Protein X-ray crystallography and glycobiology

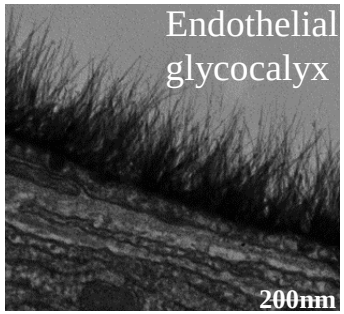
Dr Annabelle Varrot

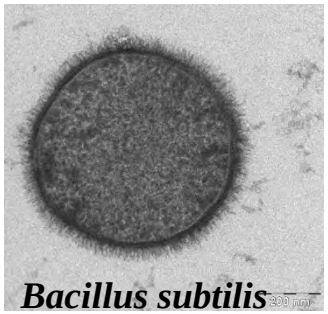
« Structural Glycosciences » Summer School

IBS, Grenoble 7th June 2023

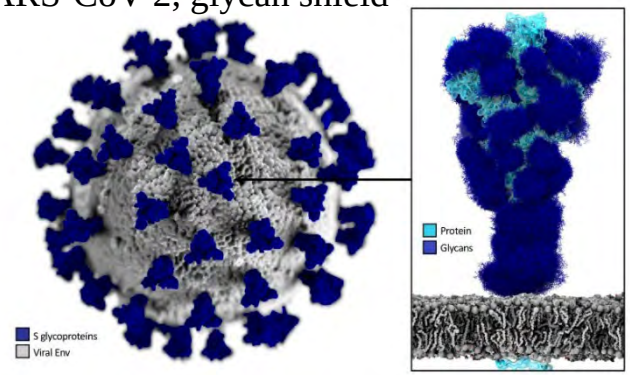
"Sweet side" of cells

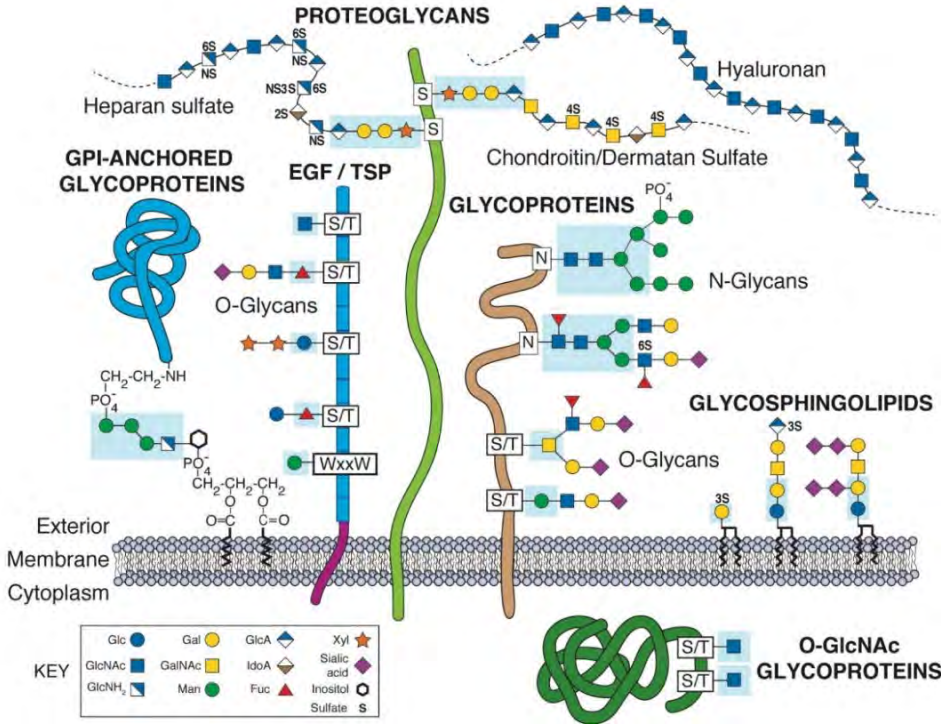
- Sugars: 3rd alphabet of life: Glycome → glycocode
- Every cell is covered by **glycoconjugates**





SARS-CoV-2, glycan shield

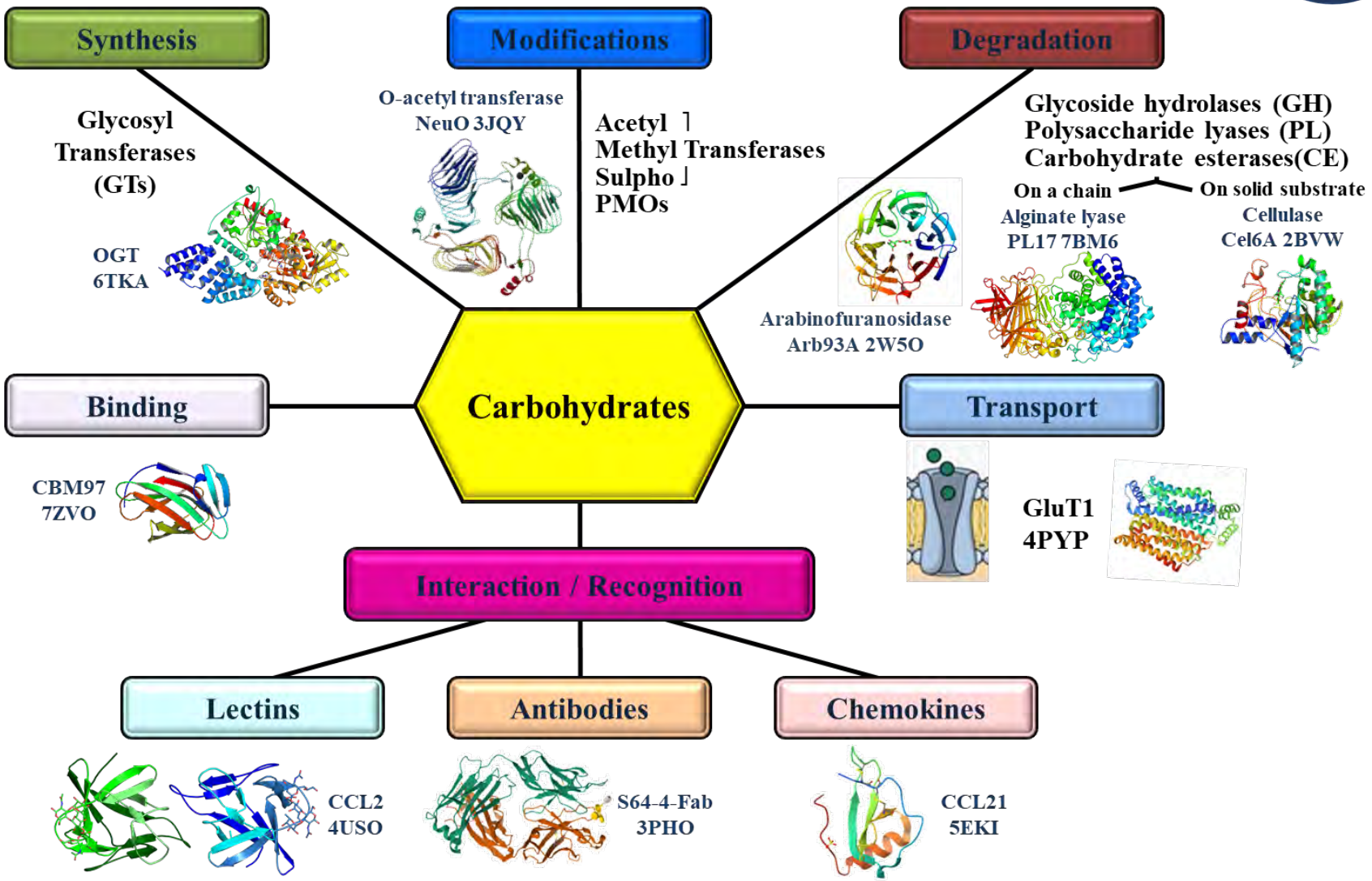




- Essentials in cell identity, fate, recognition & signaling

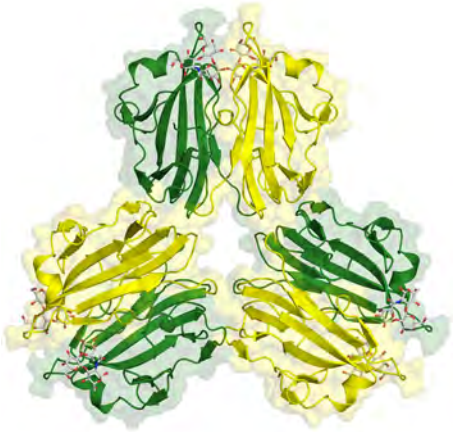
Essential in Glycobiology, The Scripps Research Institute,
Casalino *et al.*, ACS Central Sci (2020), 6: 1722.

Protein-carbohydrate interactions



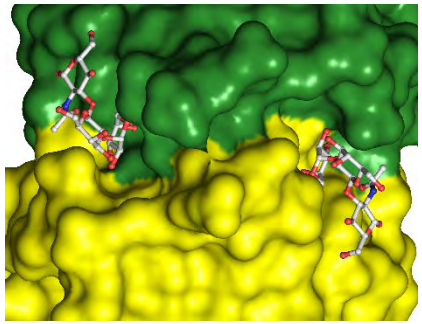
S Perez & O Makshakova, Chem. Rev., 2022, 122,15914

Which info by X-ray crystallography?

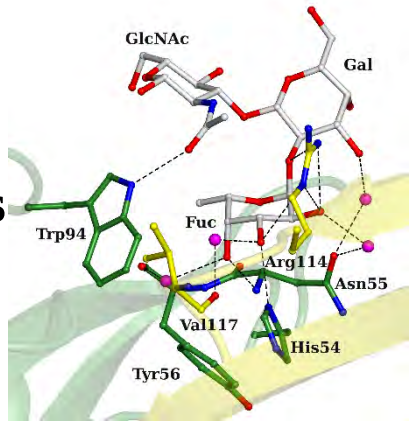


Fold/assembly

Binding site architecture

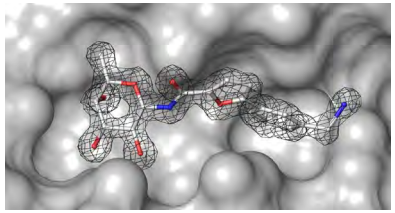


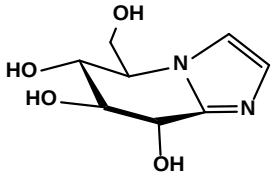
Interactions



Structure

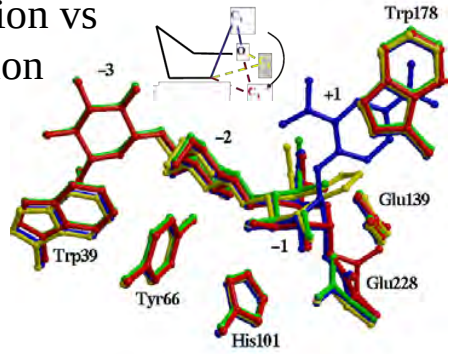
Drug design



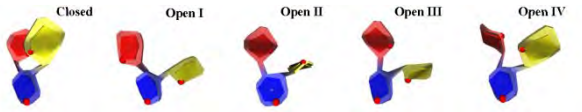


Enzyme mechanism

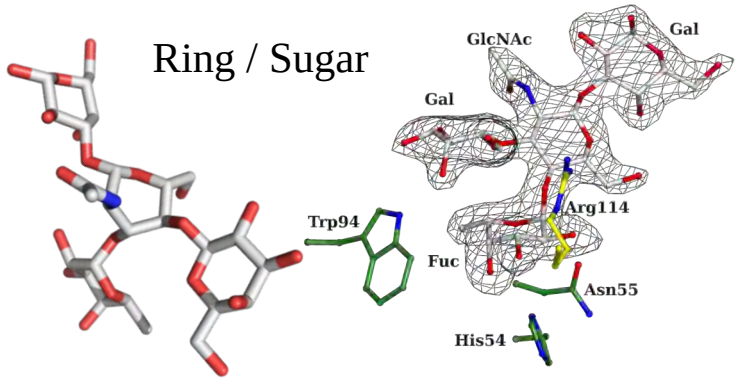
Retention vs Inversion



Conformation

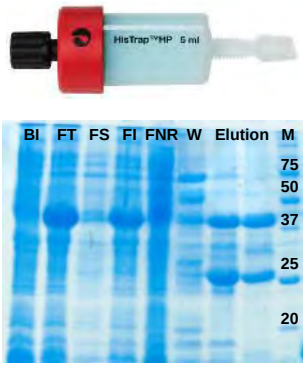


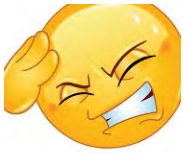
Ring / Sugar



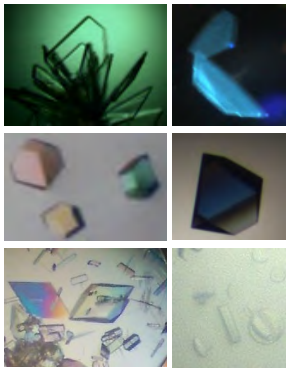
Protein X-ray crystallography

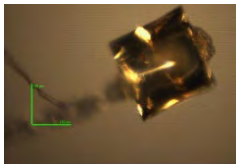
Expression
Purification



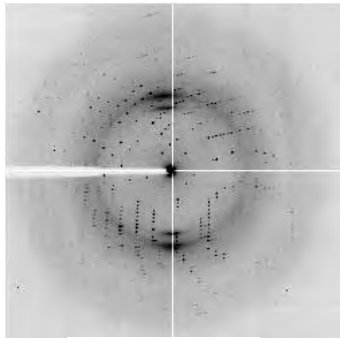


Crystallization





Freezing
& X-ray
diffraction



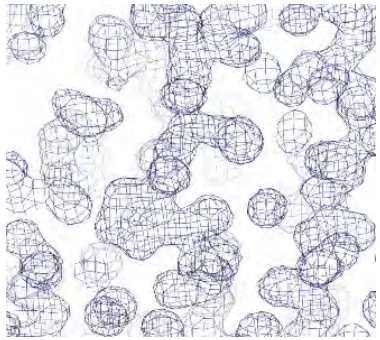
➢Advantages

- Can go to atomic resolution
- Atomic details obtained

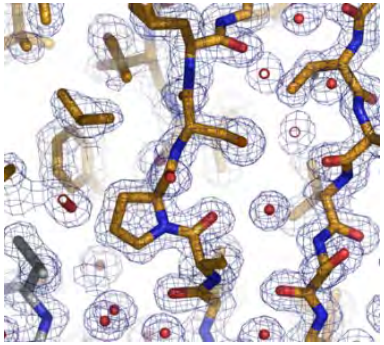
➢Disadvantages

- Molecules in solid-state environment
- Require crystals
- Requires order to diffract

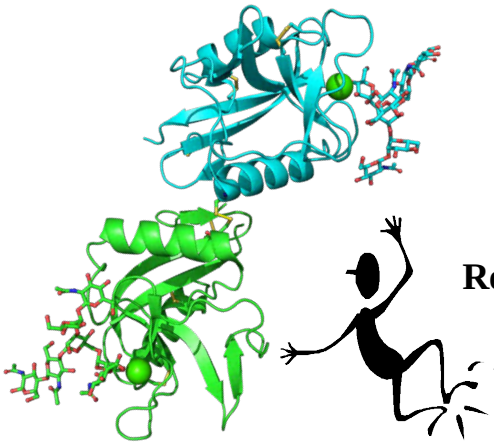
Phasing



Modelling

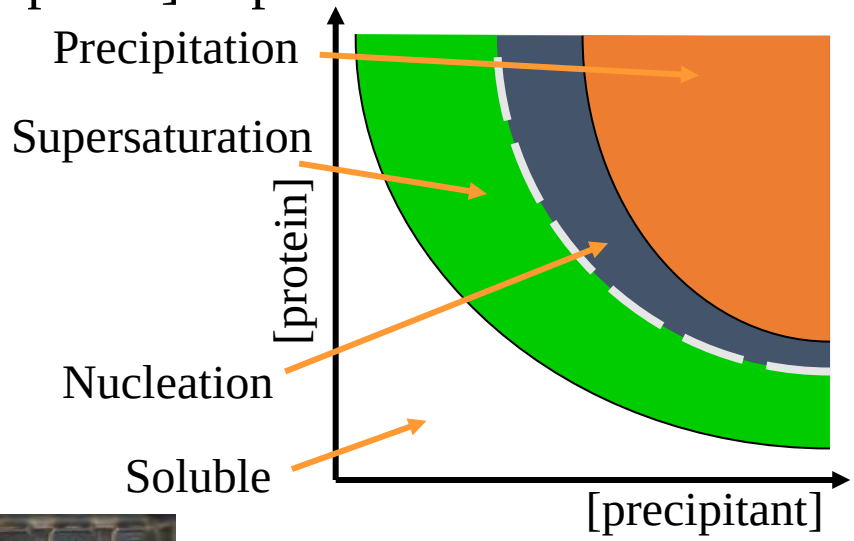


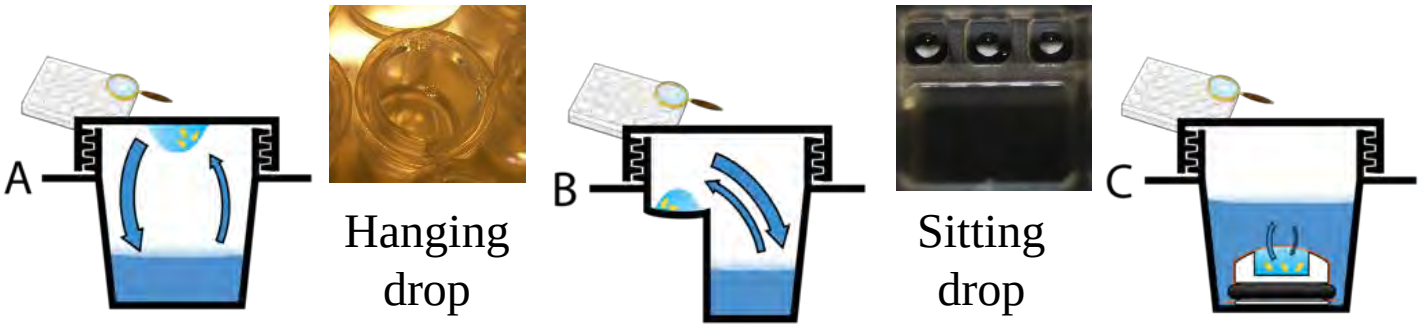
Refinement



Protein crystallogenesis

- Need pure and monodisperse protein
- Empirical
  - pH, buffer, temperature, [salt], [precipitant] dependant
- Precipitants
  - High salt concentration
  - Alcools or volatil compounds
  - Organic polymers
- Manual or robotized
- Vapor diffusion mostly used





Hanging drop

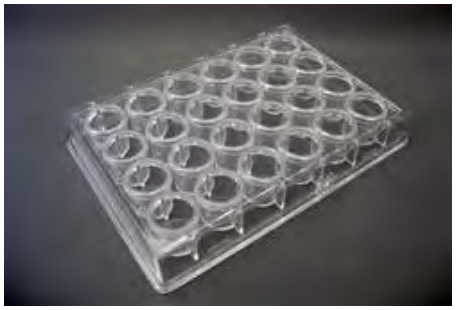
Sitting drop

microdialysis

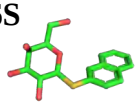
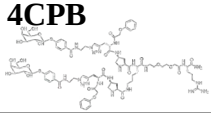
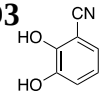
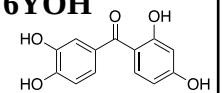
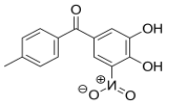
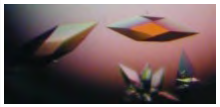
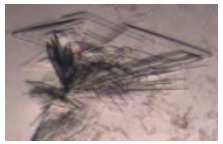
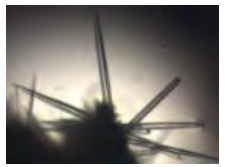
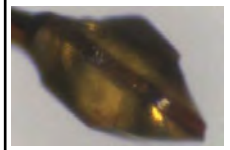
How to obtain crystals of protein complexes with ligand

Cocrystallization

- Put the ligand in reservoir
- Preincubate with the protein
  - Limit protein dilution: >1/10
  - Set duration and temperature
  - Easily reproducible & require little ligand



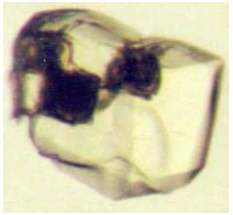
- Can have crystal and spacegroup when change ligand: LecA

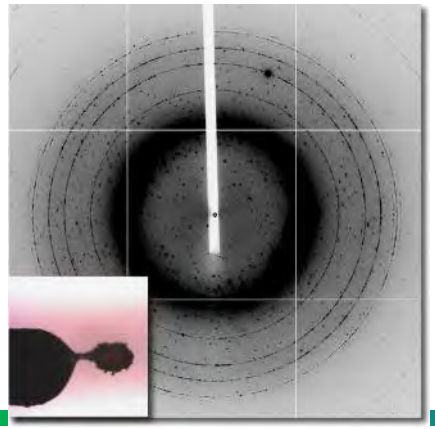
Galα1-2Galβ 2WYF	**Galα1-3Galβ1-4Glc / 2VXJ**	**Galα1-6Glc / 4AL9**	**DEG144 / 4A6S** 	**Bivalent / 4CPB** 	**Cathecol 6YO3** 	**Cathecol 6YOH** 	**Tolcapone** 
20% PEG6K 1 M LiCl 0.1 M NaAc 4	10% PEG5Kmme 25 mM KSCN 0.1 M NaAc 4.6	20% PEG2Kmme 0.2 M KBr 0.1 M NaAc 4.6	0.8 M Li_2SO4 0.1 M NaAc 4.5	10 % PEG8K 10% PEG1K 0. 2 M $MgCl_2$ 0.1 M Tris 8.5	20% PEG6K 1 M LiCl 0.1 M NaAc 4.5 1% DMSO	20% PEG6K 1 M LiCl 0.1 M NaAc 4.5 1% DMSO	24% PEG2Kmme 0.1M KSCN 0.1M NaAc 4.5
2.4 Å, $P2_1$	1.9 Å, P1 79.2 86.5 119.1 93.9 98.2 90.1	1.75 Å, P1 50.1 58.1 75.9 101.1 92.9 101	2.15 Å, $P3_221$	2.15 Å, $P2_12_12_1$	1.84 Å, I2	1.84 Å, $P2_122_1$	1.3 Å, $C222_1$
							

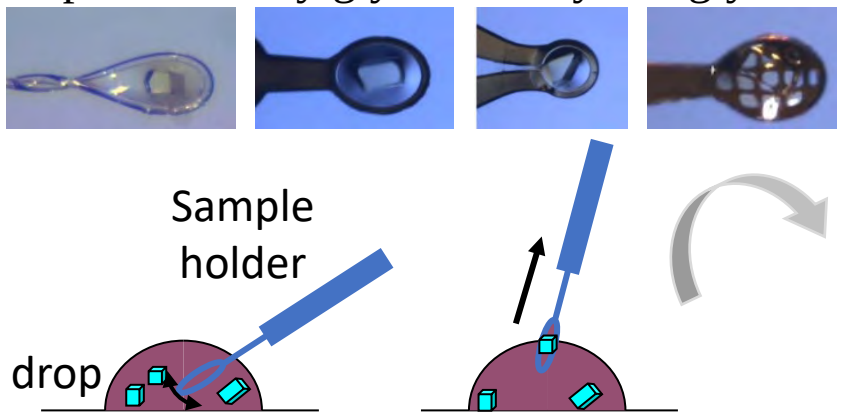
Blanchard, JMB, 2008; Nurisso, JBC, 2010; Rodrigue, Org Biomol Chem, 2013;
Blanchard, Protein, 2014; Novoa, Angewandte, 2014; Kuhaudomlarp, Angew Chem, 2021

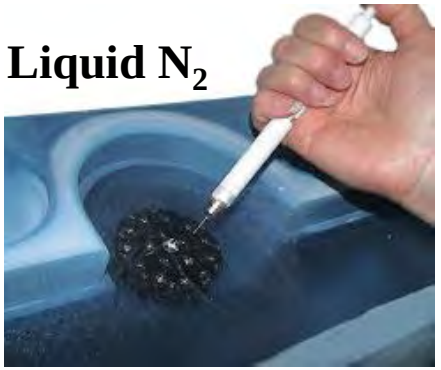
Soaking

- Dry soaking
  - Let 0.1-1µL ligand solution to dry before making drop
- Add the ligand to the reservoir solution
  - Add concentrated solution or a bit of powder
  - Transfer the crystal in solution with ligand
    - Try different concentrations and soaking time
- Soaking whilst freezing
  - Freezing limits radiation damages
  - Add ligand to the cryoprotectant solution
    - Less chance of replacement by glycerol, ethylene glycol and MPD









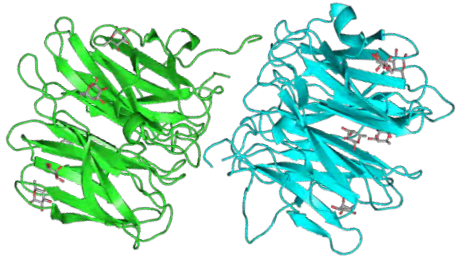
Kuhaudomlarp, Angew Chem, 2021; Mazzota, Molecules, 2023

Phasing: molecular replacement

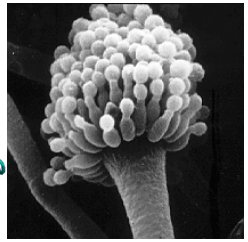
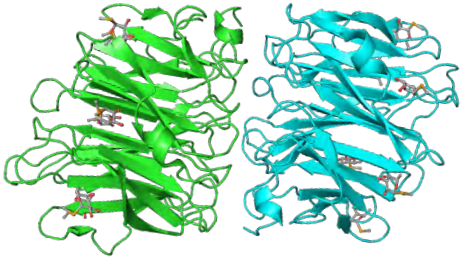
- Have a good model from homologous protein
  - Sequence identity > 25%
  - Can use alphafold if in Uniprot or fold known
  - Good conservation secondary structure: difficult for β-sheets

AAL/1OFZ/Hg
Aleuria auratia

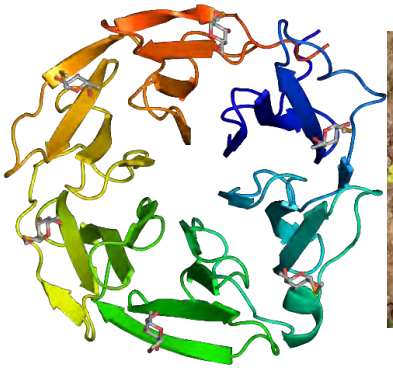


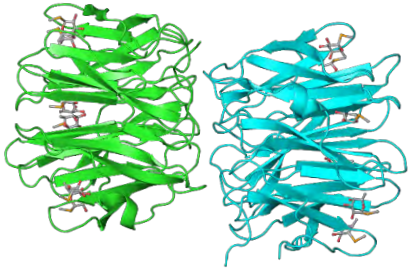


FleA/4AGI/SFU (35%)
Aspergillus fumigatus

TrfbL1/SFU (32/43%)
Terfezia boudieri

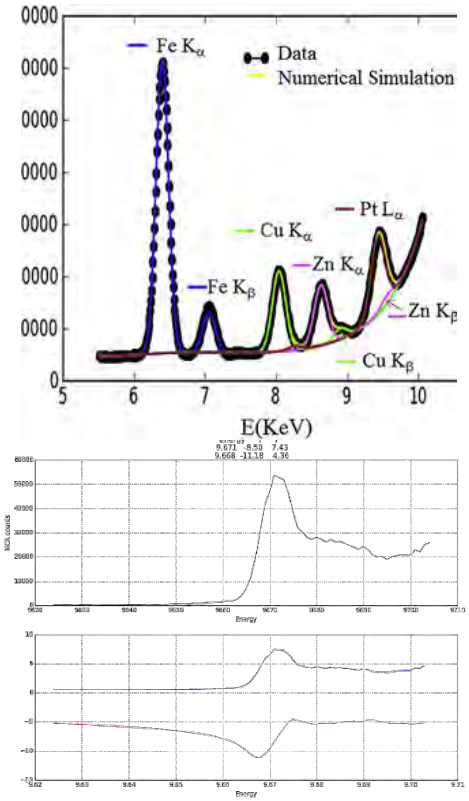




Wimmerova, JBC, 2003; Houser, PloS One, 2013, Basheer, 2023, in prep;
https://alphafold.ebi.ac.uk/

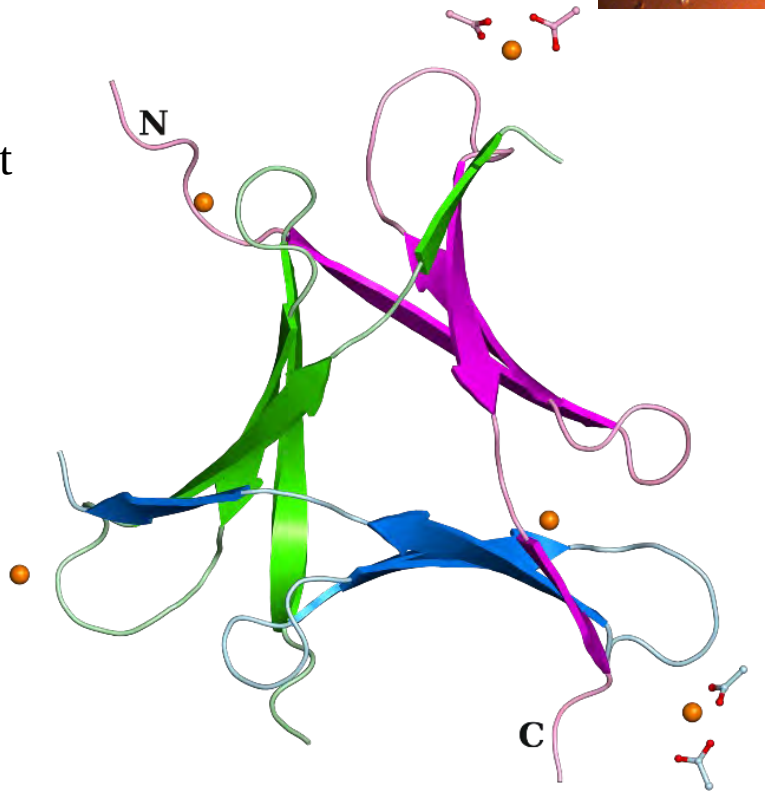
Phasing: isomorphous or anomalous

- Look in your crystallization conditions
  - PhoSL from *Pholiota squarrosa*
    - 300 mM Zinc acetate, 0.1 M Imidazole-HCl pH 6-7



X-ray fluorescent
XRF scan

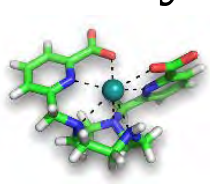
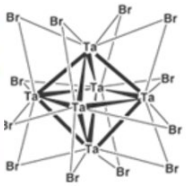
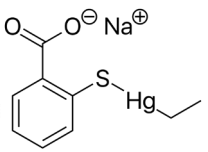
Absorption
scan



Cabanettes, Angew Chem, 2018

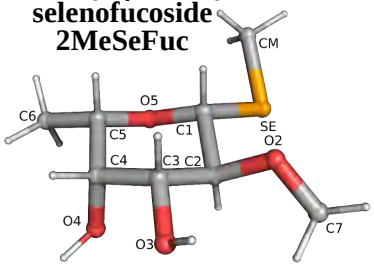
Phasing: isomorphous or anomalous-2

- Modified your protein with selenomet or selenocys
- Use heavy metals, clusters of lanthanides
- Use ligand with heavy atoms (Se, Br, S, F)

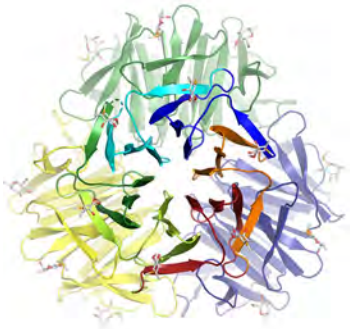




Lb-Tec2

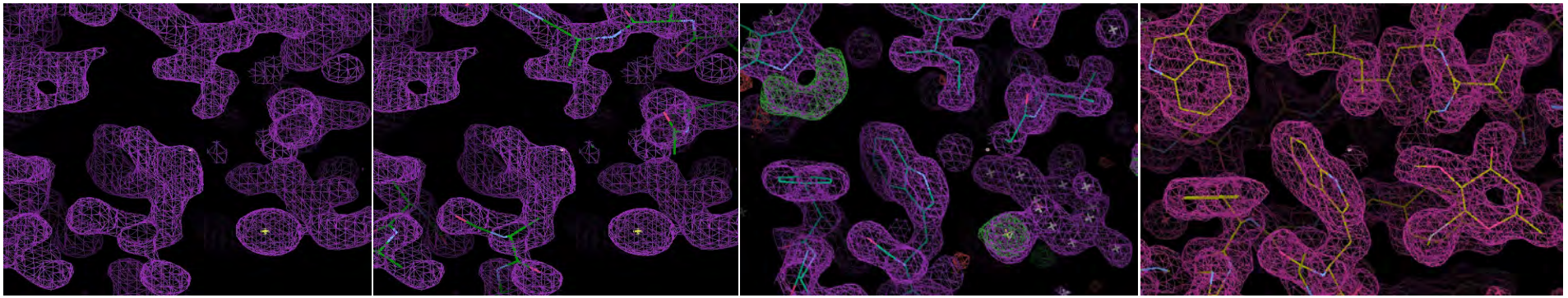
2-O-Methyl-β-methyl-L-selenofucoside 2MeSeFuc




Alex Titz
HIPS
Saarbrücken
Germany
R Sommer



Se peaks | **PolyAla building** | **Sequence assigned** | **Final**



Saboia Aragao, Proteins, 2008; Rocha, Plant Cell, 2016; Sommer, Structure, 2018;Notova, Comm Biol, 2022, https://crystallophore.fr/; Tiralongo, 2023, in prep

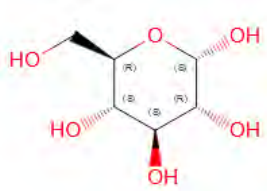
Carbohydrate refinement

- Find proper 3-letter code
  - Check ligand database http://ligand-expo.rcsb.org/ld-search.html

RCSB PDB
PROTEIN DATA BANK
RCSB PDB | Contact Us

Ligand Expo Search Result Summary

Query: glucopyranose
Query type: Molecular name (exact sub-string)
Result count: 252



GLC

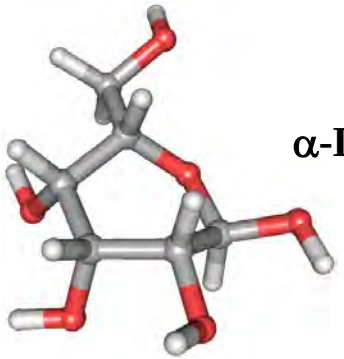
alpha-D-glucopyranose

Find entries where: GLC
- ☑ is present as a standalone ligand in 301 entries
- ☑ as a non-polymer is covalently linked to polymer or other heterogen groups 61 entries
- ☑ is present in a branched oligosaccharide 1,638 entries

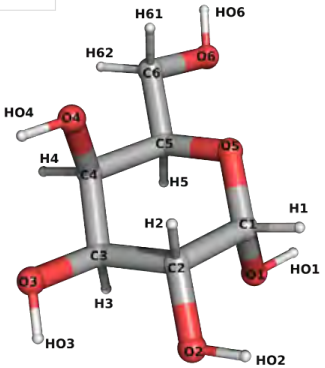
Chemical Component Summary	
Name	alpha-D-glucopyranose
Synonyms	alpha-D-glucose; D-glucose; glucose
Identifiers	(2S,3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol
Formula	$C_6 H_{12} O_6$
Molecular Weight	180.16
Type	D-SACCHARIDE, ALPHA LINKING
Isomeric SMILES	C([C@@H]1[C@H]([C@@H]([C@H]([C@H](O1)O)O)O)O)O
InChI	InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1
InChIKey	WQZGKKKJIJFFOK-DVKNGEFBSA-N

Chemical Details	
Formal Charge	0
Atom Count	24
Chiral Atom Count	5
Bond Count	24
Aromatic Bond Count	0

- 1 code per anomer
  - α-D-Glucose: GLC
  - β-D-Glucose: BGC
  - β-L-Glucose: Z8T
- Check library
  - No distorsion

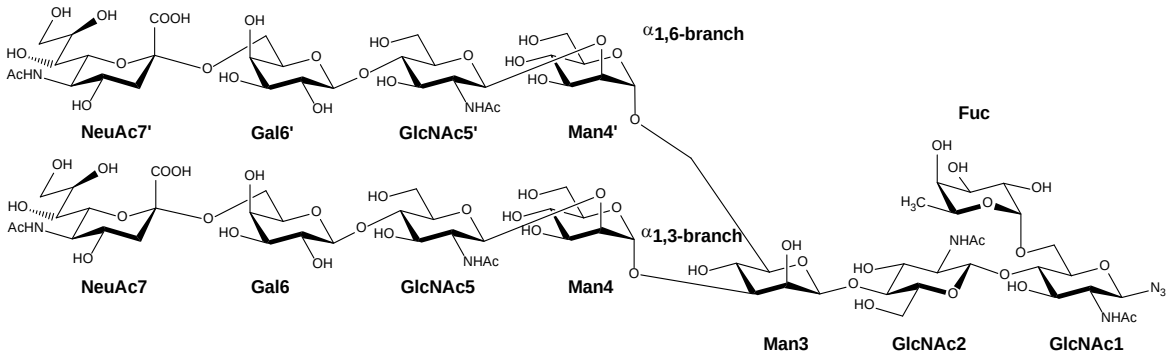


GLA
α-D-galactose

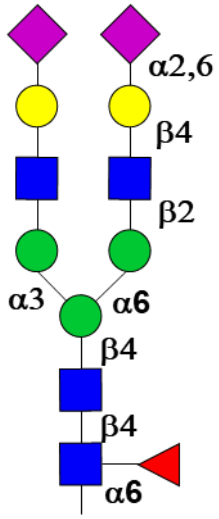


Oligosaccharide refinement

- Define correct linkage description
  - Programs do not know how to deal with L-sugars



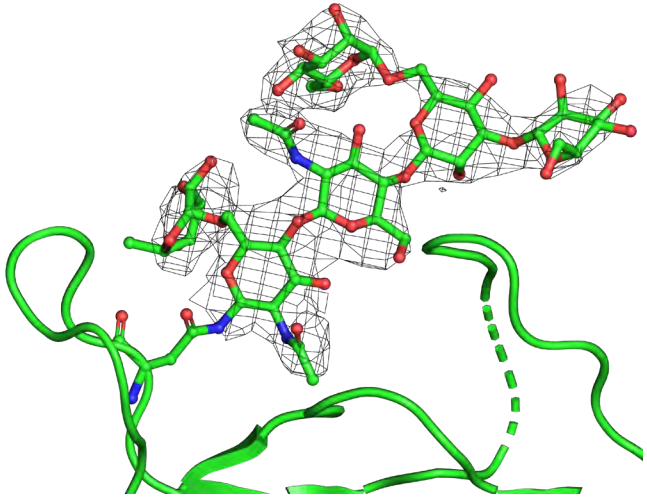
6FX3



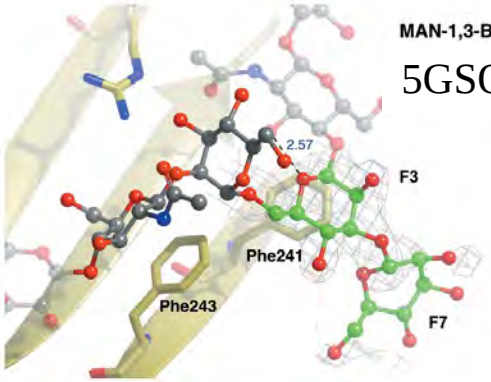

```
LINKR        O6   C4W A 102        C1   FUC A 101        BETA1-6
LINKR        O4   C4W A 102        C1   NAG A 103        BETA1-4
LINKR        O4   NAG A 103        C1   BMA A 104        BETA1-4
LINKR        O6   BMA A 104        C1   MAN A 105        ALPHA1-6
LINKR        O2   MAN A 105        C1   NAG A 106        BETA1-2
LINKR        O4   NAG A 106        C1   GAL A 107        BETA1-4
LINKR        C2   SIA A 108        O6   GAL A 107        SIA-GAL
LINKR        O3   BMA A 104        C1   MAN A 109        ALPHA1-3
LINKR        O2   MAN A 109        C1   NAG A 110        BETA1-2
```

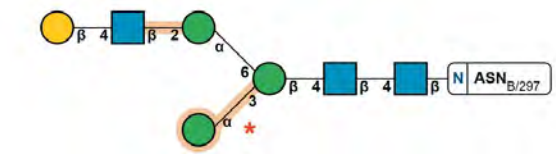
Carbohydrate validation

- Do not overfit at low resolution
- Check nomenclature, ring conformation, density fit
  - Privateer
  - PDB-REDO



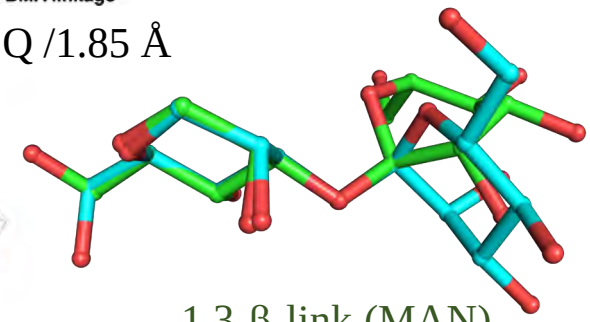
7ALK /3 Å





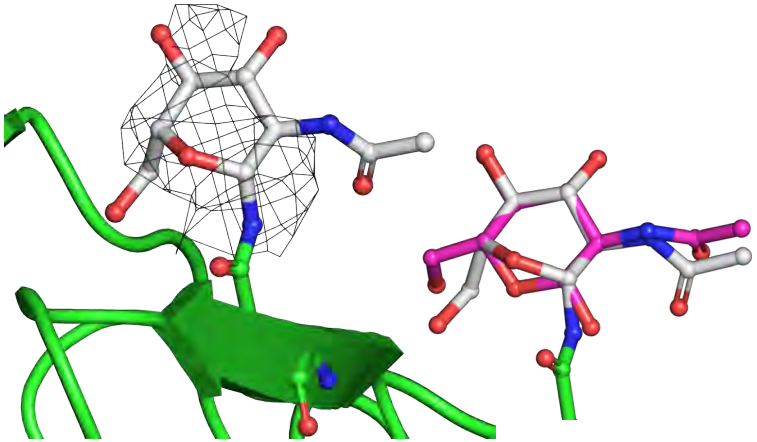
MAN-1,3-BMA linkage

5GSQ /1.85 Å



-1,3-β-link (MAN)

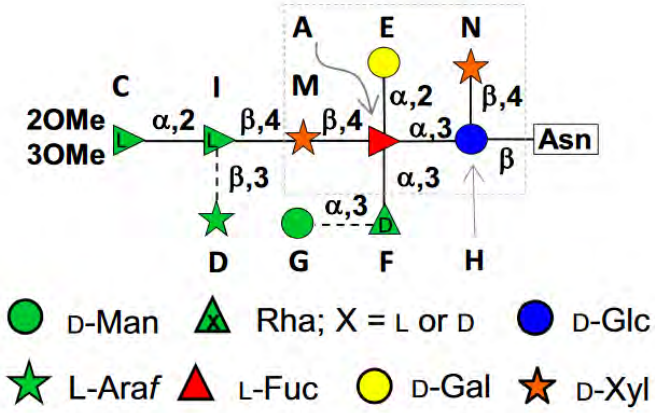
-1,3-α-link (BMA)



Agirre, Nat Struct Mol Biol, 2015; Dialpuri, Acta Crys D, 2023

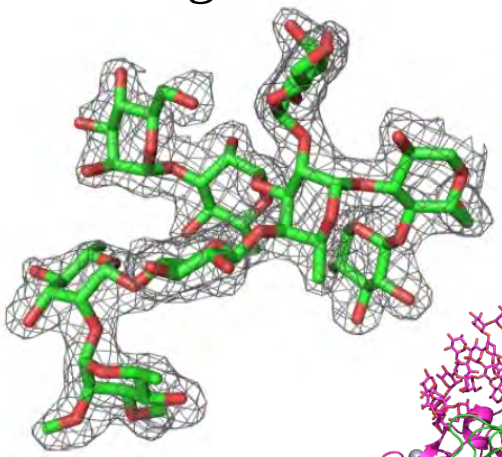
Tricky for non structural glycobiologists

- Major capsid protein (Vp54) of chlorovirus PBCV-1
  - X-ray 2002: try to fit classical N-glycans
  - Sugar NMR 2013: highly complex N-glycosylation
  - X-ray revised in 2017 + modelling



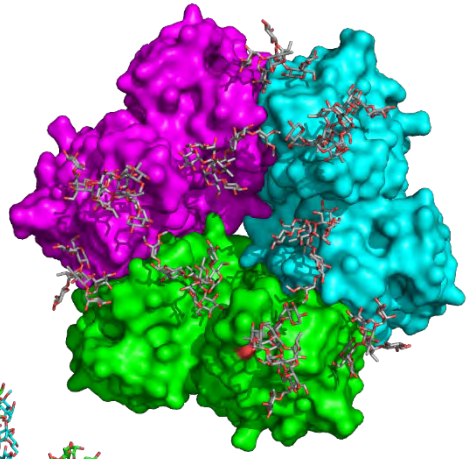
H = BGC
N/M = XYP
F = XXR
I = RM4

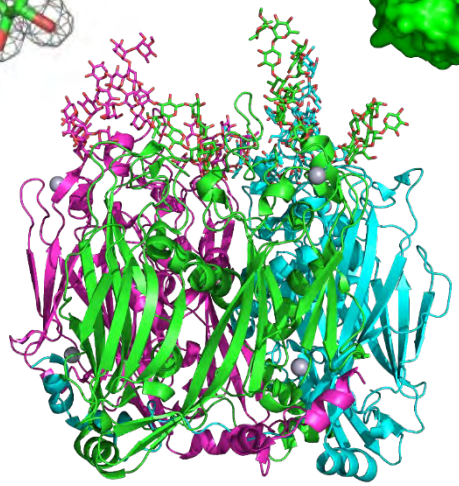
A = FUC
E = GLA
G = MAN
C = 7CV



4 glycosylation sites

5TIP

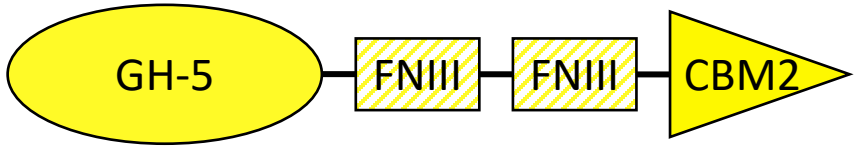


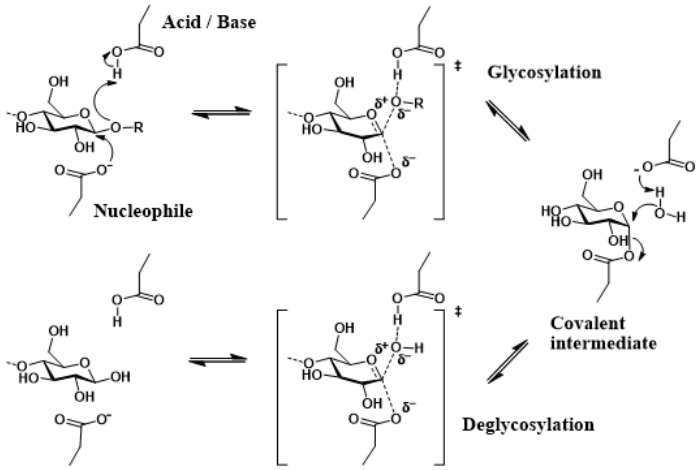


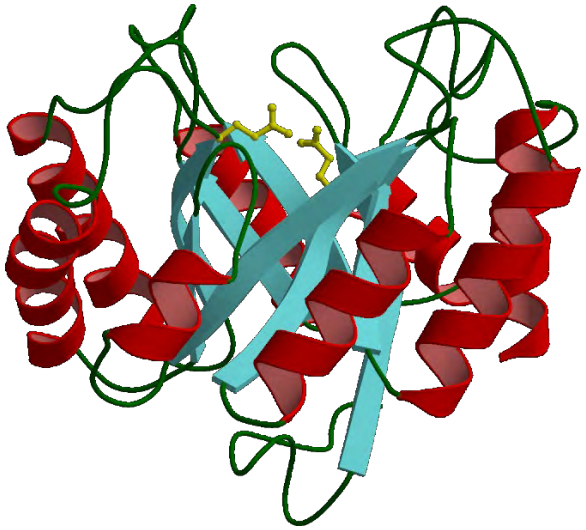
De Castro, PNAS, 2018

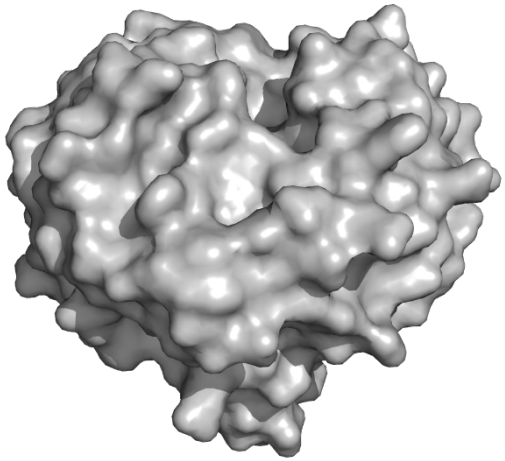
Insights in the retaining mechanism of GHs

- Cellulase Cel5A from *Bacillus agaradhaerens*
  - Modular endoglucanase
  - Active pH range 5.0-13.0
  - Glu139 (acid/base) and Glu228 (nuc)





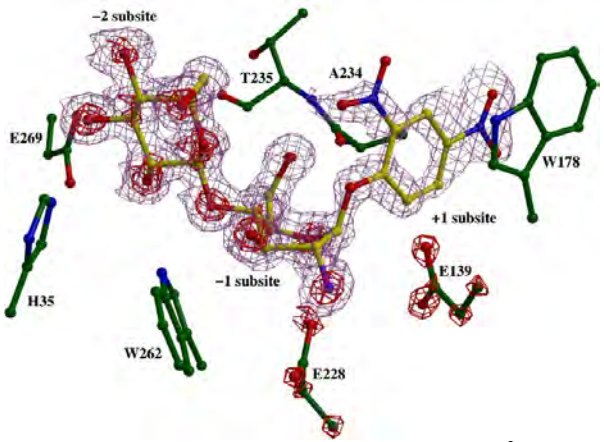




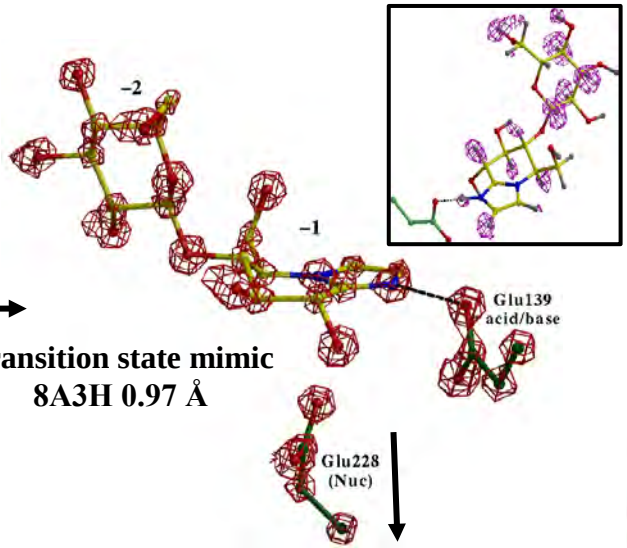
- Trapping of each step by X-ray crystallography
  - Use specific ligand, mutated protein, inactive pH

Davies, Biochemistry, 1998; Varrot, JMB, 2000; Varrot, Acta Cryst D, 2003

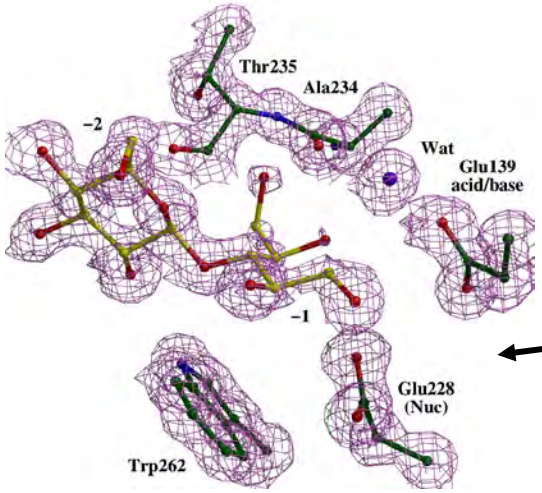
Snapshots at atomic resolution



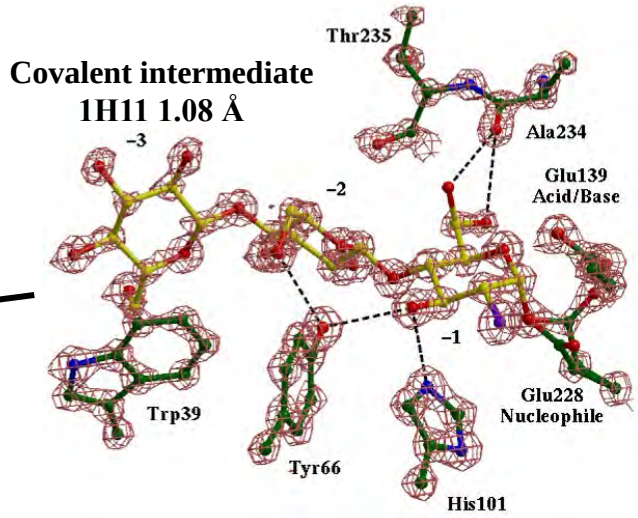
Mickaelis complex 1H2J 1.15 Å



Transition state mimic
8A3H 0.97 Å

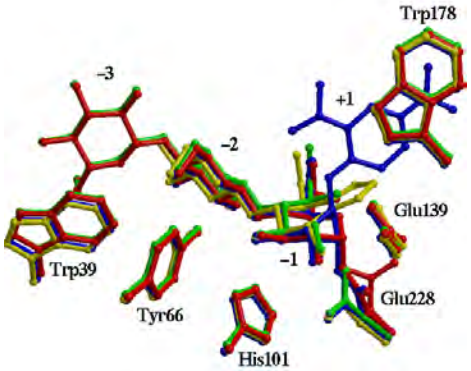


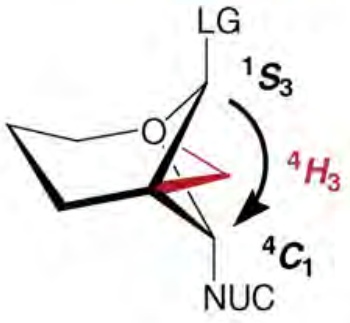
Product complex 1HF6 1.15 Å



Covalent intermediate
1H11 1.08 Å

- Static Cel5A
- Nucleophilic migration

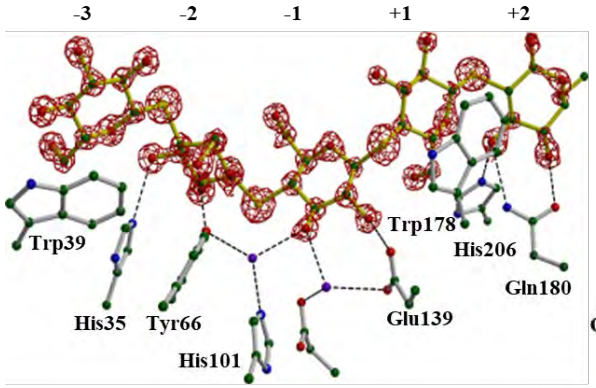




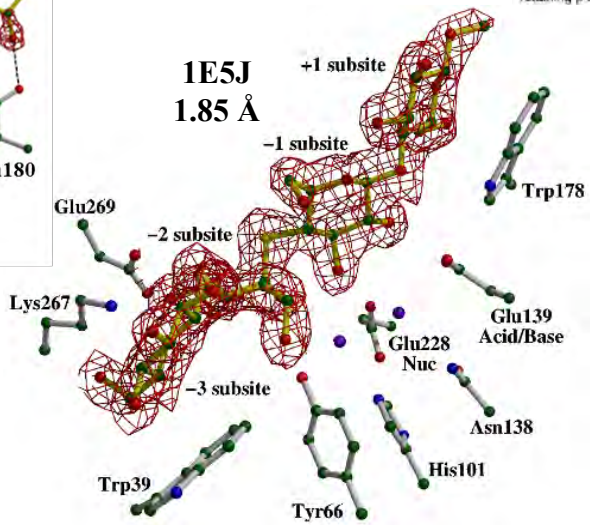
Varrot, Acta Cryst D, 2003

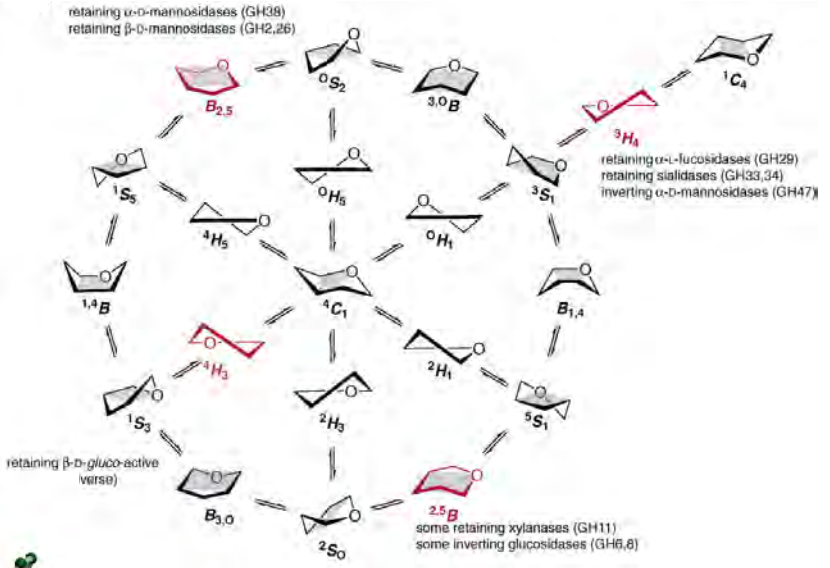
GHs inhibitors design

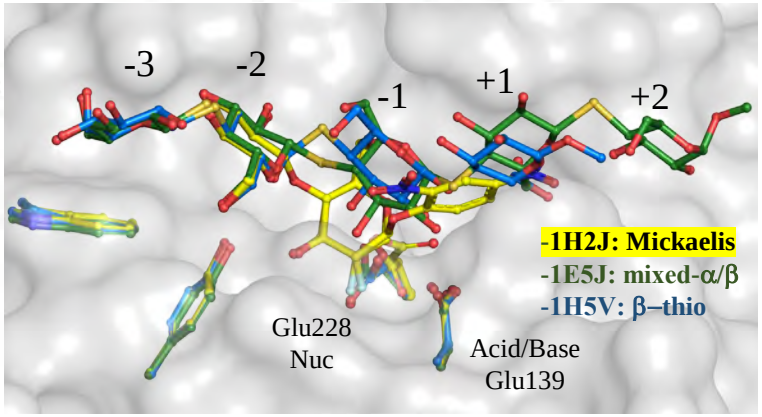
- Interconversion of sugar ring conformations
  - Highlights transition state
  - Essential for inhibitor design
- Serendipity
  - New inhibitors class for glucosidases



1H5V
1.1 Å



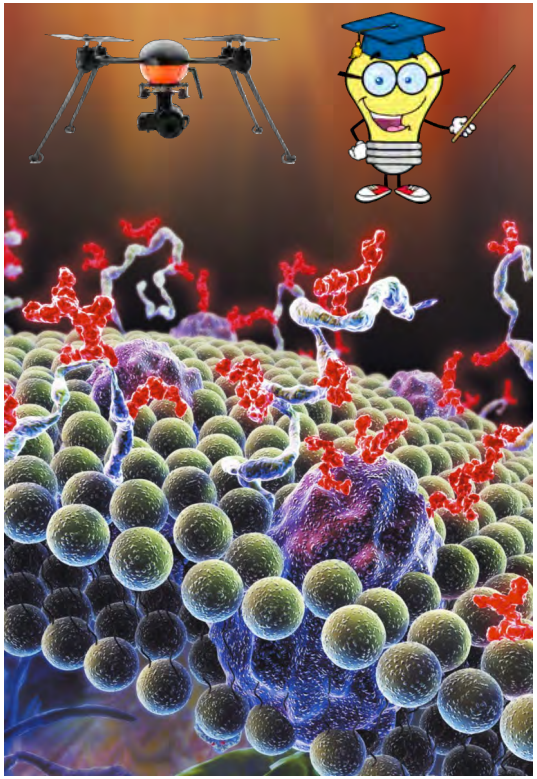




Davies, Biochem Soc Trans, 2003, Varrot Acta Cryst D, 2001; Varrot, ChemBiochem 2001

Lectins

- Ubiquitous carbohydrate binding proteins
- Specific and reversible sugar binding without modification
- Decipher the glycocode
- Implicated in many cellular processes
  - From warning their kin to poisoning their enemies
- Multivalency compensates for low affinity
  - Give ability to agglutinate cells
- New database: https://unilectin.unige.ch/



Cermav cnrs Glyco@Alps UGA Université Grenoble Alpes SIB Swiss Institute of Bioinformatics UNIVERSITÉ DE GENÈVE

Unified exploration platform for manually curated and predicted lectins

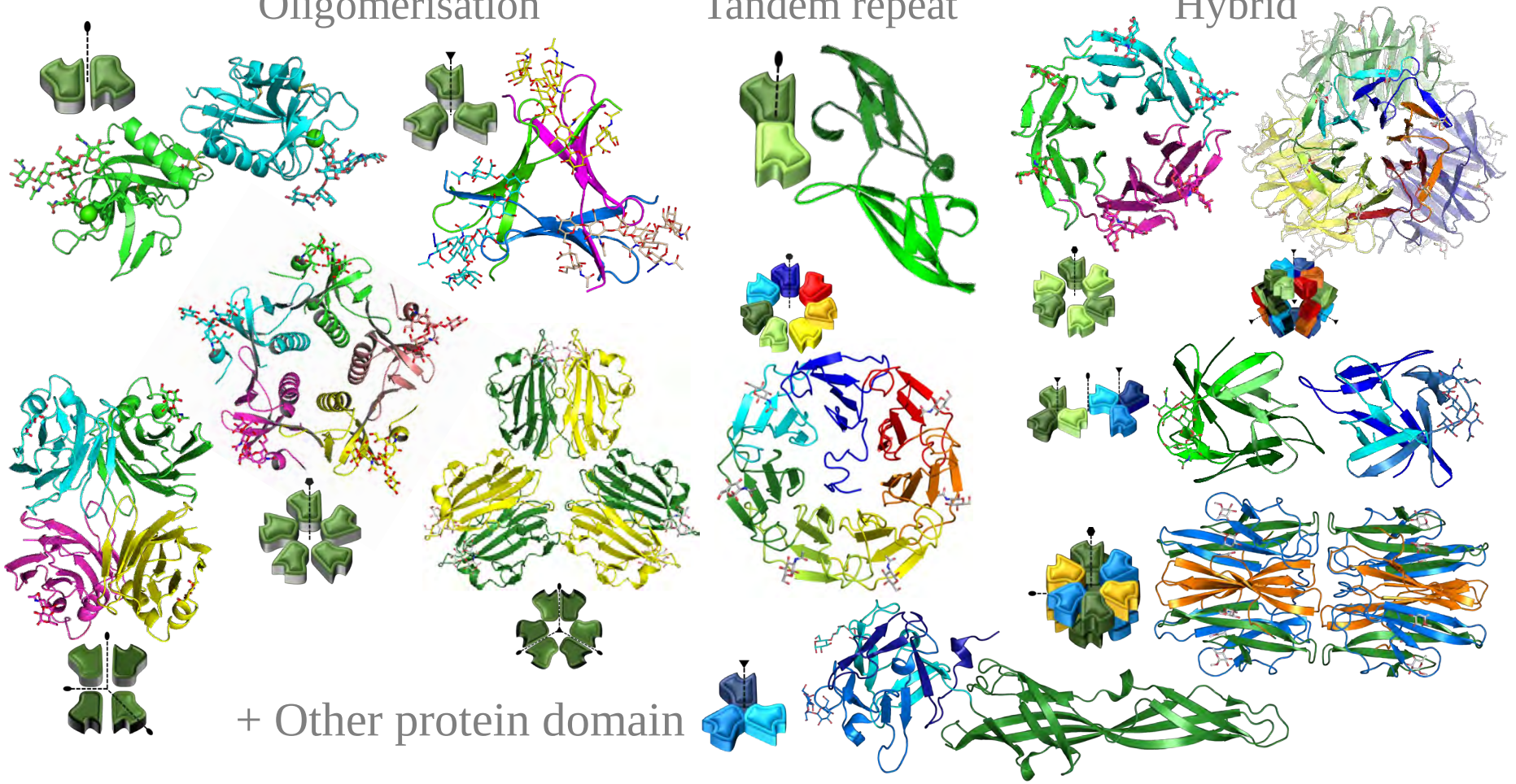
Lectin structure

- High diversity of fold and quaternary structure

Oligomerisation

Tandem repeat

Hybrid

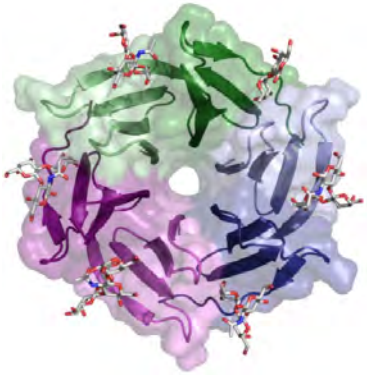


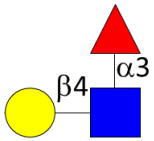
+ Other protein domain

Notova, Cur Op Struct Biol, 2020

Rare conformation

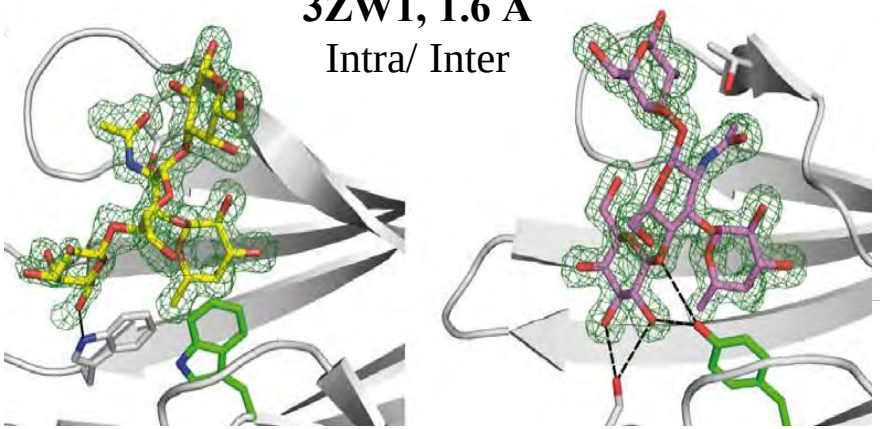
- 1 structure of lectin complex with ring distorsion
- BambL/RSL: Hidden conformation of Lewis X

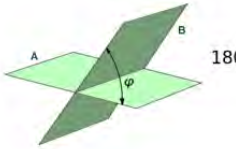




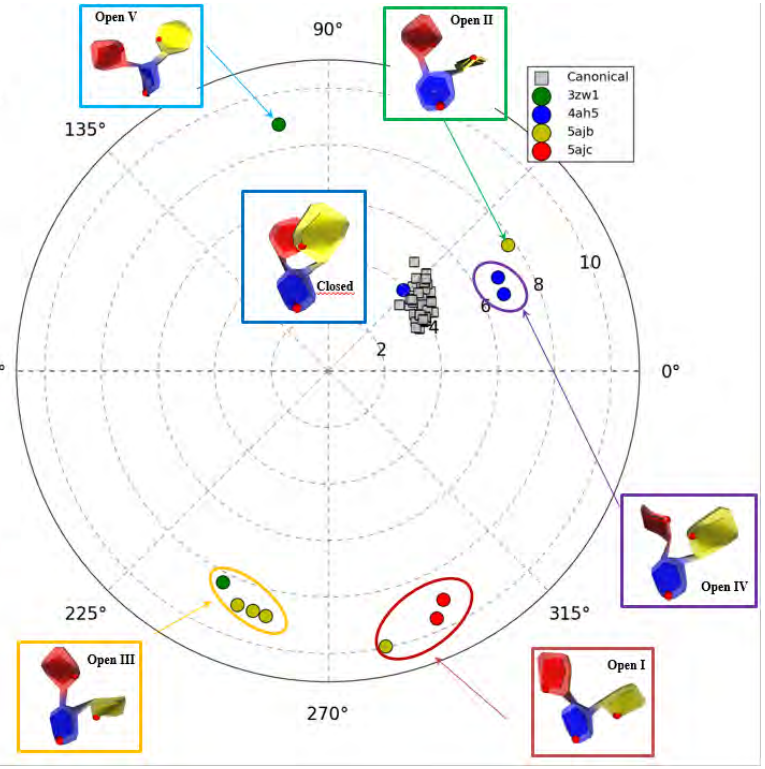
Lewis x

3ZW1, 1.6 Å
Intra/ Inter





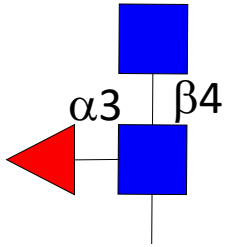
PDB analysis Polar coordinates

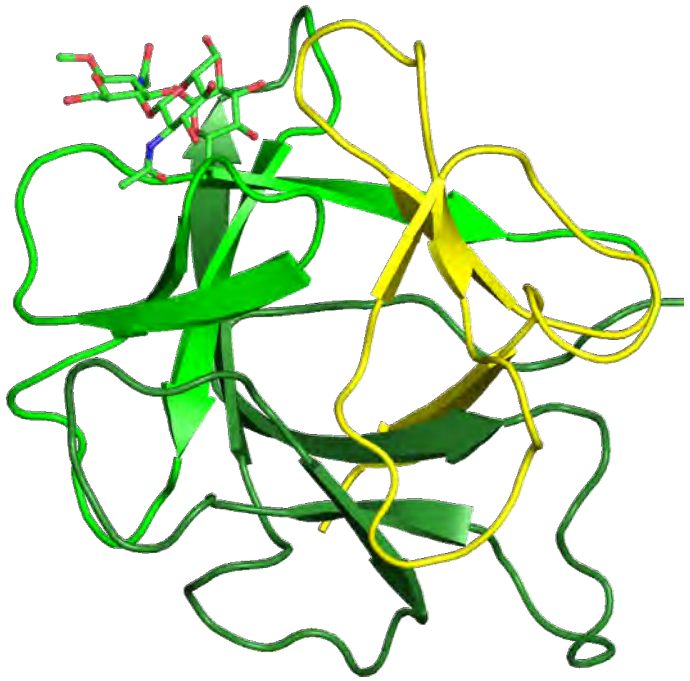


Topin, Plos One; 2013, Topin, ACS chem Biol, 2016; Lepsik, Eur J Med Chem, 2019

Multivalency and function

- CCL2 from *Coprinopsis cinerea*
  - Involved in innate immunity and defense
  - Toxic for nematodes and flies
    - Recognised 3-core fucose of midgut N-glycans



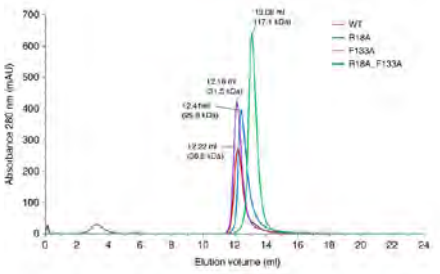


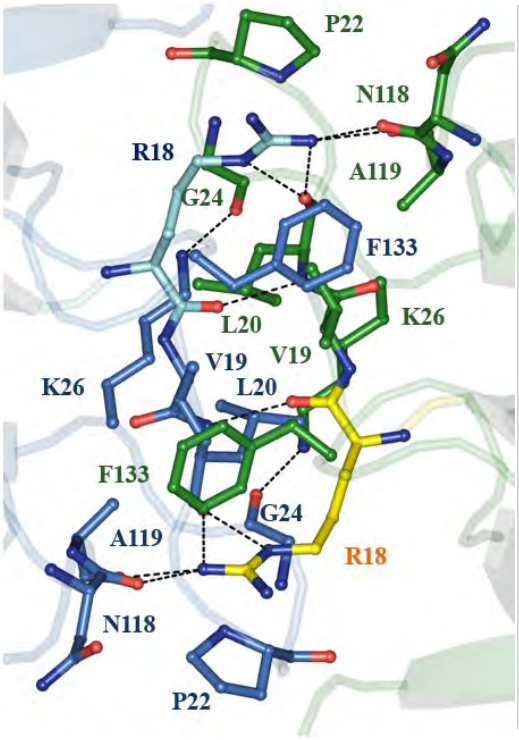
  - Monomer on Superdex75
  - RMN structure: β-trefoil fold
    - Only subsite β functional
  - Could not explain toxicity
    - Monomer and monovalent

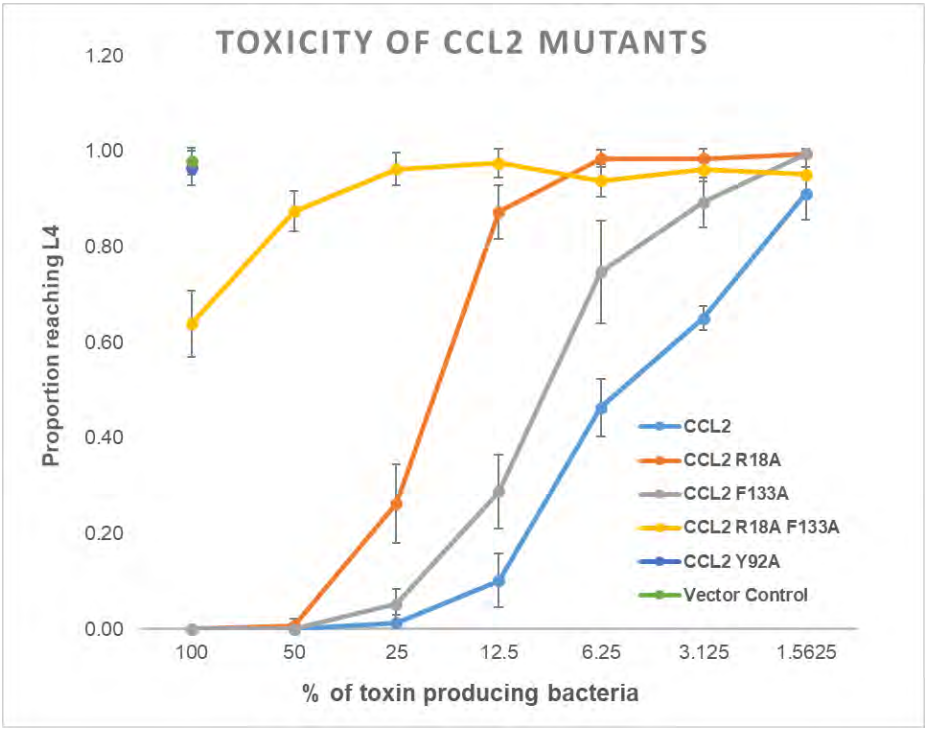
Schubert, Pmos Path, 2012; Bleuler-Martinez, Glycobiology, 2017

Multivalency and function-2

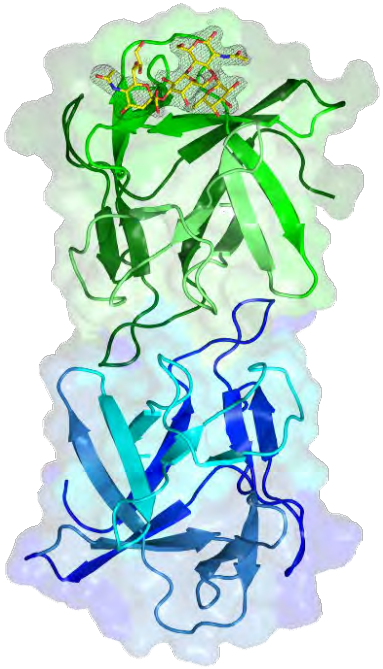
- CCL2 is a dimer
  - Confirmed by SEC on Enrich70, DLS, Native gel
  - Toxicity dependent on its dimerization
    - Disruption of dimer interface requires double mutation







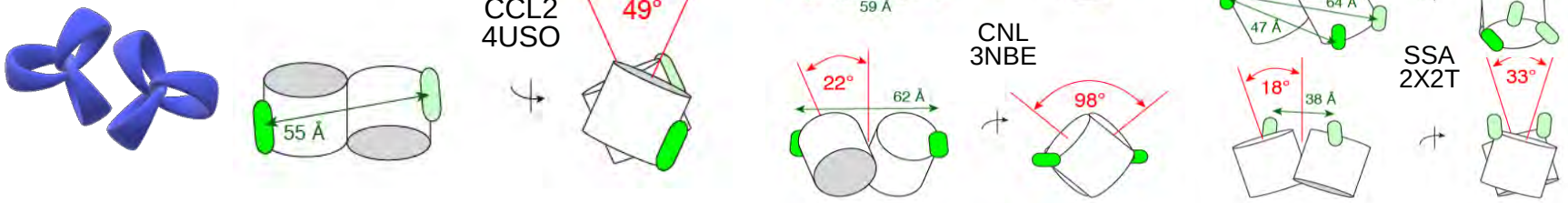
+sLe^x
4USO 1.95Å



Bleuler-Martinez, Glycobiology, 2017

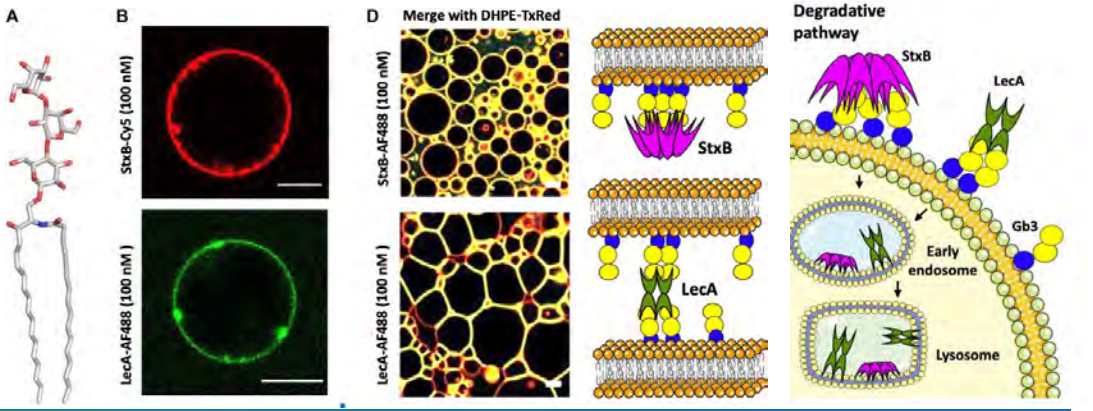
Multivalency and function-2

- Spatial distribution of binding sites impact function and recognition mode of β-trefoil lectins
  - CCL2: 1 binding surface on the side



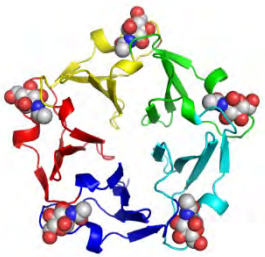
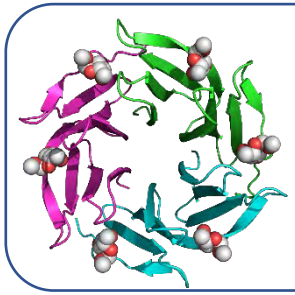
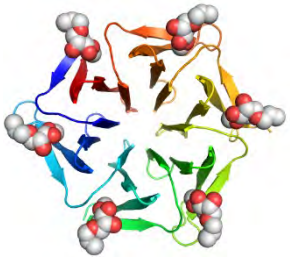
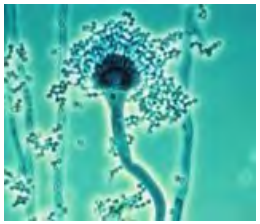
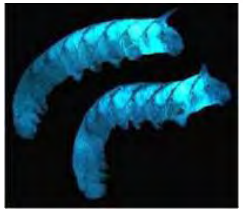
  - Others: 1/2 binding surfaces on the top

- Reorganisation of membrane glycoconjugates
  - Clustering of glycolipids
  - Crosslinking
  - Change in membrane dynamics
  - Endocytosis

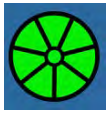


Bleuler-Martinez, Glycobiology, 2017, Siukstaite, Front Mol Biosci, 2021

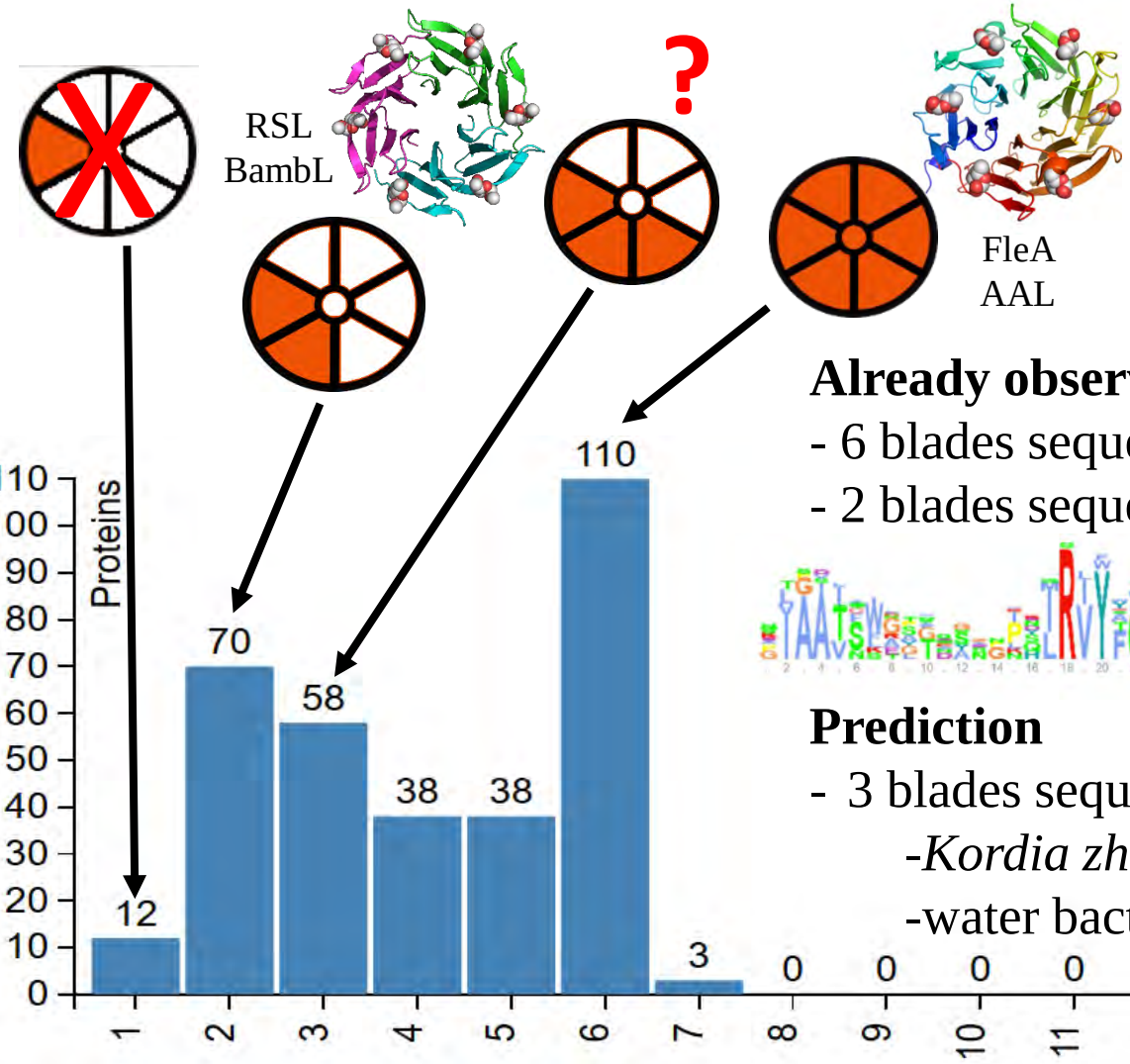
Lectins with β-propeller fold

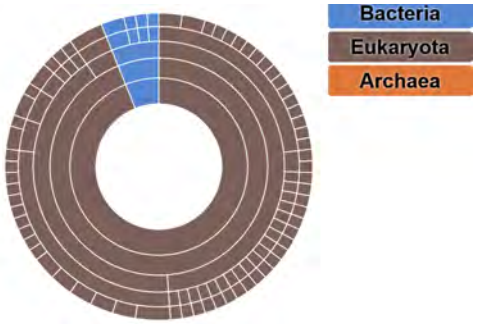
Tachylectin 2 GlcNAc	RSL fucose	FleA fucose	Tectonin/Lb-Tec2 Me-sugar	PLL fucose	PVL GlcNAc
					
					
Tachipleus tridentatus	*Ralstonia solanacearum*	*Aspergillus fumigatus*	*Laccaria bicolor*	*Photorhabdus luminescens*	*Pstatyrella velutina*
Beisel *et al.* *EMBO J.* 1999	Kostlanova *et al.* *J. Biol. Chem* 2005	Houser J, Plos One, 2013, 8:e83077	Sommer *et al.* *Structure* 2018	Kumar *et al.* *J. Biol. Chem* 2016	Cioci *et al.* *J. Mol. Biol.* 2006

- Define blade signature for prediction: PropLec in Unilectin

PropLec5A FLF	PropLec6A RVY	PropLec6B GVN	PropLec7A EVF	PropLec7B GFG	PropLec7C
					

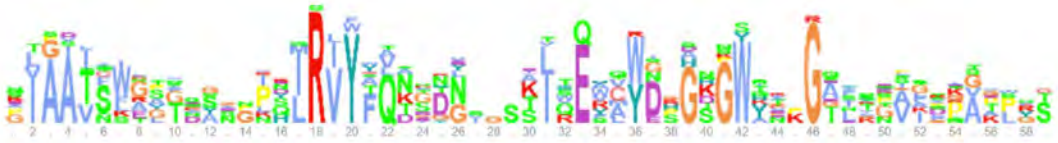
Distribution of predicted blade in PropLec6A





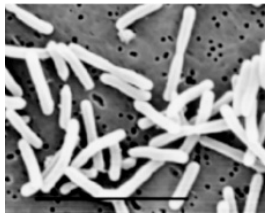
Already observed

- 6 blades sequences in fungi
- 2 blades sequences (fold : 3 x 2 blades)



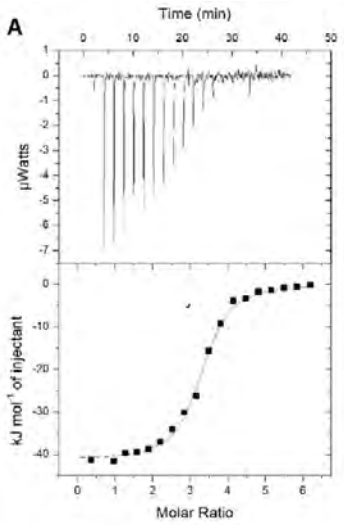
Prediction

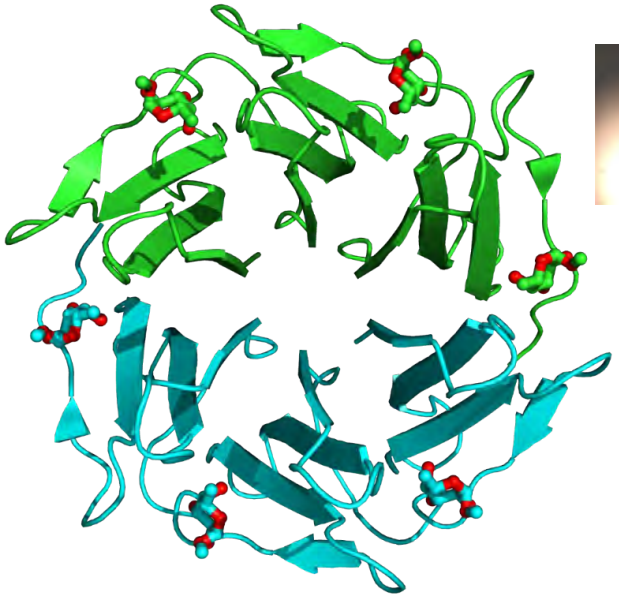
- 3 blades sequences (fold : 2 x 3 blades)
  - *Kordia zhanghouengsis*
  - water bacteria that eats algi



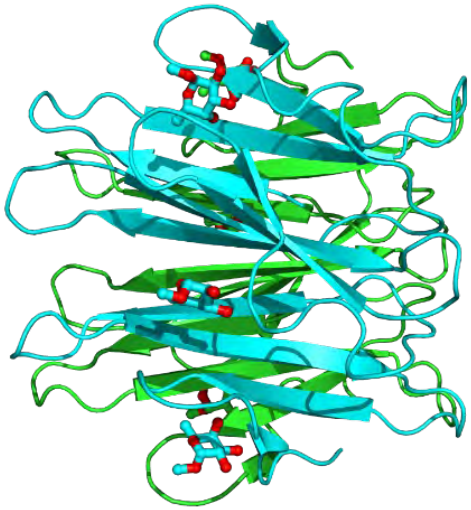
Recombinant KozL

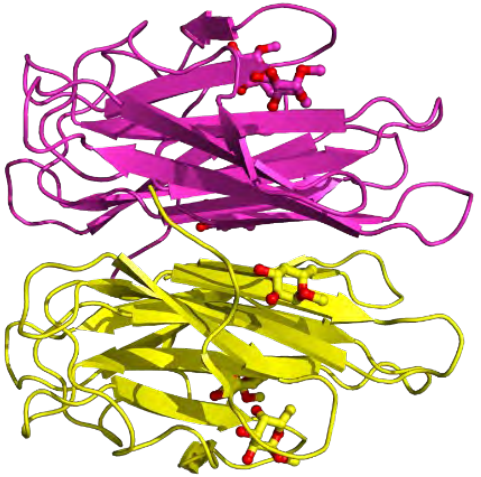
- Overexpressed in *Escherichia coli*
  - Purified on mannose agarose: 35 mg/L
  - ITC: 3 fucose binding sites
  - Tetramer by AUC
- Structure solved by SAD using αSeMeFuc
  - **Prediction corroborated:** 2 x 3 blades





6HTN
1.55 Å

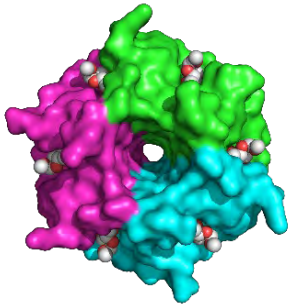




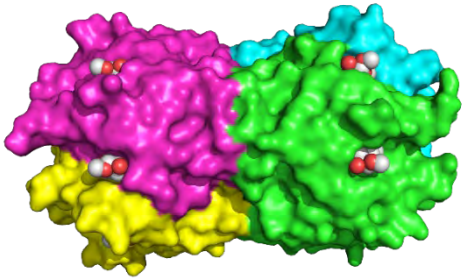
Bonnardel, Structure, 2019

PropLec6: many arrangements

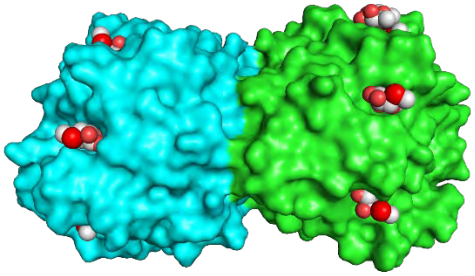
BambL/RSL
6 Fuc



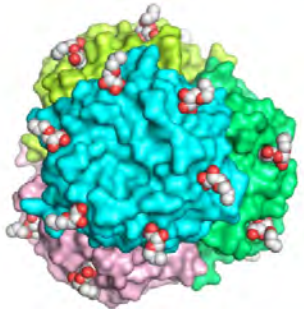
KhozL
12 Fuc



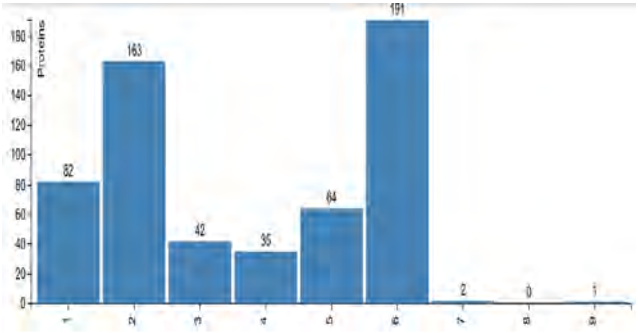
FleA/ SapL1
12 Fuc



Lb-Tect2
24 Me-sug



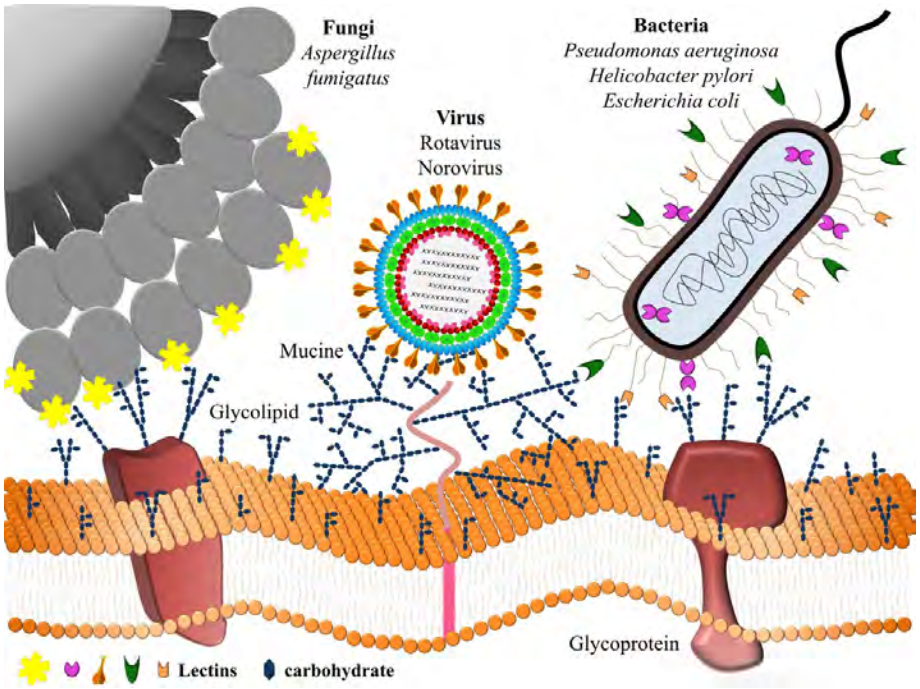
- Is 1 blade occuring?
  - Fold: 6x1 blade

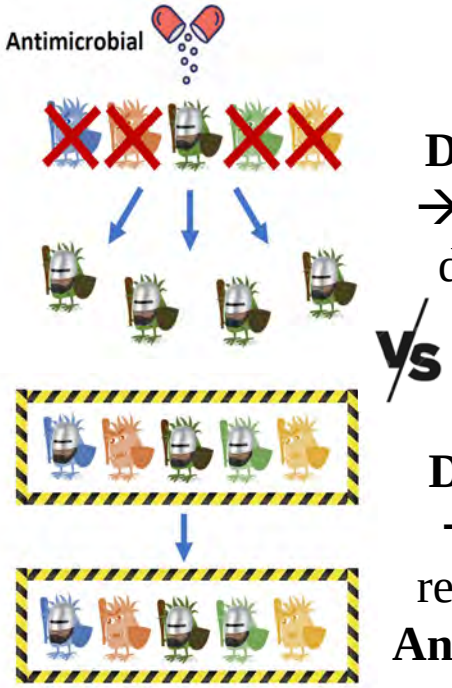


Audfray, JBC, 2012, Houser, Plos One, 2013; Sommer, Structure, 2018; Bonnardel, Structure, 2019, Martinez-Alarçon, Scie Rep, 2021

Lectins as therapeutical targets

- Lectins used host glycoconjugates as entry points
  - Mediate host recognition and adhesion
  - Blocking lectins → prevent infections





Classic:
Destroy pathogens
→ selective pressure driving **resistance**

vs

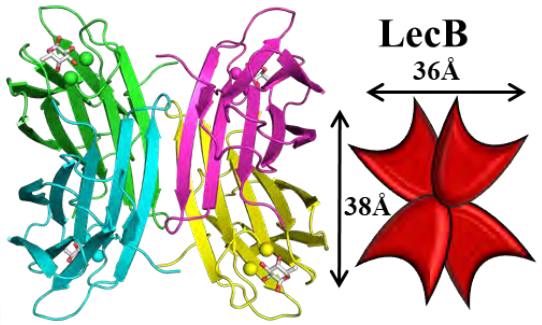
Alternative:
Disable pathogens
→ **Resistance** not required for survival
Antiadhesion therapy

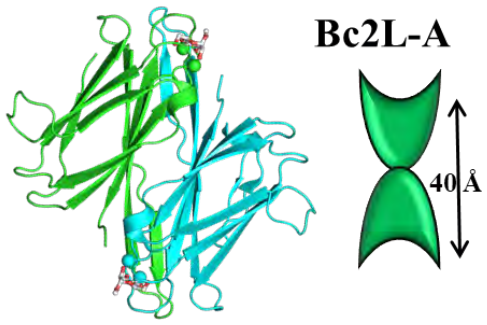
→Development of glycocompounds as antimicrobials

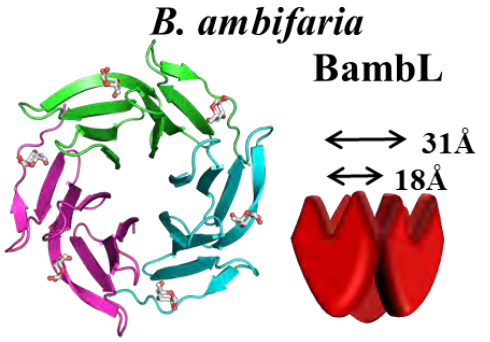
Ofek et al, FEMS Immunology & Medical Microbiology, 2003, 33: 181; Heggelund J, Curr Op Struct Biol, 2017, 44:190

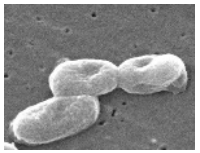
Lectin targets

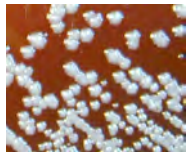
➢Opportunistic pathogens → bronchopulmonary infections







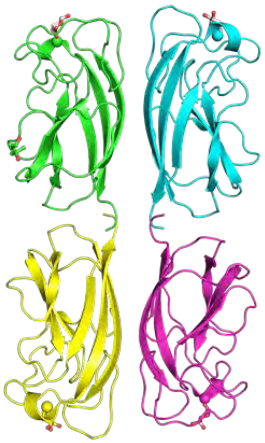


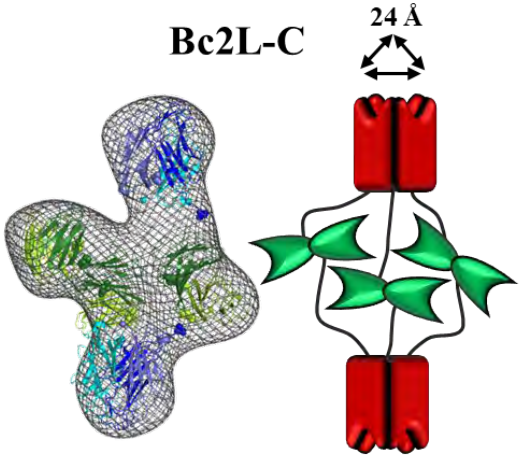


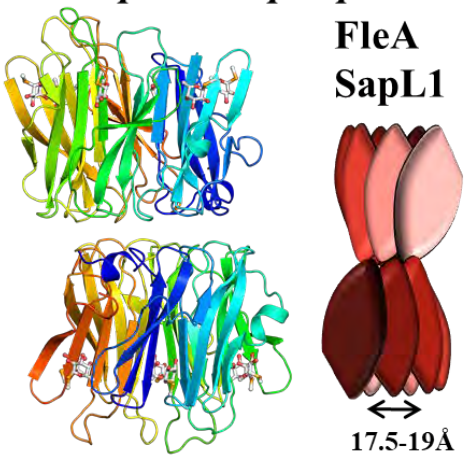
Pseudomonas aeruginosa

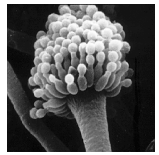
Burkholderia cenocepacia

Aspergillus fumigatus
Scedosporium apiospermum

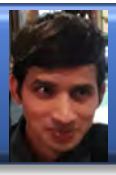






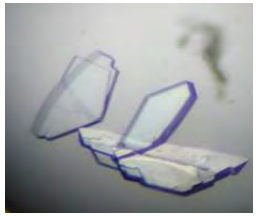


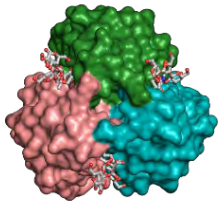
Bc2lC-nt

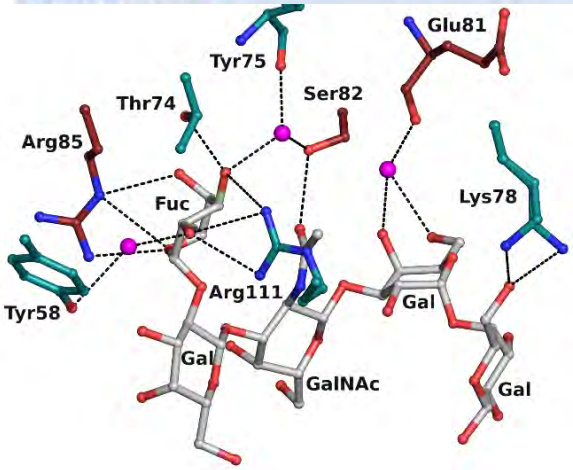



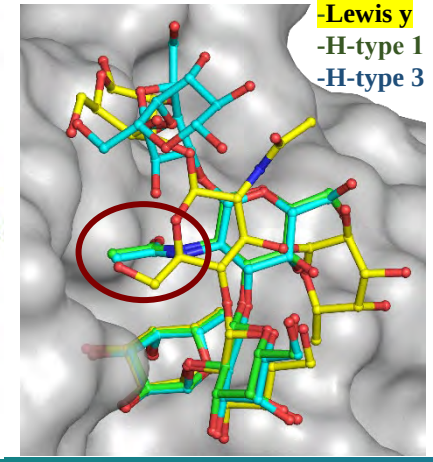
- New construct 1-132 in pCold-TEV: good yield & stable
- Complex with oligos by cocrys and with inhibitors by soaking

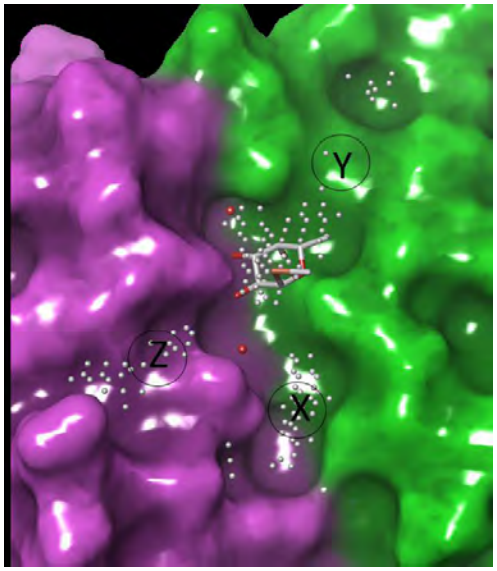
Ligand	k_D (µM)	PDB	Res (Å)
L-Fuc	2430	2WQ4	1.42
L-Gal	2000		
Disac	2500		
H-type 1 tri	25.4		
H-type 1 tetra	56.6	6TID	1.6
Lewis Y penta	53.2	7OLU	1.6
H-type 3-GloboH hexa	26.7	6TIG	1.9

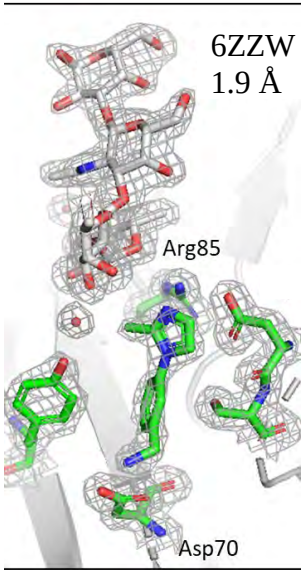








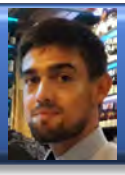




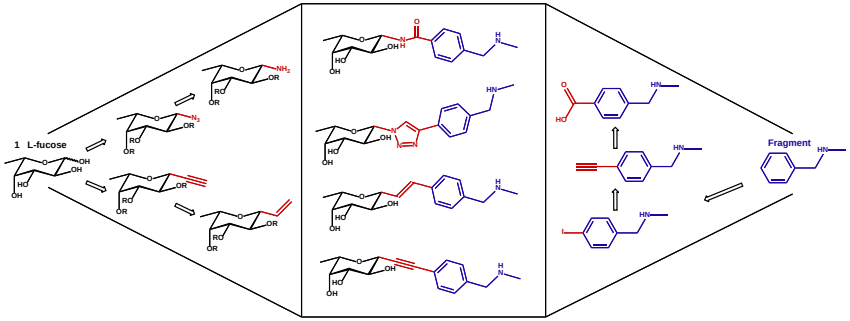
- 3 druggable sites predicted with SiteMap in Maestro

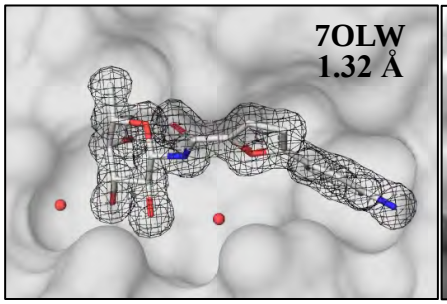
Sulak *et al*, PloS path, 2011, Bermeo *et al*, Molecules, 2020, 25:248; Lal *et al*, Chem. Eur. J. 2021, 27, 10341

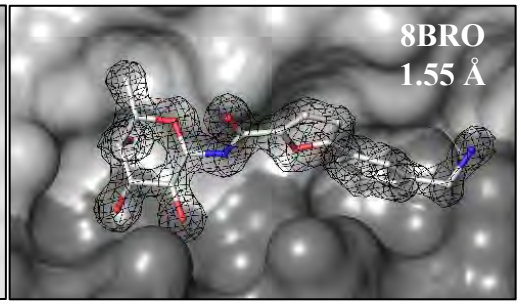
Bc2LC-nt inhibitors



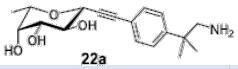
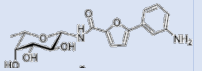
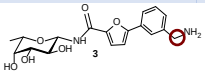
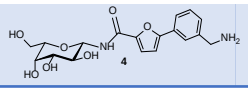
➢ Design, synthesis & evaluation of fucosides derivatives

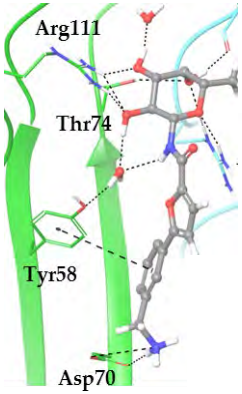




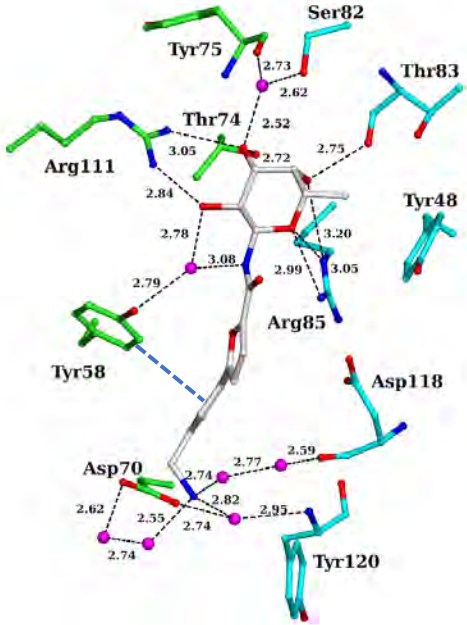


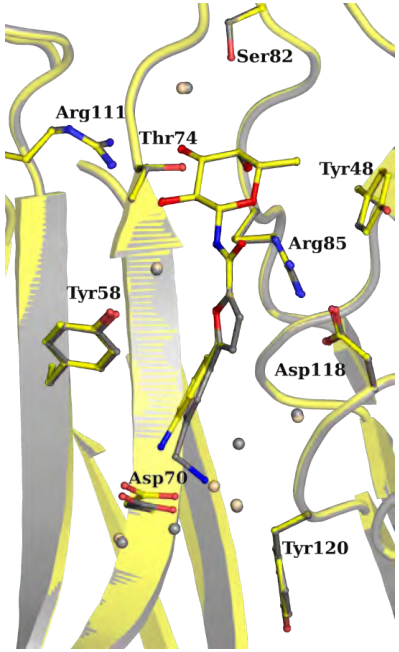
- Transform aniline in aminomethylene
  - One order of magnitude gain
  - Terminal amine does indirect Hbonds

Ligand	k_D (µM)	PDB	Res (Å)
L-Fuc 	2700		
22a 	280	7OLU	1.6
8c 	nd	7OLW	1.32
3 	159	8BRO	1.55
4 	390		



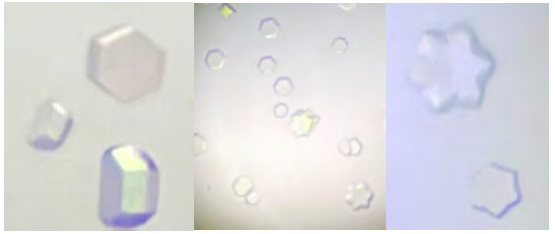
Modelling

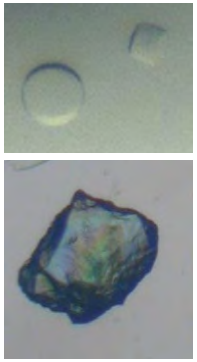




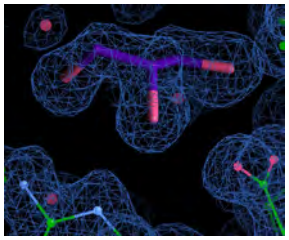
Bermeo, ACS Chem Biol, 2022; Mazzota, Molecules 2023; Antonini, Frontiers, 2023, in press

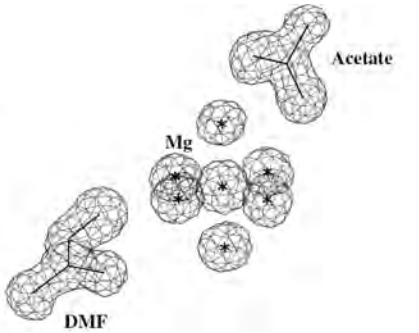
Miscellaneous





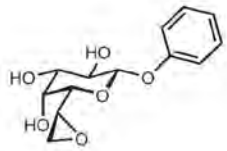
- Crystal shape can be misleading
  - Bc2lCnt 0.9-1.2 Trisodium citrate pH 7
    - Hexagons OK - flower not OK
  - Coda: 1 crystal after several month and broken coverslip
    - Not multiple, 1.3 Å

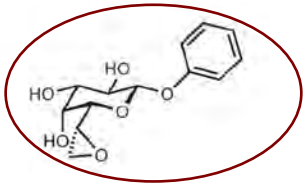


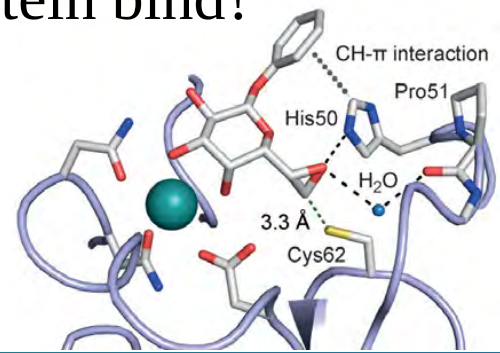


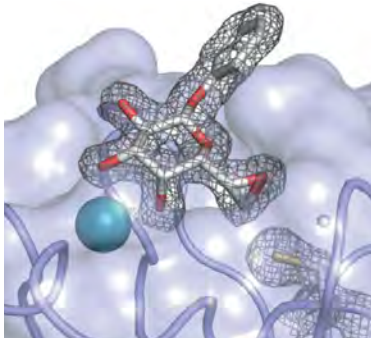
- Components from purification or crystallisation
  - Electron density of chemical can help
    - Glycerol from cryo in TrfbL1 native
    - Cel6A D405N-SDP5: 1OC7, 1.1Å

- Which diastereoisomer did the protein bind?
  - LecA with potential covalent epoxide
    - Crystal at unreactive pH : 4.5









Varrot, Structure, 2003; Gourdine, JBC, 2008; Wagner, Angewand Chem, 2017

Conclusions

- Still need for X-ray crystallography to gain access to protein-sugar interactions at the atomic level
  - Highest the resolution the better to draw accurate conclusions
    - Distorsion / High energy conformation
      - Real or artefact from user/program errors
  - Binding site location and architecture
    - → Structure based drug design
    - → Site directed mutagenesis: new specificity
  - Quaternary structure and multivalency
    - → Gain fondamental knowledge on protein folding and function
    - → Design new molecular tools

- Could be tricky for non glycobiologists
  - No « Structural glycobiology for dummies »
  - Do not hesitate to contact experts

Acknowledgements

CERMAV, Grenoble
Aymeric Audfray
Soorej Basheer
Rafael Bermeo*
Bertrand Blanchard
François Bonnardel
Aurore Cabanettes
Valérie Chazalet
Emilie Gillon
Anne Imberty
Sue Kuhaudomlarp
Atul Kumar
Mickael Lelimousin
Martin Lepsik
Jérémie Topin

NCBR, Brno, Czeck Rep
Michaela Wimmerova
Josef Houser

YSBL, York, UK
Gideon J Davies
Eleanor Dodson
Johan Turkenburg

UBC, Canada
Stephen G. Withers group

Laboratorium für Organische Chemie, Zürich
Andrea Vasella group

CERMAV Grenoble
Hugues Driguez
Sébastien Fort

ETH, Zürich, Switzerland
Markus Künzler
Therese Wohlschlager
Silvia Bleuler-Martinez
Mario Schubert

CRCINA, Nantes
Jacques Lependu
Adrien Breiman

IAB, Grenoble
Jean-Luc Coll
Benoit Busser

Bayreuth U. Germany
Lukas Perkams
Carolina Spies
Carlo Unverzagt

University of Milan, Italy
Anna Bernardi
Laura Belvisi
Kanhaya Lal*
Davide Ruggeri
Daniele Lanaro
Sarah Mazzotta
Giulia Antonini
Francesca Vasile

HIPS, Saarbrücken, Germany
Alexander Titz
Roman Sommer
Stefanie Wagner
Dirk Hauck
+++