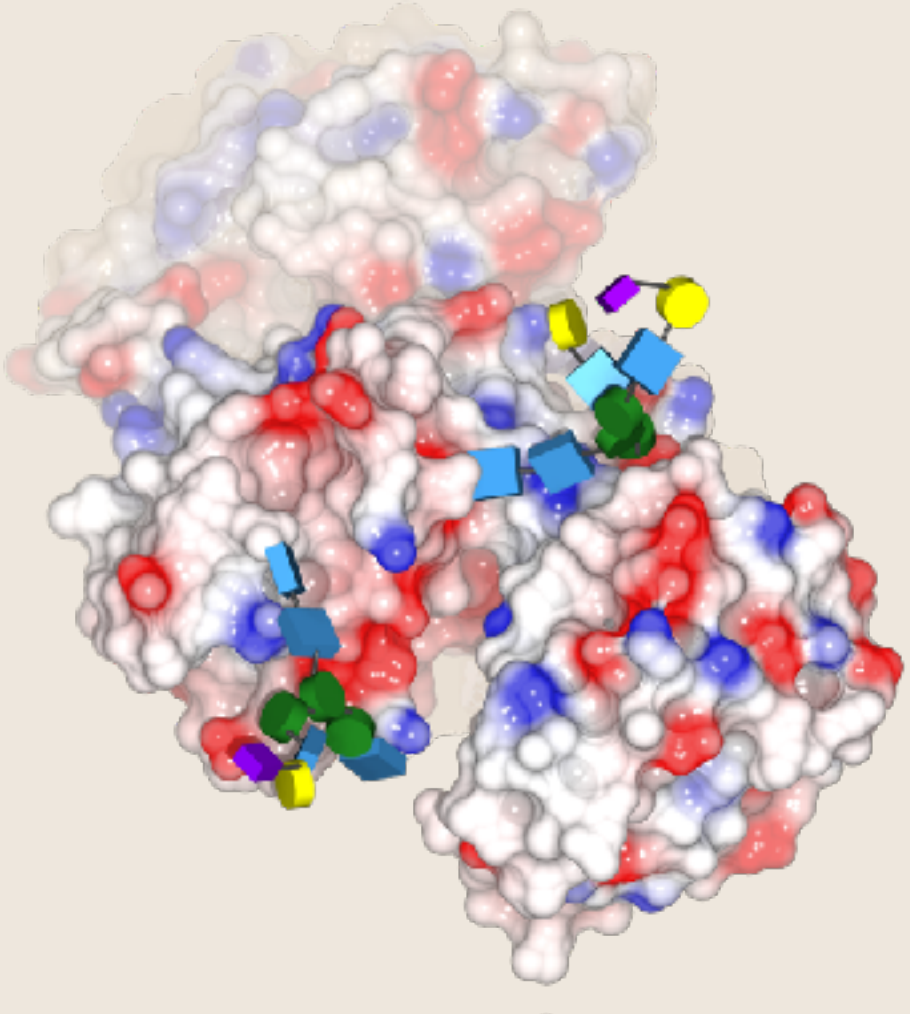
Carbohydrate crystallography

Jon Agirre

Royal Society Olga Kennard Research Fellow

York Structural Biology Laboratory, Department of Chemistry, University of York

@glycojones@mastodon.world



glyco-
/ˈglʌɪkəʊ/

GREEK
glukus → glyco-

glyco-
/ˈglʌɪkəʊ/

GREEK
glukus → glyco-

Sweet

glyco-
/ˈglʌɪkəʊ/

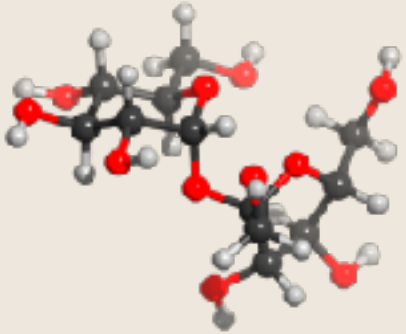
GREEK
glukus → glyco-

Sweet

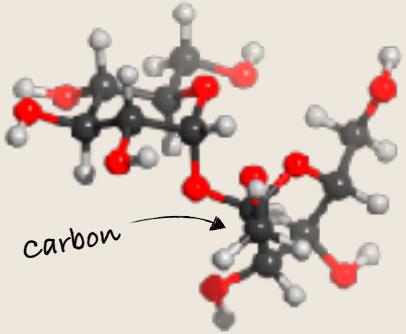
Not always sweet!

Not always sweet!

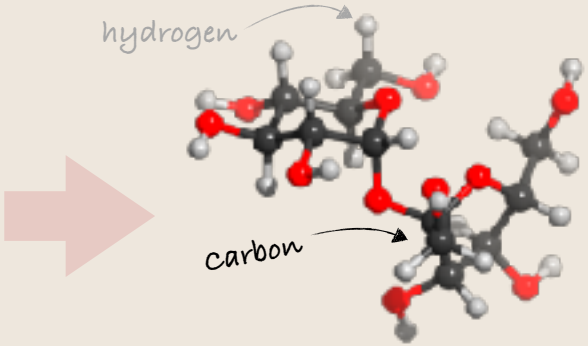
Not always sweet!



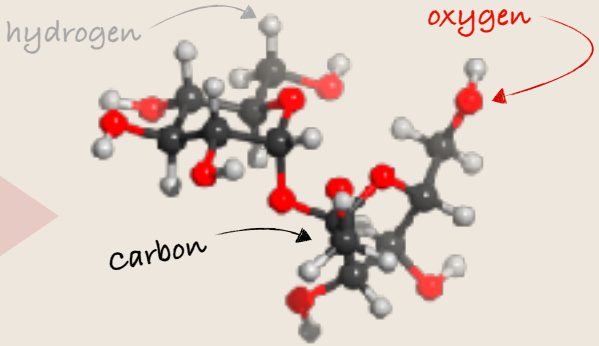
Not always sweet!



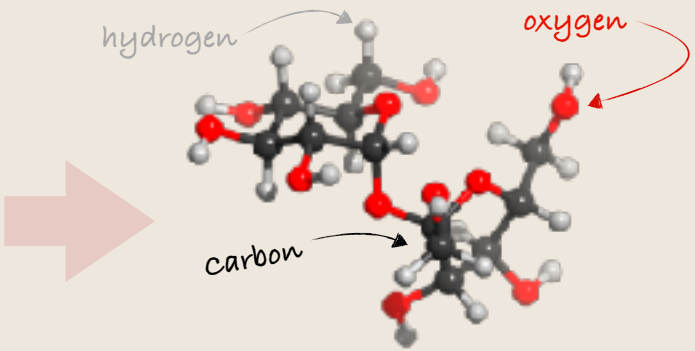
Not always sweet!



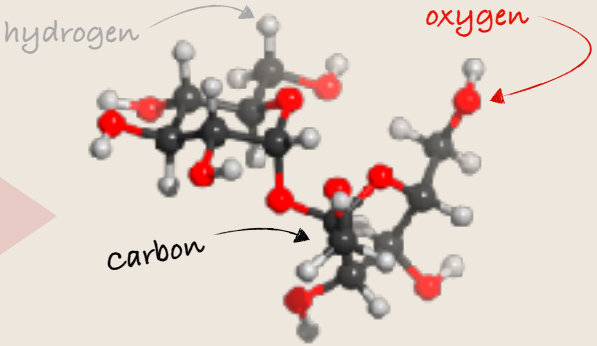
Not always sweet!



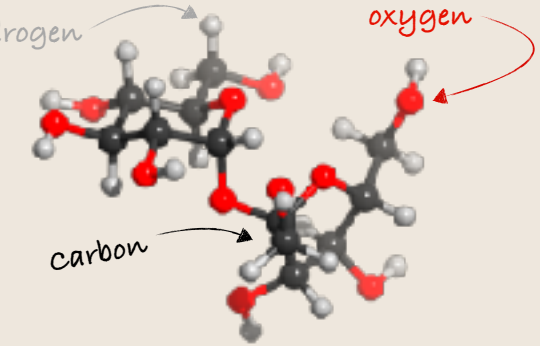
Not always sweet!



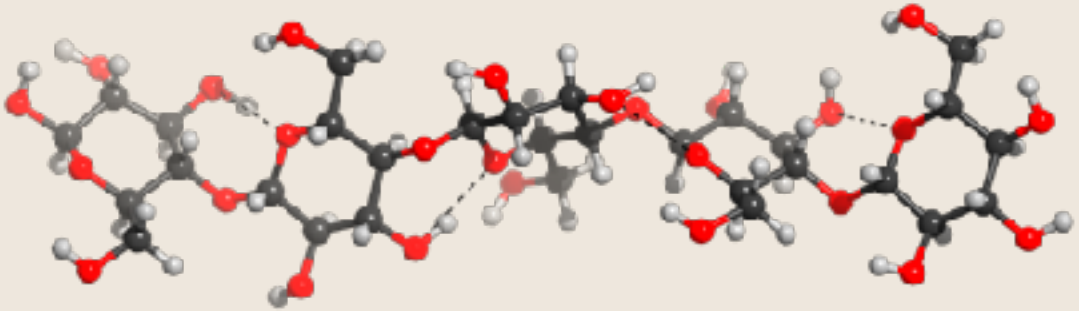
Not always sweet!

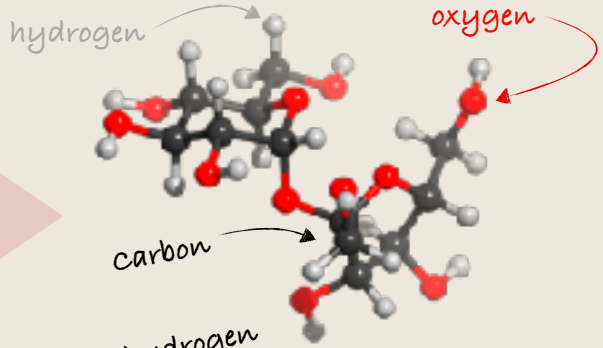


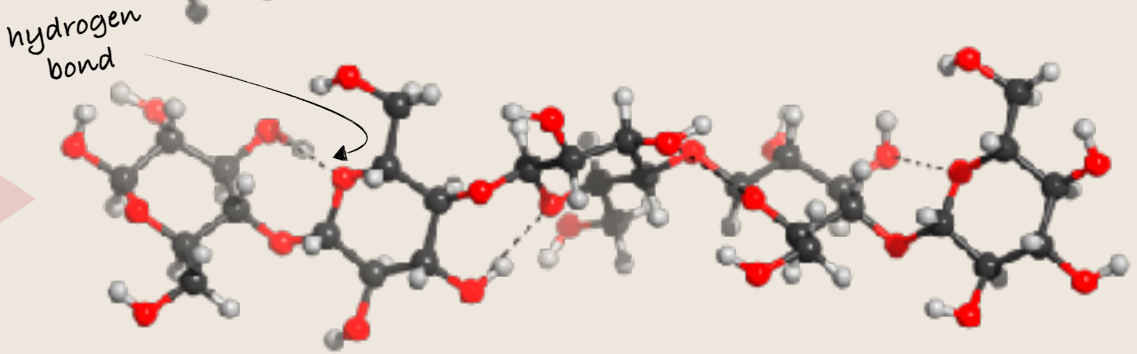
Not always sweet!



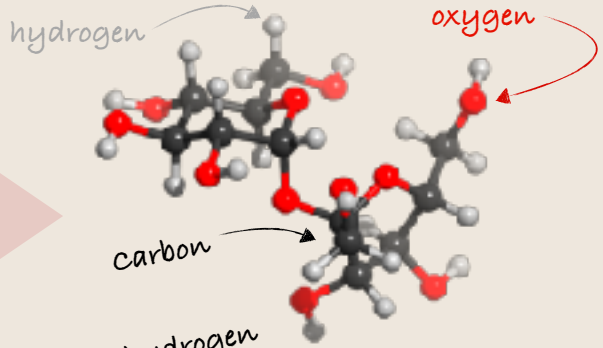
Not always sweet!

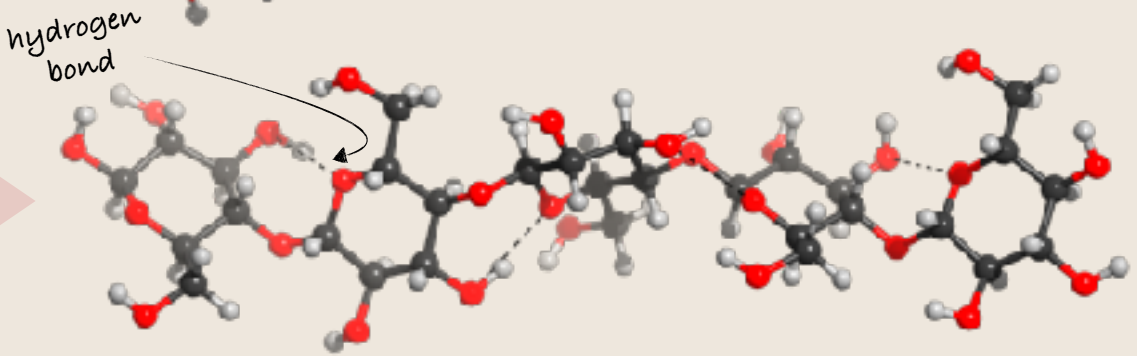




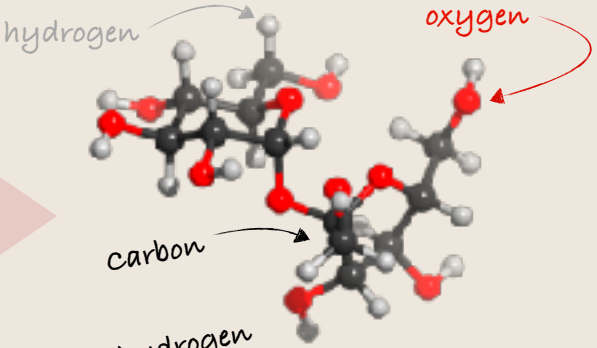


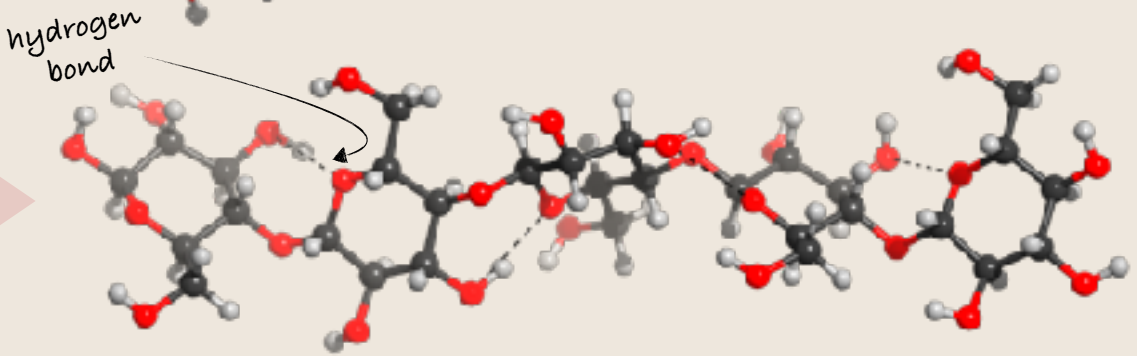
Not always sweet!



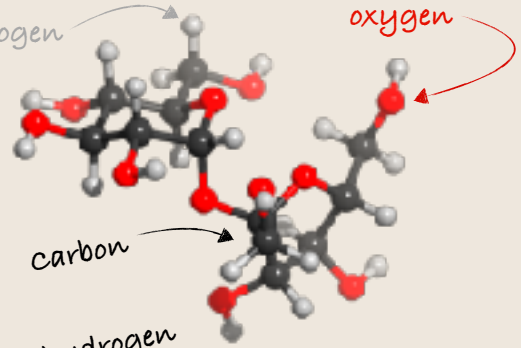


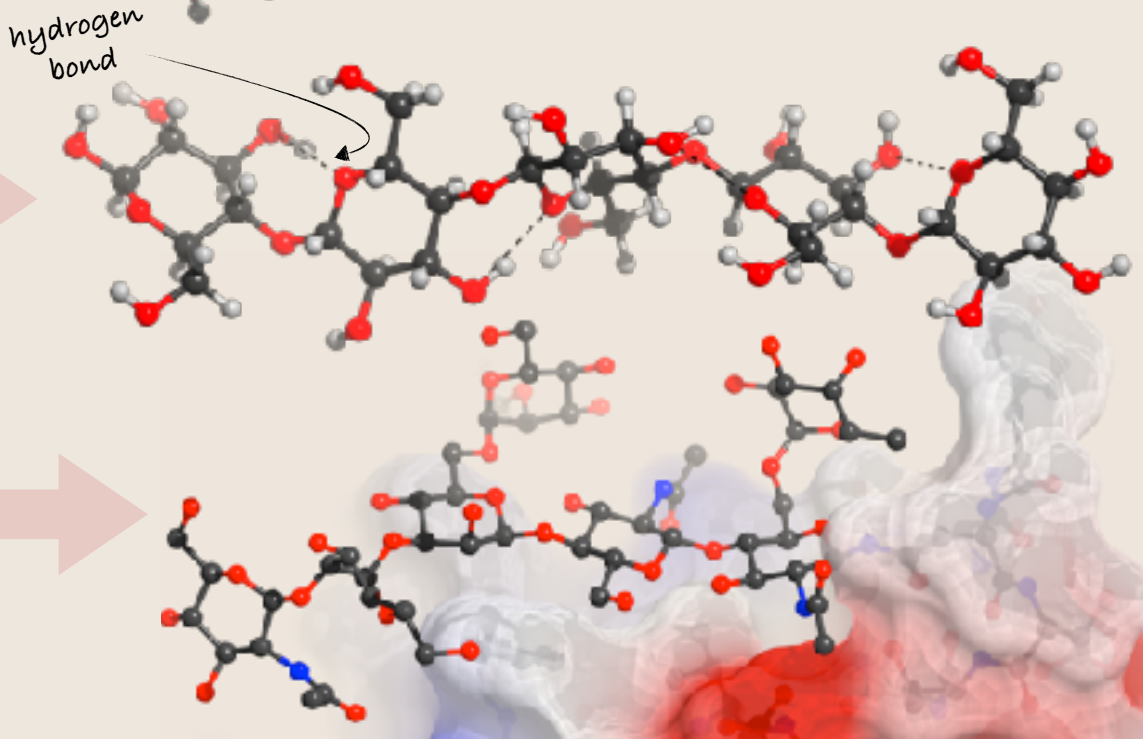
Not always sweet!





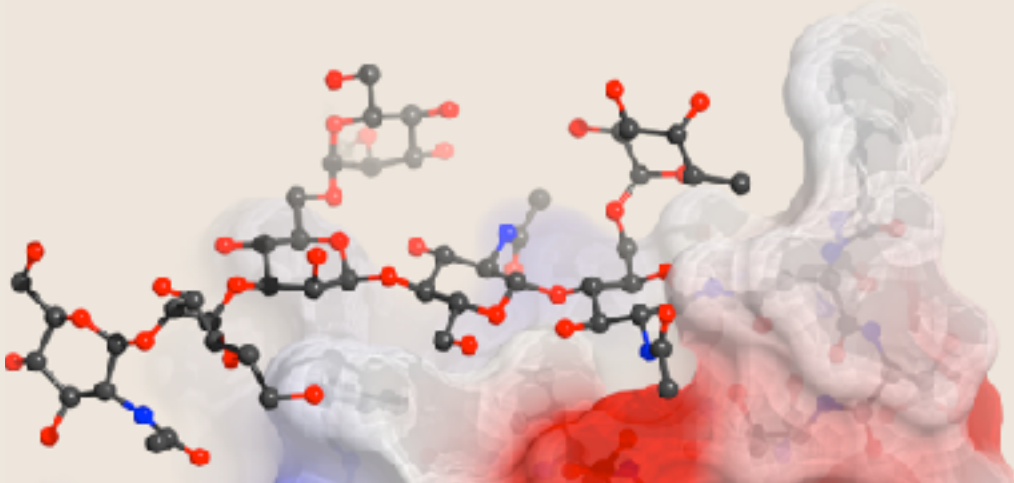
Not always sweet!



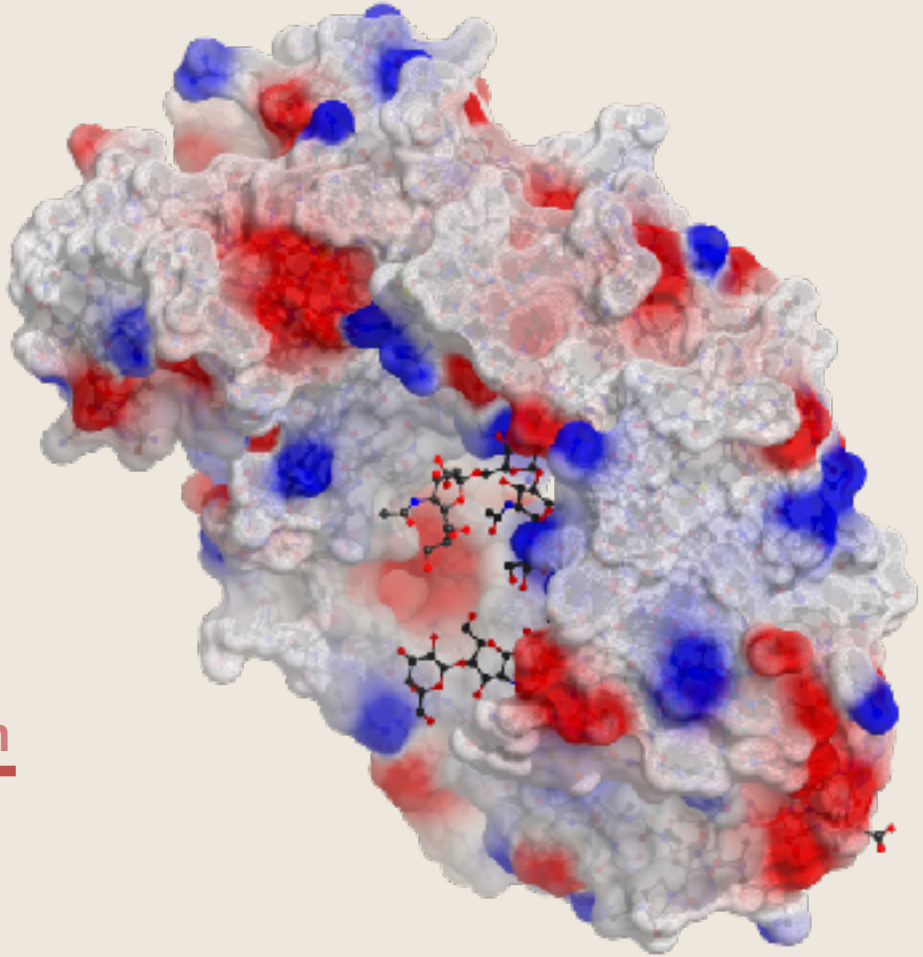


Not always sweet!

Protein glycosylation

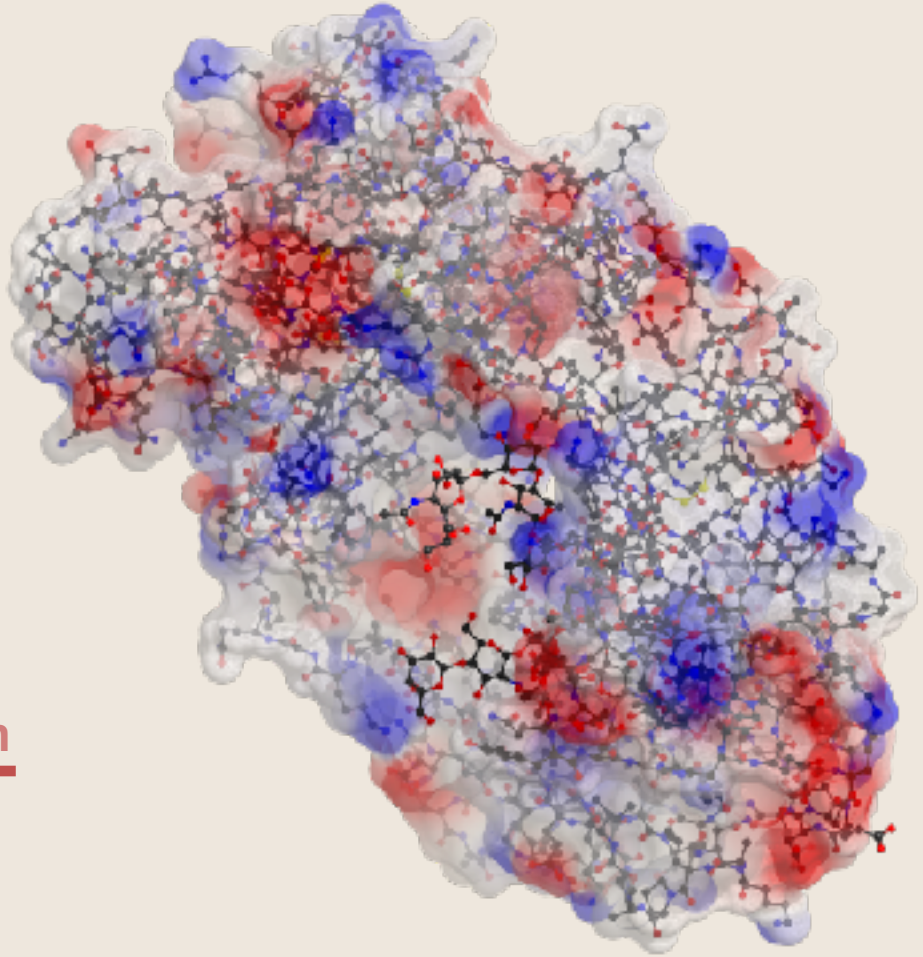


Protein glycosylation



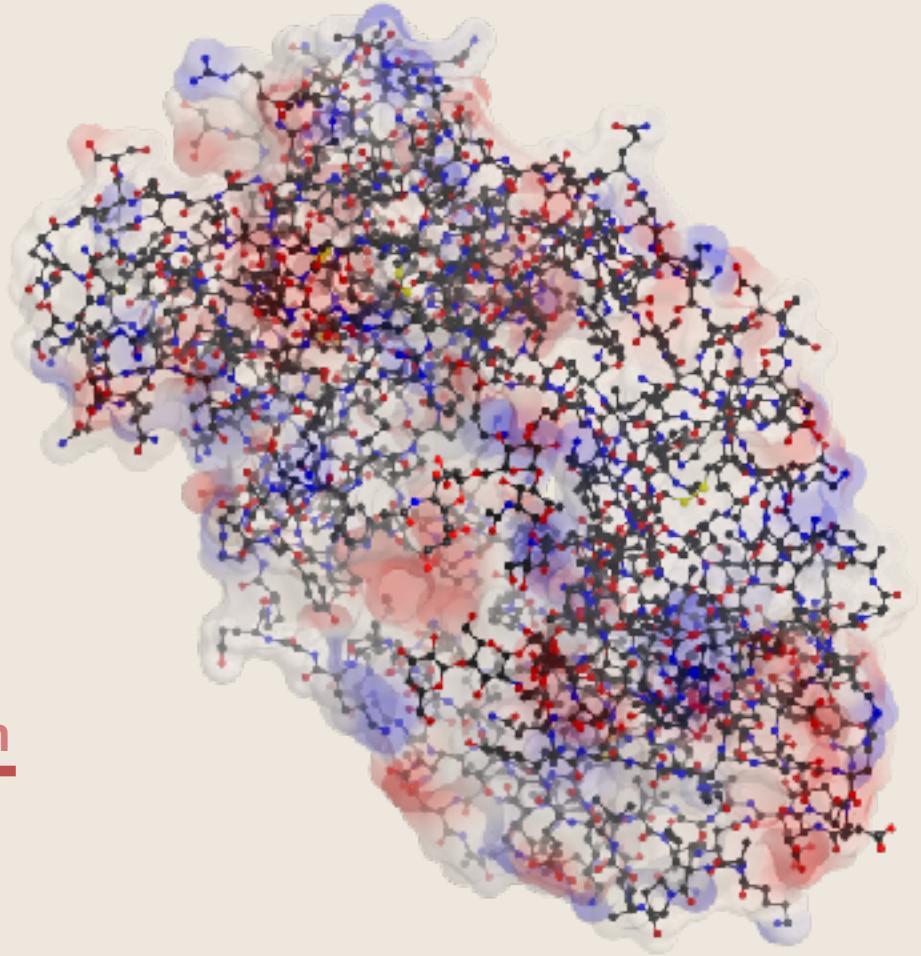
1 nm

Protein glycosylation

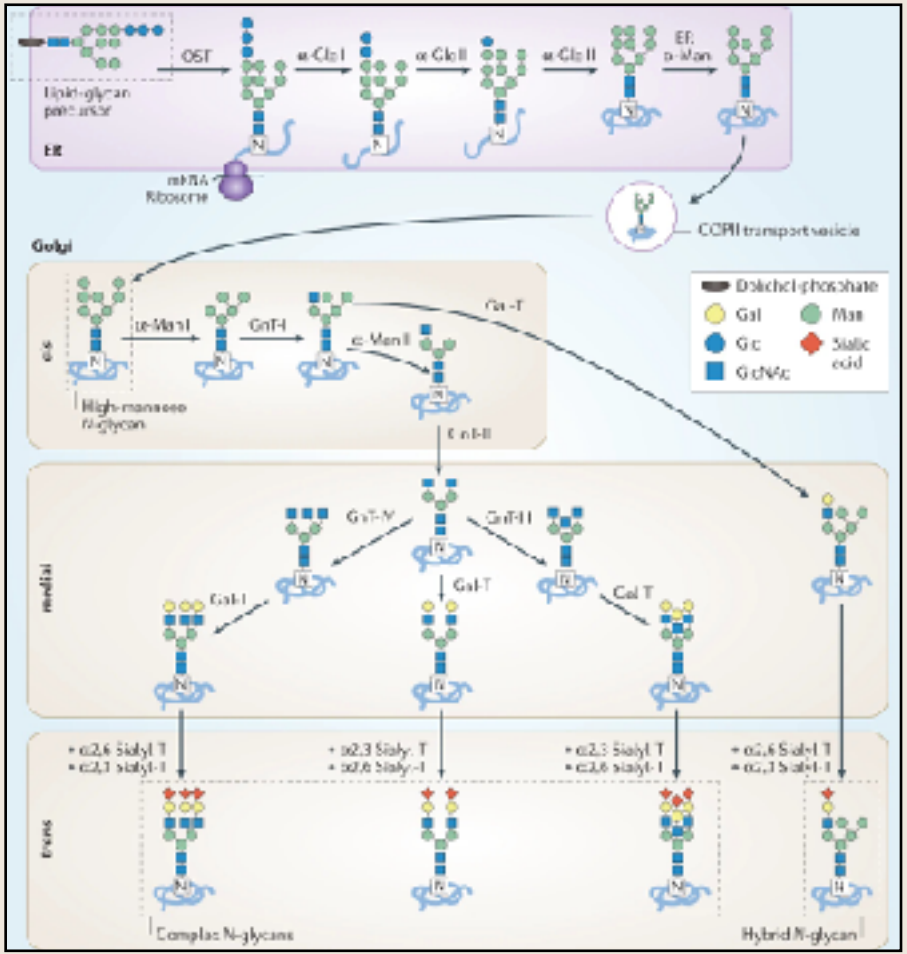


1 nm

Protein glycosylation

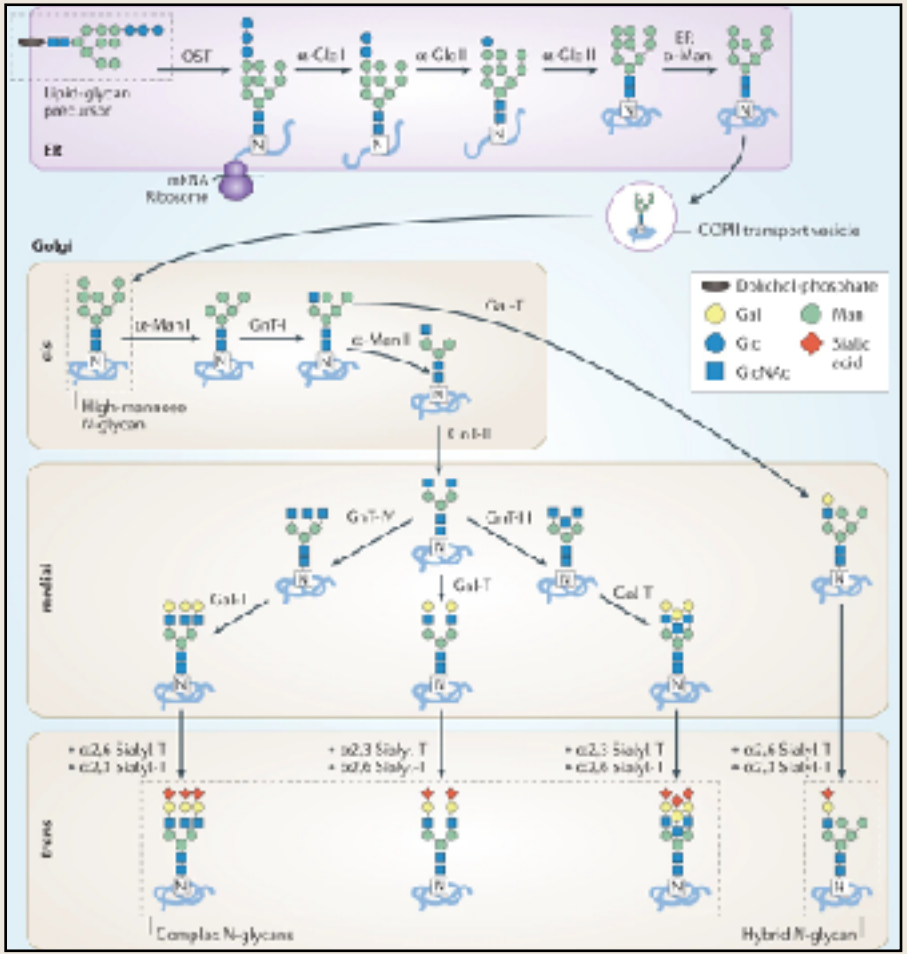


1 nm



Protein glycosylation

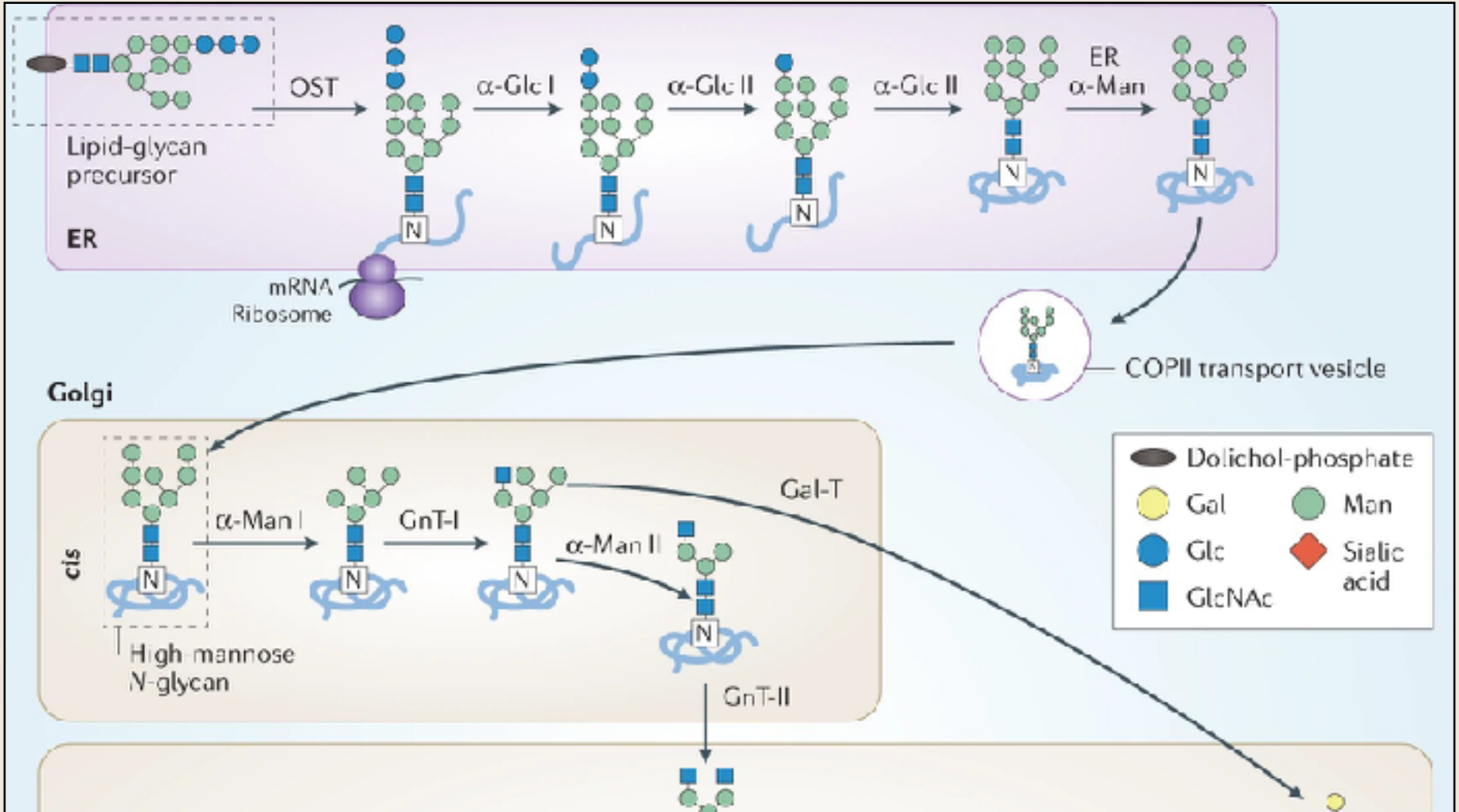
N-glycans

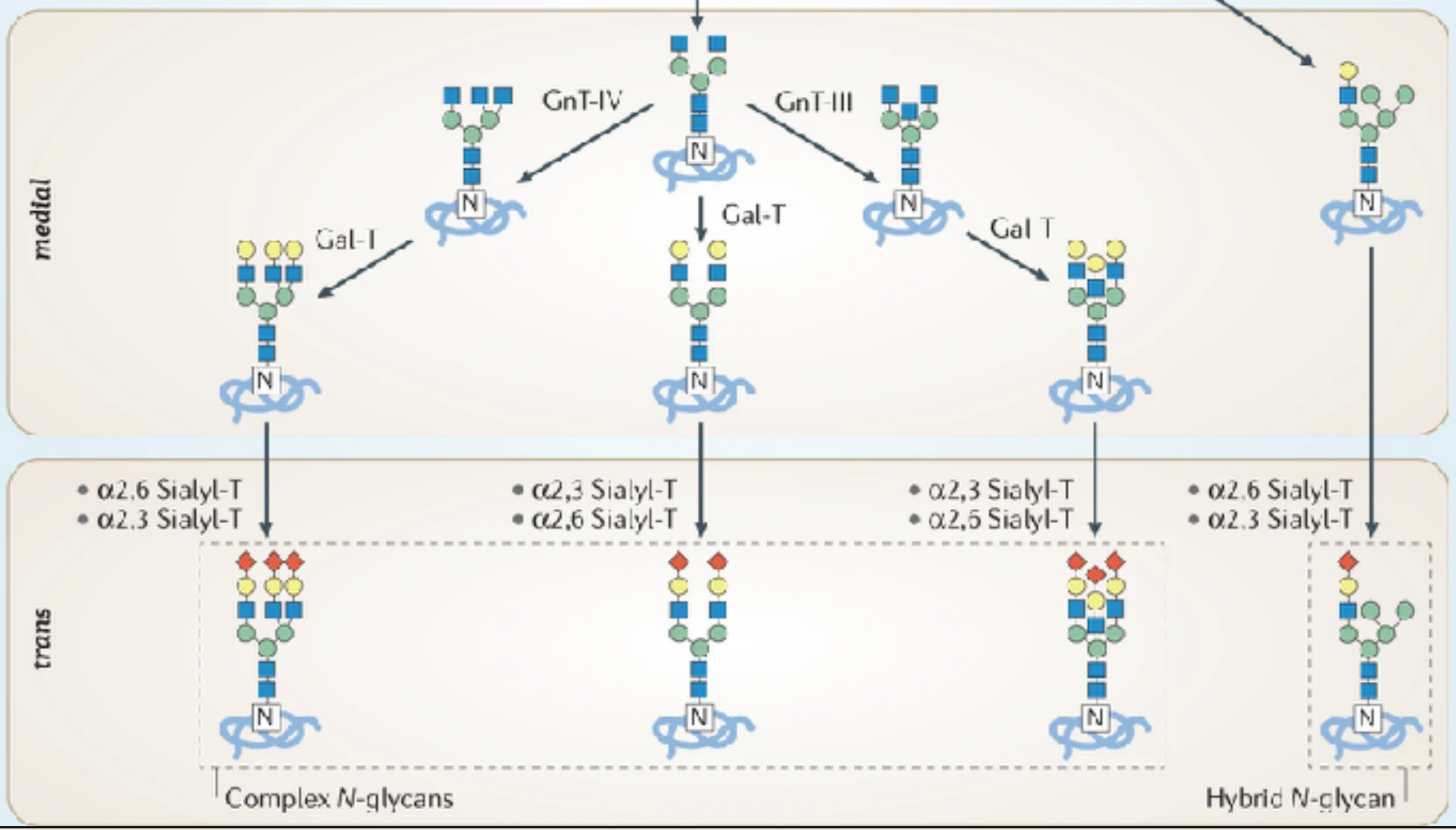


Protein glycosylation

N-glycans

Reily, C., Stewart, T.J., Renfrow, M.B. & Novak, J., 2019, Nat Rev Nephrol 15:346–366.

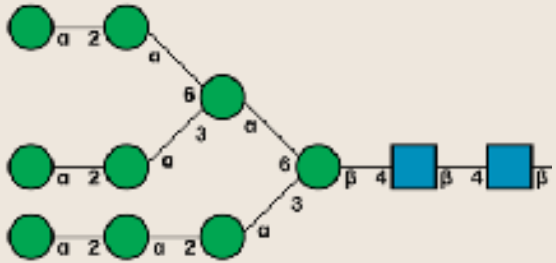

Lipid-glycan precursor
ER
OST
α-Glc I
α-Glc II
α-Glc II
ER α-Man
N
mRNA
Ribosome
COPII transport vesicle
Golgi
cis
α-Man I
GnT-I
α-Man II
Gal-T
High mannose N-glycan
GnT-II
Dolichol-phosphate
Gal
Man
Glc
Sialic acid
GlcNAc


medial
GnT-IV
GnT-III
Gal-T
Gal-T
Gal T
N
• α2,6 Sialyl-T
• α2,3 Sialyl-T
• α2,3 Sialyl-T
• α2,6 Sialyl-T
• α2,3 Sialyl-T
• α2,6 Sialyl-T
trans
Complex N-glycans
Hybrid N-glycan

N-glycan diversity

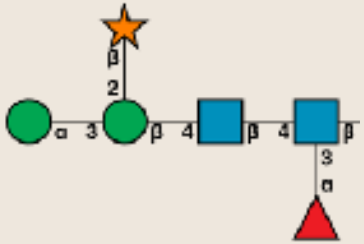
(a)

PDB code 5FJI
High mannose,
GH3 enzyme from
Aspergillus fumigatus



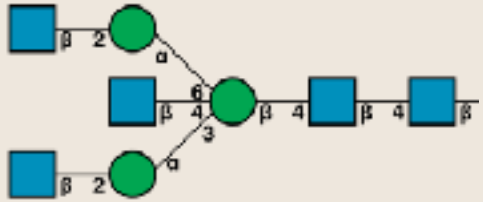
(b)

PDB code 5AOG
Plant glycan,
peroxidase enzyme from
Sorghum bicolor



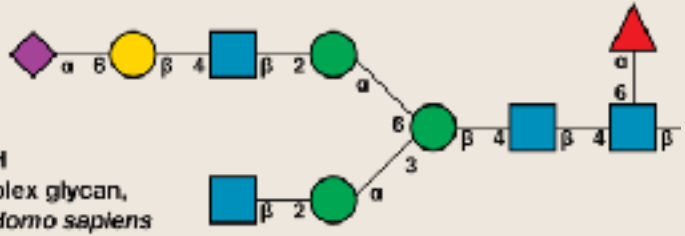
(c)

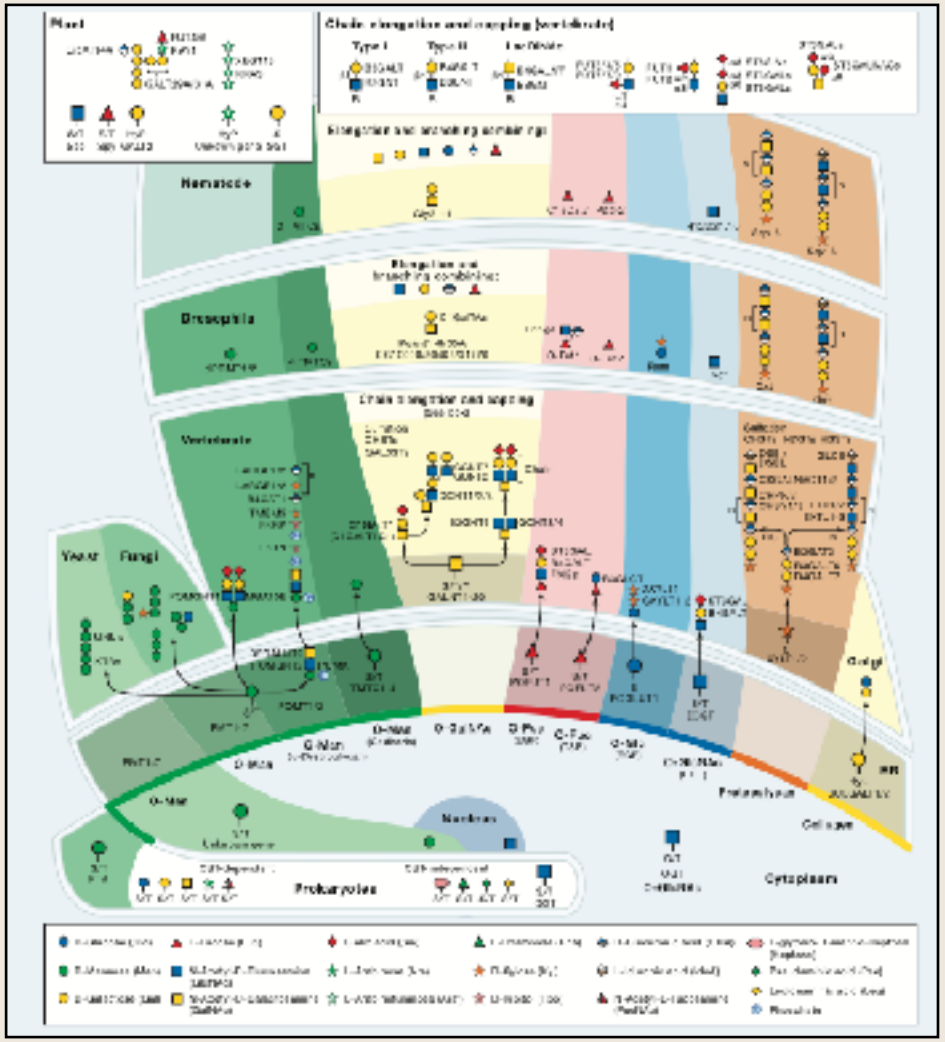
PDB code 3SGK
Complex glycan,
antibody from *Homo sapiens*,
expressed in *Cricetulus griseus*



(d)

PDB code 4BYH
Sialylated complex glycan,
antibody from *Homo sapiens*

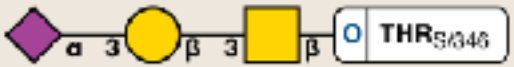




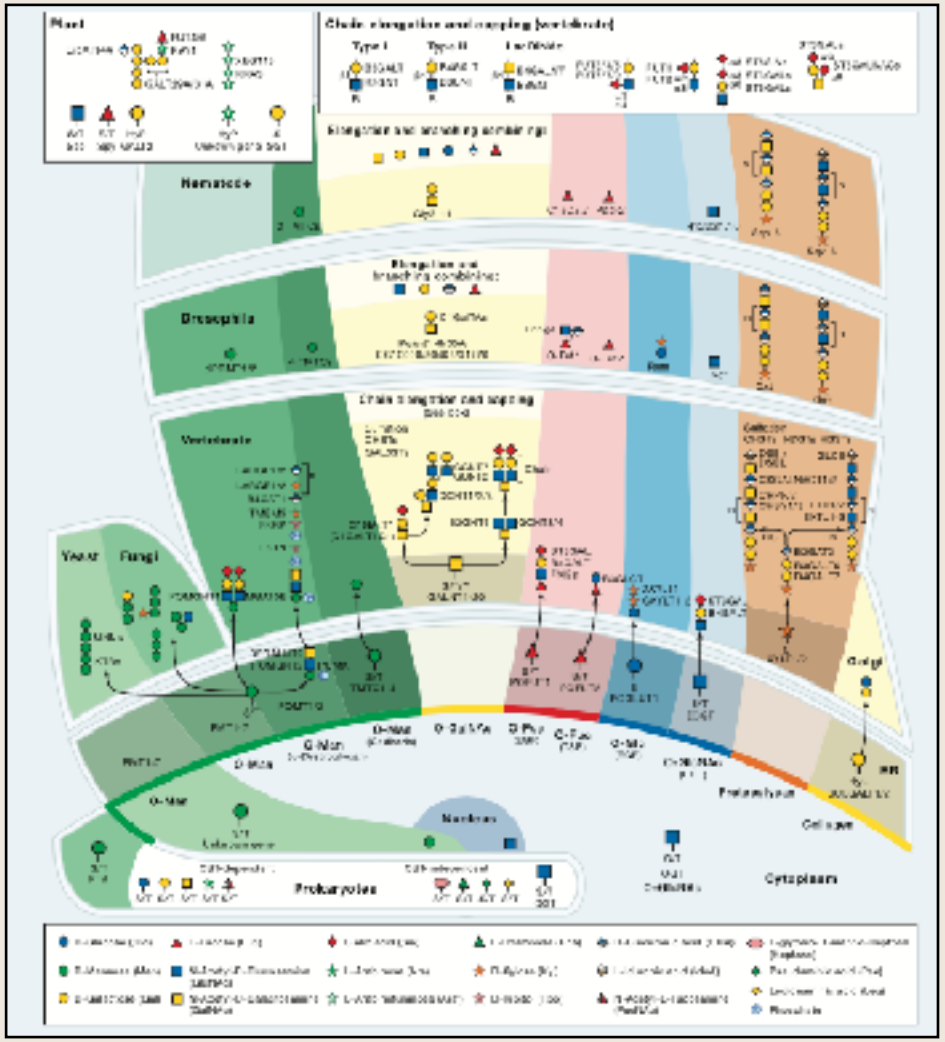
Protein glycosylation

O- and C-glycans

O-GalNAc (o-glycan)



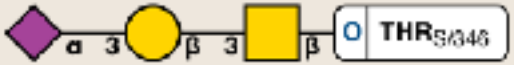
Tryptophan mannosylation (c-glycan)



Protein glycosylation

O- and C-glycans

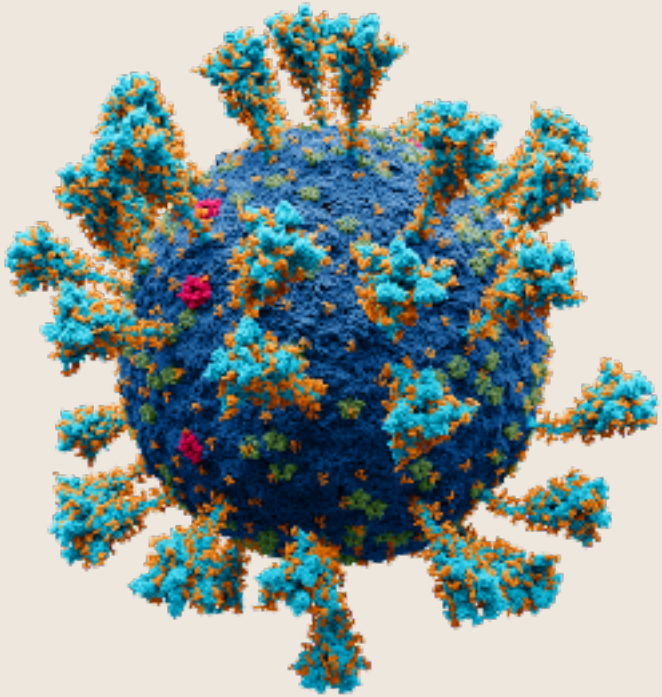
O-GalNAc (o-glycan)



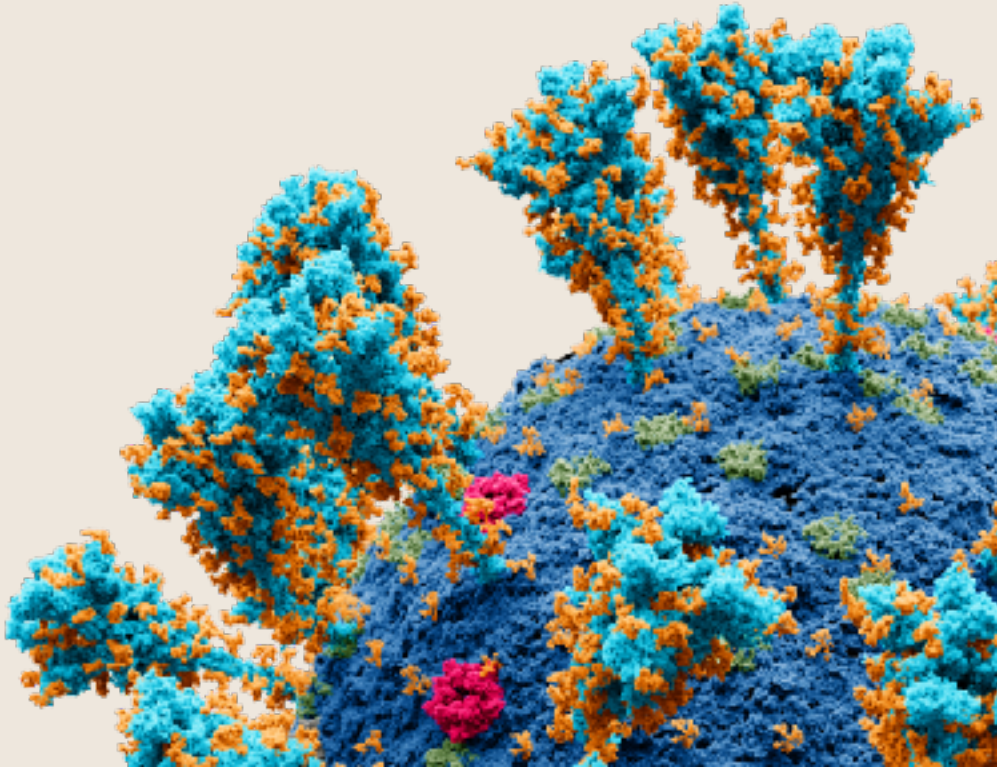
Tryptophan mannosylation (c-glycan)

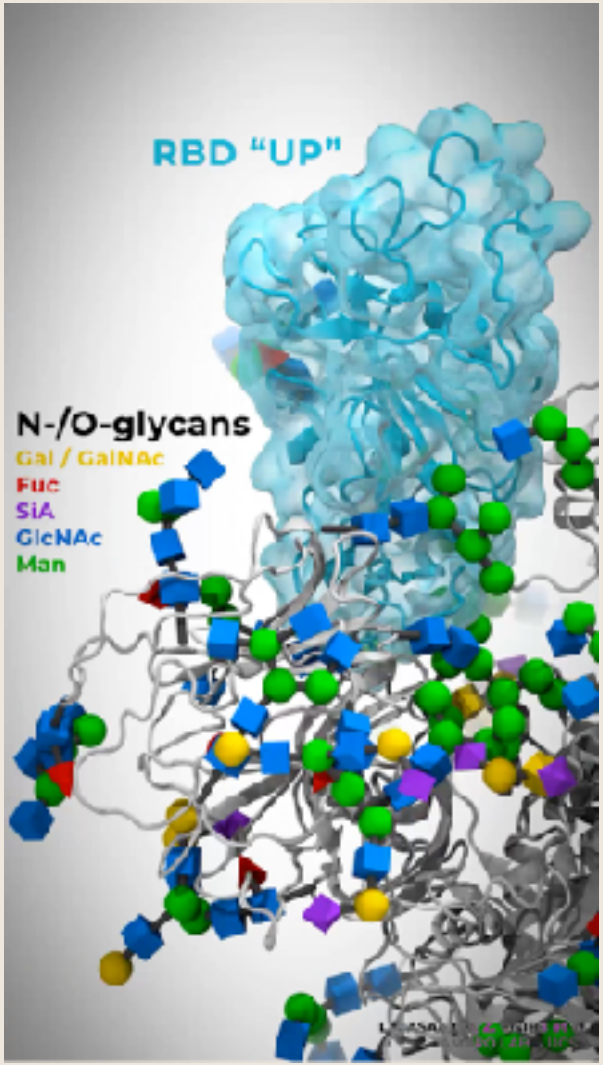
Joshi, Narimatsu, Schjoldager, Tytgat, Aebi, Clausen & Halim, 2018, Cell 172.

Glycans matter



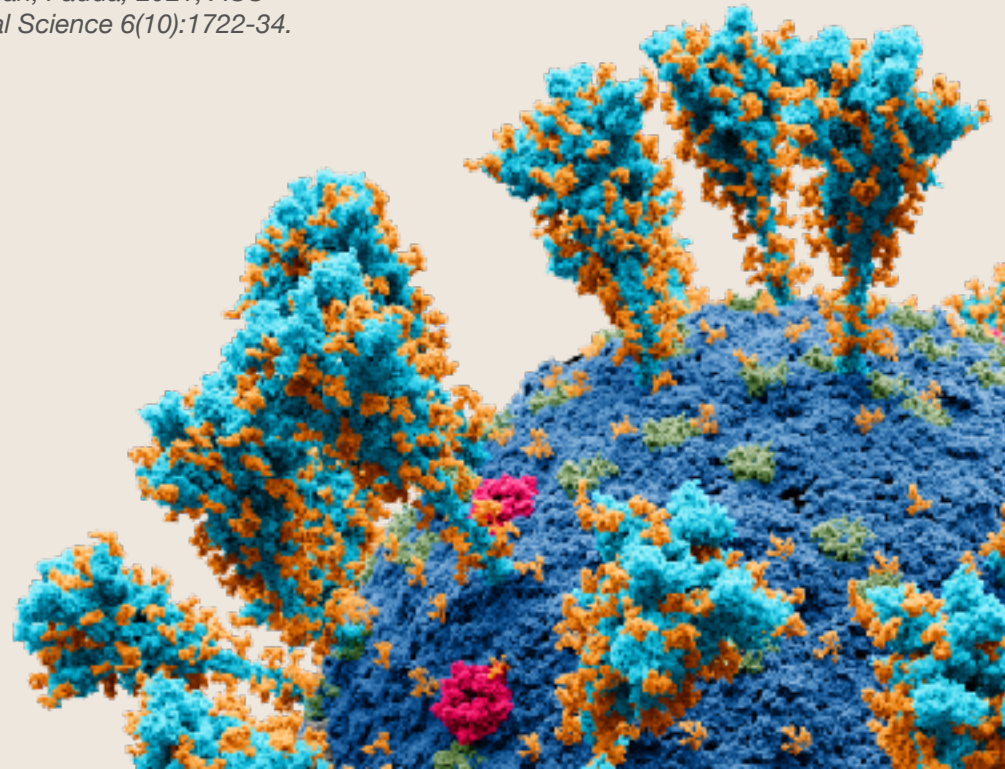
Glycans matter



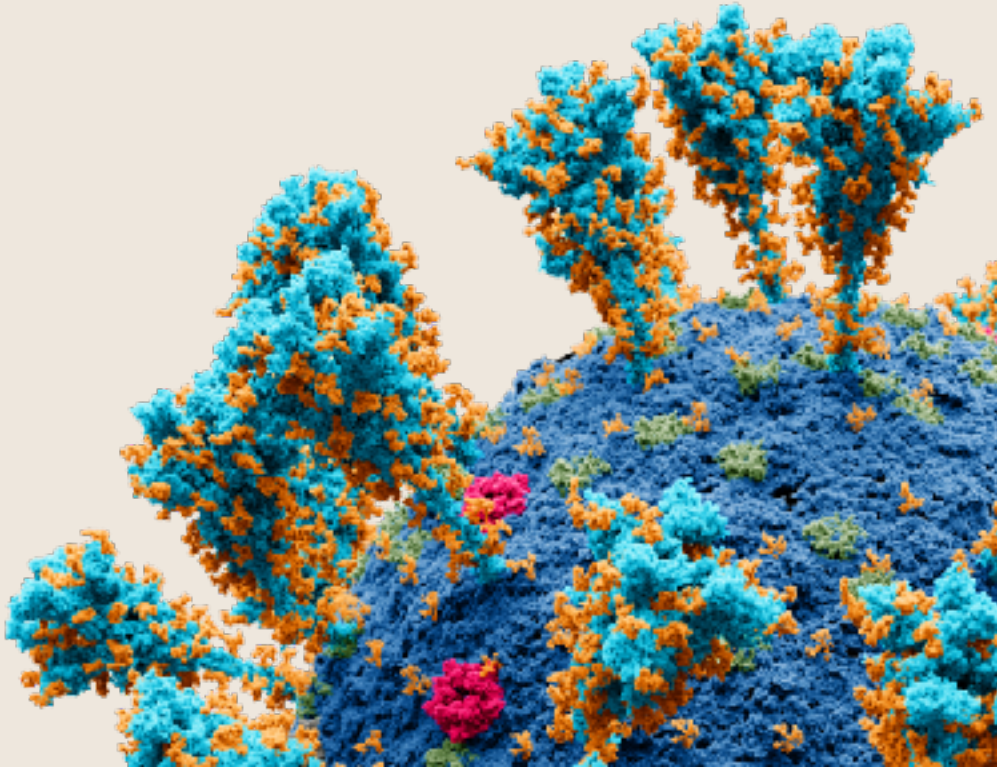


Glycans matter

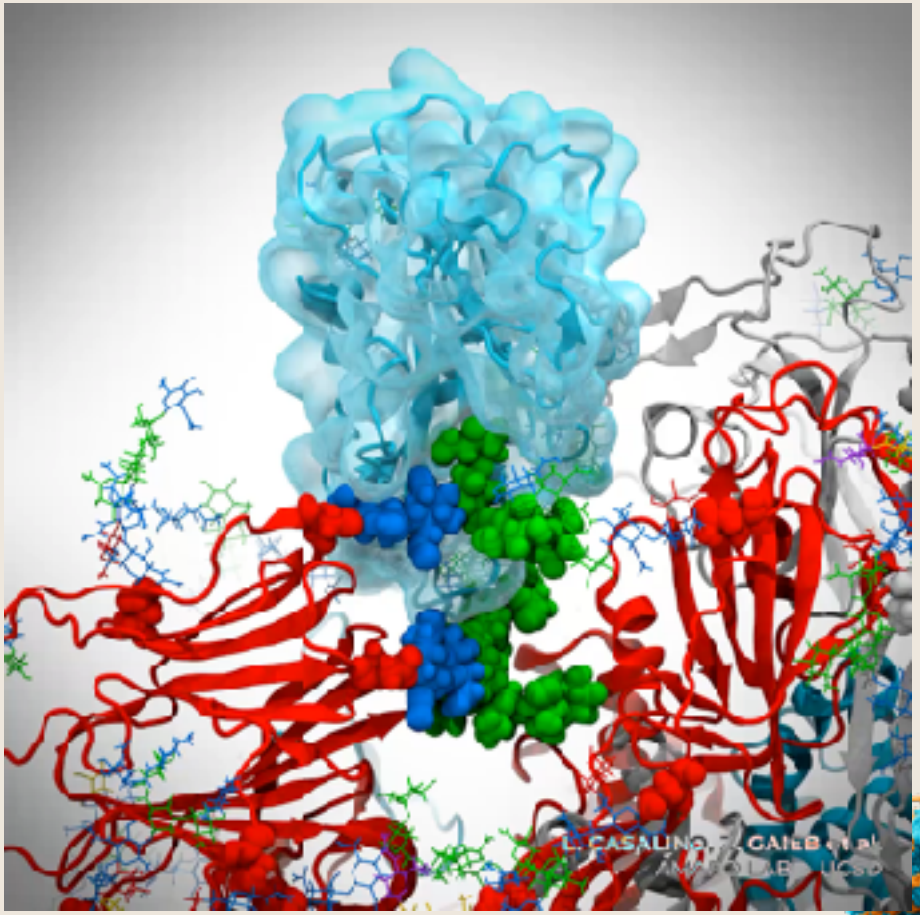
Casalino, Gaieb, Goldsmith, Hjorth, Dommer, Harbison, Fogarty, Barros, Taylor, McLellan, Fadda, 2021, ACS Central Science 6(10):1722-34.

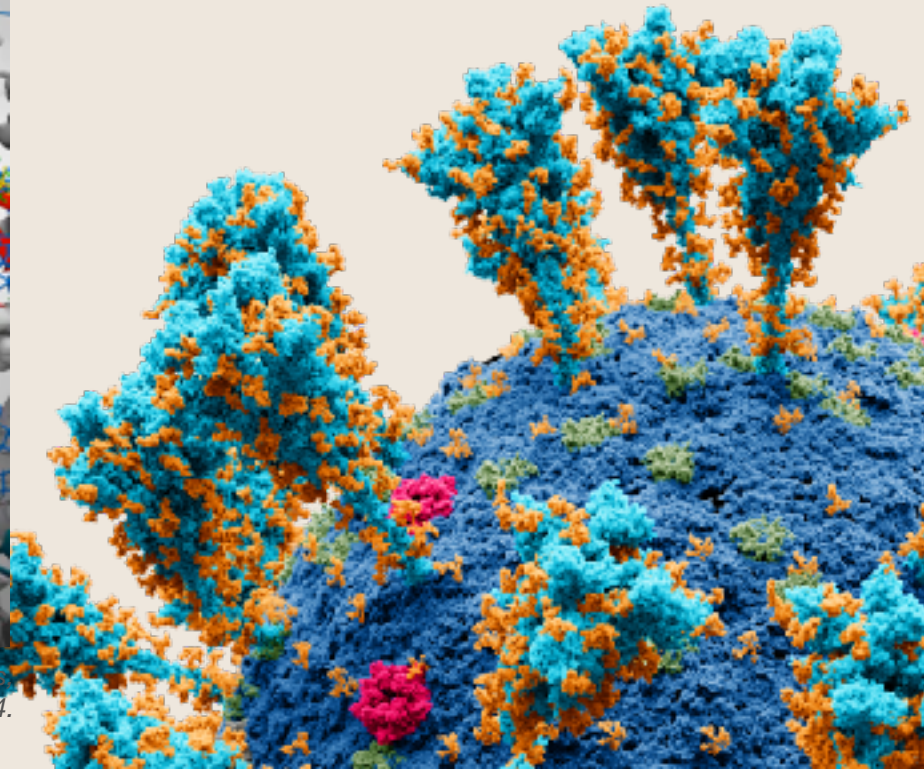


Glycans matter



Glycans matter





Casalino, Gaieb, Goldsmith, Hjorth, Dommer, Harbison, Fogarty, Barros, Taylor, McLellan, Fadda, 2021, ACS Central Science 6(10):1722-34.

Mechanistic studies show how things may work

Mechanistic studies show
how things may work

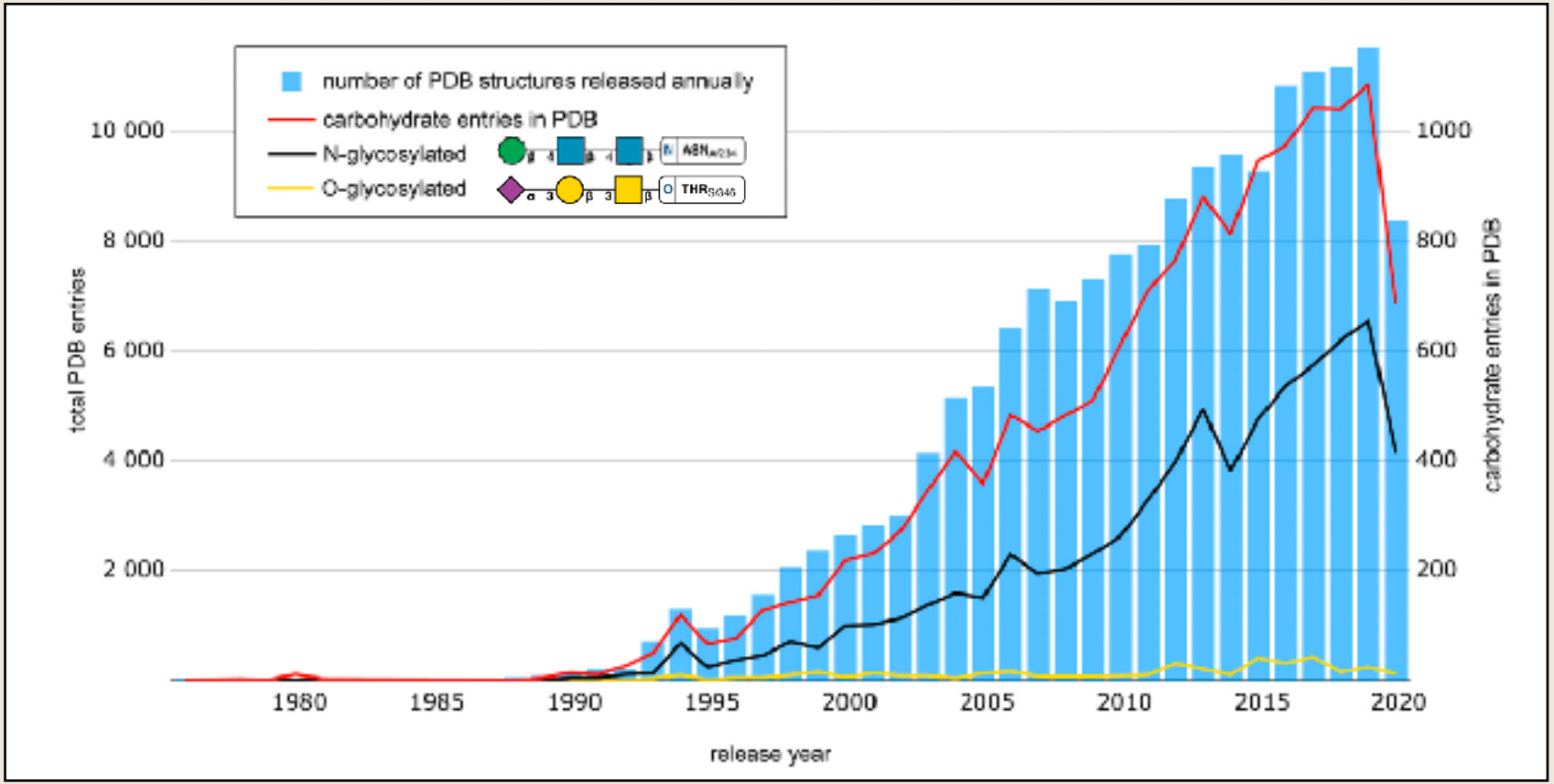
But simulations will only make sense if the atomic models used are correct

Mechanistic studies show
how things may work

But simulations will only make sense if the atomic models used are correct

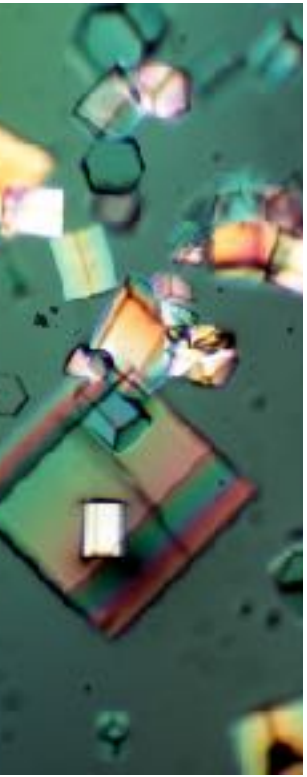
Are all atomic models correct?

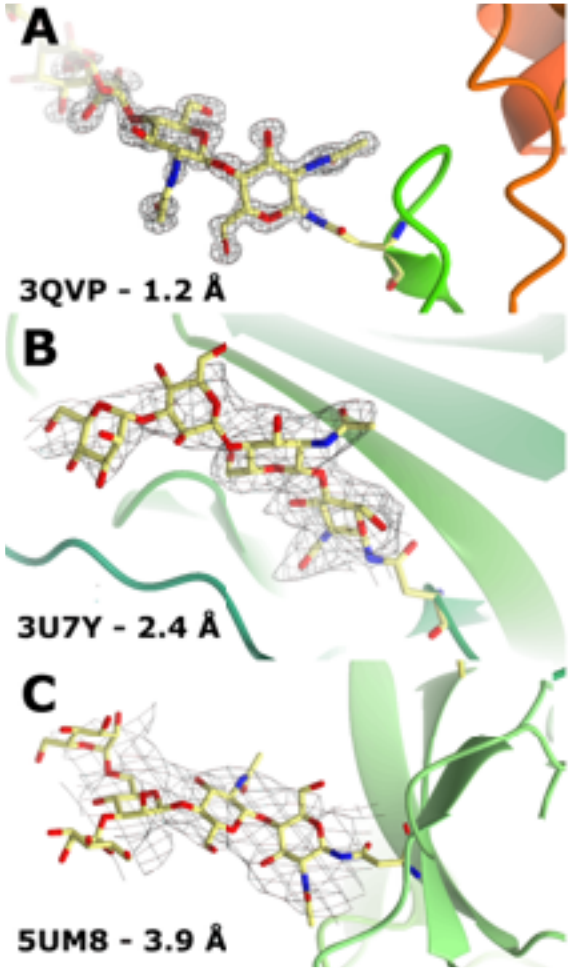
Carbohydrates in the PDB

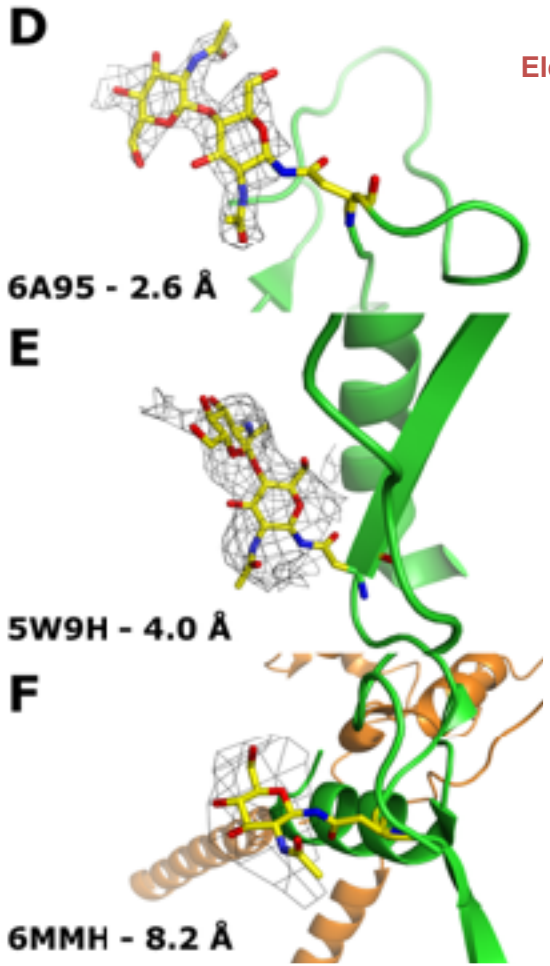


Scherbinina and Toukach, 2020 IJMS 21, 7702

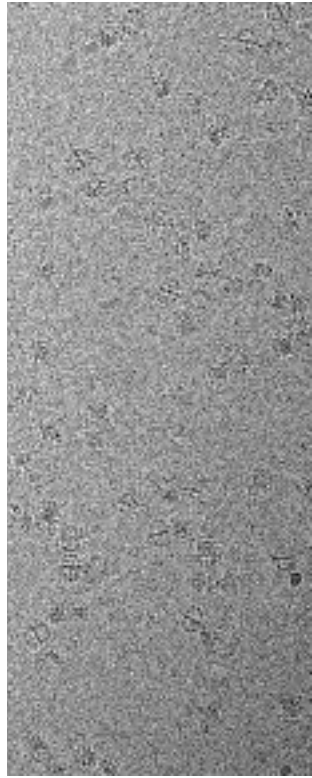
X-ray Crystallography







Electron Cryo-Microscopy



PDBe
Protein Data Bank in Europe

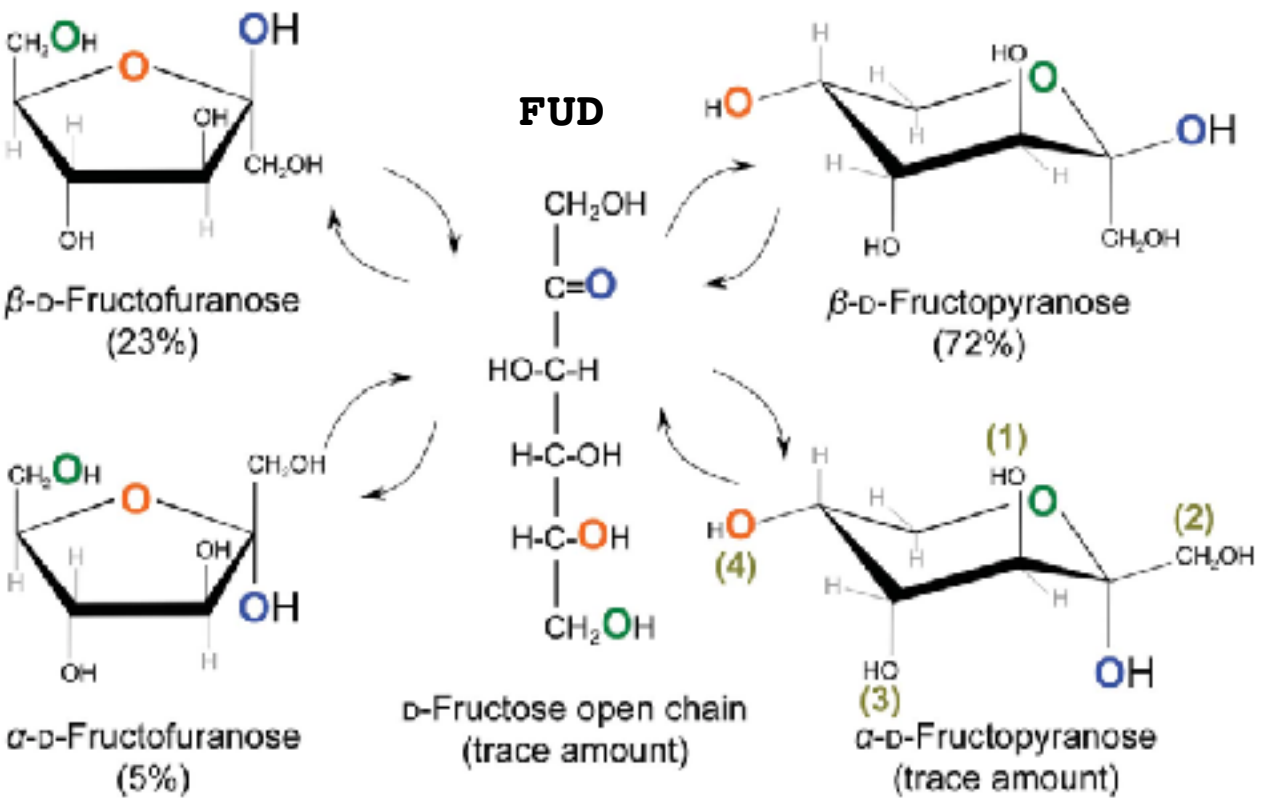
Adapted from Atanasova, Bagdonas & Agirre, 2020, Current Opinion in Structural Biology 62:70-78.

5 forms = 5 codes 👀

FRU

FUD

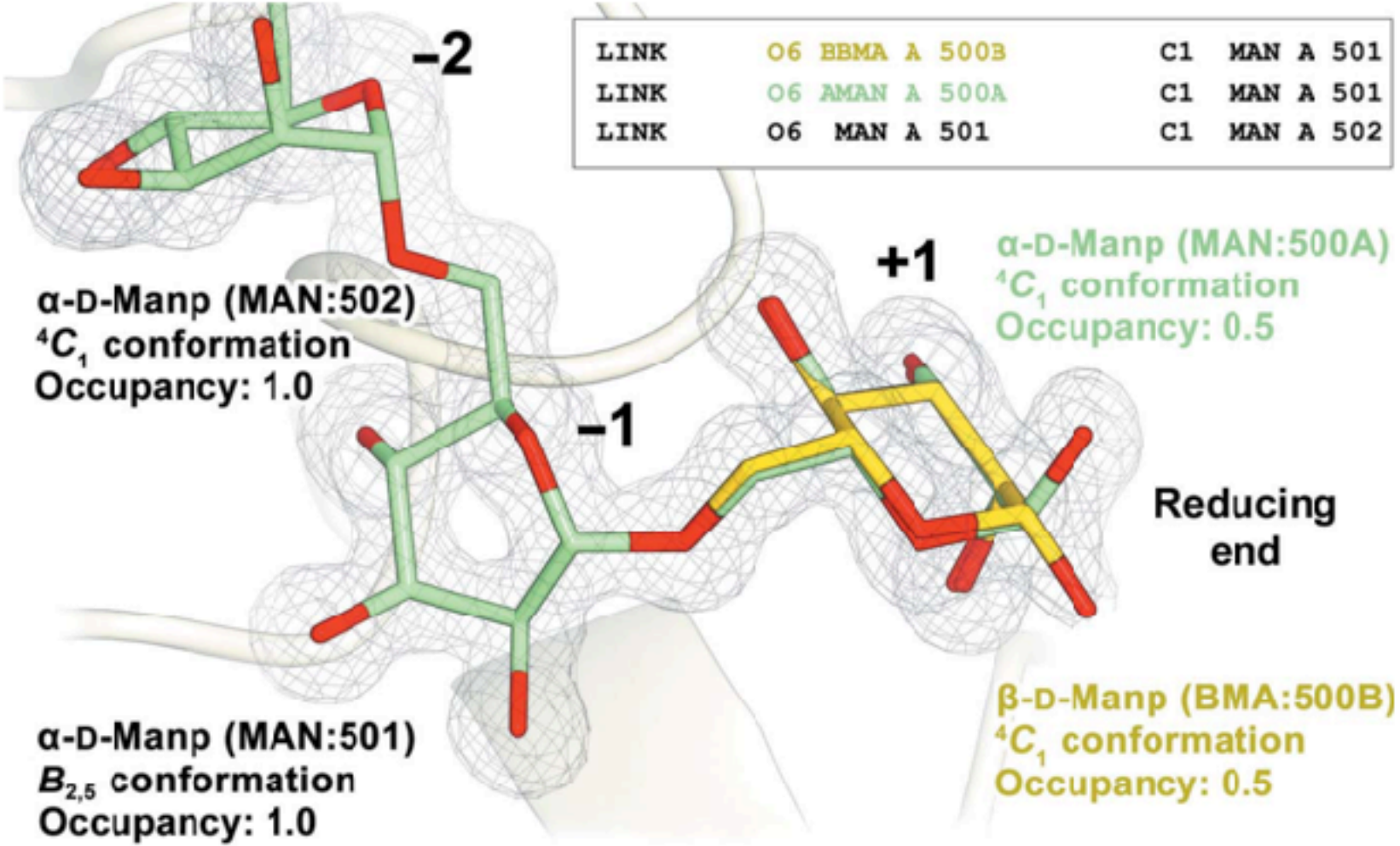
BDF



?

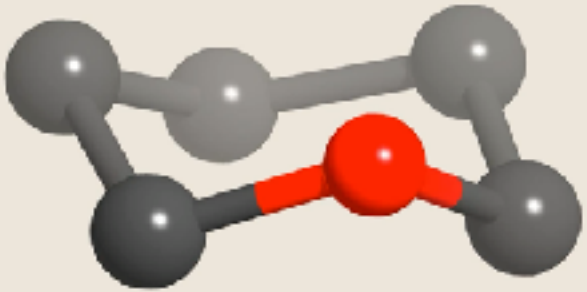
?

Agirre, 2017, Acta Cryst D73(2):171-186

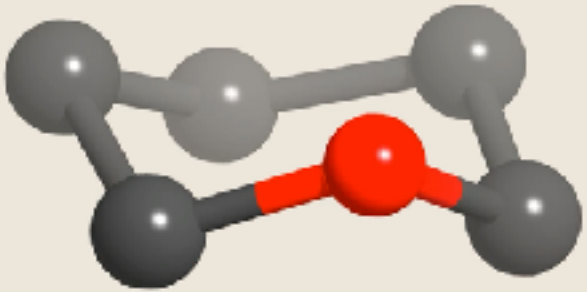


Adapted from Agirre, 2017, Acta Cryst D73(2):171-186

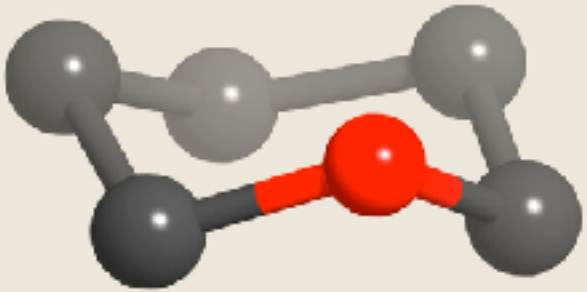
Ring conformation



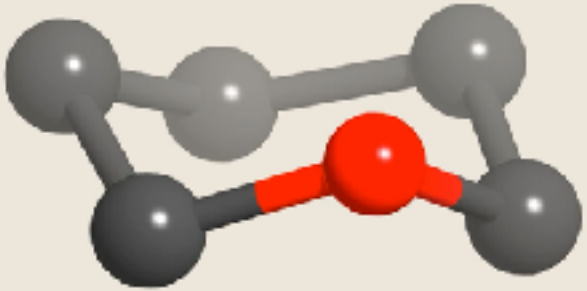
Ring conformation



Ring conformation



Ring conformation

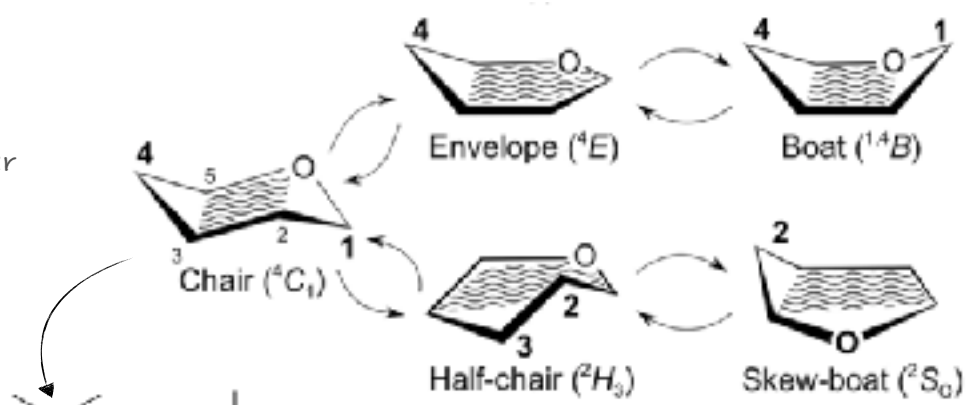


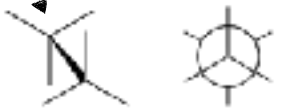
a chair

chairs are comfortable for sugars

Ring conformation

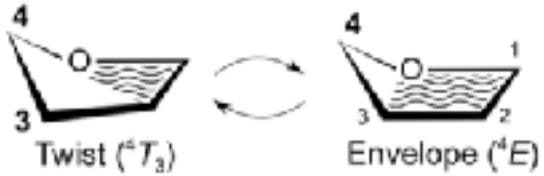
Wavy lines indicate which atoms are roughly coplanar

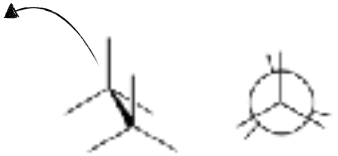




Staggered

Furanose forms
- 5-membered rings
- 2 ring puckers
- 20 conformations





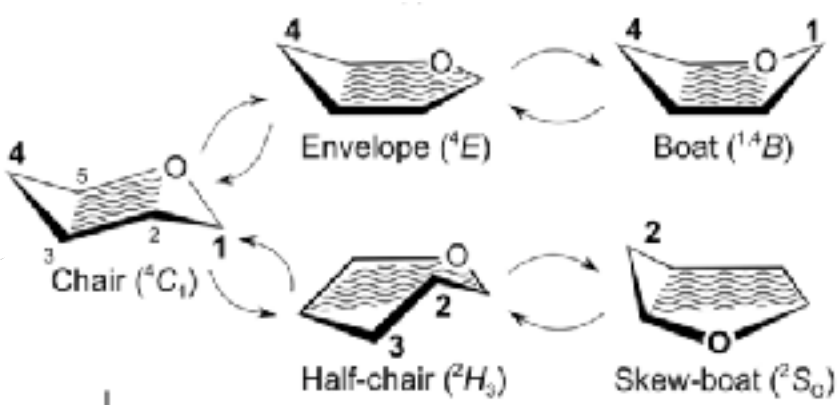
Eclipsed

Pyranose forms
- 6-membered rings
- 5 ring puckers
- 38 conformations

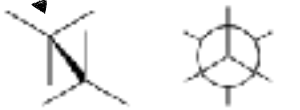
Adapted from Agirre, 2017, Acta Crystallographica D73(2):171-186.

Ring conformation

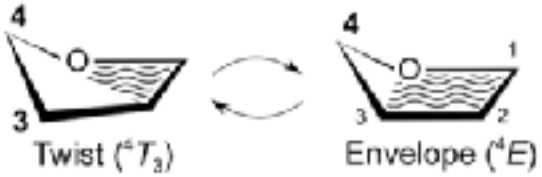
Wavy lines indicate which atoms are roughly coplanar



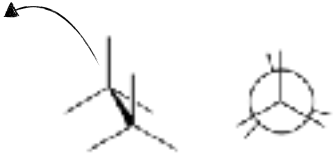
Furanose forms
- 5-membered rings
- 2 ring puckers
- 20 conformations



Staggered



Very strained :'(



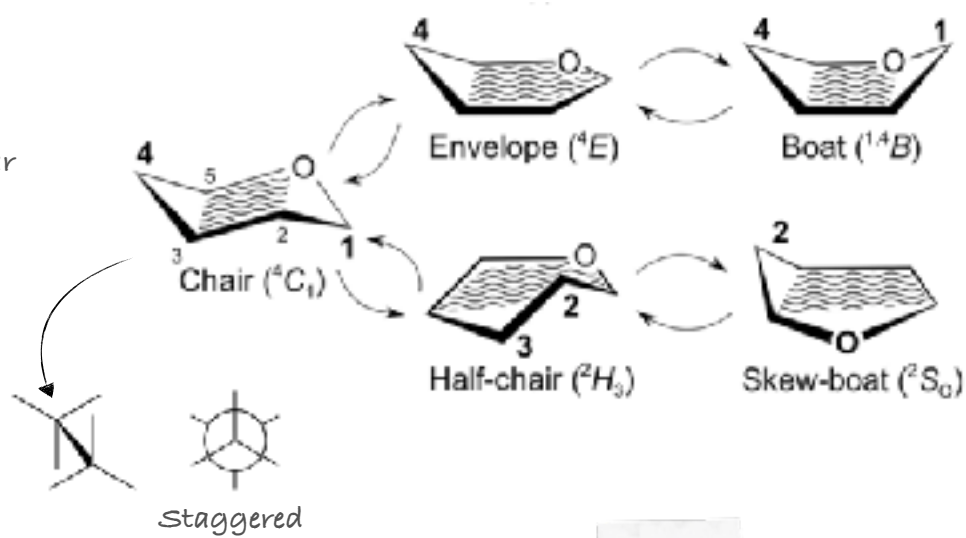
Eclipsed

Pyranose forms
- 6-membered rings
- 5 ring puckers
- 38 conformations

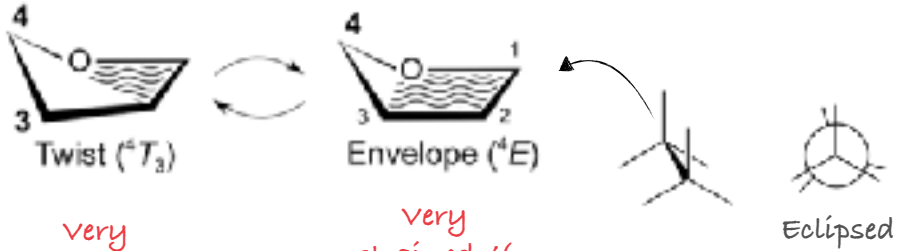
Adapted from Agirre, 2017, Acta Crystallographica D73(2):171-186.

Ring conformation

Wavy lines indicate which atoms are roughly coplanar



Furanose forms
- 5-membered rings
- 2 ring puckers
- 20 conformations



Very strained :'(

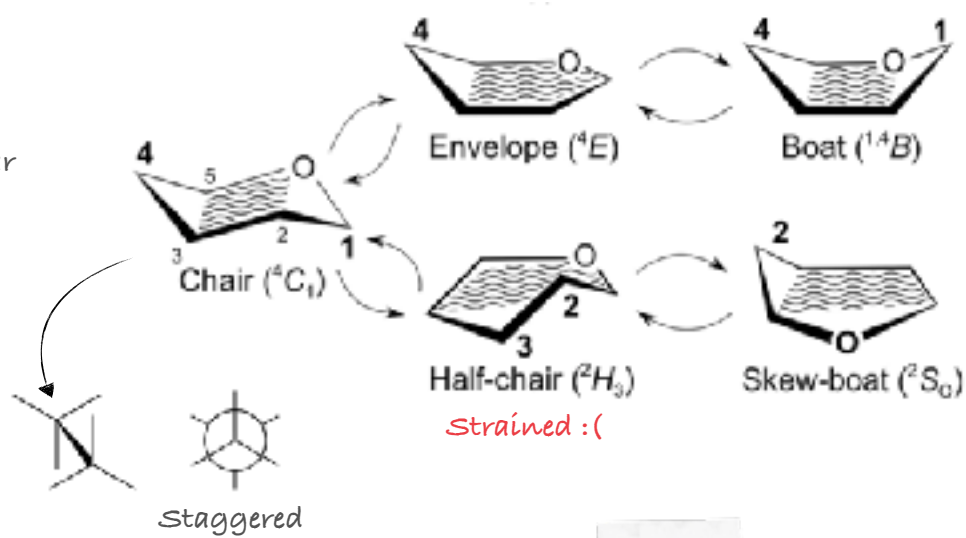
Very strained :'(

Pyranose forms
- 6-membered rings
- 5 ring puckers
- 38 conformations

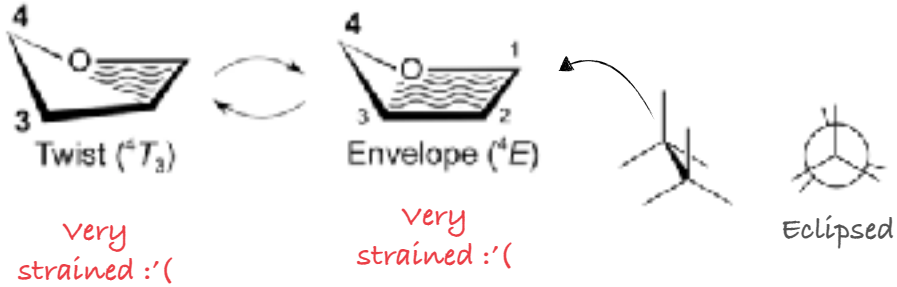
Adapted from Agirre, 2017, Acta Crystallographica D73(2):171-186.

Ring conformation

Wavy lines indicate which atoms are roughly coplanar



Furanose forms
- 5-membered rings
- 2 ring puckers
- 20 conformations

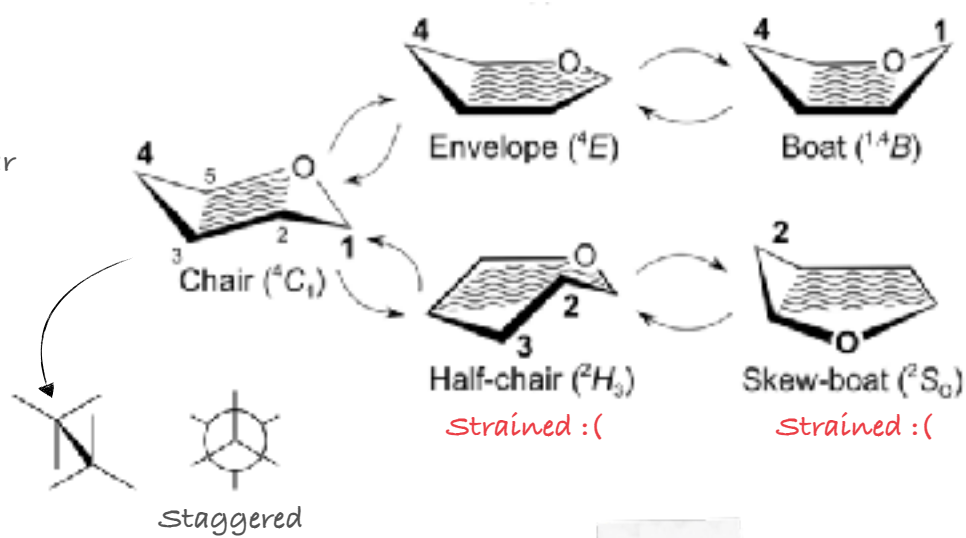


Pyranose forms
- 6-membered rings
- 5 ring puckers
- 38 conformations

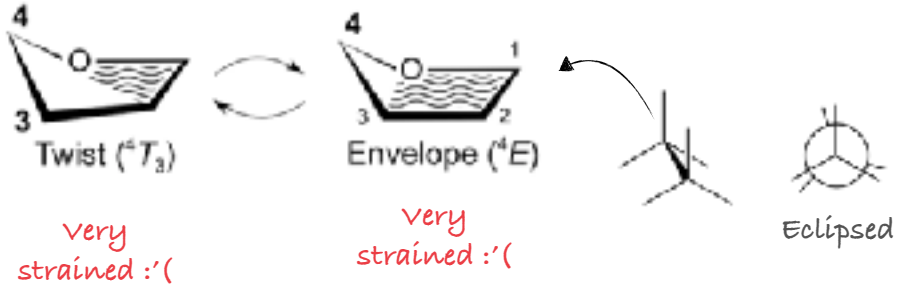
Adapted from Agirre, 2017, Acta Crystallographica D73(2):171-186.

Ring conformation

Wavy lines indicate which atoms are roughly coplanar



Furanose forms
- 5-membered rings
- 2 ring puckers
- 20 conformations

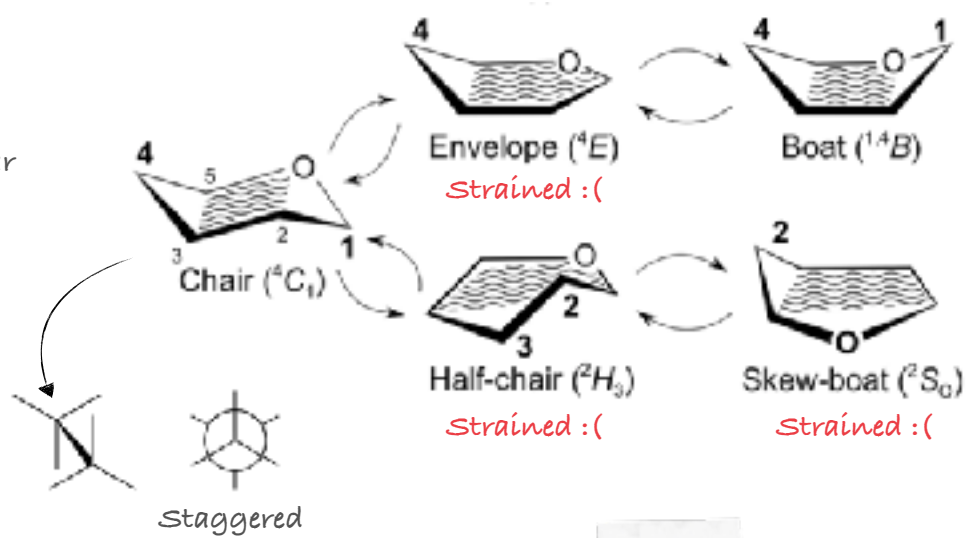


Pyranose forms
- 6-membered rings
- 5 ring puckers
- 38 conformations

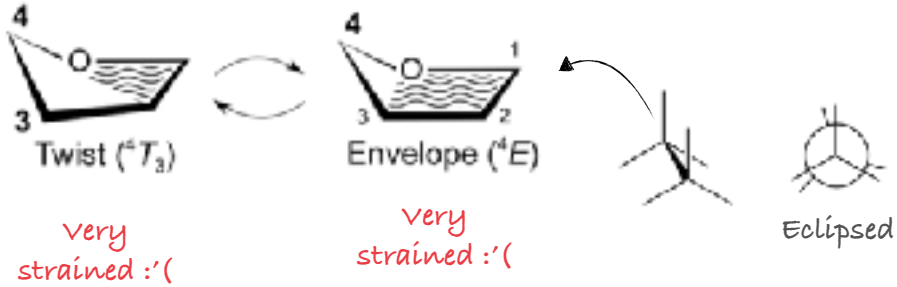
Adapted from Agirre, 2017, Acta Crystallographica D73(2):171-186.

Ring conformation

Wavy lines indicate which atoms are roughly coplanar



Furanose forms
- 5-membered rings
- 2 ring puckers
- 20 conformations

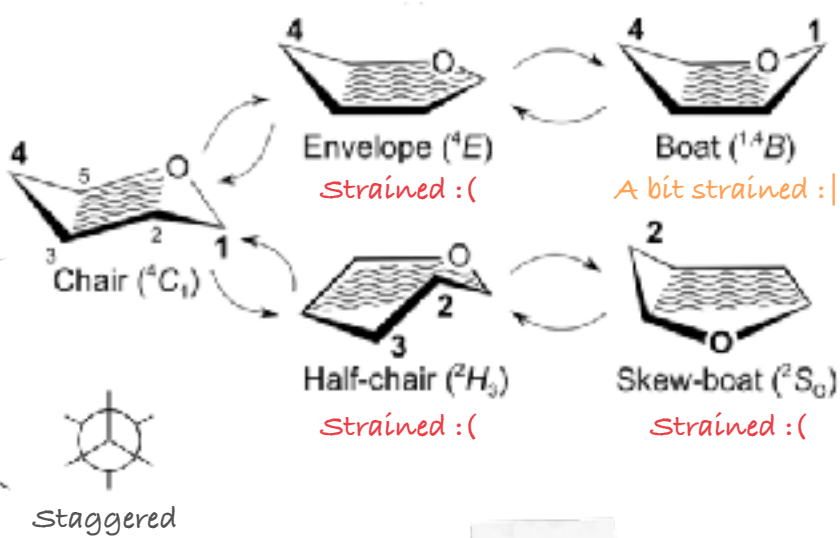


Pyranose forms
- 6-membered rings
- 5 ring puckers
- 38 conformations

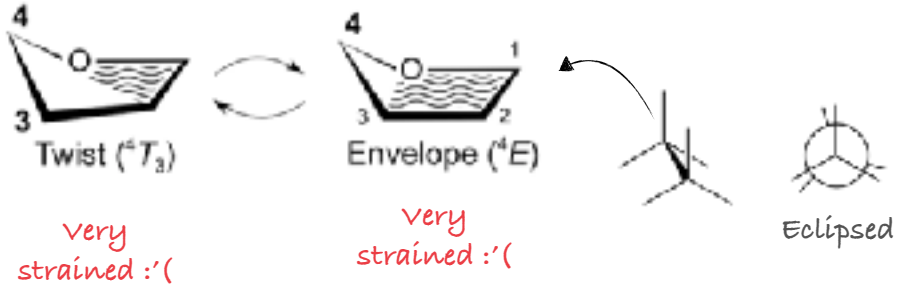
Adapted from Agirre, 2017, Acta Crystallographica D73(2):171-186.

Ring conformation

Wavy lines indicate which atoms are roughly coplanar



Furanose forms
- 5-membered rings
- 2 ring puckers
- 20 conformations

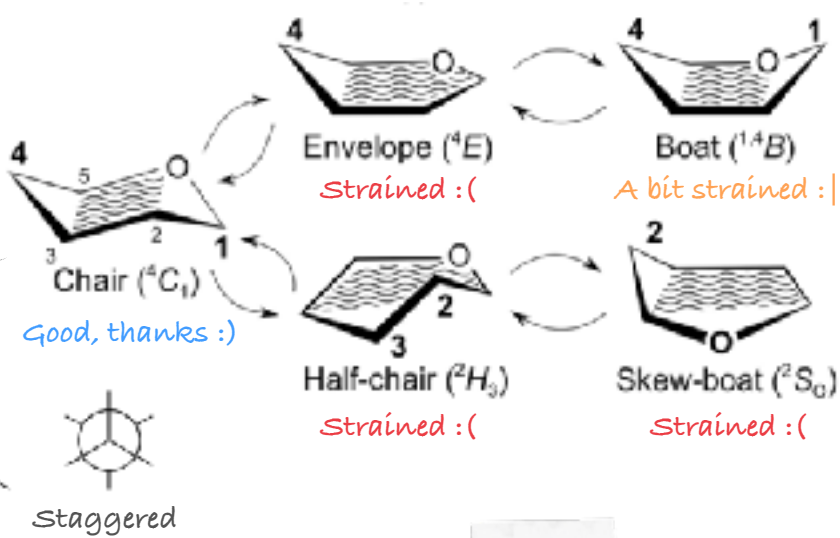


Pyranose forms
- 6-membered rings
- 5 ring puckers
- 38 conformations

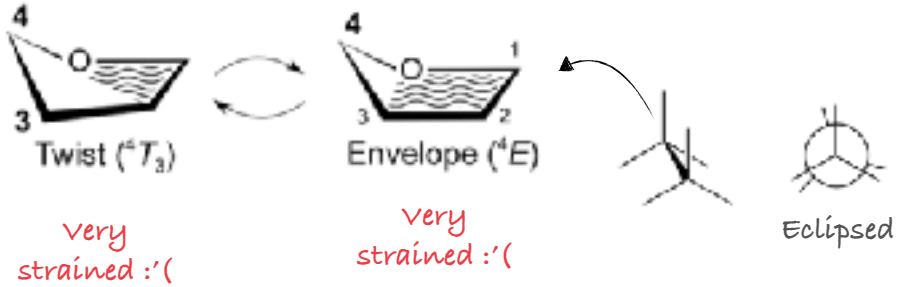
Adapted from Agirre, 2017, Acta Crystallographica D73(2):171-186.

Ring conformation

Wavy lines indicate which atoms are roughly coplanar



Furanose forms
- 5-membered rings
- 2 ring puckers
- 20 conformations



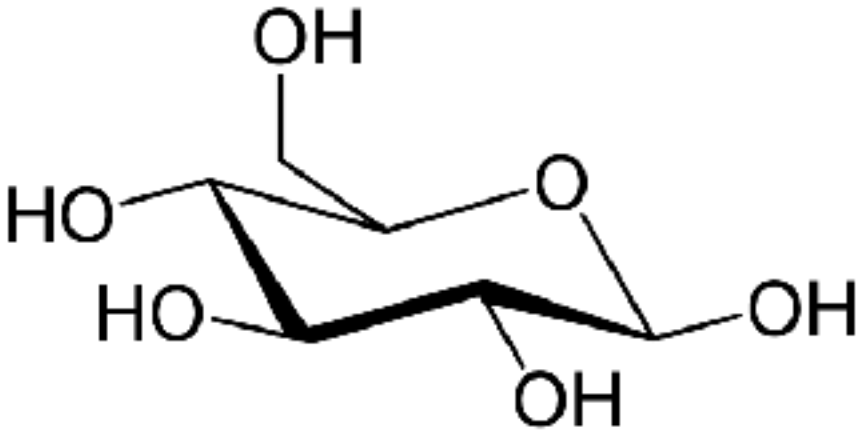
Pyranose forms
- 6-membered rings
- 5 ring puckers
- 38 conformations

Adapted from Agirre, 2017, Acta Crystallographica D73(2):171-186.

Ring conformation

The Cremer-Pople algorithm

Θ and Φ tell us which atoms move away from the mean ring plane, describing the conformation



Q tells us by how much:

$$Q = \sqrt{\sum_{i=0}^{5} Z_i^2}$$

"total puckering amplitude"

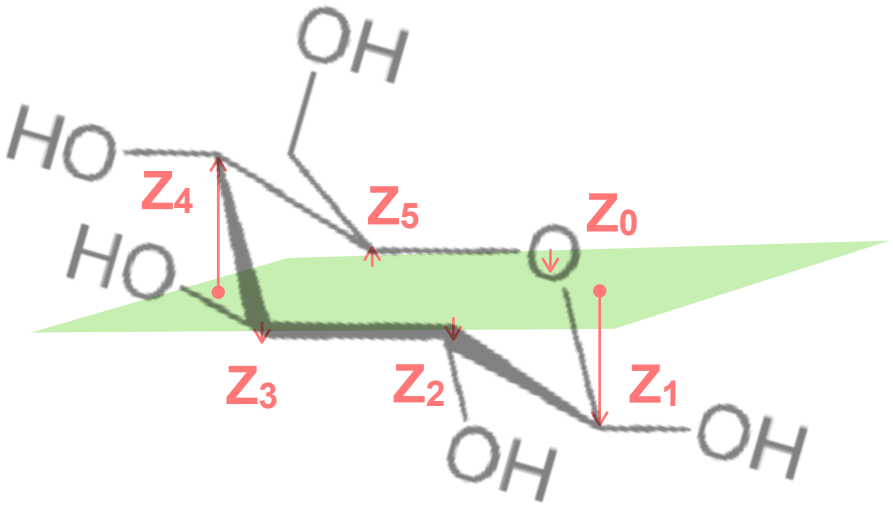
Q = 0.54Å for an ideal Glucose $^4C_1$ chair

Cremer & Pople, 1975, JACS 97(6)

Ring conformation

The Cremer-Pople algorithm

Θ and Φ tell us which atoms move away from the mean ring plane, describing the conformation



Q tells us by how much:

$$Q = \sqrt{\sum_{i=0}^{5} Z_i^2}$$

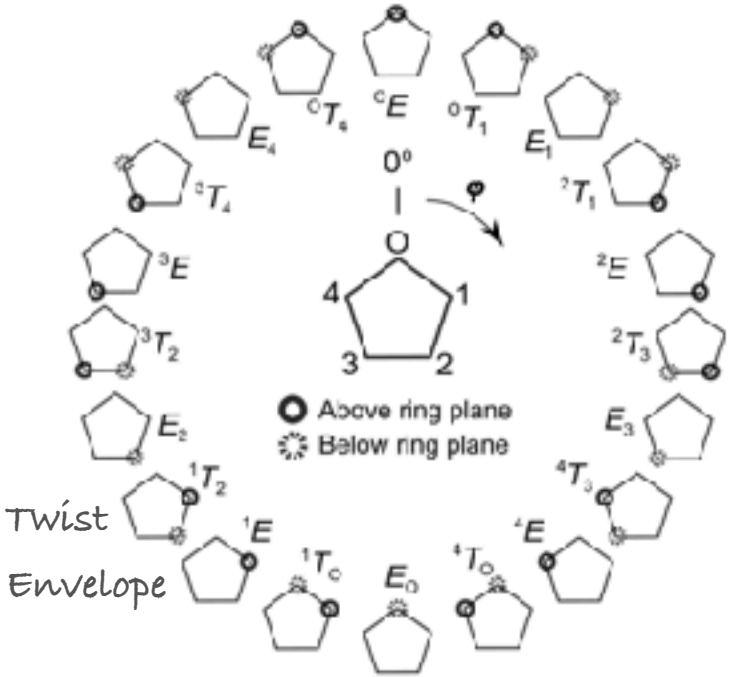
"total puckering amplitude"

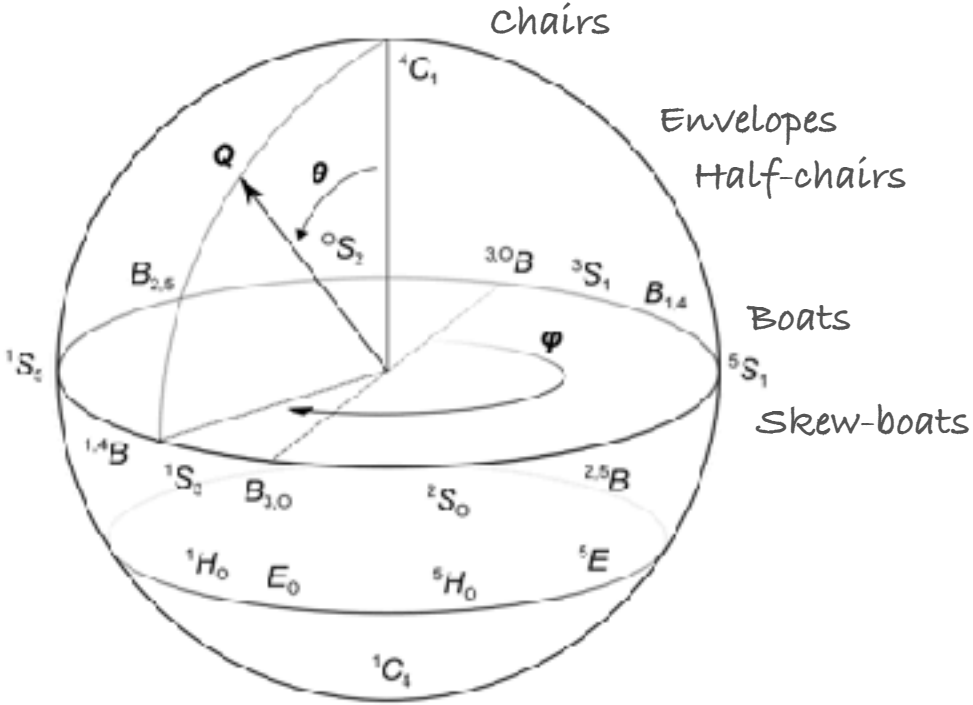
Q = 0.54Å for an ideal Glucose $^4C_1$ chair

Cremer & Pople, 1975, JACS 97(6)

Ring conformation

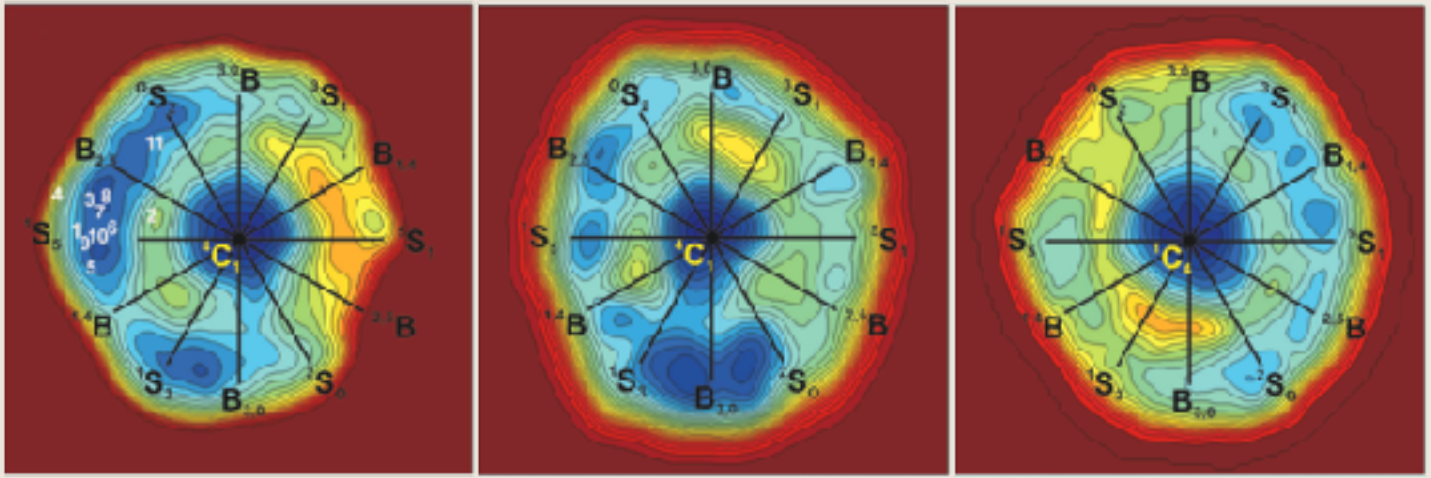
The Cremer-Pople algorithm

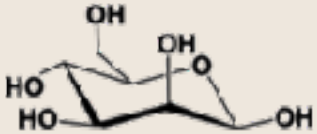




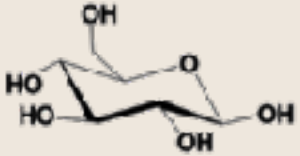
Agirre, 2017, Acta Cryst D73(2):171-186

Ring conformation

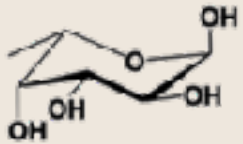




β-D-Mannopyranose



β-D-Glucopyranose

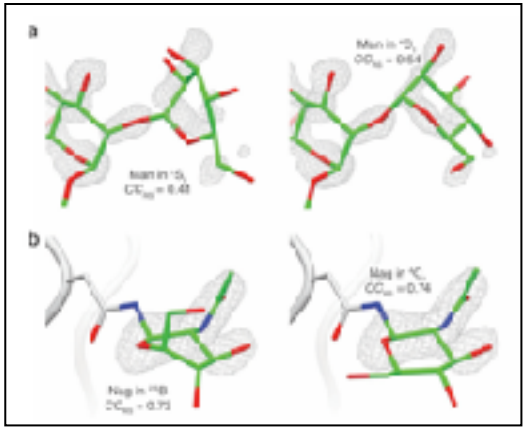


α-L-Fucopyranose

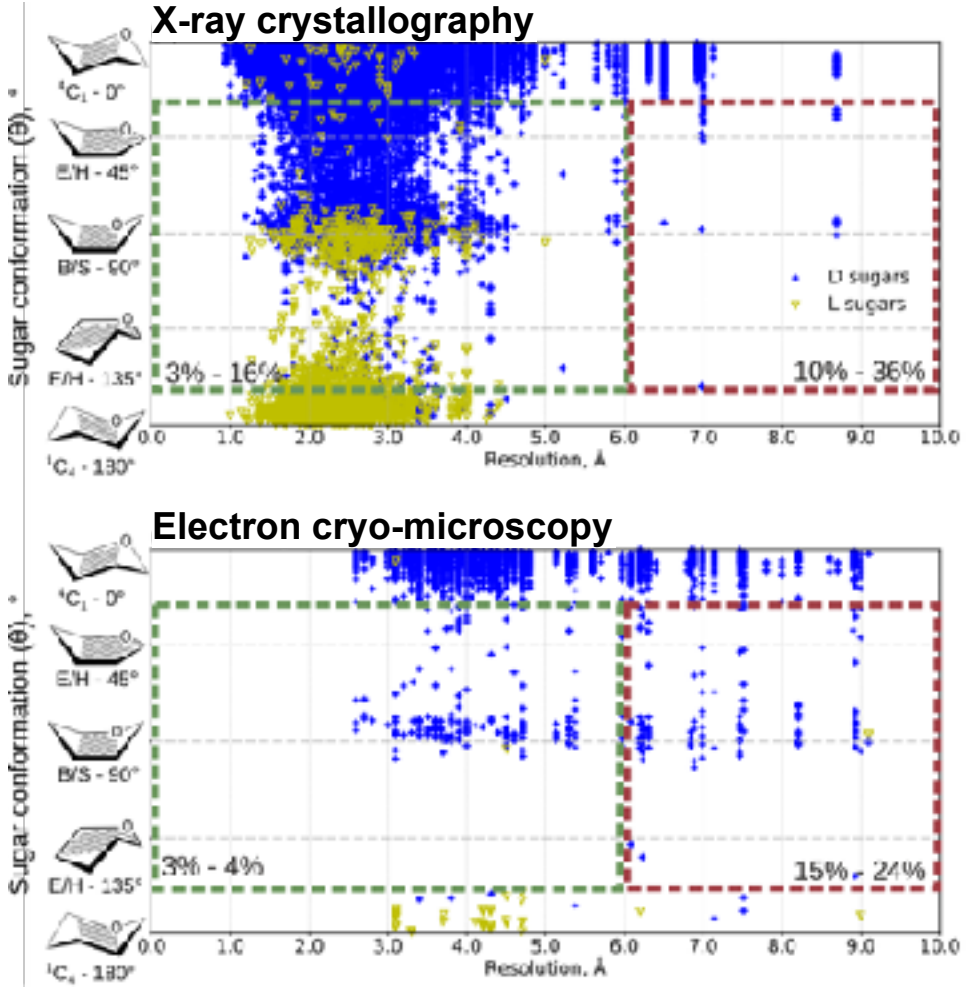
Adapted from Ardevol, Biarnes, Planas & Rovira, 2010, JACS 132(45).

Ring conformation in protein glycosylation

- Many more **high-energy** conformations than expected.
- Clear need **for carbohydrate specific methodology**.



Agirre, Davies, Wilson & Cowtan, 2015, Nature Structural & Molecular Biology 22(11):833-834.



Atanasova, Bagdonas & Agirre, 2020, Current Opinion in Structural Biology 62:70-78.

Publication and deposition of incorrect structures informs subsequent statistical analyses that suggest the deposited structures are normal

Agirre, Davies, Wilson & Cowtan, 2015, Nature Chemical Biology 11:303

Publication and deposition of incorrect structures informs subsequent statistical analyses that suggest the deposited structures are normal

Agirre, Davies, Wilson & Cowtan, 2015, Nature Chemical Biology 11:303

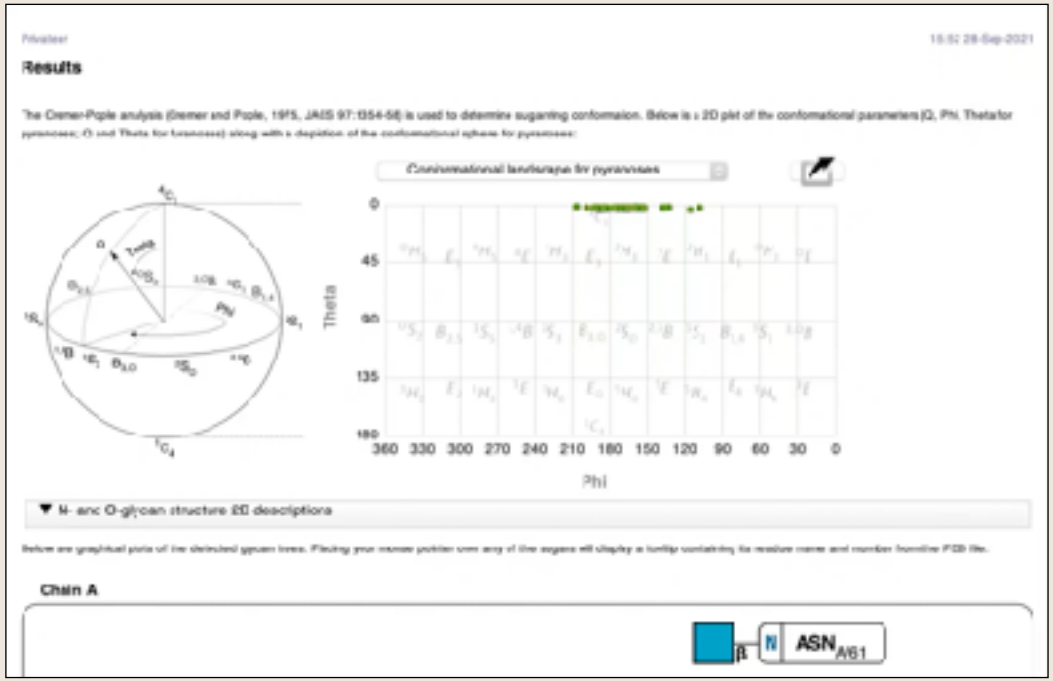
Publication and deposition of incorrect structures informs subsequent statistical analyses that suggest the deposited structures are normal

Agirre, Davies, Wilson & Cowtan, 2015, Nature Chemical Biology 11:303

Publication and deposition of incorrect structures informs subsequent statistical analyses that suggest the deposited structures are normal

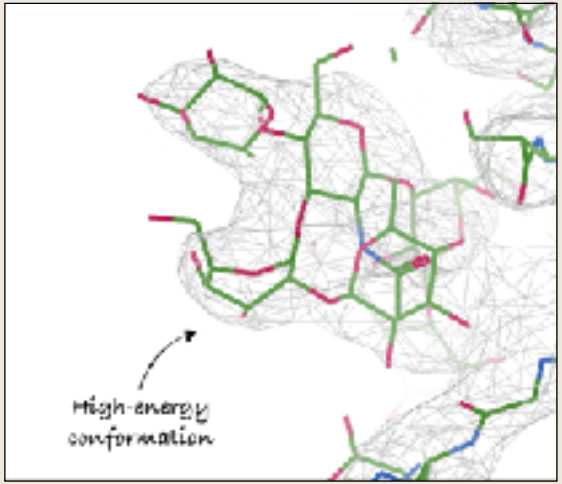
Agirre, Davies, Wilson & Cowtan, 2015, Nature Chemical Biology 11:303

Validation of structure, ring and link conformation of carbohydrate structures (MKIV)

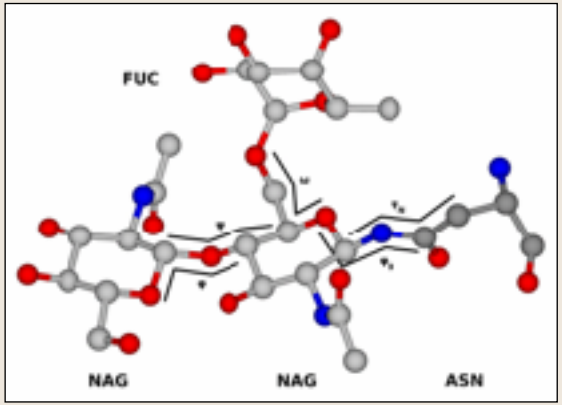


Bagdonas, Ungar & Agirre, 2020, Beilstein Journal of Organic Chemistry, 16(1):2523-2533.

Dictionaries for model improvement (MKIV)



Linkage torsion validation (MKV)

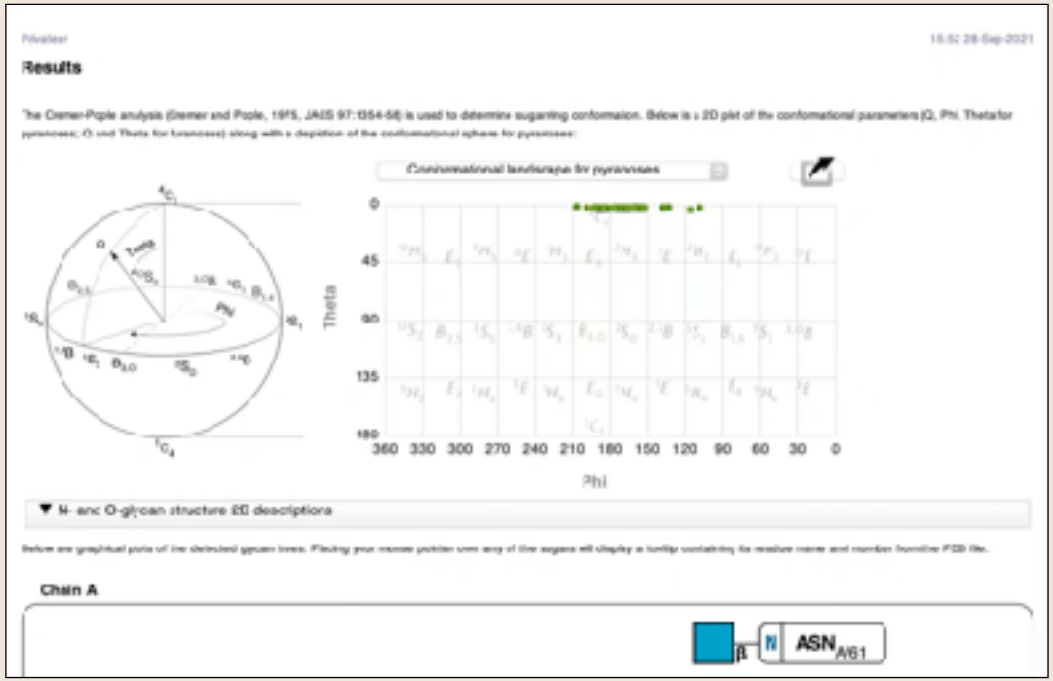


Privateer

Available in CCP4 8.0 and CCP-EM (nightly builds)

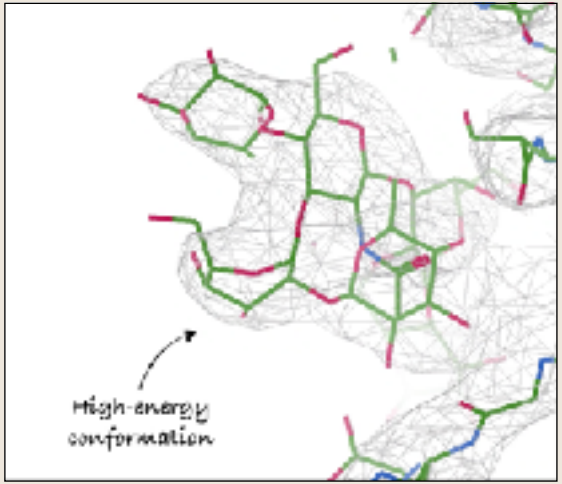
Or build it yourself and access the latest stuff:
https://github.com/glycojones/privateer

Validation of structure, ring and link conformation of carbohydrate structures (MKIV)

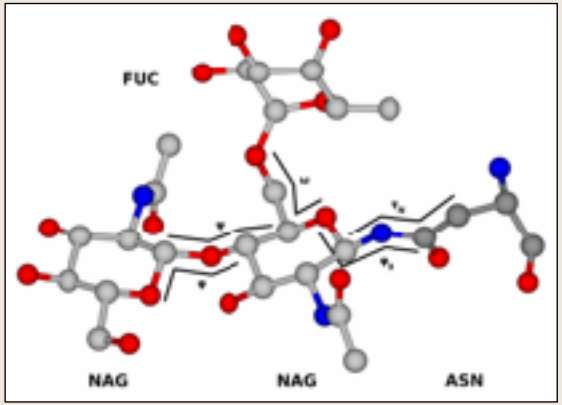


Bagdonas, Ungar & Agirre, 2020, Beilstein Journal of Organic Chemistry, 16(1):2523-2533.

Dictionaries for model improvement (MKIV)



Linkage torsion validation (MKV)

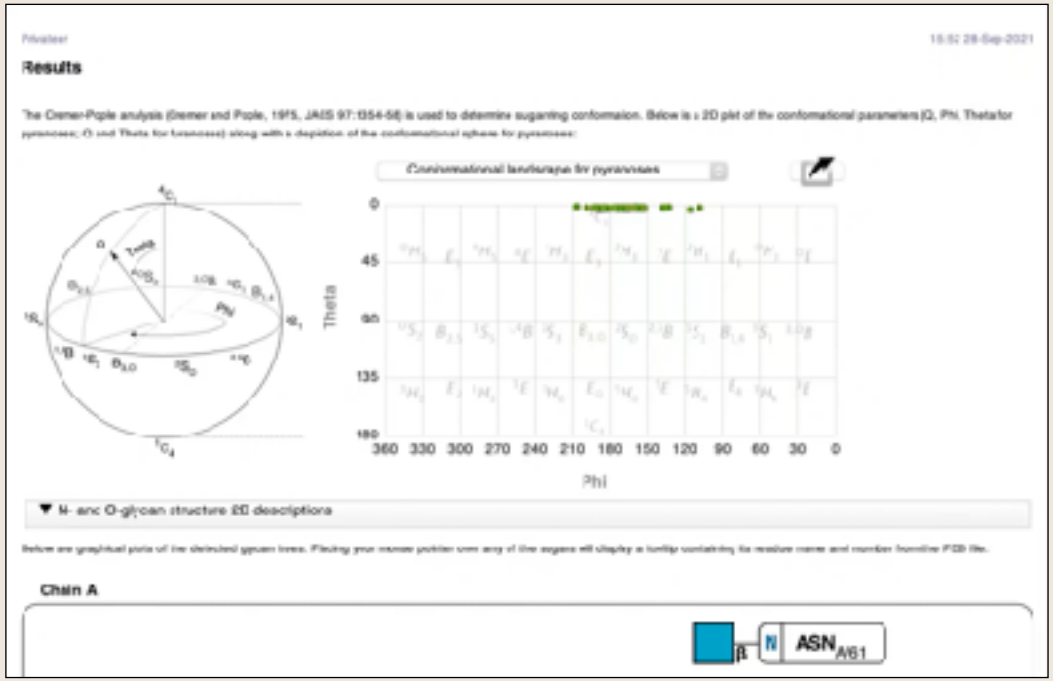


Privateer

Available in CCP4 8.0 and CCP-EM (nightly builds)

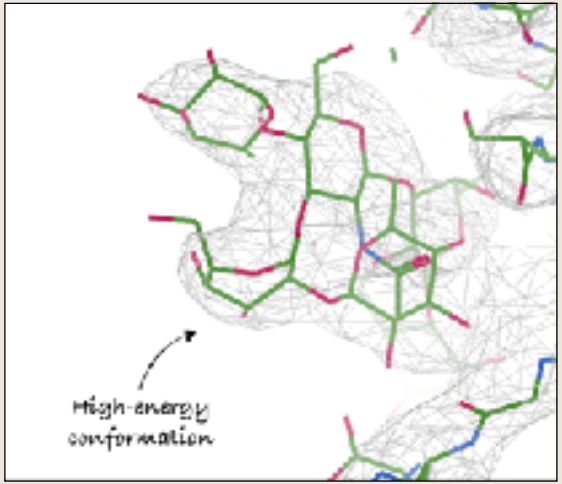
Or build it yourself and access the latest stuff:
https://github.com/glycojones/privateer

Validation of structure, ring and link conformation of carbohydrate structures (MKIV)

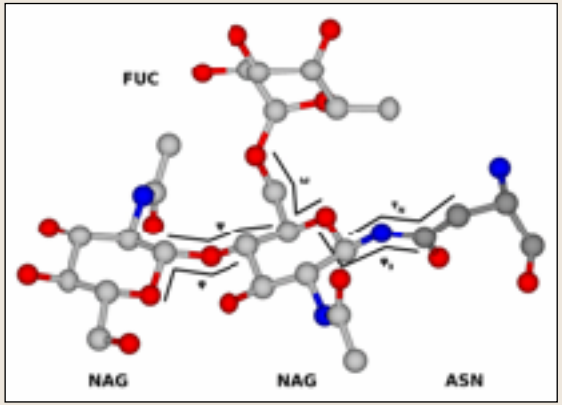


Bagdonas, Ungar & Agirre, 2020, Beilstein Journal of Organic Chemistry, 16(1):2523-2533.

Dictionaries for model improvement (MKIV)



Linkage torsion validation (MKV)



Privateer

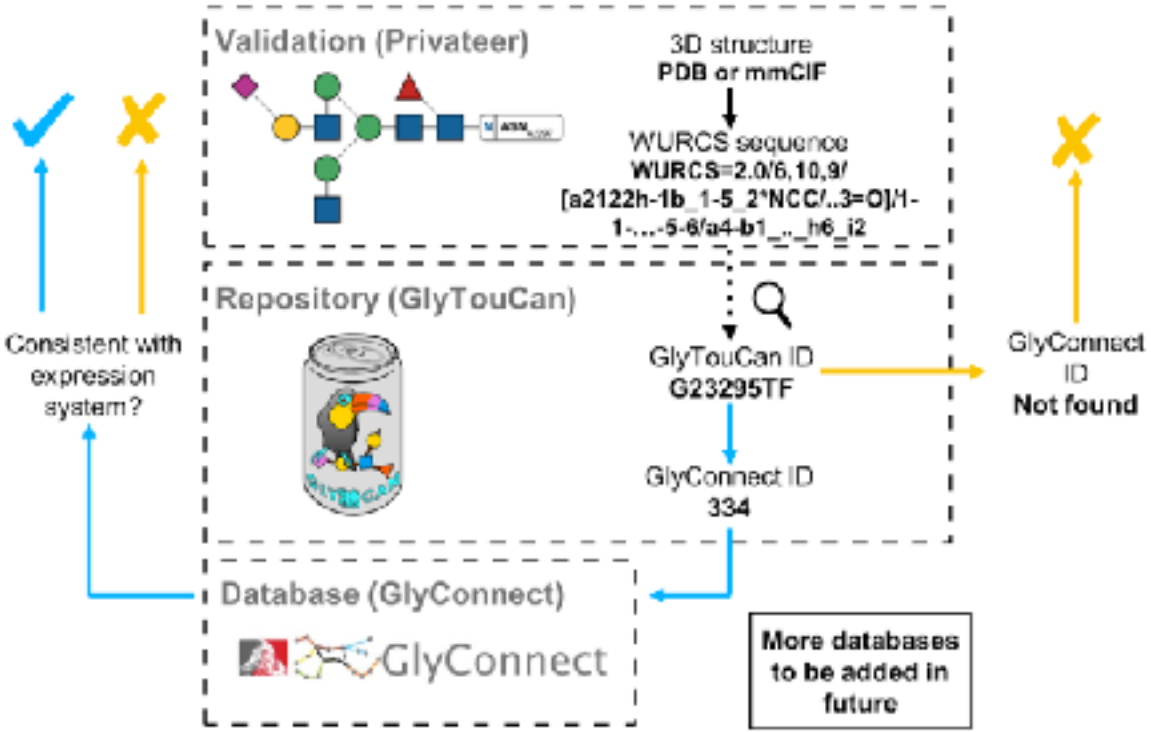
Available in CCP4 8.0 and CCP-EM (nightly builds)

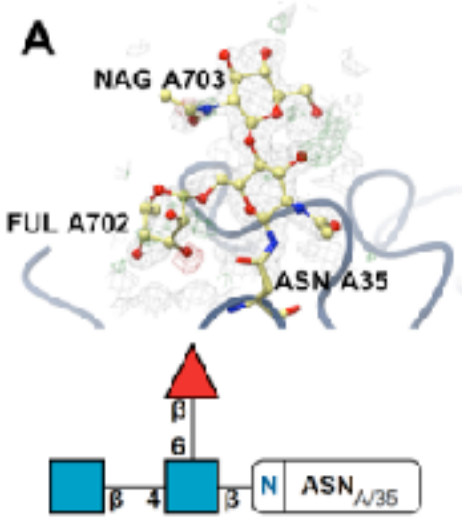
Or build it yourself and access the latest stuff:
https://github.com/glycojones/privateer

The Privateer software

- Has been useful in **every relevant context** where structural glycobiology has played a role: biofuels, immunology, cancer, fecundation and, of course, viral glycoproteins (HIV, Ebolavirus, Influenza, MERS, SARS-CoV-1 & 2).
- COVID-19 pandemic: used in constructs informing **mRNA vaccine design**, structures of **spike-antibody complexes** and **neutralising drugs**. Integrated in several analysis and validation pipelines.
- Released with main UK structural biology software suites: CCP4 (MKIV available in 8.0, MKV soon via update) and CCP-EM. New Python bindings (MKV) will bring fully interactive graphical interfaces for Coot and ChimeraX. A web server for *Expasy* is planned too.

Glycomics powered validation





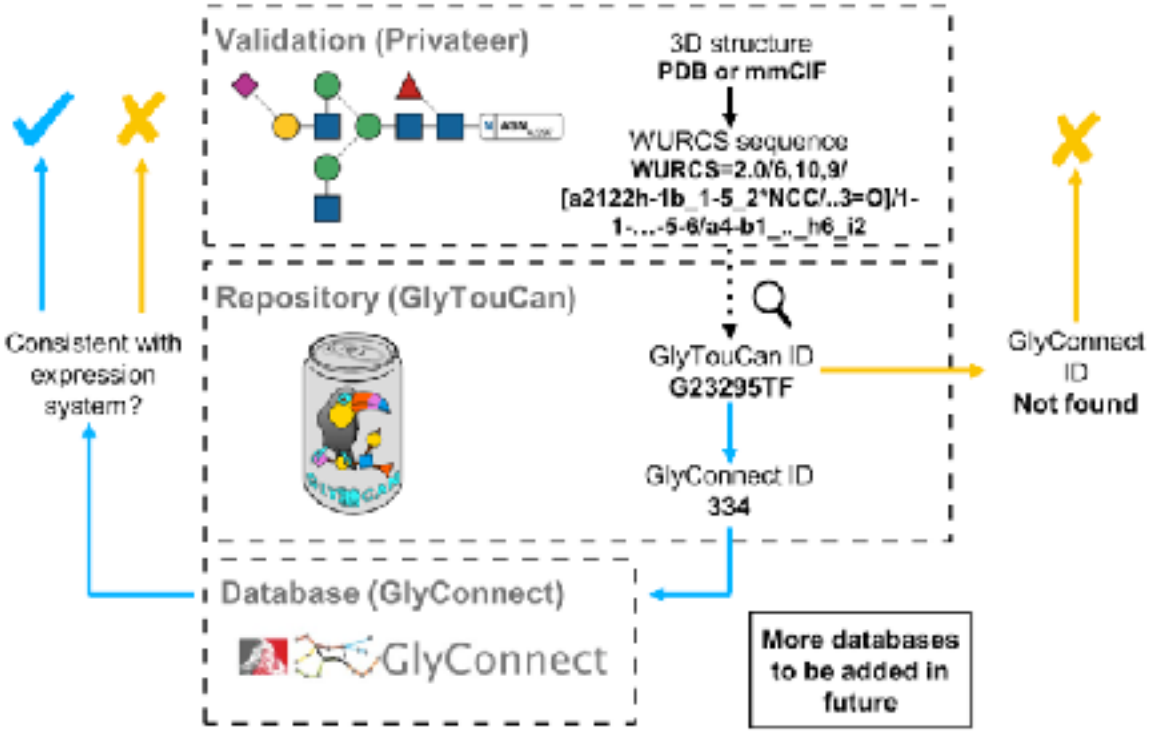
WURCS=2.0/2,3,2/[a2122h-1b_1-5_2*NCC/3=O][a1221m-1b_1-5]/1-1-2/a4-b1_a6-c1

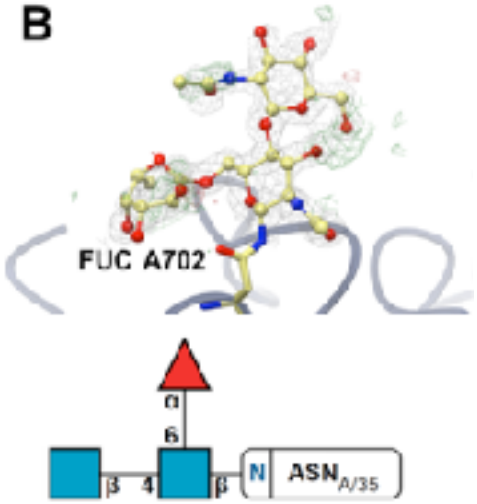
GlyTouCan ID: G28454KX

GlyConnect ID: Not Found.

Bagdonas, Ungar & Agirre, 2020, Beilstein Journal of Organic Chemistry, 16(1):2523-2533.

Glycomics powered validation





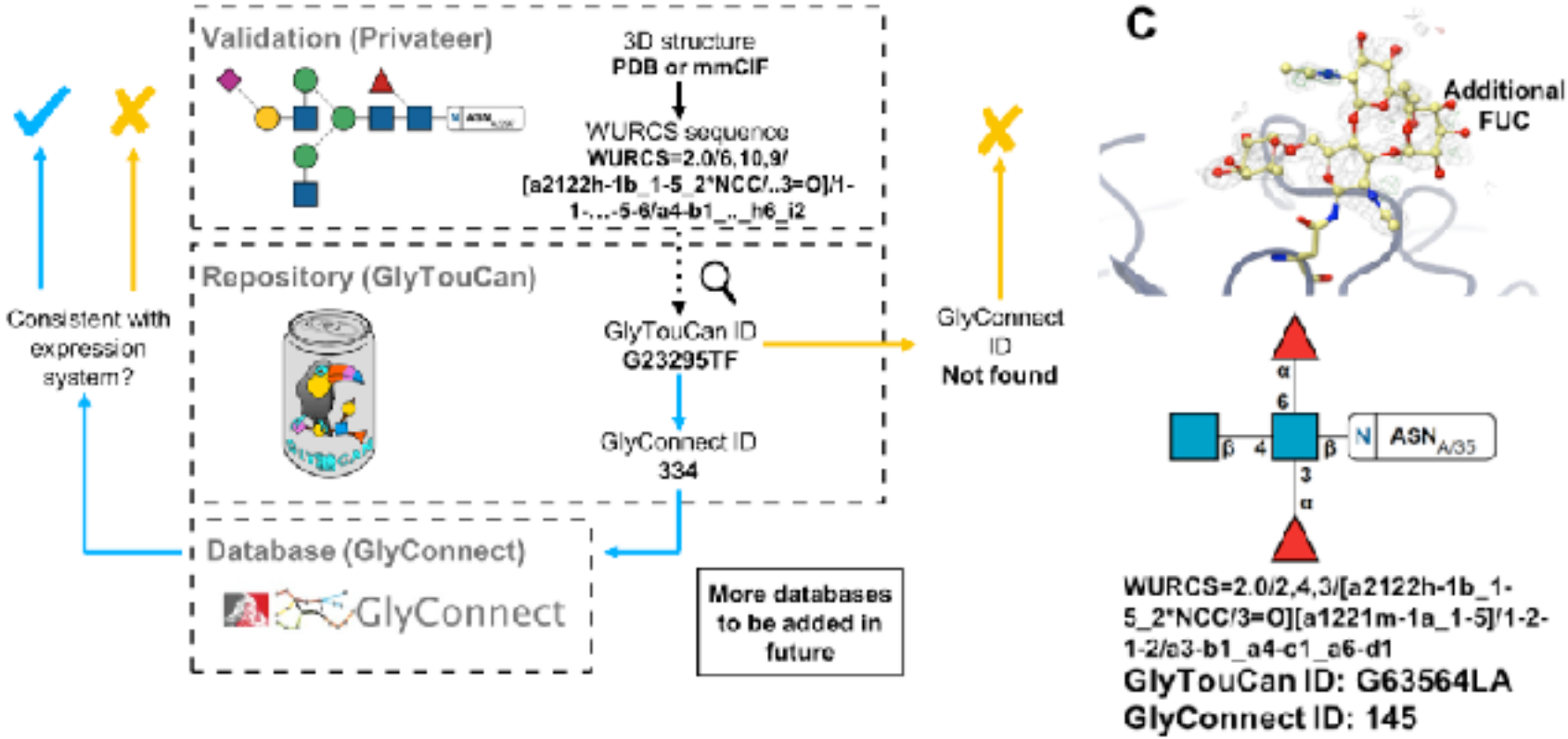
WURCS=2.0/2,3,2/[a2122h-1b_1-5_2*NCC/3=O][a1221m-1a_1-5]/1-1-2/a4-b1_a6-c1

GlyTouCan ID: G21290RB

GlyConnect ID: 54

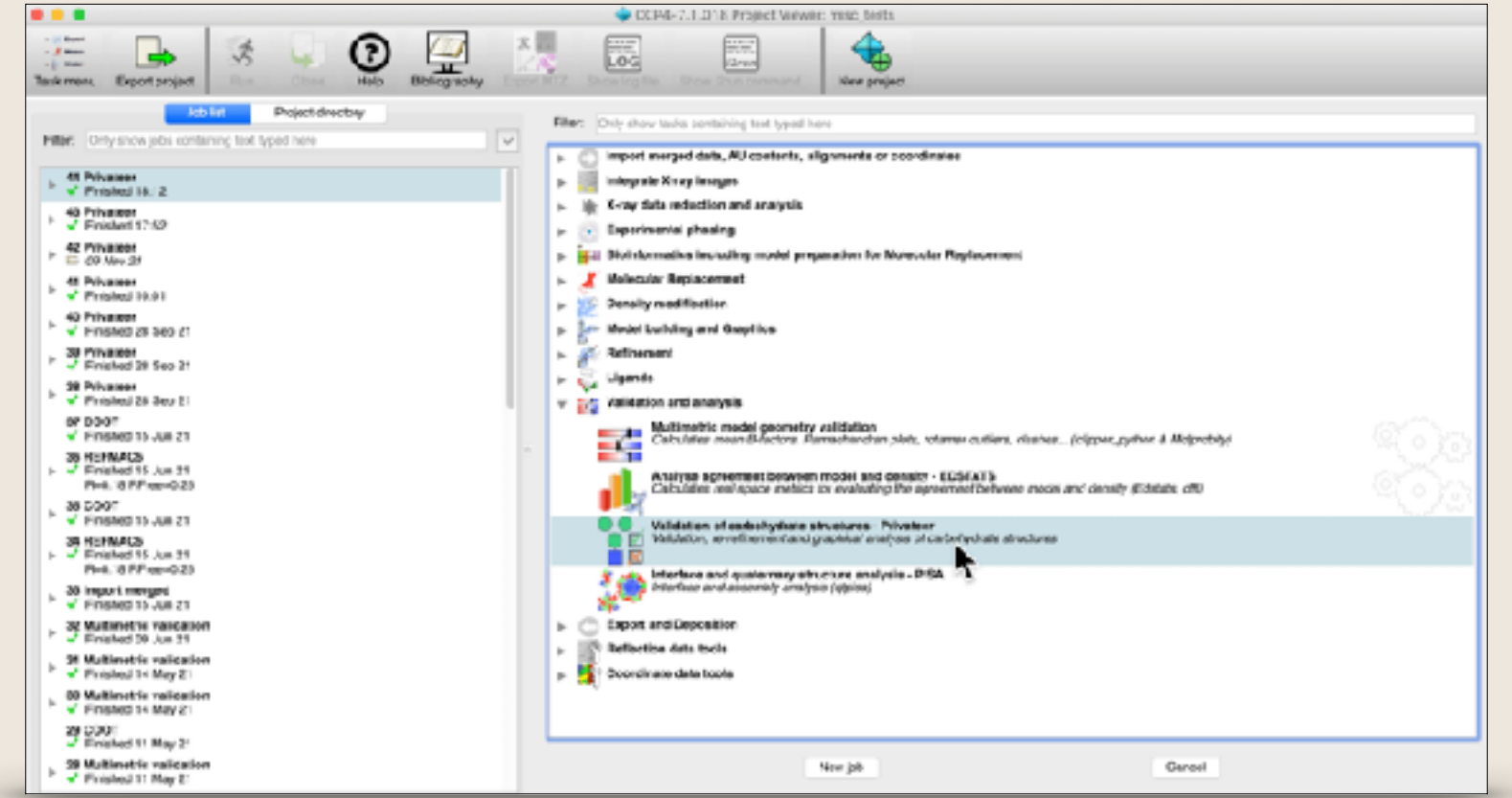
Bagdonas, Ungar & Agirre, 2020, Beilstein Journal of Organic Chemistry, 16(1):2523-2533.

Glycomics powered validation

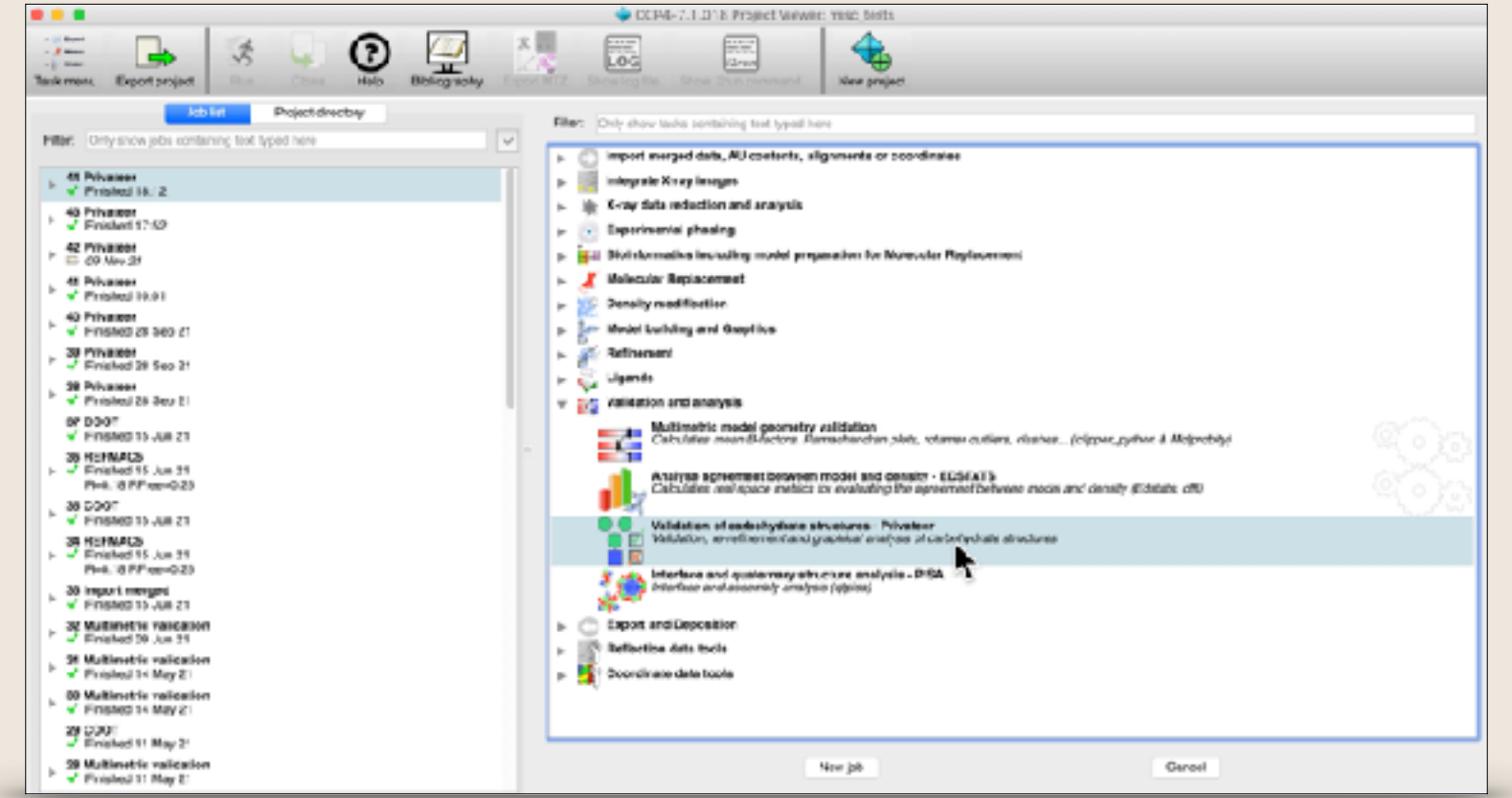


Bagdonas, Ungar & Agirre, 2020, Beilstein Journal of Organic Chemistry, 16(1):2523-2533.

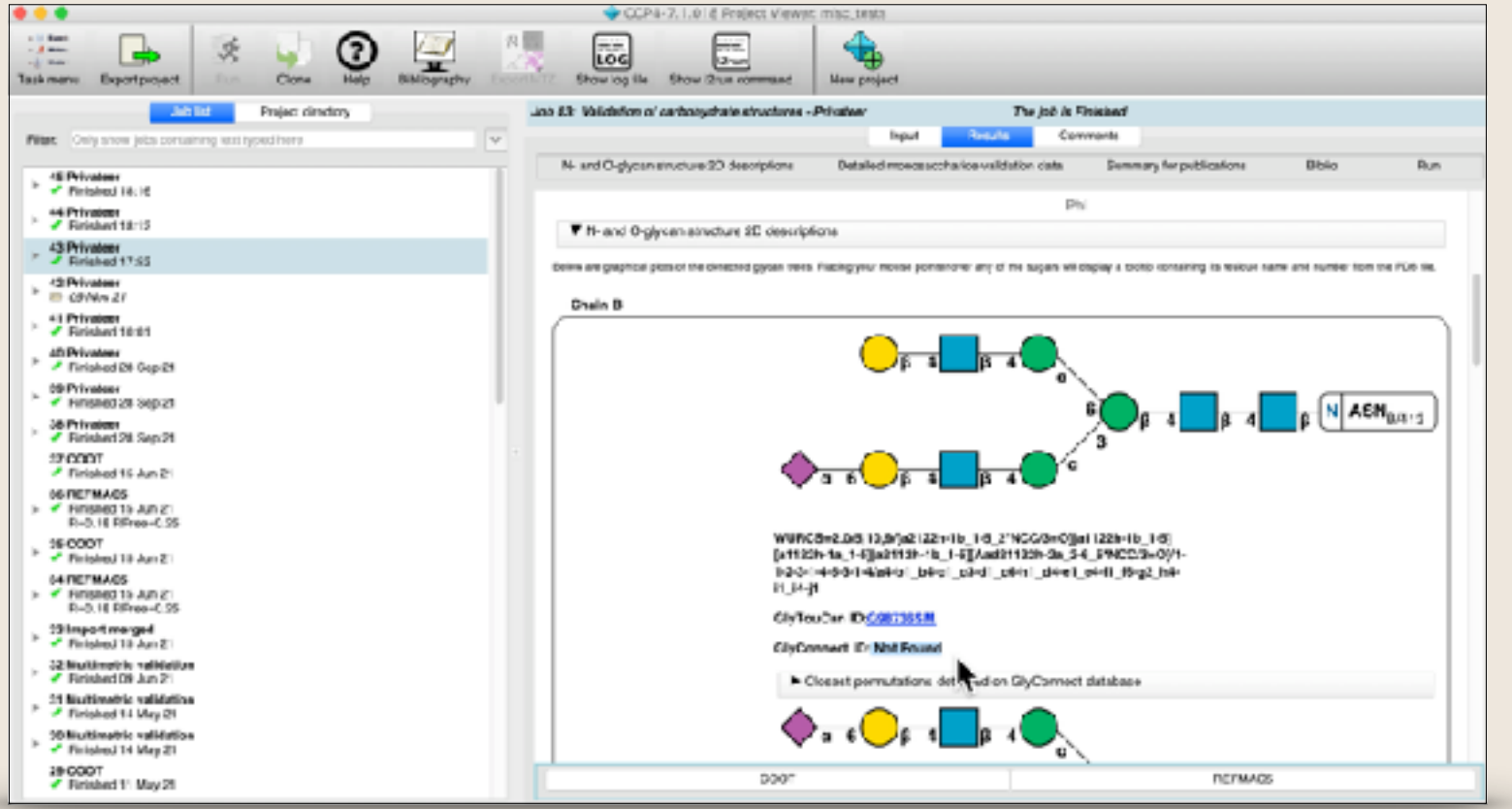
Running Privateer MKIV (CCP4i2)



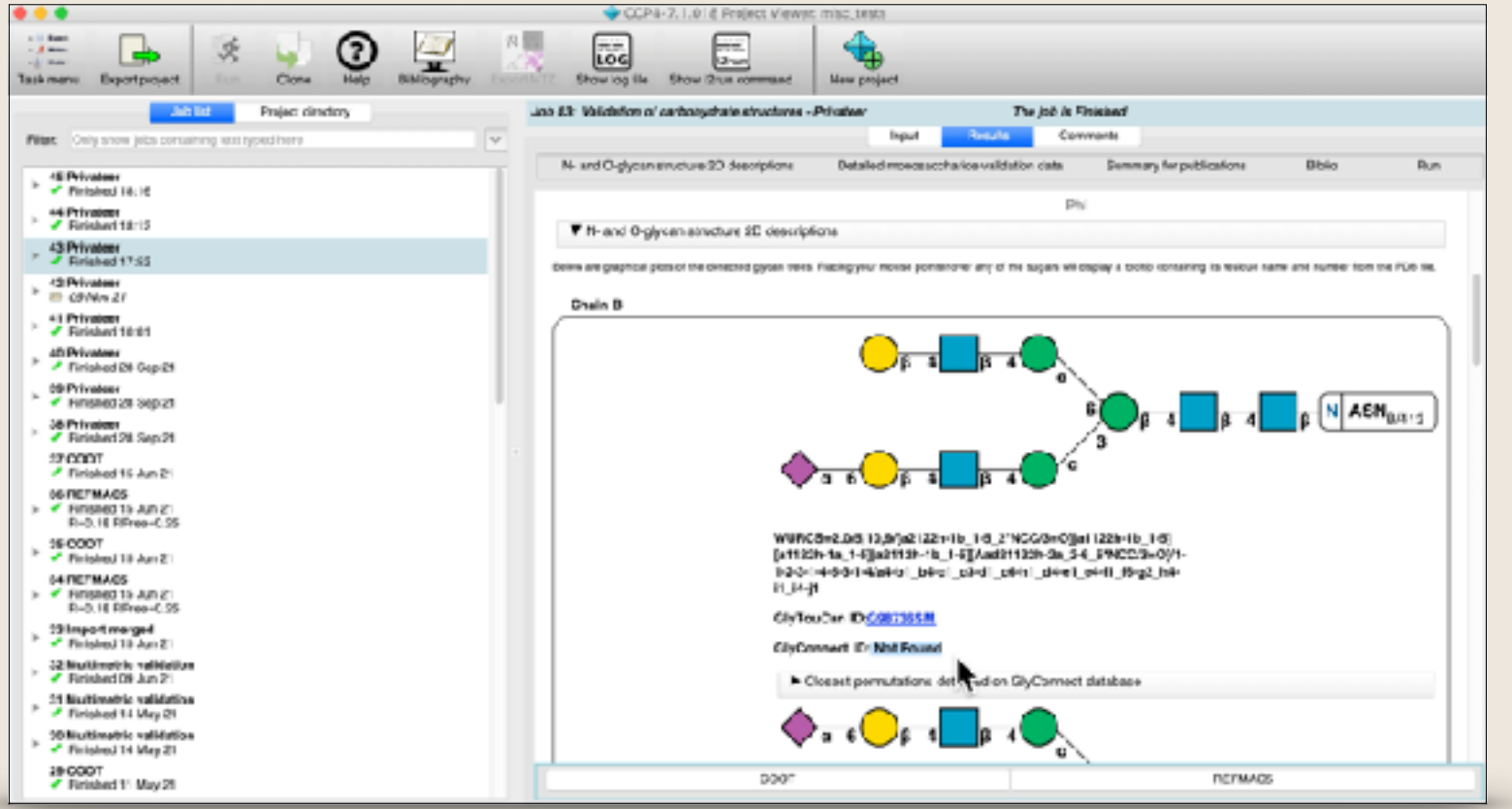
Running Privateer MKIV (CCP4i2)



Running Privateer MKIV (CCP4i2)

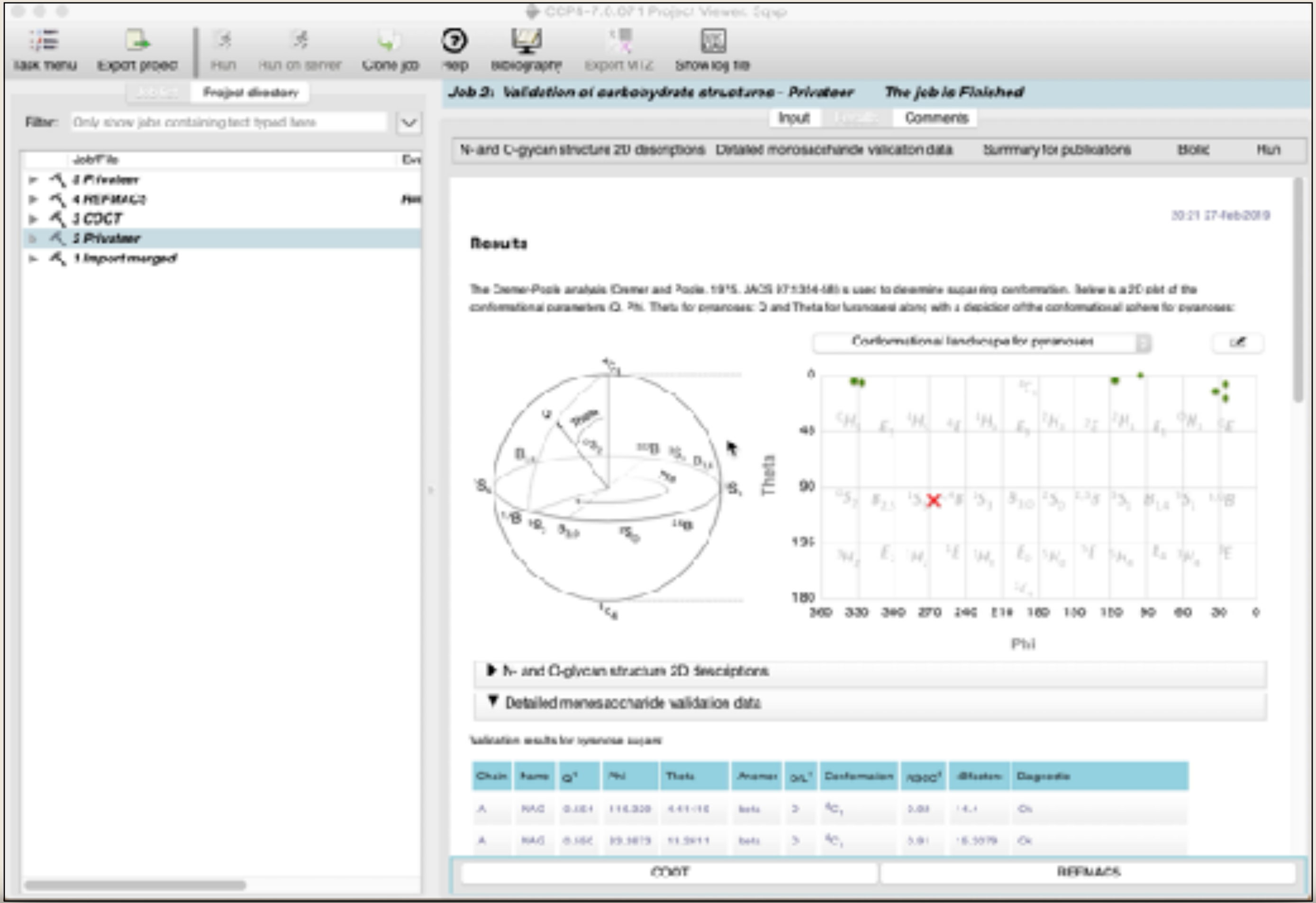


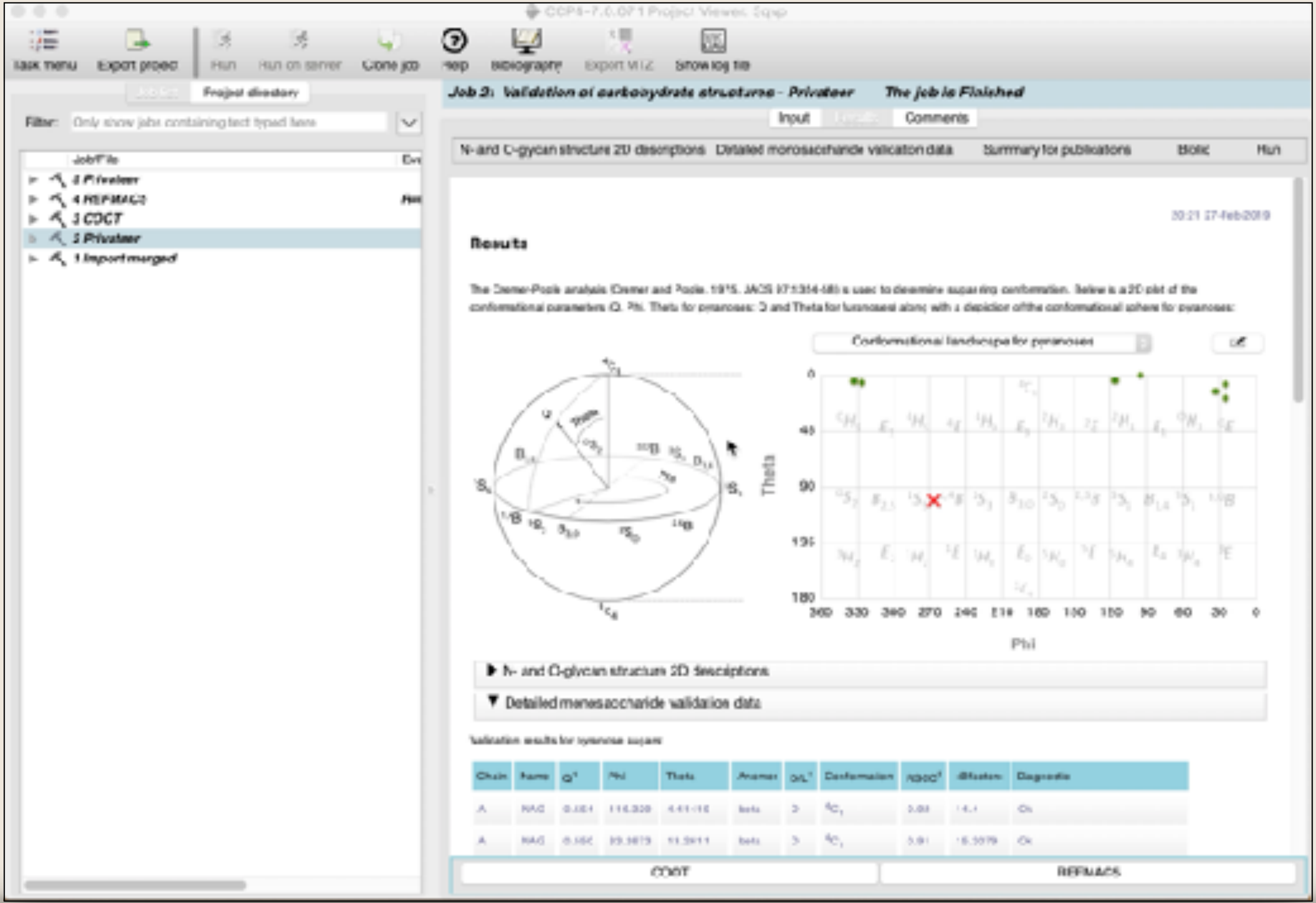
Running Privateer MKIV (CCP4i2)

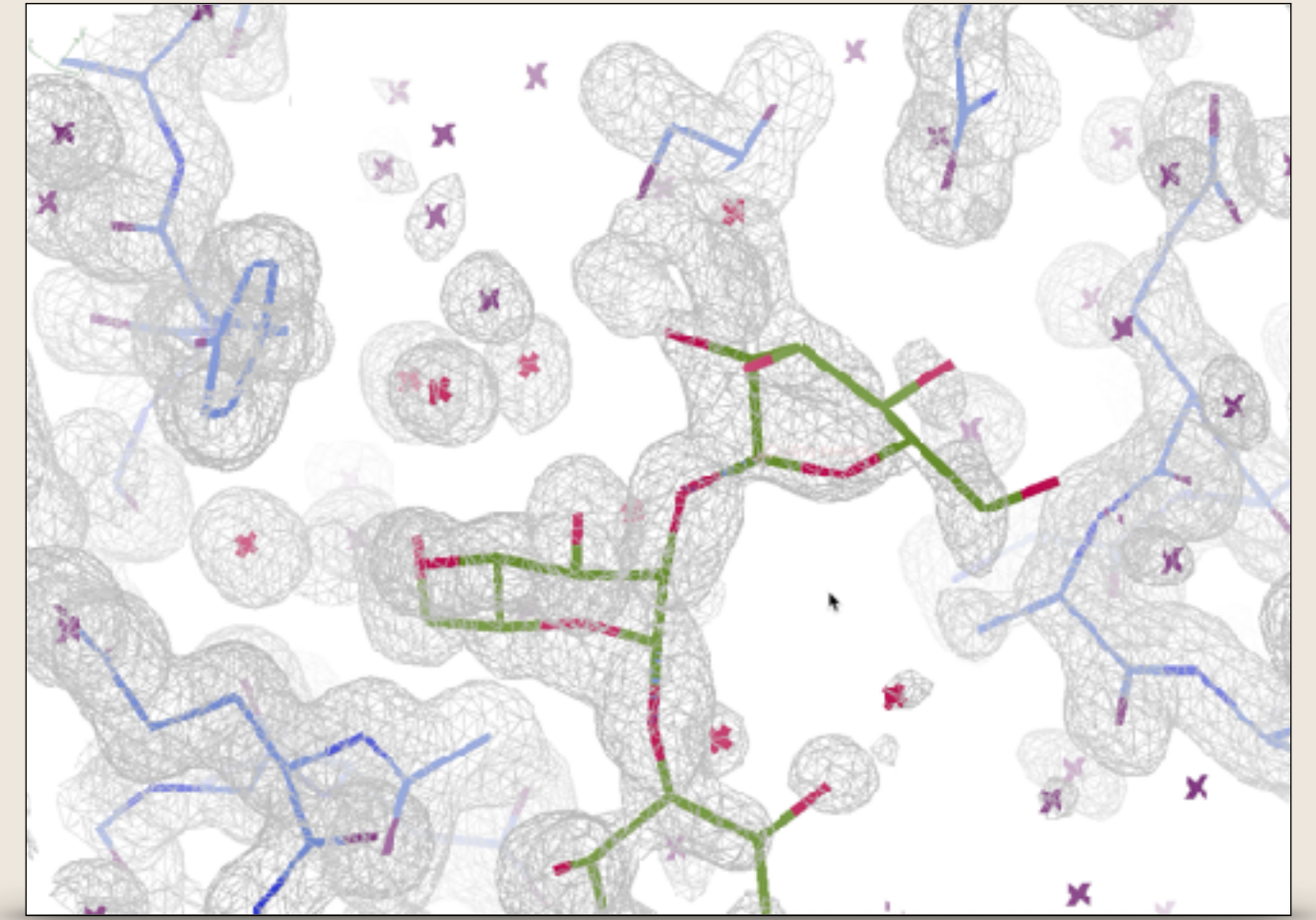


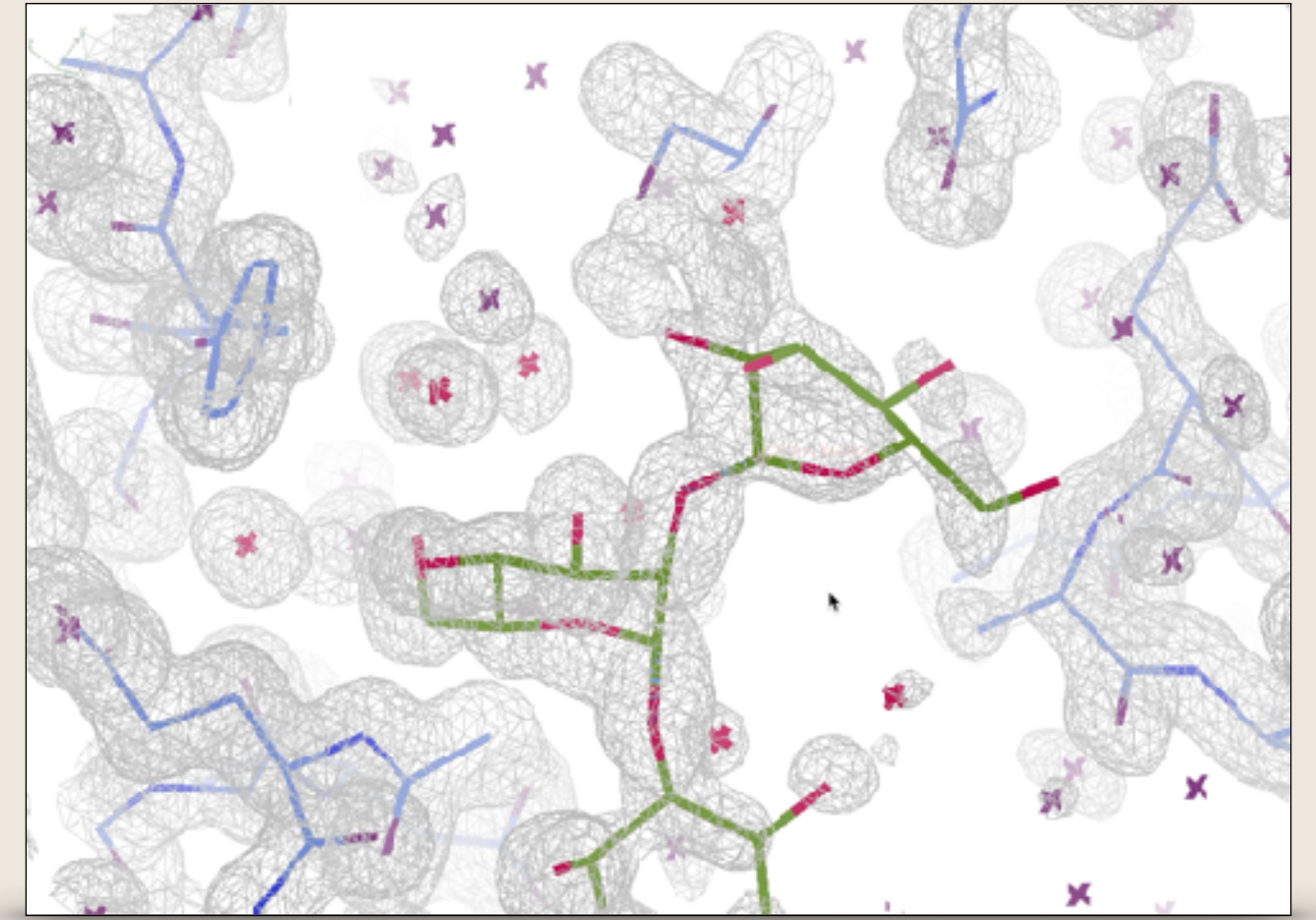
Running Privateer MKIV (CCP-EM)

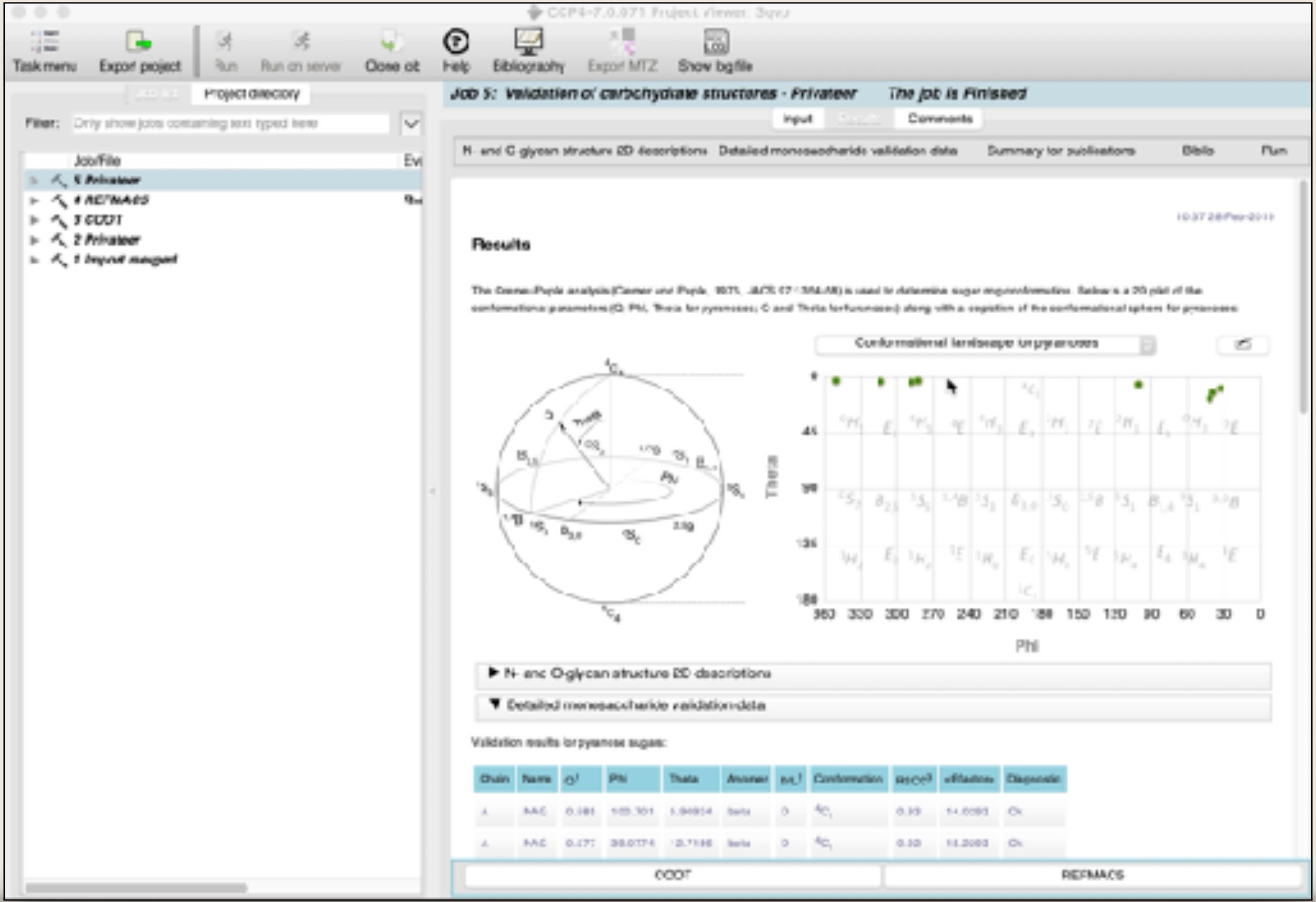
Running Privateer MKIV (CCP-EM)

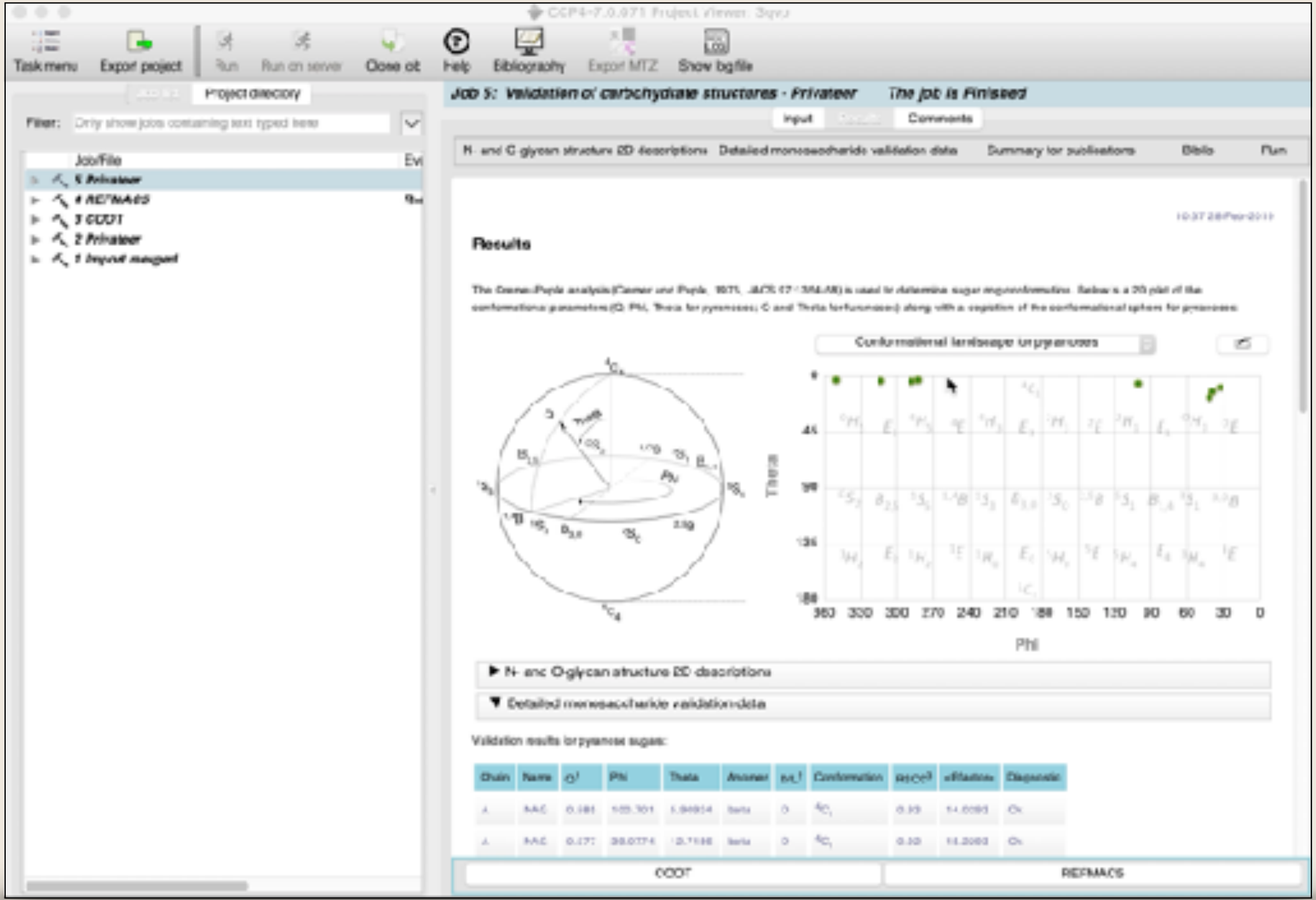

Task menu
Export project
Clone job
Help
Bibliography
Show log file
Project directory
Filter:
Only show jobs containing text typed here
4 REFMAC5
3 COOT
2 Privateer
1 Import merged
Job 2: Validation of carbohydrate structures - Privateer
The job is Finished
Input
Comments
Results
Conformational landscape for pyranoses
Theta
Phi
N- and O-glycan structure 2D descriptions
Detailed monosaccharide validation data
COOT
REFMAC5


Task menu
Export project
Run
Run on server
Clone job
Help
Bibliography
Export MTZ
Show log file
Project directory
Filter:
Only show jobs containing text typed here
5 Privateer
4 REFMAC5
3 COOT
2 Privateer
1 Import merged
Job 2: Validation of carbohydrate structures - Privateer
The job is Finished
Input
Comments
Results
Conformational landscape for pyranoses
Theta
Phi
N- and O-glycan structure 2D descriptions
Detailed monosaccharide validation data
COOT
REFMAC5

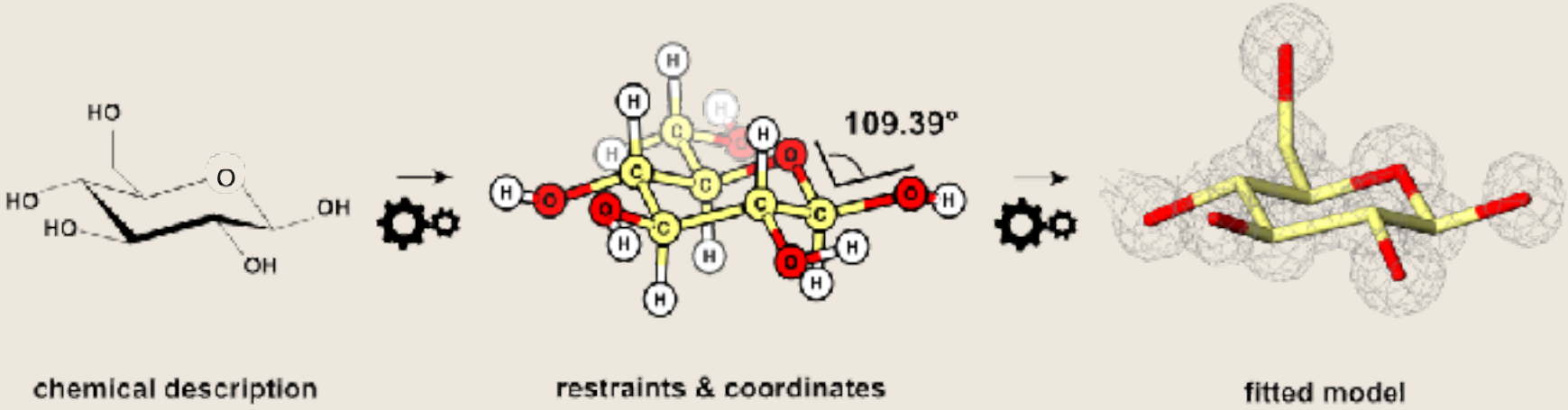




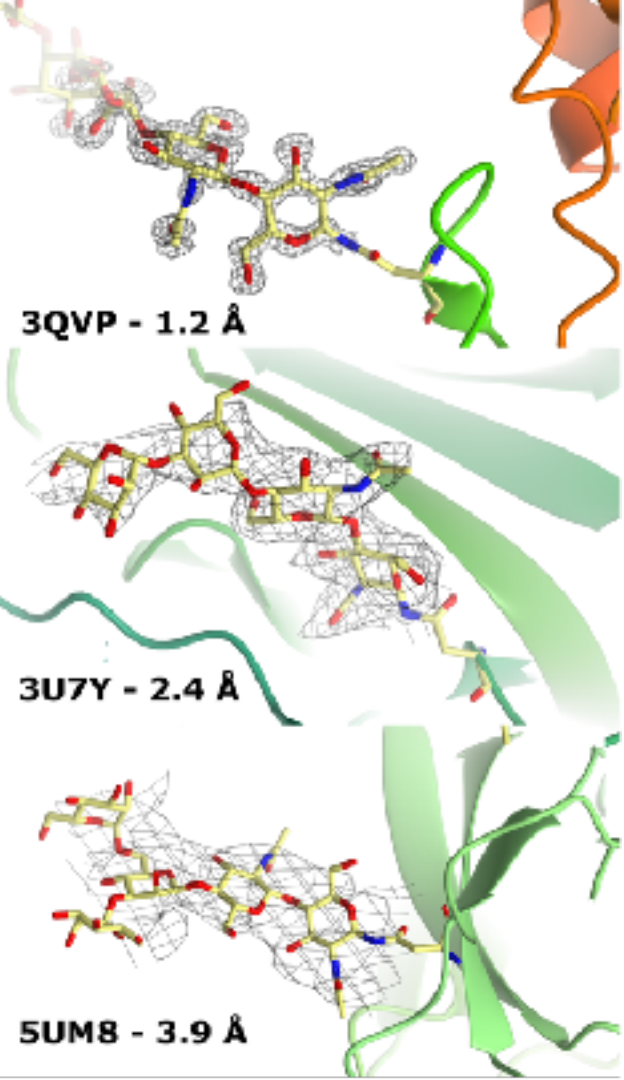

Task menu
Export project
Run
Run on server
Help
Bibliography
Export MTZ
Show bgfile
Project directory
Filter:
Job/File
2 Privateer
Input
Comments
Summary for publications
Biblio
Results
Theta
Phi
N- and O-glycan structure 2D descriptions
Detailed monosaccharide validation data
Validation results for pyranose sugars:
COOT
REFMACS

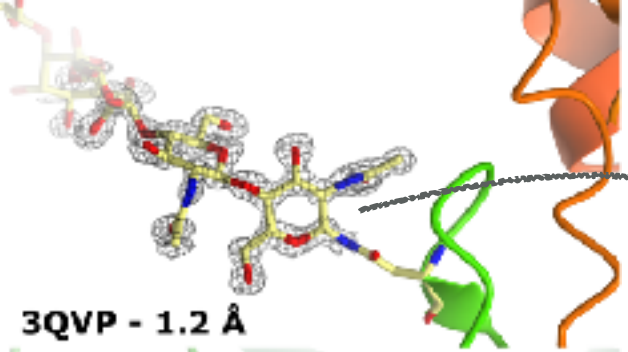

Task menu
Export project
Run
Run on server
Help
Bibliography
Export MTZ
Project directory
Filter:
Job/File
Job 5: Validation of carbohydrate structures - Privateer
The job is Finished
Input
Comments
Detailed monosaccharide validation data
Summary for publications
Biblio
Results
Conformational landscape for pyranoses
Theta
Phi
N- and O-glycan structure 2D descriptions
Detailed monosaccharide validation data
Validation results for pyranose sugars:
COOT

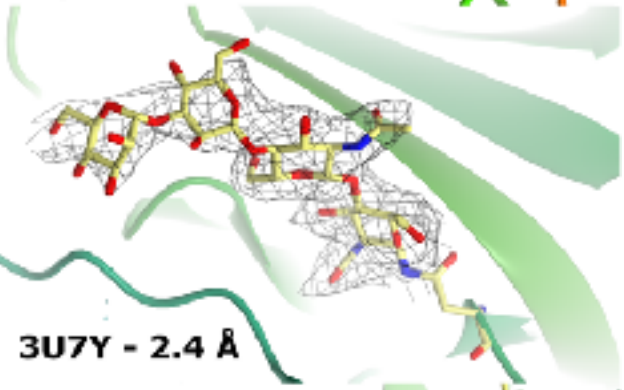
New carbohydrate dictionaries

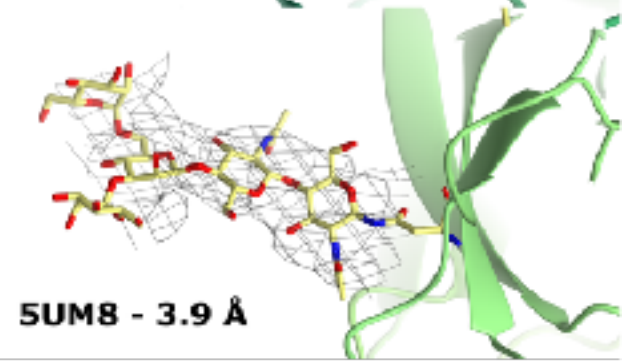


New carbohydrate dictionaries









New carbohydrate dictionaries

All this information to restrain one GlcNAc

Compound definition, with **3-letter code**, **full name**, **type** and number of atoms, **all** and **non-hydrogen**

```
NAG      NAG 'N-ACETYL-D-GLUCOSAMINE                ' pyranose              30  15 .
```

Atom by atom definition of the conformer, with **name**, symbol, charge and **cartesian coordinates**

```
NAG      O1      O     OH1        0.000      1.203     0.420     0.648
```

A bond between a **pair of atoms**, **cardinality**, **expected distance** and **standard deviation**

```
NAG      O1      C1         single       1.432     0.020
```

An angle between **three atoms**, with **expected value** and **standard deviation**

```
NAG      O1      C1      C2       109.470    3.000
```

A **torsion angle** between **four atoms**, with **expected value**, **standard deviation** and **periodicity**

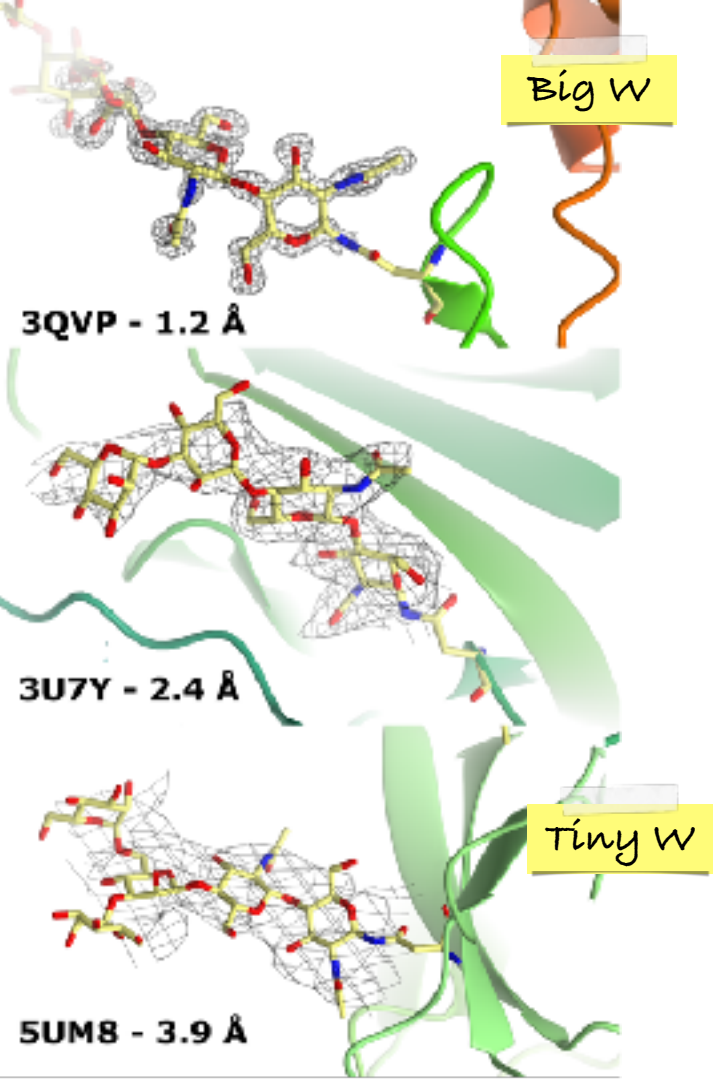
```
NAG      var_7   C5      C4      C3      C2          60.000   20.000   3
```

A **chiral volume**, defined by **four atoms** in a tetrahedral arrangement, and a **sign defining orientation**

```
NAG      chir_01 C1      C2      O1      O5          positiv
```

A **series of atoms** taking part of a **planar arrangement**, plus **standard deviation**

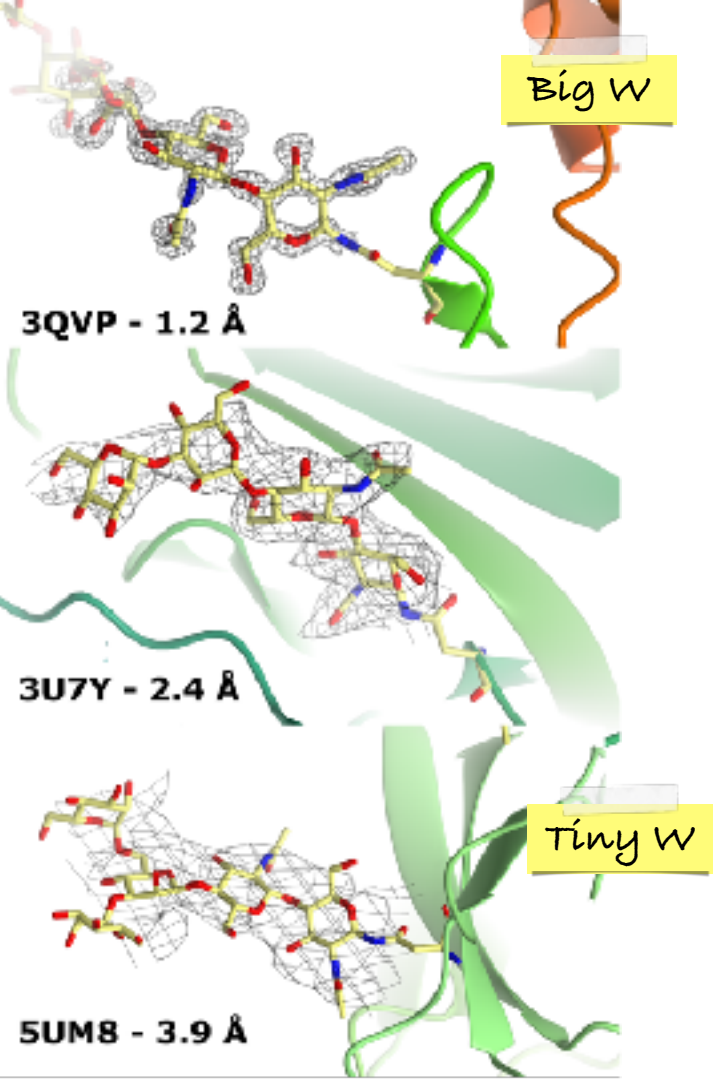
```
NAG      plan-2      N2          0.020
NAG      plan-2      C2          0.020
NAG      plan-2      C7          0.020
NAG      plan-2      HN2         0.020
```



What about ring conformation?

With weak/incomplete density, a ring's lowest-energy conformation may have to be explicitly restrained:

$$L(p) = wL_X(p) + L_G(p)$$



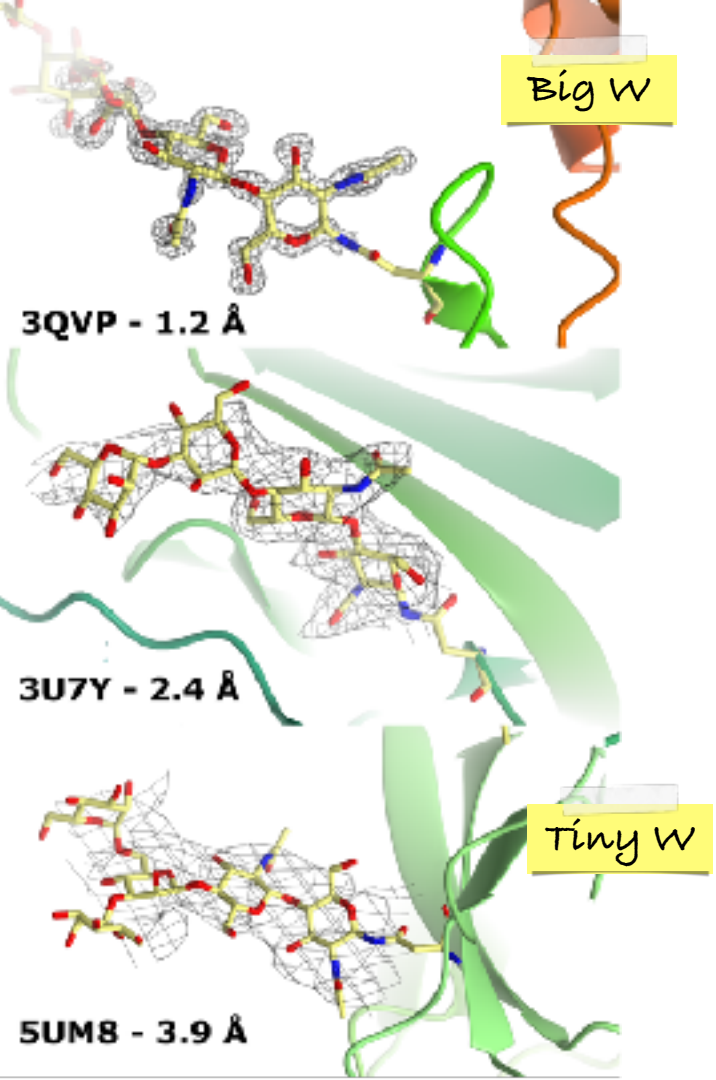
What about ring conformation?

With weak/incomplete density, a ring's lowest-energy conformation may have to be explicitly restrained:

$$L(p) = wL_X(p) + L_G(p)$$

Yes!

bond lengths



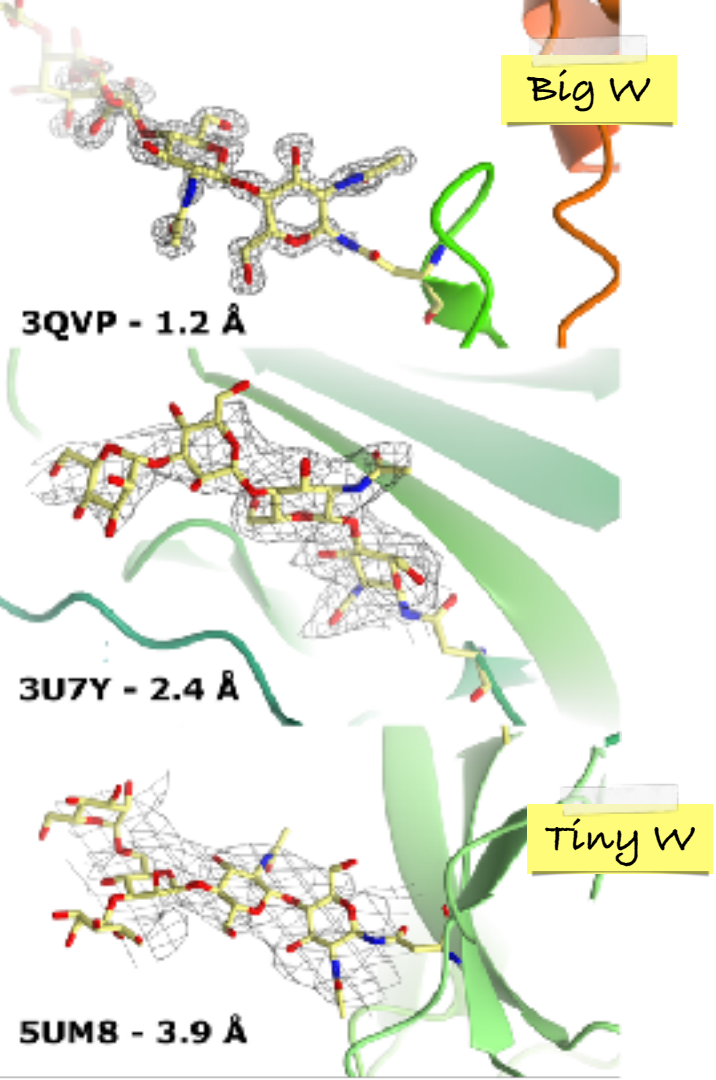
What about ring conformation?

With weak/incomplete density, a ring's lowest-energy conformation may have to be explicitly restrained:

$$L(p) = wL_X(p) + L_G(p)$$

Yes! bond lengths

Yes! bond angles



What about ring conformation?

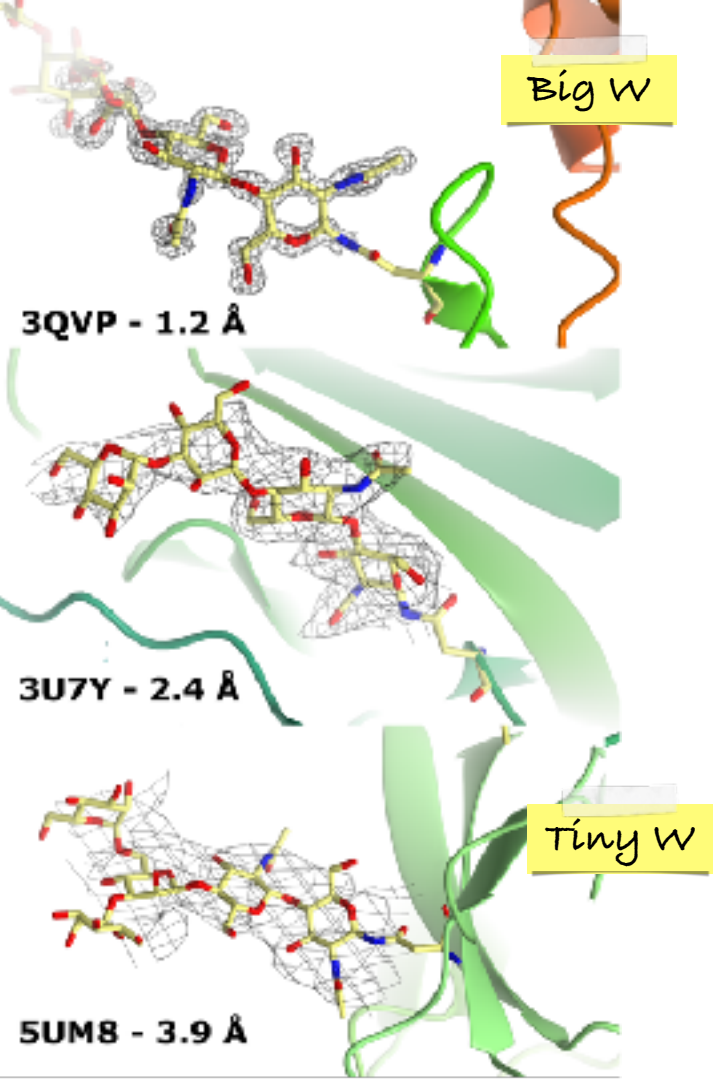
With weak/incomplete density, a ring's lowest-energy conformation may have to be explicitly restrained:

$$L(p) = wLX(p) + LG(p)$$

Yes! bond lengths

Yes! bond angles

Yes! Planar groups

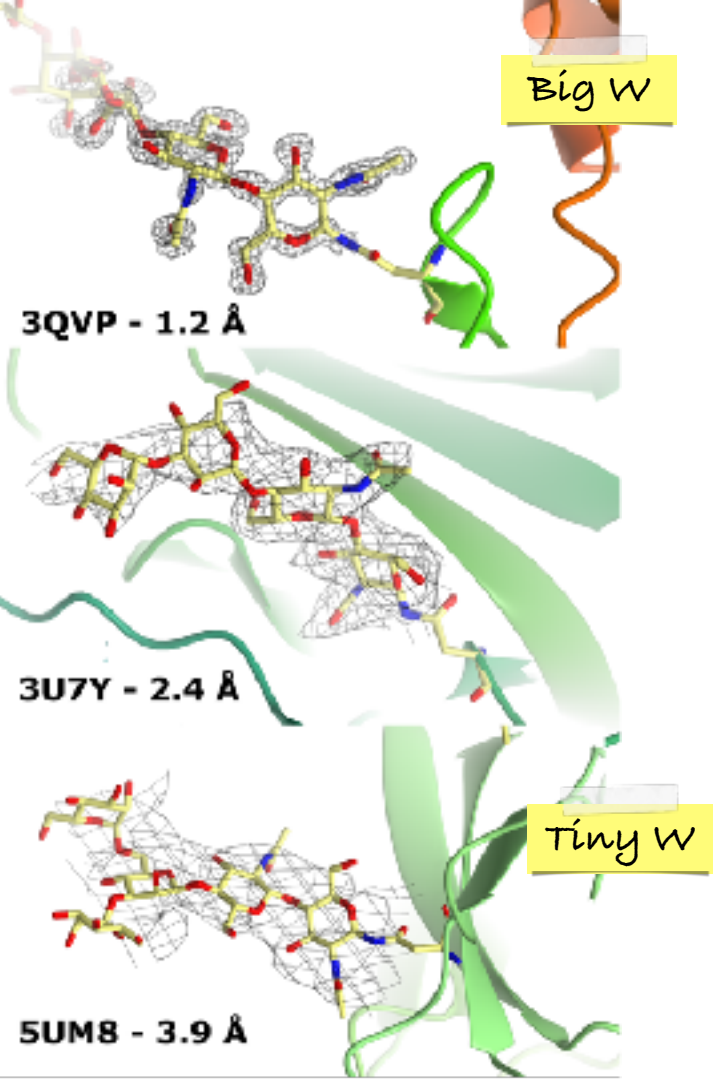


What about ring conformation?

With weak/incomplete density, a ring's lowest-energy conformation may have to be explicitly restrained:

$$L(p) = wL_X(p) + L_G(p)$$

- Yes! bond lengths
- Yes! bond angles
- Yes! Planar groups
- ? harmonic torsions

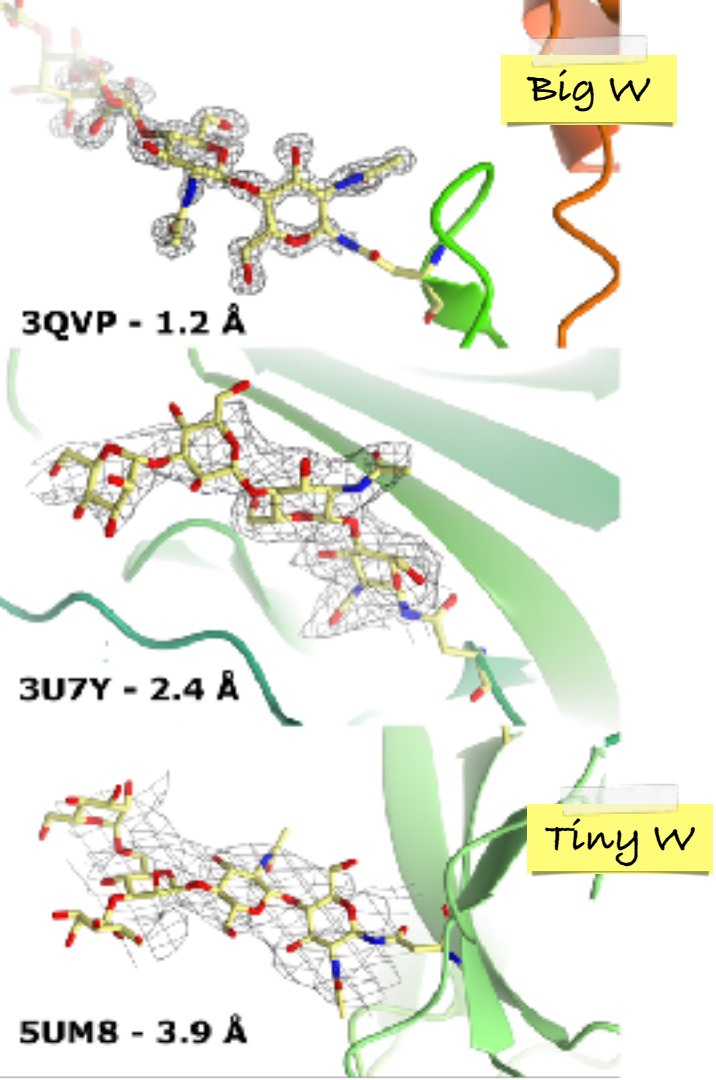


What about ring conformation?

With weak/incomplete density, a ring's lowest-energy conformation may have to be explicitly restrained:

$$L(p) = wL_X(p) + L_G(p)$$

- Yes! bond lengths
- Yes! bond angles
- Yes! Planar groups
- ? harmonic torsions
- ? NCS, etc



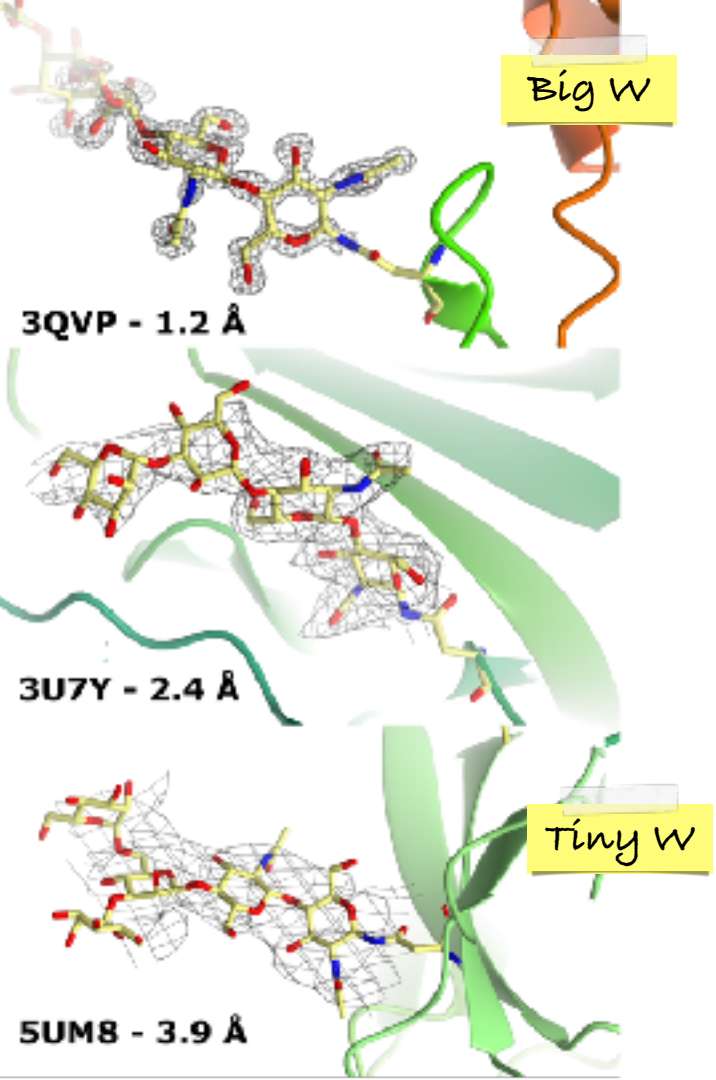
What about ring conformation?

With weak/incomplete density, a ring's lowest-energy conformation may have to be explicitly restrained:

$$L(p) = wL_X(p) + L_G(p)$$

- Yes! bond lengths
- Yes! bond angles
- Yes! Planar groups
- ? harmonic torsions
- ? NCS, etc

Staggered Eclipsed



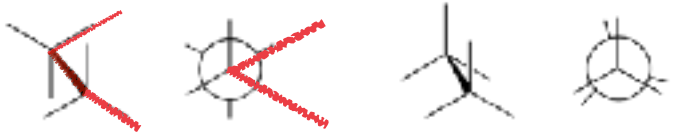
What about ring conformation?

With weak/incomplete density, a ring's lowest-energy conformation may have to be explicitly restrained:

$$L(p) = wL_X(p) + L_G(p)$$

- Yes! bond lengths
- Yes! bond angles
- Yes! Planar groups
- ? harmonic torsions
- ? NCS, etc

~60°/-60°



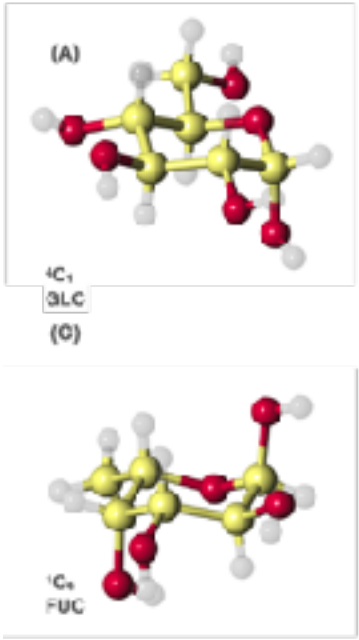
Staggered

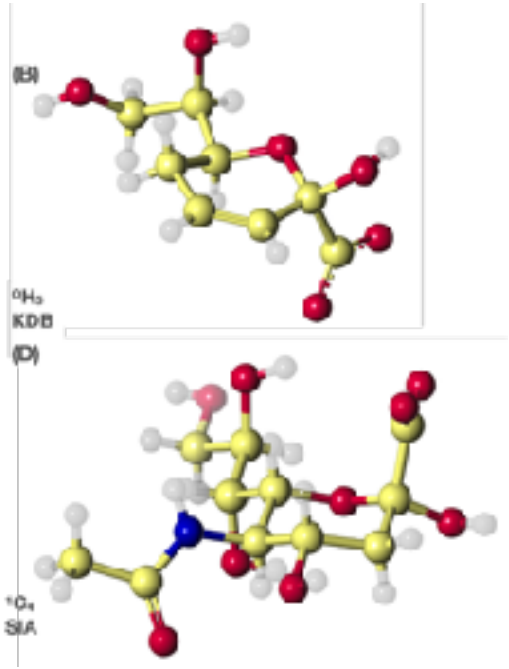
Eclipsed

Our new restraint dictionaries *understand* ring conformation

Patched _chem_comp_tor section of a restraint dictionary separating **ring torsion angles** from the **rest**, as specified by **four atoms**, **target value**, **uncertainty** and **periodicity**

NAG	ring_1	C5	O5	C1	C2	-59.675385	3.0	1
NAG	ring_2	O5	C1	C2	C3	53.650513	3.0	1
NAG	ring_3	C1	C2	C3	C4	-52.014420	3.0	1
NAG	ring_4	C2	C3	C4	C5	54.096725	3.0	1
NAG	ring_5	C3	C4	C5	O5	-56.921230	3.0	1
NAG	ring_6	C4	C5	O5	C1	61.200516	3.0	1
NAG	tors_1	C8	C7	N2	C2	-175.114227	10.0	2
NAG	tors_2	N2	C7	C8	H81	-13.703261	10.0	6
NAG	tors_3	C5	C6	O6	HO6	-177.520996	10.0	3
NAG	tors_4	C4	C5	C6	O6	61.135471	10.0	3
NAG	tors_5	C6	C5	O5	C1	-175.561295	10.0	3
NAG	tors_6	O4	C4	C5	C6	63.707928	10.0	3
NAG	tors_7	C3	C4	O4	HO4	-61.268230	10.0	3
NAG	tors_8	O3	C3	C4	O4	-63.830528	10.0	3
NAG	tors_9	C2	C3	O3	HO3	-169.485916	10.0	3
NAG	tors_10	C7	N2	C2	C1	124.894669	10.0	6
NAG	tors_11	N2	C2	C3	O3	61.918137	10.0	3
NAG	tors_12	C2	C1	O1	HO1	163.115189	10.0	3
NAG	tors_13	O1	C1	O5	C5	179.557251	10.0	3
NAG	tors_14	O1	C1	C2	N2	-62.214077	10.0	3

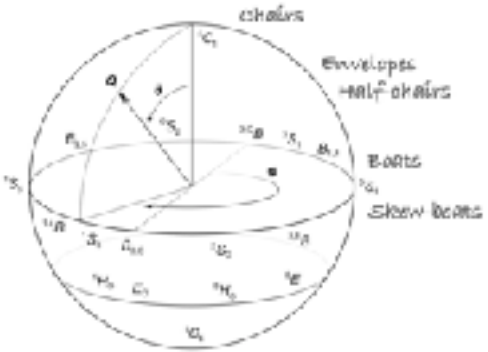




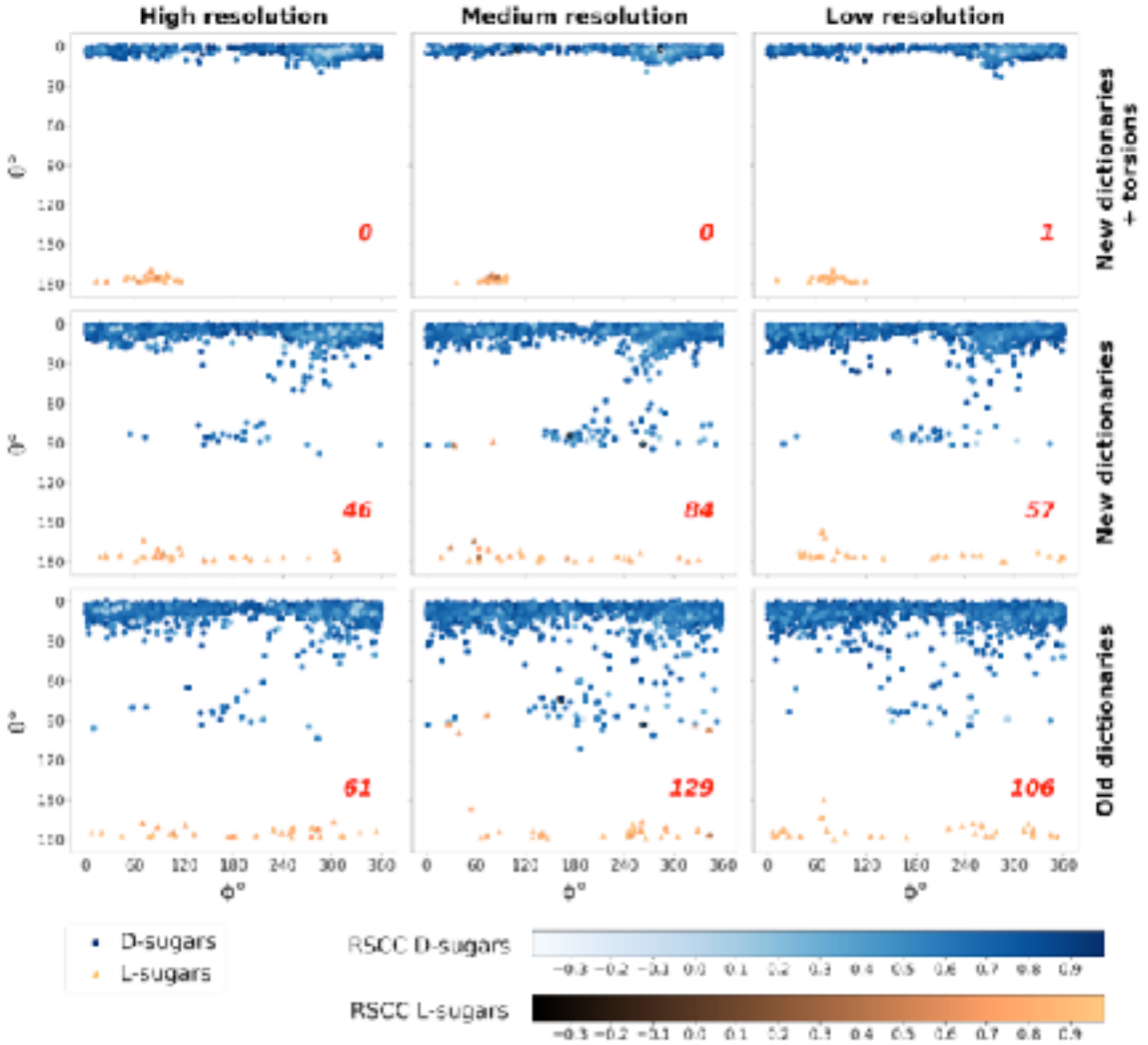
Atanasova, Joosten, Nicholls & Agirre, 2022, Acta Crystallographica D(78):455-465

Conformational restraints
N-glycosylation

No conformational distortions expected!

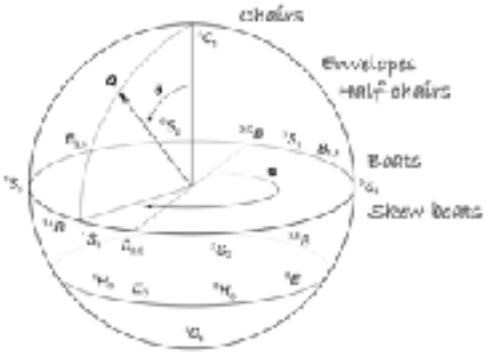


Atanasova, Joosten, Nicholls & Agirre, 2022, Acta Crystallographica D(78):455-465

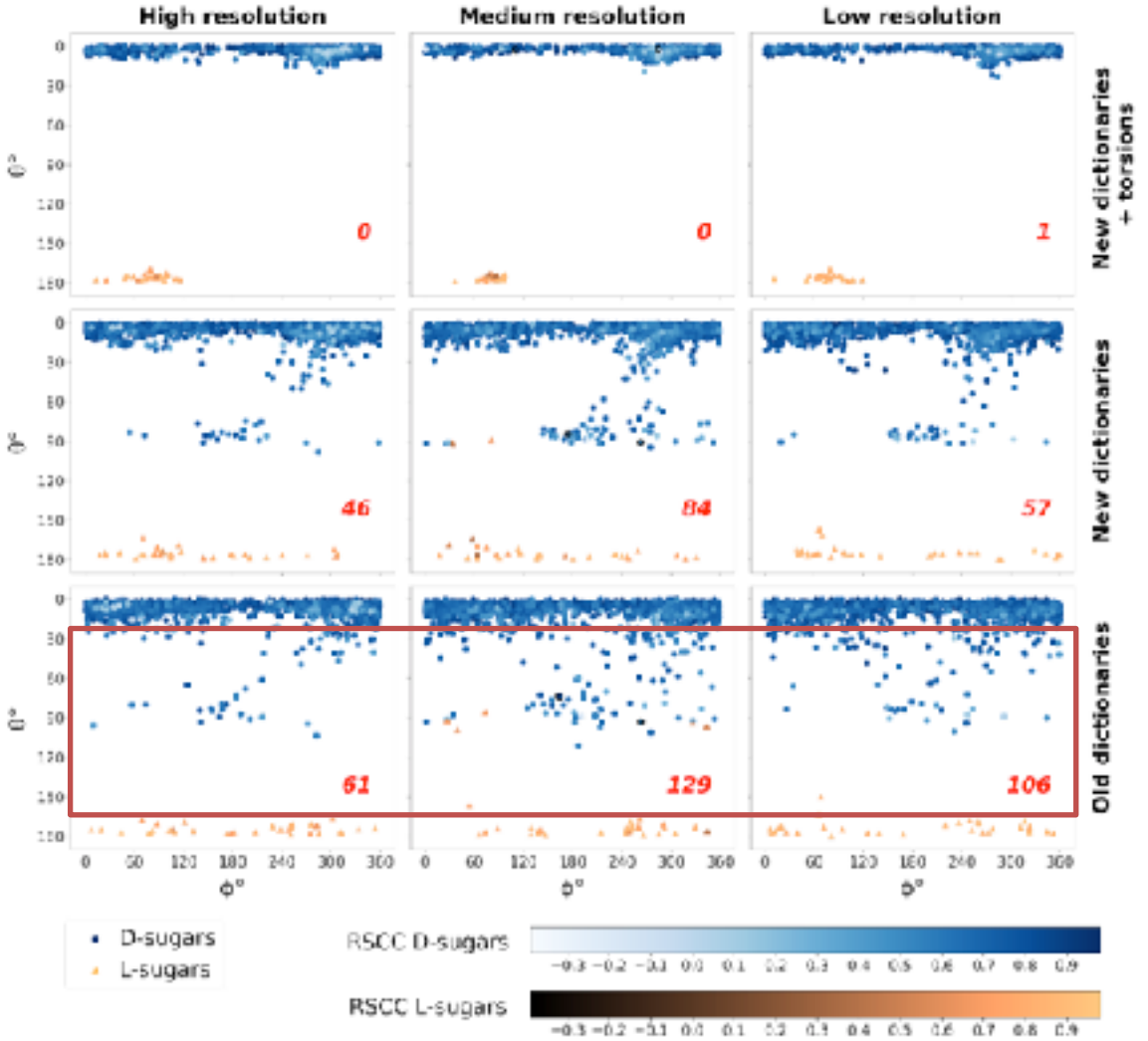


Conformational restraints
N-glycosylation

No conformational distortions expected!

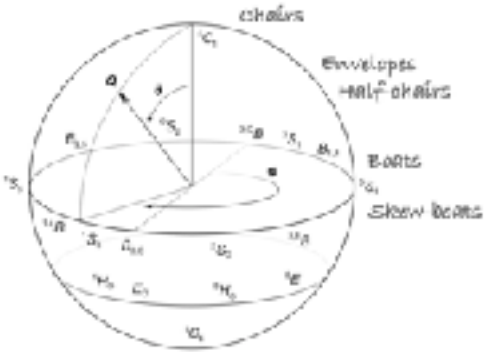


Atanasova, Joosten, Nicholls & Agirre, 2022, Acta Crystallographica D(78):455-465

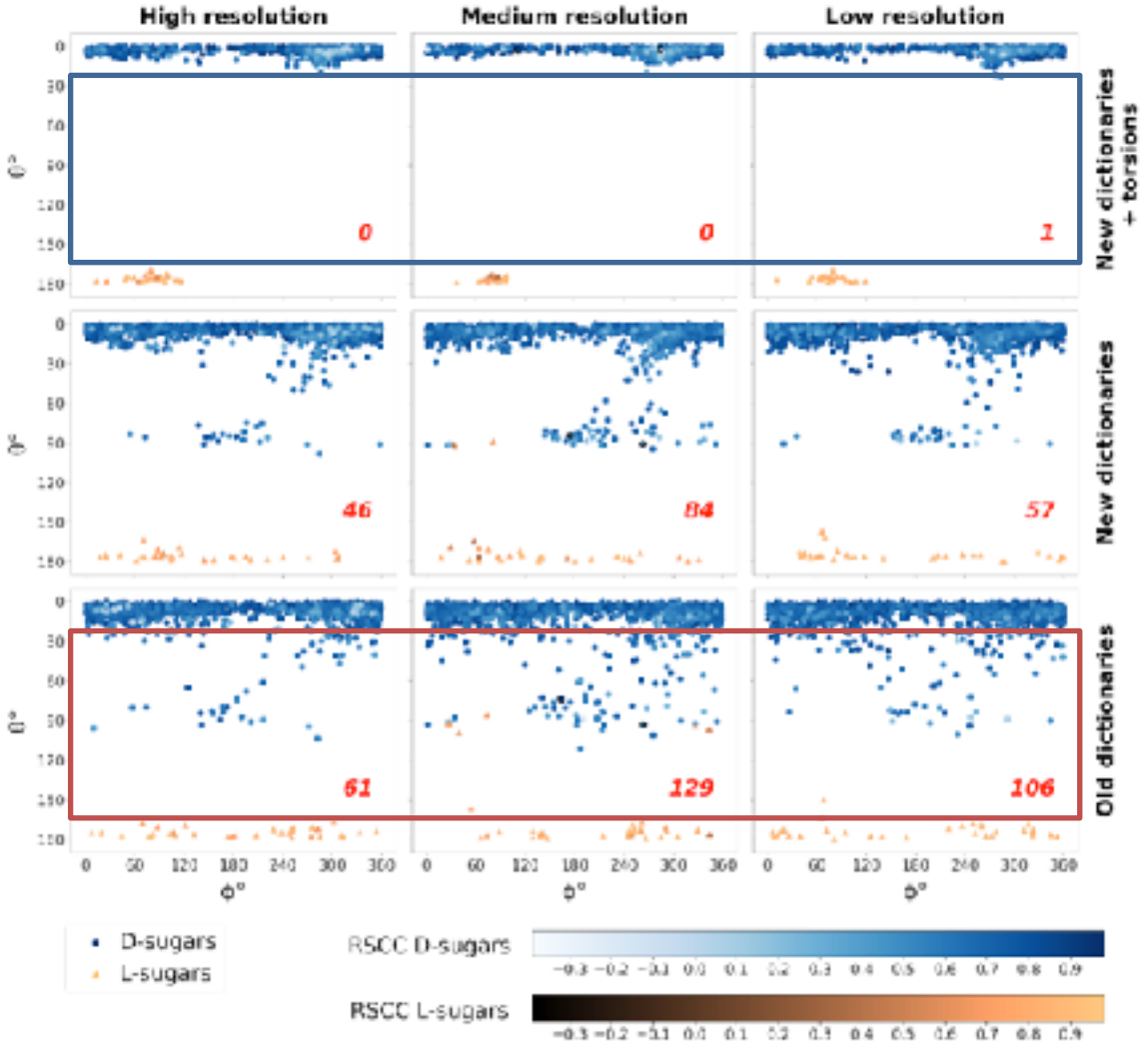


Conformational restraints
N-glycosylation

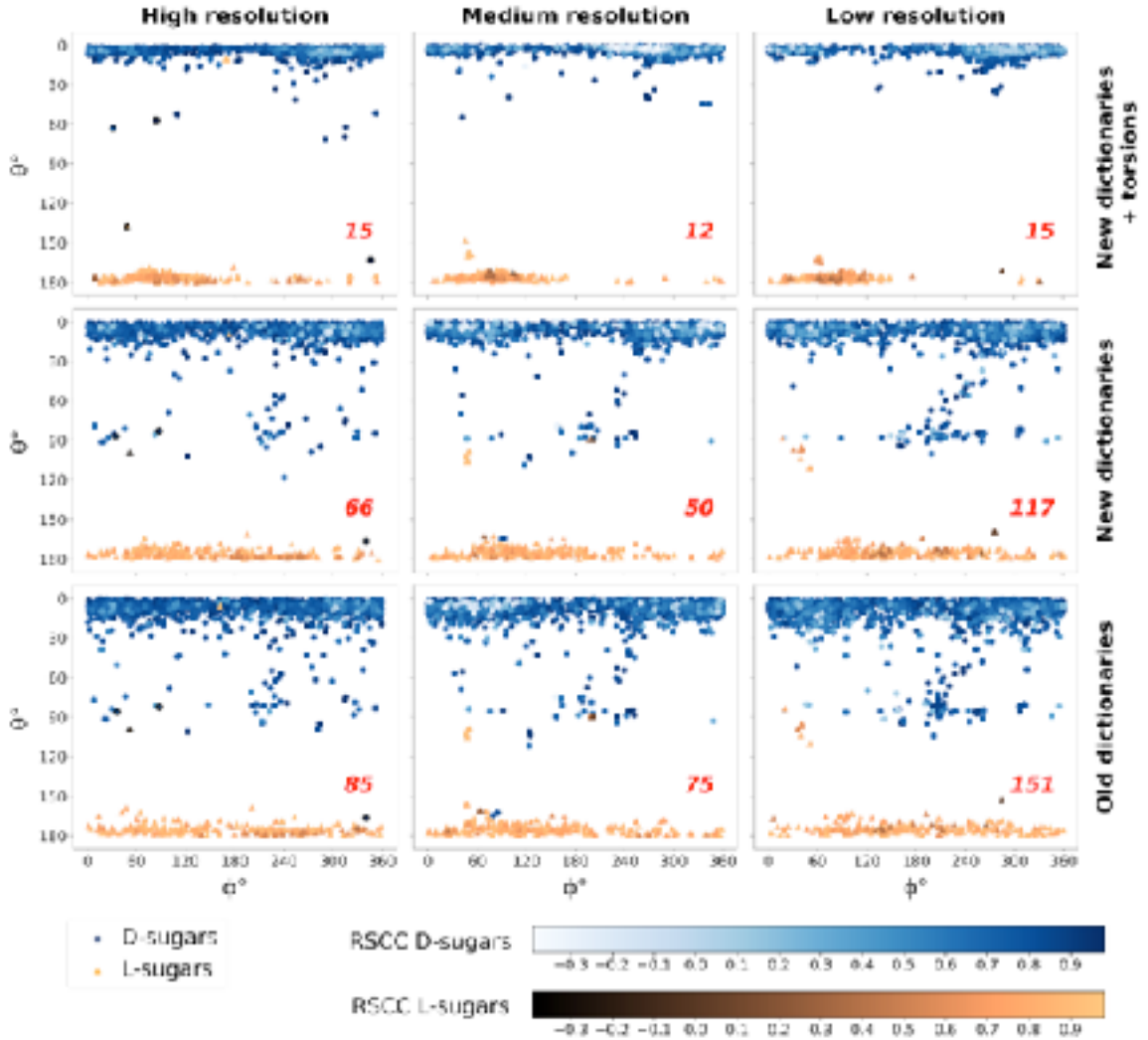
No conformational distortions expected!



Atanasova, Joosten, Nicholls & Agirre, 2022, Acta Crystallographica D(78):455-465

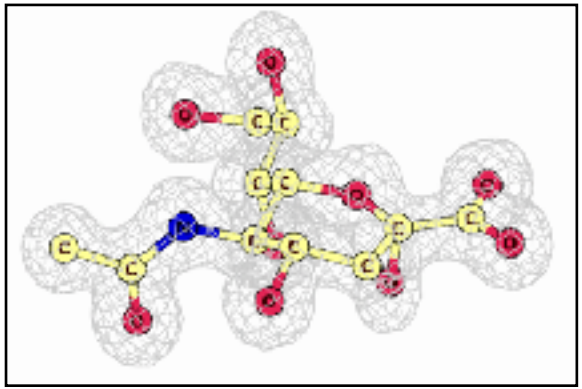


Conformational restraints
Ligands



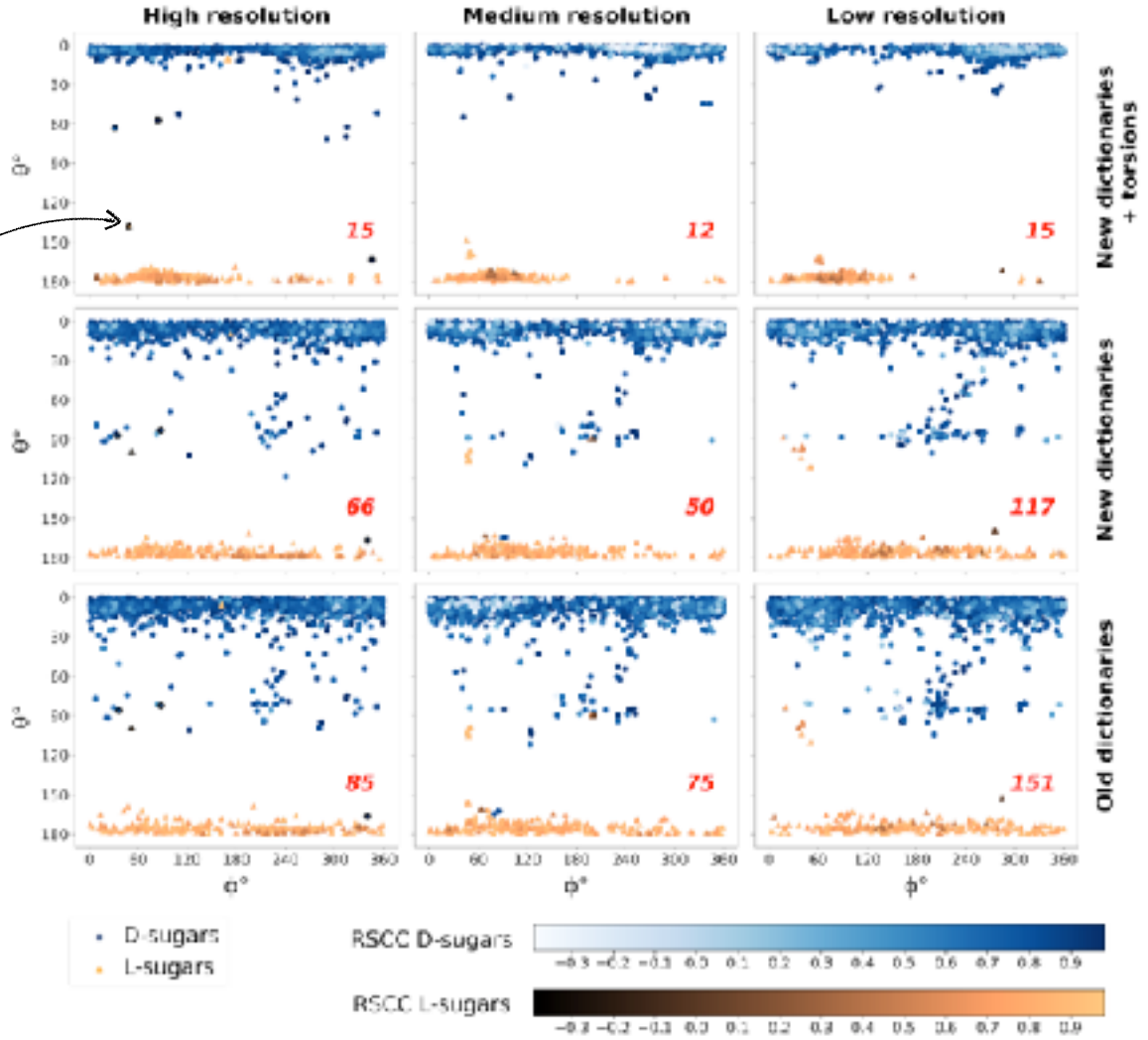
Atanasova, Joosten, Nicholls & Agirre, 2022,
Acta Crystallographica D(78):455-465

Conformational restraints
Ligands

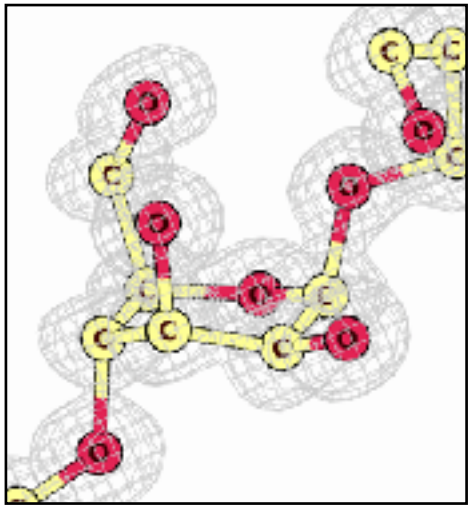


6HG0 (SIA-A-522) in ^{4}E
1.30 Å resolution, RSCC=0.94

Atanasova, Joosten, Nicholls & Agirre, 2022,
Acta Crystallographica D(78):455-465

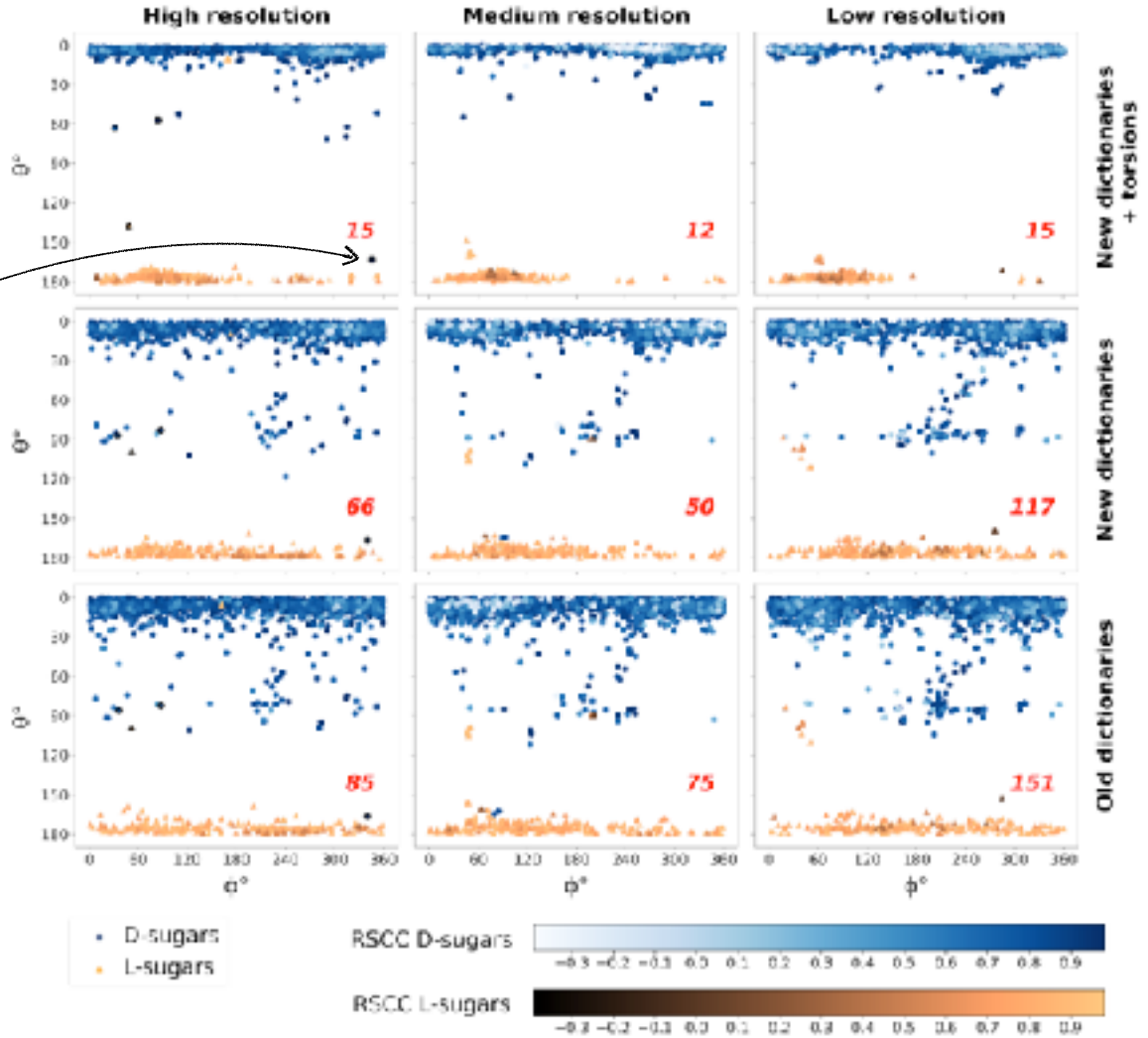


Conformational restraints
Ligands

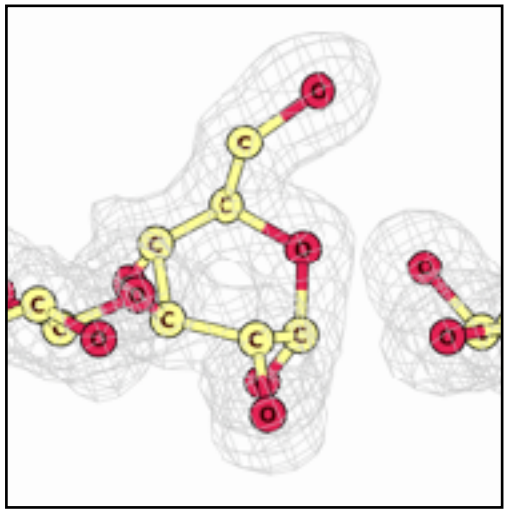


5JUG (BMA-B-3) in $^1C_4$
0.95 Å resolution, RSCC=0.95

Atanasova, Joosten, Nicholls & Agirre, 2022,
Acta Crystallographica D(78):455-465

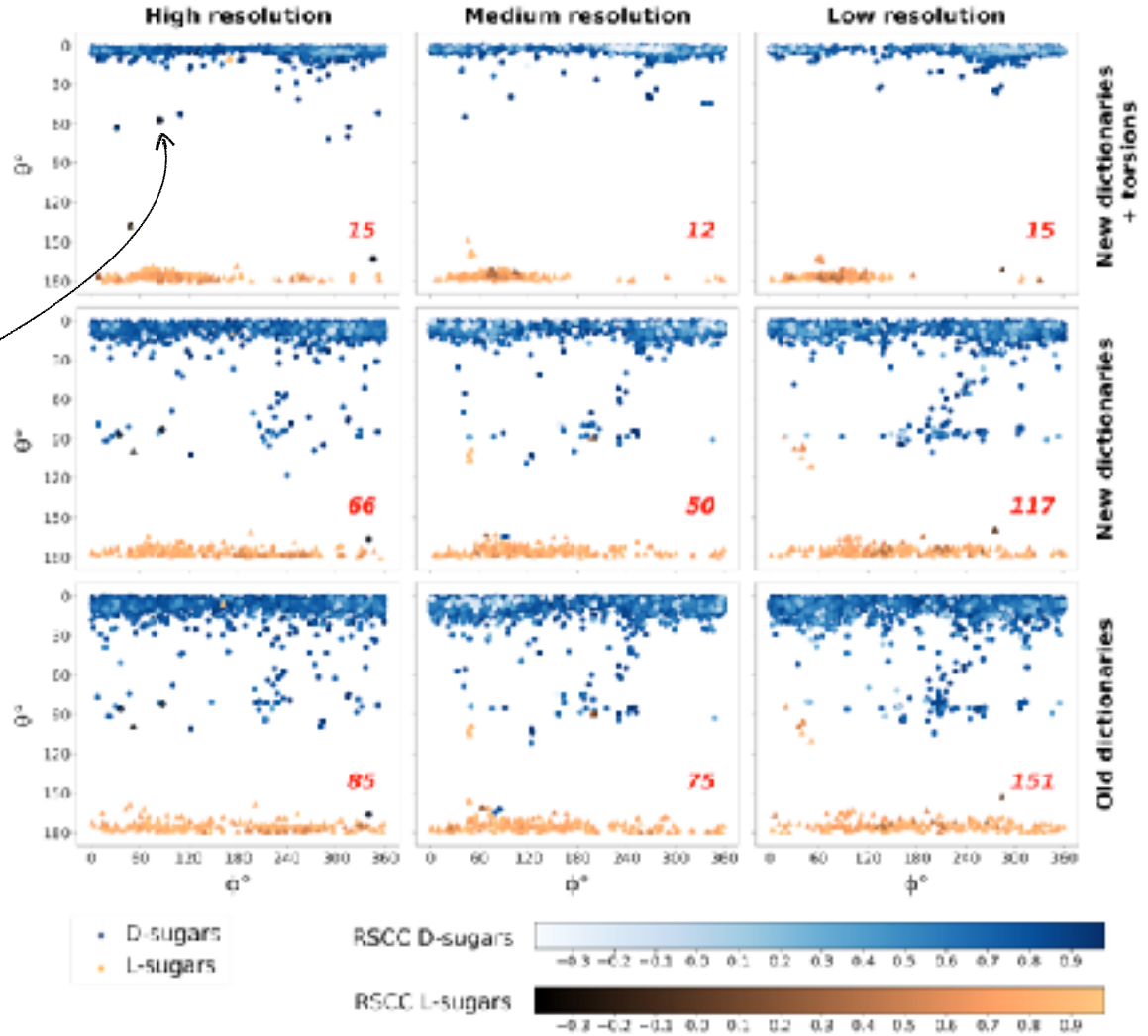


Conformational restraints
Ligands

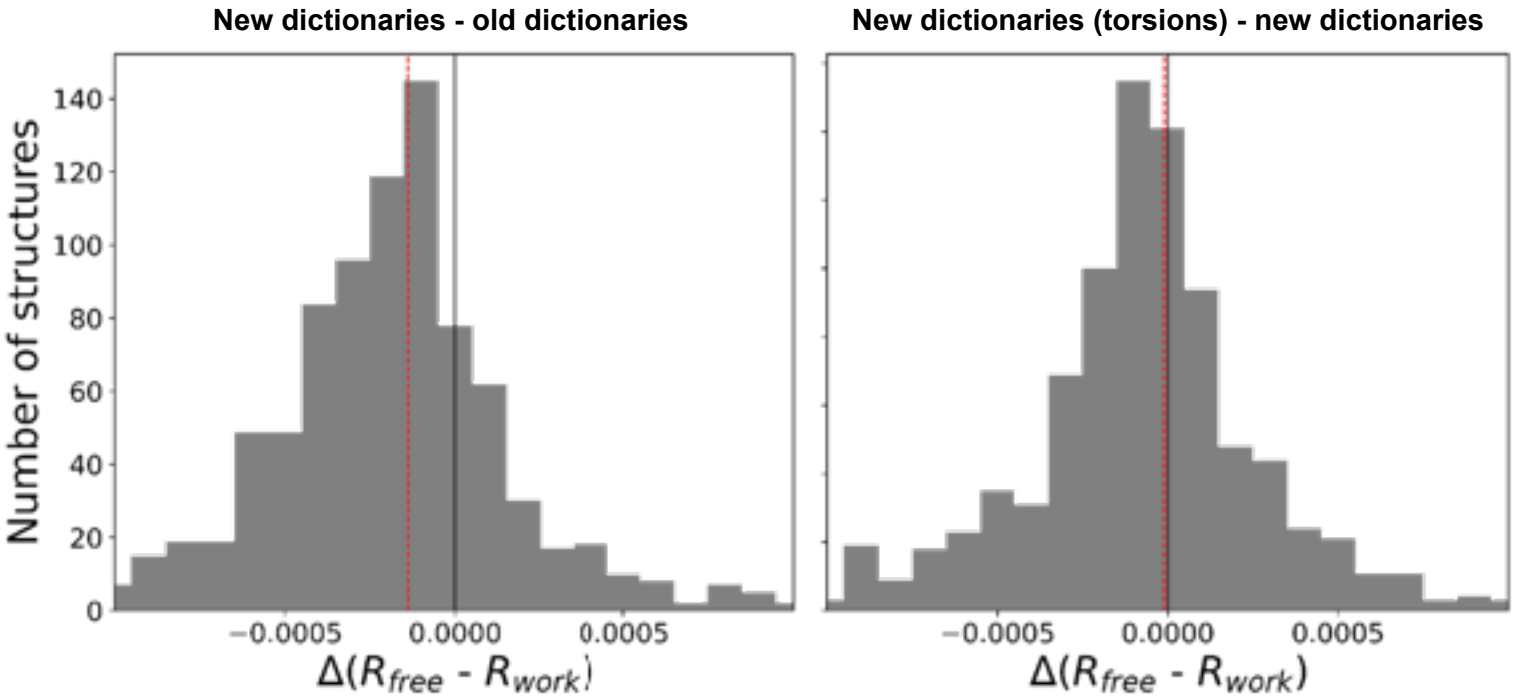


5UPM (GLC-C-1) in $^{2}H_{1}$
1.70 Å resolution, RSCC=0.96

Atanasova, Joosten, Nicholls & Agirre, 2022,
Acta Crystallographica D(78):455-465

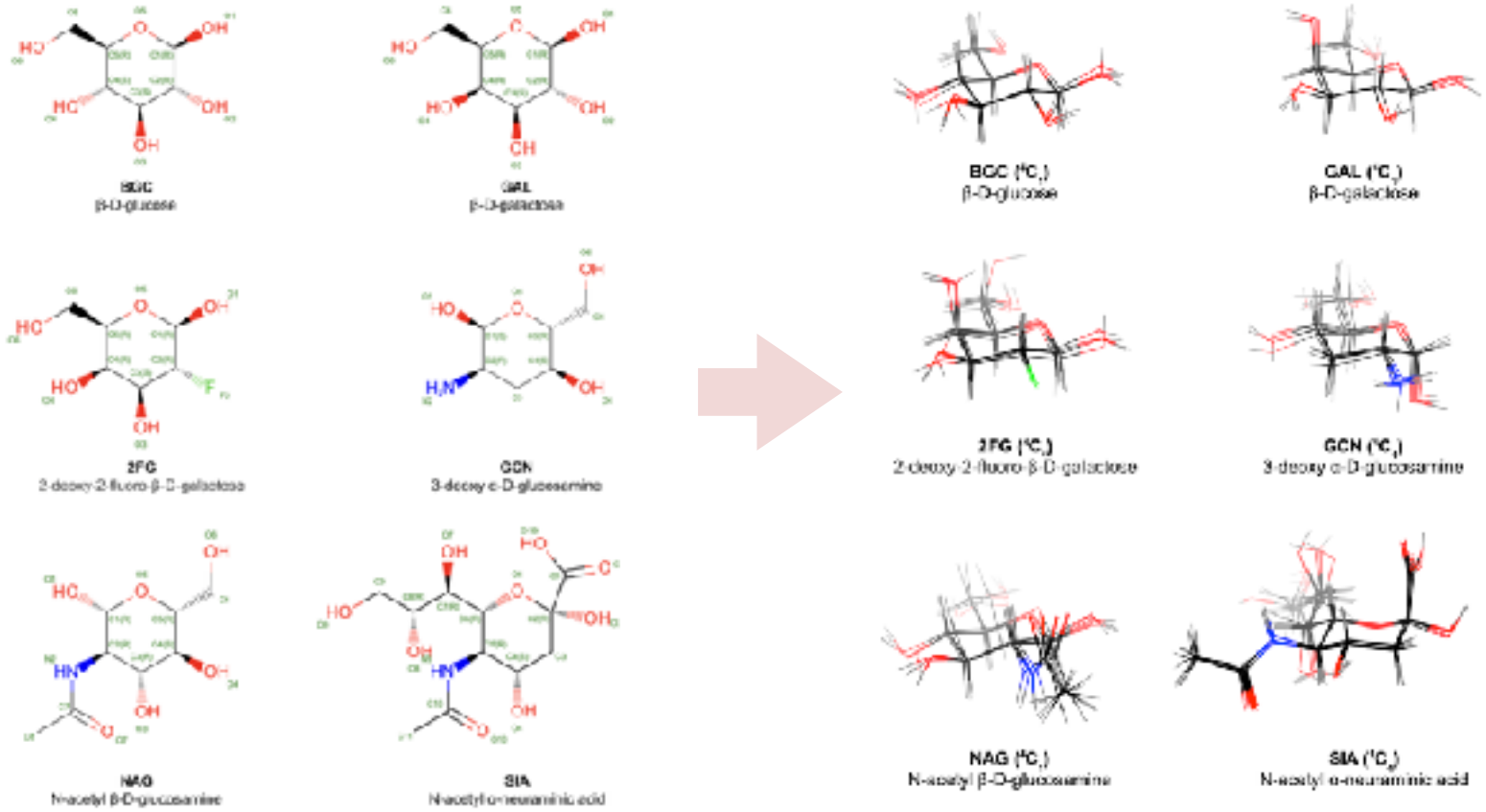


New dictionaries with conformational restraints
Reduction in R_{gap} = less overfitting



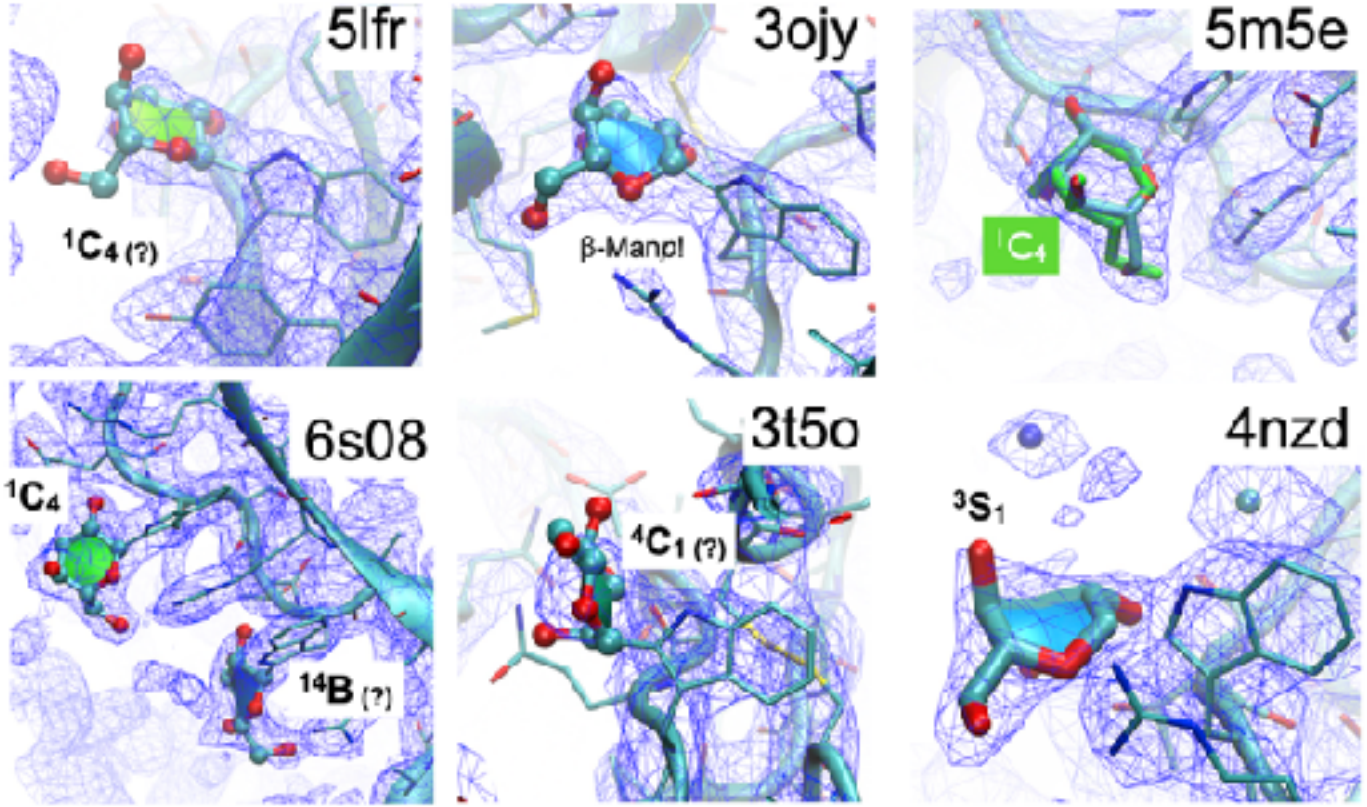
Atanasova, Joosten, Nicholls & Agirre, 2022, Acta Crystallographica D(78):455-465

Improving restraint consistency across software: getting in touch with developers



Joosten, Nicholls & Agirre, 2022, Current Medicinal Chemistry 29(4).

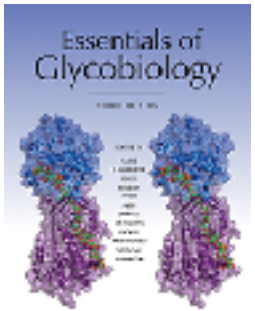
Restraints for C-mannosylation (MKV)

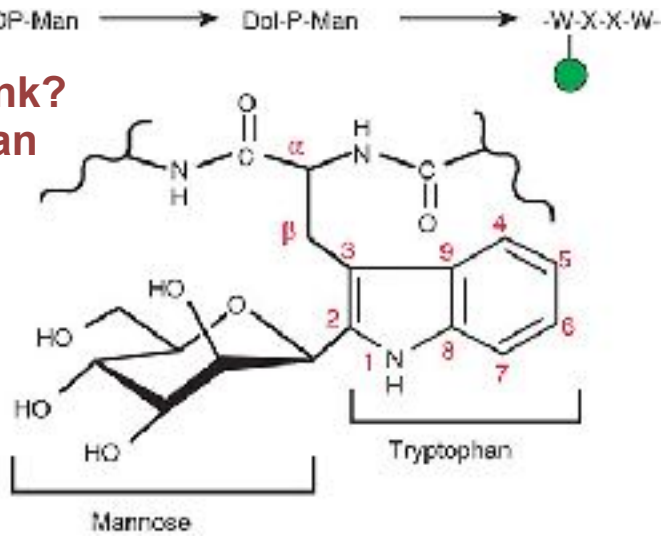


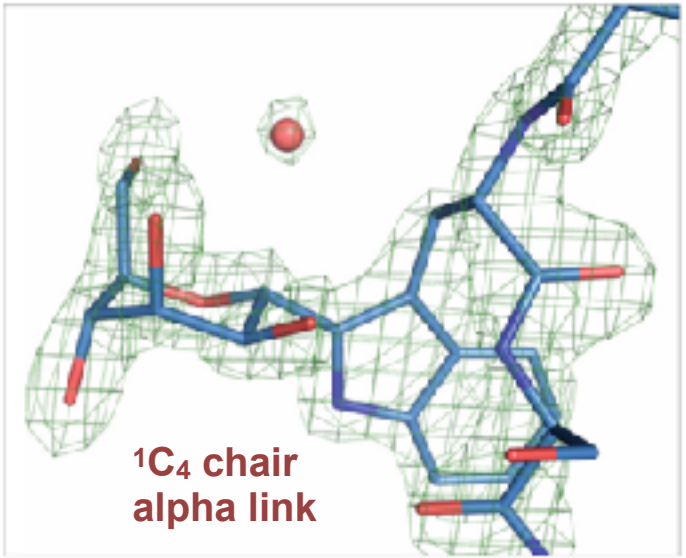
Frank, Beccati, Leeflang, Vliegenthart, 2020, iScience 23(8):101371.

Restraints for C-mannosylation (MKV)

$^4C_1$ chair, beta link? Enzyme makes an alpha link…







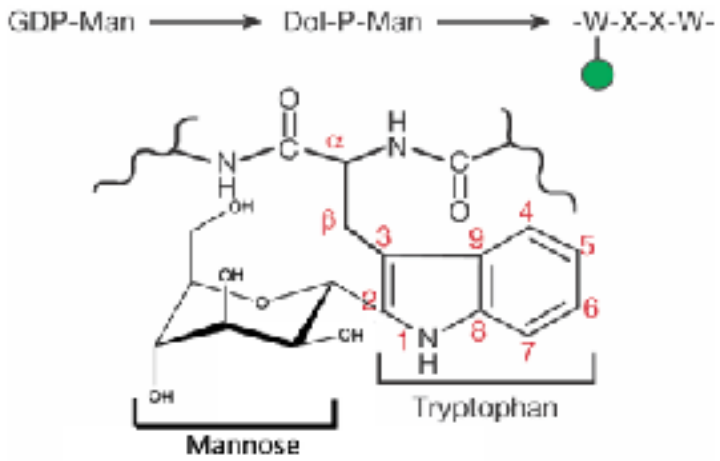
PDB 6PLH, omit mFo-DFc map at 1.5σ, 1.6 Å data

John, Jarva, Shah, Mao, Chappaz, Birkinshaw, Czabotar, Lo, Scott & Goddard-Borger, 2021, Nature Chemical Biology 17: 428-437

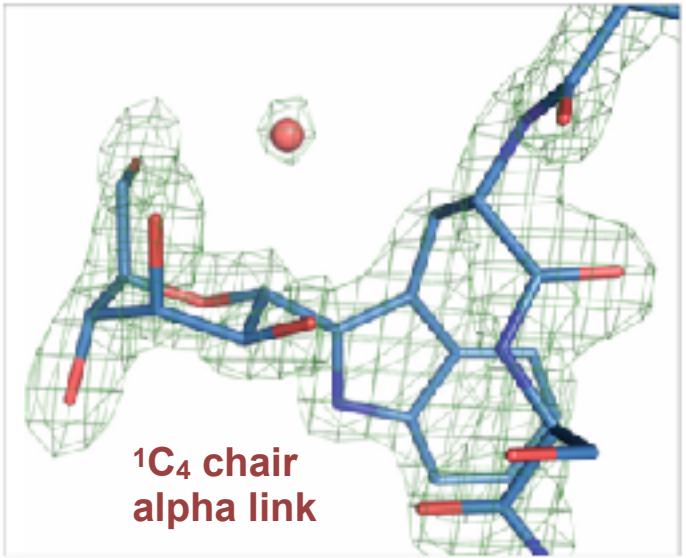
Essentials of Glycobiology [Internet]. 3rd edition.
Varki A, Cummings RD, Esko JD, et al., editors.
Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.

Restraints for C-mannosylation (MKV)

$^1C_4$ chair alpha link



Updated figure courtesy of Prof. Pamela Stanley (Einstein College, USA)

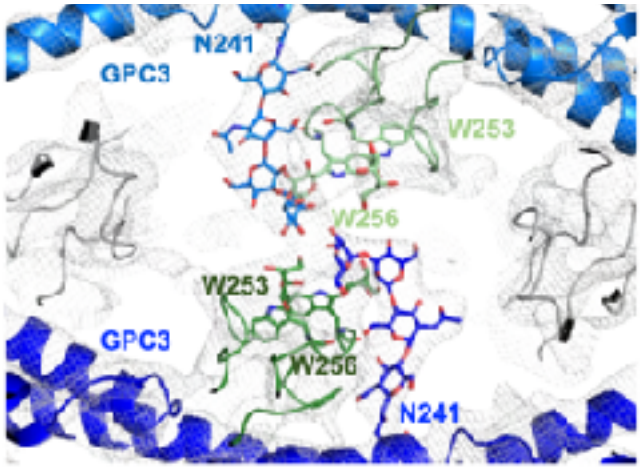


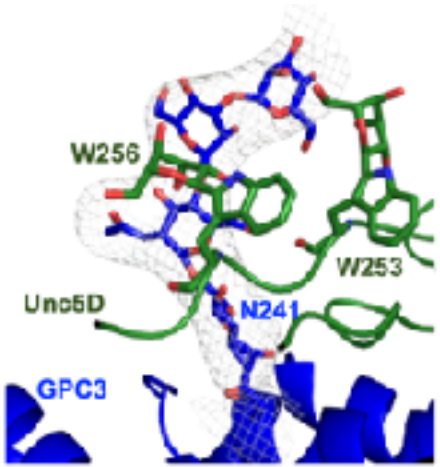
PDB 6PLH, omit mFo-DFc map at 1.5σ, 1.6 Å data

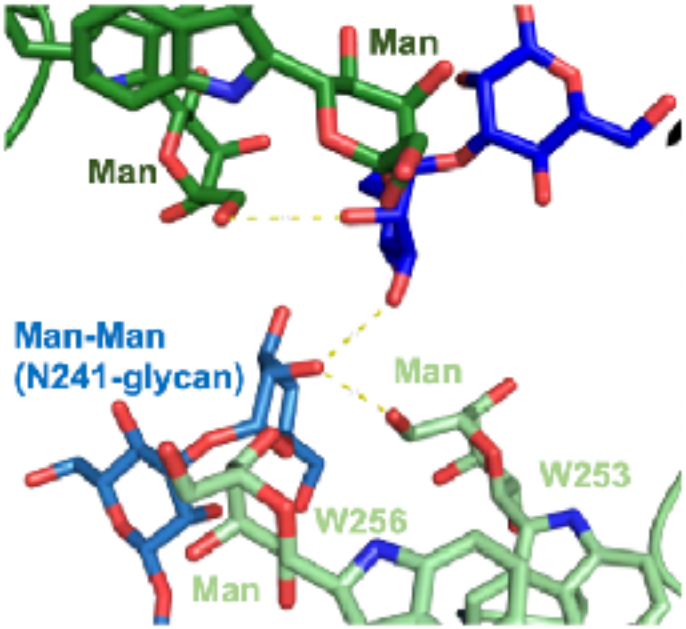
John, Jarva, Shah, Mao, Chappaz, Birkinshaw, Czabotar, Lo, Scott & Goddard-Borger, 2021, Nature Chemical Biology 17: 428-437

Essentials of Glycobiology [Internet]. 4th edition.
Varki A, Cummings RD, Esko JD, et al., editors.
Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.

Restraints for C-mannosylation (MKV)







GPC3

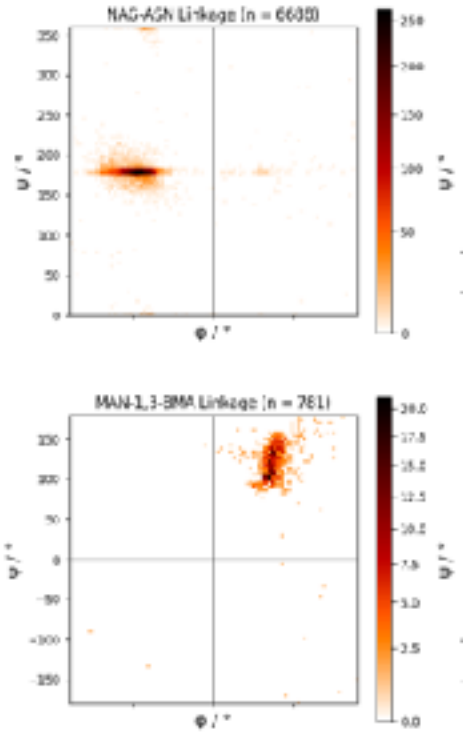
Unc5D

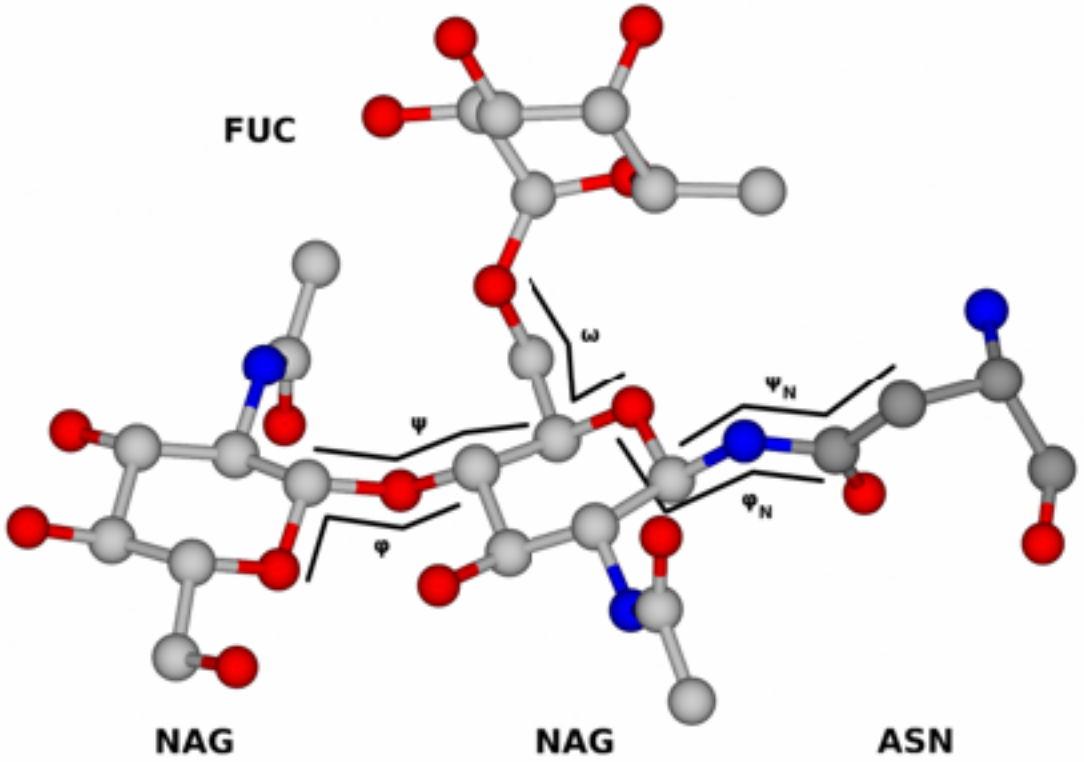
Akkermans, Delloye-Bourgeois, Peregrina, Carrasquero, Kokolaki, Santana, Chavent, Reynaud, Raj, Agirre, Aksu, White, Lowe, Ben Amar, Zaballa, Huo, McCubbin, Comoletti, Owens, Robinson, Castellani, del Toro & Seiradake, 2022, Cell 185(21): 3931-3949.

Support for C-mannosylation (MKV)

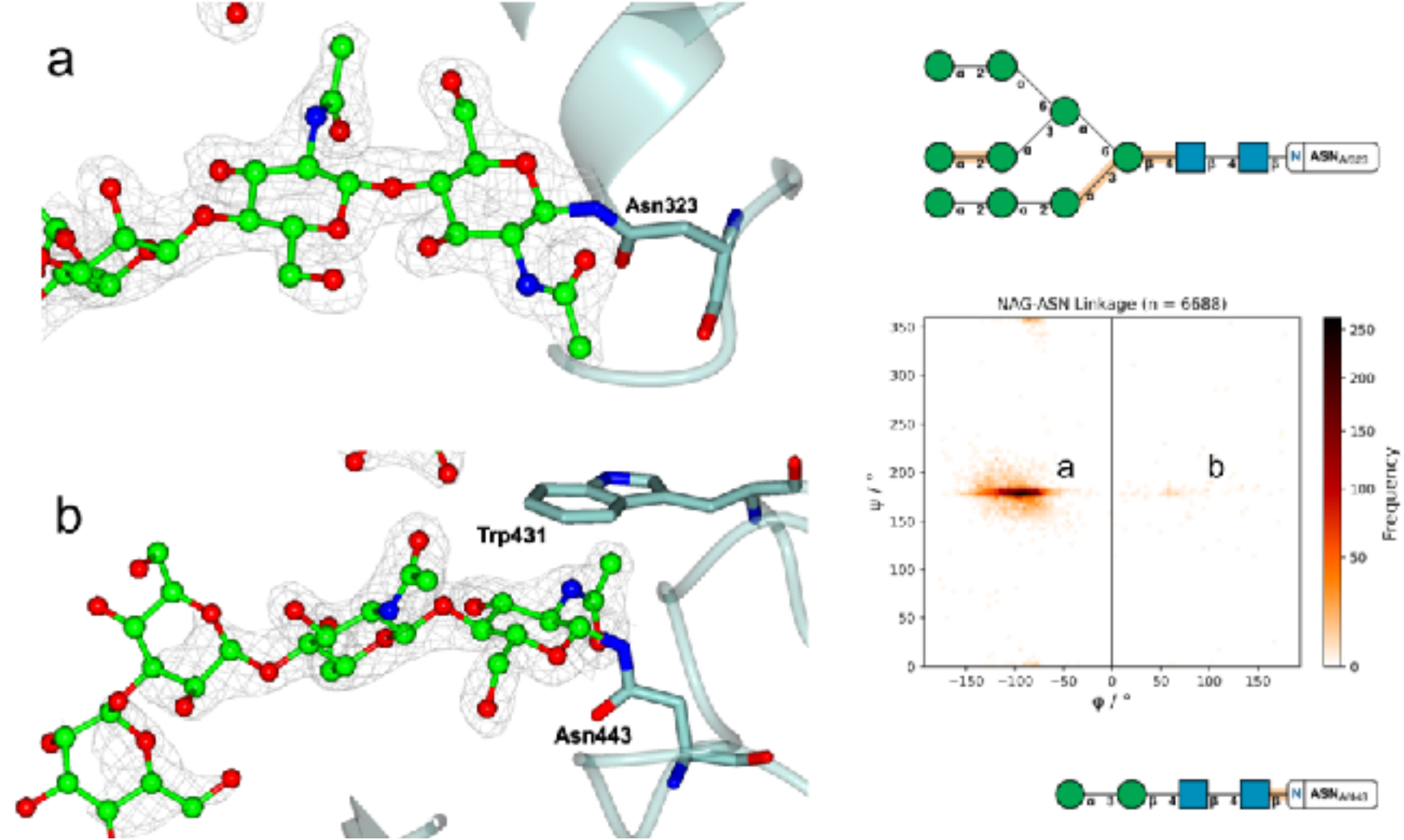
- **Privateer MKV (CCP4 8.0 via update)**
    - checks that mannopyranose ring is $^1C_4$
    - makes sure linkage is alpha, shouts if it isn't
- For refinement at **low resolution with refmac5**
    - External torsion restraints for link and $^1C_4$ conformation (tested, working well)
    - Modify linkage dictionary to impose torsions for $^1C_4$ conformation on the mannose (work in progress)

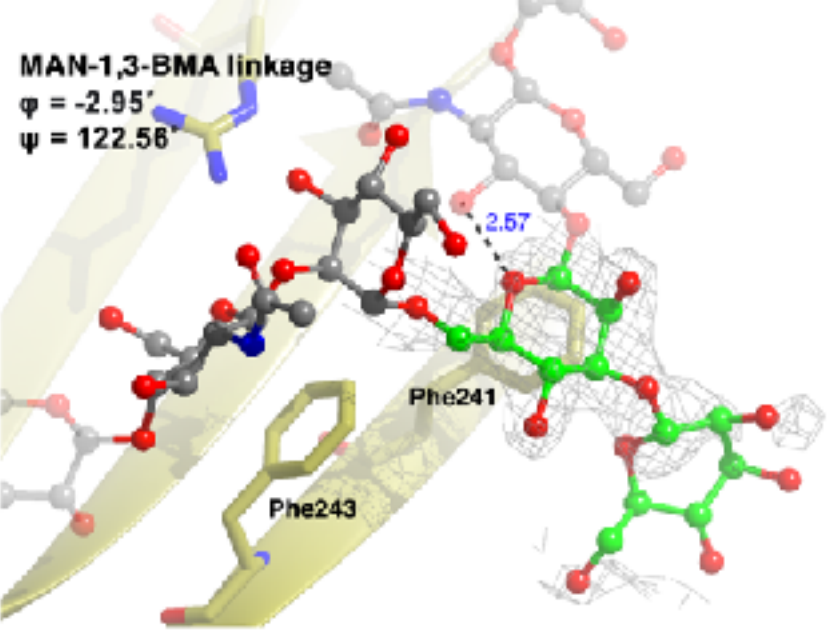
Validation of linkage torsions (MKV)

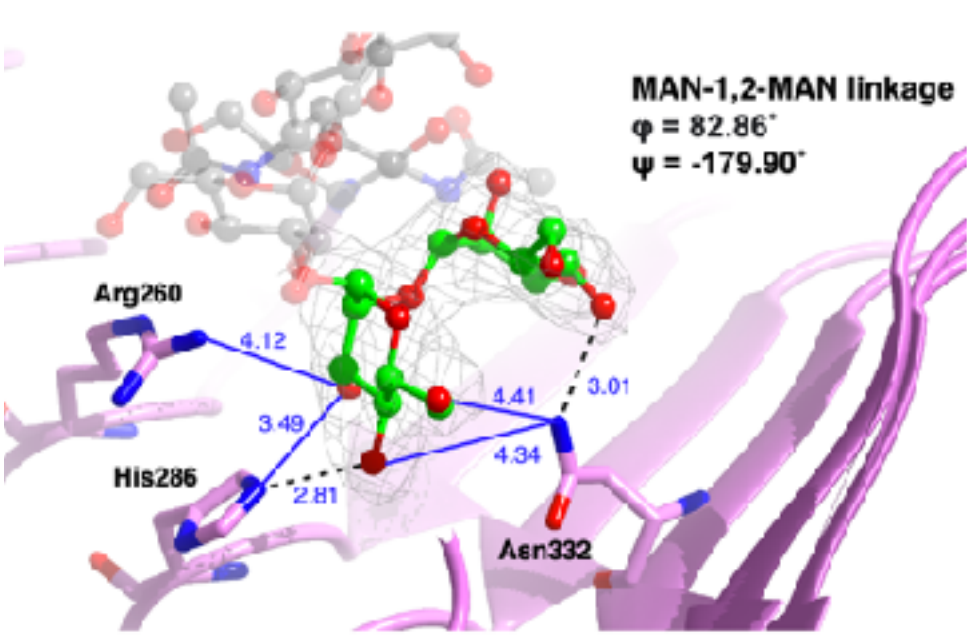


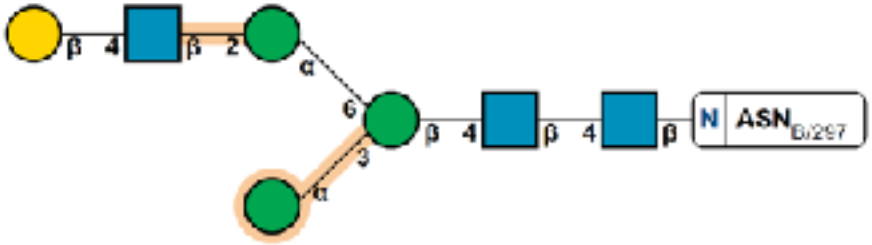


Dialpuri, Bagdonas, Atanasova, Schofield, Hekkelman, Joosten & Agirre, under review.


a
Asn323
b
Trp431
Asn443
NAG-ASN Linkage (n = 6688)
a
b
ψ / °
φ / °
Frequency
ASN







Anomaly!

Well supported by map and interactions

Conclusions

- **Glycan composition**
    - Always make sure your glycans **match biosynthetic pathways**
    - Privateer will check your glycans against glycomics data and suggest **alternatives** if there are inconsistencies
- **Ring conformation**
    - High-energy puckers are **almost never true**
    - These are usually the result of **modelling errors** or **refinement against poor density**, and need to be corrected
- **Glycosidic link torsions**
    - Modelling errors may force links into surprising conformations
    - Not all standout conformations are wrong – **check interactions**!


Manal Alzahrani


Charlie Campbell


Lucy Schofield


Mihaela Atanasova


Ali Darius Khan


Haroldas Bagdonas


Jordan Dialpuri


Katie Tlusty

Acknowledgements

Collaborators

Robbie Joosten (NKI, The Netherlands)
Robert Nicholls (MRC-LMB, Cambridge)
Elisa Fadda (Maynooth University, Ireland)
Hiren Joshi (University of Copenhagen)
Elena Seiradake (University of Oxford)
Martin Frank (Biognos, Sweden)
Frédérique Lisacek (SIB, Switzerland)
Sameer Velankar & Gerard Kleywegt (PDBe & AFDB)
CCP4 & CCP-EM core teams

Stock figures: Wikimedia Commons