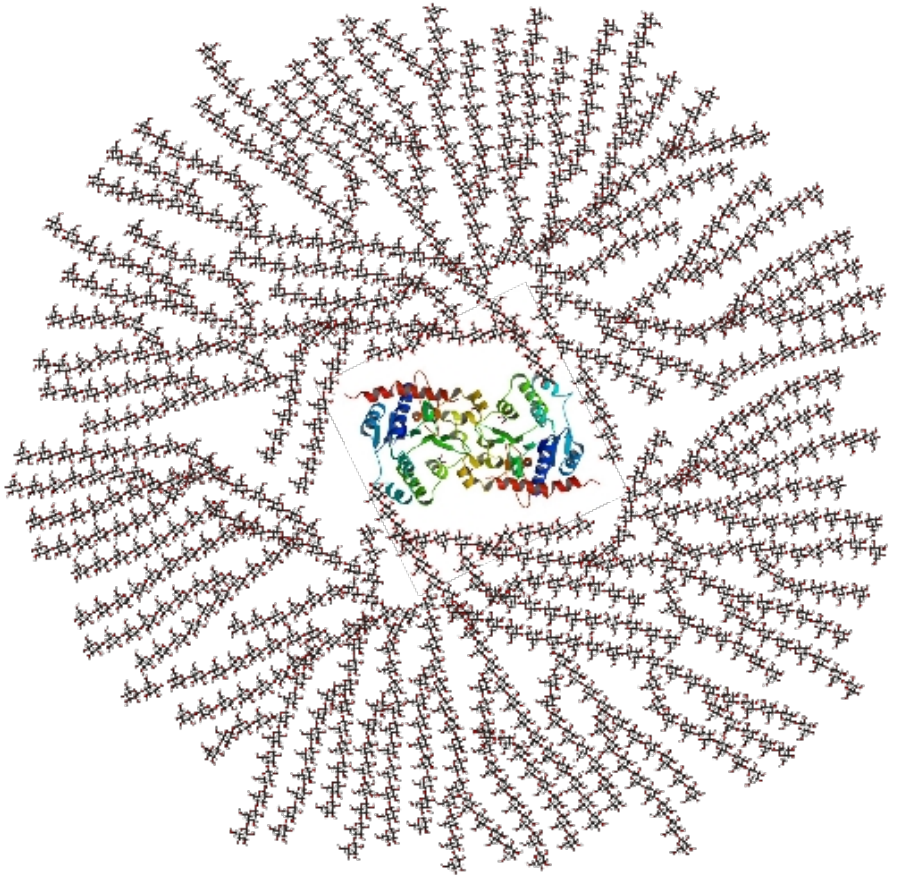
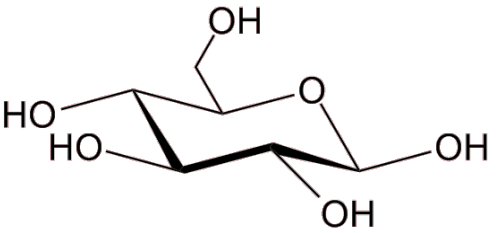
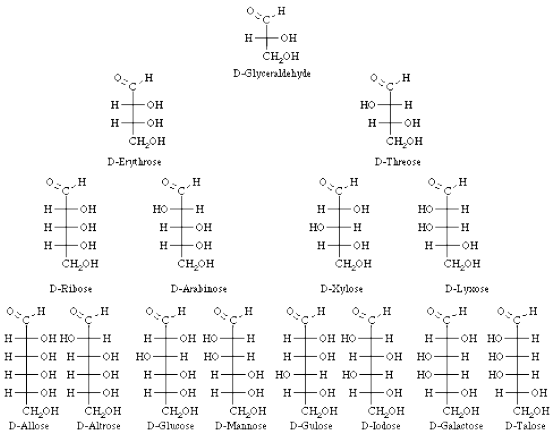
Wintersemester 2016/2017
Biomolecular Engineering/Nanobiophysics Module

LECTURE 5:

GLYCOBIOINFORMATICS/COMPUTATIONAL BIOLOGY OF SACCHARIDES







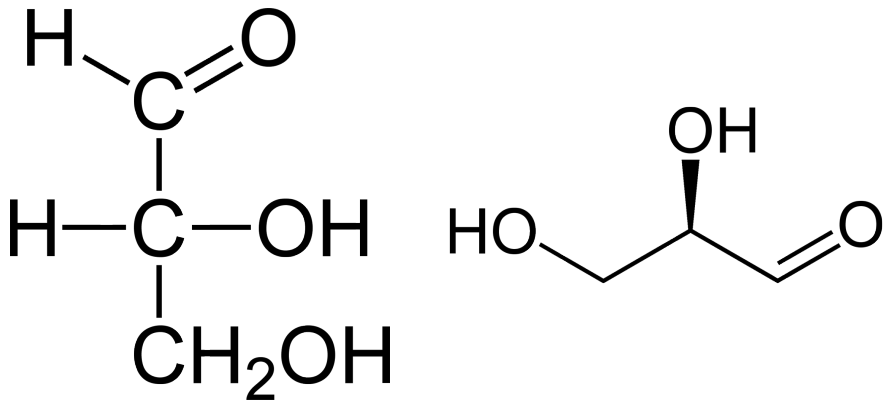
LECTURE 5:

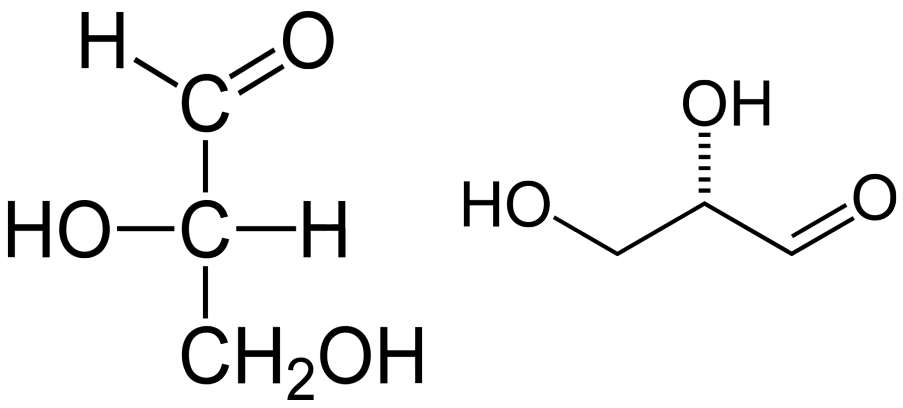
GLYCOBIOINFORMATICS/COMPUTATIONAL BIOLOGY OF SACCHARIDES

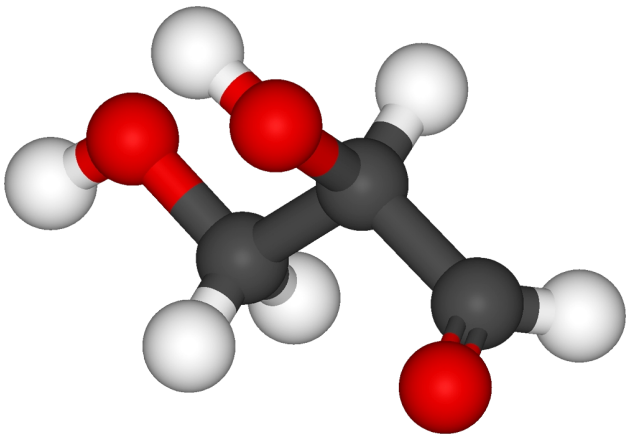
- Basics of saccharides chemistry, nomenclature
- Saccharides related open databanks
- Saccharides in the PDB
- Challenges in computational analysis of saccharides
- Tools for saccharides analysis: MD, docking
- Case study: docking glycosaminoglycans with Dynamic Molecular Docking

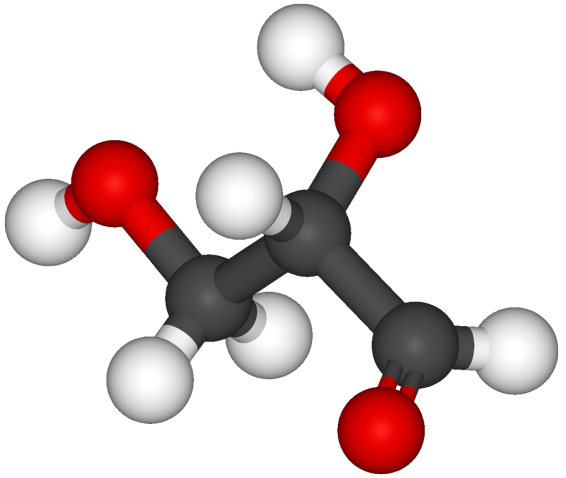
CARBOHYDRATES: GLYCERALDEHYDE

- **Saccharides/carbohydrates: $C_n(H_2O)_m$; polyhydroxy aldehydes/ketones**





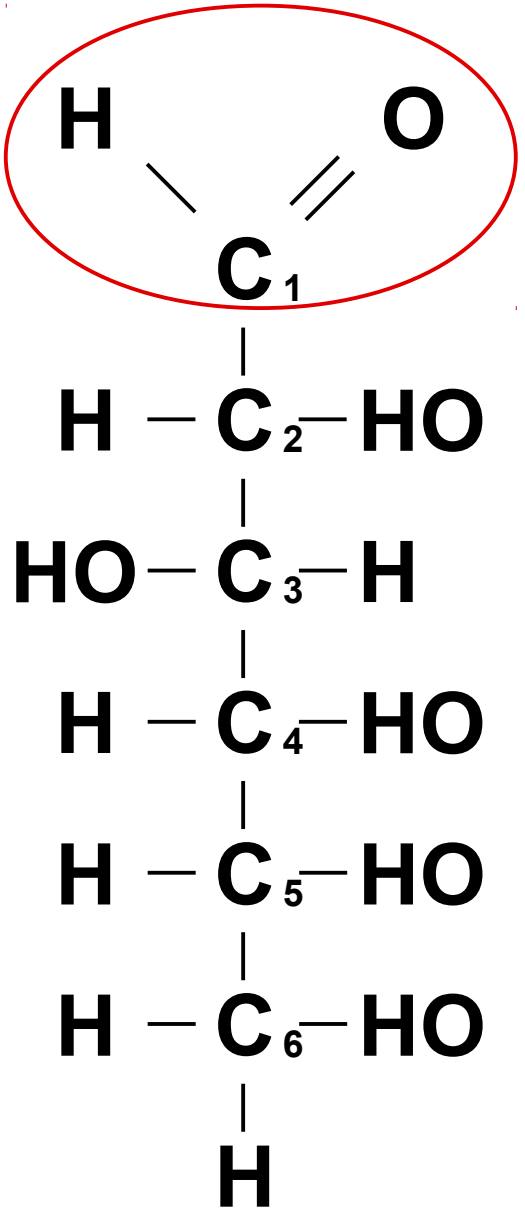




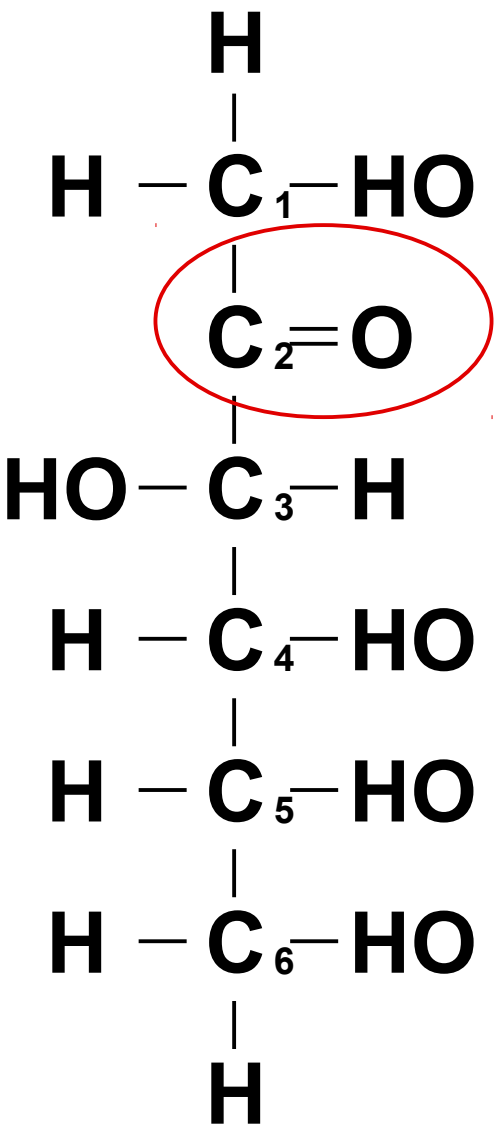
D-glyceraldehyde (R, +)

L-glyceraldehyde (S, -)

CARBOHYDRATES: ALDOSES AND KETOSES

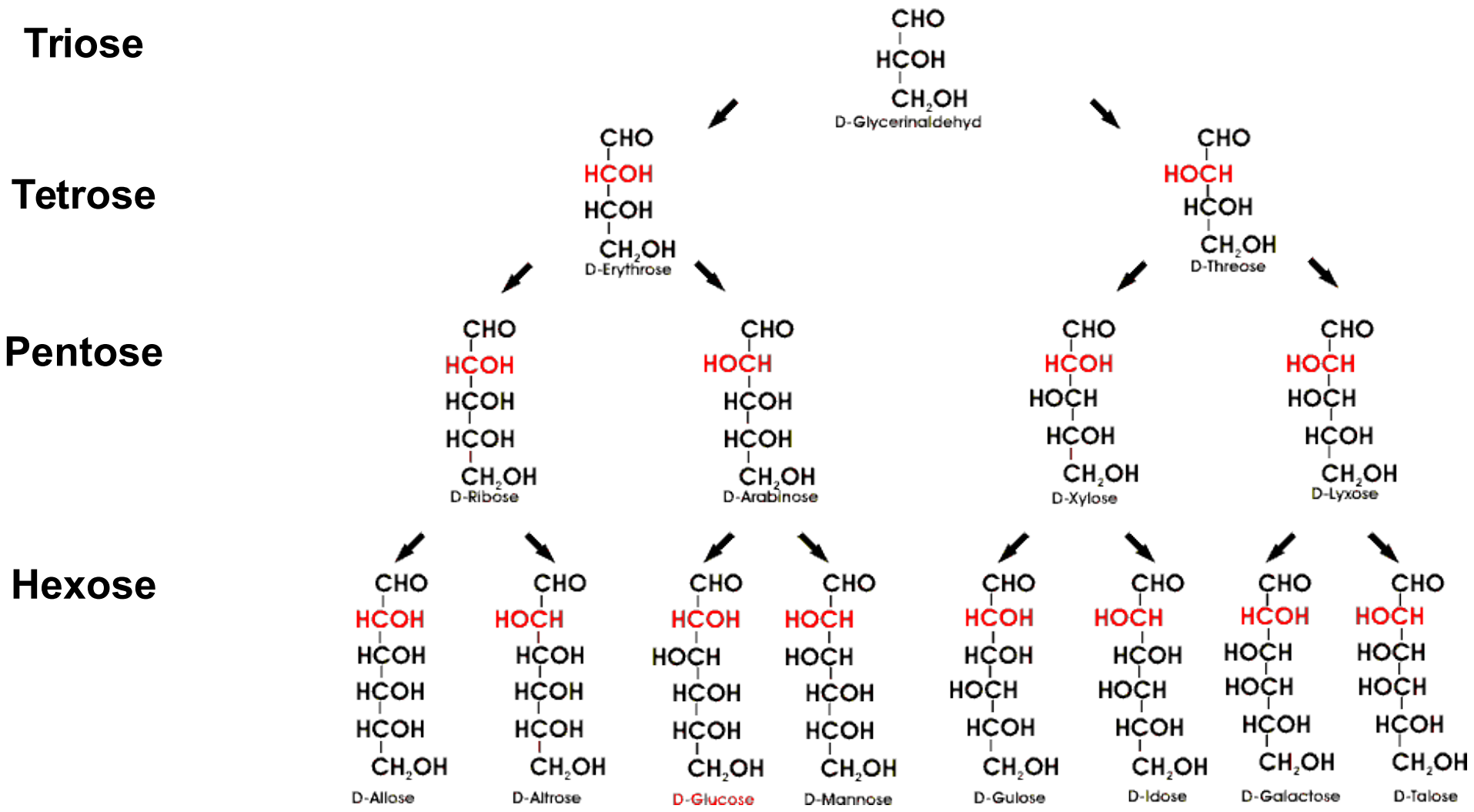


Aldose: D-glucose



Ketose: D-fructose

CARBOHYDRATES: ALDOSE TREE



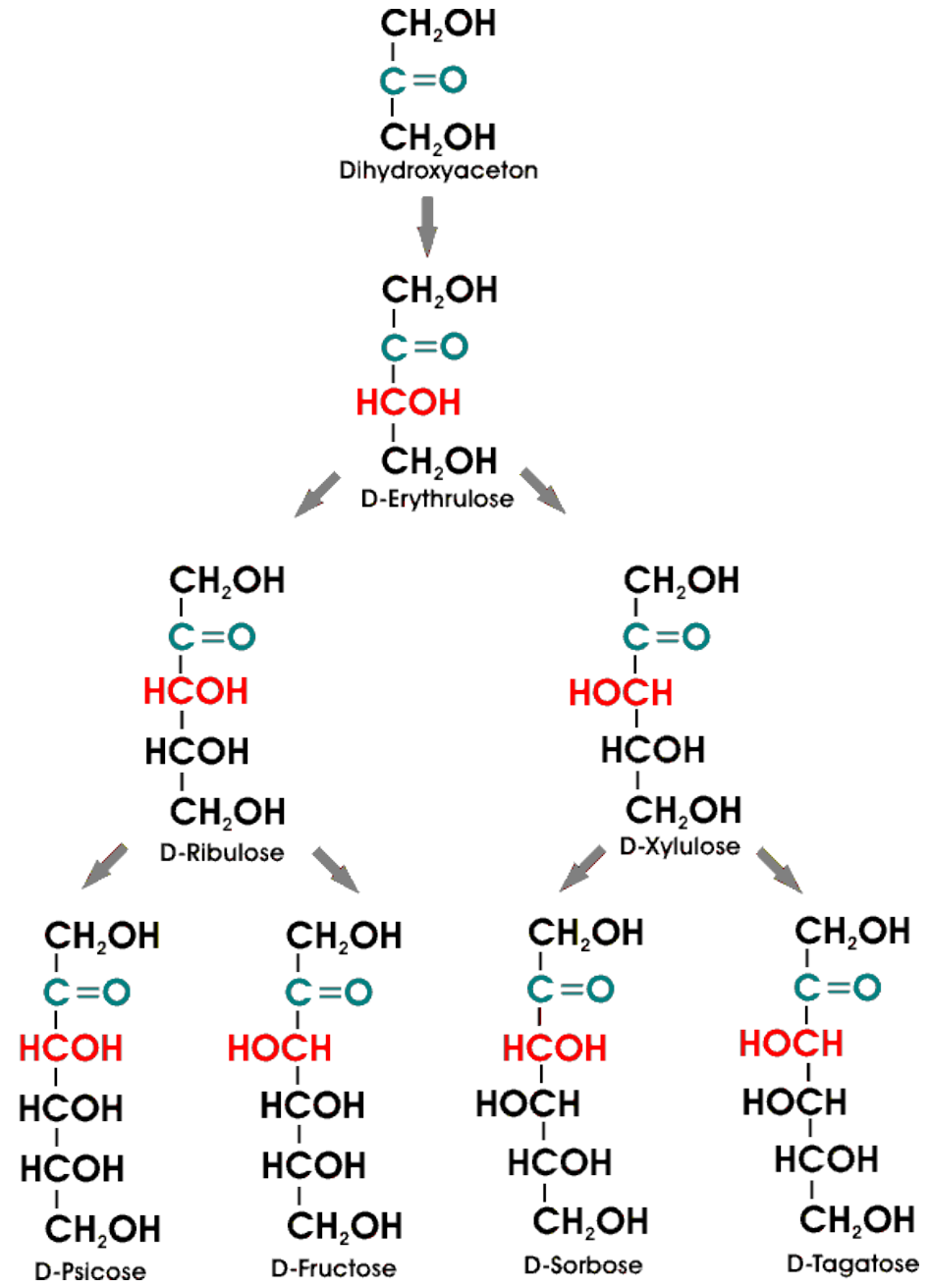
CARBOHYDRATES: KETOSE TREE

Triose

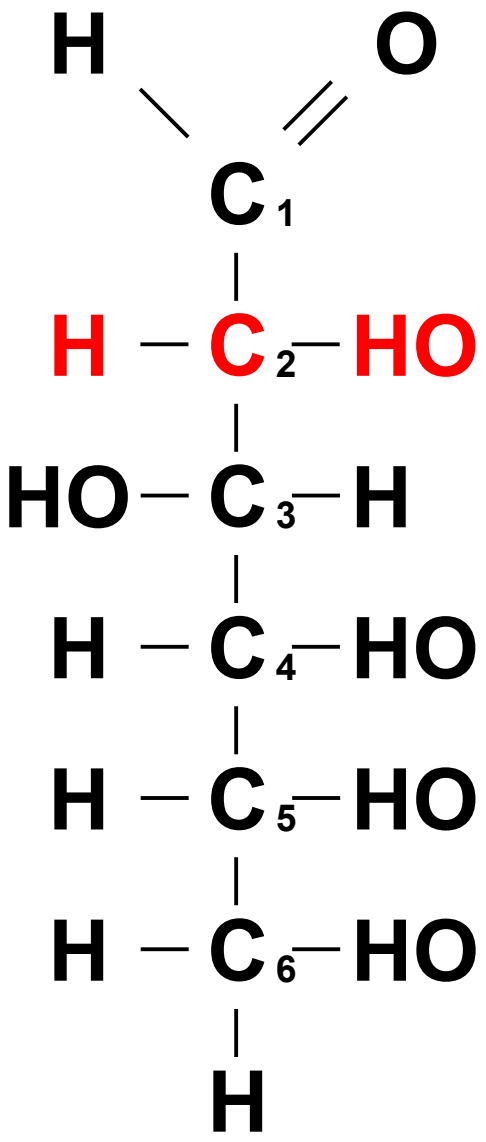
Tetrose

Pentose

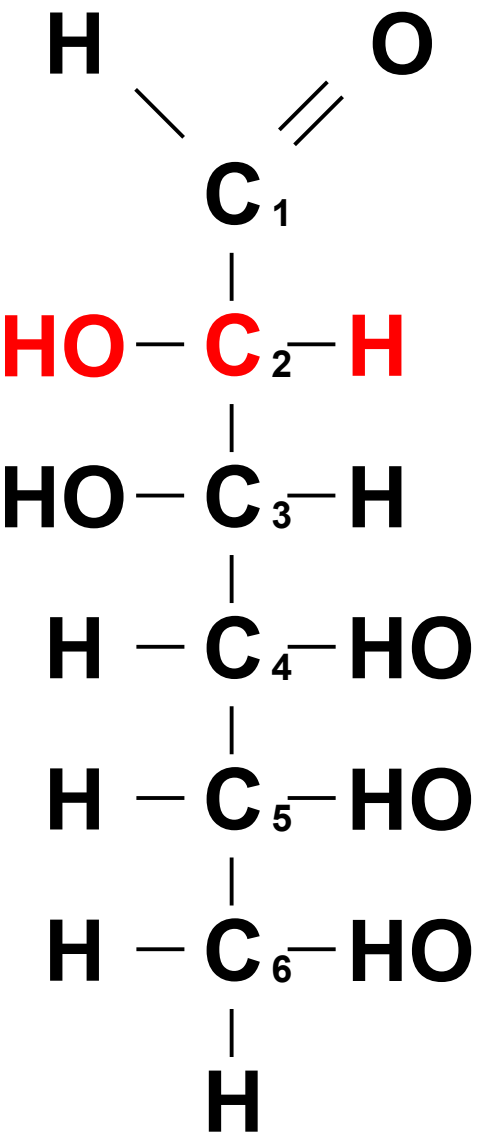
Hexose



CARBOHYDRATES: DIASTEREOMERS

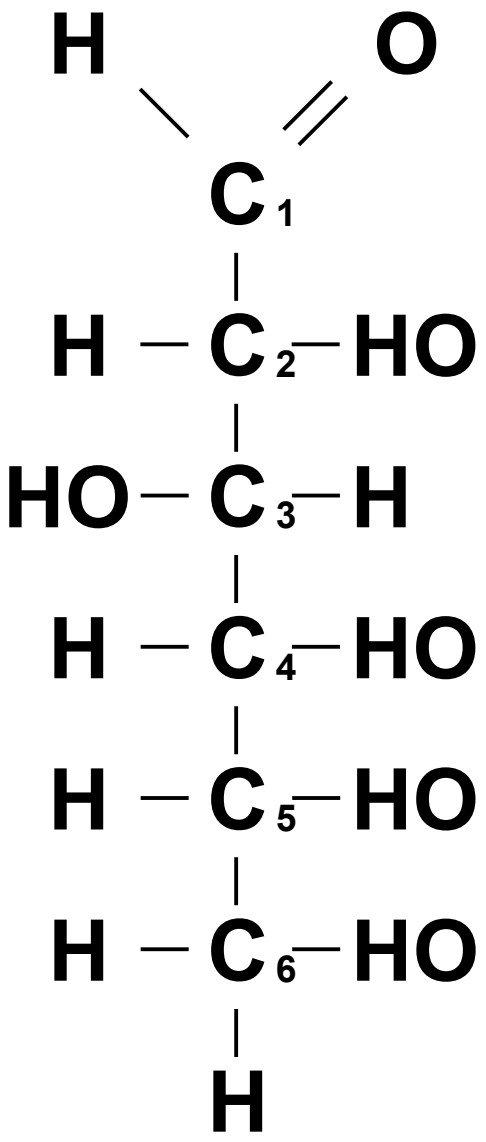


D-glucose

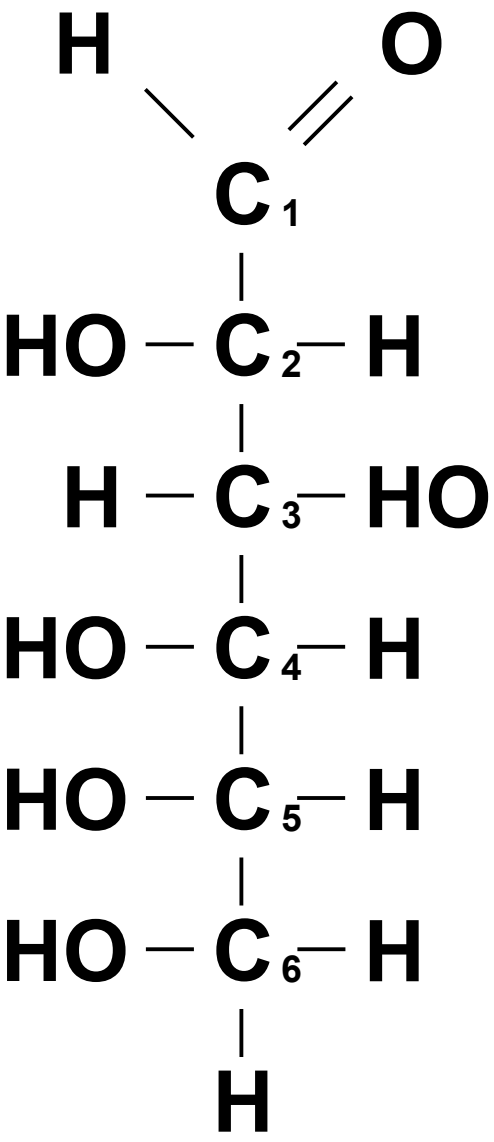


D-mannose

CARBOHYDRATES: D/L-ISOMERS

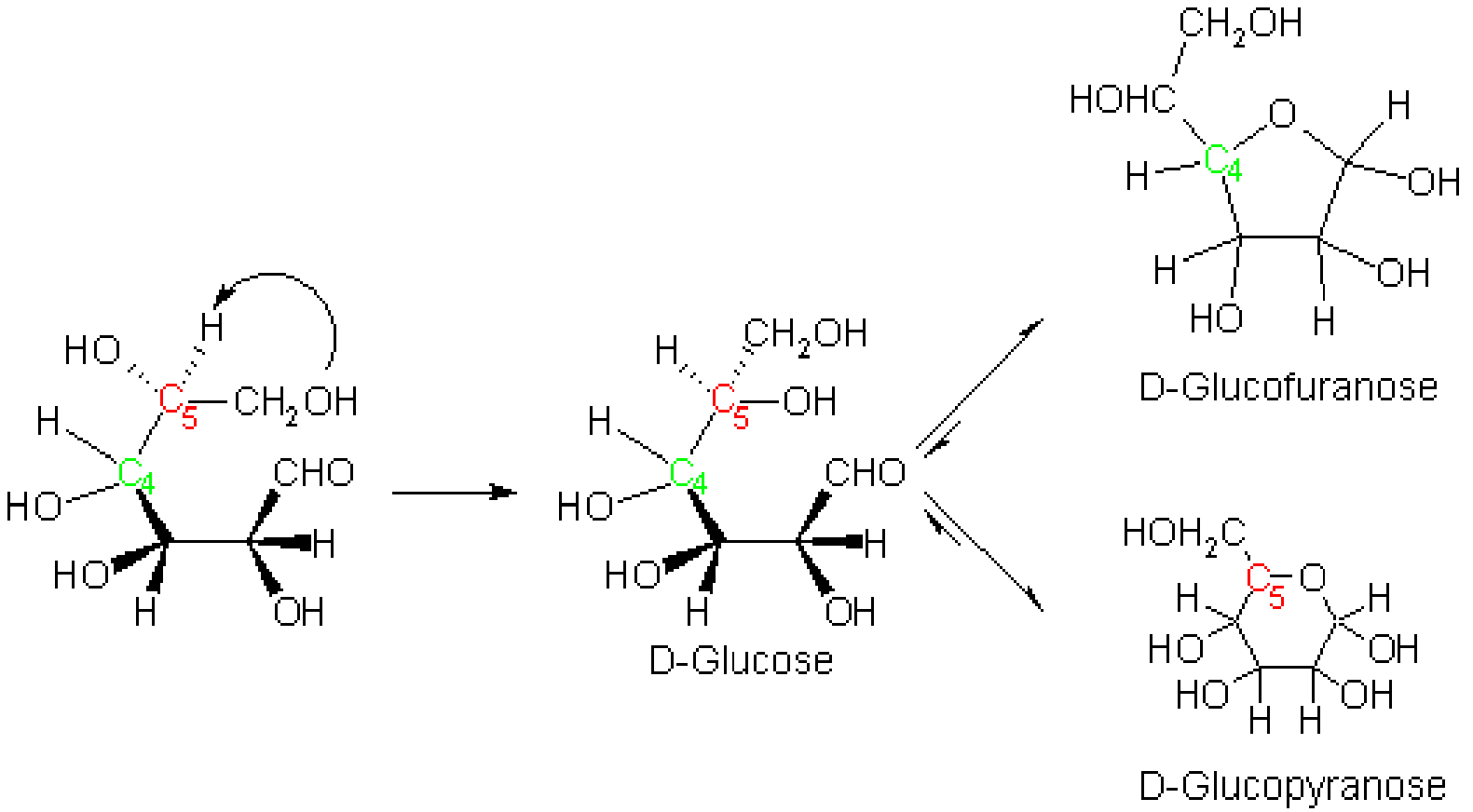


D-glucose



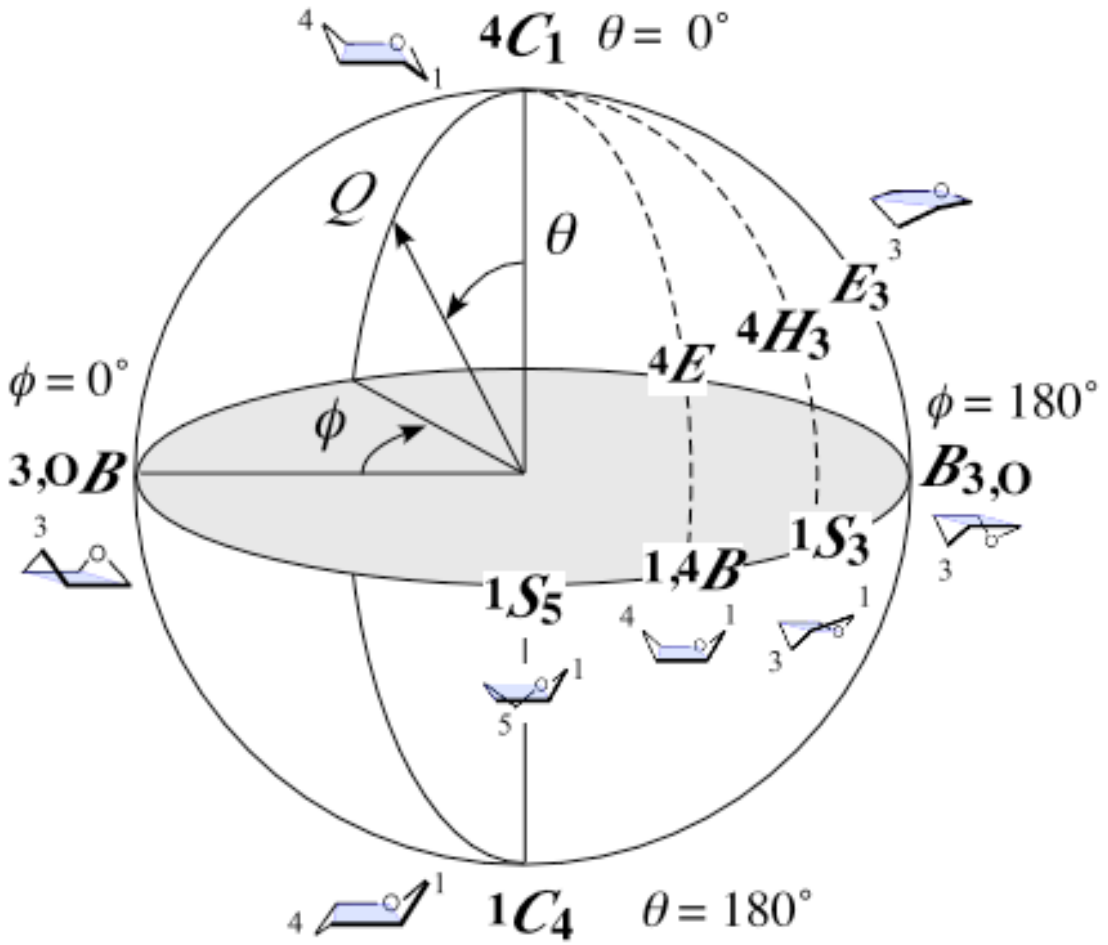
L-glucose

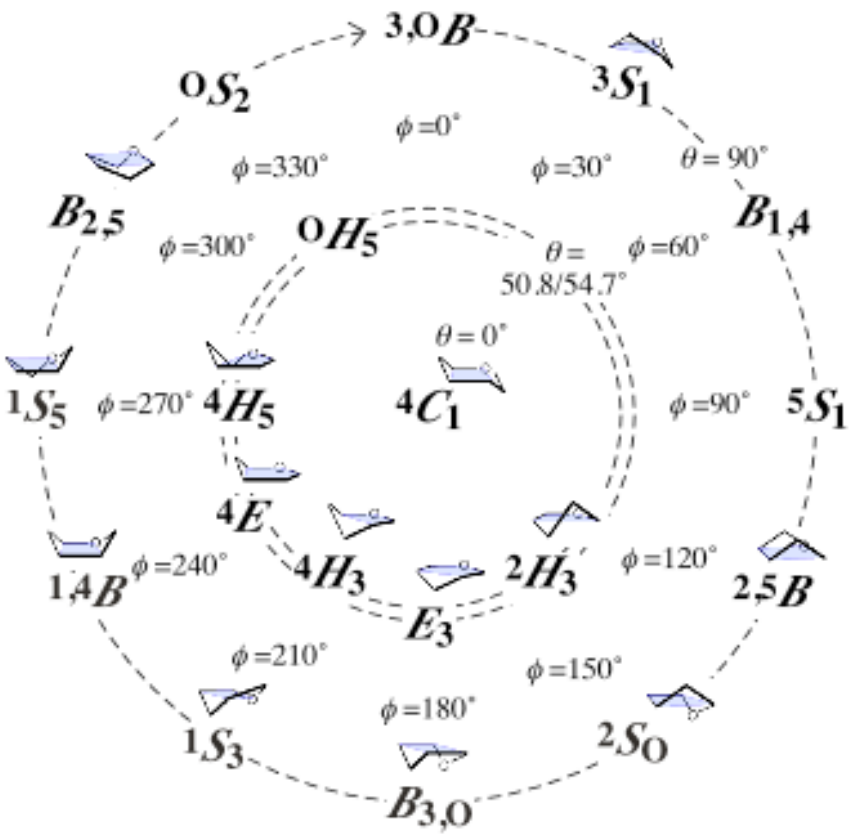
CARBOHYDRATES: CYCLIC FORMS



- **Furanose is a 5-membered ring**
- **Pyranose is a 6-membered ring**

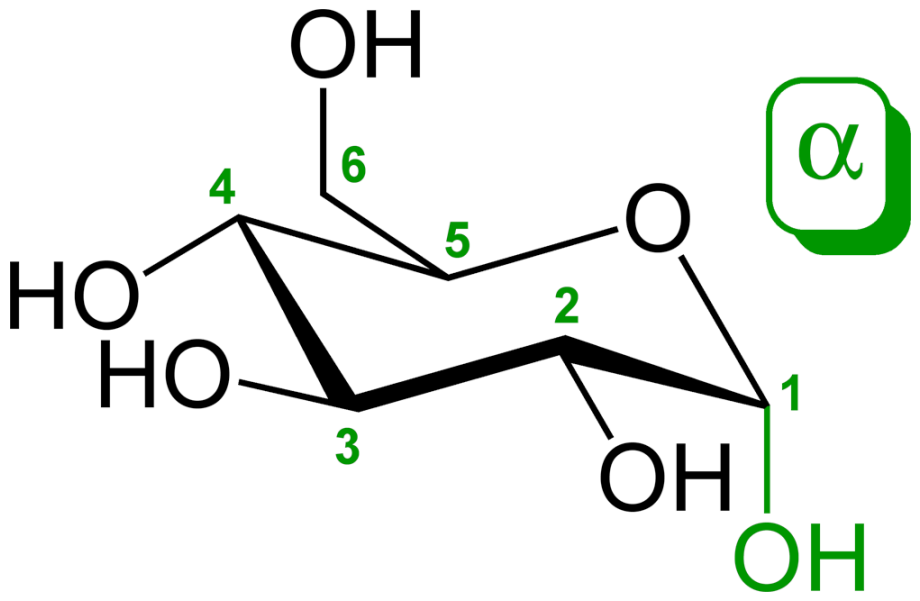
CYCLIC FORMS: PYRANOSE RING CONFORMATIONS



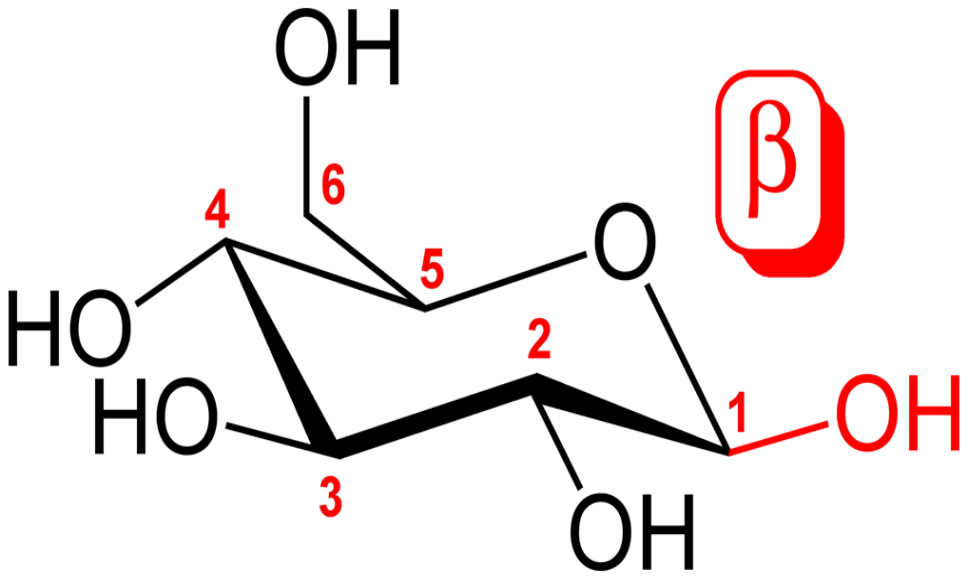


CARBOHYDRATES: α/β- FORMS

- C1- anomeric center
- α-form: CH_2OH and OH_{c1} are on the opposite sides of the ring
- β-form: CH_2OH and OH_{c1} are on the same sides of the ring



α-D-glucopyranose



β-D-glucopyranose

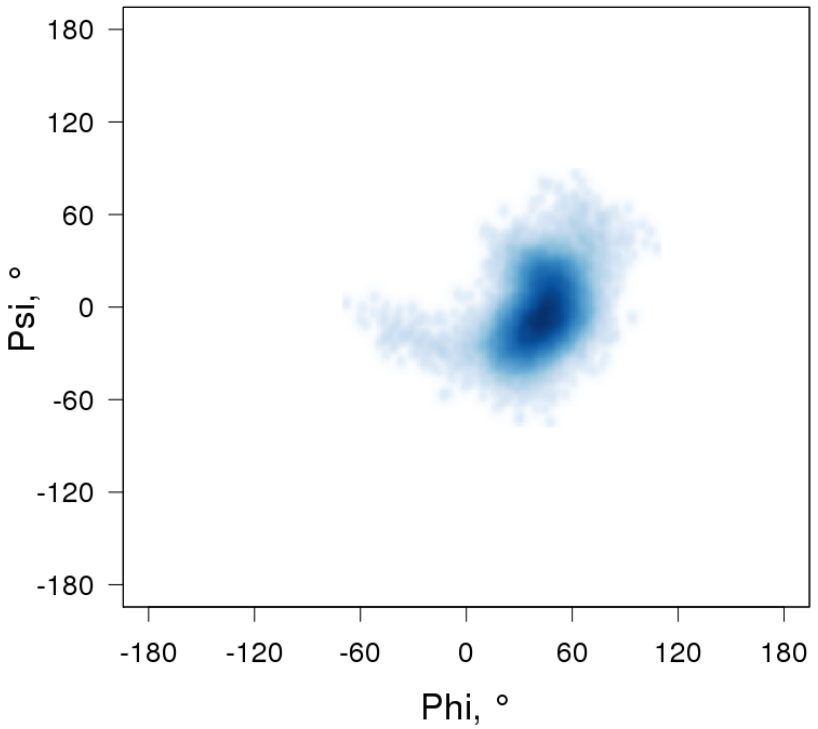
GLYCOSIDIC LINKAGE

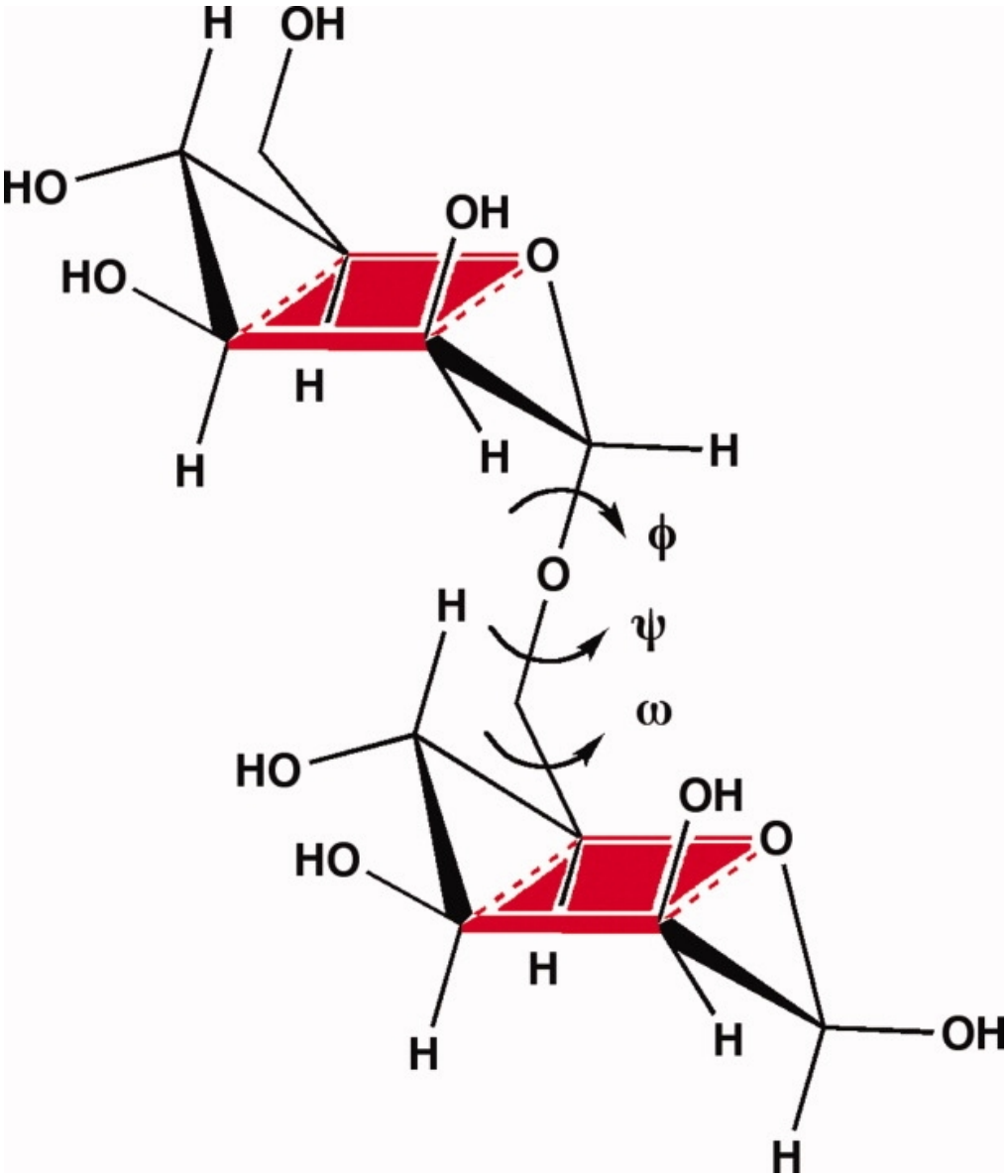
- **Glycosidic linkage: covalent bond between 2 sugars**

➢ **Maps for glycosidic linkages:**

- 2D

- 3D (1-6 linkages)





GLYCOSYLATION

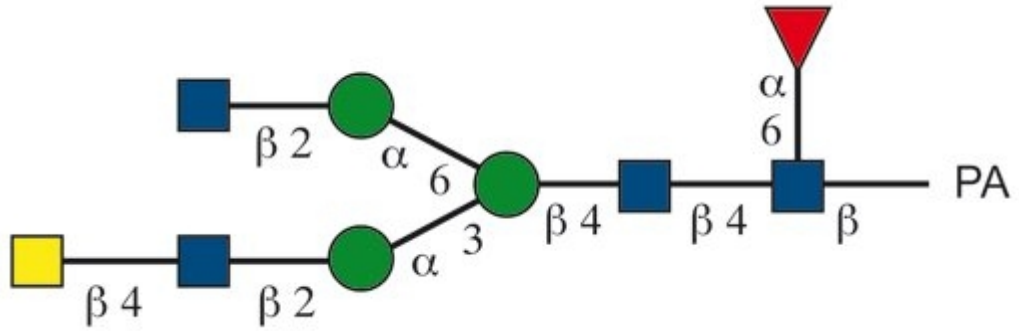
- **Glycosylation: reaction between saccharide and saccharide acceptor**

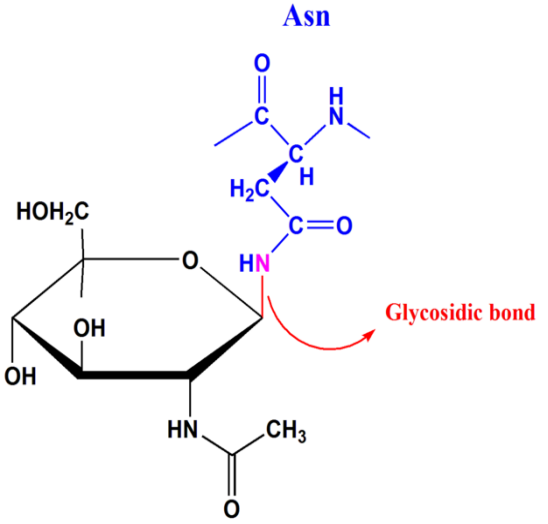
$$A + sugar = A\text{-}sugar + H_2O$$

- **Lipids**
- **Proteins:**

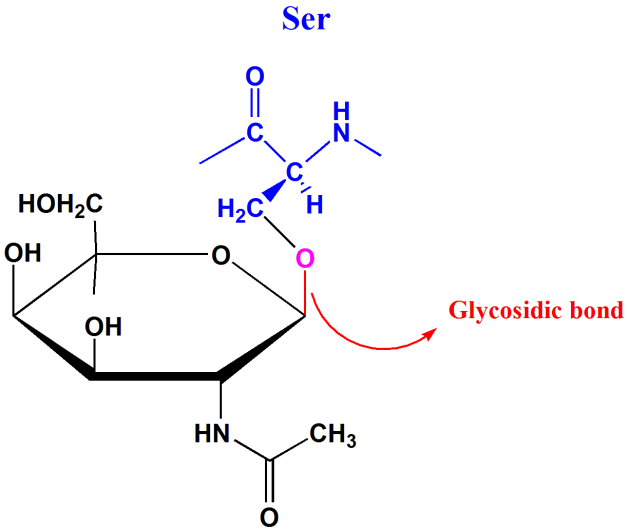
  - N-glycosylation (Asn)

  - O-glycosylation (Thr, Ser, Tyr, HO-Lys, HO-Pro)





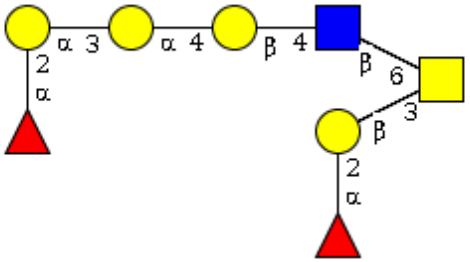
N-linked GlcAC

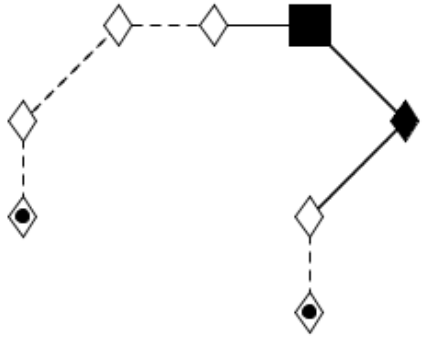


O-linked GalNAC

CARBOHYDRATES SYMBOLIC REPRESENTATION

	CFG	CFG B&W	UOXF		CFG	CFG B&W	UOXF
Xylose	☆	☆	△	Galactosamine			
Fucose	▲	▲	◆	Glucosamine			
Hexose	○	○	○	Galacturonic acid			
Galactose	●	○	◇	Glucuronic acid			
Glucose	●	●	□	Iduronic acid			
Mannose	●	●	○	Mannuronic acid			
N-acetyl hexosamine	□	□	⬢	KDN			
N-acetyl galactosamine	■	□	♦	N-acetyl neuraminic acid			★
N-acetyl glucosamine	■	■	■	N-glycolyl neuraminic acid			☆

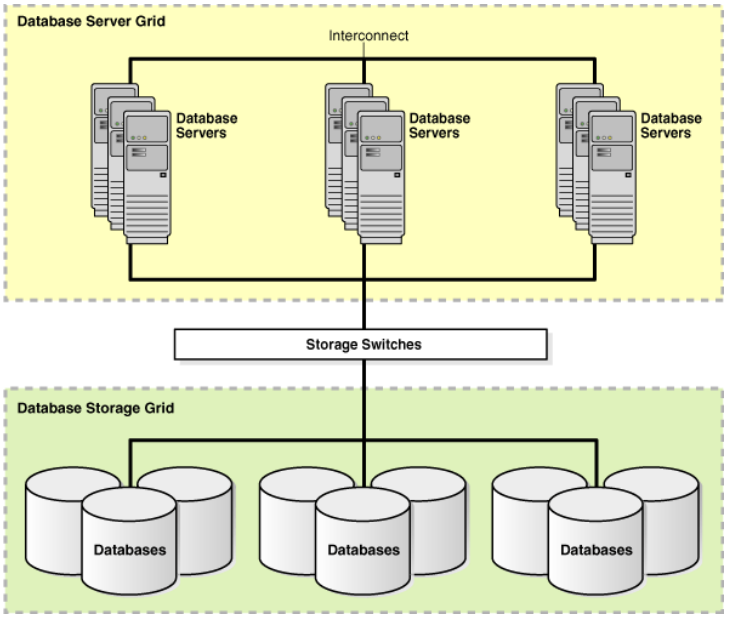




Representation of monosaccharides with geometric shapes as described in the notations used by the Consortium for Functional Glycomics (CFG) and the Oxford Glycobiology Institute (UOXF).

SACCHARIDES DATABASES

- **Structure:**
  - **Primary (MS, enzymatic reactions)**
  - **3D (X-ray, NMR)**
- **Interactions:**
  - **Microarrays**
  - **FRET**
  - **Enzymatic reactions**
  - **...**



UniCarbDB

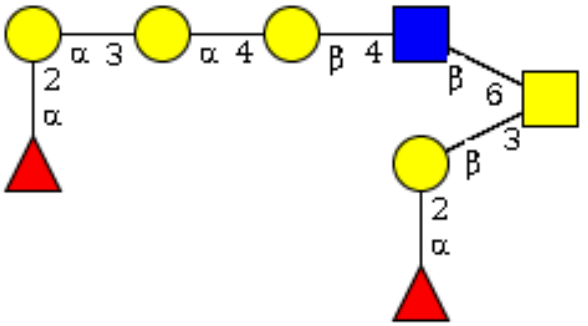
http://www.unicarbkb.org/

UniCarb-DB a glycomics initiative

A platform to build on the success of recent resources and to bring together leading researchers.

Database Access - first release of a LC-MS/MS database

- **Mass-spectroscopy data**
- **Substructure search**
- **Equivalent structures**
- **Stereochemical equivalents**
- **Superstuctures**



UniCarbDB: QUERY

Notation
CFG CFGI Text UOXF UOXFCOL

LC-MS Database
Home - View all structures

A LC-MS Data Model for Mucin Structures

To navigate the site either select a publication below to list structures published or select a structure entry in table listing to retrive MS scan and further information.

References

- Everest-Dass, **Fully Characterised Saliva N-link glycans**, (submitted), See listings
- Karlsson NG; Thomsson KA,**Salivary MUC7 is a major carrier of blood group I type O-linked oligosaccharides serving as the scaffold for sialyl Lewis x**, (Pubmed link 19043084) , See listings
- Issa S; Moran AP; Ustinov SN; Lln JHH; Ligtenberg AJ; Karlsson NG,**O-linked oligosaccharides from salivary agglutinin: Helicobacter pylori binding sialyl-Lewis x and Lewis b are terminating moieties on hyperfucosylated oligo- N-acetyllactosamine** , (Pubmed link 20466654) , See listings
- Estrella RP; Whitelock JM; Packer NH; Karlsson NG,**The glycosylation of human synovial lubricin: implications for its role in inflammation**, (Pubmed link 20443780) , See listings

Data collections are limited to negative mode MS.

Mass Find:
charge -1 -2 -3 Submit

Structure Listing Overview

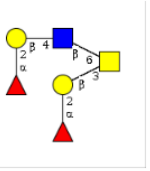
Show All

1 2 3 >> last

Glycan Name	Retention Time and Mass Information	Composition
	• RetentionTime: 26.3 (Precursor ion: 530.2 [M-H]¯)	1-aldi-D-GalNAc: 1 D-Gal: 1 L-Fuc: 1
	• RetentionTime: not available (Precursor ion: 384.3 [M-H]¯)	1-aldi-D-GalNAc: 1 D-Gal: 1
	• RetentionTime: 28.3 (Precursor ion: 1041.4 [M-H]¯)	1-aldi-D-GalNAc: 1 D-Gal: 2 D-GlcNAc: 1 L-Fuc: 2
	• RetentionTime: not available (Precursor ion: 425.3 [M-H]¯)	1-aldi-D-GalNAc: 1 D-GlcNAc: 1

UniCarbDB: QUERY => RESULTS

Glycan sequence detail



EurocarbDB Glycan Sequence ID	369
Originally contributed	Apr 15, 2008 2:21:58 PM, by Carbbank
Evidence for this sequence	LC-MS details
Biological contexts in which this sequence has been observed	• *genus*: Rattus; *tissue*: unknown/unspecified; (Detail) • *species*: Homo sapiens; *tissue*: unknown/unspecified; (Detail) • *species*: Sus scrofa; *tissue*: unknown/unspecified; (Detail) • *species*: Homo sapiens; *tissue*: Saliva; (Detail) • *species*: Bos taurus; *tissue*: unknown/unspecified; (Detail)
References	• Carbbank entry 997 • Carbbank entry 999 • Carbbank entry 1000 • Carbbank entry 1001 • Carbbank entry 2691 • Carbbank entry 7254 • Carbbank entry 7255 • Carbbank entry 7256 • Carbbank entry 12030 • Carbbank entry 20490 • Carbbank entry 23547 • Carbbank entry 24589 • Carbbank entry 29668 • Bock K, Pedersen C, Pedersen H, **Carbon-13 nuclear magnetic resonance data for oligosaccharides**, *Adv Carbohydr Chem Biochem (1984) 42; 193-225* (detail, Pubmed) • Issa S, Moran A, Ustinov S, Lln J, Ligtenberg A, Karlsson N, **O-linked oligosaccharides from salivary agglutinin: Helicobacter pylori binding sialyl-Lewis x and Lewis b are terminating moieties on hyperfucosylated oligo-N-acetyllactosamine** , *Glycobiology (2010) 0; 0-0* (detail, Pubmed id 20,466,654) • glycosciences.de entry 377
Composition	1-aldi-D-GalNAc: 1 D-Gal: 2 D-GlcNAc: 1 L-Fuc: 2
Sequence	RES 1b:o-dgal-HEX-0:0|1:aldi 2s:n-acetyl 3b:b-dgal-HEX-1:5 4b:a-lgal-HEX-1:5|6:d 5b:b-dglc-HEX-1:5 6s:n-acetyl 7b:b-dgal-HEX-1:5 8b:a-lgal-HEX-1:5|6:d LIN 1:1d(2+1)2n 2:1o(3+1)3d 3:3o(2+1)4d 4:1o(6+1)5d 5:5d(2+1)6n 6:5o(4+1)7d 7:7o(2+1)8d

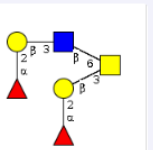
Notation
CFG CFGI Text UOXF UOXFCOL

Actions
Sub-structure search

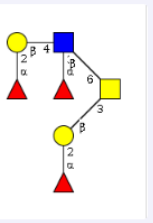
Equivalent structures
none, structure is definite

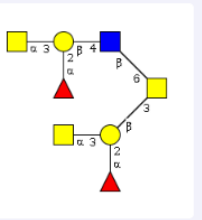
Stereochemical equivalents
none

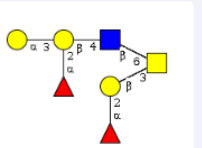
Linkage isomers
1 structure(s)



Superstructures of this structure
8 superstructures(s)



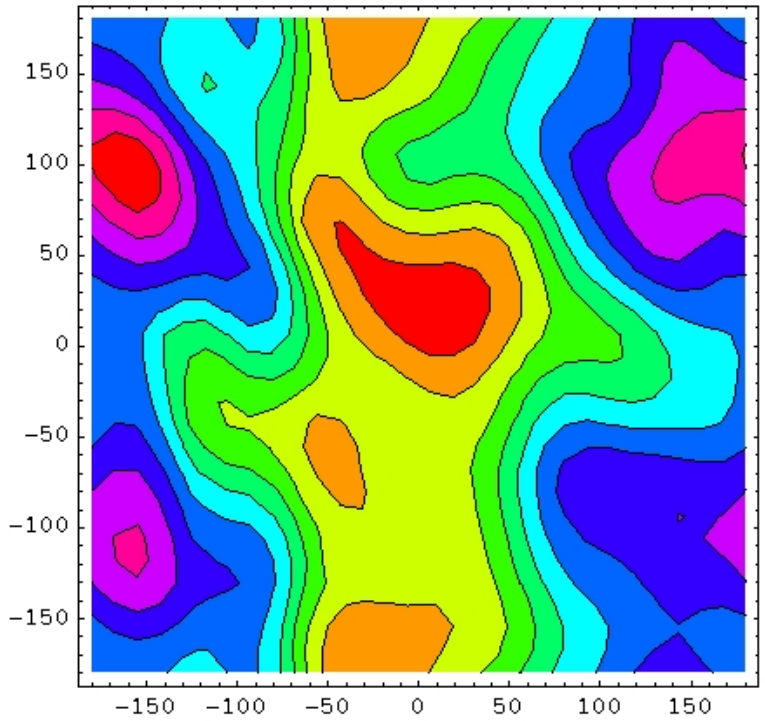


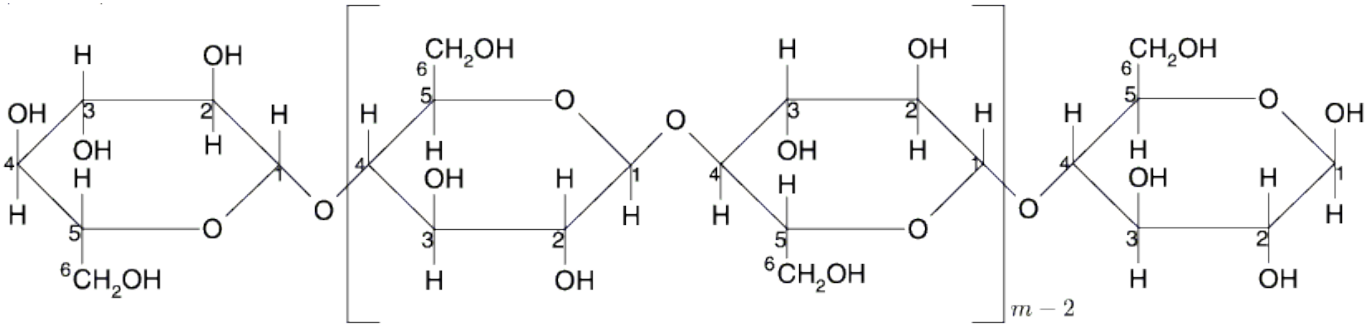


GLYCOSCIENCES.DE

http://www.glycosciences.de

- **Glycomics related databases**
- **Glycomics analysis and modelling tools**
- **Links to other databases and tools**





GLYCOSCIENCES.DE: DATABASES

•bibliography •structure •nmr •ms •pdb

/ databases

Bibliography

Author query

query author normal
query author fuzzy

Title query

query title normal
query title fuzzy

advanced query

Structure

substructure search (beginner)
substructure search (advanced)
exact structure search

composition
molecular formula
classification (CarbBank)
n-glycan classification
motifs (o-glycan, Lewis, ...)

NMR

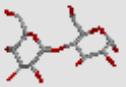
atom search
peak search
shift estimation

Mass Spectroscopy

glyco-search-ms
profiling

Direct Query

Enter LinucsID: Submit Query



PDB

search pdb data

GLYCOSCIENCES.DE: DATABASES

GLYCOSCIENCES.DE

Home | Databases | Modeling | Tools | Links

• bibliography • structure • nmr • ms • pdb

/ databases / bibliography

Bibliography Search

Author query

•normal •fuzzy

Title query

•normal •fuzzy

advanced query

back to top

Structure Search

Select query type:

substructure search (beginner)

(preselected residues only)

substructure search (advanced)

(free text input of residues)

exact structure search

```
b-D-Galp-(1-4)+
                |
          b-D-GlcpNAc-(1-3)-b-D-Galp
                |
a-L-Fucp-(1-3)+
```

molecular formula

		#C	#H	#O	#N	#P	#S	#F	#I	#Cl	#Br
Chemical Formula:	$C_{19}H_{32}F_2O_{14}$	19	32	14	0	0	0	2	0	0	0
Molec. Weight:	522										
# Atoms:	67										
# Residues:	4										
# Heavy Atoms:	35										

composition

Neu		Hex		Other	
NeuAc	1	Hex	4	Me	1
		HexNAc	2		

other:

- classification (CarbBank)
- n-glycan classification
- motifs (o-glycan, Lewis, ...)

Query by LinucsID: [ ] submit LinucsID

GLYCOSCIENCES.DE: DATABASES

atom search
displays a histogram of all 1H- or 13C-NMR shifts assigned to a certain atom (e.g. H-1 of Galactose) contained in the databank.

peak search
compares a list of NMR-shifts (1H- or 13C) given by the user with all spectra contained in the database. A hit-list of spectra and structures in descending order of their spectral similarity is displayed.

shift estimation
estimates 1H- or 13C spectra of given structures based on the assumption, that similar structural environments exhibit also similar spectra.

back to top

NMR Information / Advanced Search

Peaks :
92.699997
79.699997
70.300003
67.300003
61.299999
100.900002
79.099998
70.300003
67.300003
73.5
61.299999

○ Proton ◉ Carbon

Tolerance : 0.5 PPM

Search now

Example 1: Proton Search

Example 2: Carbon Search

Searched peaks (spectra view).

Structures Details
Spectrum matches query with 89.23%
Explore theor. 3D Co-ord.
180.0 160.0 140.0 120.0 100.0 80.0 60.0 40.0 20.0 0.0

Structures Details
Spectrum matches query with 86.92%
Explore theor. 3D Co-ord.
180.0 160.0 140.0 120.0 100.0 80.0 60.0 40.0 20.0 0.0

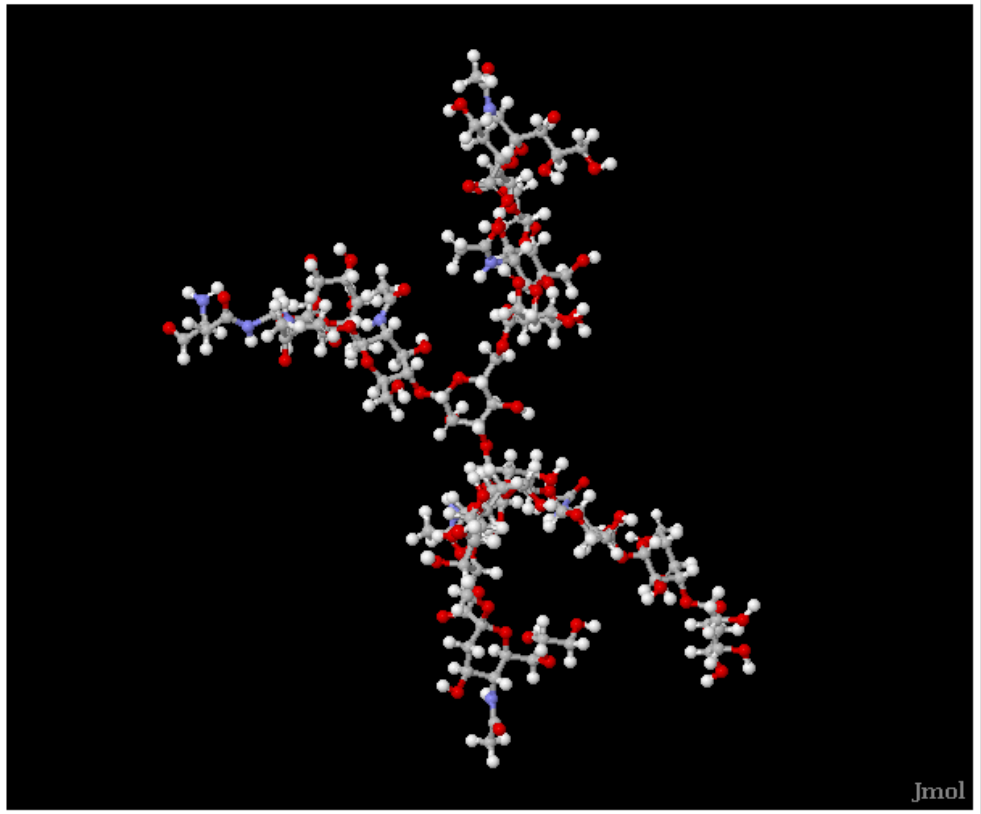
Structures Details
Spectrum matches query with 86.15%
Explore theor. 3D Co-ord.

Structures Details
Spectrum matches query with 70.00%
Explore
180.0 160.0 140.0 120.0 100.0 80.0 60.0 40.0 20.0 0.0

Structures Details

GLYCOSCIENCES.DE: DATABASES

3D Coordinates for LinucsID 3162



Display:

wireframe

sticks

balls and sticks

spacefill

Background:

white

black

RasMol/Chime script command:

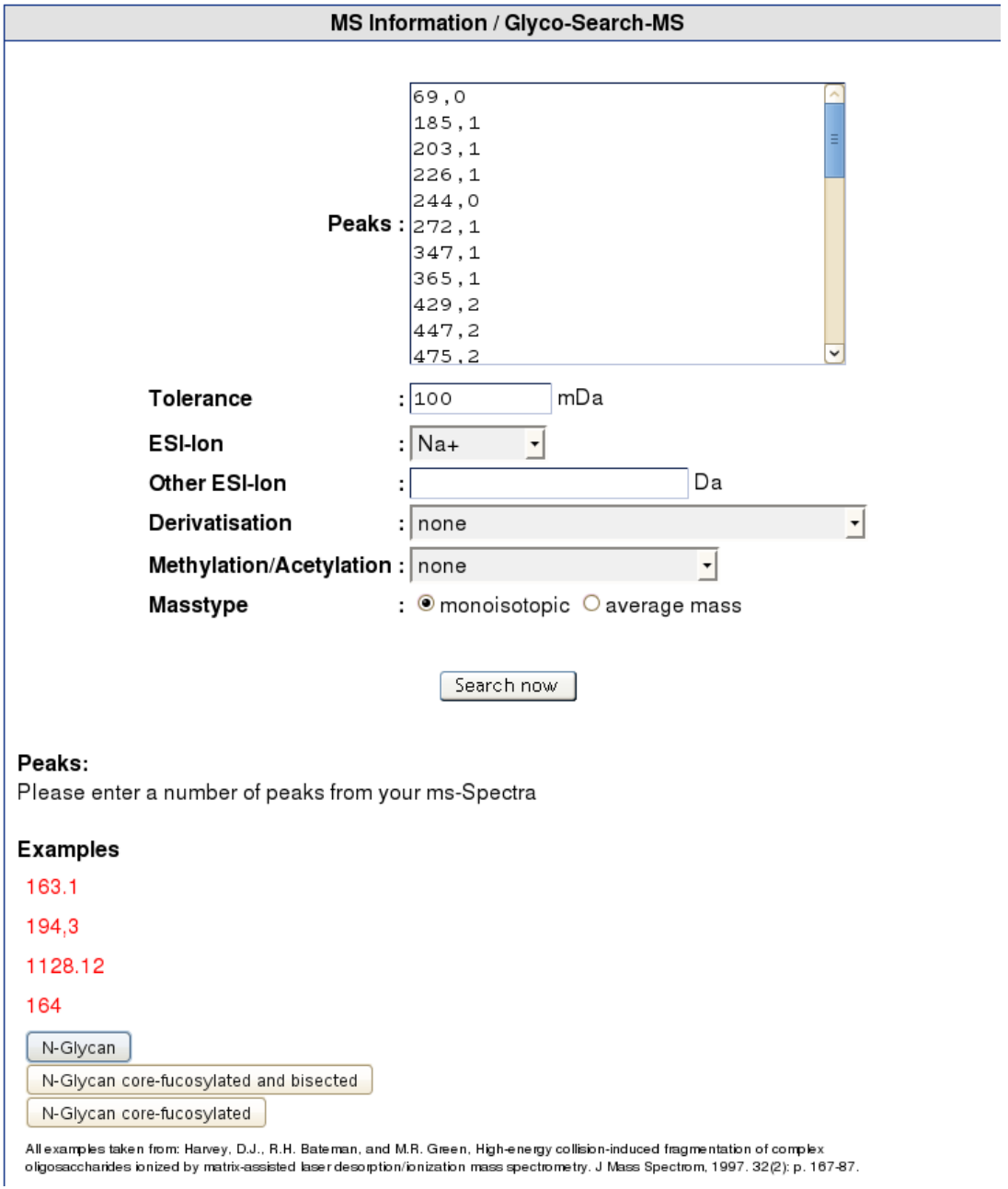
Execute Command

Download pdb file

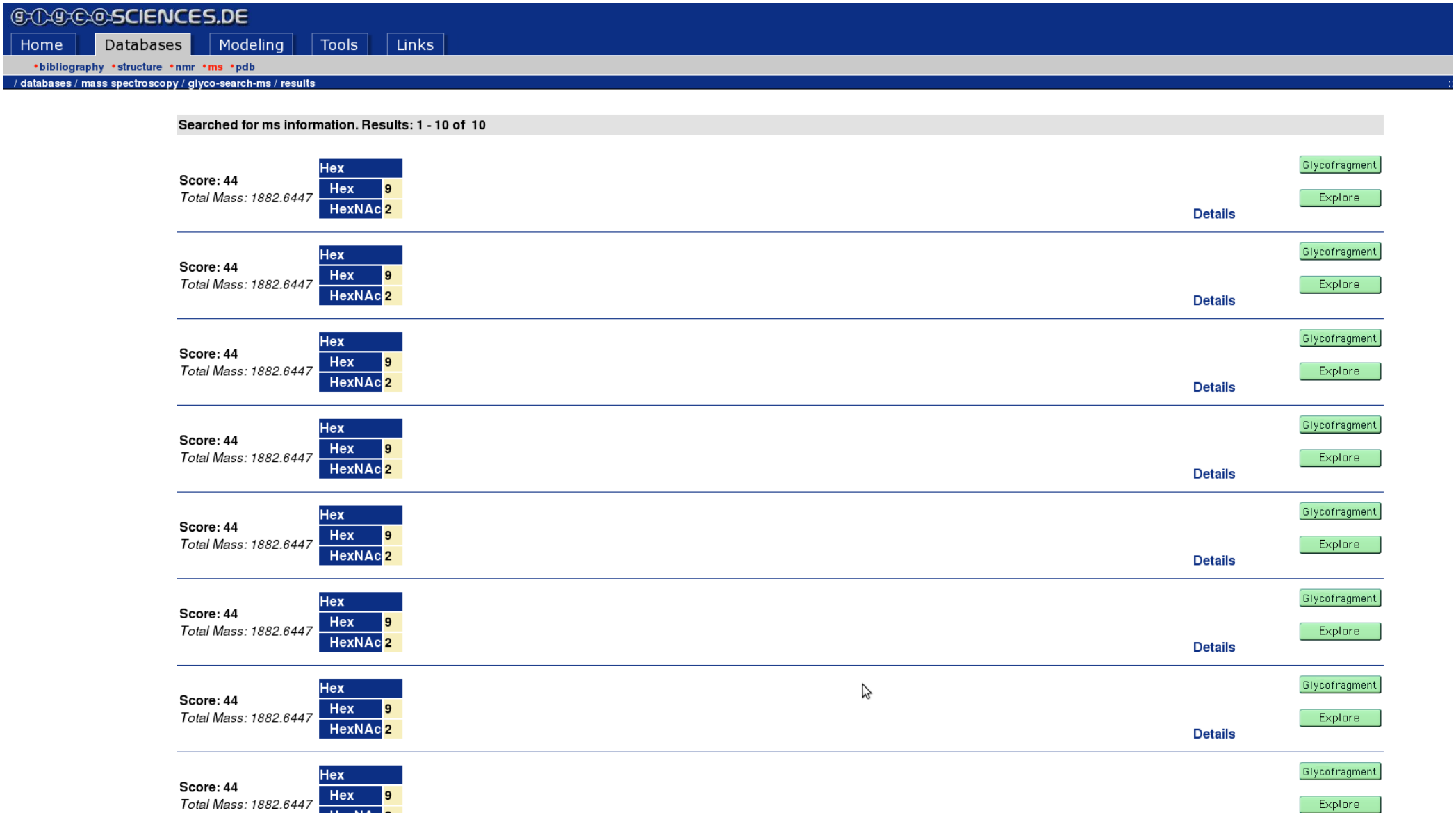
Structure for LinucsID 3162:

```
                                                                    a-L-Fucp-(1-6)+
                                                                                  |
a-D-Neup5Ac-(2-6)-b-D-Galp-(1-4)-b-D-GlcpNAc-(1-2)-a-D-Manp-(1-6)+                b-D-GlcpNAc-(1-4)-Asn
                                                                |                 |
         a-D-Galp-(1-3)-b-D-Galp-(1-4)-b-D-GlcpNAc-(1-4)+   b-D-Manp-(1-4)-b-D-GlcpNAc-(1-4)+
                                                         |        |
                                                    a-D-Manp-(1-3)+
                                                                  |
      a-D-Neup5Ac-(2-6)-b-D-Galp-(1-4)-b-D-GlcpNAc-(1-2)+
```

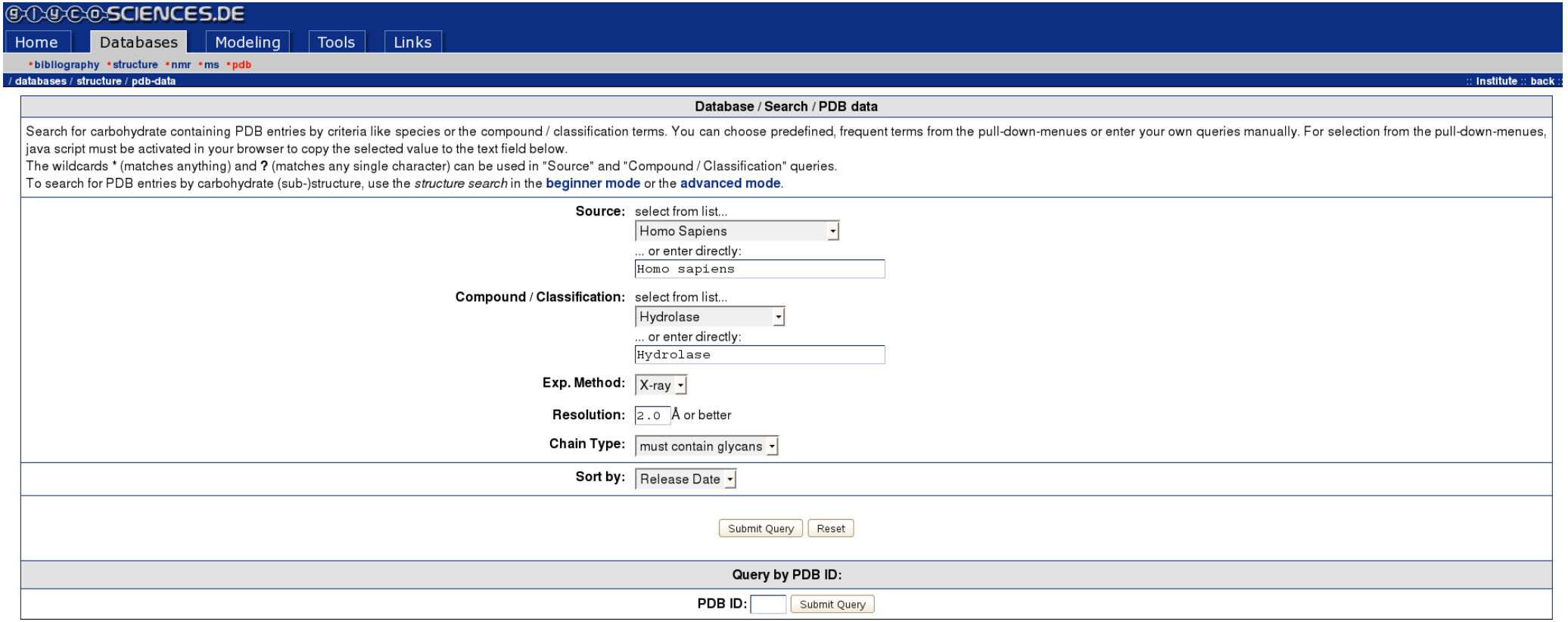
GLYCOSCIENCES.DE: DATABASES



GLYCOSCIENCES.DE: DATABASES



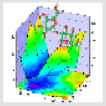
GLYCOSCIENCES.DE: DATABASES



GLYCOSCIENCES.DE: TOOLS

SWEET2

rapidly converts the primary sequence of a complex carbohydrate, as defined by standard nomenclature, directly into a reliable 3D molecular model by linking together preconstructed 3D molecular templates of monosaccharides in the manner specified by the sequence and then optimizing the 3D structure using the MM3 force field.

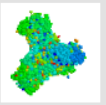


GlycoMaps DB

is a data base system for the management of conformational maps and profiles, The system allows conformational maps to be archived in a standard format, and it will provide search and comparison facilities. An interface to structures from Sweet-DB is implemented. We also offer scientists the possibility of adding their own publicized structures to the database via a web interface.

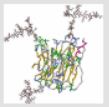
Dynamic Molecules

Dynamic Molecules is the first Internet portal which provides interactive access to the techniques of molecular dynamics simulations and tutorials via standard Web technologies and using only publicly available software. The 'expert mode' has been specially developed to explore the conformational space of oligosaccharides.



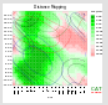
PDB2MultiGif

Visualization of chemical 3D structures on the web comes with problems because the web browser cannot display chemical structures without the help of additional software. If you create a page with a 3D structure of a molecule and the visitor of your page does not use this special viewer software for displaying molecules it cannot get the whole information of the page which should be meditated.
PDB2MultiGIF takes the 3D structure and generates an animated image which can be displayed using any browser. Thus every visitor of your page can get the whole information.



GlyProt

performs an in silico glycosylation of proteins. The 3D structure of protein is required as input. Potential N-glysylations site are automatically detected. The attached glycan are constructed with SWEET-II.



Distance Mapping

Commonly, computational methods, which explore the conformational space of oligosaccharides, are discussed in conjunction with experimental results mostly derived from NMR data. The nuclear Overhauser enhancement (NOE) allows one to detect the proximity in space between protons that may be located in different, yet spatially neighbouring residues of oligosaccharides.

The DISTANCE MAPPING approach allows to draw distances of equal r as a function of the appropriate F,Y coordinates. A single pair of contours drawn for the lower and upper limits of r for one NOE encloses a torus-like region which still covers an infinite number of F,Y conformations.

- **Translation of sequence to 3D (SWEET)**
- **Conformational maps and profiles (GlycoMaps DB)**
- **Preparation of MD inputs**
- **Animated Gif from PDB (PDB2MultiGif)**
- ***In silico* glycosylation of proteins (GlyProt)**
- **NMR parameters relation to structures (Distance Mapping)**

GLYCOSCIENCES.DE: TOOLS

Input / Work

Classes of complex Saccharides

Templates

Background

Integration of helper tools (RasMol)

Examples

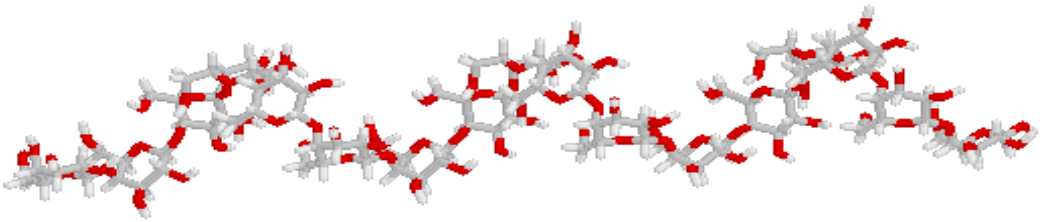
Guestbook

Our-Homepage

Sweet is a program for constructing 3D models of saccharides from their sequences using standard nomenclature.

SWEET

beginner version	expert version	direct input
example page	example page	example page
work page	work page	work page



Web-Services: SOAP

Input / Work

Classes of complex Saccharides

Templates

Background

Integration of helper tools (RasMol)

Examples

Guestbook

Our-Homepage

Example page for the beginner version:

Saccharide in nomenclature:

a-D-Manp - (1-4) - a-D-Manp - (1-4) - a-D-Manp - (1-4) - a-D-Manp - (1-4) - a-D-Manp - (1-4) - a-D-Manp - (1-4) - a-D-Manp

Input for the web-interface:

A-D-MANP | 1-4 | A-D-MANP | 1-4 | A-D-MANP | 1-4 | A-D-MANP | 1-4 | A-D-MANP | 1-4 | A-D-MANP

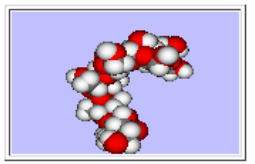
SEND

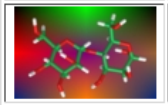
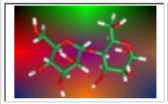
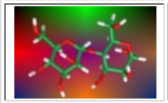
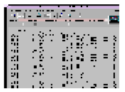
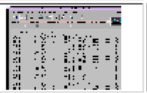
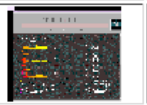
back to top

GLYCOSCIENCES.DE: TOOLS

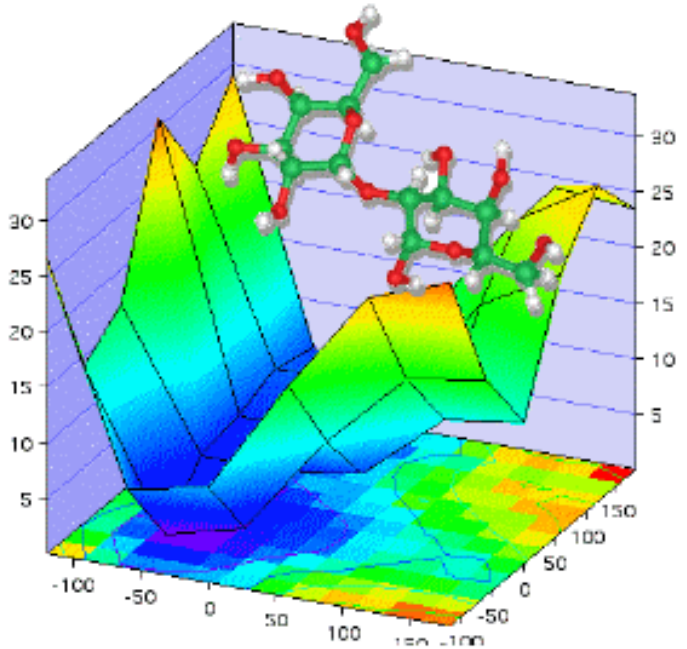
Sweet Result

You called the programm **SWEET**:
The Input was interpreted as followed: []][A-D-MANP]{[(1-4)][A-D-MANP]{[(1-4)][A-D-MANP]{[(1-4)][A-D-MANP]{[(1-4)][A-D-MANP]{[(1-4)][A-D-MANP]{}}}}}}



Link	Topic	MIME type	Comment
	Saccharide PDB file	(chemical/x-pdb)	(use RasMol or Chime for viewing)
	View saccharide with JMol-Applet	Java	JMol
	View saccharide with Chemis3D-Applet	Java	Chemis3D
	View molecule by AISMIG	no additional software	AISMIG (An Interactice Server-side Molecule Image Generator)
	PDB file with ATOM PDB file with HETATM PDB file with ATOM&HETATM	(ASCII)	
	Saccharide PDB file UMF (universal molecular format)	(ASCII)	
Optimize	Method: Full MM3(96) parameters Gradient: 1.0		will take a little time - please wait
	Conformation map file		
Structure View	Structure view program from Wolf-D. Ihlenfeldt		http://www2.ccc.uni-erlangen.de/services/gif.html
Babel 1.3	Call Babel in order to get more output formats		Will take a little time - please wait

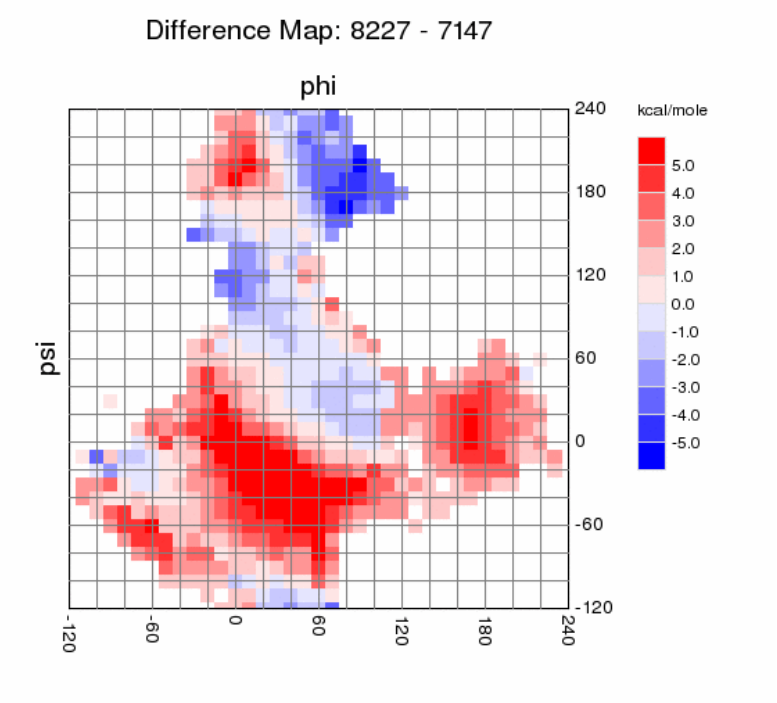
GLYCOSCIENCES.DE: TOOLS

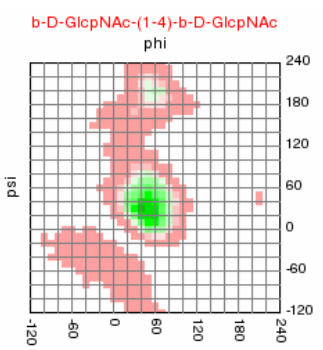


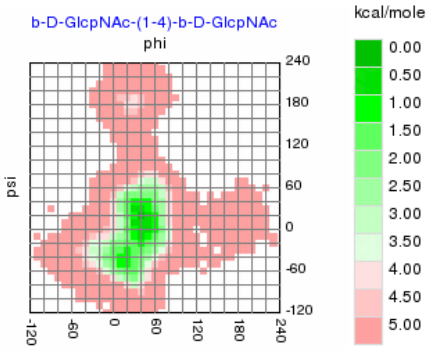
GlycoMaps Database

This database currently contains 2585 conformational maps.

GLYCOSCIENCES.DE: TOOLS







Map ID:	8227	7147
Disaccaride Fragment:	b-D-GlcpNAc-(1-4)-b-D-GlcpNAc	b-D-GlcpNAc-(1-4)-b-D-GlcpNAc
Complete Structure:	b-D-GlcpNAc-(1-4)+ | b-D-GlcpNAc | a-L-Fucp-(1-3)+	b-D-GlcpNAc-(1-4)-b-D-GlcpNAc
Linkage Path:	4	4
Calculation Method:	MD	MD
Forcefield / QM Method:	MM3(1996)	MM3(1996)
Details:	HTMD, 1000K, 30ns	HTMD, 1000K, 10ns
Software used:	Tinker 4	Tinker 4
Map Type:	FGE	FGE
Date added:	2004-04-22	2004-04-21
Comment:	Automatically generated using CAT	Automatically generated using CAT
Authors:	M.Frank, DKFZ Heidelberg	M.Frank, DKFZ Heidelberg
References:	GlycoMapsDB	GlycoMapsDB

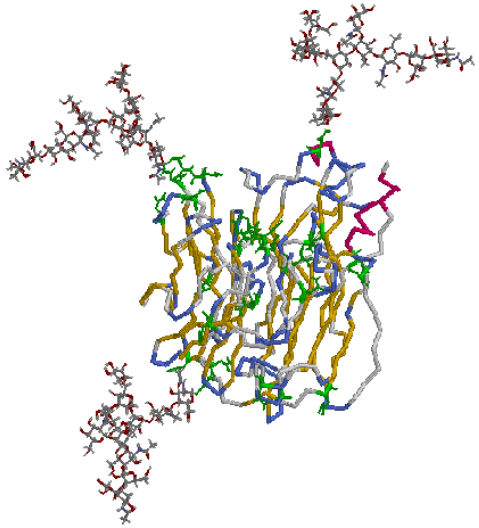
GLYCOSCIENCES.DE: TOOLS

GlyProt - *In Silico* Glycosylation of Proteins

Introduction

It is estimated that over 50% of all the proteins are glycosylated. But most of the 3D structures of proteins stored in PDB do have no attached glycans. GlyProt is capable to connect N-glycans *in silico* to a given 3D protein structure.

Contact: Andreas Bohne — References



Enter PDB-ID

PDB ID.: [ ]

Submit

Example: GP120

1G9M

More Examples

1AOL | Test

Upload a file in PDB format

File: [ ] Browse...

Upload File

Distance Mapping

Introduction

Commonly, computational methods, which explore the conformational space of oligosaccharides, are discussed in conjunction with experimental results mostly derived from NMR data. The nuclear Overhauser enhancement (NOE) allows to detect the proximity in space between protons that may be located in different, yet spatially neighbouring residues of oligosaccharides.

The DISTANCE MAPPING approach allows to draw distances of equal r as a function of the appropriate ϕ,ψ coordinates. A single pair of contours drawn for the lower and upper limits of r for one NOE encloses a torus-like region which still covers an infinite number of ϕ,ψ conformations.

Please note: This application only works if Java Script is activated in your web browser.

Contact: Martin Frank

Step 1: Create Disaccharide with Sweet II:

(Select residues from menu or enter residue names manually)

a-D-Galp | 1-2 | a-D-Galp

a-D-Galp -(1-2)- a-D-Galp

Analyzable atoms:

- (•) Hydrogens only (e.g. for interpretation of NOESY spectra)
- ( ) Hydrogens, Oxygens and Nitrogens (e.g. for H-bond analysis)

Submit Query

GLYCOSCIENCES.DE: TOOLS

Distance Mapping

Introduction

Commonly, computational methods, which explore the conformational space of oligosaccharides, are discussed in conjunction with experimental results mostly derived from NMR data. The nuclear Overhauser enhancement (NOE) allows to detect the proximity in space between protons that may be located in different, yet spatially neighbouring residues of oligosaccharides.
The DISTANCE MAPPING approach allows to draw distances of equal r as a function of the appropriate ϕ,ψ coordinates. A single pair of contours drawn for the lower and upper limits of r for one NOE encloses a torus-like region which still covers an infinite number of ϕ,ψ conformations.

Please note: This application only works if Java Script is activated in your web browser.

Contact: Martin Frank

Step 1: Create Disaccharide with Sweet II:

(Select residues from menu or enter residue names manually)

a-D-Galp | 1-2 | a-D-Galp

a-D-Galp -(1-2)- a-D-Galp

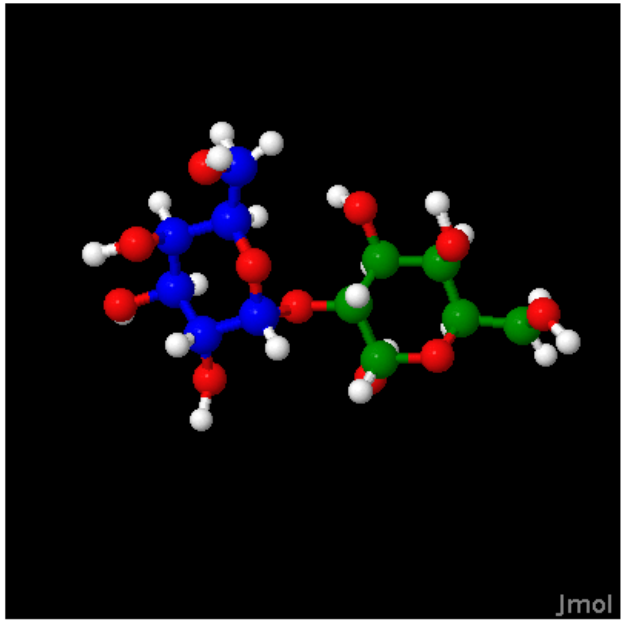
Analyzable atoms:

- (•) Hydrogens only (e.g. for interpretation of NOESY spectra)
- ( ) Hydrogens, Oxygens and Nitrogens (e.g. for H-bond analysis)

Step 2: Select pairs of hydrogen atoms and min/max distances to be mapped:

(Atoms can be picked either by selection from the drop-down menus or by clicking on the atoms in the 3D structure view on the left side)

a-D-Galp-(1-2)-a-D-Galp



Atom labels:

(•) Off ( ) Hydrogens ( ) All atoms

Set Residue Colors | Clear Residue Colors

a-D-Galp (GLA2) No.

Torsions | Weighting

Angle definitions | -60° | +60° | 180°

a-D-Galp (GLA1) No.

Torsions | Weighting

Angle definitions | -60° | +60° | 180°

Min. Max.

Distance (Å)

Add Dataset | Reset

Data sets:

Atoms GLA2.HO2 / GLA1.H3 (5 - 34Å) | Delete | Edit |

Torsion HO2, OH2, C2 , H2 (31,30,29,32), weighting 1,1,1

GlycoMapsDB ID for background map: 7040 **info**

Create Distance Map

GLYCOSCIENCES.DE: TOOLS

pdb-care: PDB CArbohydrate REsidue check

Introduction

A recent study revealed that about 30% of the carbohydrate-containing PDB entries comprise at least one error within the carbohydrate moieties [1].
To reduce this high rate of errors in further entries, *pdb-care* aids experimentalists in detecting discrepancies in connectivities and nomenclature.

A beta-version of a new, improved version of pdb-care is available now. Please consider that this is a beta-version and thus still contains some errors.

Contact: Thomas Lütteke

Enter PDB ID: (⇒Examples)

or select file to upload: Browse...

or insert pdb-file below:

Select Options:

Find carbohydrates in HETATM only

Assign connections by atom distances off

Choose checks to perform:

Connections

☑ **Bond length / valence check**

Tolerance: 20%

☑ **Ignore connections to ions**

Connections between ions and other residues are mostly no covalent bonds but complex bond. Therefore, they should be ignored when checking bond lengths and atom valences.

GLYCOSCIENCES.DE: TOOLS

pdb-care: PDB CArbohydrate REsidue file check - description and examples

General information about this text:

The examples on this page show possible output of the *pdb-care* pdb file check software.

```
For output excerpts, a monospace font is used.
Informations are printed in black, warnings in blue and errors in red.
```

For explanation of the output, the standard font is used.

Output examples:

First of all, *pdb-care* checks the connections of HETATMs given in the CONECT tags of the pdb-file. Bond lengths outside a valid range or atoms whose valence is beyond the maximum number of connections allowed for the respective element are listed (excerpt from the output for PDB entry **1dzg**):

```
Distance check: Connection 3183-3194 (C-O) is 4.9. (expected: 1.1 - 1.7) ###

Atom NAG_801L C2 (max. bonds: 4) is connected to 5 atoms: ASN_96L ND2, NAG_801L C1, NAG_801L C3, NAG_801L N2, NAG_801L O6

Distance check: Connection 3245-3252 (C-N) is 66.6. (expected: 1.2 - 1.7) ###

Distance check: Connection 6558-6566 (C-C) is 38.2. (expected: 1.2 - 1.8) ###

Atom NAG_842I N2 (max. bonds: 4) is connected to 6 atoms: NAG_842I C2, NAG_842I C4, NAG_842I C7, NAG_842I N2, NAG_842I O3, NAG_861I C2
```

Afterwards, the residue names in the HETATM section of the pdb-file are checked for those where carbohydrate residues are to be expected (PDB entry **1gz9**):

```
check pdb residue names for carbohydrate rings to be expected:

   FUC 1559A: monosaccharide ?-?-Fucp

   LAT 1560A: lactose, 2 rings, linucs [][?-D-Glcp]{[(4+1)][b-D-Galp]{}}

Expected number of carbohydrate rings from pdb residues: 3
```

In the next step the HETATMs are searched for potential carbohydrate rings, and the number of detected rings is displayed. For rings lacking an Oxygen or respective atom attached to the anomeric carbon, the software tries to assign connections to atoms in the vicinity of the anomeric carbon. In this case, a warnig message like

```
   Ring 1 (c1: 4368, NAG601A) assigned by c5-atom.
```

is displayed. If potential connections are found, they are listed below in the following form:

```
   Found 1 possible connection(s) for atom 4368:
   atom 729 (ASN92 ND2) dist.-dev. 0.032 score 1.123
```

The first number is the atom number in the pdb-file, followed by residue name, residue number and atom name. The next number shows the deviation of the bond distance from the average value, the last number contains a quality score derived from bond distance and bond angles (the lower, the better).

GLYCOSCIENCES.DE: TOOLS

carp: CArbohydrate Ramachandran Plot

Introduction

The "Ramachandran Plot", where backbone torsion angles are plotted against each other, is a frequently used tool to evaluate the quality of a protein 3D structure. For carbohydrate structures, linkage torsions can be evaluated in a similar way. Preferred Phi/Psi values of the torsion angles of glycosidic bonds depend strongly on the types of monosaccharides involved in the linkage, the kind of linkage (1-3, 1-4, etc) as well as the degree of branching of the structure.

CARP analyses carbohydrate data given in PDB files using the **pdb2linucs** algorithm. For each different linkage type a separate plot is generated. The user can choose between two sources for plot background information for comparison: data obtained from PDB provided by **GlyTorsion** (Fig.1) or from **GlycoMapsDB** (Fig.2). GlycoMapsDB provides calculated conformational maps, which show energetically preferred regions for a specific linkage, while PDB data are based on experimentally solved structures. For seldom occuring linkages, however, PDB data are often rare, so maybe not sufficient background information for comparison will be available from this source.

Contact: Thomas Lütteke

References

Enter PDB ID:

or select file to upload: Browse...

Select data source for plot background:

- (•) PDB (NMR-like definition)
- ( ) PDB (crystallographic definition) ?
- ( ) GlycoMapsDB

Select graphics type:

- (•) GIF
- ( ) SVG ?

Find carbohydrates in HETATM only

Assign connections by atom distances off

Select structure viewer / size:

- (•) Jmol
- ( ) Chime

/ 700 x 500 pixel

Submit Query Reset

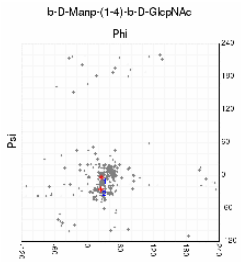


Fig.1: Background data source PDB

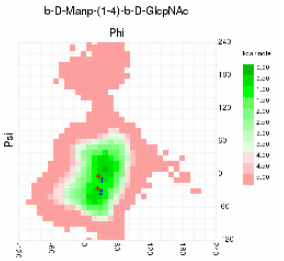
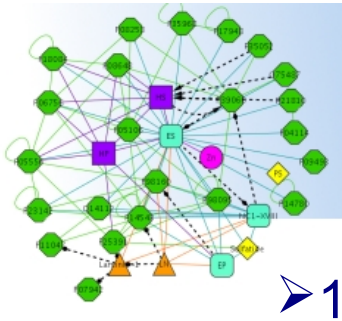


Fig.2: Background data source GlycoMapsDB

http://matrixdb.ibcp.fr

MatrixDB



MatrixDB

Extracellular Matrix interactions DataBase

- 1987 extracellular matrix interactions including 1706 protein-protein
- 111 protein-glycosaminoglycan interactions

TOTAL: **1987** interactions involving **263** extracellular molecules (total: **1032** molecules)
corresponding to **4447** experiments

Interactions by interactor types:

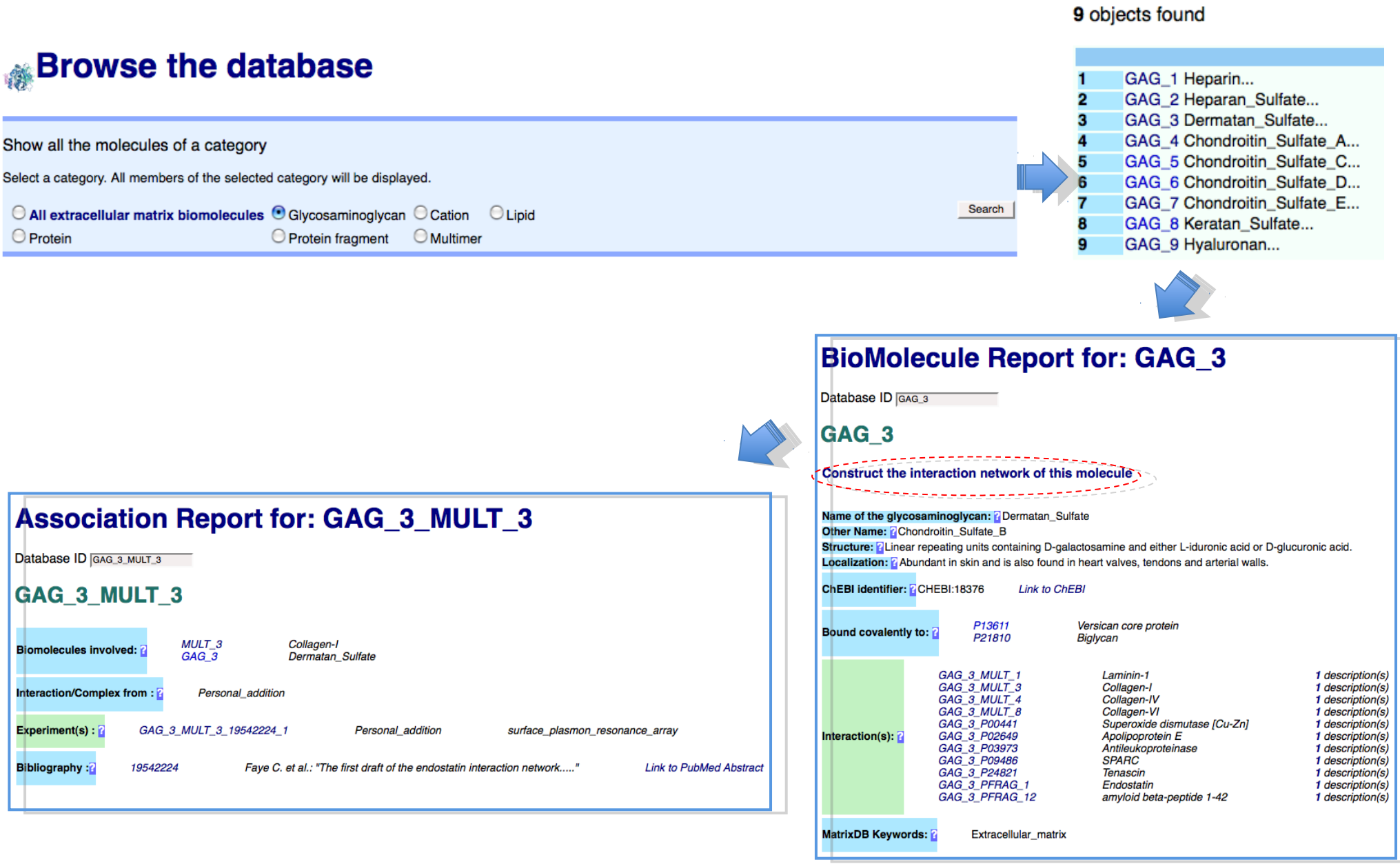
	Interaction(s) between					
	Protein	Glycosaminoglycan	Fragment	Multimer	Lipid	Cation
Protein	1706	111	51	32	3	5
Glycosaminoglycan	111	0	14	30	0	0
Fragment	51	14	4	22	3	2
Multimer	32	30	22	3	1	0
Lipid	3	0	3	1	0	0
Cation	5	0	2	0	0	0
TOTAL	1908	155	96	88	7	7

Sources of interaction data:

	Interaction(s)	Shared interaction(s)						
	TOTAL	MatrixDB SPR	MatrixDB LC	MINT	IntAct	DIP	DIP-IMEx	HPRD	BioGrid
MatrixDB SPR	65	-	15	0	0	0	0	0	0
MatrixDB LC	303	15	-	8	3	1	0	9	17
MINT	211	0	8	-	84	0	0	140	82
IntAct	232	0	3	84	-	4	1	127	96
DIP	46	0	1	0	4	-	14	29	24
DIP-IMEX	14	0	0	0	1	14	-	3	1
HPRD	1378	0	9	140	127	29	3	-	716
BioGrid	839	0	17	82	96	24	1	716	-

MatrixDB SPR: MatrixDB Surface Plasmon Resonance arrays
MatrixDB LC: MatrixDB Literature Curation
MINT: Molecular INTeraction database
IntAct: IntAct database
DIP: Database of Interacting Proteins
DIP-IMEx: DIP interactions following the IMEx guidelines
HPRD: Human Protein Reference Database
BioGRID: Biological General Repository for Interaction Datasets

MatrixDB: BROWSING



MatrixDB: INTERACTOME

Interactome Construction

Select the interaction data you want to represent.

Warning: some queries may take a while, please be patient...

1 - By cellular localization
The cellular localization can also be combined to a keyword
(e.g. Extracellular molecules implicated in Alzheimer disease)

- [x] Extracellular matrix BioMolecules
- [ ] Secreted BioMolecules
- [ ] Membrane BioMolecules

2 - By UniProtKB/Swiss-Prot keyword(s)
You can select one or more keyword(s) in this list.
All relevant proteins will be displayed with their partners.

Osteogenesis
Proteoglycan
Stickler syndrome
Tissue remodeling

BUILD THE INTERACTOME | CLEAR

3 - Partners of BioMolecule(s)
Enter the EXACT identifier of the BioMolecule(s):
- *UniProtKB Accession Number* for proteins (P39060),
- *MatrixDB identifier* for other biomolecules (PFRAG_1, GAG_2, MULT_3, LIP_1, CAT_2),
separated by a ",".

GAG_3

(•) 1- Direct partners of the BioMolecule(s)
( ) 2- Partners of the direct partners

Hint: identifiers can be retrieved using the BioMolecule Search.

BUILD THE INTERACTOME | CLEAR

4 - Tissue specific interactomes
Construction of tissue specific interaction networks using mRNA expression data (EST profiles) stored in UniGene. EST (Expressed Sequence Tag) profiles are approximate gene expression inferred from EST count and the cDNA library source. You can modify the EST profile threshold for a tissue (1: low, 1000: high selection).
Warning: if the threshold is too high or if the molecule is not present in the tissue, no interaction network will be constructed.

A/ Threshold:
(TPM: Transcripts per million)

- (•) ≥1 TPM (*low*)
- ( ) ≥10 TPM
- ( ) ≥100 TPM
- ( ) ≥1000 TPM (*high*)

B/ You can select one or more tissue(s) in this list.
All relevant proteins will be displayed with their partners.

adipose_tissue
adrenal_gland
ascites
bladder
blood
bone
bone_marrow

C/ Enter the EXACT identifier of the BioMolecules:
- *UniProtKB Accession Number* for proteins (P39060),
- *MatrixDB identifier* for other biomolecules (PFRAG_1, GAG_2, MULT_3, LIP_1, CAT_2),
separated by a ",".

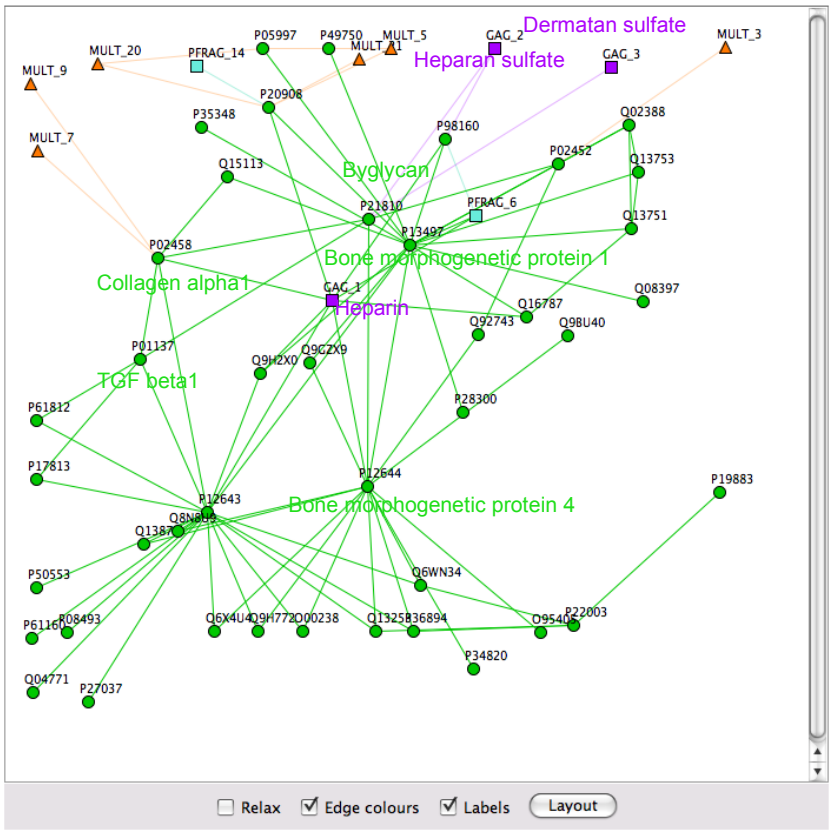
GAG_3

(•) 1- Direct partners of the BioMolecule(s)
( ) 2- Partners of the direct partners

Hint: identifiers can be retrieved using the Search page.

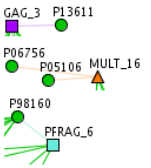
BUILD THE INTERACTOME | CLEAR

Interaction network of GAG_3, level 1 for tissue(s) bone (EST threshold 1).

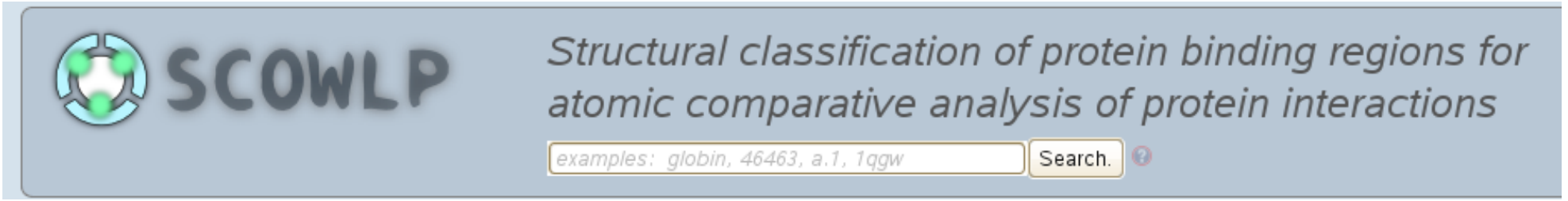


Protein
Protein fragment
Glycosaminoglycan
Multimer
Cation
Lipid

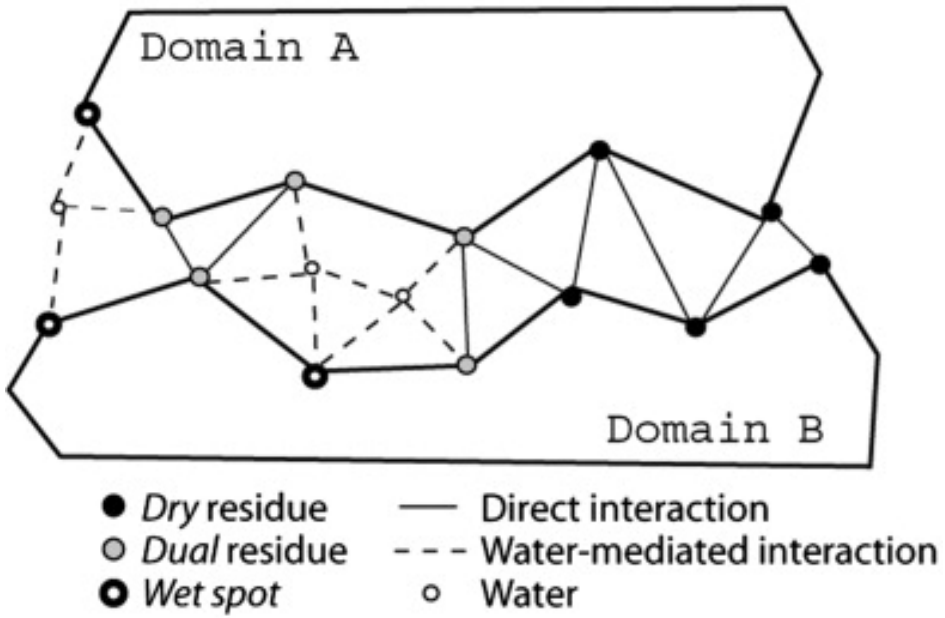
Glycosaminoglycan covalently linked to a protein (proteoglycan)
Monomer-Multimer
Fragment-Parent protein



SACCHARIDES IN THE PDB: SCOWLP



www.scowlp.org



Interactions definition:

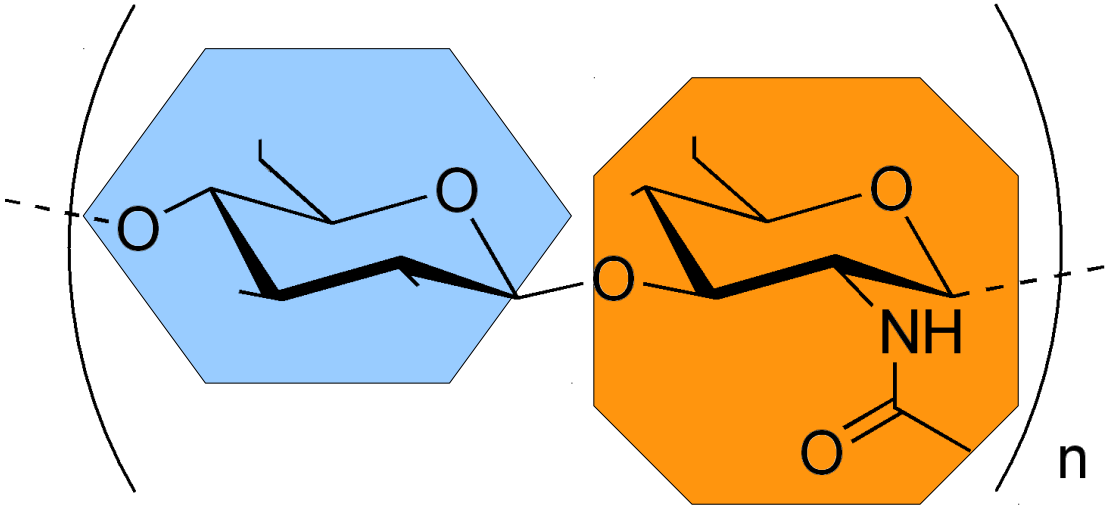
- **H-bond: 3.2 Å**
- **Salt bridge: 4.0 Å**
- **VDW: $R_{1\ VDW}+R_{2\ VDW}$**

- **PDB: 110790 structures (4.08.2015)**
- **~2000 protein-saccharide interfaces**

SACCHARIDES IN THE PDB: HYDRATION

Interfaces dataset	Number of interfaces	Water molecules/interface area (1/1000 Å^2)
GAGs-protein	57	10.8
Sugar-protein (not GAGs)	1910	9.5
Protein-protein	176	3

GAGs:



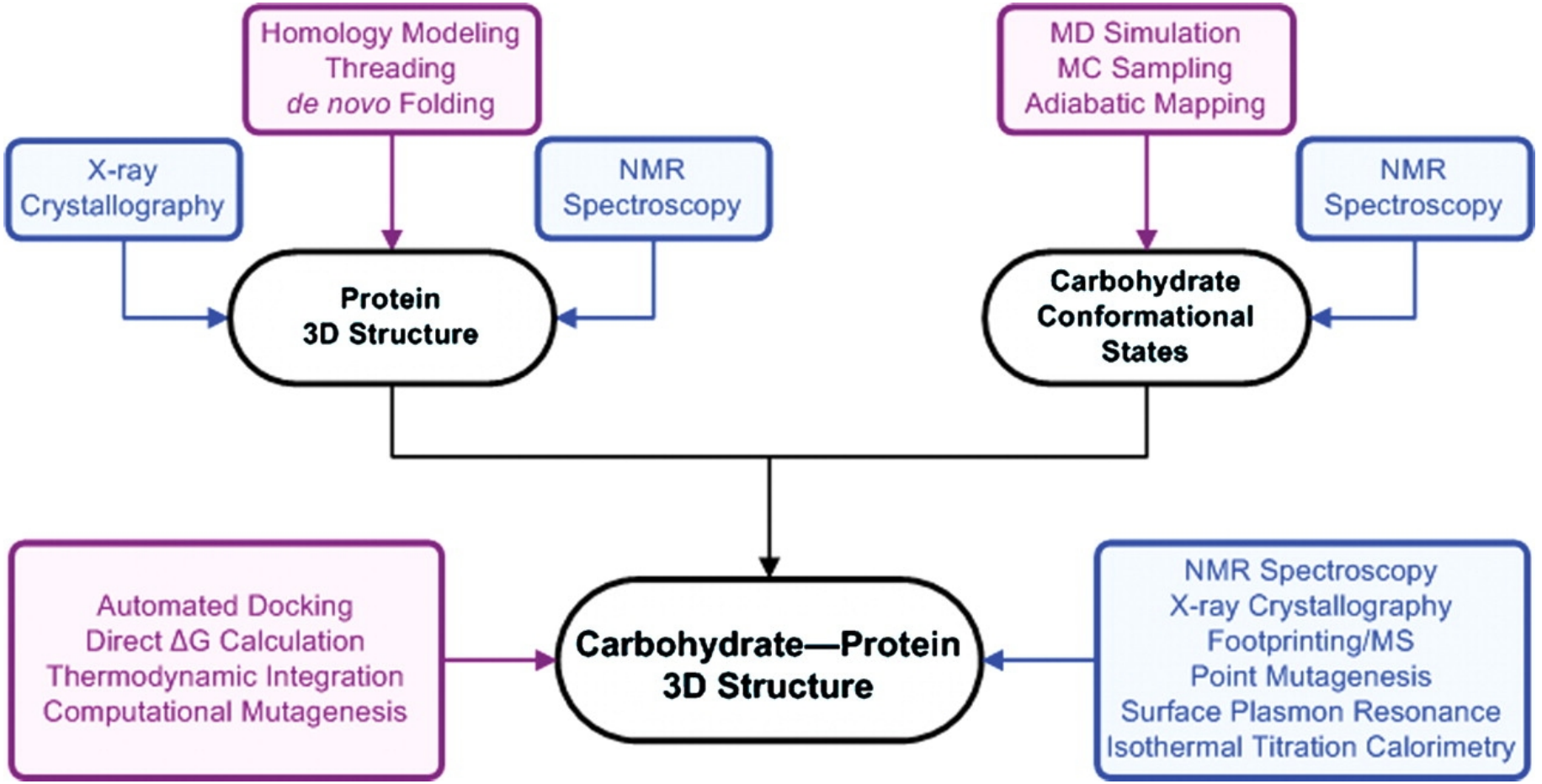
Hexose/Hexuronic acid:

- GlcU
- IdoU
- Gal
- Sulfated derivatives

Hexosamine:

- GlcNAc
- GalNAc
- Sulfated derivatives

THEORETICAL AND EXPERIMENTAL METHODS FOR SACCHARIDES ANALYSIS



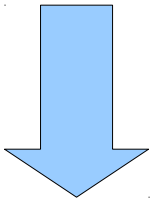
De Marco, Woods; *Glycobiology* 2008.

COMPUTATIONAL CHALLENGES

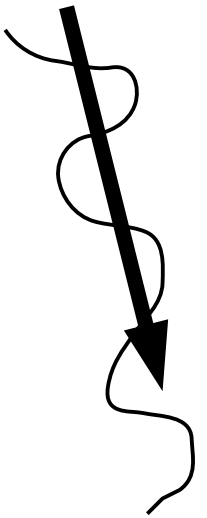
- Experimental data quality

- Variety of carbohydrates (f.i. tetramer):
  - DNA: $4^4 = 256 \sim 10^2$
  - Proteins: $20^4 \sim 10^5$
  - Carbohydrates: $\sim 10^{12}$

- Huge conformational space/flexibility
  - size challenge
  - time challenge



- "Average parameters" problem

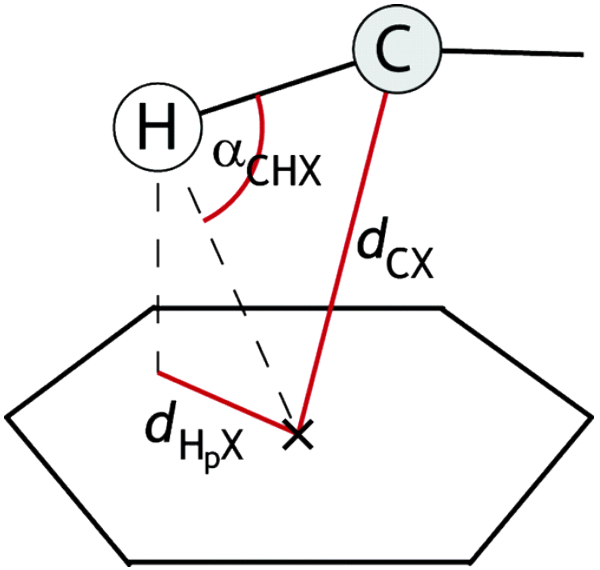


- Induced fit while binding other molecules

- The role of solvent, entropic component:
  - high solubility (-OH groups)
  - anisotropic solvent properties near saccharides
  - min ASA algorithms challenges

SACCHARIDE-PROTEIN INTERACTIONS

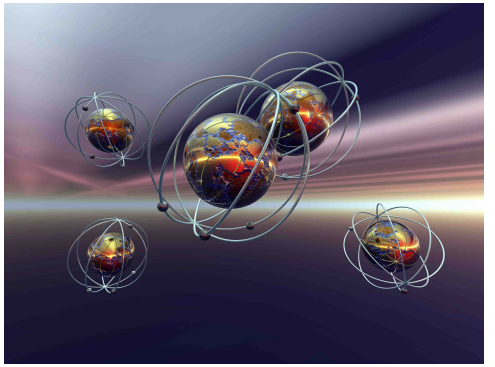
- Bind to receptor proteins, antibodies, lectins, enzymes
- Relatively low affinities (up to ~μM)
- Importance of polyvalence
- Contributions to free energy:
  - Hydrophobic effect: classical and non-classical, ~25-100% of enthalpy
  - CH/π interactions
  - Hydrogen bonds (also water-mediated)
  - Electrostatics+vdW
  - Solvation/Desolvation
- Not biased to cavities binding



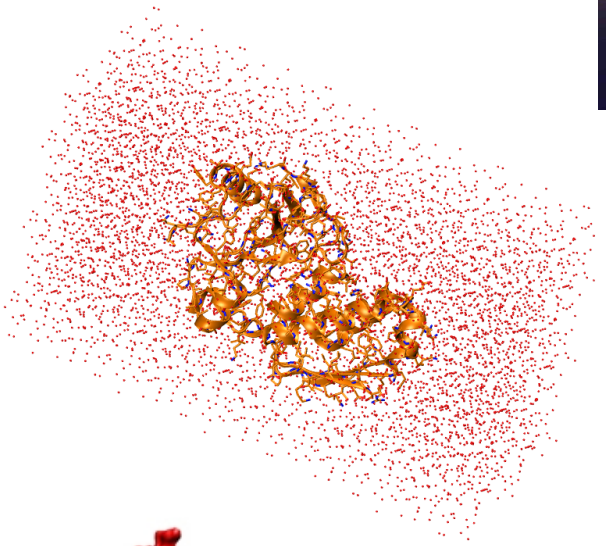
Kerzmann et al; *J Chem Inf Model* 2006.

COMPUTATIONAL METHODS FOR SACCHARIDES ANALYSIS

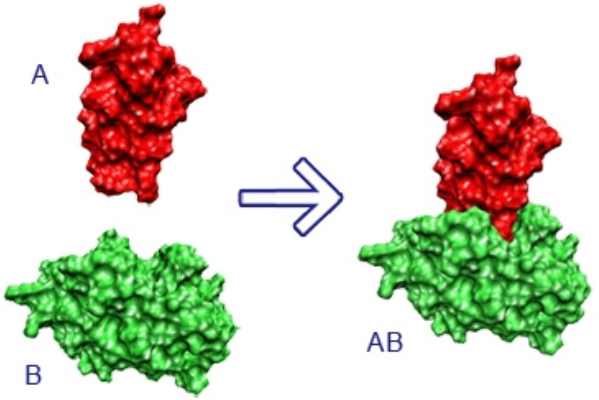
- QM methods



- MD

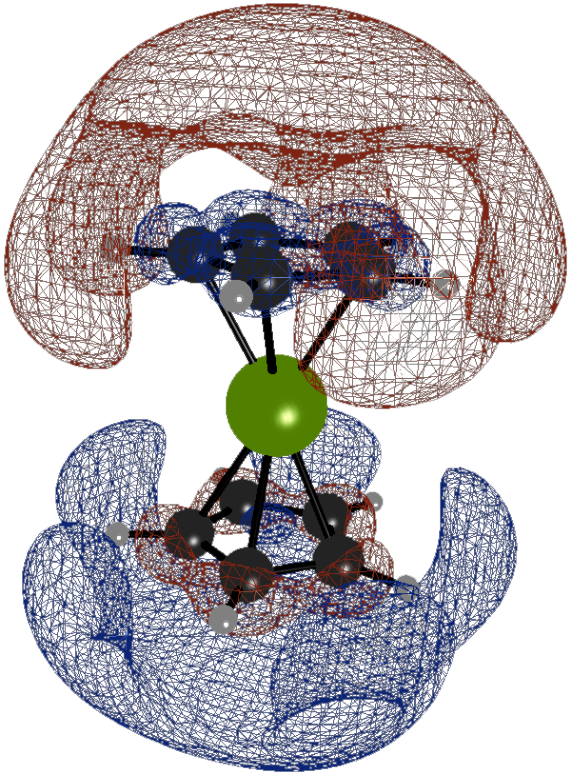


- Docking



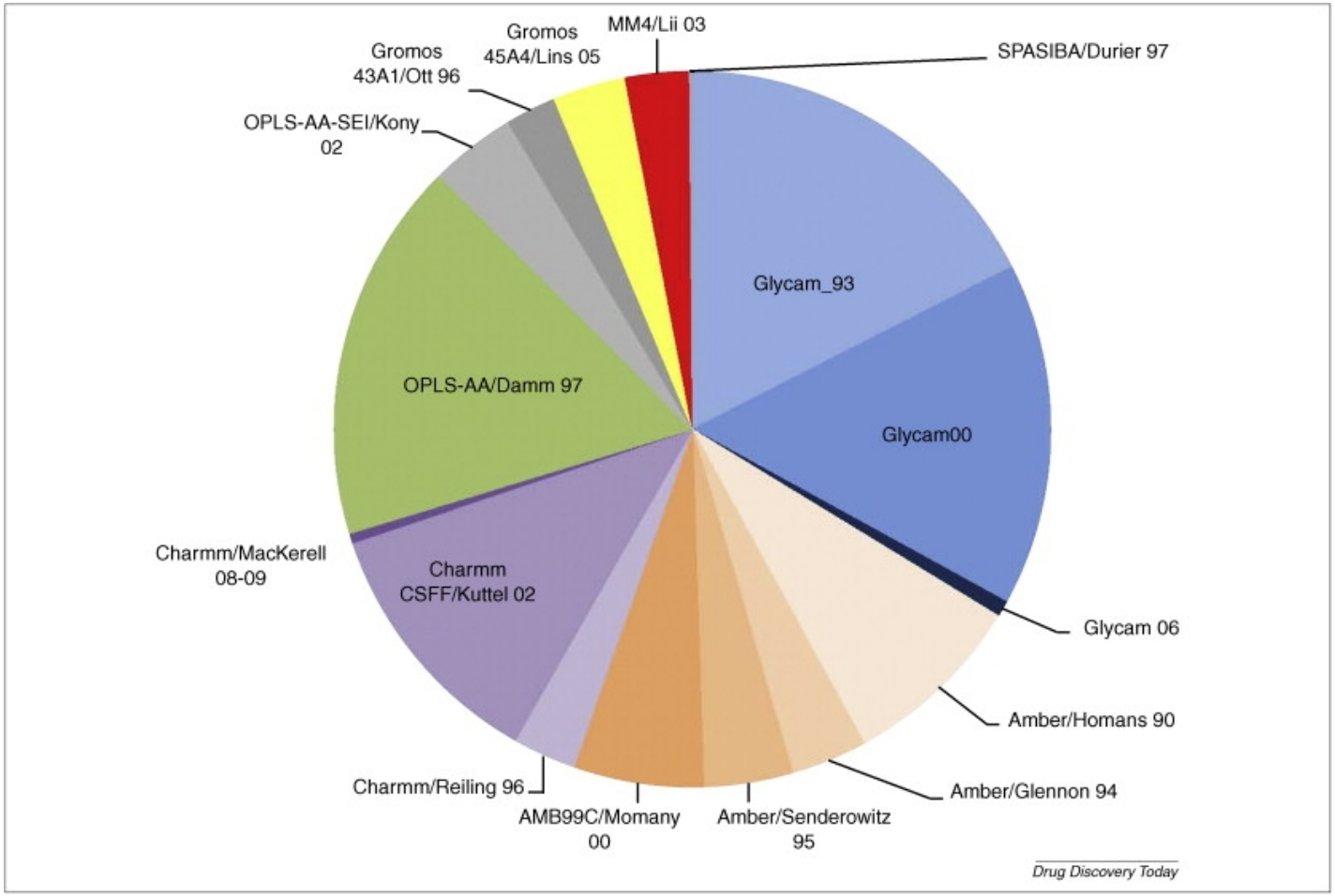
QM OF SACCHARIDES

- Maximum size of system: di-,trisaccharides
- Limitations of explicit solvent introduction
- Counterions are needed but it is challenge
- Conformational analysis with HF or DFT
- NMR parameters
- Reactions studies (ONIOM, MM/QM)
- Need for specific functionals for saccharides



MD OF SACCHARIDES

- Relatively large systems
- Force fields for saccharides: compatibility with proteinic force fields
- Time-scales challenge for flexibility challenge: f.i. rings conformations
- Complementary to NMR-studies:
  - Chemical shifts
  - Karplus equation
- Glycosylation issue
- Free energy:
  - MM-PBSA
  - FEP
  - Steered dynamics



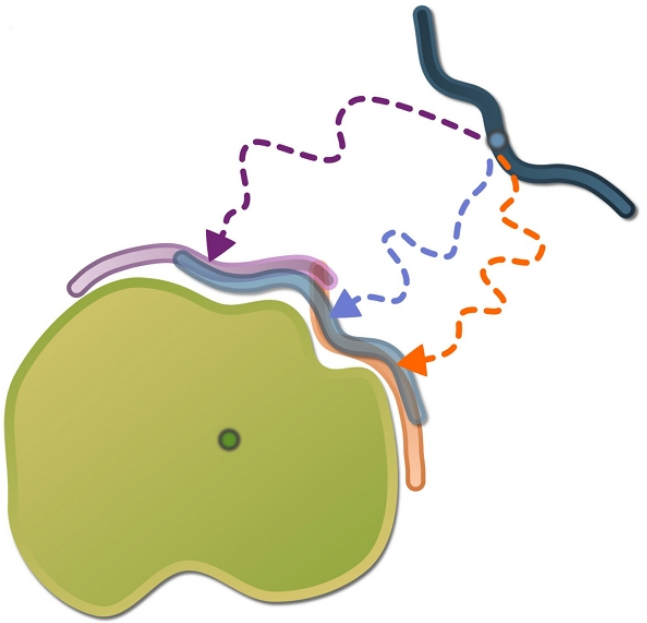
Fadda, Woods; *Drug Discovery Today* 2010.

DOCKING OF SACCHARIDES

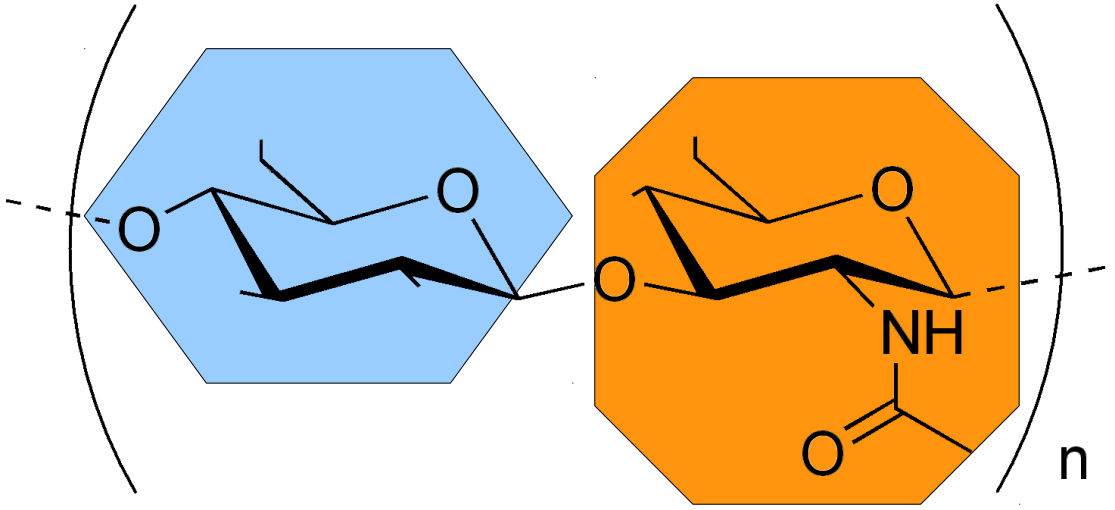
- Most of programs are not tuned for saccharides, docking alone fails
- Size limitations: tetrasaccharides is often the border
- Incremental reconstruction algorithms fail in general (DOCK, FlexX)
- Addition of explicit solvent helps
- BALLDock SLICK:
  - 2 scoring functions
  - CH/π plays crucial role
  - Electrostatics+vdW from GLYCAM ff
  - Desolvation is explicitly taken into account

CASE STUDY

Dynamic Molecular Docking (DMD): a new approach to treat flexibility and explicit solvent in docking of protein-glycosaminoglycan systems

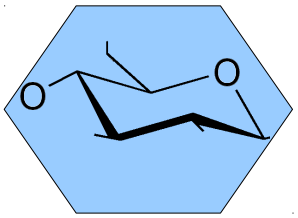


GLYCOSAMINOGLYCANS (GAGs)



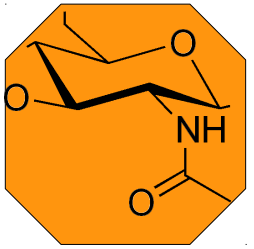
GAGs:

- Hyaluronan
- Chondroitin sulfate
- Heparin
- Heparan sulfate
- Keratan sulfate
- Dermatan sulfate



Hexose/Hexuronic acid:

- GlcU
- IdoU
- Gal
- Sulfated derivatives

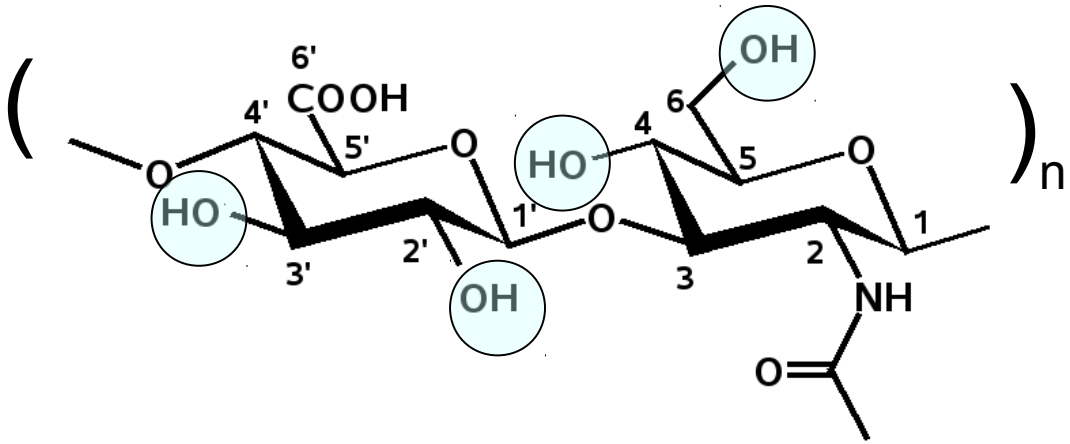


Hexosamine:

- GlcNAc
- GalNAc
- Sulfated derivatives

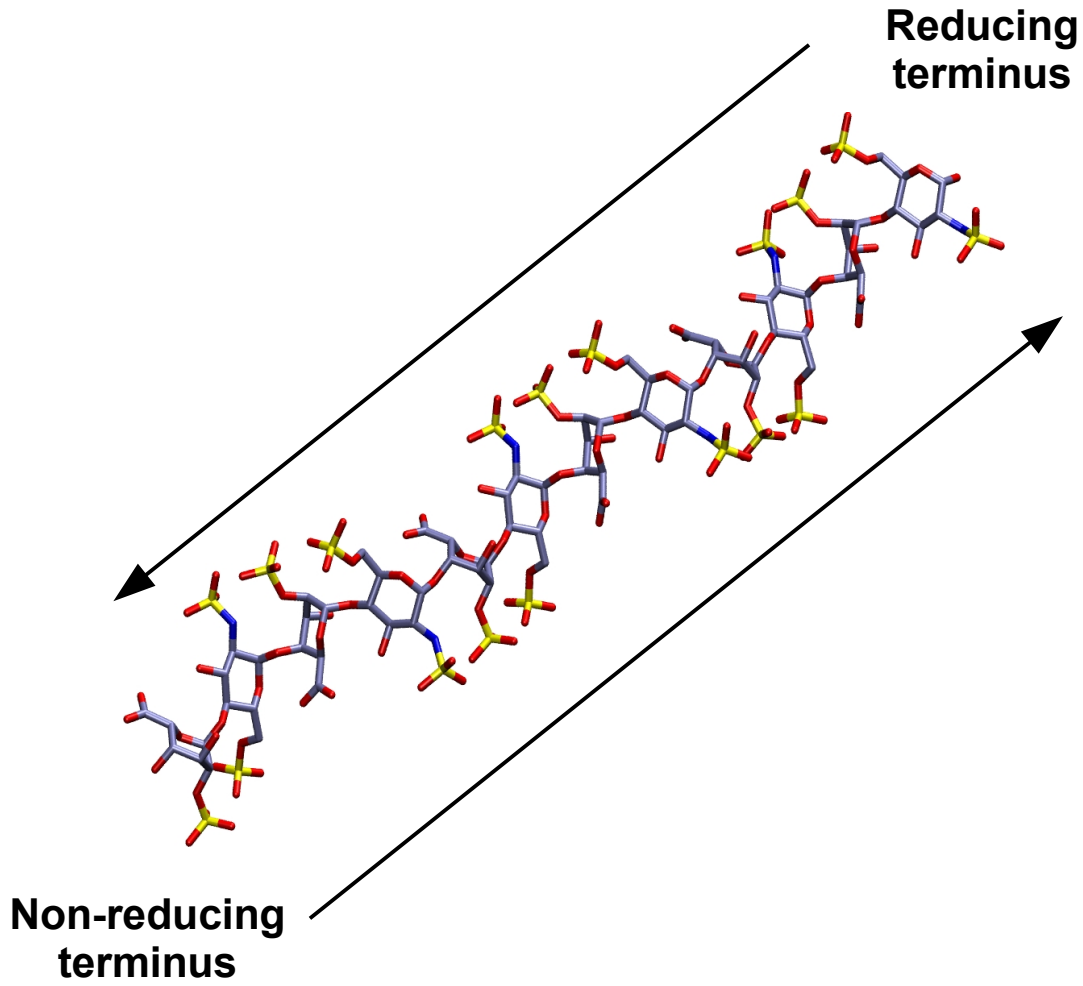
GAGs: COMPUTATIONAL CHALLENGES

- **Sulfation pattern**
- **Symmetry**
- **Electrostatics and solvent**
- **Flexibility**
- **Binding not in cavities**



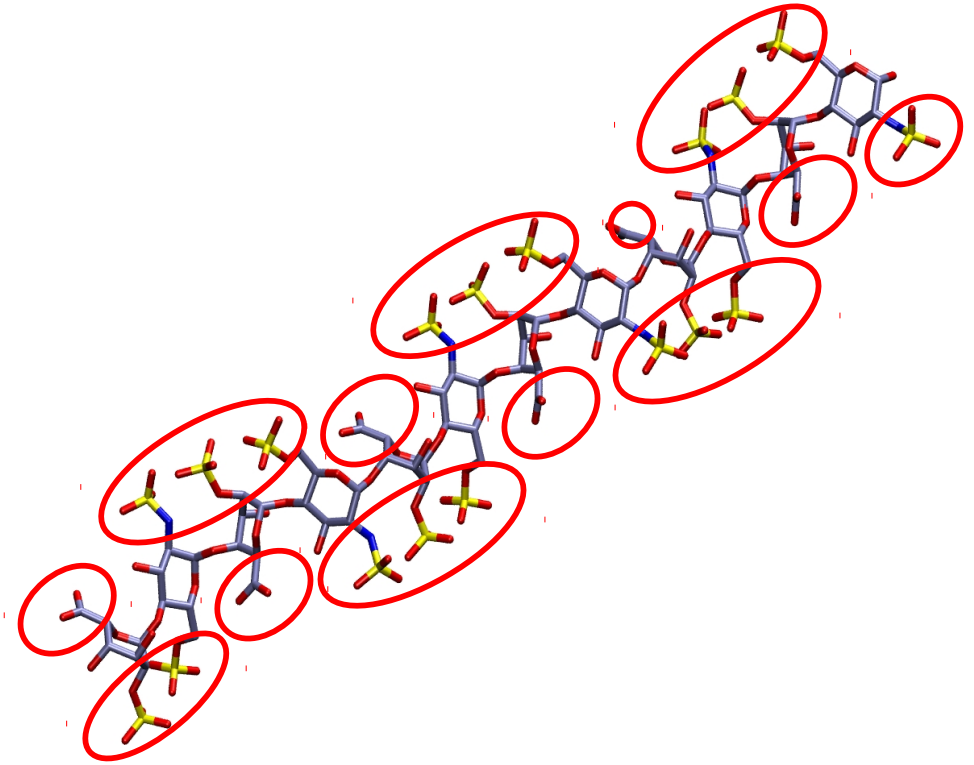
GAGs: COMPUTATIONAL CHALLENGES

- Sulfation pattern
- Symmetry
- Electrostatics and solvent
- Flexibility
- Binding not in cavities



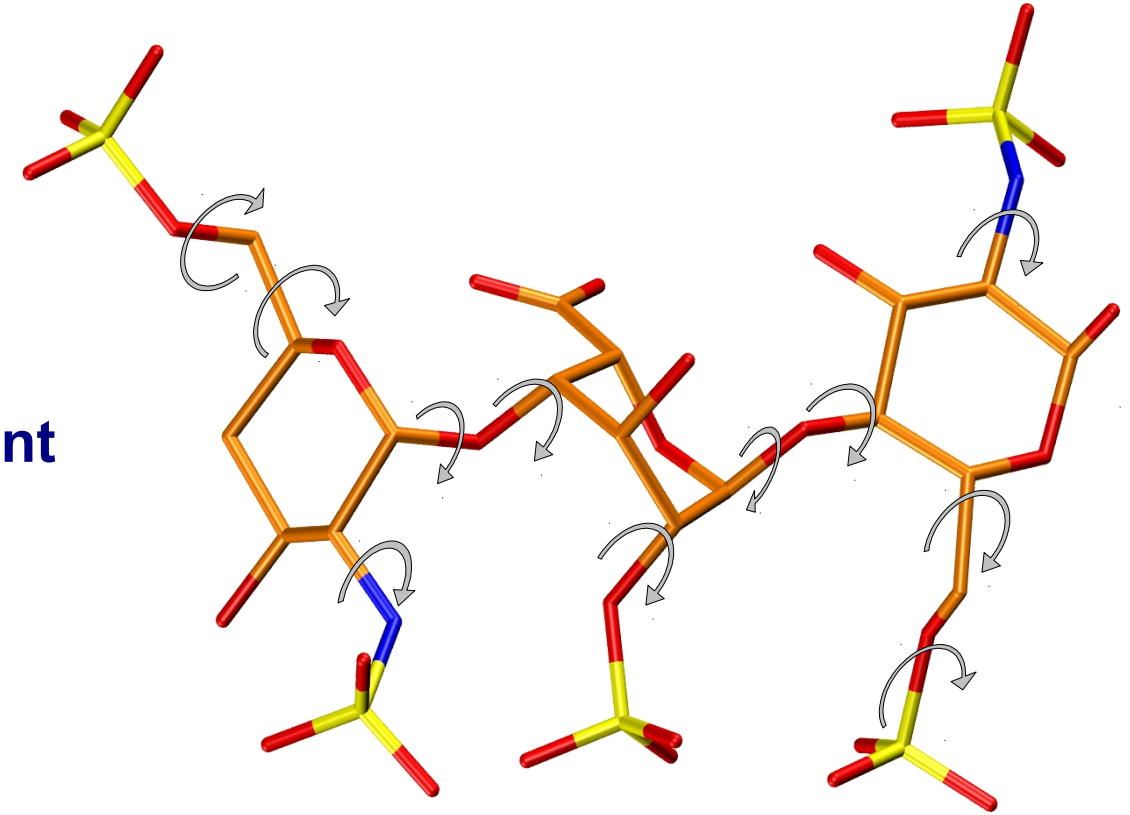
GAGs: COMPUTATIONAL CHALLENGES

- Sulfation pattern
- Symmetry
- Electrostatics and solvent
- Flexibility
- Binding not in cavities



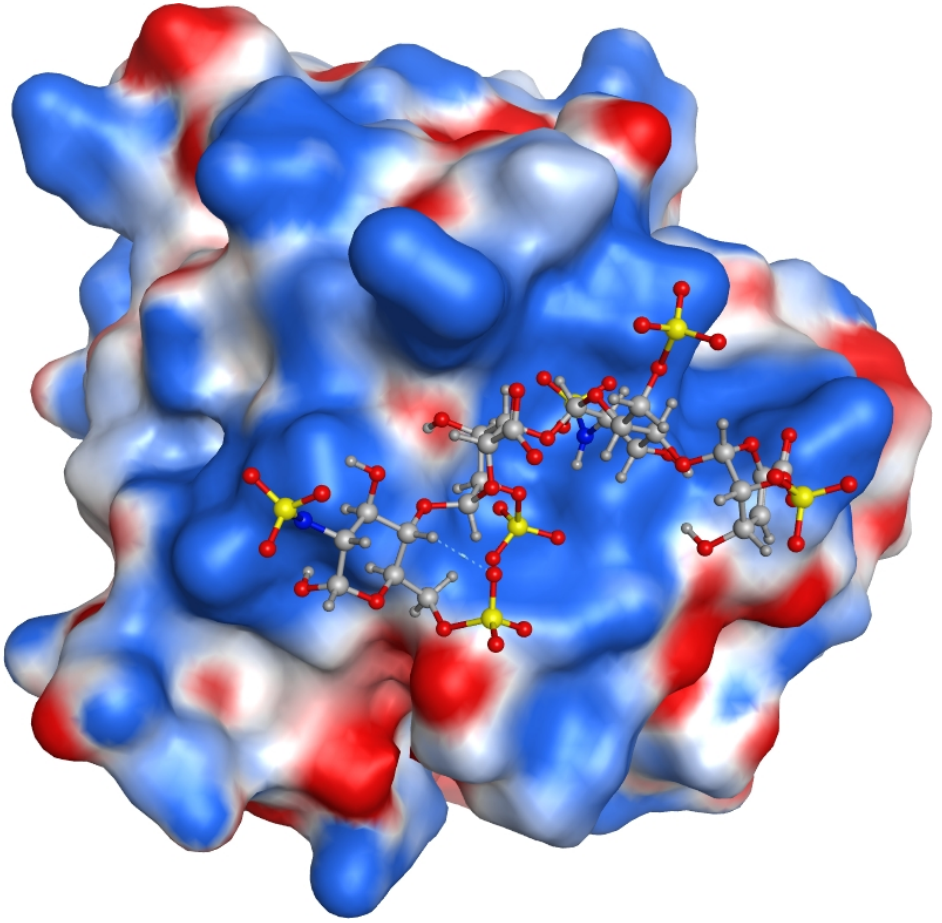
GAGs: COMPUTATIONAL CHALLENGES

- Sulfation pattern
- Symmetry
- Electrostatics and solvent
- Flexibility
- Binding not in cavities



GAGs: COMPUTATIONAL CHALLENGES

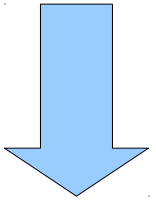
- Sulfation pattern
- Symmetry
- Electrostatics and solvent
- Flexibility



- Binding not in cavities but to positively charged protein patches

MOTIVATION AND GOAL

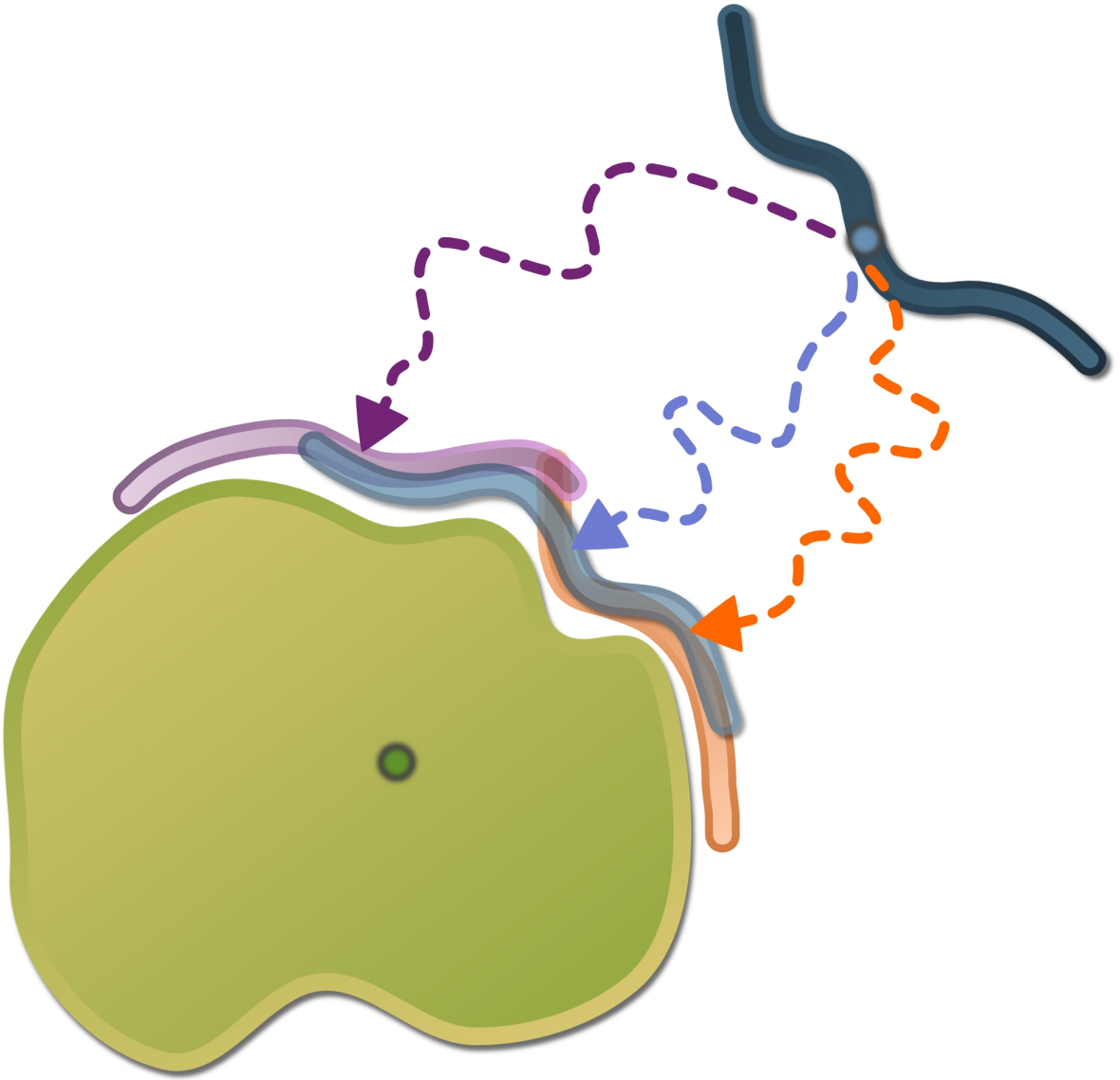
There are no specialized docking tools for highly flexible electrostatics-driven molecular systems such as protein-GAG systems.



The goal is to develop a docking approach, which considers:

- **Receptor and ligand flexibility**
- **Explicit solvent**

THE CONCEPT OF DMD



THE CONCEPT OF DMD

- **Local docking**
- **MD-based**
- **Predicting anchoring residues**

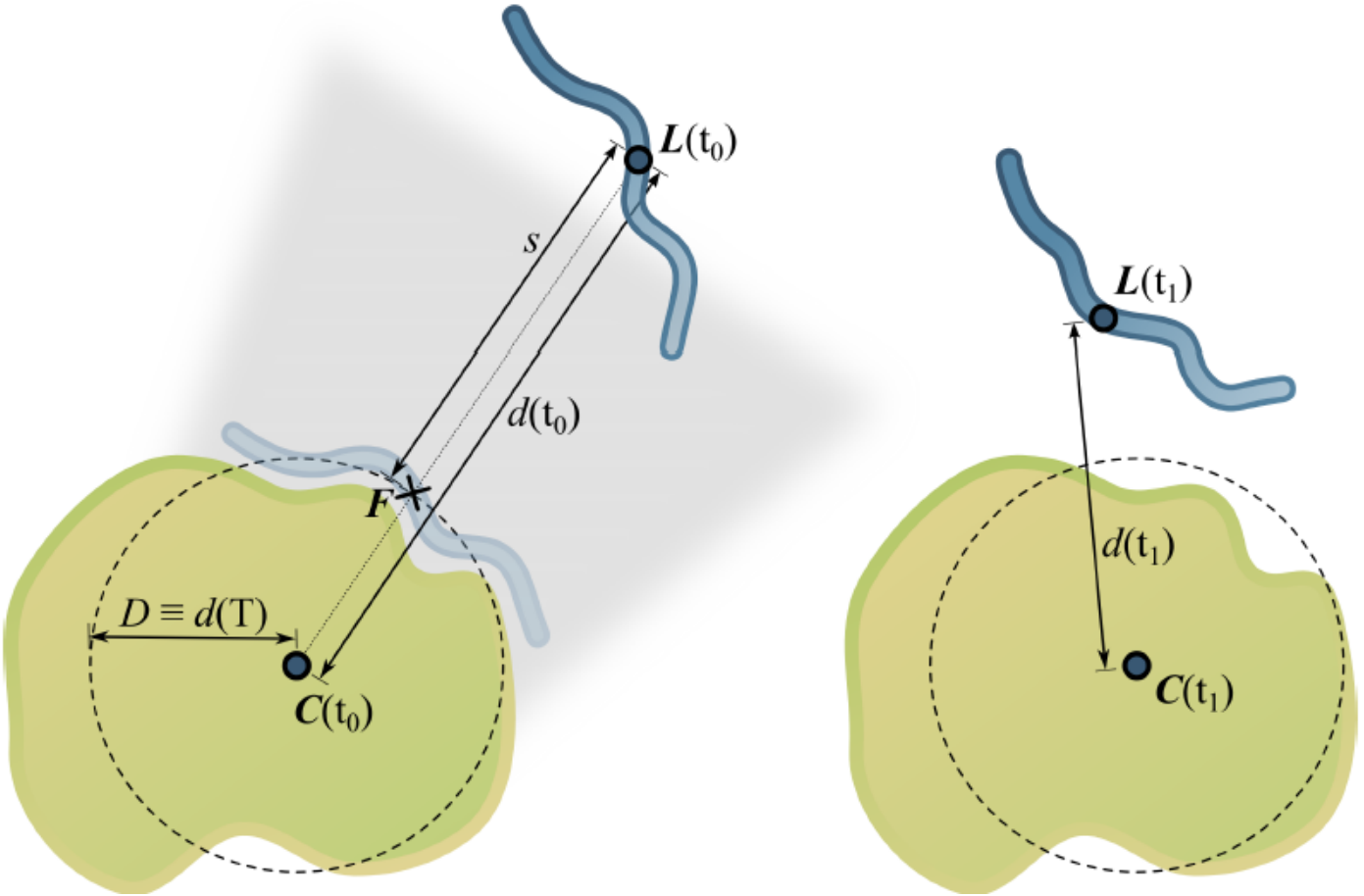
F: focus point

C: core atom

$$d(T) \equiv D \equiv \| \vec{F} - \vec{C} \|$$

$$L(\vec{t_0}) = \vec{F} + s\frac{\vec{F}-\vec{C}}{D}$$

$$d(t) = \| L(\vec{t}) - C(\vec{t}) \|$$



DMD PROTOCOL

I. tMD step

- $U(t) = \frac{1}{2}\ k(d(t) - d(t_0) + vt)^2$
- $k = 200$ kcal mol^{-1}Å^{-1}
- $s \sim 30$ Å
- $T = 4$ ns
- $v = \frac{s}{T} \sim 0.01$ Å/ps
- ff99SB + Glycam06g
- NTP ensemble
- TIP3P water model

II. Free MD step

- 10 ns
- Scoring: MM-PBSA analysis of last 200 ps

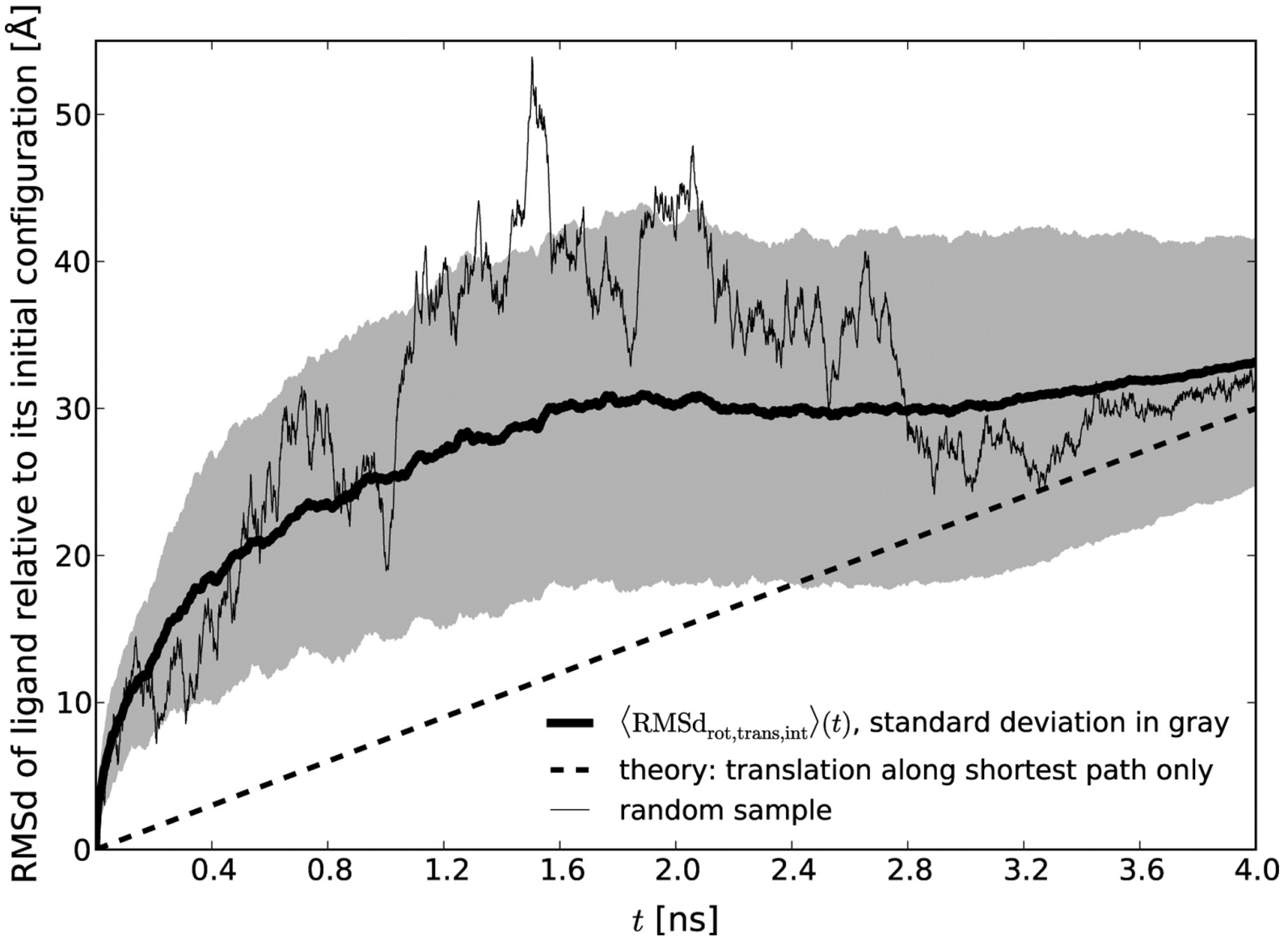
x 100 times

TEST DATASET

- 5 protein-GAG
- 1 protein-peptide
- 1 protein-small molecule

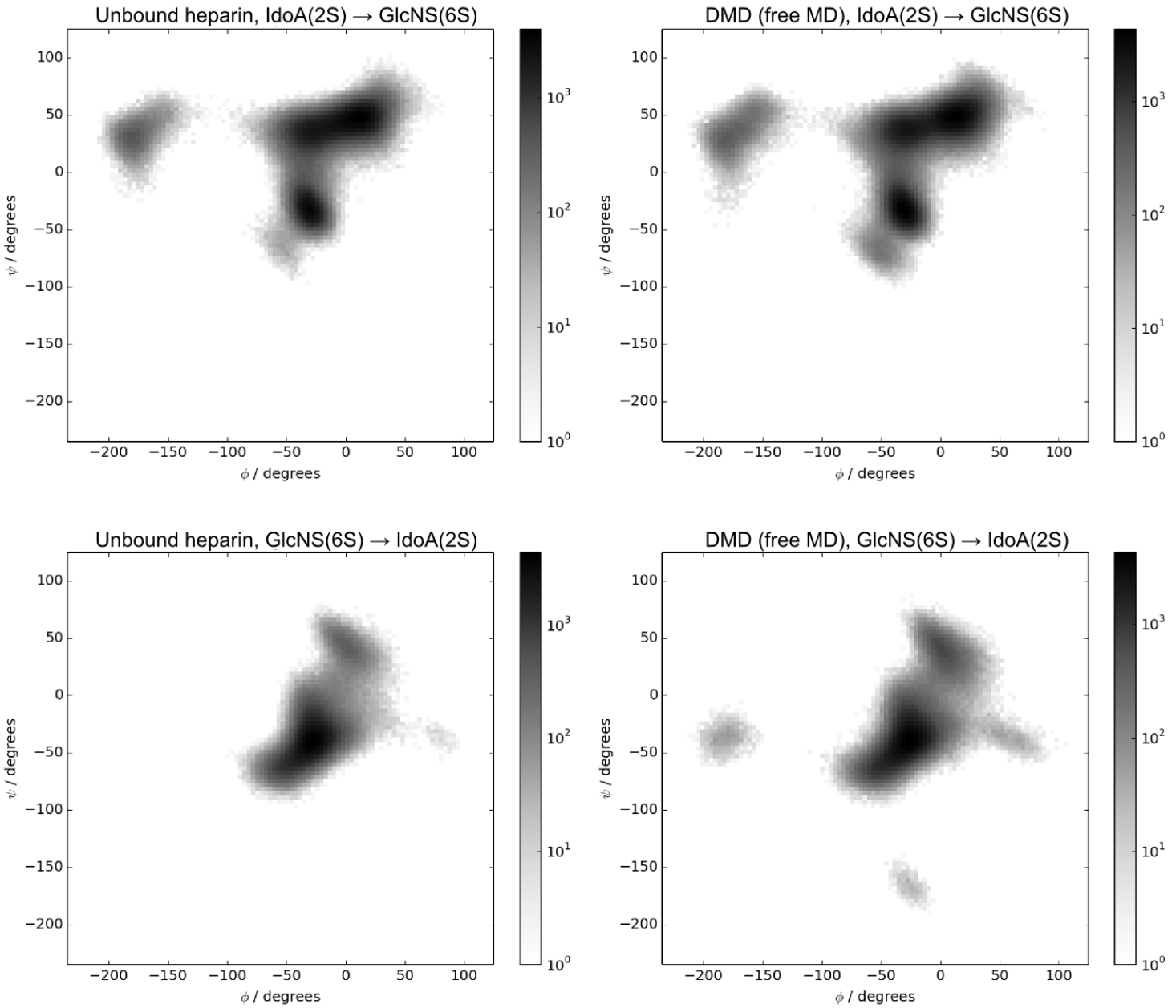
Complex	Ligand length, dp	PDB ID	Res, Å
SDF-1 – HP	2	2NWG	2.1
CathK – CS4	6	3C9E	1.8
CathKmut – CS4	6	3H7D	2.2
FGF2 – HP	4	1BFB	1.9
CD44 – HA	7	2JCQ	NMR
SH3 – p41	11	1BBZ	1.7
Trypsin – Inh.	-	DINGO dataset	-

LIGAND'S DEGREE OF FREEDOM IN DMD



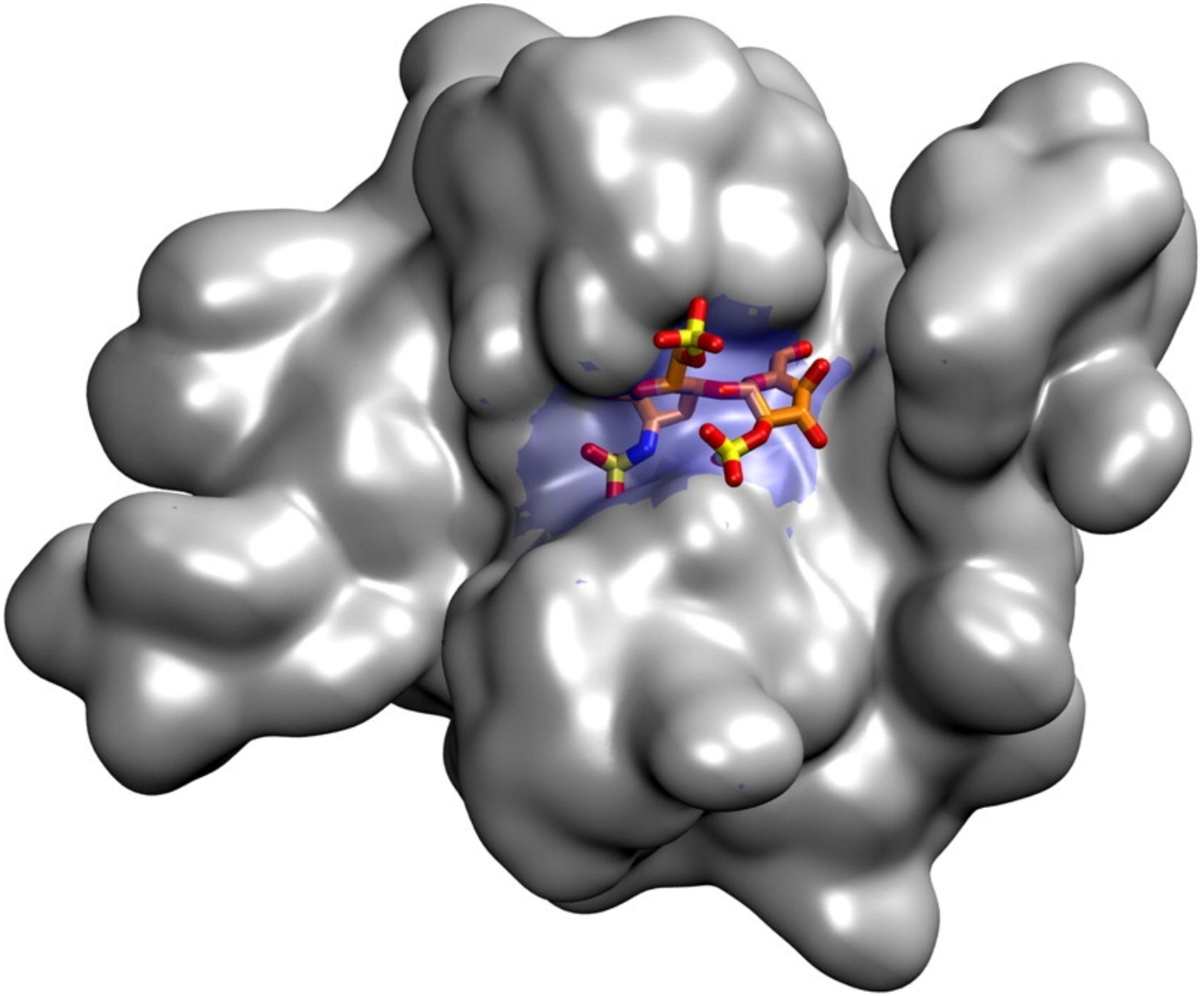
Fair sampling of ligand's degrees of freedom.

GLYCOSIDIC LINKAGE SAMPLING



Fair sampling of glycosidic linkage conformational space.

MEP FOR BINDING SITE DEFINITION

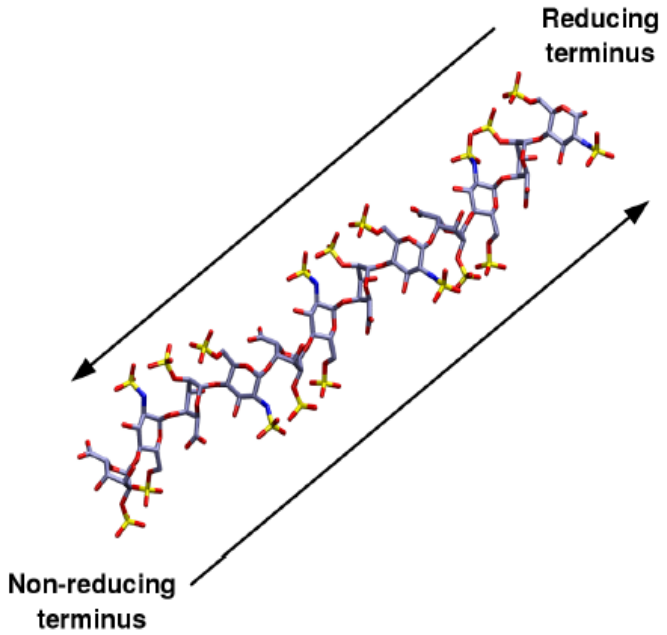


EVOLUATION OF DMD

I. Distance metric:

$$RMSd = \sqrt{\frac{\sum_{i}^{N1}(R_i - R_{ref})^2}{\sum_i m_i}} \; ; \; i\text{: atomic ID}$$

$$\delta \equiv RMSatd = \sqrt{\frac{\sum_{i}^{N2}(R_i - R_{ref})^2}{\sum_i m_i}} \; ; \; i\text{: atomic type}$$



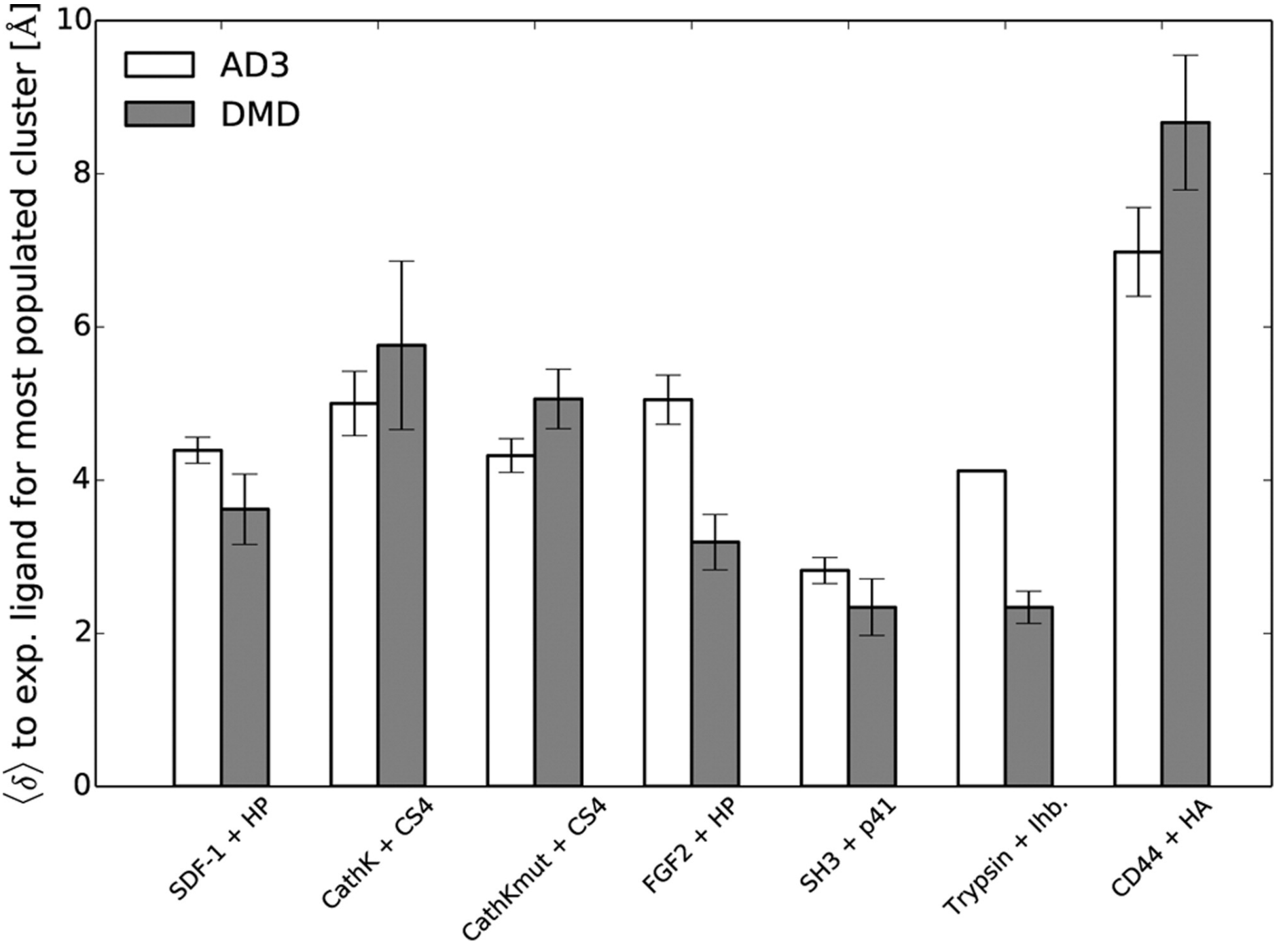
δ accounts for:

- Periodicity
- "Symmetry"

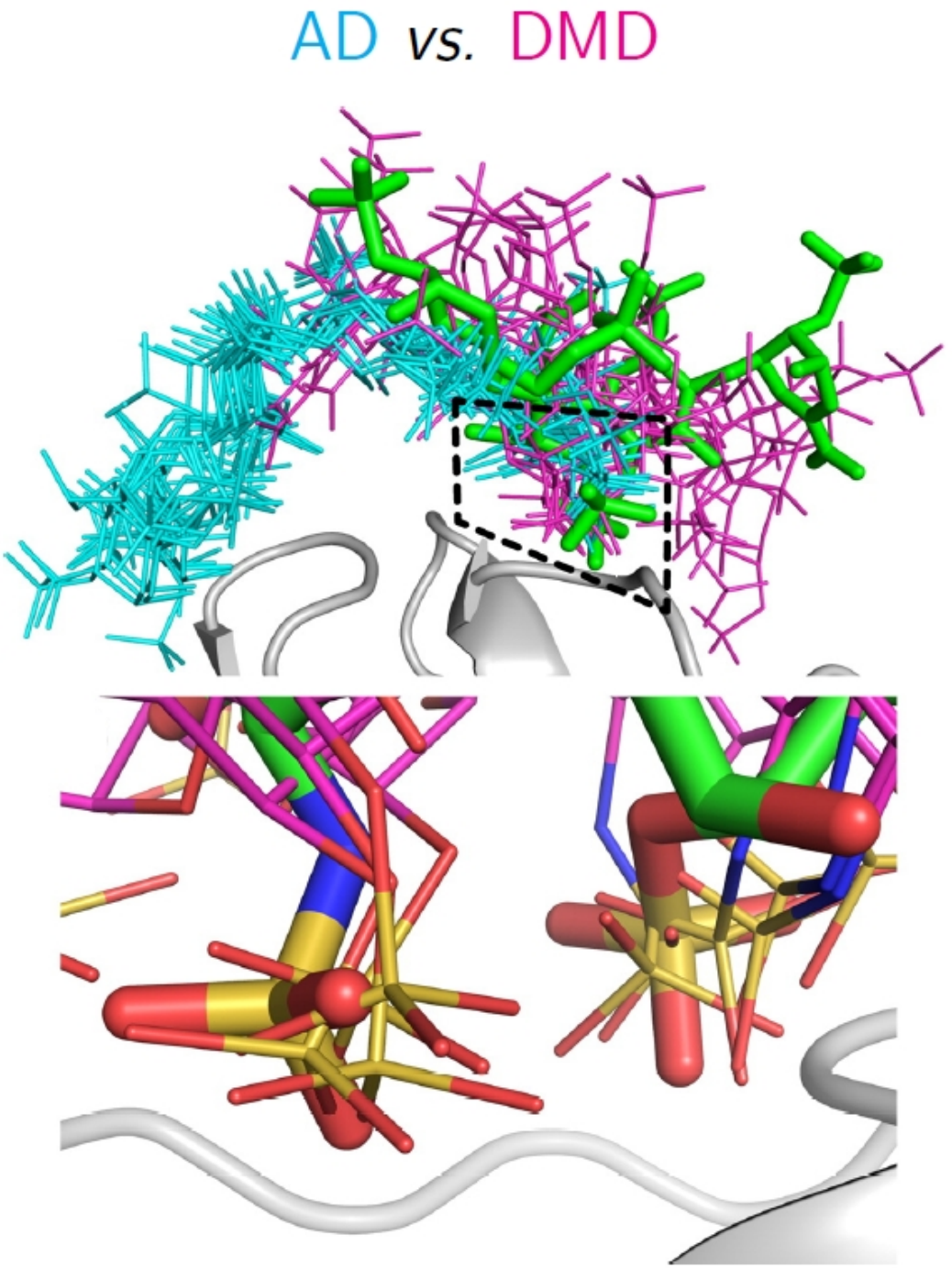
II. Clustering: density-based spatial clustering of applications with noise (DBSCAN)

III. Comparison with Autodock

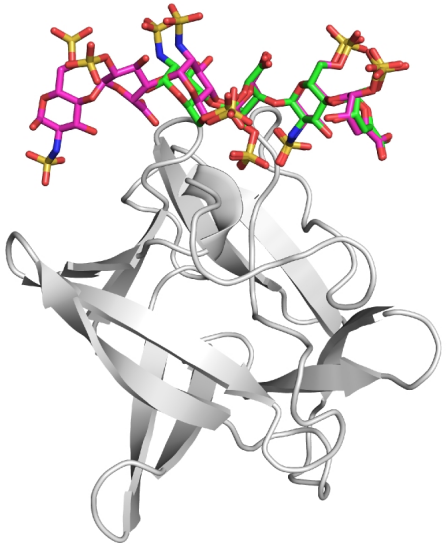
COMPARISON WITH AUTODOCK

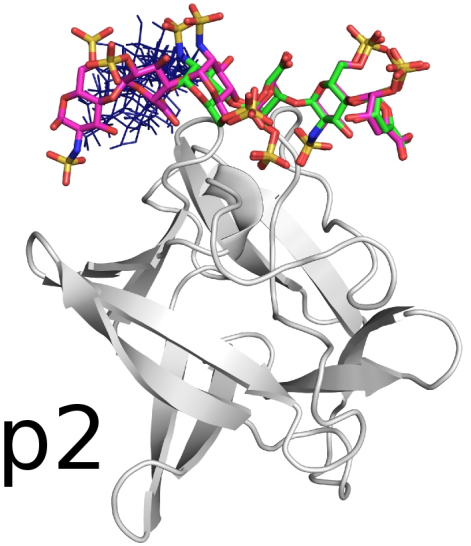


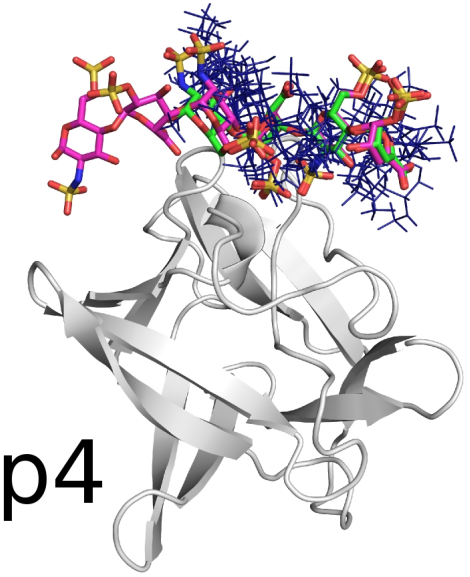
COMPARISON WITH AUTODOCK

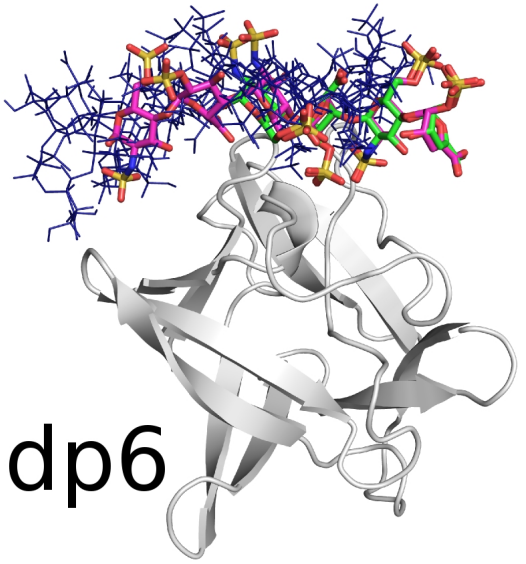


INCREASING GAG'S LENGTH IN DOCKING









Calculated free energies decrease with GAG's length

DEFINING ANCHORING RESIDUES

10 best ranked residues (MM-GBSA)

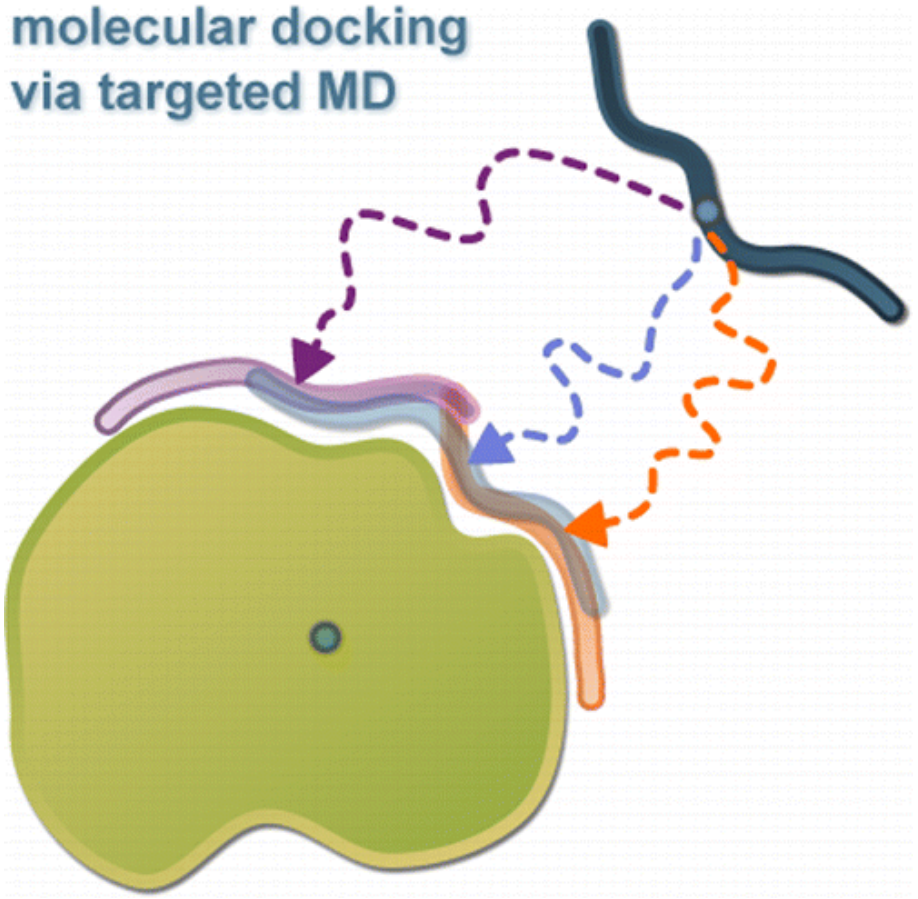
DMD ensemble *vs.* MD of experimental structure

Complex	N_{res}	N_{+}	$N_{neutral,pol}$	$R_{Spearman}$
SDF-1 – HP	7 of 10	7	0	0.52
CathK – CS4	6 of 10	4	2	-0.21
CathKmut – CS4	6 of 10	1	1	-0.12
FGF2 – HP	9 of 10	6	2	0.84
SH3 – p41	2 of 5	0	0	-
Trypsin – Inh.	5 of 10	2	2	0.73
CD44 – HA	1 of 7	1	0	-

DMD correctly indentifies anchoring resides for the systems dominated by electrostatics.

CASE STUDY CONCLUSIONS

We have developed and characterized DMD, an MD-based protocol for local docking for highly flexible electrostatics-driven systems.



LECTURE 5:

GLYCOBIOINFORMATICS/COMPUTATIONAL BIOLOGY OF SACCHARIDES

- Basics of saccharides chemistry, nomenclature
- Saccharides related open databanks
- Saccharides in the PDB
- Challenges in computational analysis of saccharides
- Tools for saccharides analysis: MD, docking
- Case study: docking glycosaminoglycans with Dynamic Molecular Docking