Application of protein X-ray crystallography to molecular glycobiochemistry

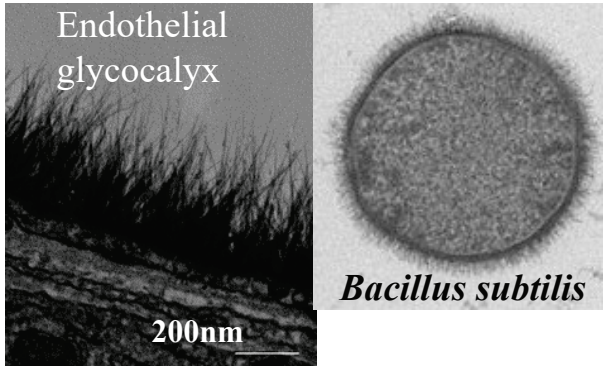
Dr Annabelle Varrot

« Structural Glycosciences » Summer School

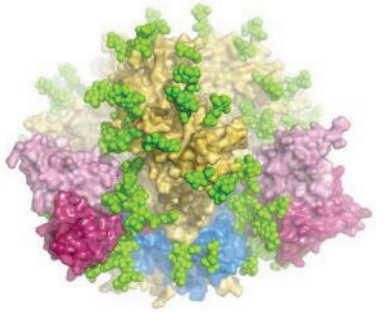
CERMAV, Grenoble 3rd July 2018

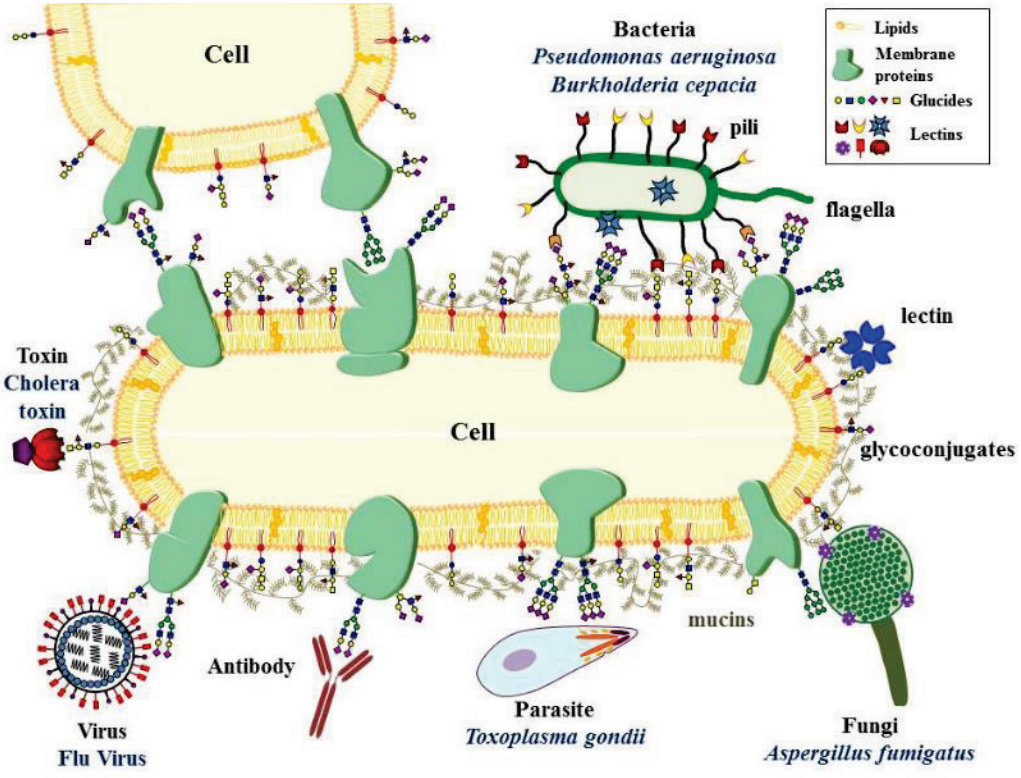
The “sweet side” of the cell surface

- Every cell surface is covered by **glycoconjugates**
- 1st cell information received and sent: **glycocode**
  - Essentials in cell identity, recognition and signaling



CryoEM, HIV GP120 glycan shield

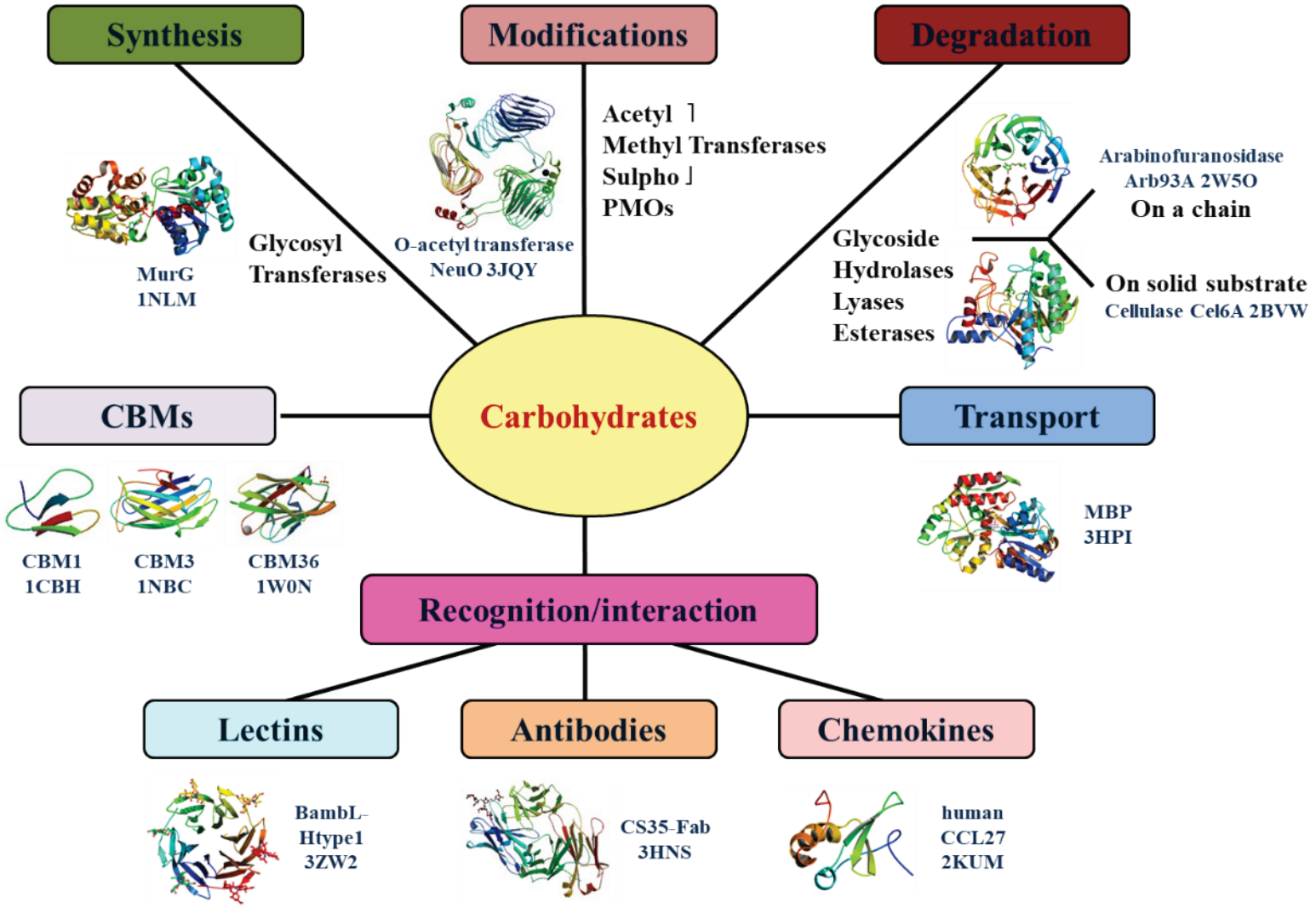




The Scripps Research Institute, Wikipedia, Varrot A & Imberty A Curr Op Struct Biol 2008 18:567

Cermav

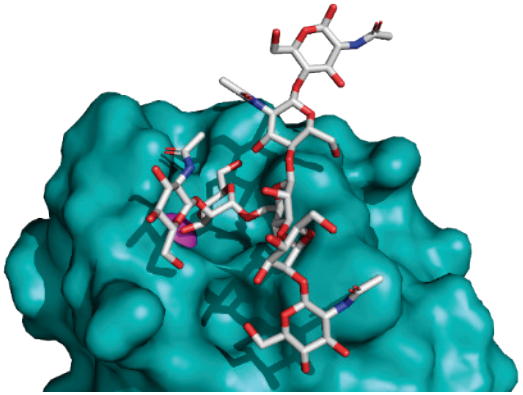
Protein-sugar interactions



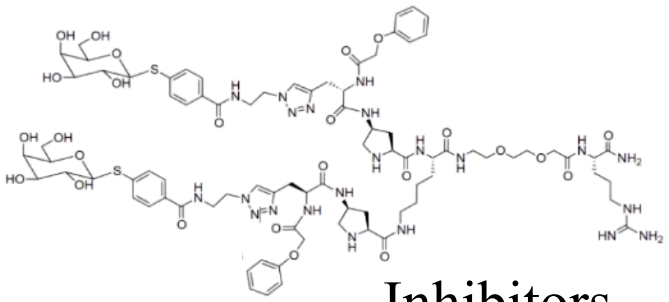
Sarkar A & Pérez S Structural Glycobiology CRC Press 2012 71

Cermav

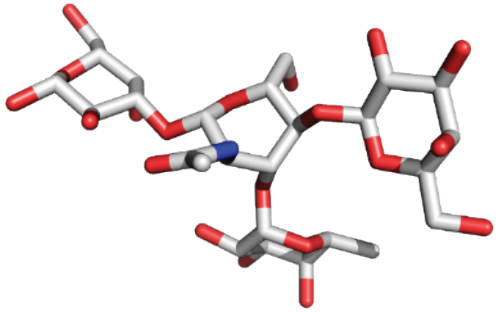
Which info do you get from crystallography?



Binding site architecture



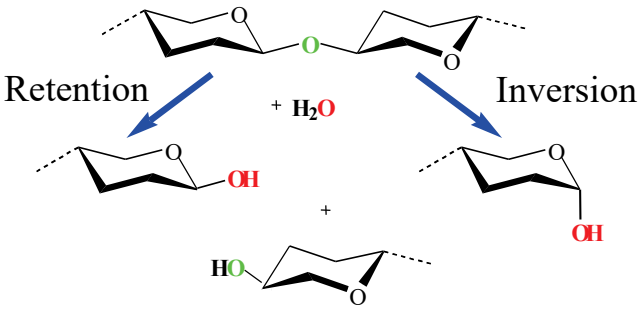
Inhibitors



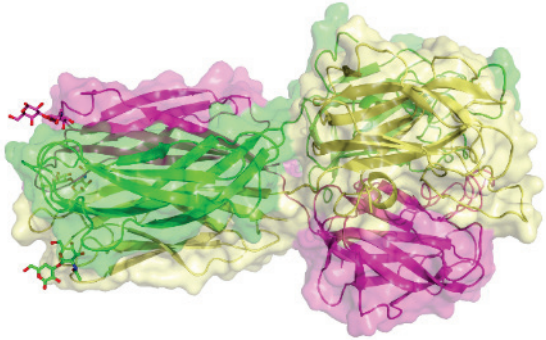
Ring conformation

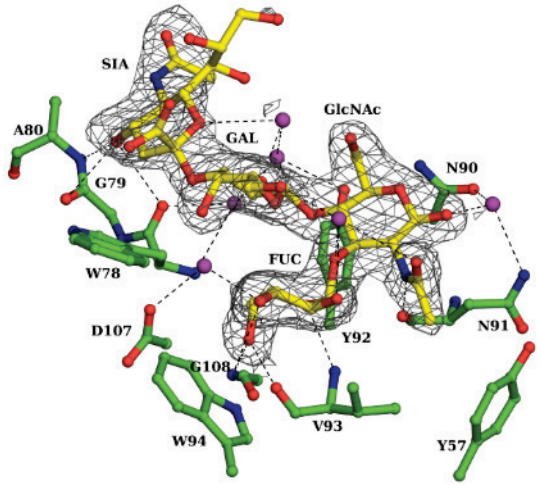
structure

Mechanism



Multivalency



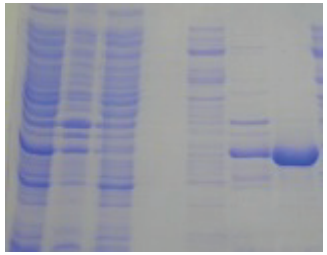


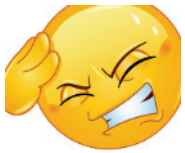
Interactions

Cermav

Protein X-ray crystallography

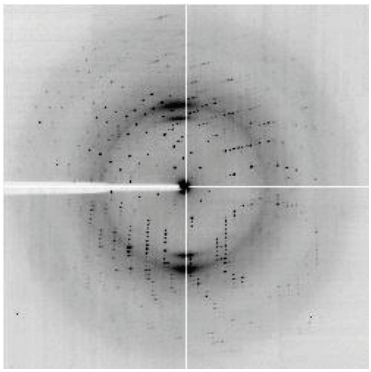
Expression
Purification



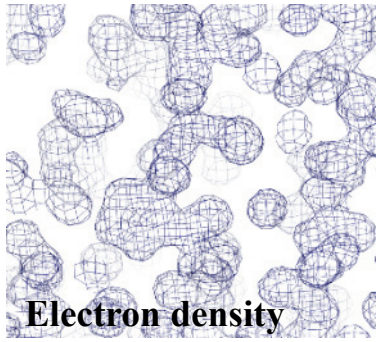


Crystallization

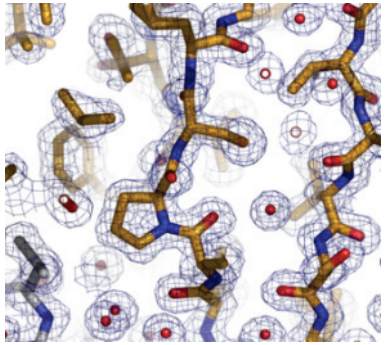
X-ray
diffraction



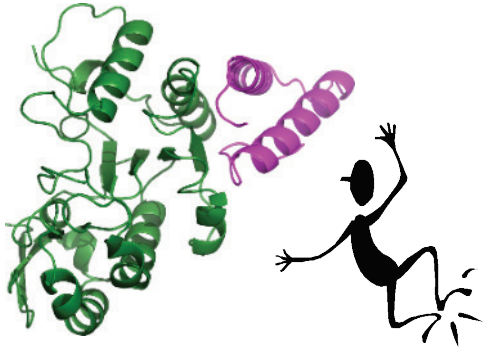
Phasing



Modelling



Refinement



➢Advantages

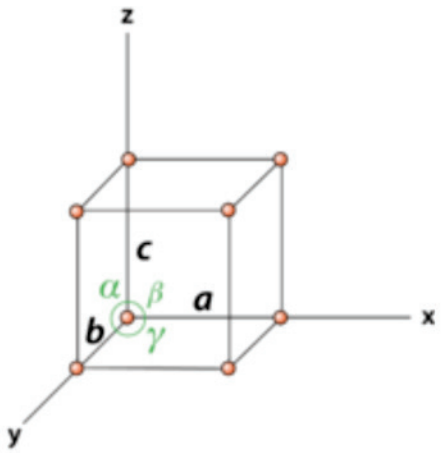
- High resolution structure
- Study of complex at the atomic level

➢Disadvantages

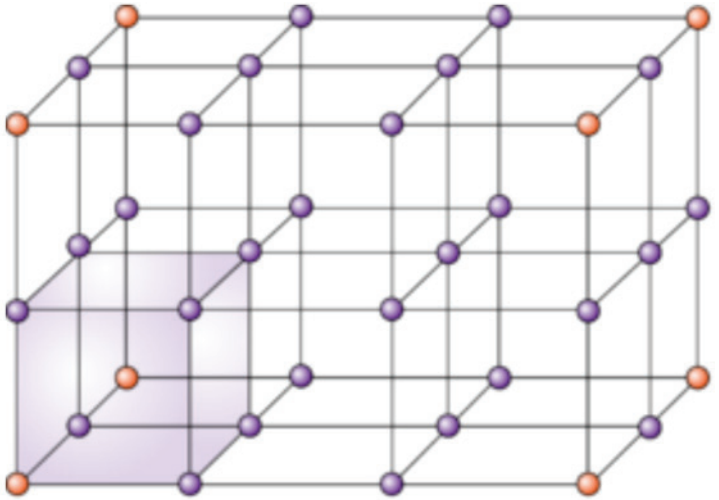
- Molecules in solid-state environment
- Require crystals

Cermav

What is a crystal?



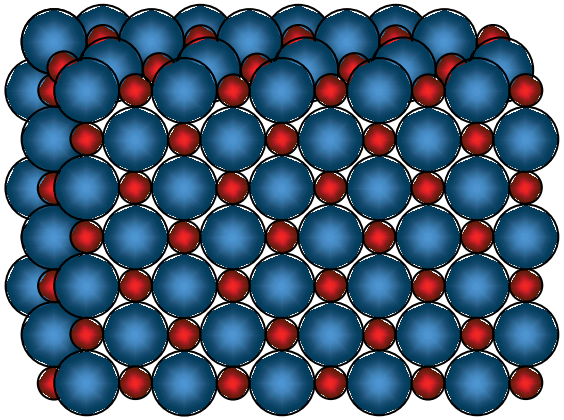
translation



Unit cell

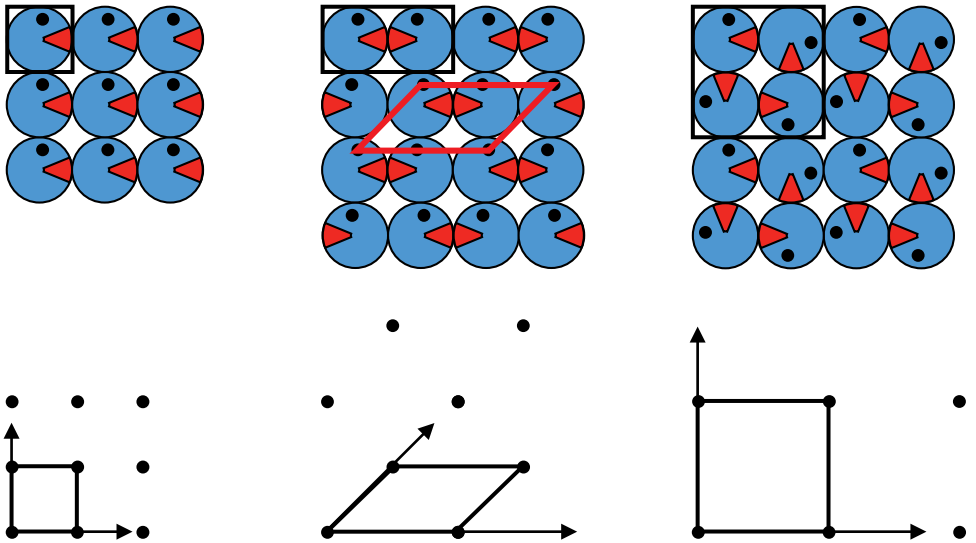
Crystal lattice

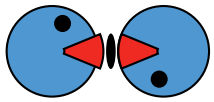
NaCl



Cermav

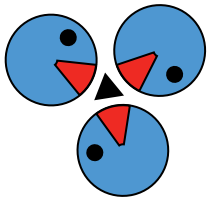
Crystal packing





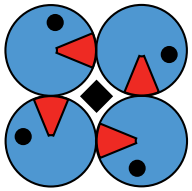
2 fold axis

2_1



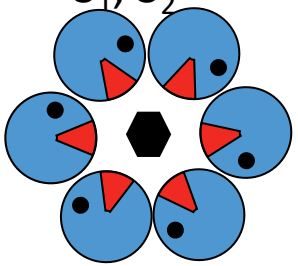
3 fold axis

3_1, 3_2



4 fold Axis

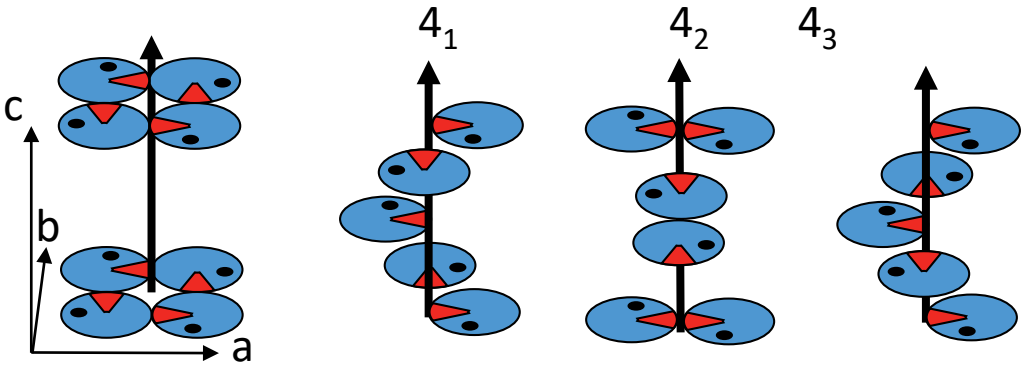
4_1, 4_2, 4_3



6 fold axis

6_1, 6_2, 6_3, 6_4, 6_5

Screw axis

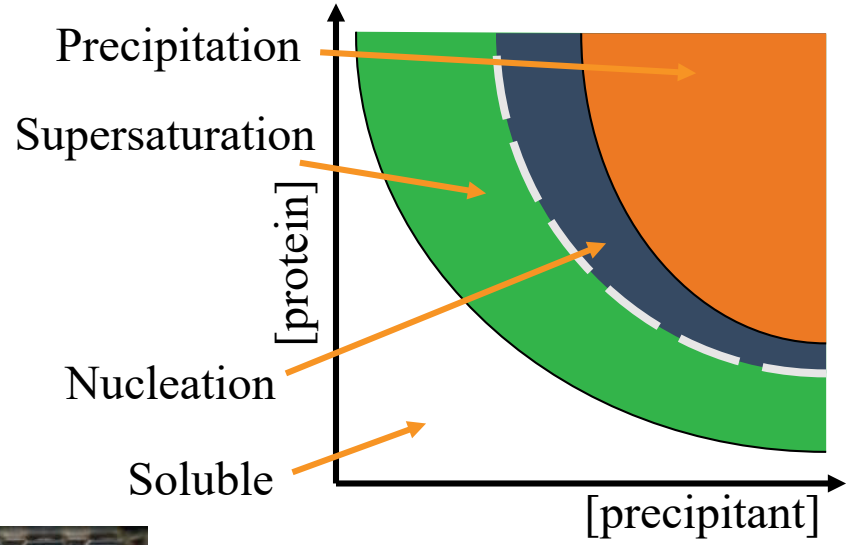


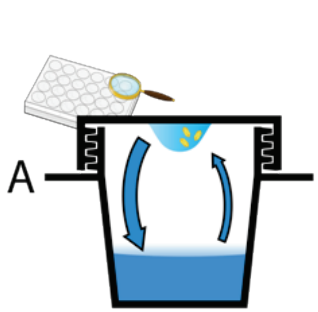
65 spacegroups for protein crystals

Cermav

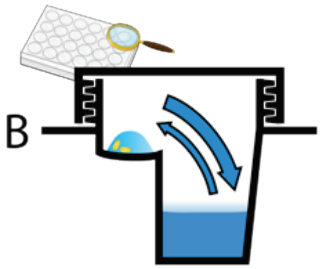
Protein crystallogenesis

- Need pure and homogenous 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

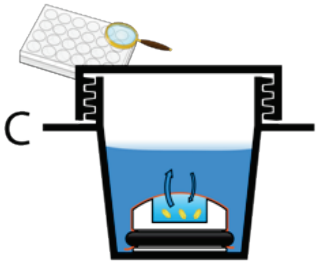




Hanging drop



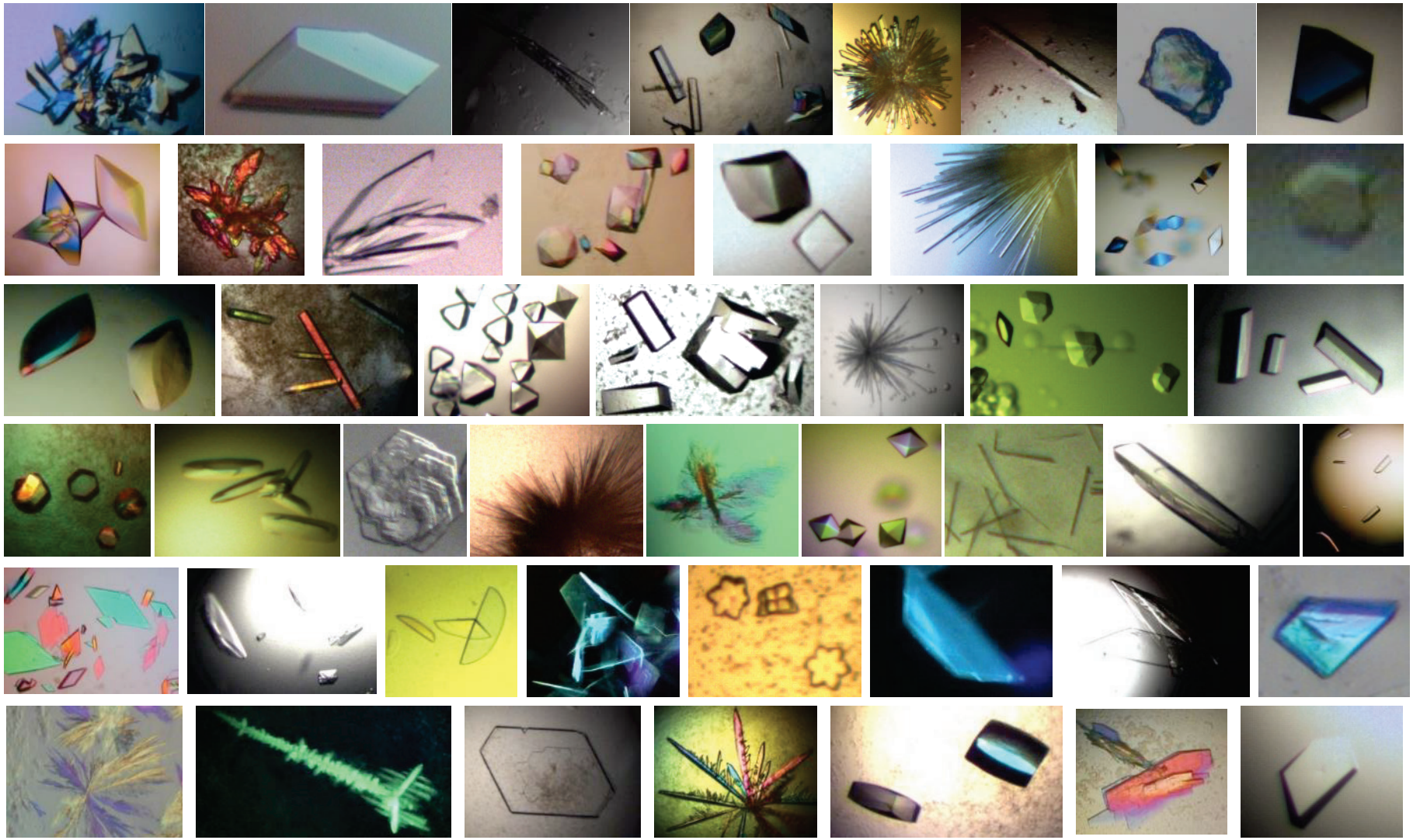
Sitting drop



microdialysis

Cermav

CERMAV protein crystals



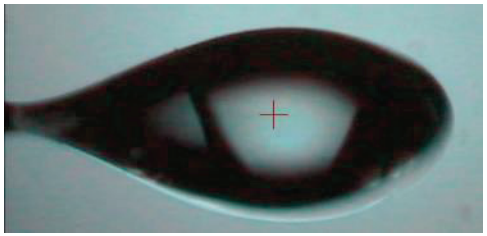
Cermav

How to obtain crystals of protein complexes with ligand

Soaking

- Add the ligand to the reservoir solution
  - Add concentrated solution to limit dilution
  - Transfer the crystal in solution with ligand
  - Add bit of powder

- Try different concentrations and soaking time
- Can soak the crystal after data collection
  - For resistant crystals diffracting well
  - BambL: lectin from *Burkholderia ambifaria*
    - Soaking with 2 mM methyl-α-L-fucopyranoside (MFU)
      - Data collected at 1.3 Å BM14, ESRF: empty
    - Resoaked with 20mM
      - Data collected at 1.3 Å BM14: still empty

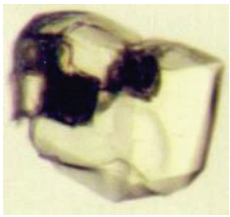


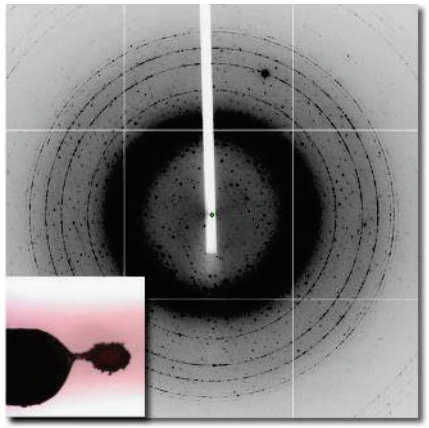
Audfray A JBC 2012 287:4335

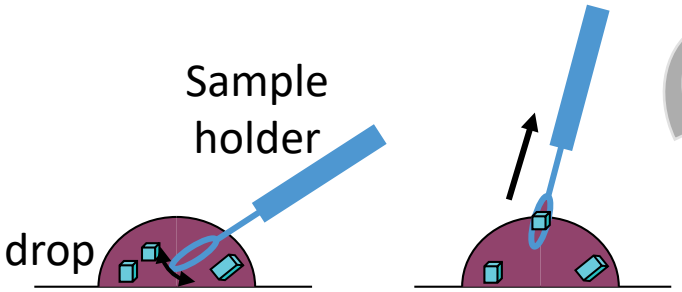
Cermav

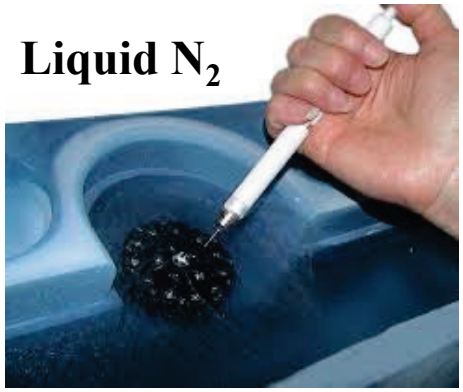
Soaking while freezing

- Freezing limits radiation damages
- Add ligand to the cryoprotectant solution
  - Less chance of replacement by glycerol, ethylene glycol and MPD

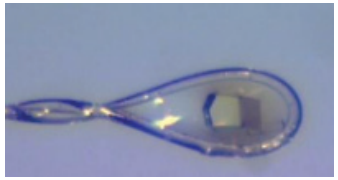


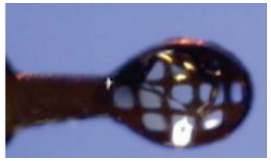






Hampton **Litholoop Molecular Dimensions** **Mitigen**



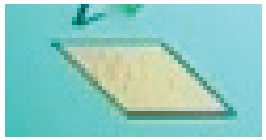
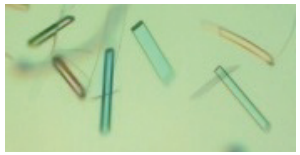


https://www.youtube.com/watch?v=FENUWRYXMOM

Cermav

Cocrystallization

- Excess of pure ligand > 2X K_D
- Put the ligand in reservoir
- Or preincubation with the protein
  - Limit protein dilution: >1/10
  - Set duration and temperature
  - Easily reproducible
  - Require little ligand
- Can have to screen for each complex: LecA

Gal	**Galα1-2Galβ-O-Me**	**Galα1-3Galβ1-4Glc**	**Galα1-6Glc**	**Inhibitor**
1.5 M $(NH_4)_2SO_4$ pH 4.7, 5% MPD, 2% glycerol	20% PEG6K 1 M LiCl 0.1 M NaAc pH 4	10% PEG 5Kmme 25 mM KSCN 0.1 M NaAc pH 4.6	20% PEG 2Kmme, 0.2 M KBr 0.1 M NaAc pH 4.6	0.8 M Li_2SO4 pH 4.5
				

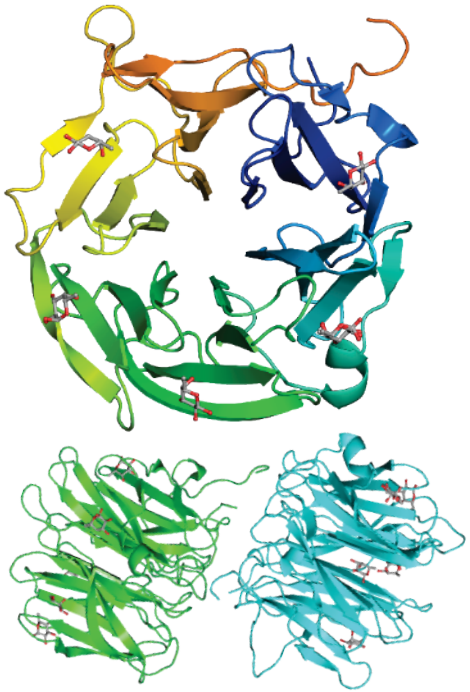
Blanchard B JMB 2008 383:837, Blanchard B Proteins 2014 82:1060; Novoa A Angewandte Chemie 2014 53:8885

Cermav

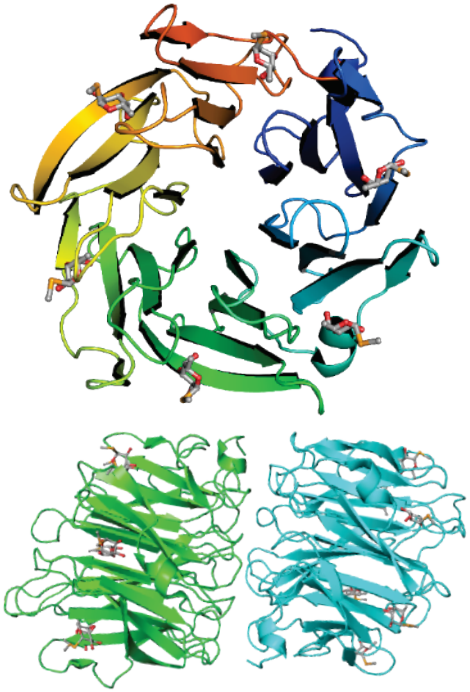
Phasing: molecular replacement

- Have a good model from homologous protein
  - Sequence identity > 25%
  - Good conservation secondary structure: difficult for β-sheets

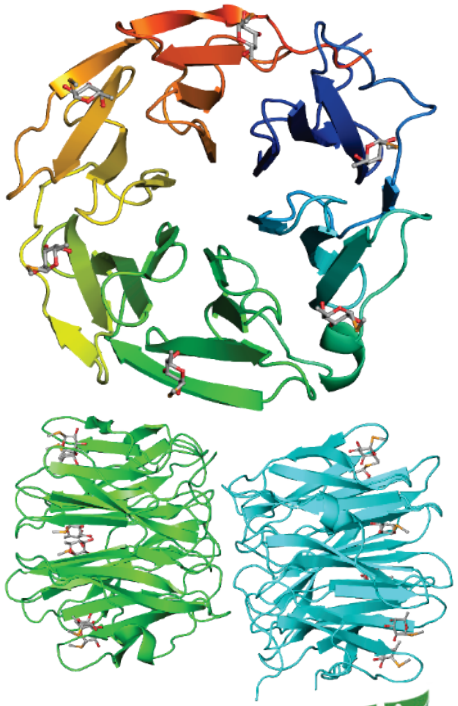
AAL/1OFZ/Hg
Aleuria aurantia



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



Trfbl1/not deposited/SFU (32/43%)
Terfezia boudieri

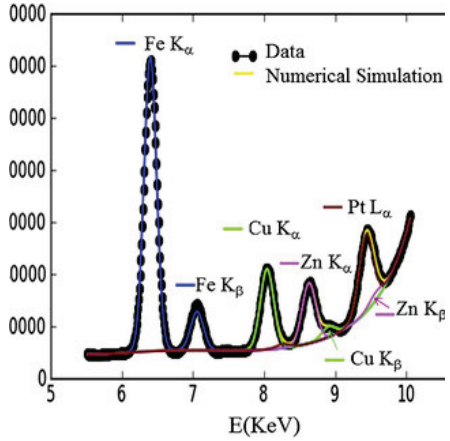


Wimmerova M, JBC 2003 278:27059; Houser J Plos One 2013 8:e83077; Basheer S, unpublished

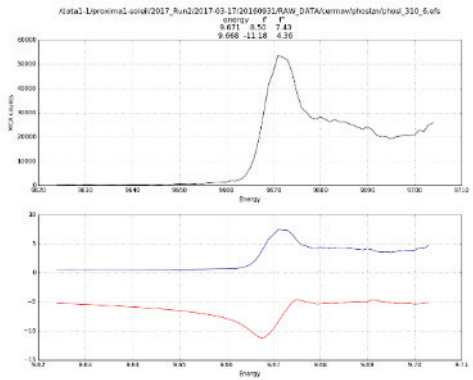
Cermav

Phasing: isomorphous or anomalous

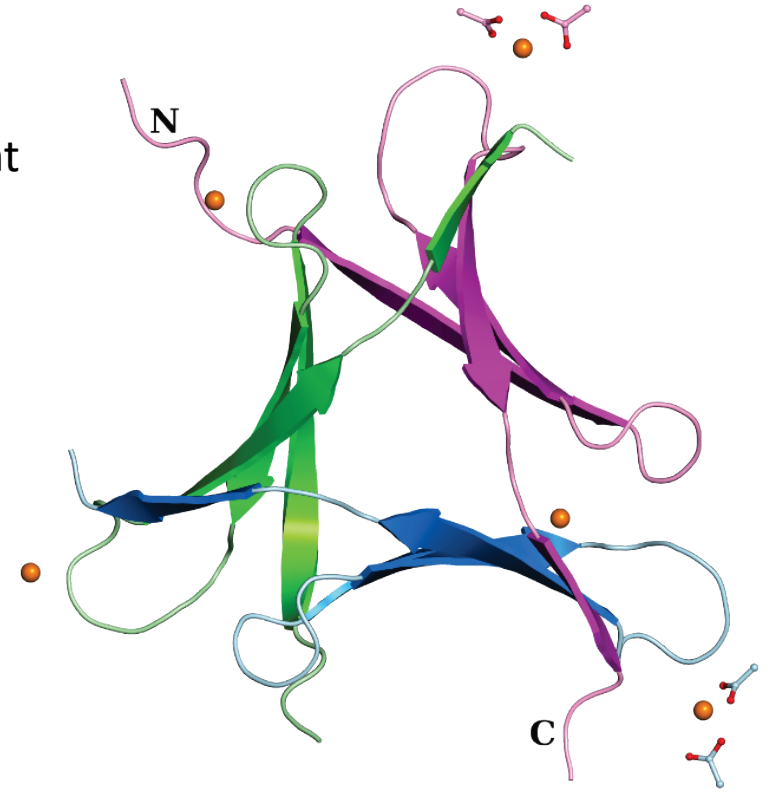
- Look in your crystallization conditions
  - PhoSL from *Pholiota squarrosa*
    - 300 mM Zinc acetate, 0.1 M Imidazole-HCl pH 6-7



X-ray fluorescent
XRF scan



Absorption
scan



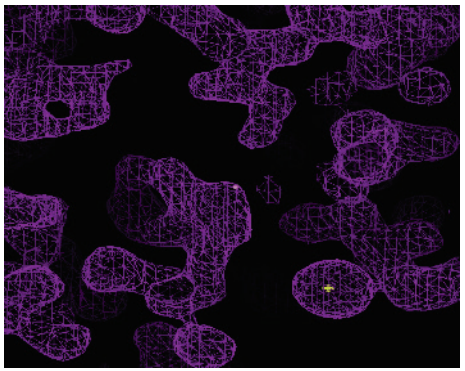
Cabanettes A Angewandte Chemie 2018 in press

Cermav

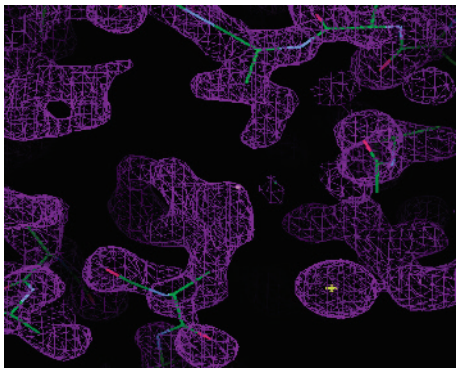
Phasing: isomorphous or anomalous-2

- Use selenomethionin protein, heavy metals
- Use ligand with heavy atoms (Se, Br, S, F)

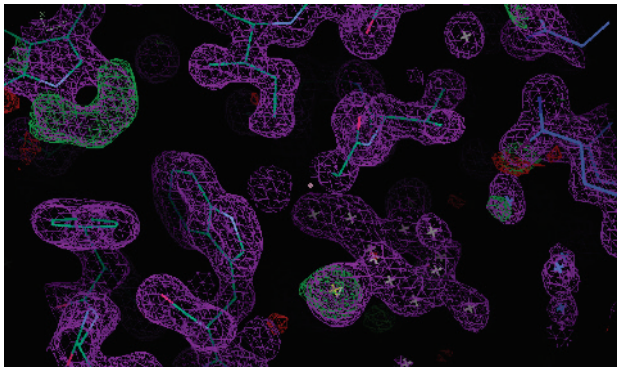
Se peaks



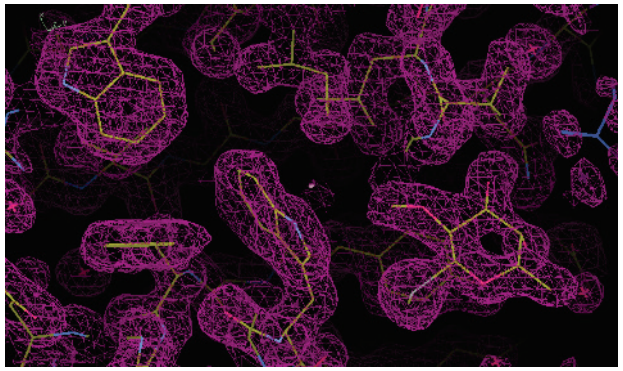
PolyAla building

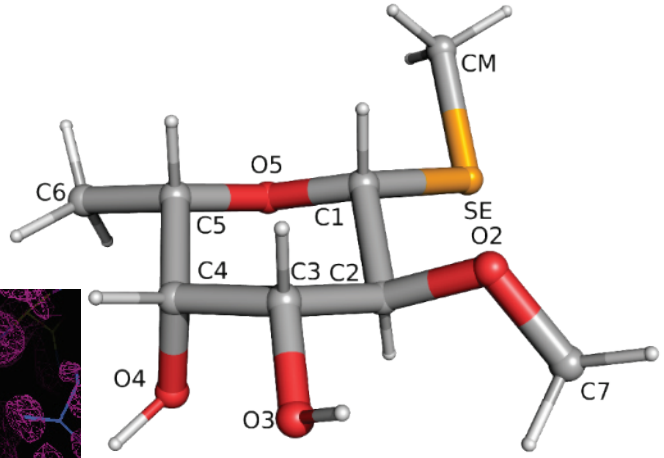


Sequence assigned



Final





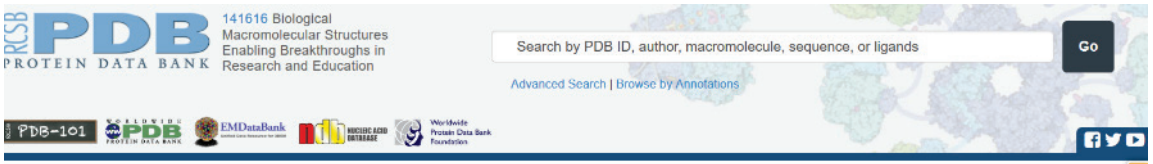
2-O-Methyl-β-methyl-L-selenofucoside
2MeSeFuc

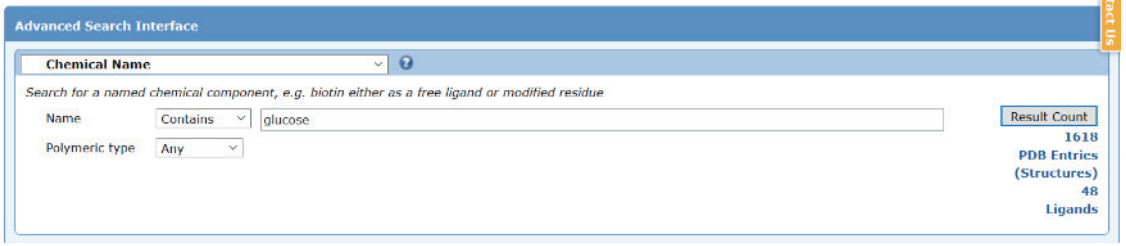
Buts L Acta Cryst D 2003 59:1012-1015, Sulak O Structure 2010 18:59-72; Sommer R Beilstein J Org Chem 2016 12:2828

Cermav

Refinement

- 3-letter code for each monosaccharide and protein residues
  - Check ligand database
    - HIC-Up: http://xray.bmc.uu.se/hicup/
    - PDB: http://www.pdb.org/pdb/home/home.do





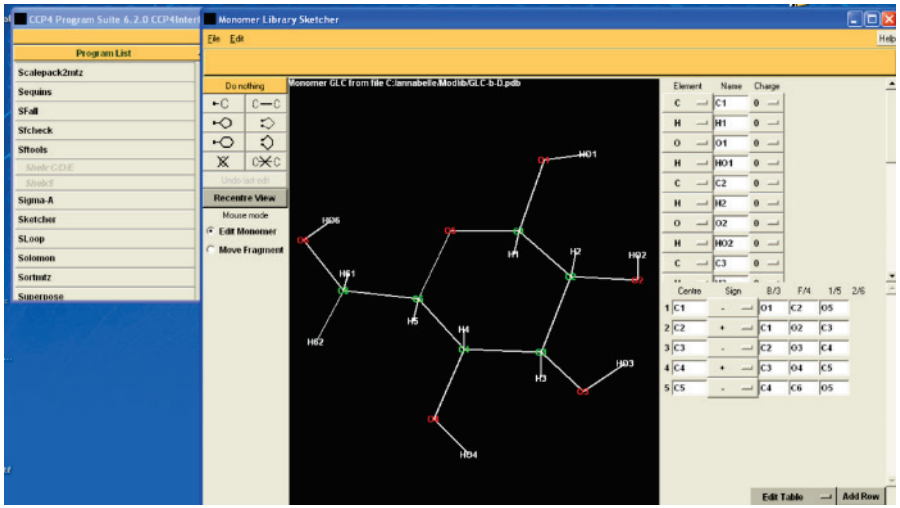
Ligand Structure	ID / Formula / Name	Structures with Specific Ligands
	GLC C6 H12 O6 ALPHA-D-GLUCOSE	686 PDB Structures contains GLC (1A47,1AC0,1ACZ ...)
	GLD C6 H12 O4 4,6-DIDEOXYGLUCOSE	9 PDB Structures contains GLD (1E3Z,1UA7,1UH3 ...)
	GLF C6 H11 F O5 1-FLUORO-ALPHA-1-DEOXY-D-GLUCOSE	6 PDB Structures contains GLF (1CXL,3IJ7,3SCO ...)
	GLO C6 H12 O6 D-GLUCOSE IN LINEAR FORM	16 PDB Structures contains GLO (1AC0,1EZ9,1FBO ...)

- Main monosaccharide have same code than common name
- 1 code for each anomer
  - α-D-Glucose: GLC
  - β-D-Glucose: BGC

Cermav

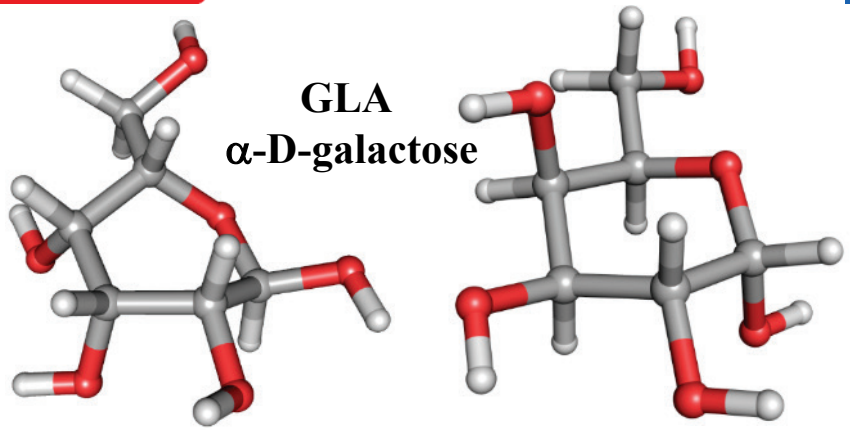
Refinement : ligand library

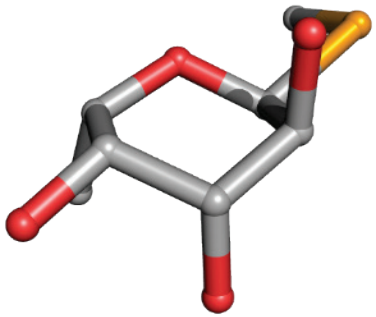
- Contains atom type and name, tree, distances, angles, torsions, chiral center, planes
- Creating new ligand
  - Old: Jligand or Sketcher
  - New: Acedrg



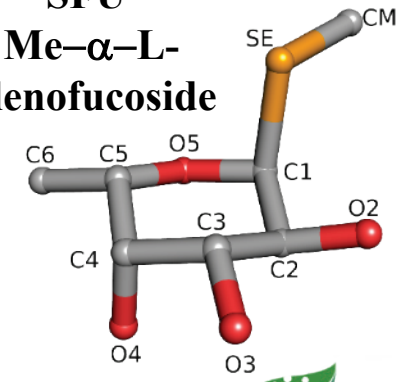
Check library is correct: no distorsion

GLA
α-D-galactose





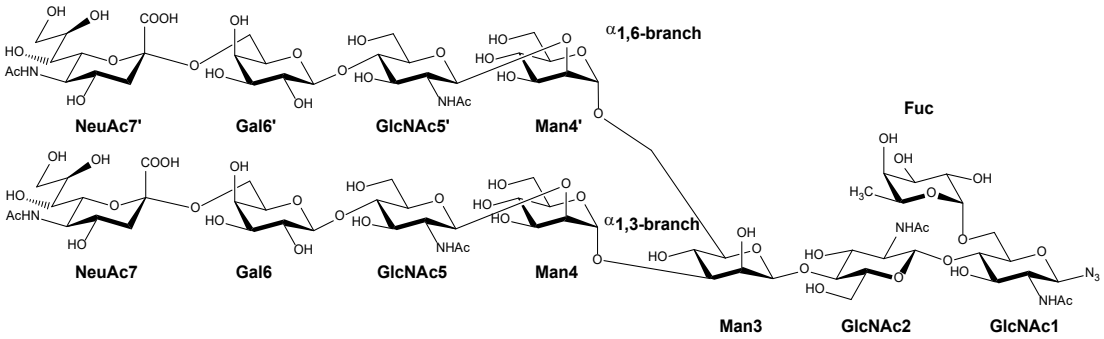
SFU
Me–α–L-selenofucoside

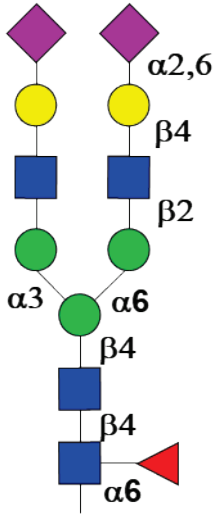


Cermav

Refinement

- Add description of the linkages
  - code could depend on program






```
LINKR        O6   LOG A 102      C1   FUC A 101      BETA1-6
LINKR        O4   LOG 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
```

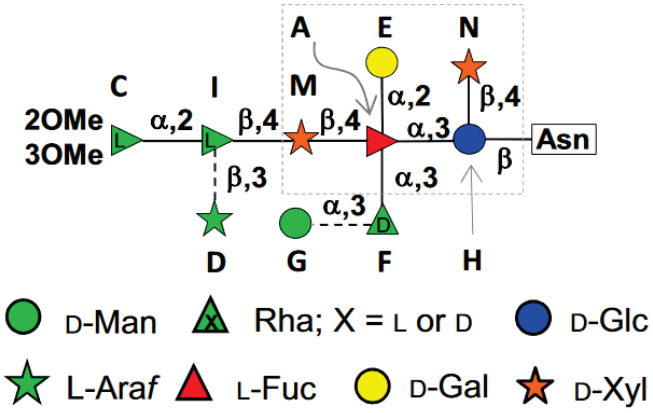
- Programmers do not know how to deal with L-sugars

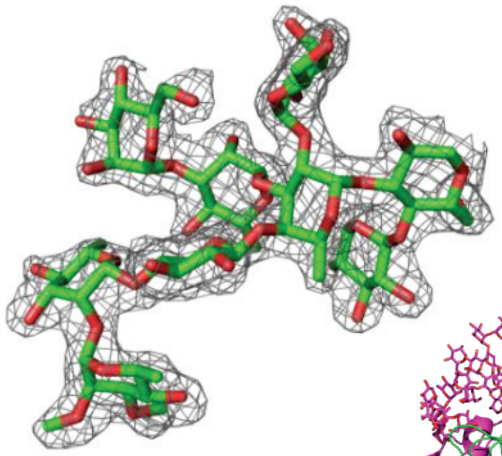
Cabanettes A Angewandte Chemie 2018 in press

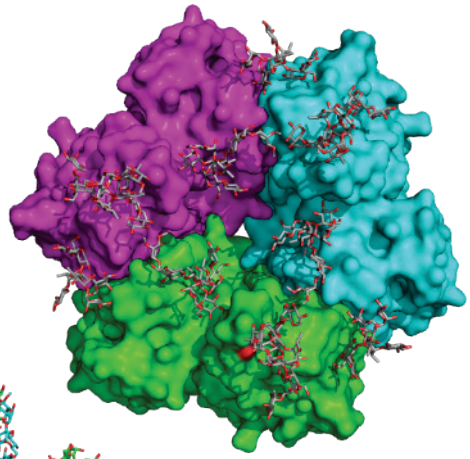
Cermav

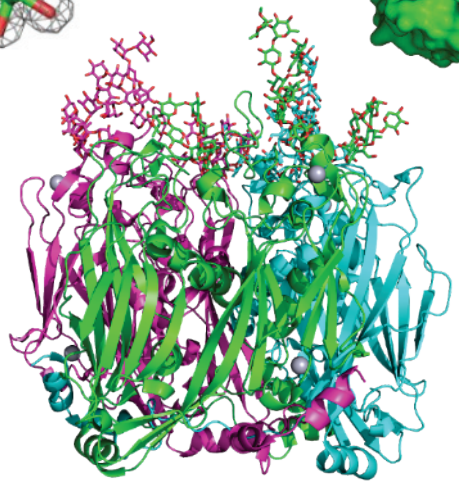
Can be tricky for non glycobiologist

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









4 glycosylation sites

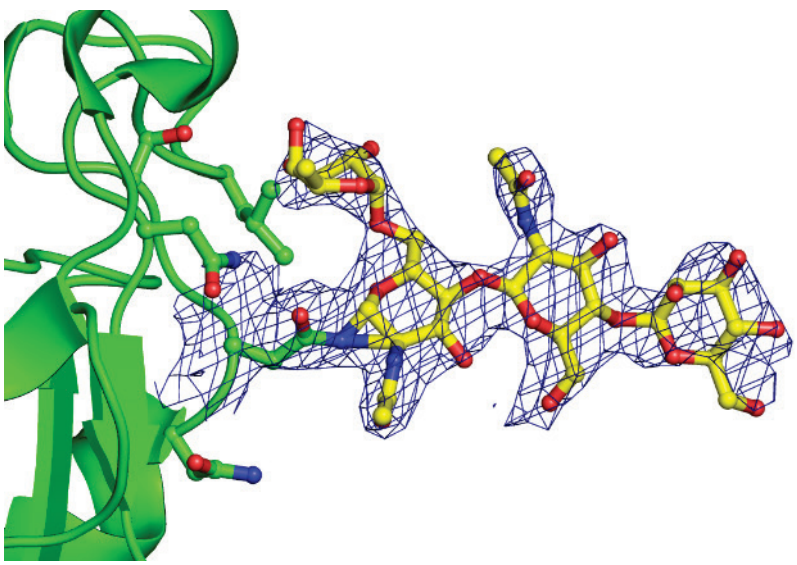
H = BGC
N/M = XYP
F = XXR
I = RM4
A = FUC
E = GLA
G = MAN
C = 7CV

Nandhagopal N PNAS 2002 99:14758, De Castro C PNAS 2013 90:2840; De Castro C PNAS 2017 115:E44

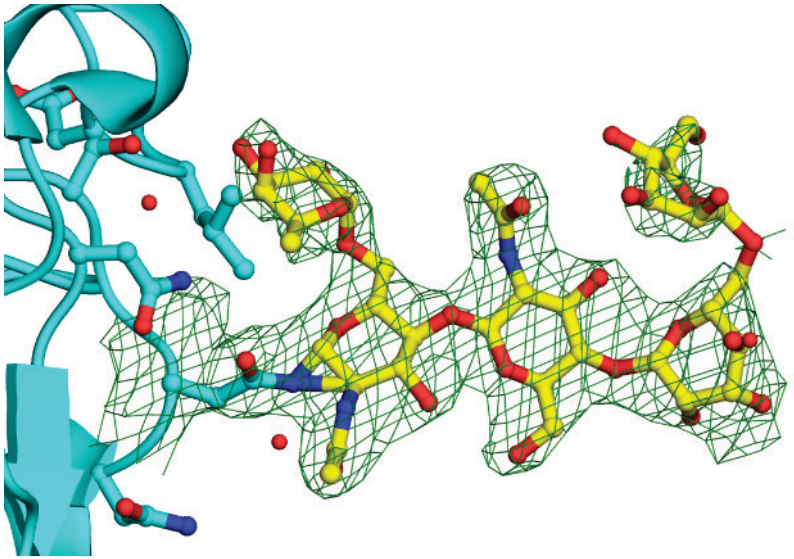
Cermav

Carbohydrate validation

- Do not overfit at low resolution



2.8 Å

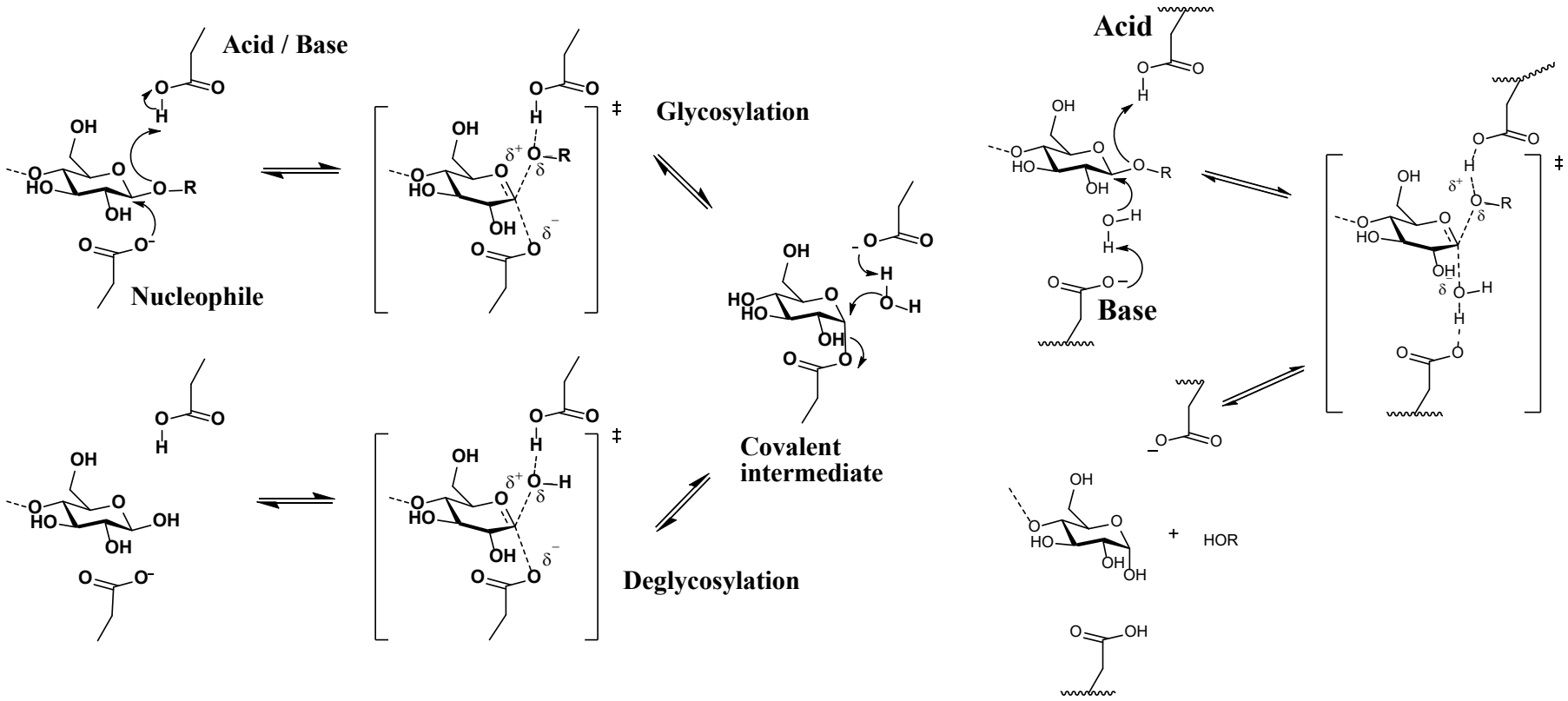


- Check nomenclature and ring conformation
  - PDB-care
  - Privateer: ccp4

Lütteke T BMC Bioinformatics 2004 5:69; Agirre NSMB 2015 22:833; Agirre J Acta Cryst D 2017 D73:171

Cermav

Glycoside hydrolases mechanism



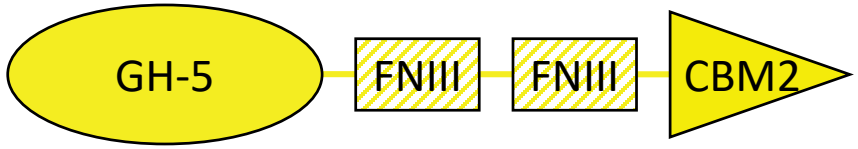
- Try to trap each step by X-ray crystallography
  - Use specific ligand, mutated protein, inactive pH

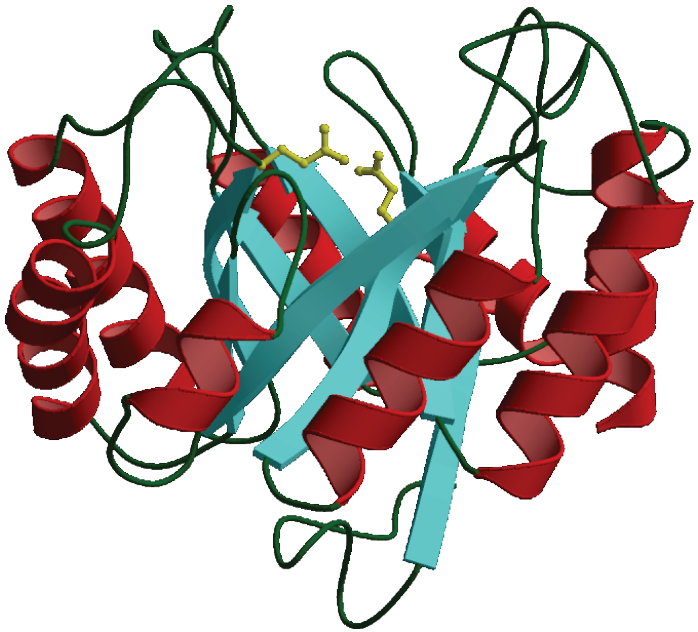
Koshland Biol Rev 1953 28:416

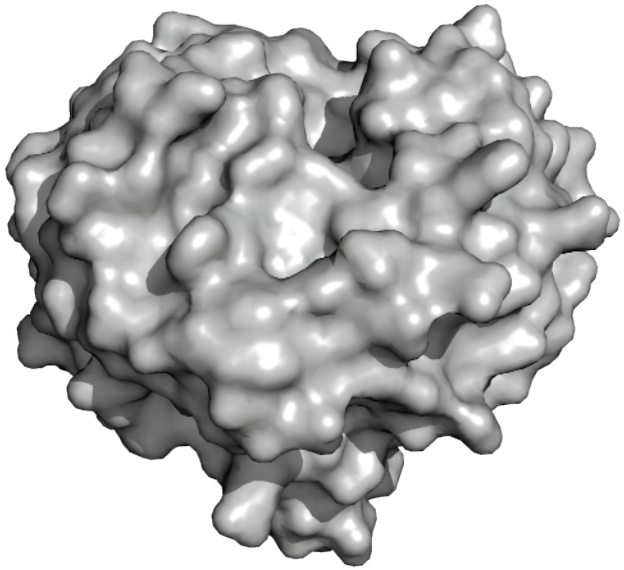
Cermav

Retaining mechanism

- Cellulase Cel5A from *Bacillus agaradhaerens*
  - Modular endoglucanase
  - Active pH range 5.0-13.0
  - Glu139 (acid/base) and Glu228 (nuc)
  - Shallow active site cleft with 5 subsites



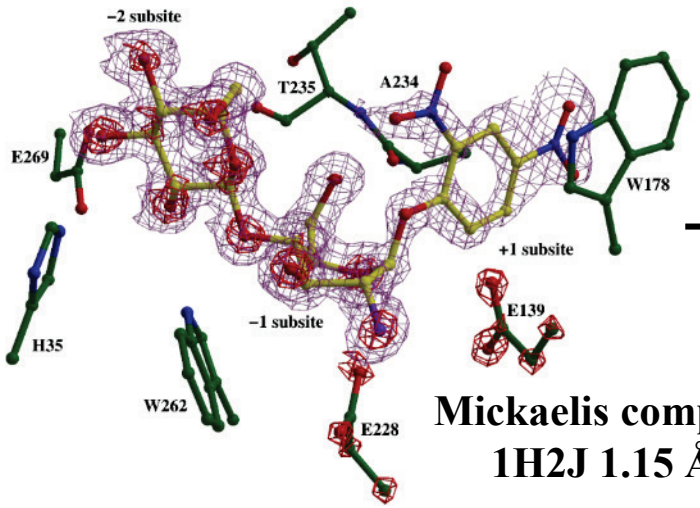




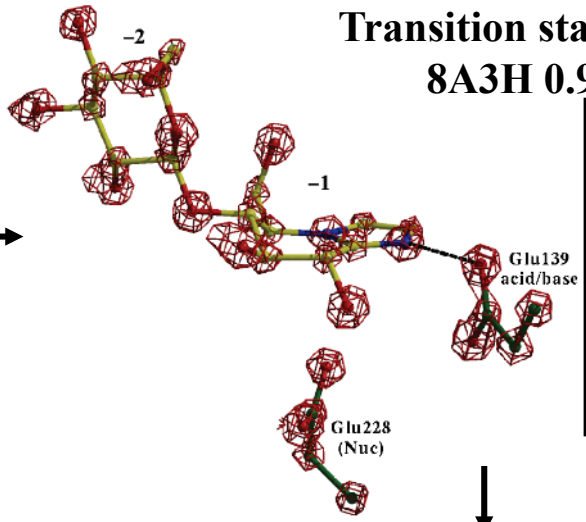
Davies GJ Biochemistry 1998 37:1926

Cermav

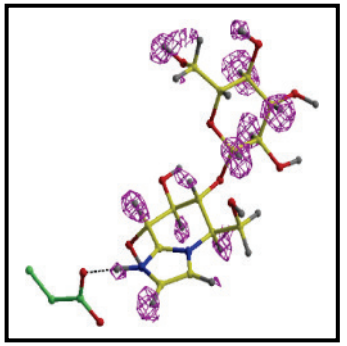
Cel5A snapshot at atomic resolution

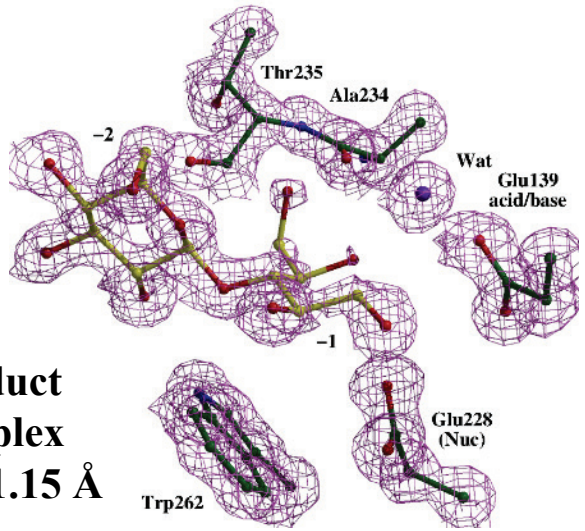


Mickaelis complex
1H2J 1.15 Å

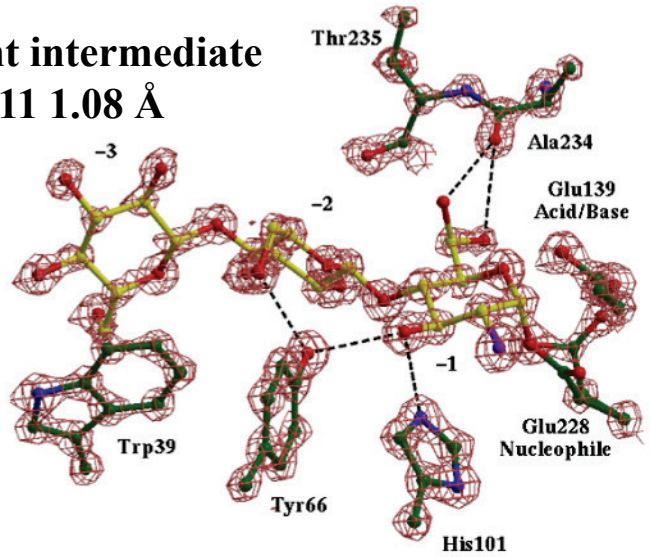


Transition state mimic
8A3H 0.97 Å





Product complex
1HF6 1.15 Å



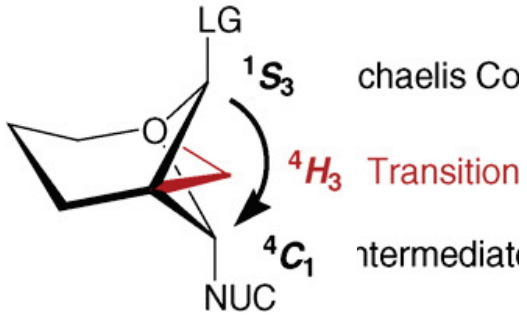
Covalent intermediate
1h11 1.08 Å

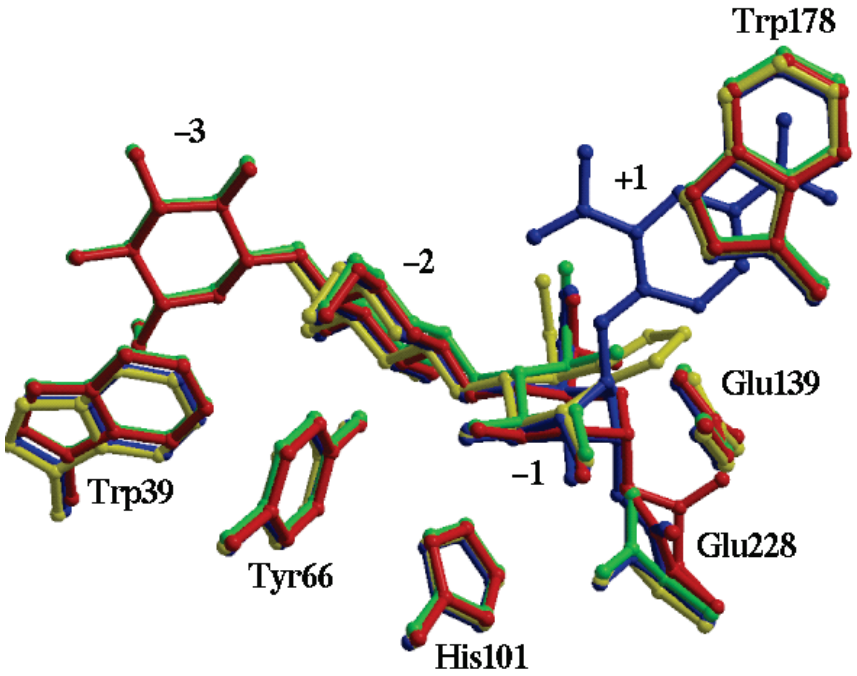
Varrot A JACS 1999 121: 2621; Varrot A Acta Cryst D 2001 D57: 1739, Vocadlo D, Nature 2001

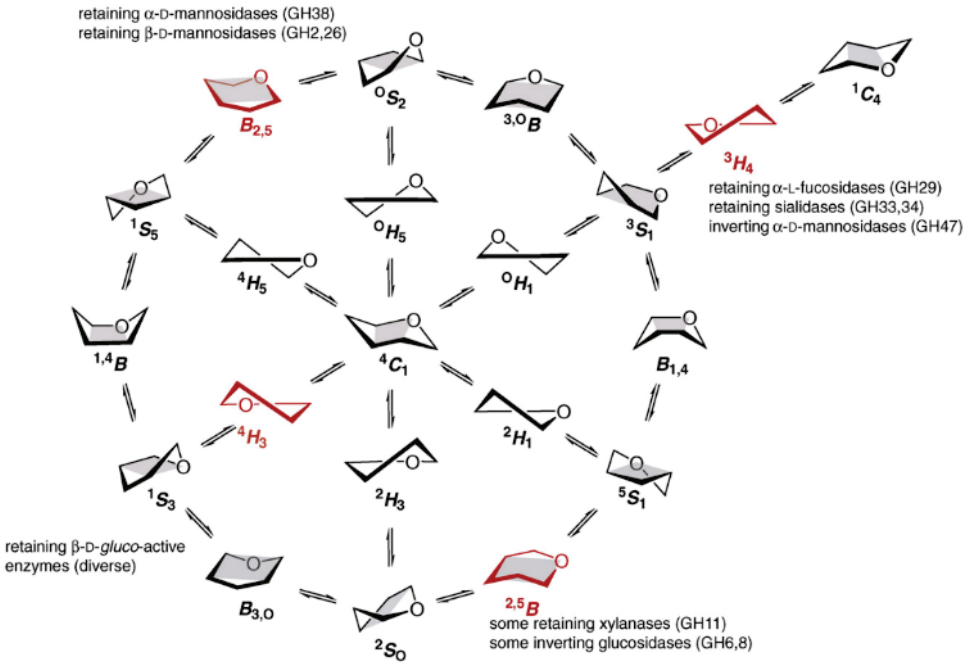
Cermav

Importance of ring distorsion

- Protein quite static
- Nucleophilic migration





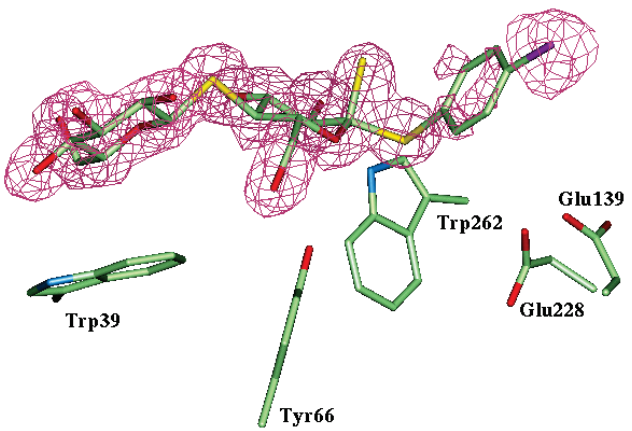


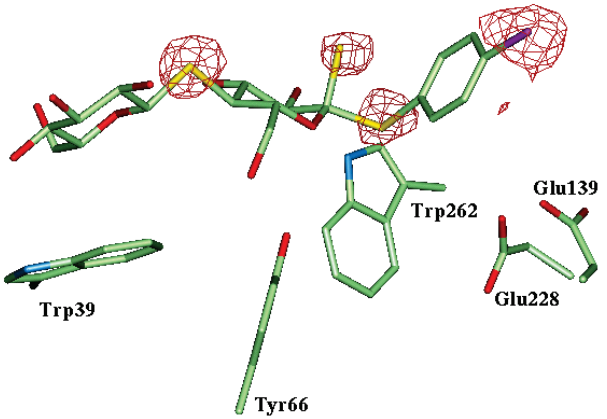
Varrot A Acta Cryst D 2001 D57: 1739; Vocadlo and Davies COSB 2008 12:539

Cermav

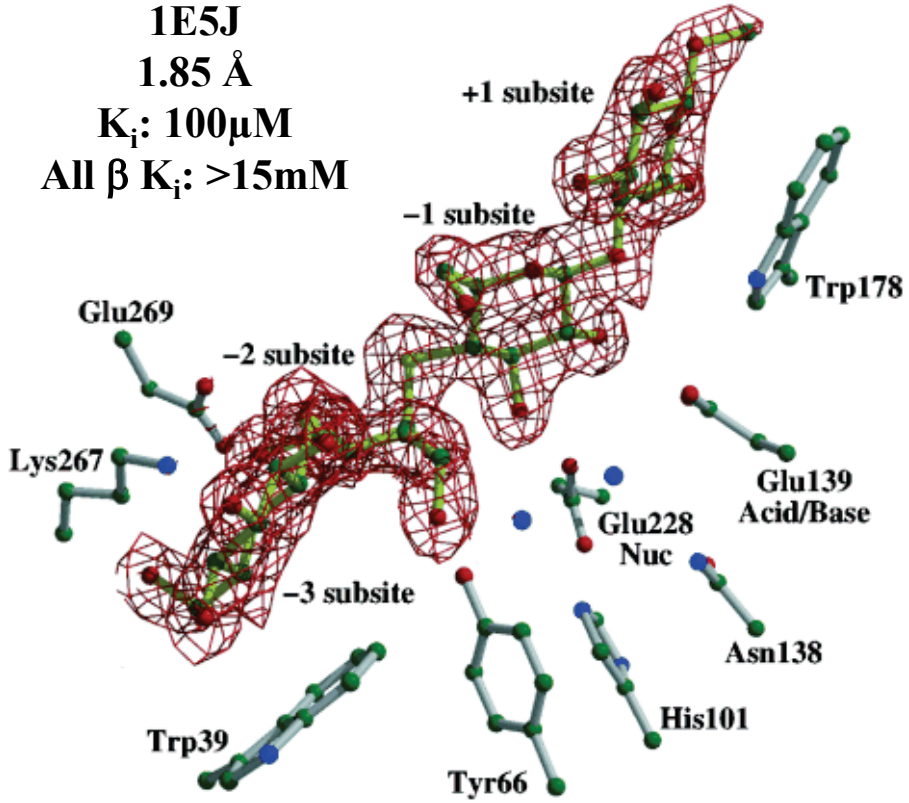
Serendipiry

- Strange density with TBR 1.68 Å on home source
  - New inhibitor class





1E5J
1.85 Å
K_i: 100µM
All β K_i: >15mM

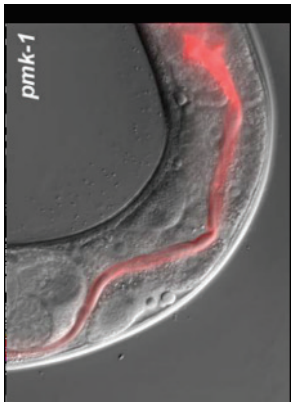


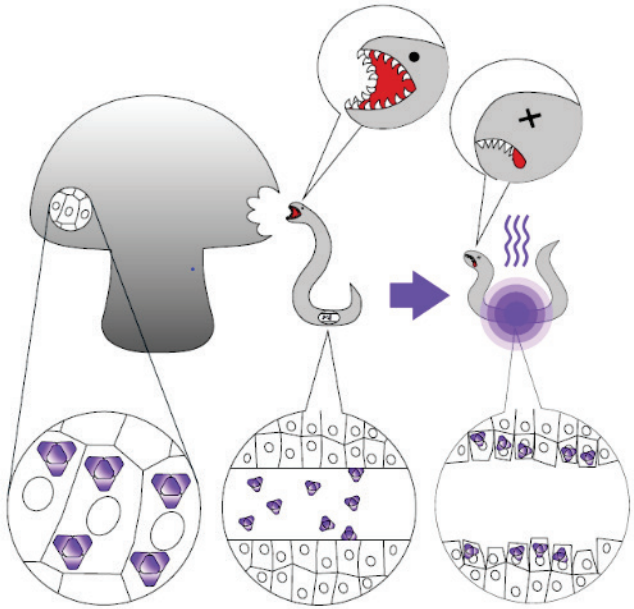
Unpublished and Fort S ChemBiochem 2001 2: 319

Cermav

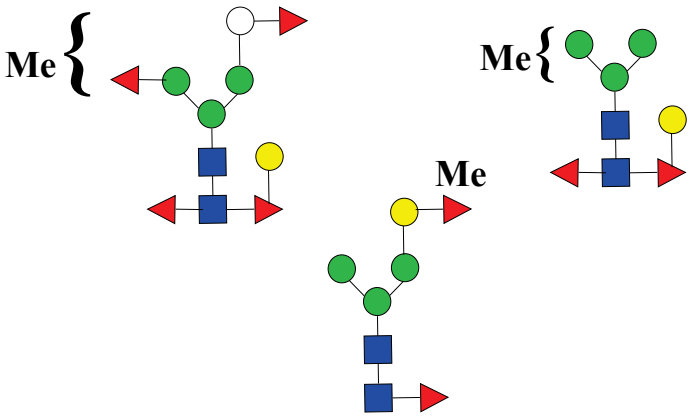
Lb-Tec2 from *Laccaria bicolor*

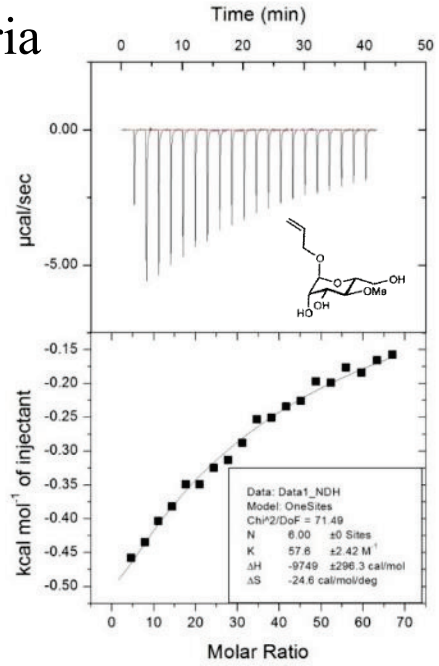
- Defense protein
  - Agglutinates Gram- bacteria
  - Nematotoxic
    - Binds to the worm gut





- Recognize O-methylated sugars
  - On worms, LPS bacteria
  - None in mammals





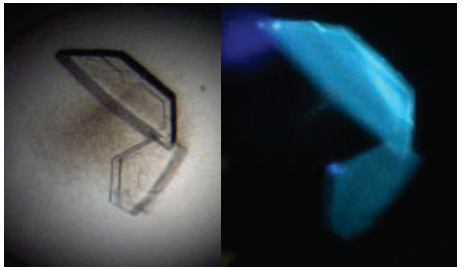
Sugar	K_d (mM)
Fuc	75
Allyl-2MeFuc	4
Allyl-3MeMan	20
Allyl-4MeMan	16

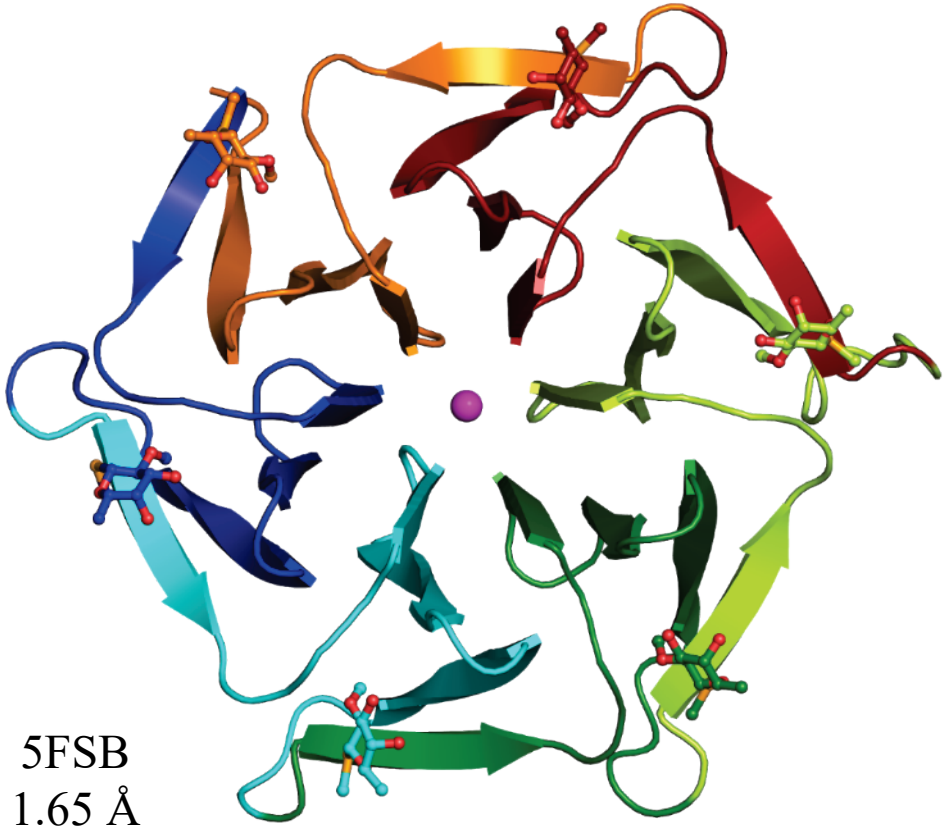
Wohlschlager T PNAS 2014 E2787, Staudacher E Biol Chem 2012 393:675; Yan S Mol Cell Proteomics. 2015 14:2111; Cordara G Structure 2018 26:369

Cermav

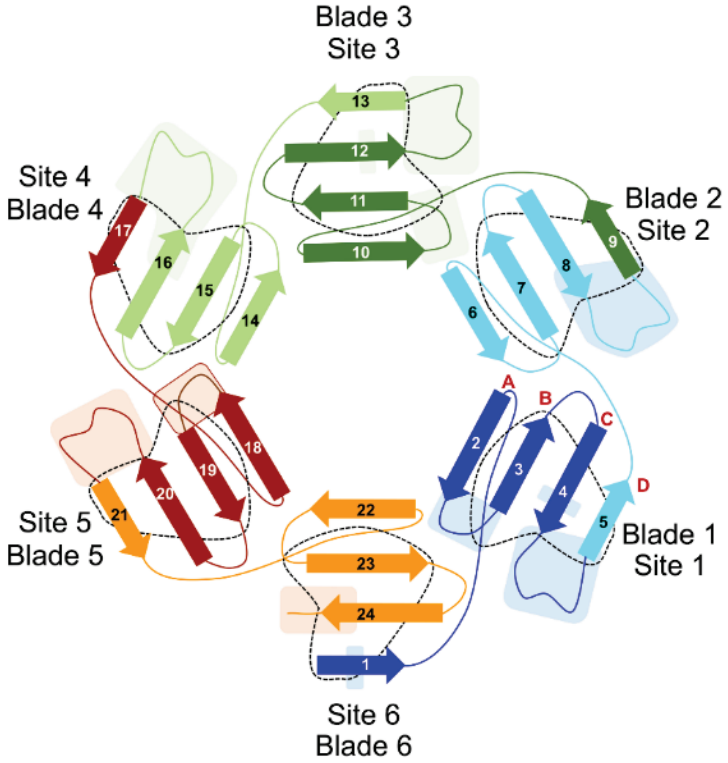
Lb-Tec2 structure

- 4-6 months to grow crystals
- Solved by MAD with soaked 2MeSeFuc





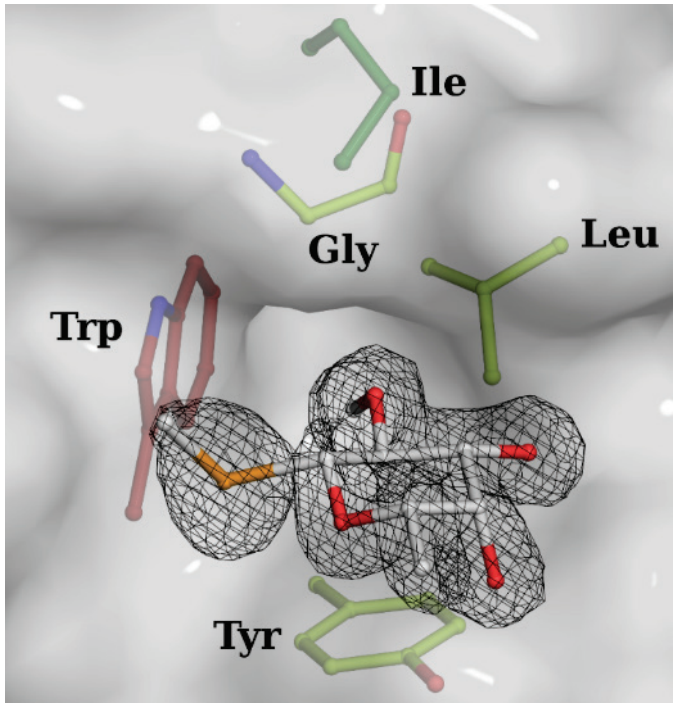
5FSB
1.65 Å

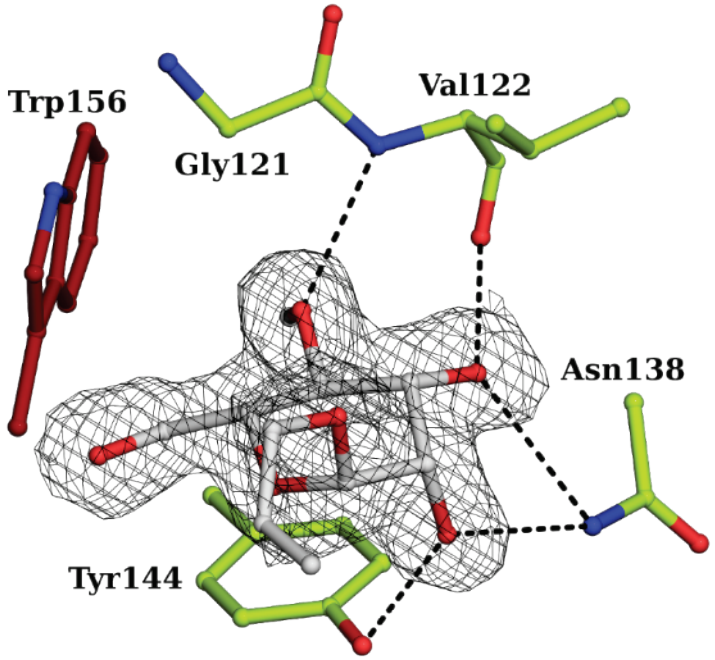


Sommer R Beilstein J Org Chem 2016 12:2828; Sommer R Structure 2018 26:391

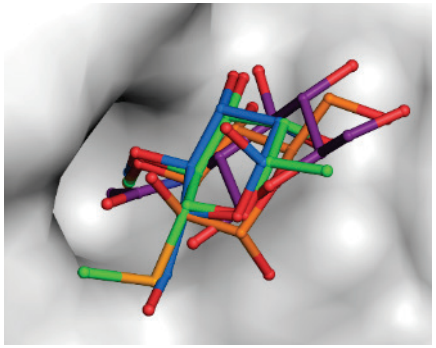
Cermav

Methyl recognition pocket

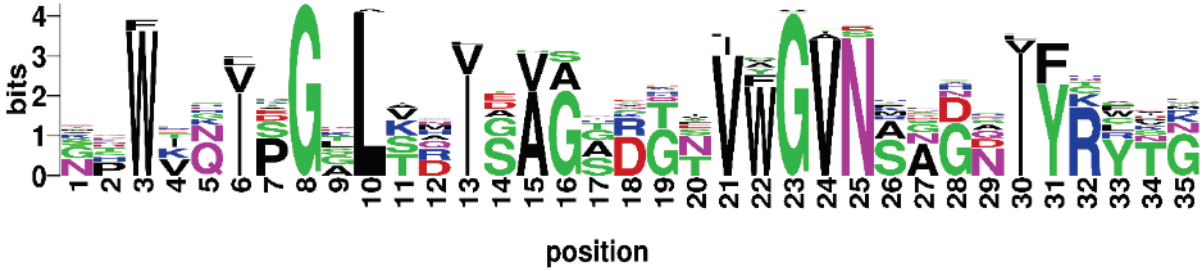




5FSC
1.95 Å



Highly conserved residues

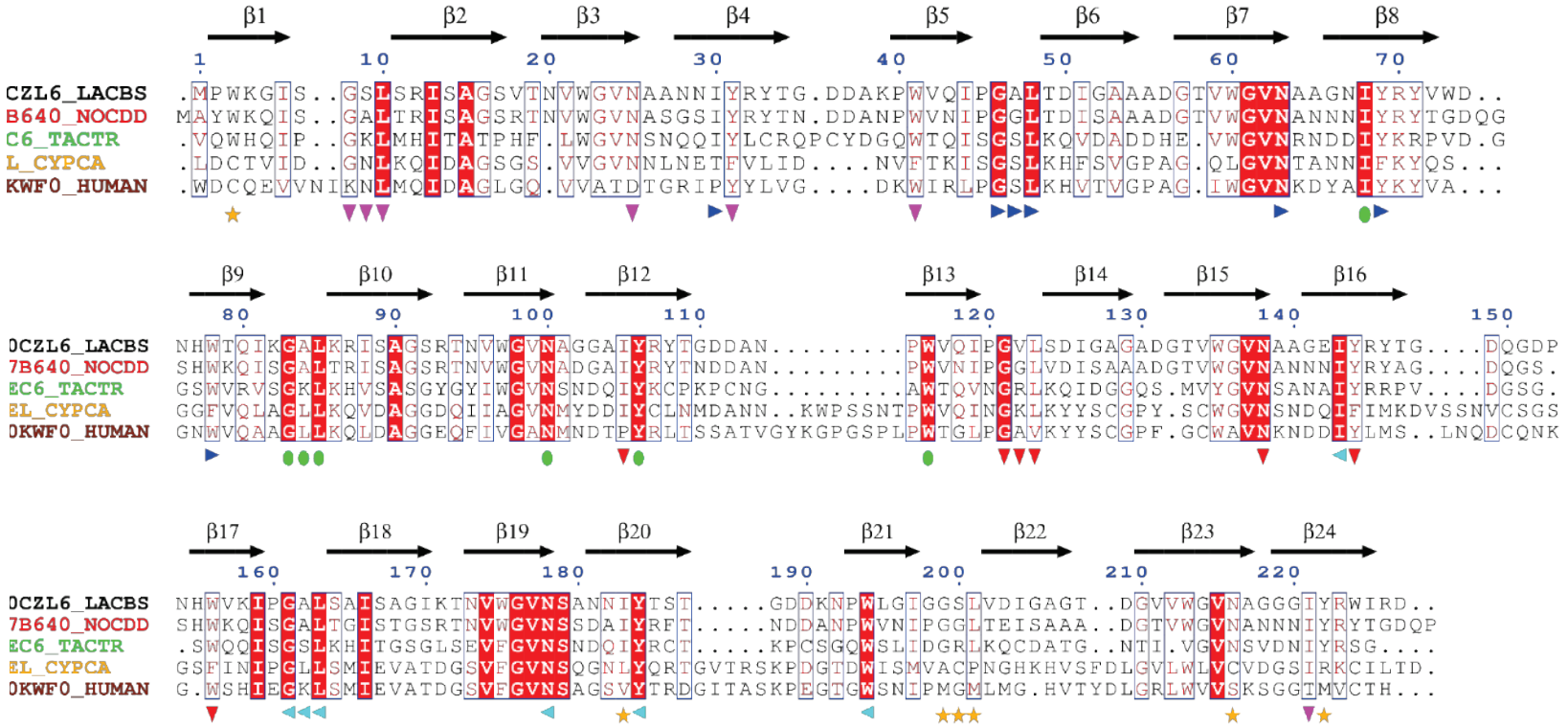


Sommer R Structure 2018 26:391

Cermav

Binding site formation

- Requires 3 TECPR repeats



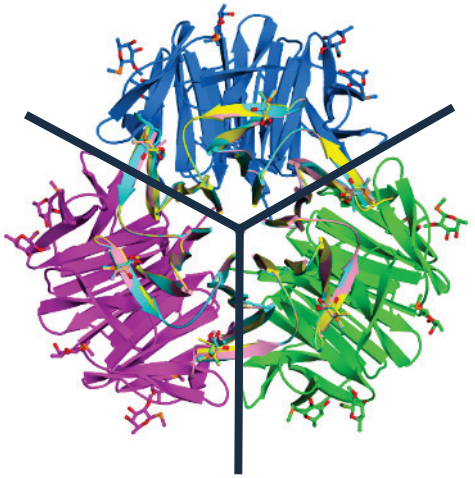
- Specificity conserved through evolution
  - Importance in innate immunity

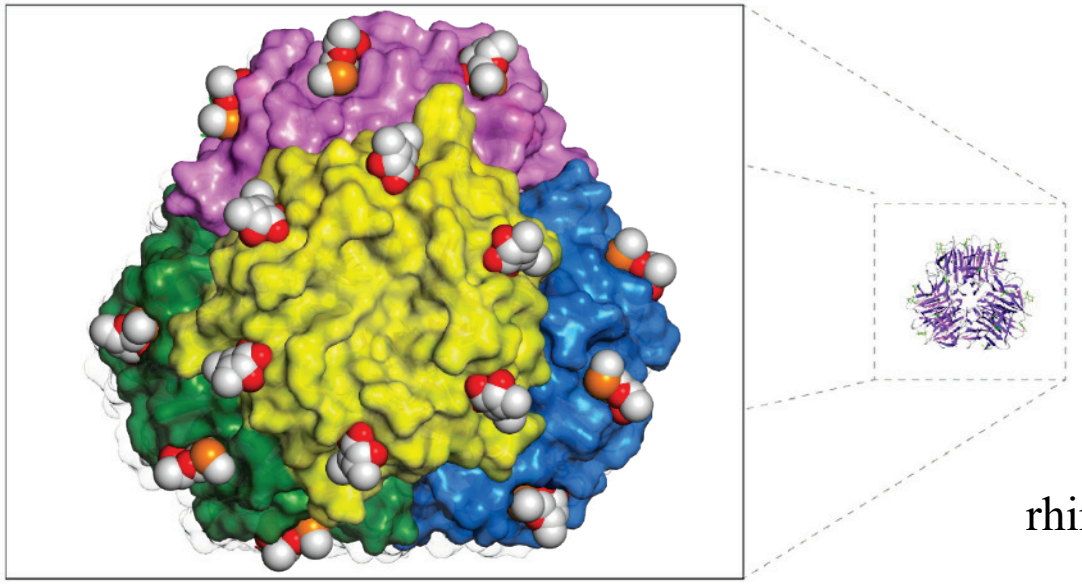
Sommer R Structure 2018 26:391

Cermav

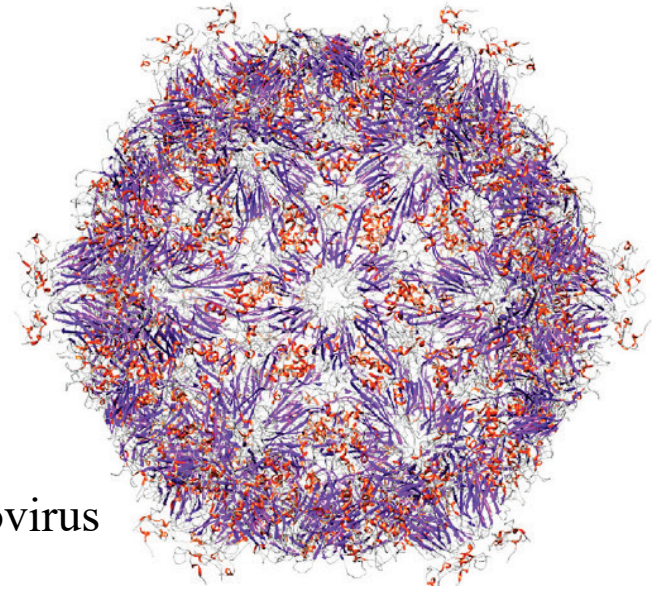
Unique quaternary structure

- Ball-shaped tetramer:
  - Unique with three fold axis in the PDB
  - Confirmed by SAXS
- 24 sites distributed on the surface
- Ressembles mini viral capsid





rhinovirus



Cordara G Structure 2018 26:369; Sommer R Structure 2018 26:391

Cermav

Mini lectin rPhoSL

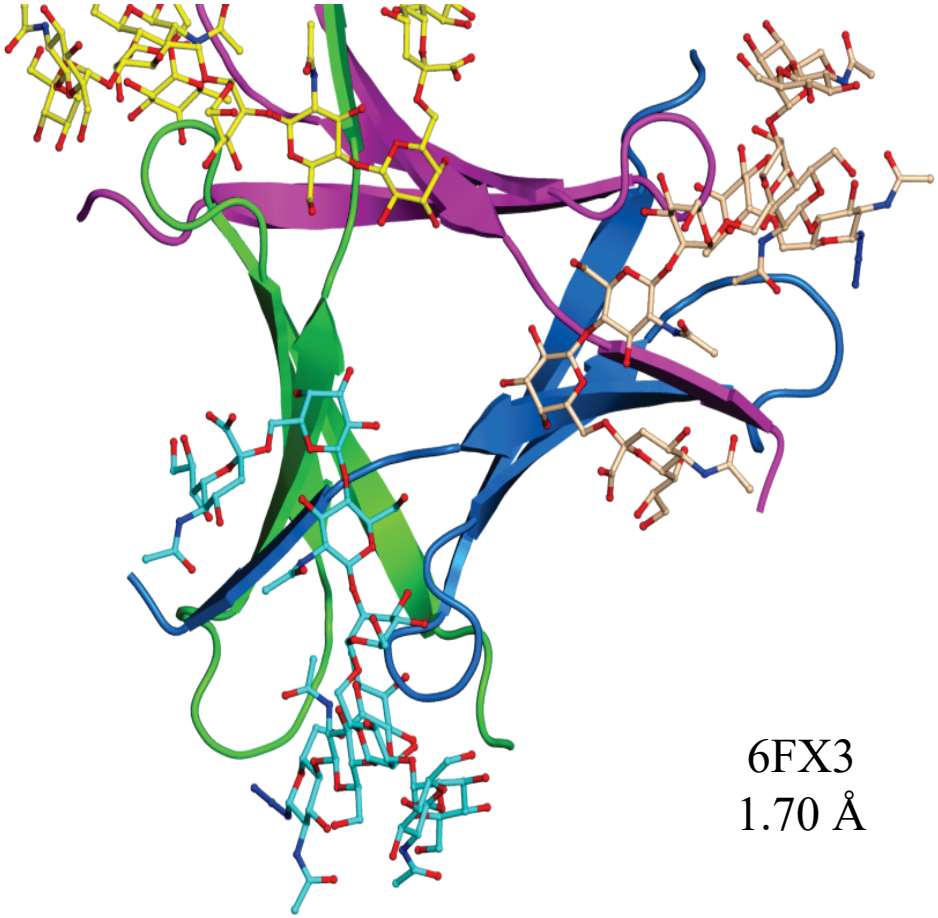
- 40 AA long
- Described as strictly specific for α1,6 core fucosylation
- Great potential as biomarker
  - Core fucose associated with cancer: hepatocarcinoma,
  - Diminish bioactivity of mAbs
    - See Tony's talk tomorrow
- Unknown fold
- Unknown recognition motif
- Produced recombinantly

Kobayashi Y JBC 2012 287:41; Yamasaki Y Scien Rep 2018 8:7740; Cabanettes A Angewandte Chemie 2018 in press

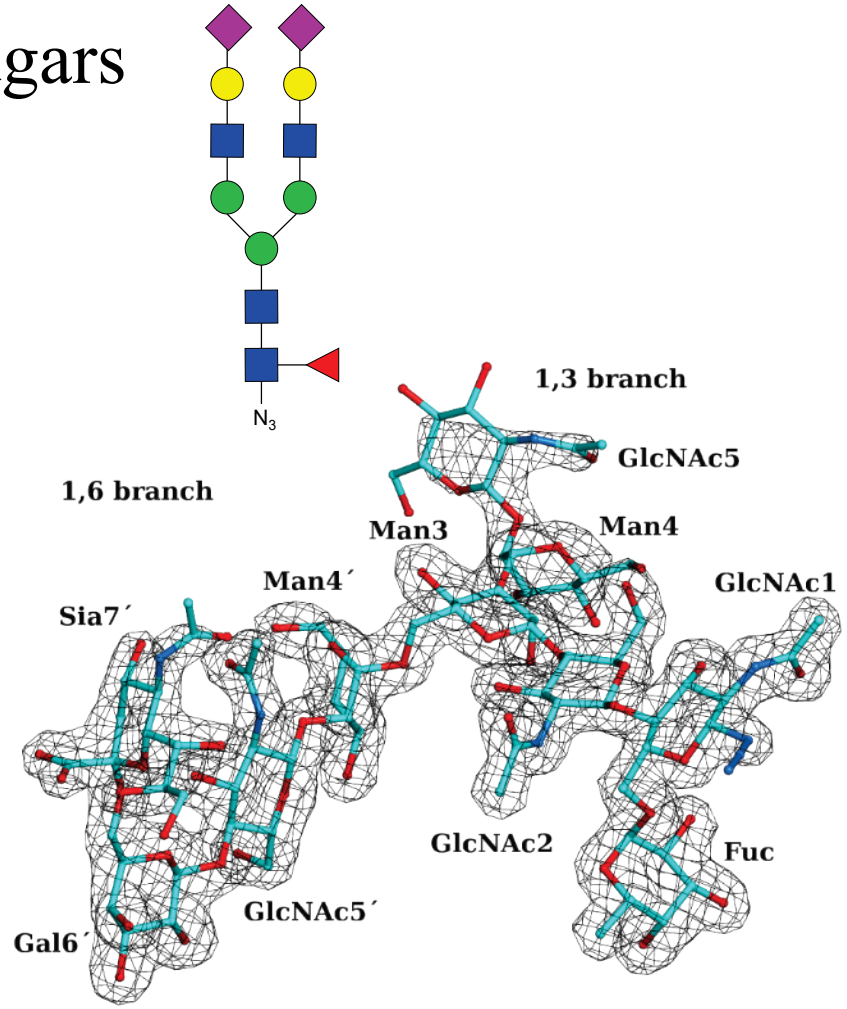
Cermav

Novel fold: β-prism III

- 1/3 of the ASU occupy by sugars



6FX3
1.70 Å

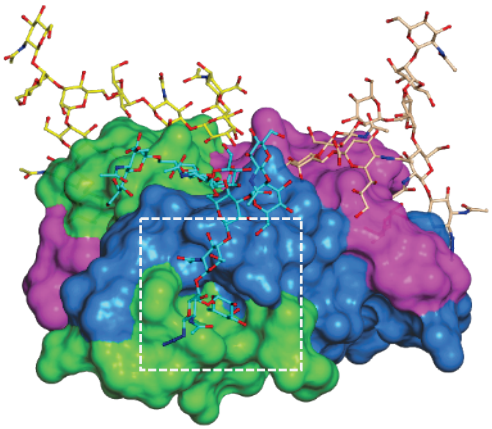


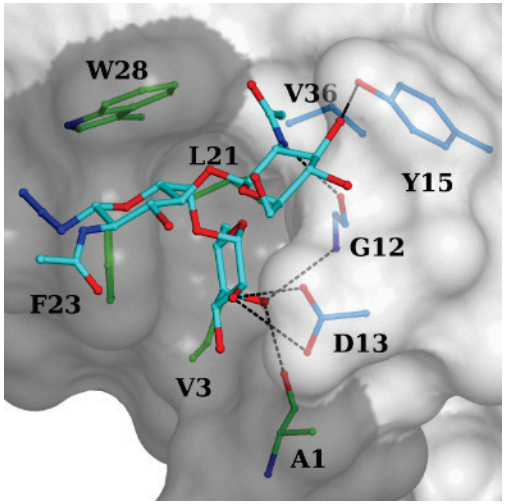
Cabanettes A Angewandte Chemie 2018 in press

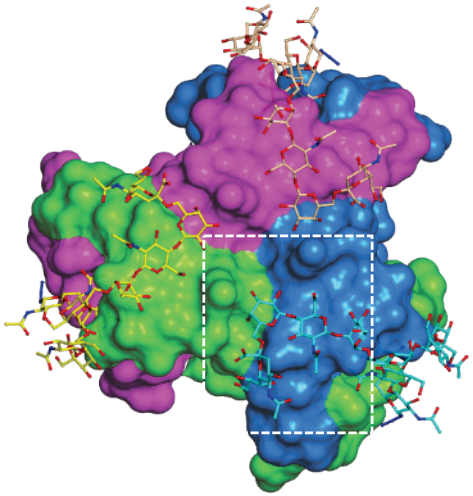
Cermav

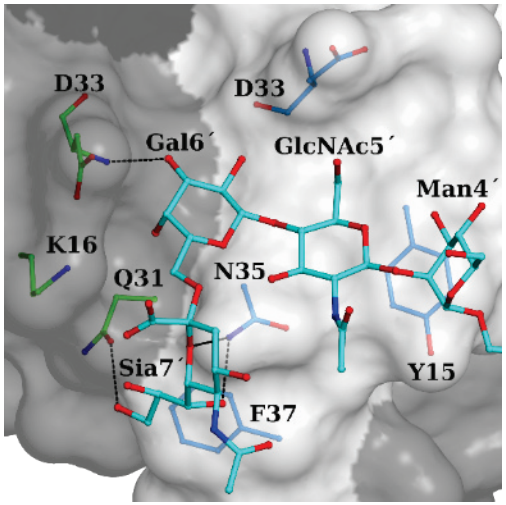
Binding sites

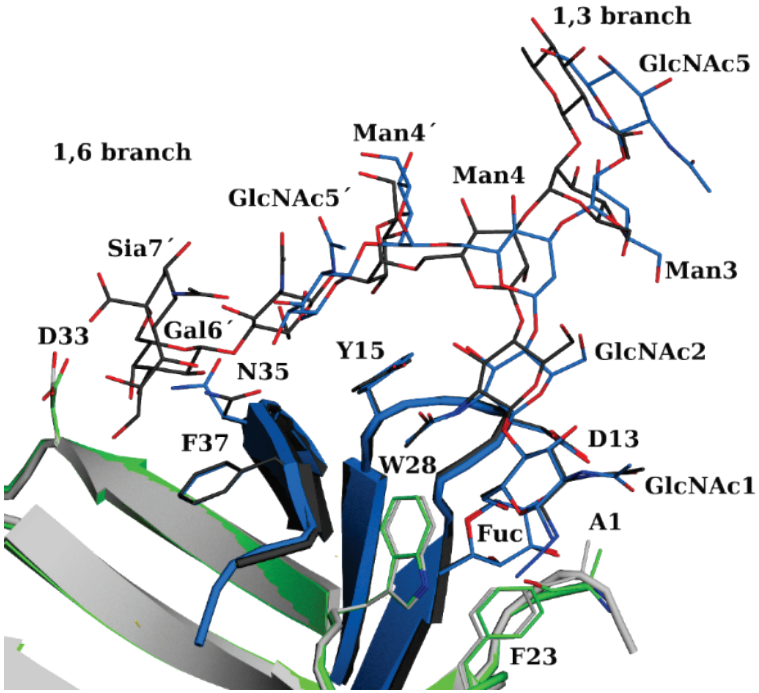
- 2 binding pockets: Fuc: conserved / Gal: can vary







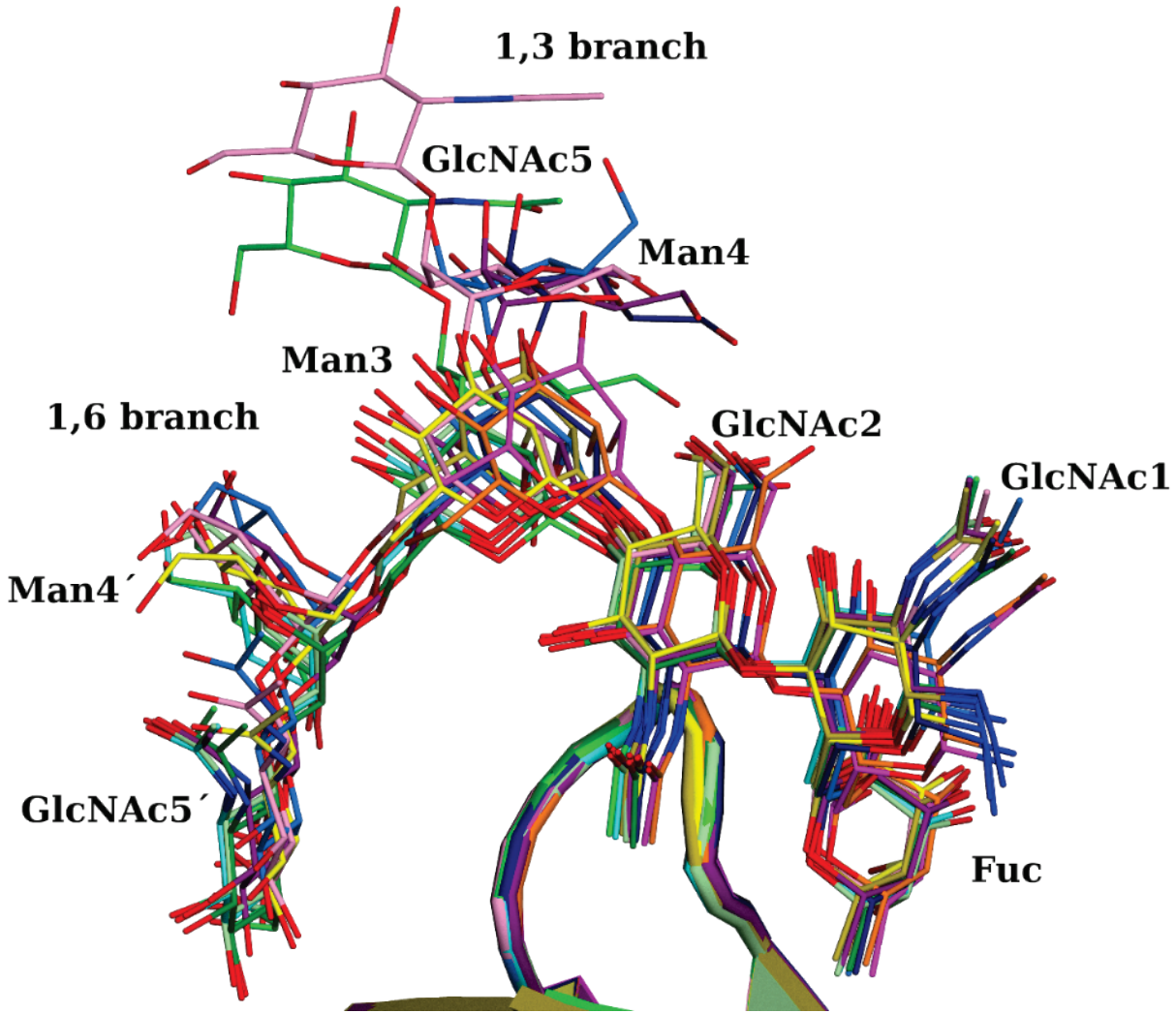


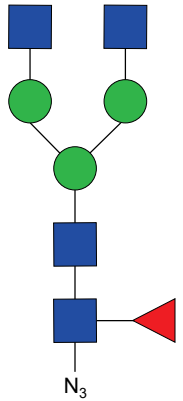


Cabanettes A Angewandte Chemie 2018 in press

Cermav

Flexibility in crystal structures





Cabanettes A Angewandte Chemie 2018 in press

Cermav

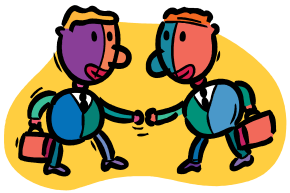
Conclusions

- X-ray crystallography gave access to protein sugar interaction at the atomic level
  - Highest the resolution the better
- Could be tricky for non glycobiologists
- Distorsion
  - Real or artefact from user/program errors
- No « Structural glycobiology for dummies »
  - Do not hesitate to contact expert of CCP4bb

Acknowledgements

CERMAV, Grenoble

Anne Imberty
Valérie Chazalet
Emilie Gillon
Aymeric Aufray
Aurore Cabanettes
Olga Makshakova
Bertrand Blanchard
Soorej Basheer
Hugues Driguez
Sébastien Fort



YSBL, York, UK

Gideon J Davies
Eleanor Dodson
Johan Turkenburg

ETH, Zürich, Switzerland

Markus Künzler
Therese Wohlschlager

Helmholtz, Saarbrücken, Germany

Alexander Titz
Roman Sommer

Bayreuth U. Germany

Lukas Perkams
Carolina Spies
Carlo Unverzagt

British Columbia University, Canada

Stephen G. Withers

Laboratorium für Organische Chemie, Zürich

Andrea Vasella