From Monosaccharides to Polysaccharides From Structures to 3D Databases

Serge Pérez, Grenoble, June 2016
From Monosaccharides to Polysaccharides
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 nuclei.

**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

Unsubstituted disaccharides  60 structures
Unsubstituted trisaccharides  30 structures
Unsubstituted 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 Intermolecular hydrogen bonds

Analysis of
Hydration features
Packing features

Emil Fisher  Johannes Bijvoet
Molecular & Crystal Structures of Carbohydrates

Hydrogen Bonding in Crystalline Oligosaccharides

Analysis of high accurate X-ray analysis – Neutron diffraction

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

CHO bonds

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

Molecular & Crystal Structures of Carbohydrates

Packing Features
Molecular & Crystal Structures of Carbohydrates

Powder Diffraction

1. Identification of Crystalline Polymorphs

2. Solving Crystal Structures – Rietvelt Method + Molecular Modelling

- Synthetic Pentasaccharide ID31@ESRF, $\lambda = 0.8 \, \text{Å}$
- Monoclinic $P2_1$
  - $a=15.54$, $b=8.83$, $c=17.67$, $\beta=94.6$


Crystalline Conformations of Oligosaccharides in Proteins

Experimental Conditions and Limitations

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

- Micro-crystal in 2010
- Micro-crystal in 1998

Unit Cell motif

10 μm
Yearly Growth of Total Protein Structures (Deposit / Total)
Crystalline Conformations of Oligosaccharides in Proteins

Protein- Carbohydrate Interactions

- Synthesis
- Modifications
- Degradation
- Binding
- Transport

Crystalline Conformations of Oligosaccharides in Proteins

Oligosaccharides –Lectin Complexes

Protein Data Bank: http://www.rcsb.org/pdb/home/home.do
Crystalline Conformations of Oligosaccharides in Proteins

Crystalline Conformations of Polysaccharides

X-Ray Fiber Diffraction of Polysaccharides

\[
\text{hp} = n \times 1.05 \text{ nm}
\]
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

Elephants 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.
**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° conformations

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

> 5 \times 10^{10}

αGal(1-3)βGal(1-4)βGlcNAc
Conformational Space of Flexible Oligosaccharides

The explorer....

Does not “see” the location of other minima.

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

Monte Carlo Calculations

Population

Replacement of the population by the successful individuals

Selection of individual as a function of fitness score.

Offspring


Best individuals

Parents
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

Residue Letter Name: Rib, Ara, Xyl, Lyx, All, Alt, Glc, Man, 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 4C1 chair conformation; those in the L-configuration are assumed to have 1C4 chair conformation. Otherwise, the ring conformation is indicated in the symbol, as αL 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
5s:n-acetyl
6b:b-dglc-HEX-1:5
7s:n-acetyl
8b:a-lgal-HEX-1:5|6:d
9b:b-dgaloHEX-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
Needs for Standardizations

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.

Glyco3D

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

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

Disaccharides

Source: Molecules or
Building 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
Bio-Oligosaccharides : 3D / NMR

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

- KAUSA: Blood group α antigen: pentasaccharide type 2
- Sequence: Fucα1-2 Galα1-3(Fucα1-3) Galα1-3 Galα1-4 Glc

- Woman representation

Molecular Weight: 553.79
Category: Blood group α antigen: Blood group A
Glycan sequence: Fuc1: Ser: 2 GlcNAc: 1 Glc: 1
Oligosaccharide Details
Reference: "Ketela, 2016"

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

- Glycan Product
- Total: 150 entries


Curations

Annotations
Lectins

**Source:** X-ray - PDB

**Classification of Lectins**
- Based on their origin: Algae, 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

**Molecule Information**
- Enzyme name
- Short name
- Origin
- Organism
- Resulting linkage
- Fold
- Cazy Family
- 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

<table>
<thead>
<tr>
<th>Molecule Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>Family</td>
</tr>
<tr>
<td>Species</td>
</tr>
<tr>
<td>View representation</td>
</tr>
<tr>
<td>PDB Code</td>
</tr>
<tr>
<td>Resolution</td>
</tr>
<tr>
<td>Sequence</td>
</tr>
<tr>
<td>Reference</td>
</tr>
<tr>
<td>Links (Medline, PDB, Swiss Prot.)</td>
</tr>
</tbody>
</table>

**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
We Can Build Your Glycans

Building Sugars

POLYS 2.0

**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**

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
Glycoproteins – Protein Carbohydrate Interactions

The Hidden Conformations of LewisX
From 1D to 3D Representations

\[ \text{D-Manp} \quad \text{D-Glcp} \quad \text{D-Galp} \quad \text{L-Fucp} \quad \text{D-Xylp} \quad \text{L-Rhap} \quad \text{D-Rhap} \quad \text{L-Araf} \]

\[ \text{Galp} \quad \text{Glc} \quad \text{Man} \]

\[ \text{Asn} \]