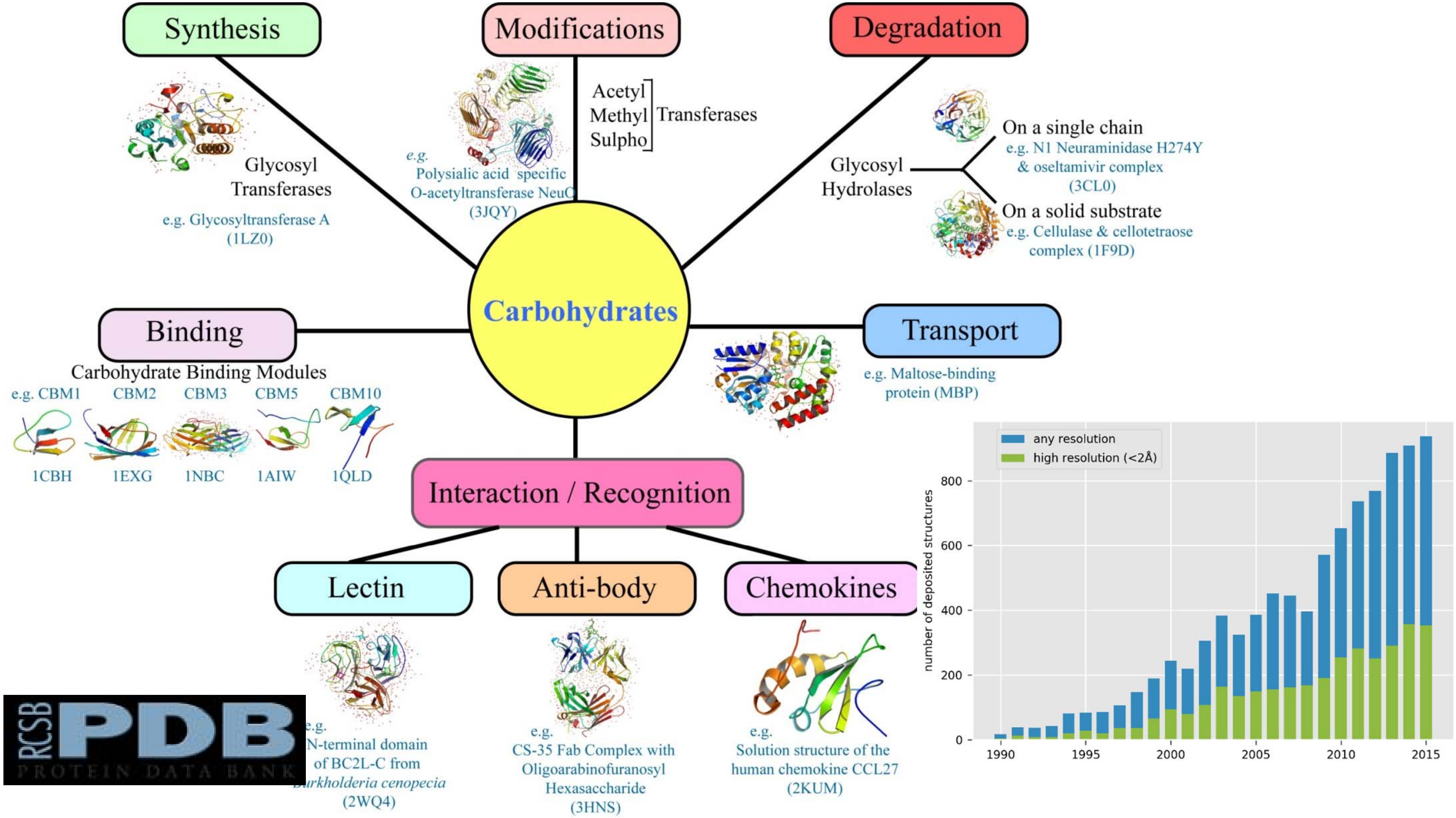


Yearly Growth of Total Protein Structures (Deposit / Total)



Crystalline Conformations of Oligosaccharides in Proteins

Protein-Carbohydrate Interactions



Carbohydrates

Synthesis

Glycosyl Transferases

e.g. Glycosyltransferase A (1LZ0)

Modifications

Acetyl
Methyl
Sulpho
Transferases

e.g. Polysialic acid specific O-acetyltransferase NeuO (3JQY)

Degradation

Glycosyl Hydrolases

On a single chain
e.g. N1 Neuraminidase H274Y & oseltamivir complex (3CL0)

On a solid substrate
e.g. Cellulase & cellotetraose complex (1F9D)

Binding

Carbohydrate Binding Modules

e.g. CBM1	CBM2	CBM3	CBM5	CBM10
1CBH	1EXG	1NBC	1AIW	1QLD

Transport

e.g. Maltose-binding protein (MBP)

Interaction / Recognition

Lectin

e.g. N-terminal domain of BC2L-C from *Burkholderia cenopecia* (2WQ4)

Anti-body

e.g. CS-35 Fab Complex with Oligoarabinofuranosyl Hexasaccharide (3HNS)

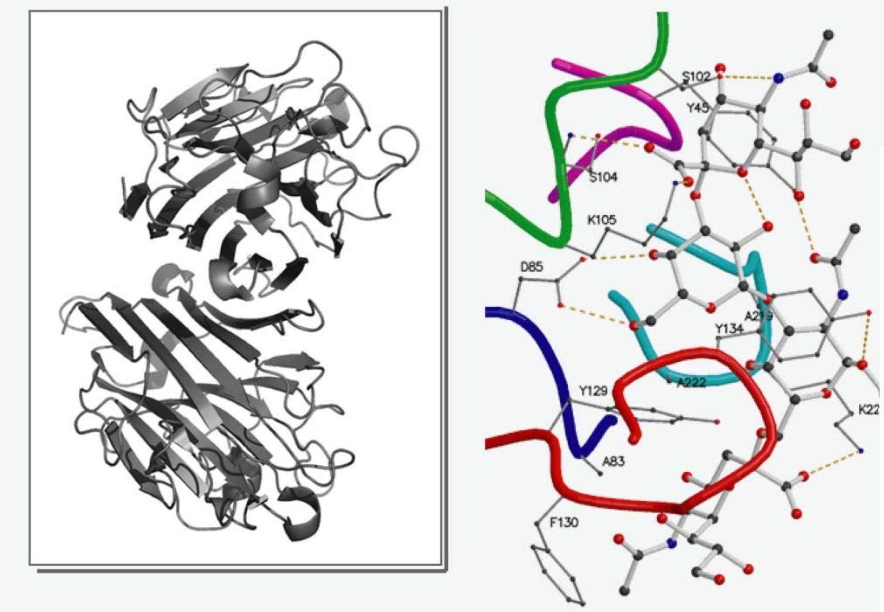
Chemokines

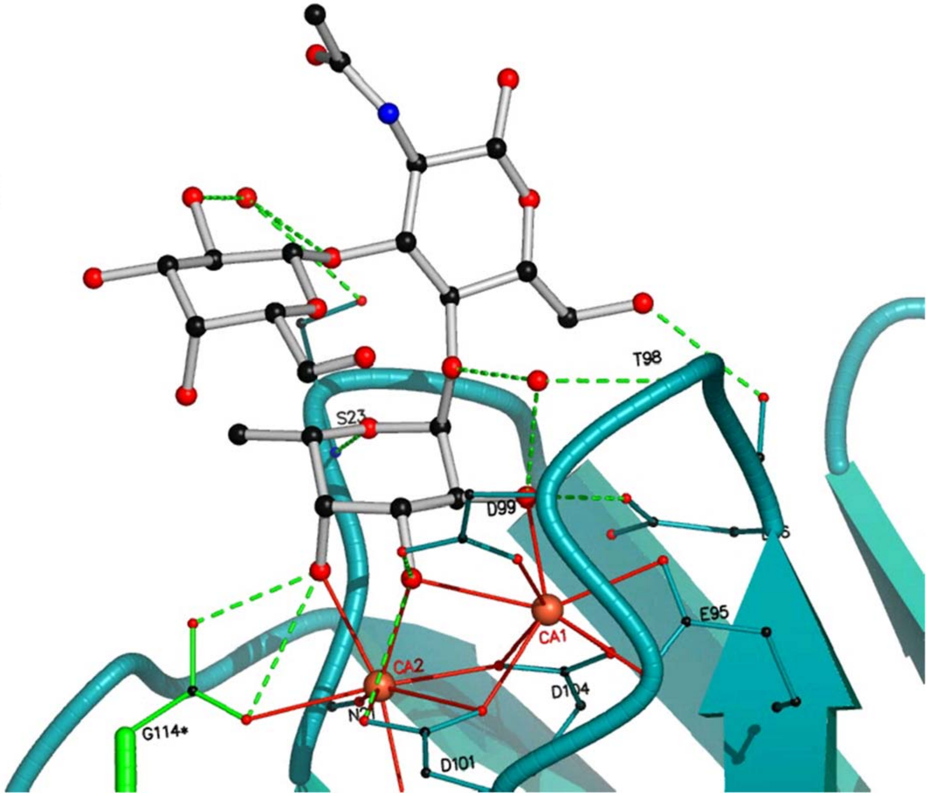
e.g. Solution structure of the human chemokine CCL27 (2KUM)

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

Crystalline Conformations of Oligosaccharides in Proteins

Oligosaccharides –Lectin Complexes





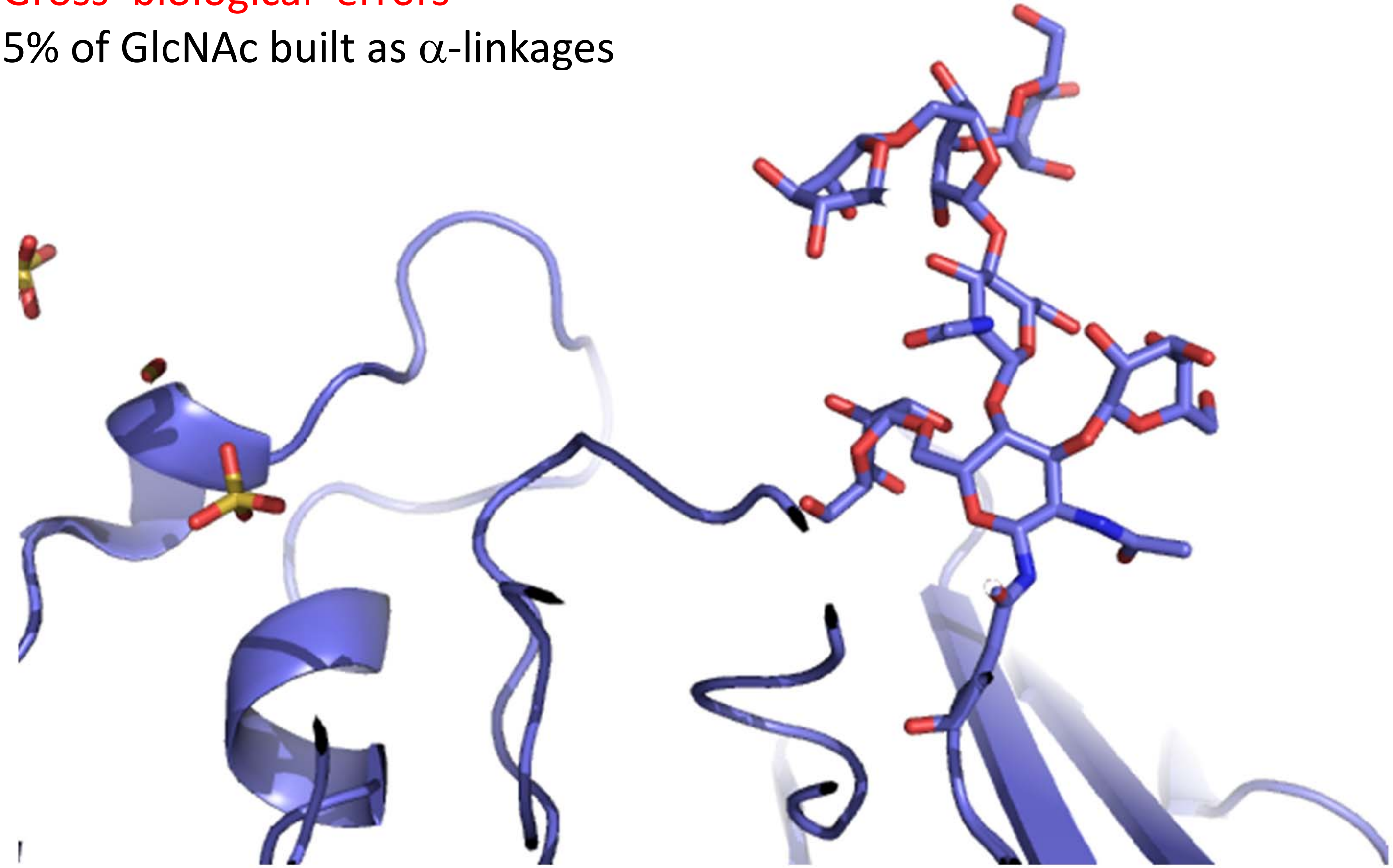
Errors in Modelling of Carbohydrates in Xstal Structures

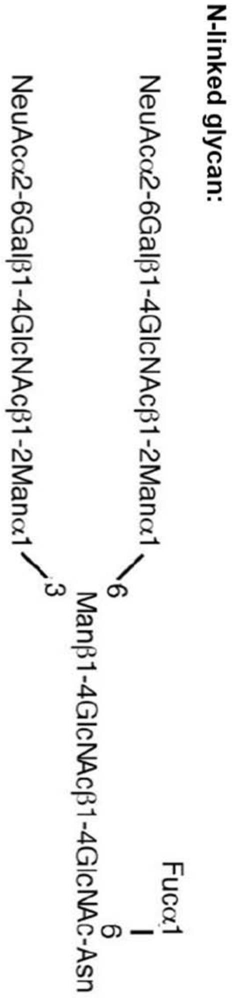
Analysis of N-glycan D-pyranosides

Poor density

Gross ‘biological’ errors

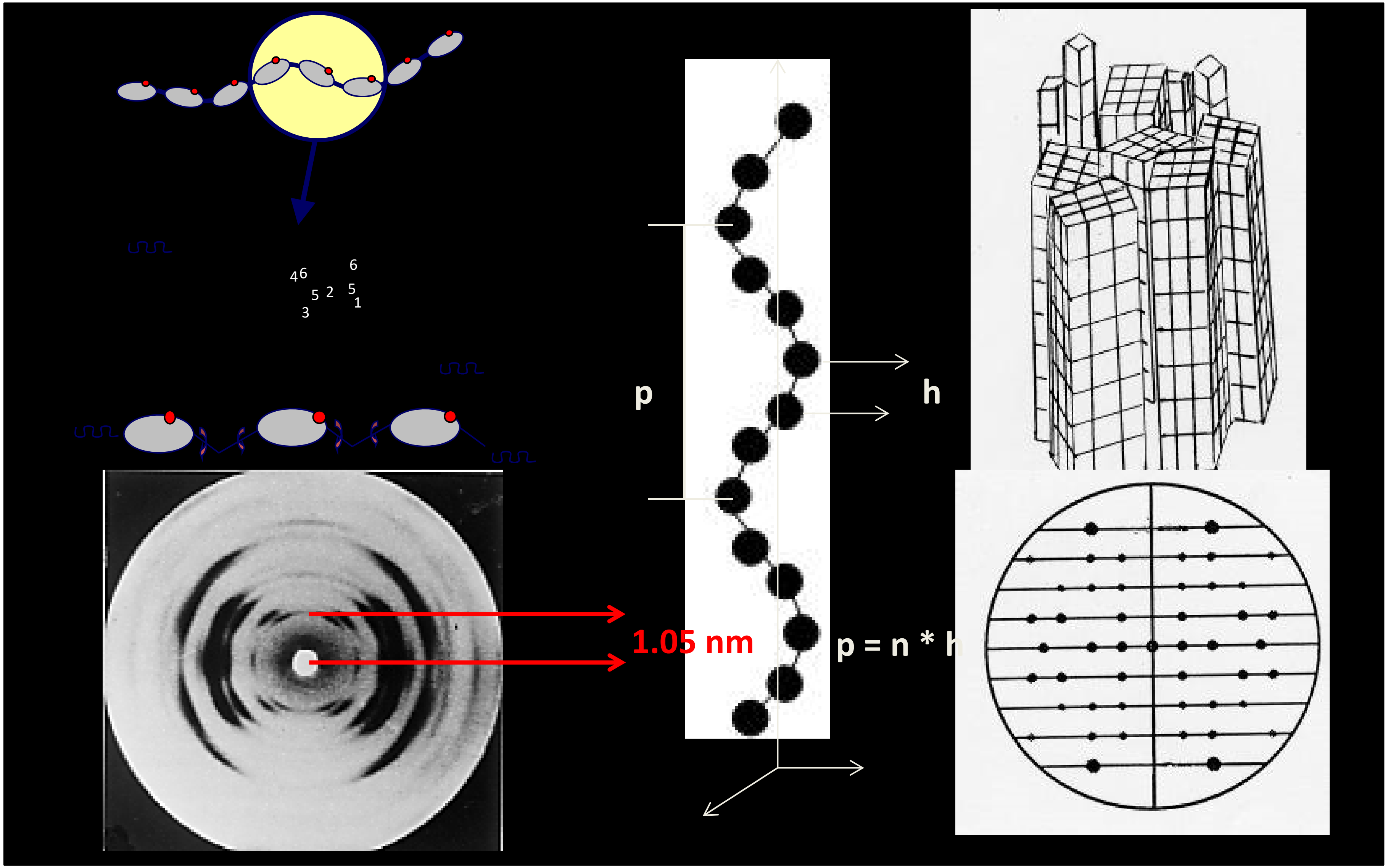
5% of GlcNAc built as α-linkages





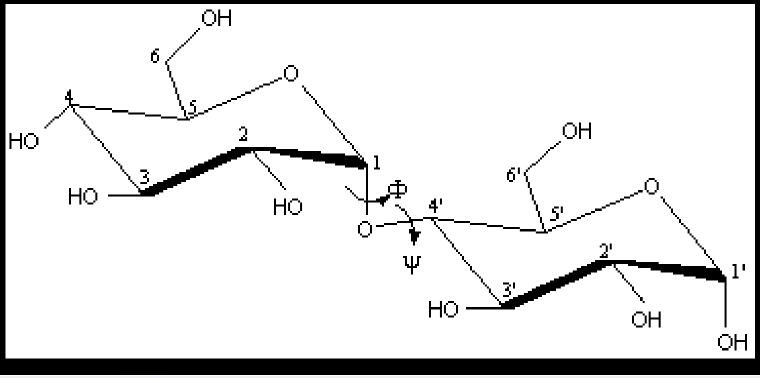
Crystalline Conformations of Polysaccharides

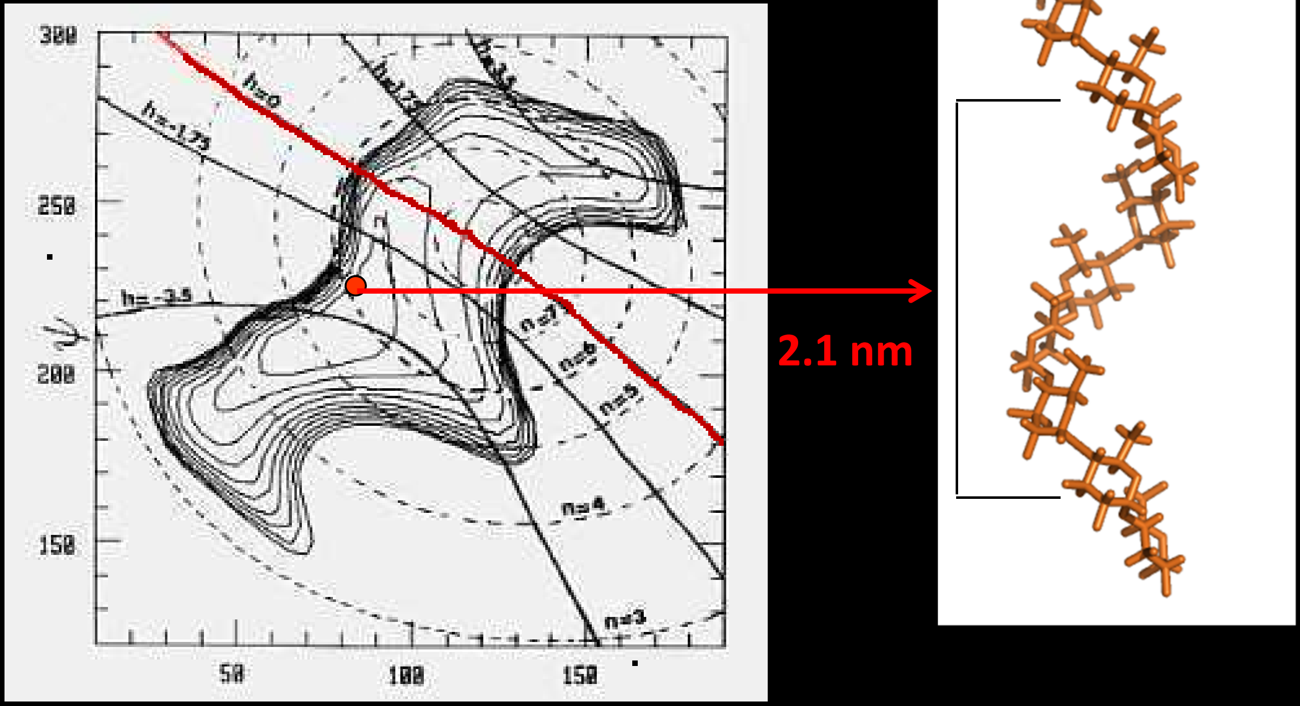
X-Ray Fiber Diffraction of Polysaccharides

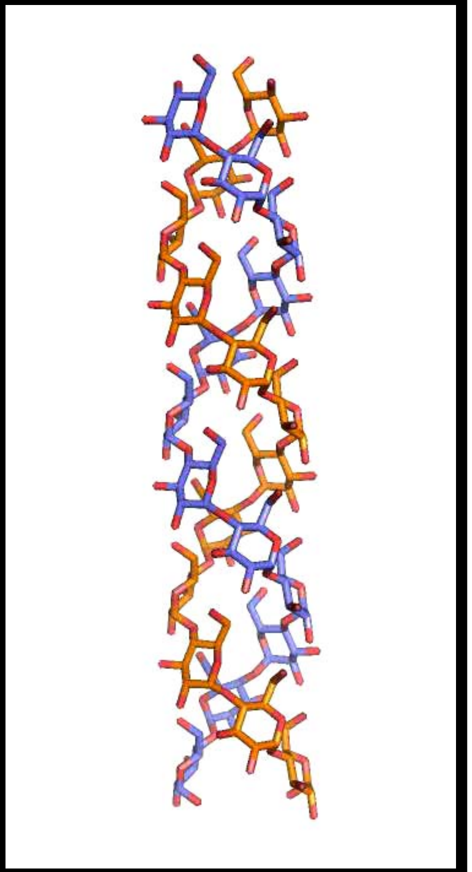


Crystalline Conformations of Polysaccharides

X-Ray Fiber Diffraction of Polysaccharides

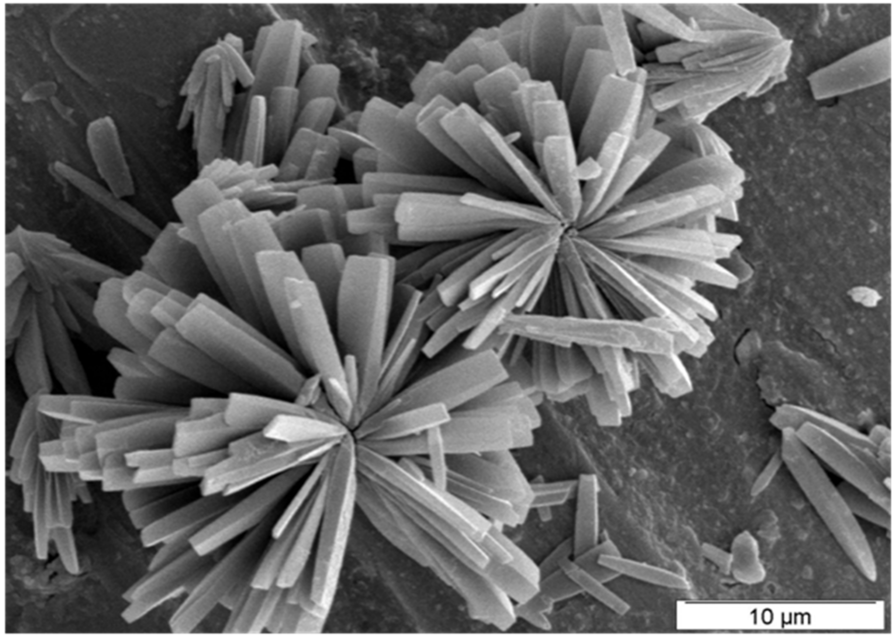


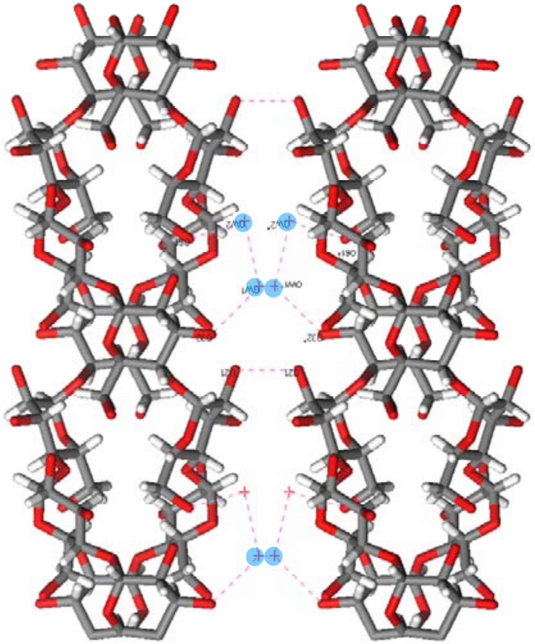


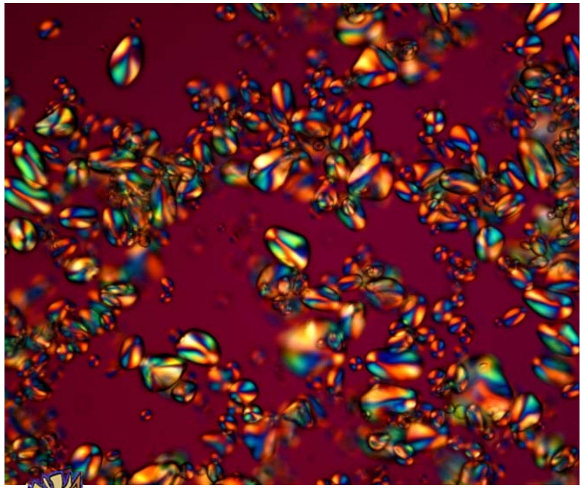


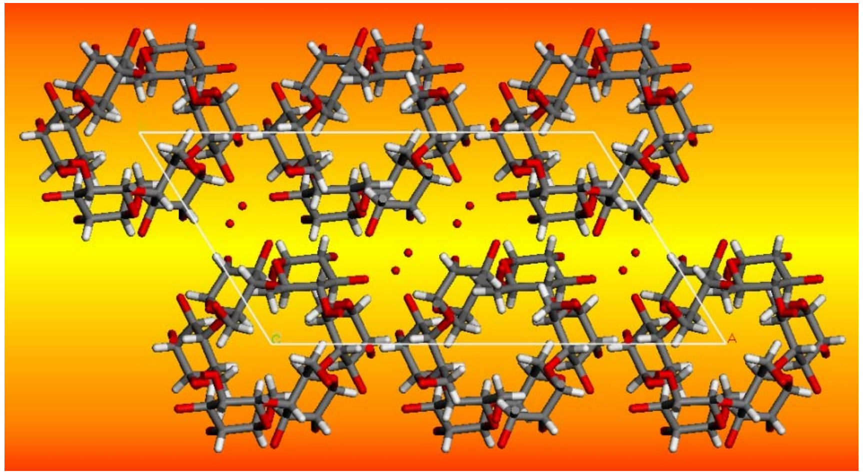
Crystalline Conformations of Polysaccharides

Synchrotron X-Ray Diffraction of Polysaccharides





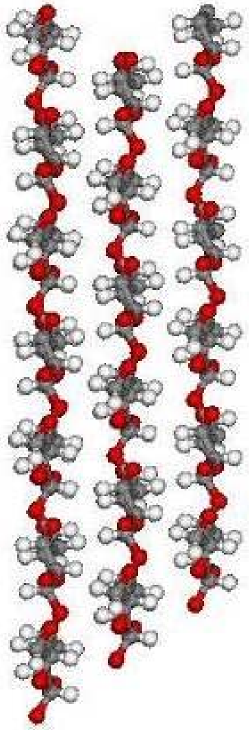


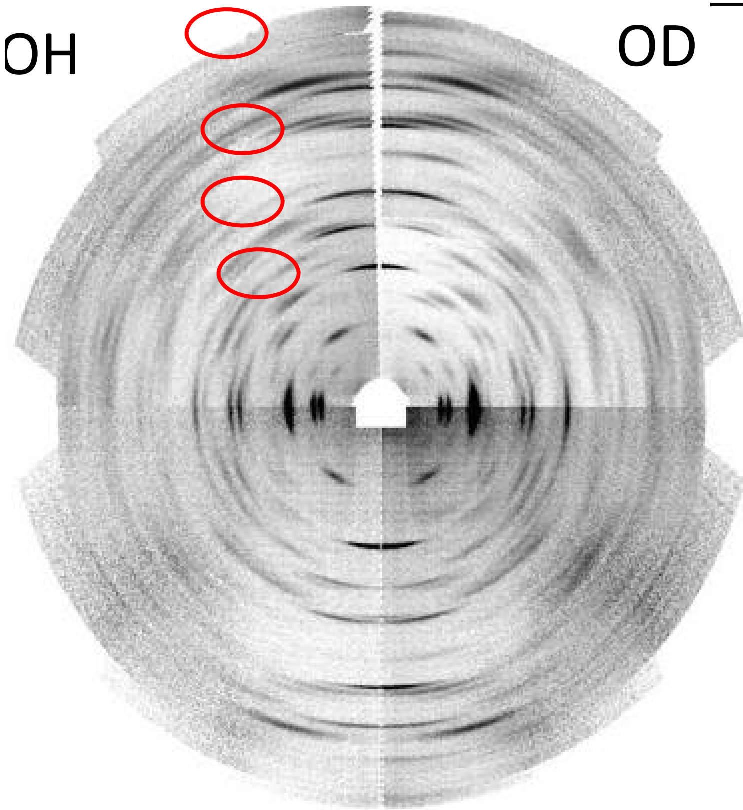


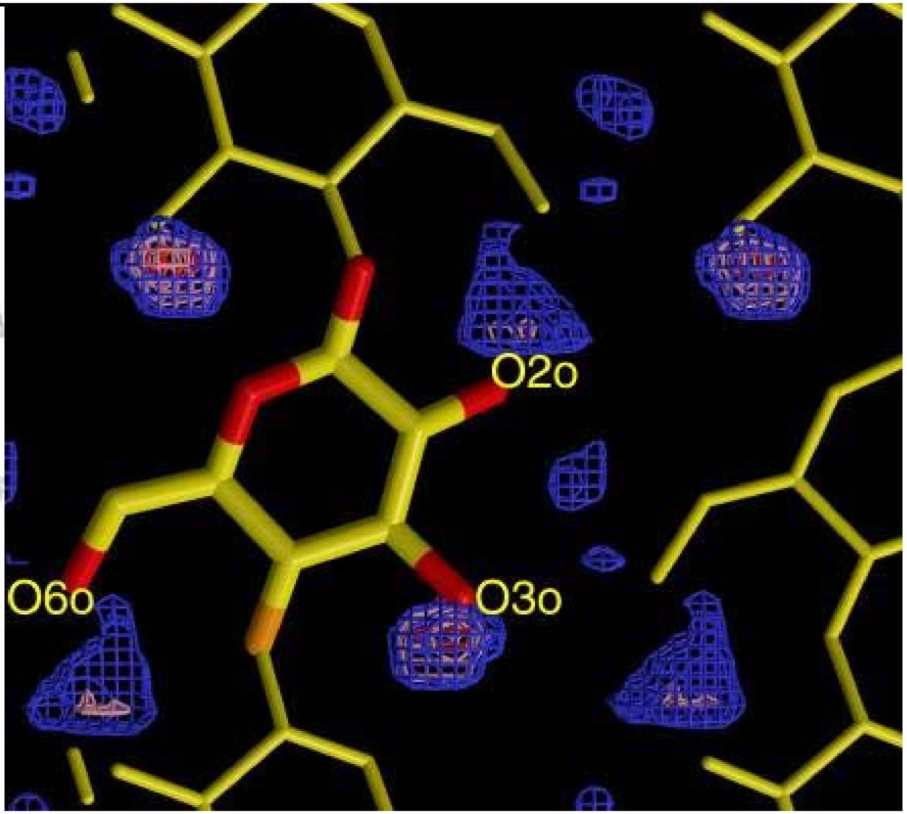
D. Popov, 2009, Macromolecules, 42, 1167-1174

Crystalline Conformations of Polysaccharides

X-Ray Fiber Diffraction using Synchrotron and Neutron Radiations







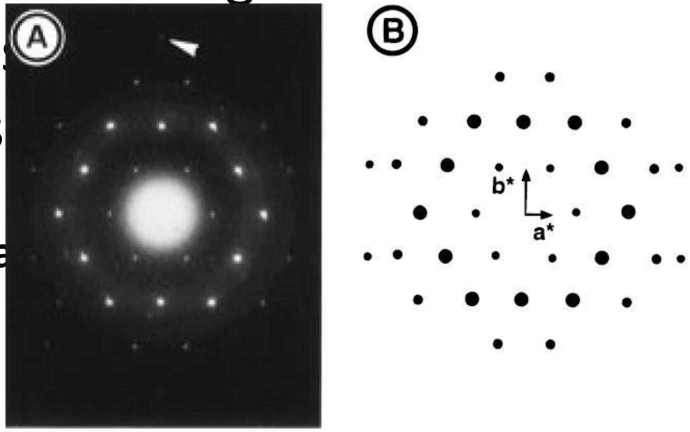
Y. Nishiyama et al., 2002, J. Am. Chem. Soc., 124, 9074-9082

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
photographic film, on imaging plates or us

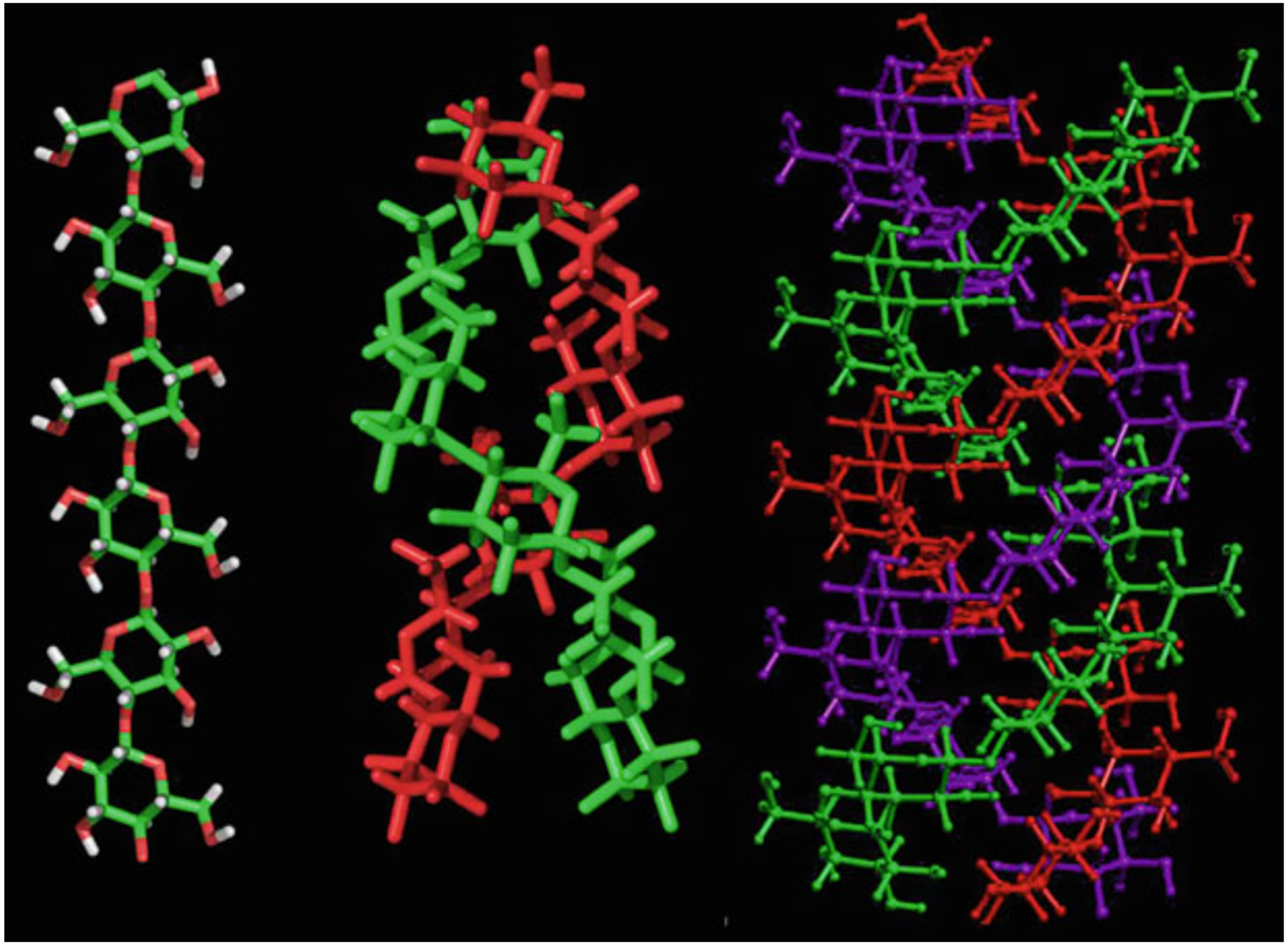


Electron diffraction in TEM is subject to several importa

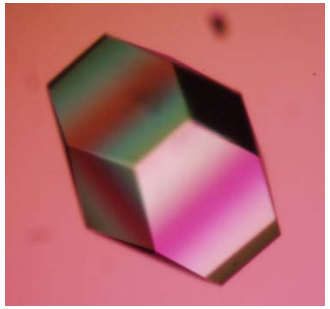
The sample to be studied must be electron transparent, meaning the sample thickness must be of the order of 100 nm or less.

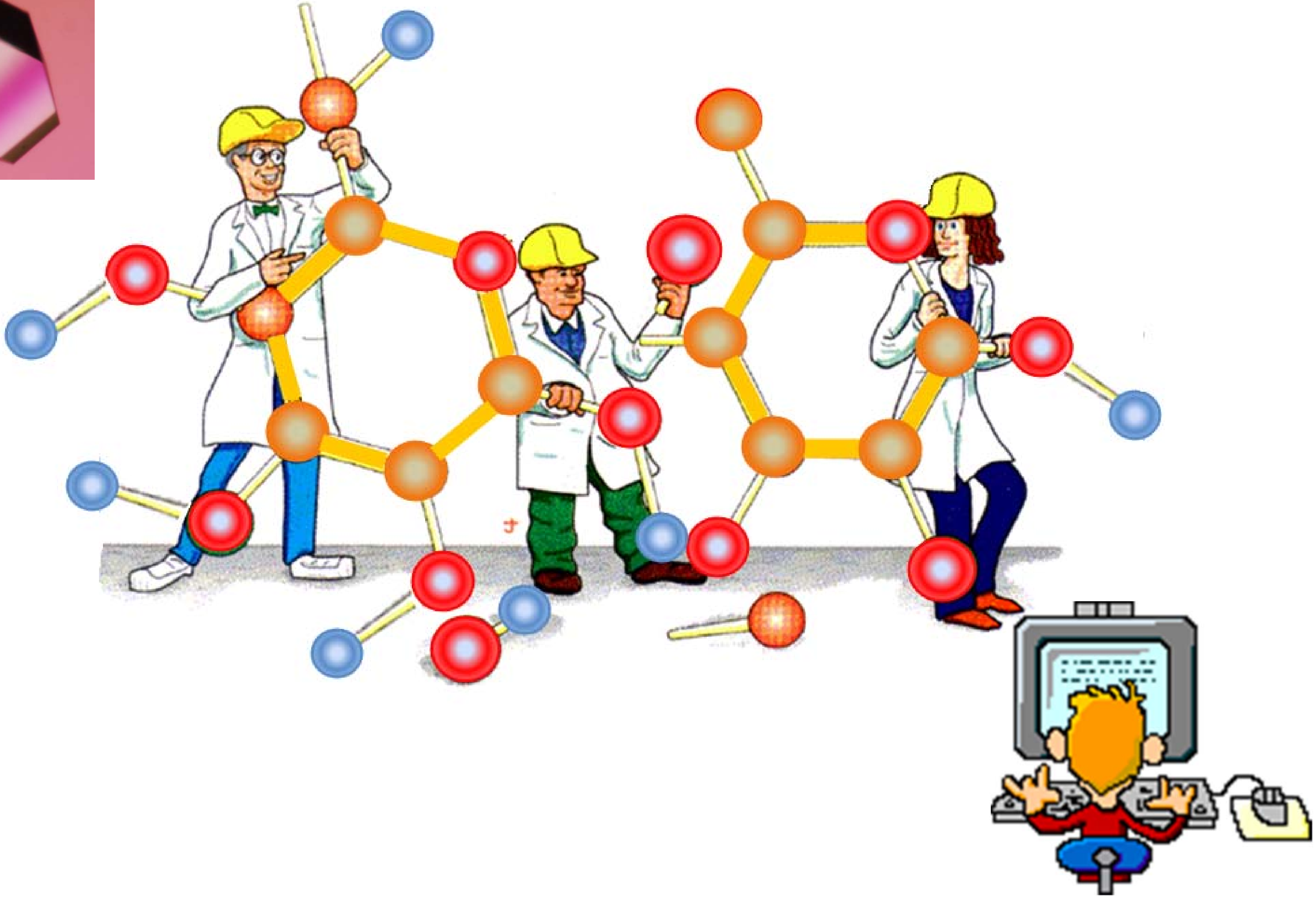
- Careful and time consuming sample preparation are needed

Helical Structures of Polysaccharides

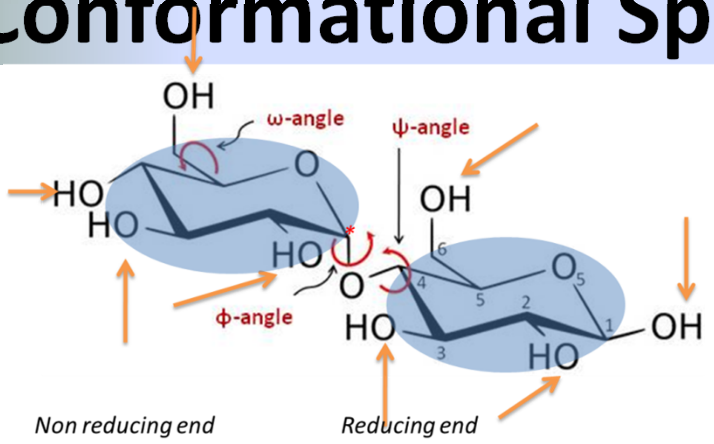


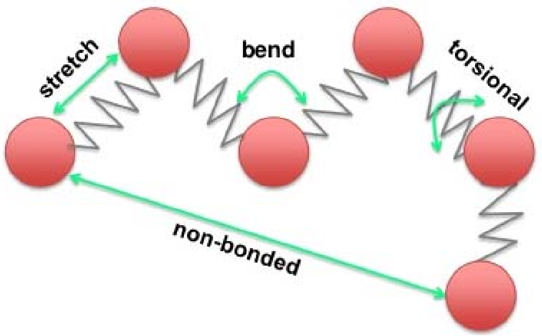
From *in cubo* to *in silico*



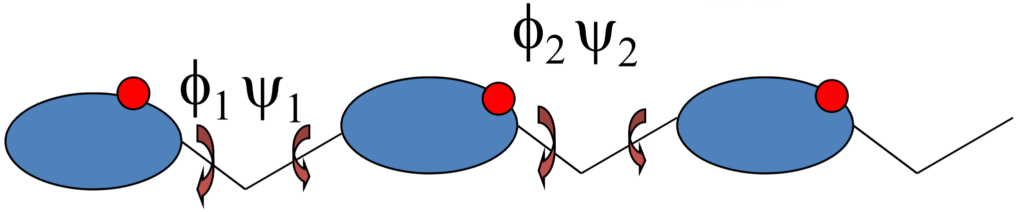


Conformational Space of Oligosaccharides





Combinatorial building



Assumption:

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

Methods :

Combine the lowest energy minima of each disaccharide map

Not true for

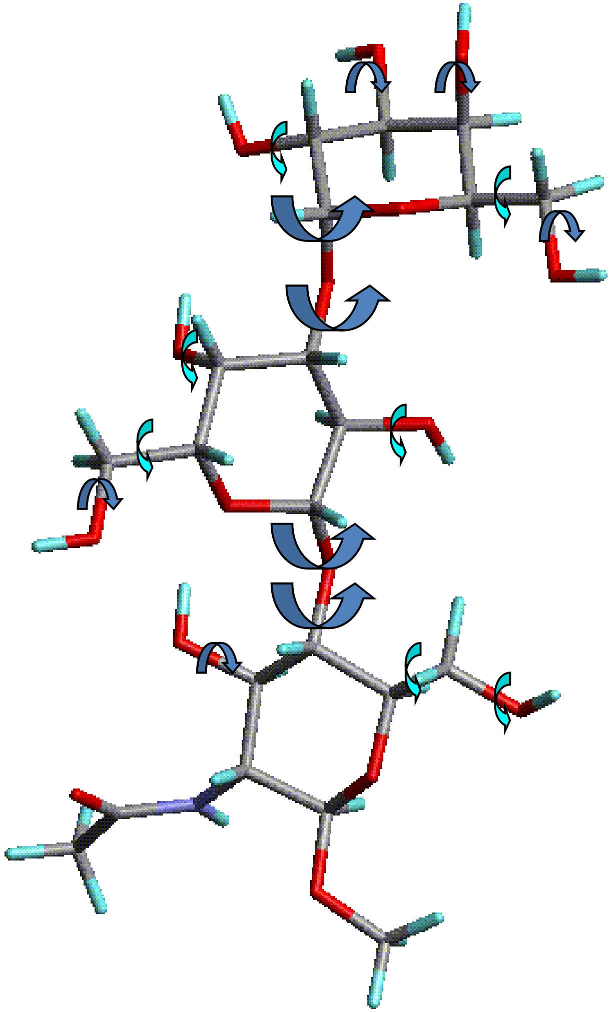
- **long range interactions**
- **branched structures**

….

But very useful for building starting structures!

Conformational Space of Flexible Oligosaccharides

Systematic search of all possible conformations ?



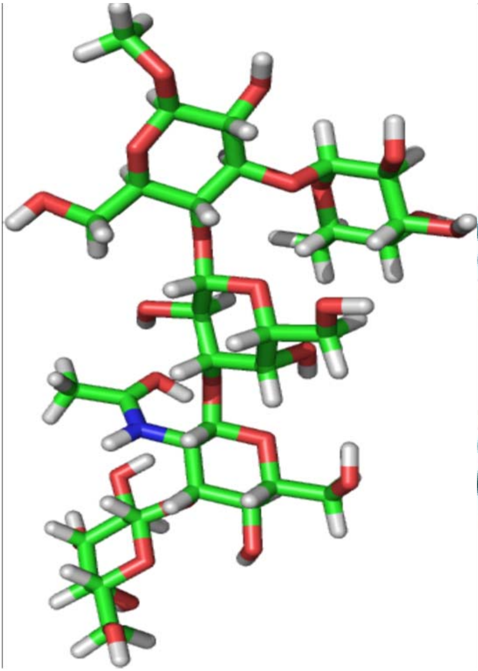
For a trisaccharide:

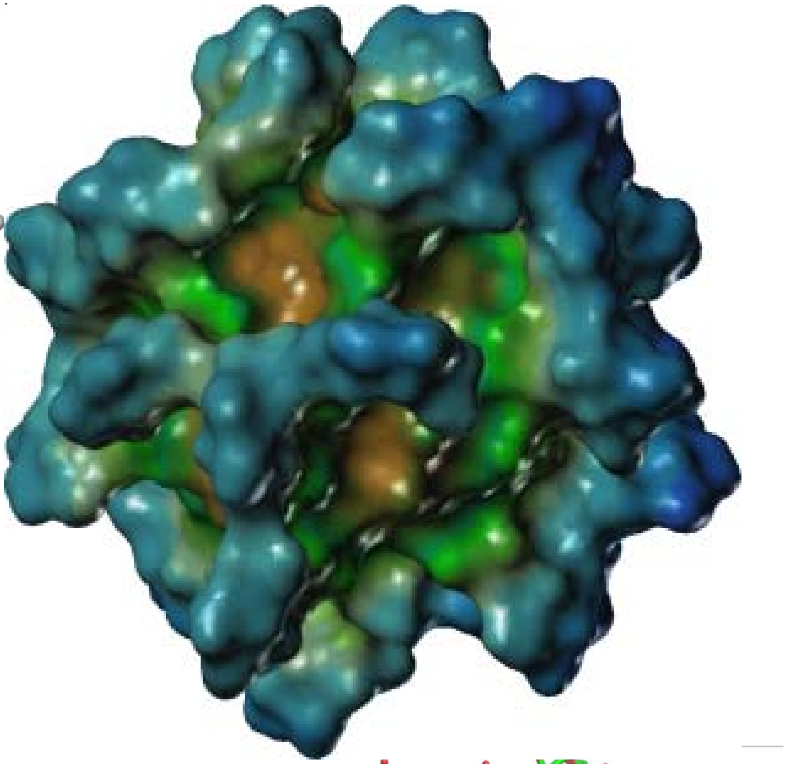
4 torsions to be searched
with 20° steps
18^4 conformations

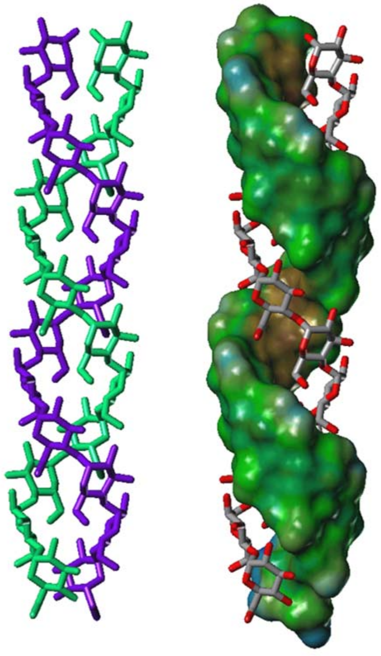
12 pendant groups
Staggered orientations:
3^{12} combinations

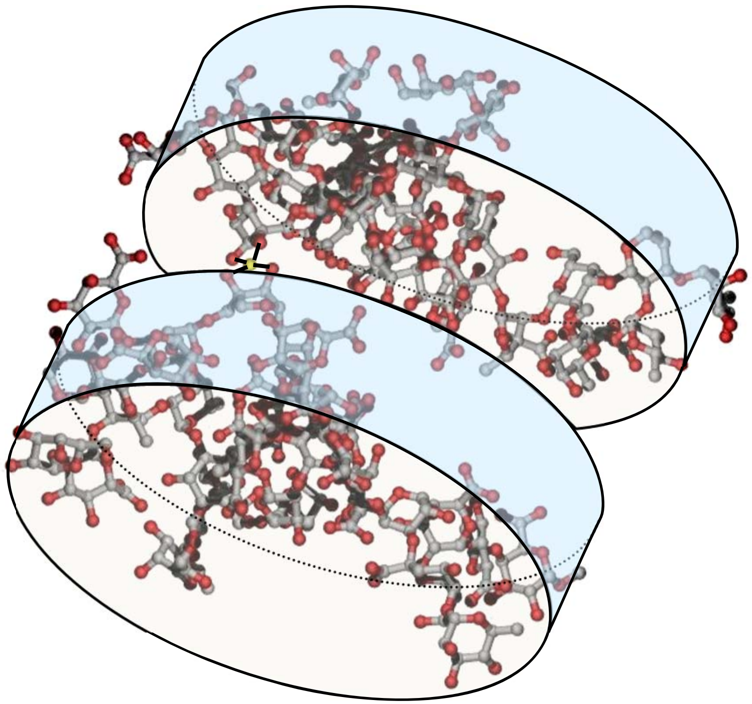
$> 5\ 10^{10}$

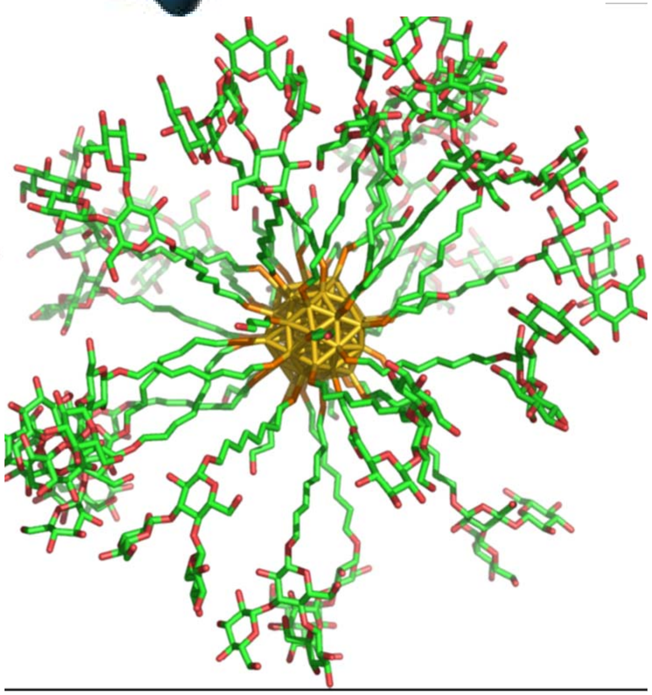
αGal(1-3)βGal(1-4)βGlcNAc

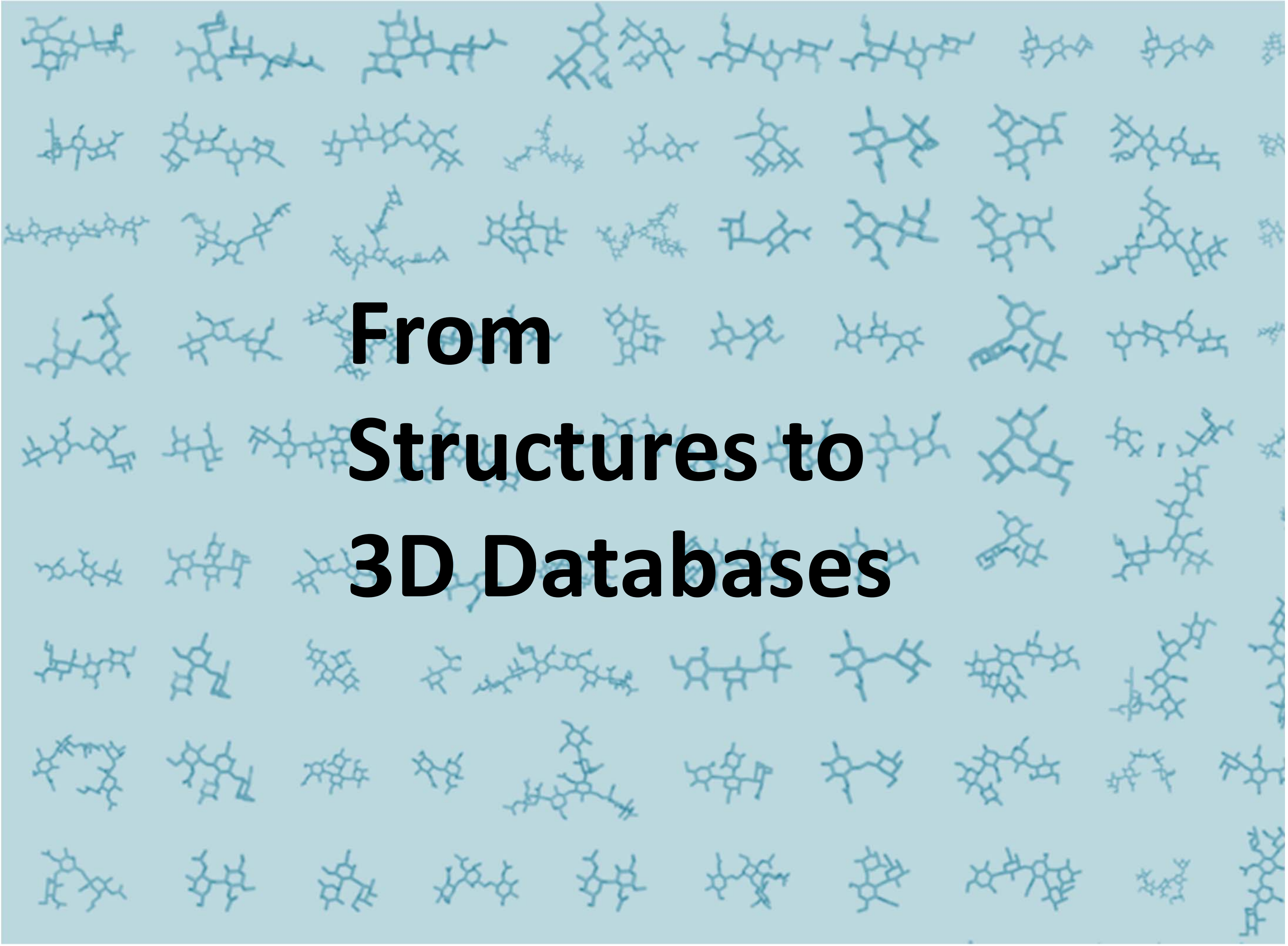






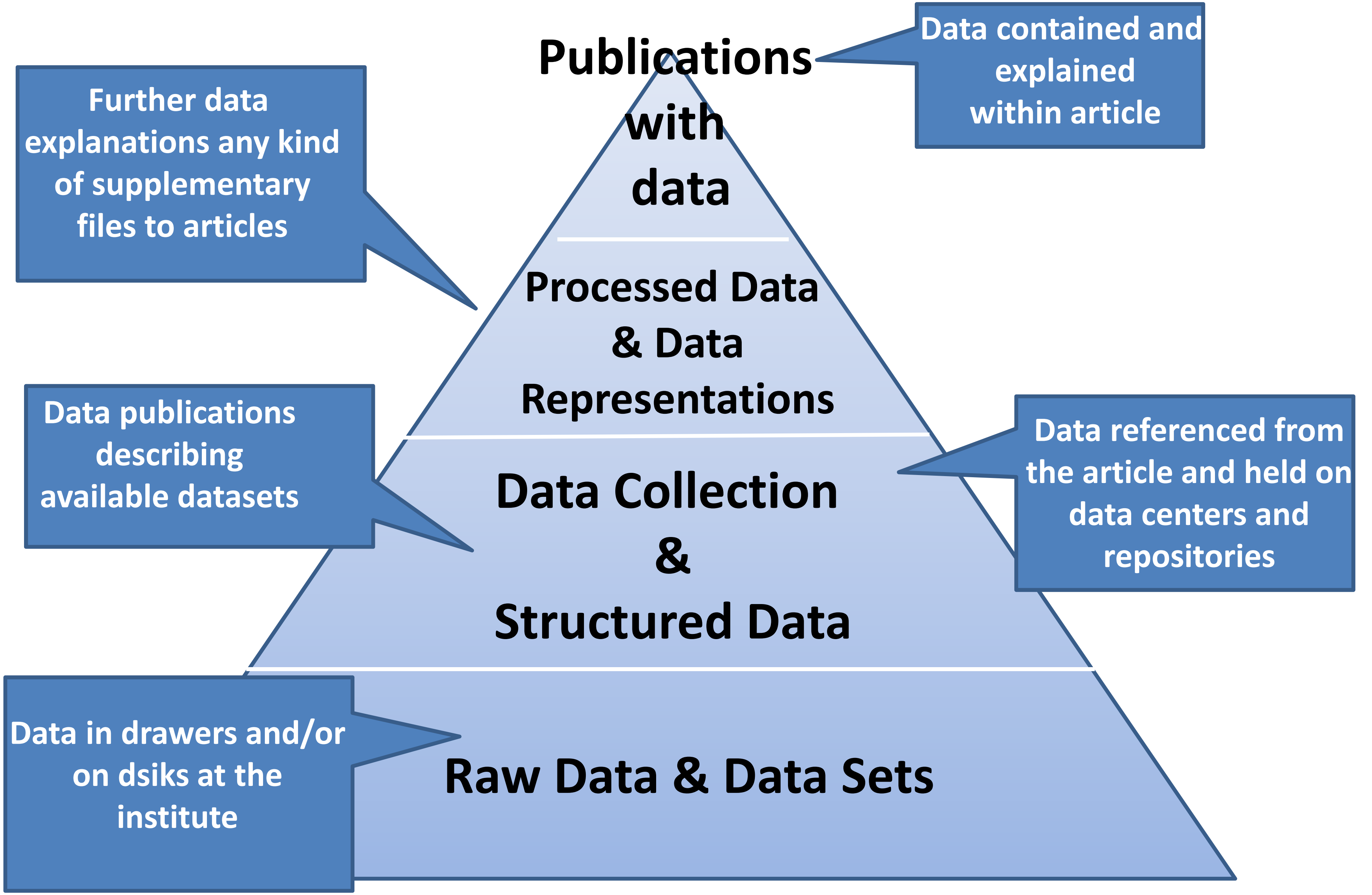






From
Structures to
3D Databases

An avalanche of data...



Publications with data

Data contained and explained within article

Further data explanations any kind of supplementary files to articles

Processed Data & Data Representations

Data publications describing available datasets

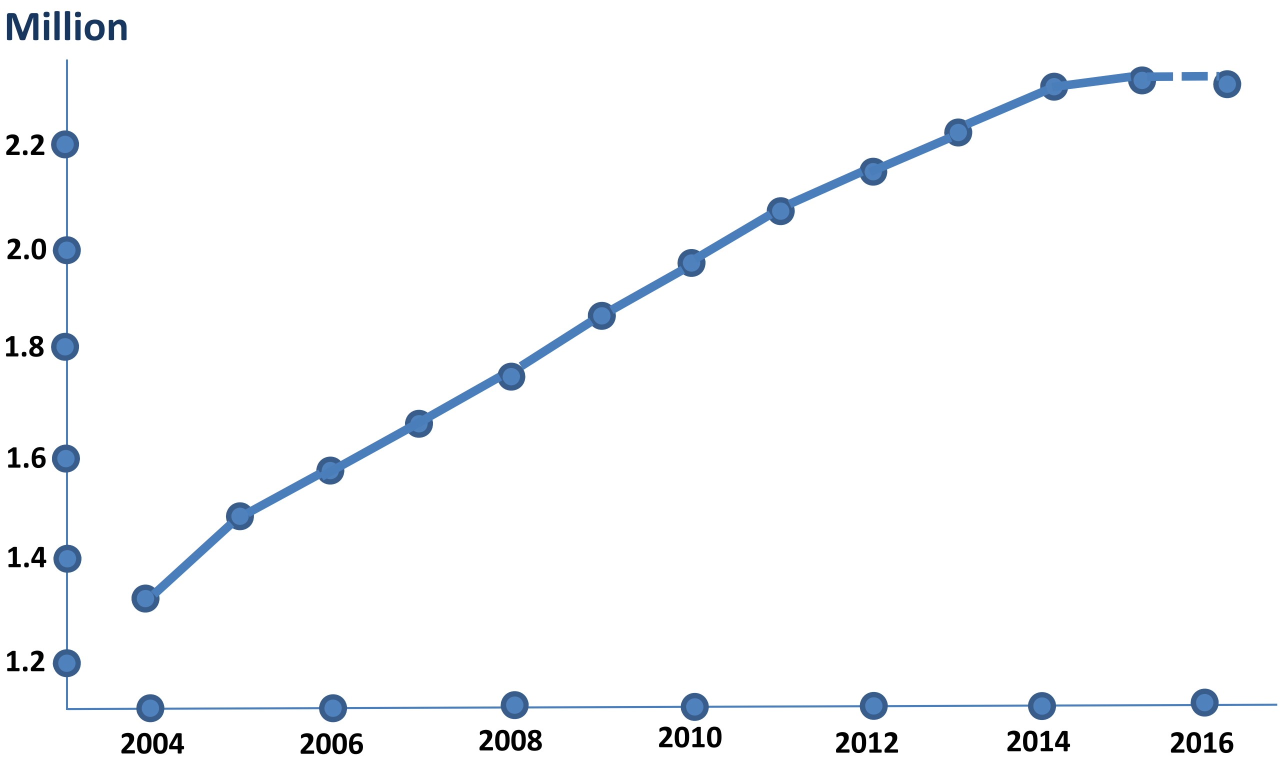
Data Collection & Structured Data

Data referenced from the article and held on data centers and repositories

Data in drawers and/or on dsiks at the institute

Raw Data & Data Sets

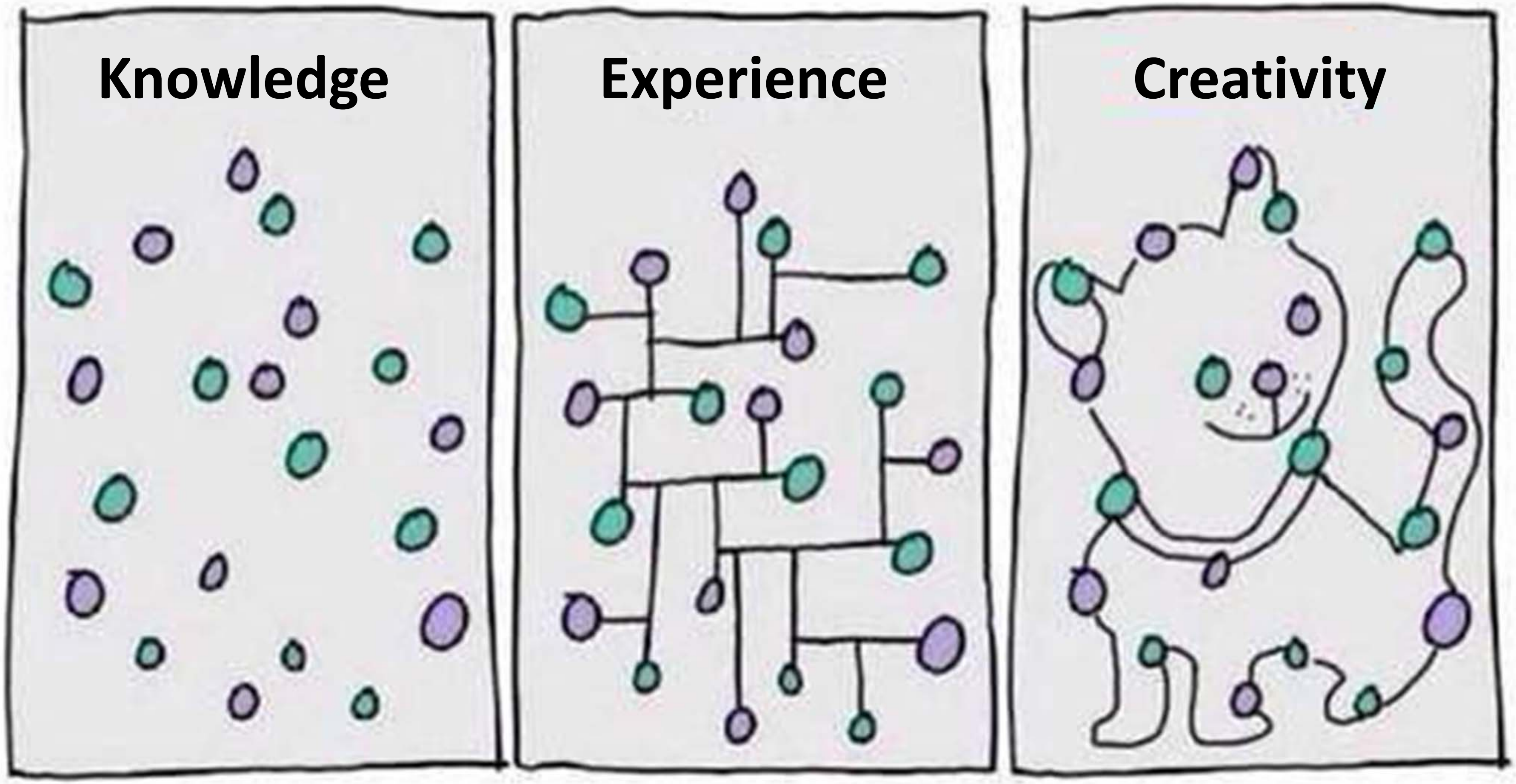
An avalanche of data…



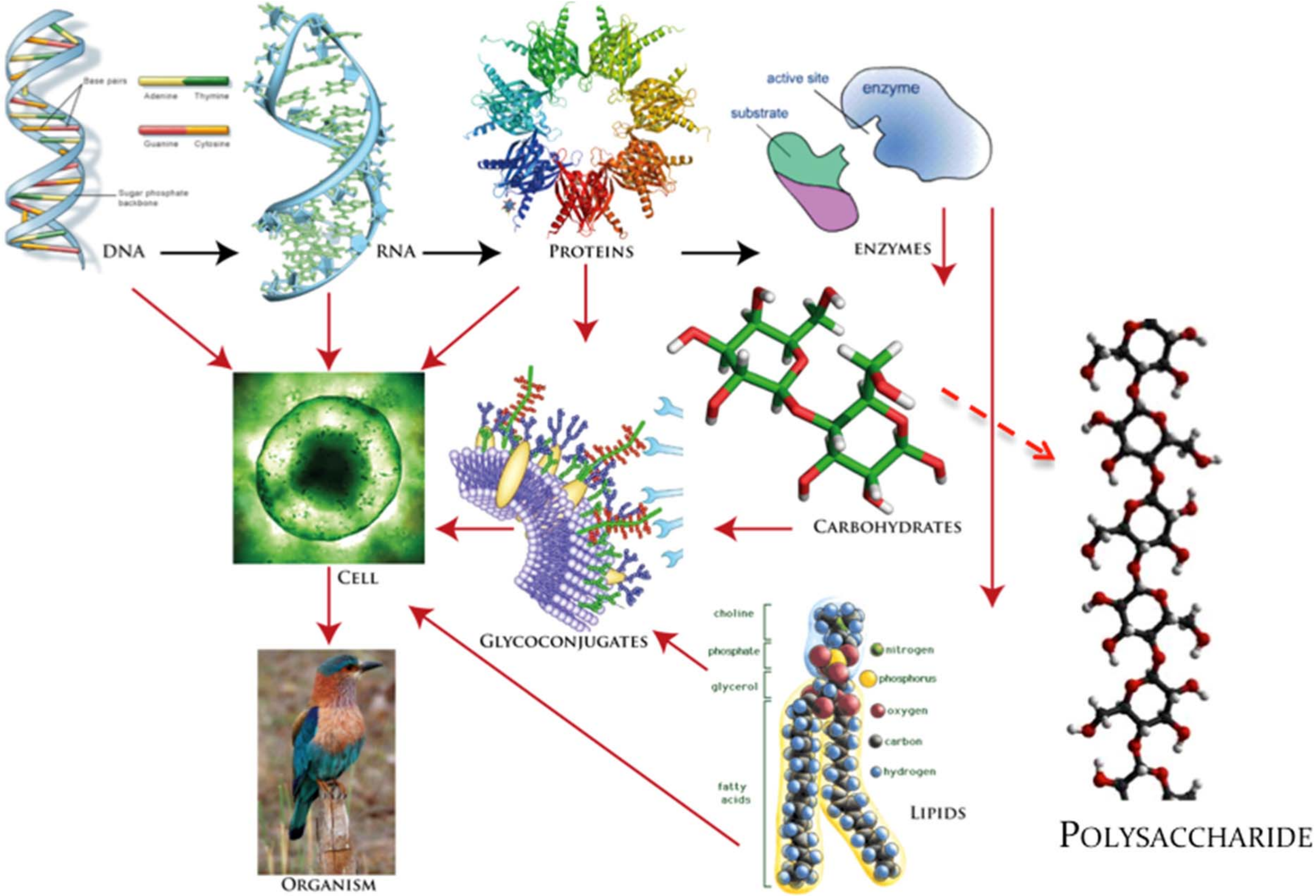
Global scientific output doubles every nine-years

Number of active researchers world-wide 8 Millions

Knowledge, Experience, Creativity

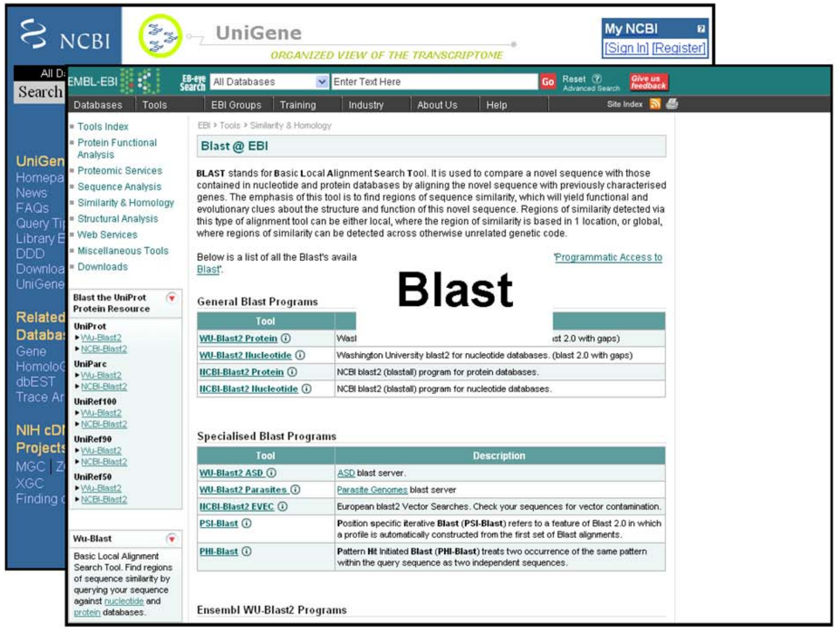


Carbohydrates in the Scheme of the Central Dogma of Life



Glycoinformatics

Genomics



Blast

Blast @ EBI

BLAST stands for **B**asic **L**ocal **A**lignment **S**earch **T**ool. It is used to compare a novel sequence with those contained in nucleotide and protein databases by aligning the novel sequence with previously characterised genes. The emphasis of this tool is to find regions of sequence similarity, which will yield functional and evolutionary clues about the structure and function of this novel sequence. Regions of similarity detected via this type of alignment tool can be either local, where the region of similarity is based in 1 location, or global, where regions of similarity can be detected across otherwise unrelated genetic code.

Below is a list of all the Blast's availa

'Programmatic Access to Blast'.

General Blast Programs

Tool	
WU-Blast2 Protein	Wasl ıst 2.0 with gaps)
WU-Blast2 Nucleotide	Washington University blast2 for nucleotide databases. (blast 2.0 with gaps)
NCBI-Blast2 Protein	NCBI blast2 (blastall) program for protein databases.
NCBI-Blast2 Nucleotide	NCBI blast2 (blastall) program for nucleotide databases.

Specialised Blast Programs

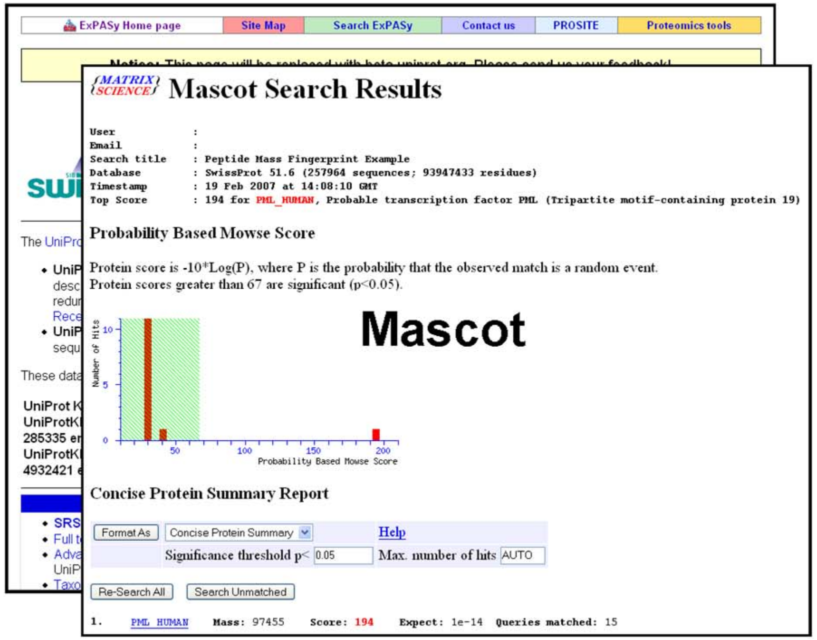
Tool	Description
WU-Blast2 ASD	ASD blast server.
WU-Blast2 Parasites	Parasite Genomes blast server
NCBI-Blast2 EVEC	European blast2 Vector Searches. Check your sequences for vector contamination.
PSI-Blast	**P**osition **s**pecific **i**terative **Blast** (**PSI-Blast**) refers to a feature of Blast 2.0 in which a profile is automatically constructed from the first set of Blast alignments.
PHI-Blast	**P**attern **H**it **I**nitiated **Blast** (**PHI-Blast**) treats two occurrence of the same pattern within the query sequence as two independent sequences.

Ensembl WU-Blast2 Programs

Wu-Blast

Basic Local Alignment Search Tool. Find regions of sequence similarity by querying your sequence against nucleotide and protein databases.

Proteomics



Mascot Search Results

Mascot

```
User          :
Email         :
Search title  : Peptide Mass Fingerprint Example
Database      : SwissProt 51.6 (257964 sequences; 93947433 residues)
Timestamp     : 19 Feb 2007 at 14:08:10 GMT
Top Score     : 194 for PML_HUMAN, Probable transcription factor PML (Tripartite motif-containing protein 19)
```

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 67 are significant ($p<0.05$).

Concise Protein Summary Report

Significance threshold p< 0.05 Max. number of hits AUTO

```
1.   PML_HUMAN   Mass: 97455   Score: 194   Expect: 1e-14  Queries matched: 15
```

Glycomics

Challenges for Glycoinformatics

Structures as Primary Access Key

Bioinformatic

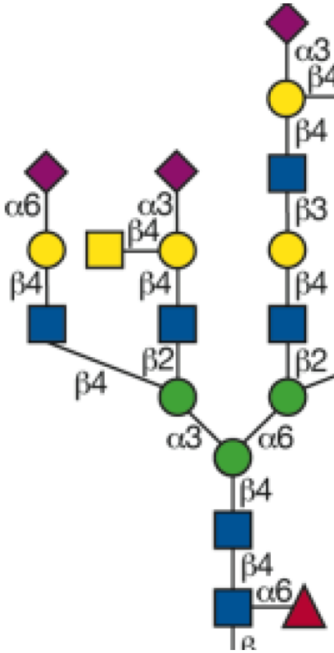
Galectin-1
source organism="Homo sapiens"
gene gene="LGALS1"
Site /site_type="binding"
/note="Beta-galactoside (Potential)."

Sequences of residues

1 MACGLVASNL NLKPGECLRV RGEVAPDAKS
31 FVLNLGKDSN NLCLHFNPRF NAHGDANTIV
61 CNSKDGGAWG TEQREAVFPF QPGSVAEVCI
91 TFDQANLTVK LPDGYEFKFP NRLNLEAINY
121 MAADGDFKIK CVAFD

Glycoinformatic

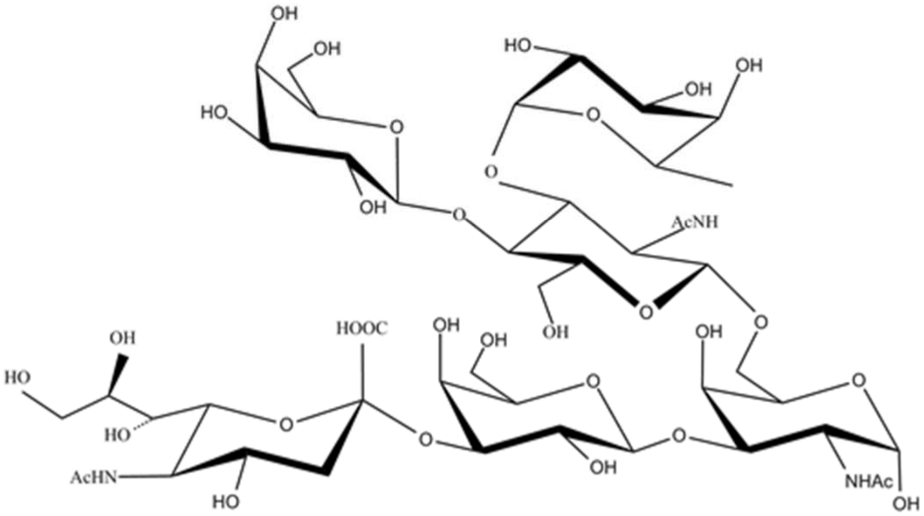
Topology of Residues

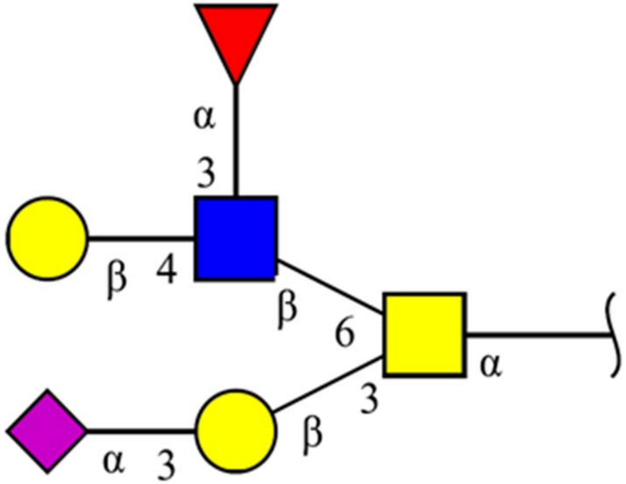


Encoding of Glycan Structures

Lewis X and Sialyl Acid on Core 2

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



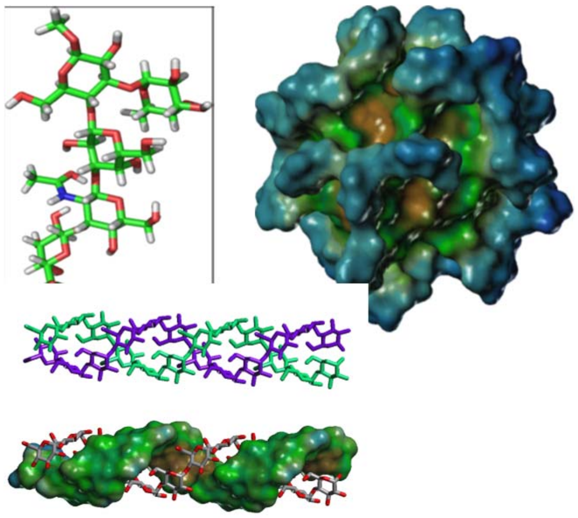


```
RES
1b:a-dgal-HEX-1:5
2s:n-acetyl
3b:b-dgal-HEX-1:5
4b:a-dgro-dgal-NON-2:6|1:a|2:keto|3:d
5s:n-acetyl
6b:b-dglc-HEX-1:5
7s:n-acetyl
8b:a-lgal-HEX-1:5|6:d
9b:b-dgal-HEX-1:5
LIN
1:1d(2+1)2n
2:1o(3+3)3d
3:3o(3+2)4d
4:4d(5+1)5n
5:1o(6+1)6d
6:6d(2+1)7n
7:6o(3+1)8d
8:6o(4+1)9d
```

GlycoCT

e-Glycoscience

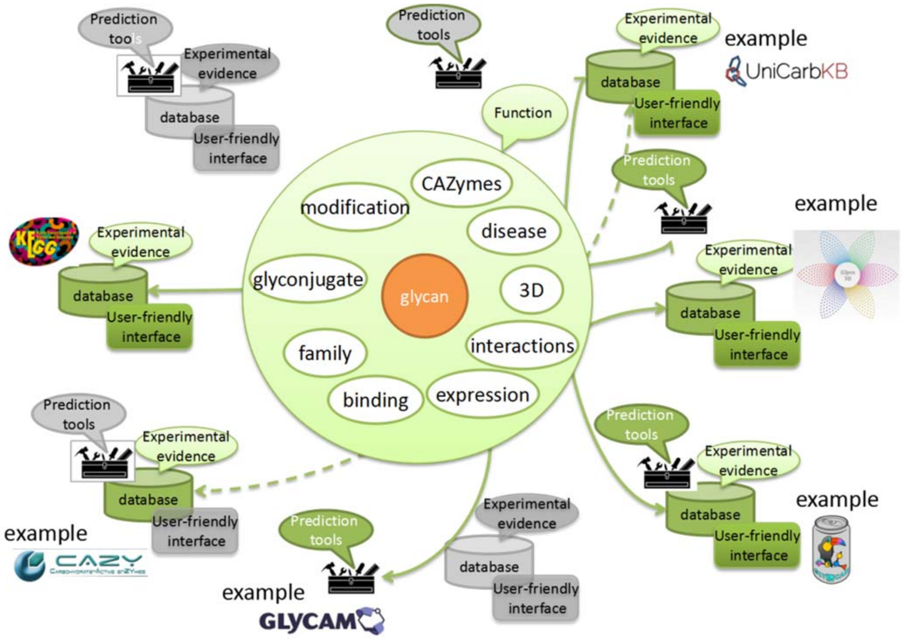
Continued advances in molecular modeling has generated insights for understanding glycan structures and properties. Robust, validated informatics tools are developed in to enable accurate and fast determination of complex **carbohydrate and glycoconjugate structural prediction, computational modeling,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

Protocole: Search on Internet Tools and DataBases in Glycosciences

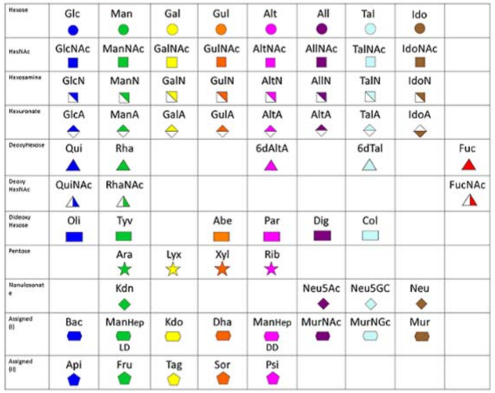
Sites should be up to date (checked) and freely available.

Results: 94 Tools 32 DataBases

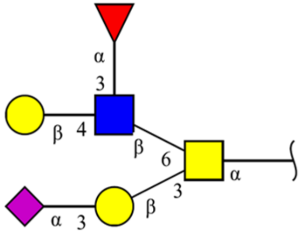
Analysis and Clustering

Representations

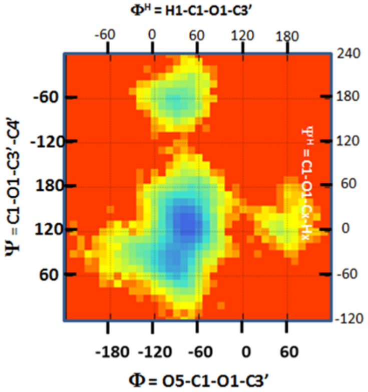
Symbol Nomenclature for Glycans: SNFG



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 ⬟				

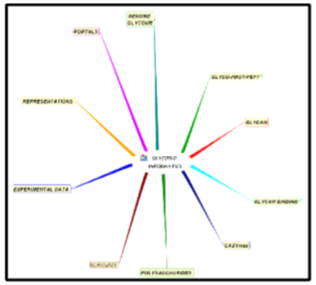


```
RES
1b:a-dgal-HEX-1:5
2s:n-acetyl
3b:b-dgal-HEX-1:5
4b:a-dgro-dgal-NON-
2:6|1:a|2:keto|3:d
5s:n-acetyl
6b:b-dglc-HEX-1:5
7s:n-acetyl
8b:a-lgal-HEX-1:5|6:d
9b:b-dgal-HEX-1:5
LIN
1:1d(2+1)2n
2:1o(3+3)3d
3:3o(3+2)4d
4:4d(5+1)5n
5:1o(6+1)6d
6:6d(2+1)7n
7:6o(3+1)8d
8:6o(4+1)9d
```

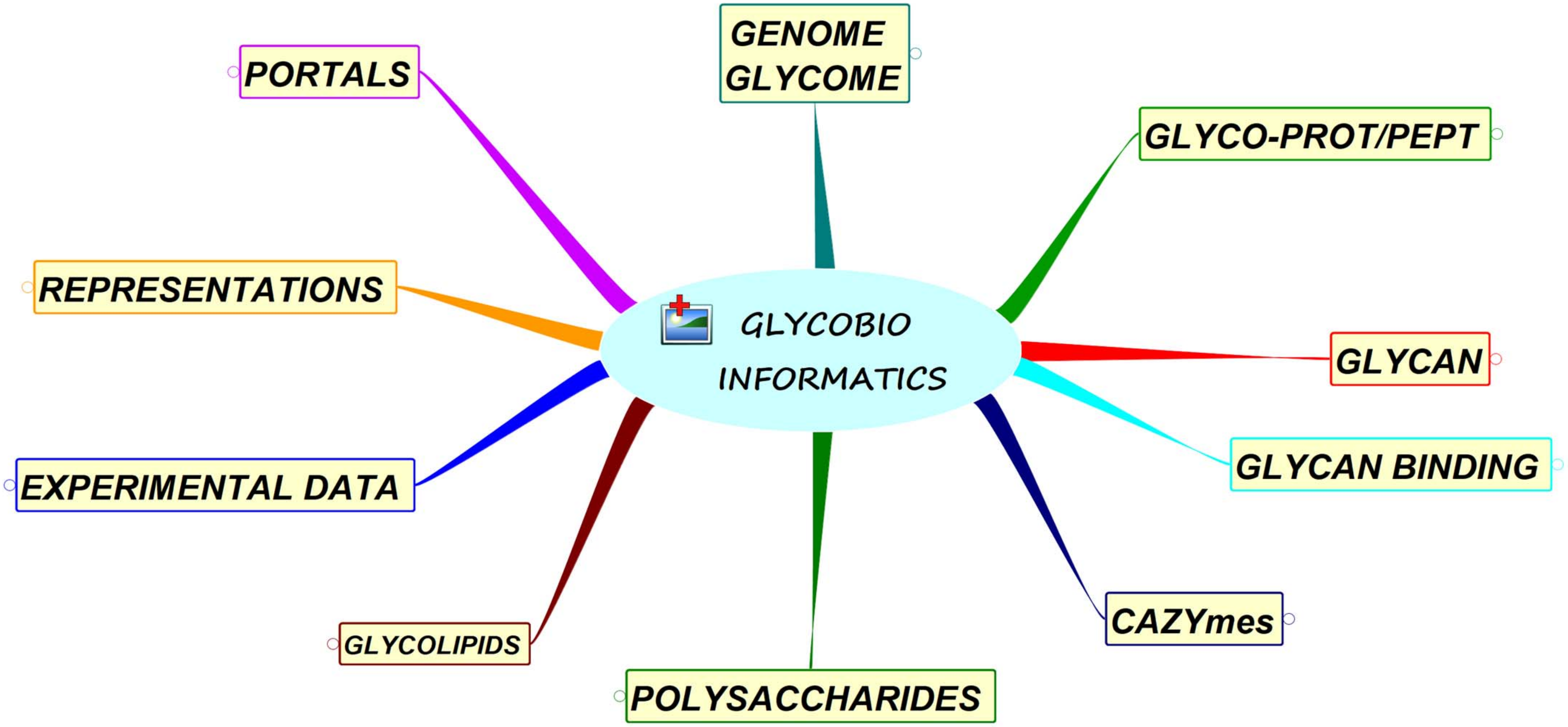


REPRESENTATIONS

- **Ontology**
  - GlycoRDF
- **Nomenclature**
  - LiGaph
  - GlycoCT
  - LINUCS
  - Glyde II
  - KCF
  - IUPAC
  - SNFG
- **Composition**
  - Glycan Mass
  - SwissMassAbaccus
- **Drawing**
  - Glycan Builder
  - Draw Glycan SNFG
  - DrawRINGS
- **3D-Analysers**
  - Glytorsion
  - Carp
  - Azahar
  - pdb-care
  - pdb2linux
- **3D-Builders**
  - doGlycans
  - Glycal-Web
  - Sweet
  - PolysGlycanBuilder
  - Others
- **3D-VIewers**
  - LiteMol
  - GlycamWeb
  - Sweet Unity Mol
  - Others...

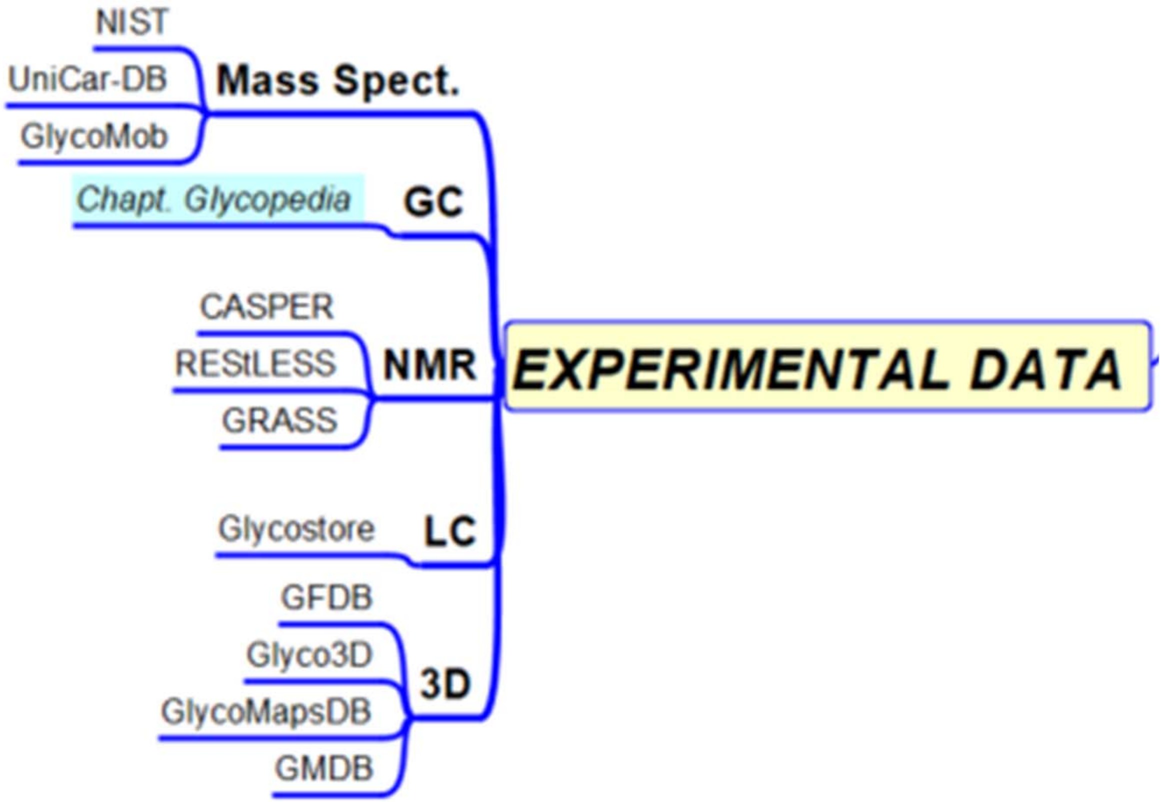


Tools and DataBases



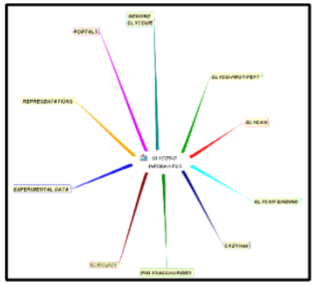
mentalmap

A Traveller-Guide in Glycoscience Cyber-Space

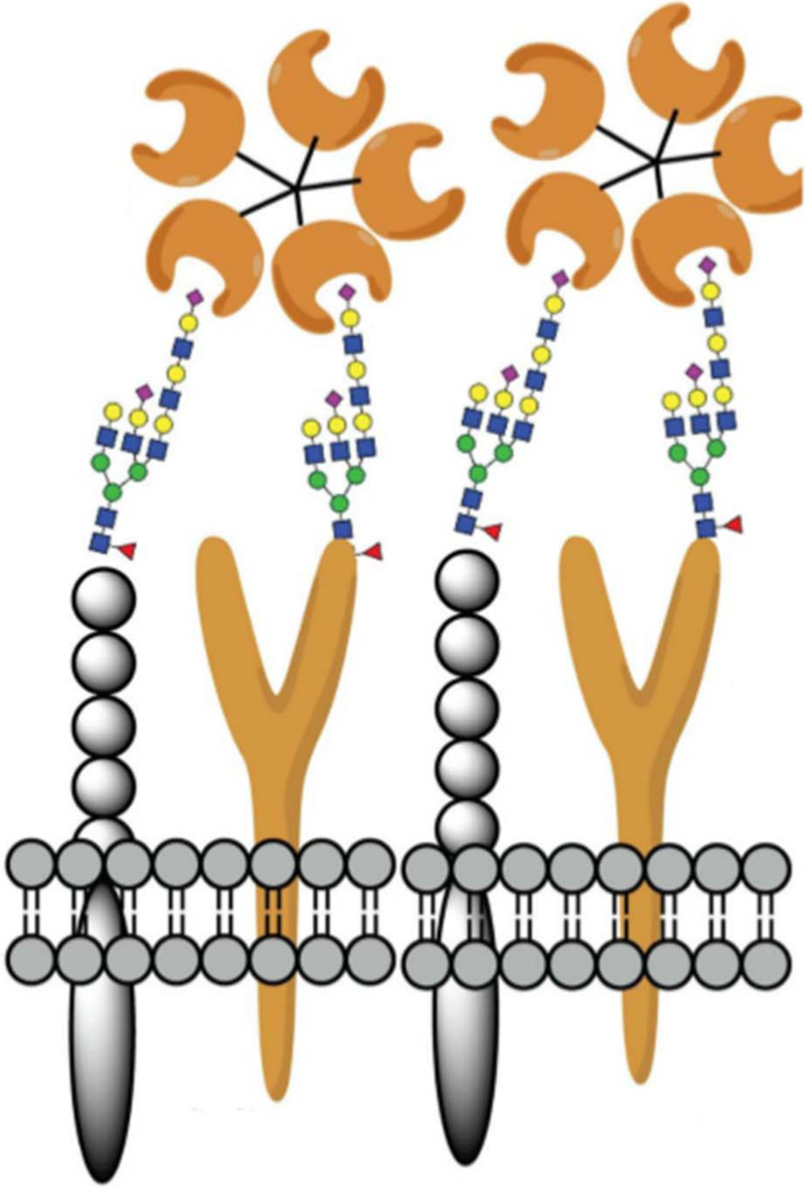


EXPERIMENTAL DATA

- **Mass Spect.**
  - NIST
  - UniCar-DB
  - GlycoMob
- **GC**
  - *Chapt. Glycopedia*
- **NMR**
  - CASPER
  - REStLESS
  - GRASS
- **LC**
  - Glycostore
- **3D**
  - GFDB
  - Glyco3D
  - GlycoMapsDB
  - GMDB

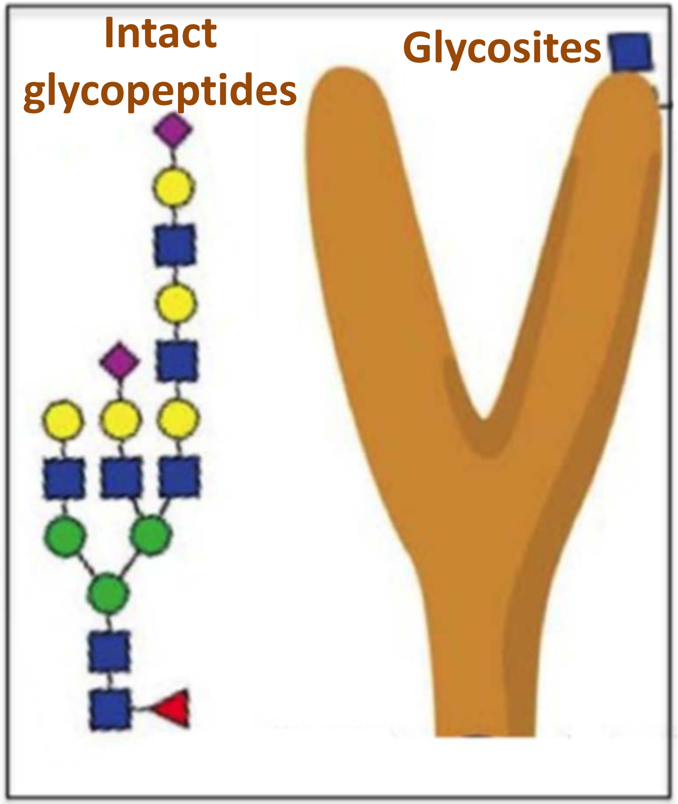


Glycomics

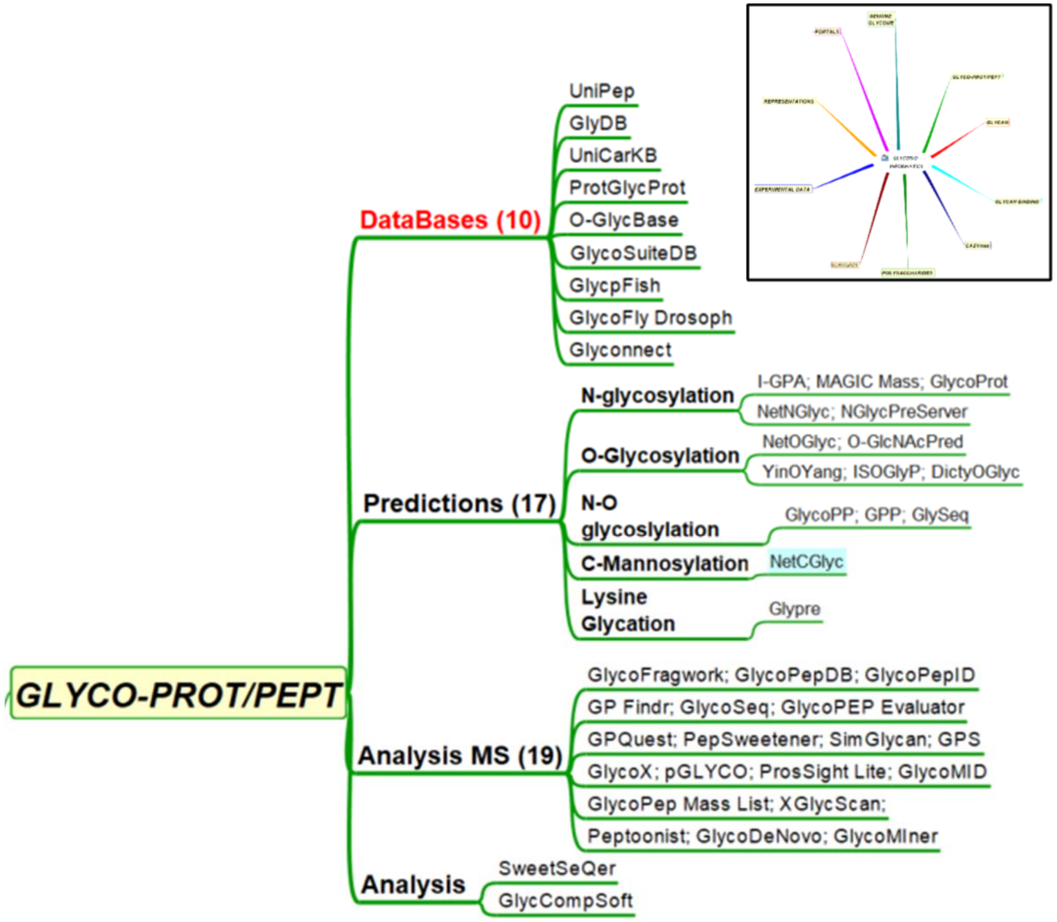


Glycoproteomics

Glycoproteomics

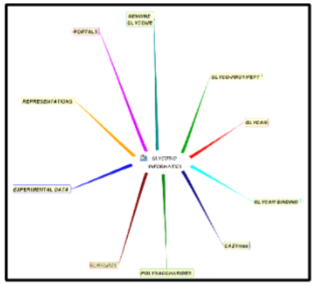


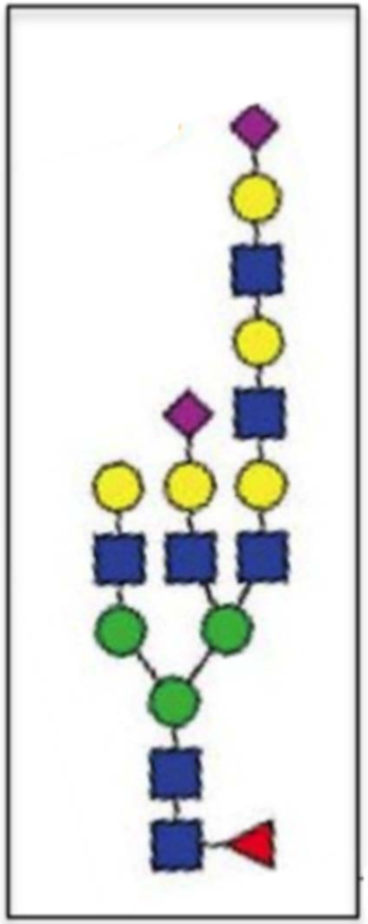
Glycan profiles

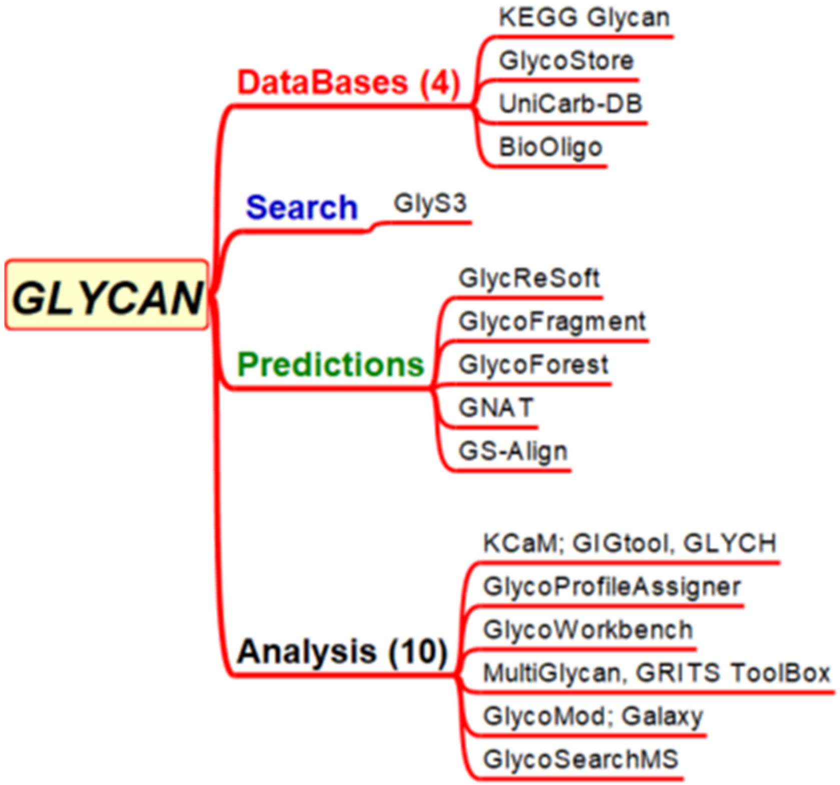


GLYCO-PROT/PEPT
- **DataBases (10)**
  - UniPep
  - GlyDB
  - UniCarKB
  - ProtGlycProt
  - O-GlycBase
  - GlycoSuiteDB
  - GlycpFish
  - GlycoFly Drosoph
  - Glyconnect
- **Predictions (17)**
  - **N-glycosylation**
    - I-GPA; MAGIC Mass; GlycoProt
    - NetNGlyc; NGlycPreServer
  - **O-Glycosylation**
    - NetOGlyc; O-GlcNAcPred
    - YinOYang; ISOGlyP; DictyOGlyc
  - **N-O glycoslylation**
    - GlycoPP; GPP; GlySeq
  - **C-Mannosylation**
    - NetCGlyc
  - **Lysine Glycation**
    - Glypre
- **Analysis MS (19)**
  - GlycoFragwork; GlycoPepDB; GlycoPepID
  - GP Findr; GlycoSeq; GlycoPEP Evaluator
  - GPQuest; PepSweetener; SimGlycan; GPS
  - GlycoX; pGLYCO; ProsSight Lite; GlycoMID
  - GlycoPep Mass List; XGlycScan;
  - Peptoonist; GlycoDeNovo; GlycoMIner
- **Analysis**
  - SweetSeQer
  - GlycCompSoft

Glycans



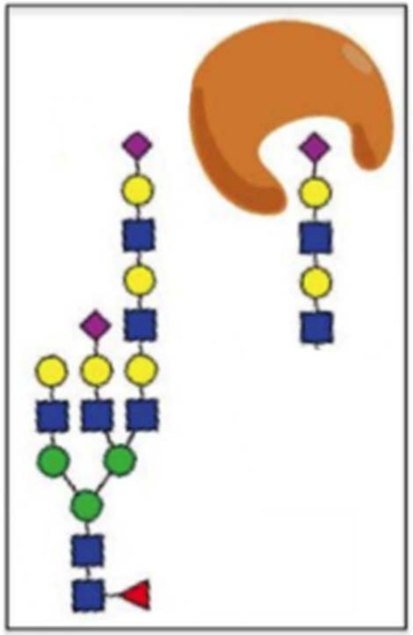


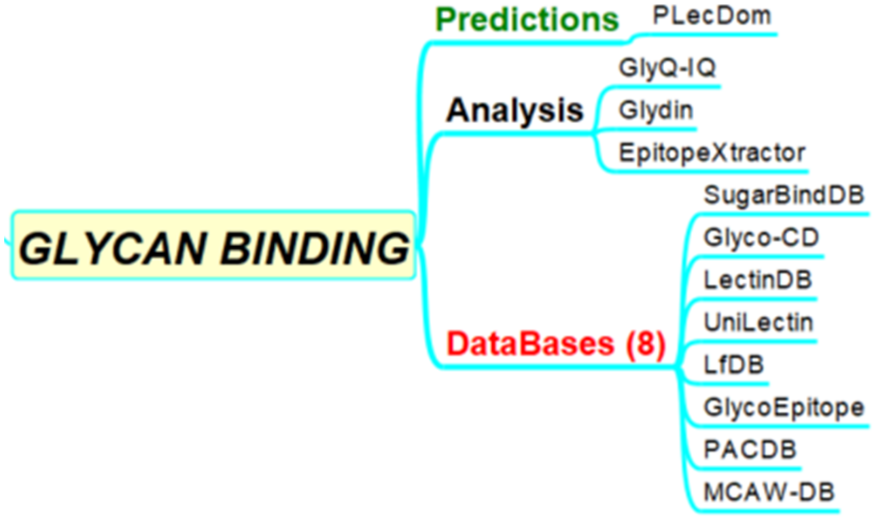


GLYCAN

- **DataBases (4)**
  - KEGG Glycan
  - GlycoStore
  - UniCarb-DB
  - BioOligo
- **Search**
  - GlyS3
- **Predictions**
  - GlycReSoft
  - GlycoFragment
  - GlycoForest
  - GNAT
  - GS-Align
- **Analysis (10)**
  - KCaM; GlGtool, GLYCH
  - GlycoProfileAssigner
  - GlycoWorkbench
  - MultiGlycan, GRITS ToolBox
  - GlycoMod; Galaxy
  - GlycoSearchMS

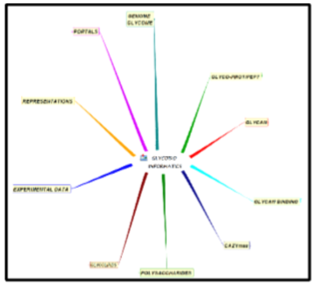
Functional Glycomics





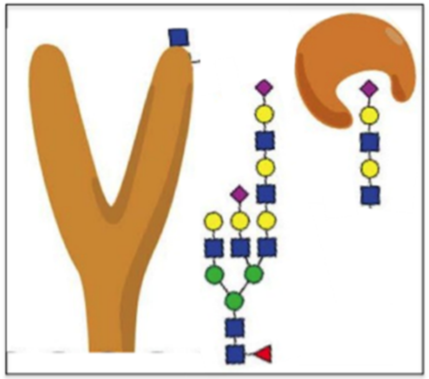
GLYCAN BINDING

- **Predictions**
  - PLecDom
- **Analysis**
  - GlyQ-IQ
  - Glydin
  - EpitopeXtractor
- **DataBases (8)**
  - SugarBindDB
  - Glyco-CD
  - LectinDB
  - UniLectin
  - LfDB
  - GlycoEpitope
  - PACDB
  - MCAW-DB



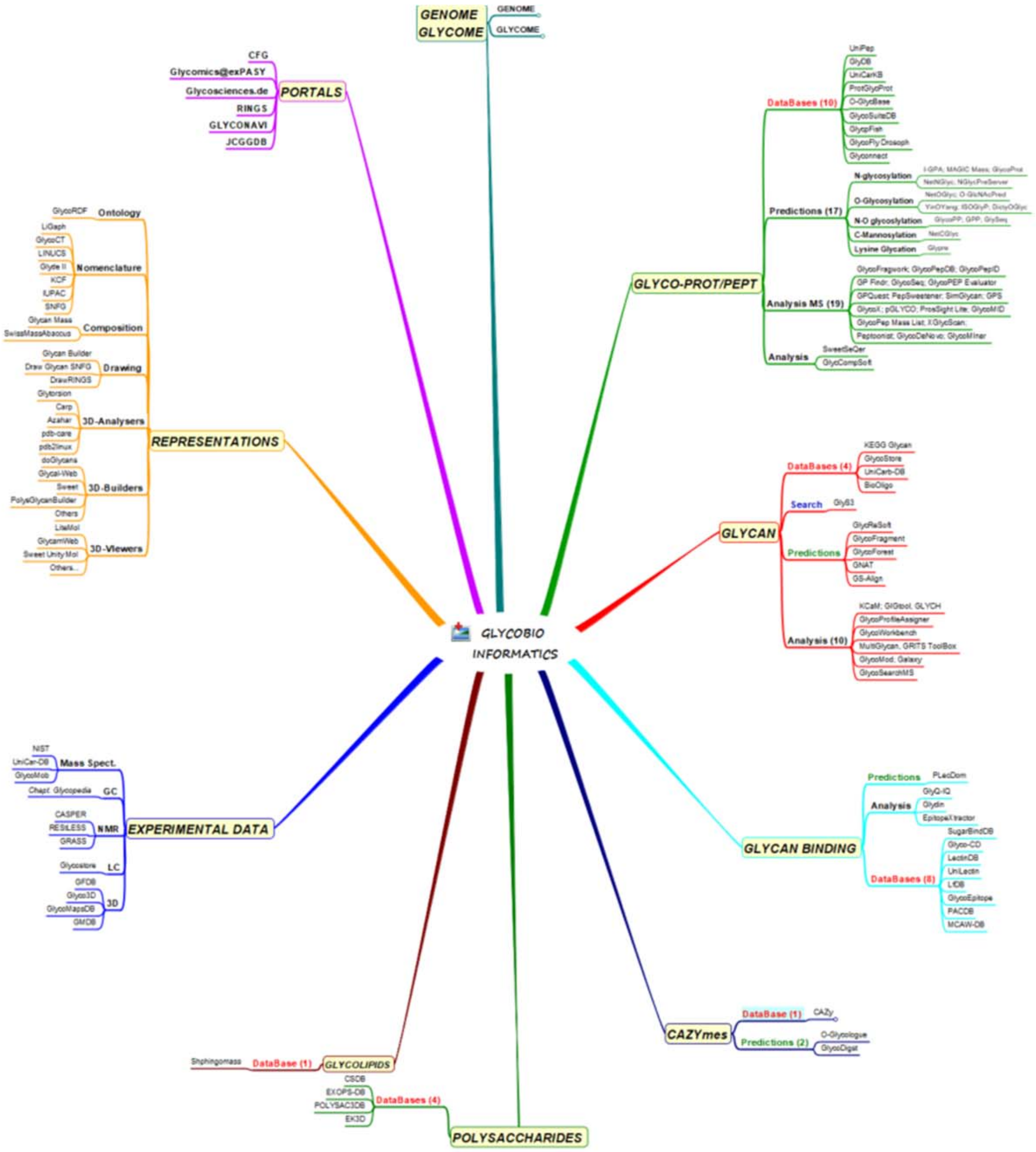
Glycoproteomics

Functional Glycomics



Glyconnect

A Traveller-Guide in Glycoscience Cyber-Space



available at : https://glycopedia.eu end July 2018


CHAMBERY
ALBERTVILLE
GENEVE
TURIN