Privateer: validation of carbohydrate structures

YSBL
York Structural Biology Laboratory

University of York

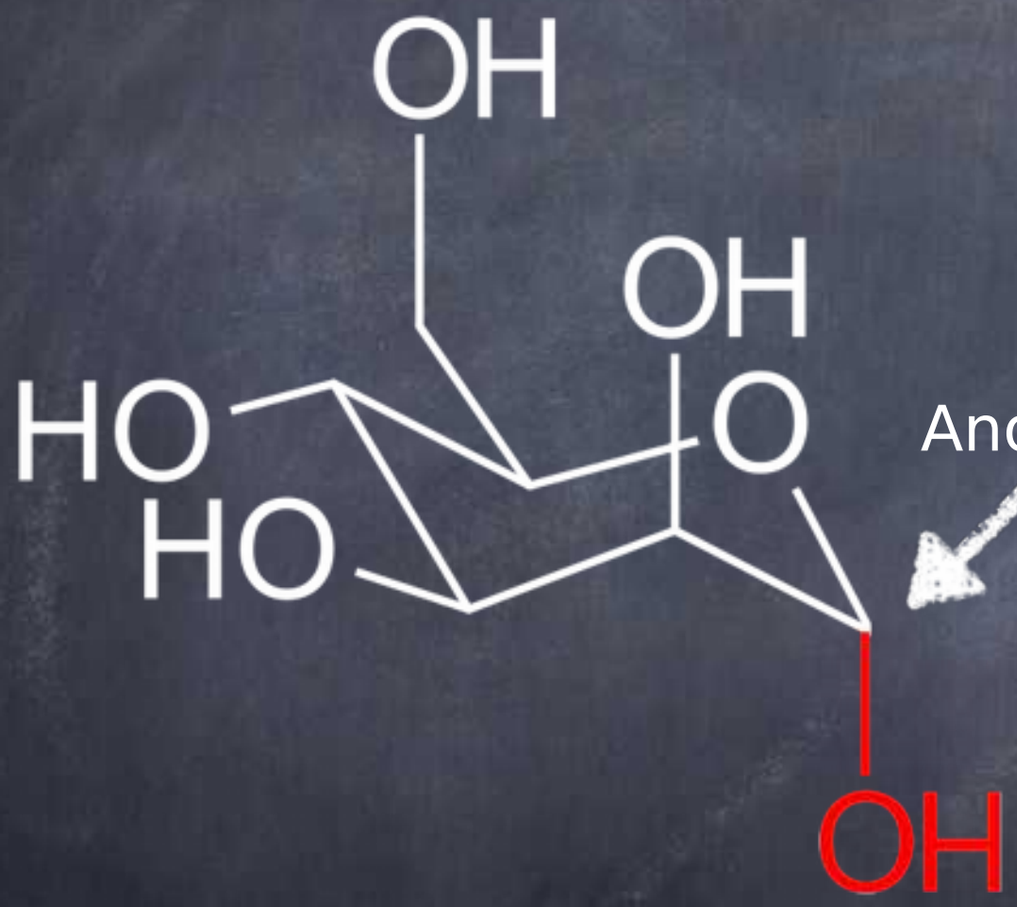
Carbohydrates

- Biological molecules composed of Carbon, Oxygen and Hydrogen atoms
- Can be found covalently linked to protein, but they are not encoded in genome!
  - They are added/removed by enzymes
- Nearly always found in cyclic form

Carbohydrates

- Stereochemistry defines the sugar
- Clear conformational preferences
  - Only one “rotamer” 99.9% of the time
  - Transitions usually due to external factors (e.g. in enzyme’s active site)

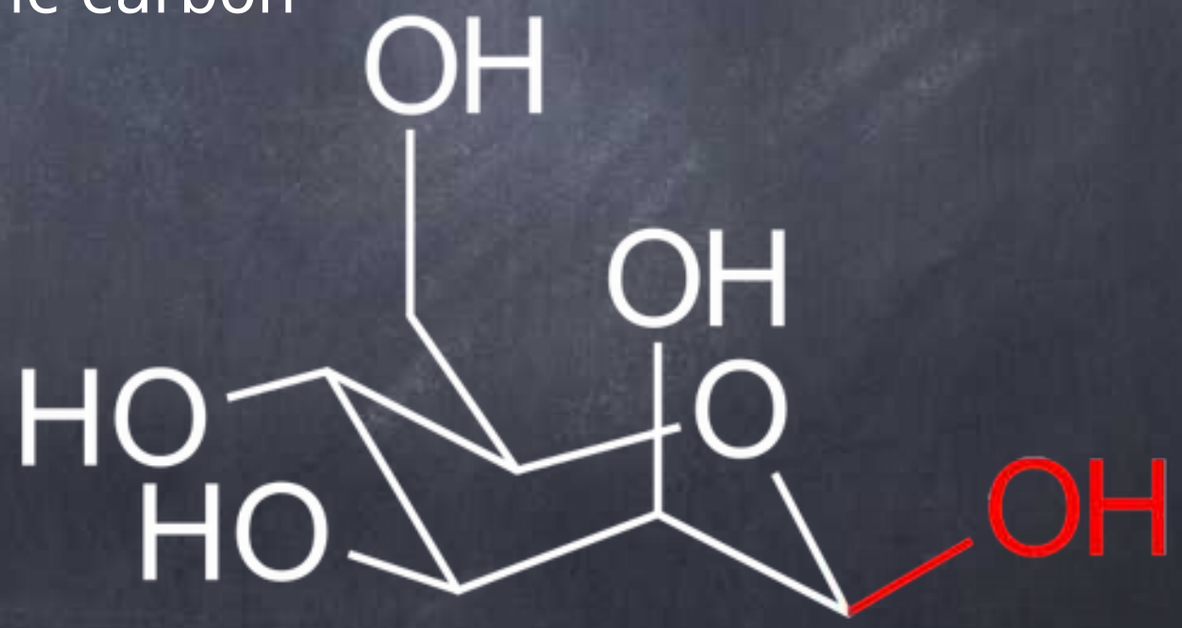
Stereochemistry



Anomeric carbon

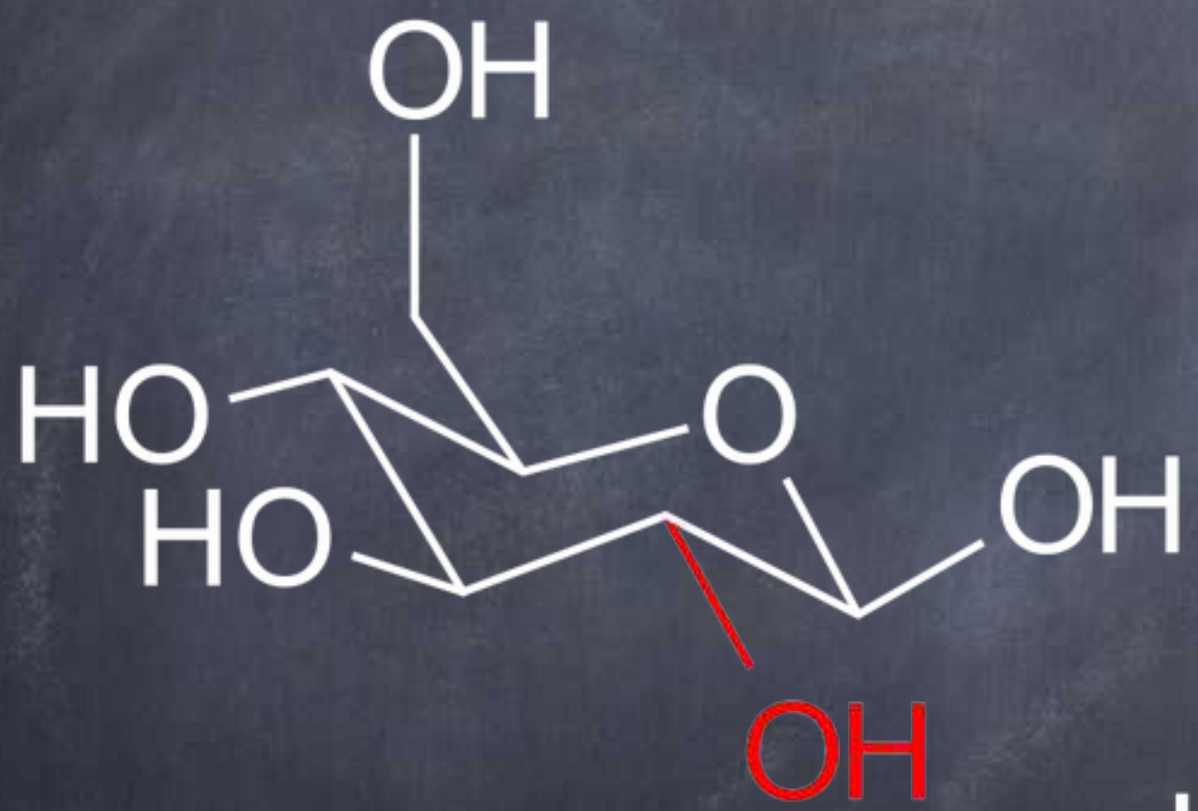
alpha-d-mannose
‘MAN’ in the PDB

Same chemical formula!
$C_6 H_{12} O_6$



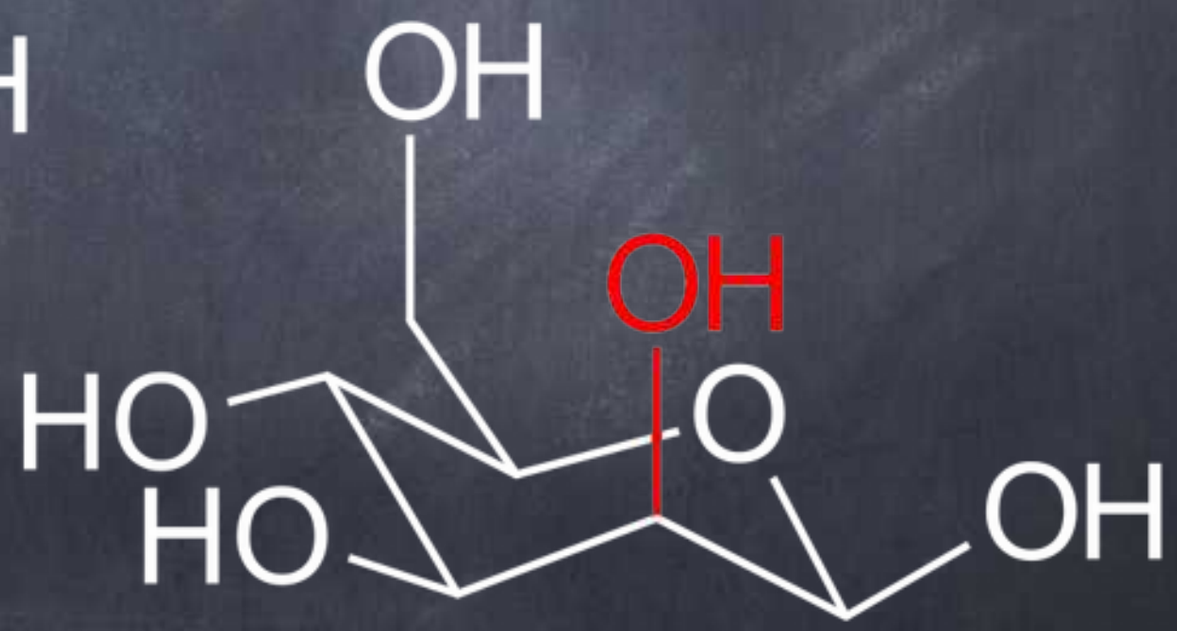
beta-d-mannose
‘BMA’ in the PDB

Stereochemistry



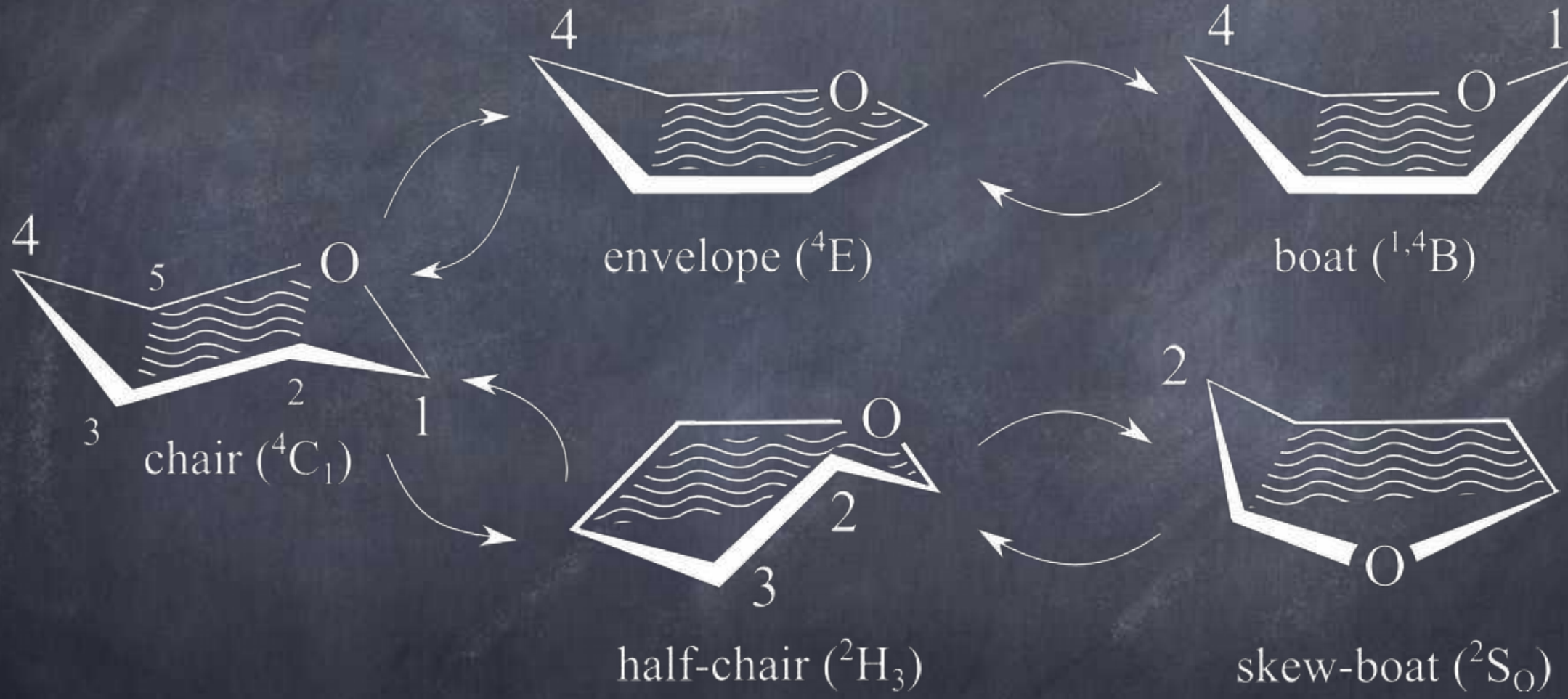
beta-d-glucose
‘BGC’ in the PDB

Same chemical formula!
$C_6H_{12}O_6$



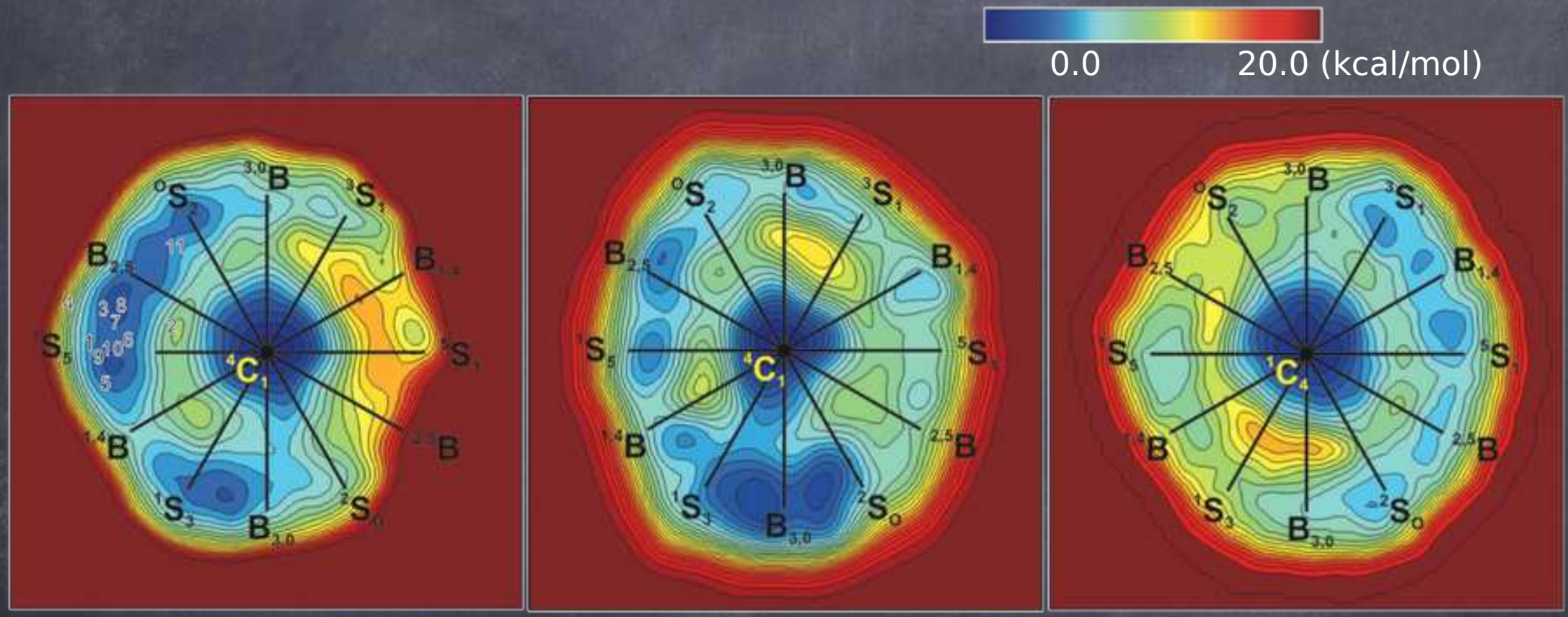
beta-d-mannose
‘BMA’ in the PDB

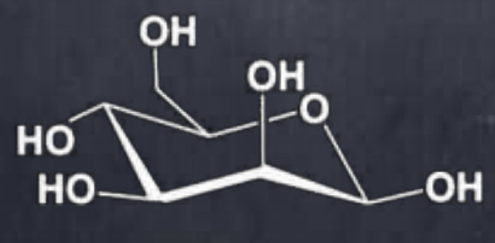
Conformations



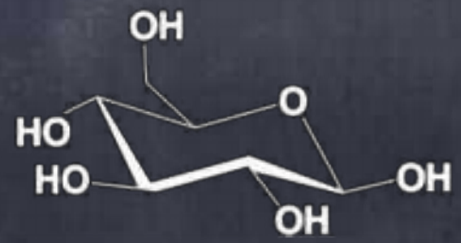
Wavy lines = atoms are roughly coplanar

Conformations

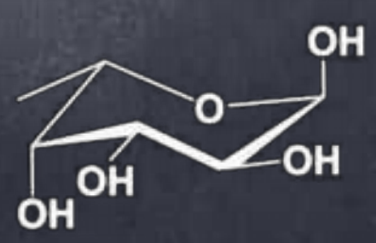




beta-D-mannopyranose



beta-D-glucopyranose



alpha-L-fucopyranose

Ardevol et al., JACS 132 (2010)

In context

- Covalently linked to protein
  - N-glycans: GlcNac linked to Nitrogen in ASN
  - O-glycans: GalNac, Fucose, Glucose, Mannose or GlcNac. Linked to Oxygen in SER or THR
- As ligands
  - Bound to proteins by stacking interactions and/or H-bonds

N- and O-glycans

N-Acetyl-glucosamine
(NAG)

Mannose
(BMA, MAN)

Fucose
(FCA, FCB, FUC)

Galactose
(GAL)

N-Acetyl-galactosamine
(NGA)

Glucose
(GLC, BGC)

2-keto-3-deoxynononic acid
(KDN)

Galacturonic acid
(ADA, GTR)

Glucosamine
(GCS)

Xylose
(XYS, XYP)

Iduronic acid
(IDR)

Mannuronic acid
(BEM)

N- and O-glycans

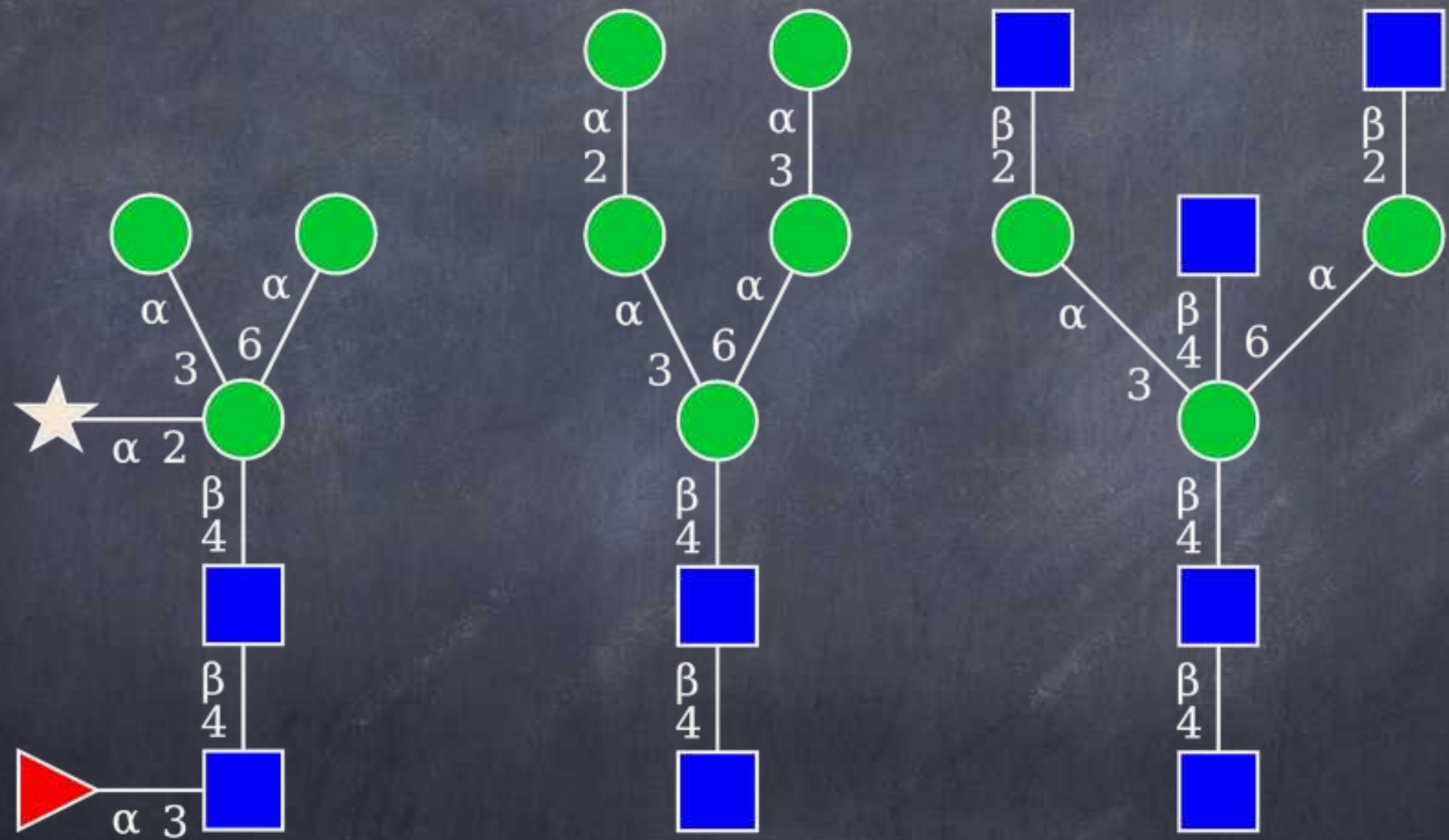
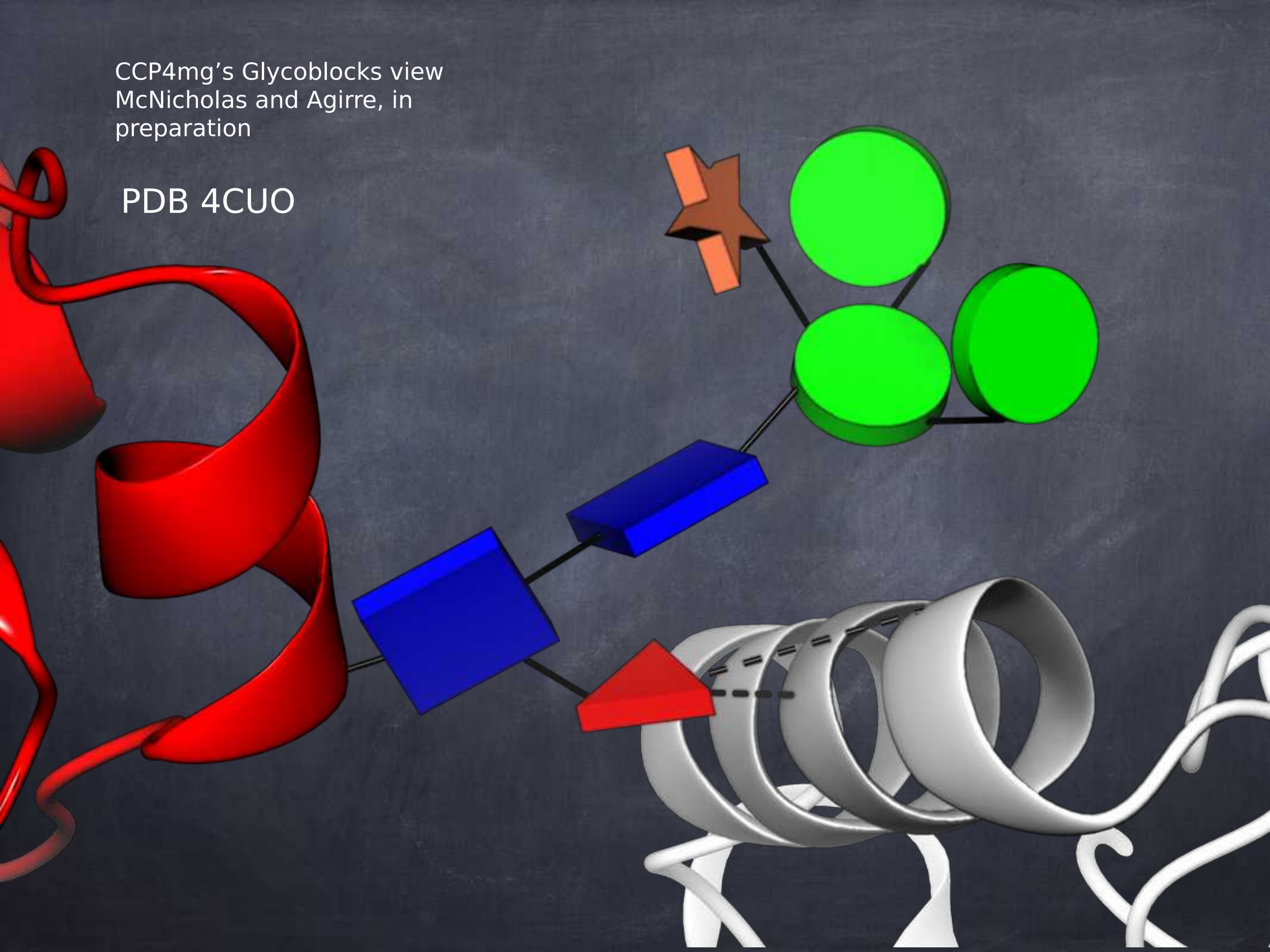


Figure made with GlycanBuilder

CCP4mg’s Glycoblocks view
McNicholas and Agirre, in preparation

PDB 4CUO

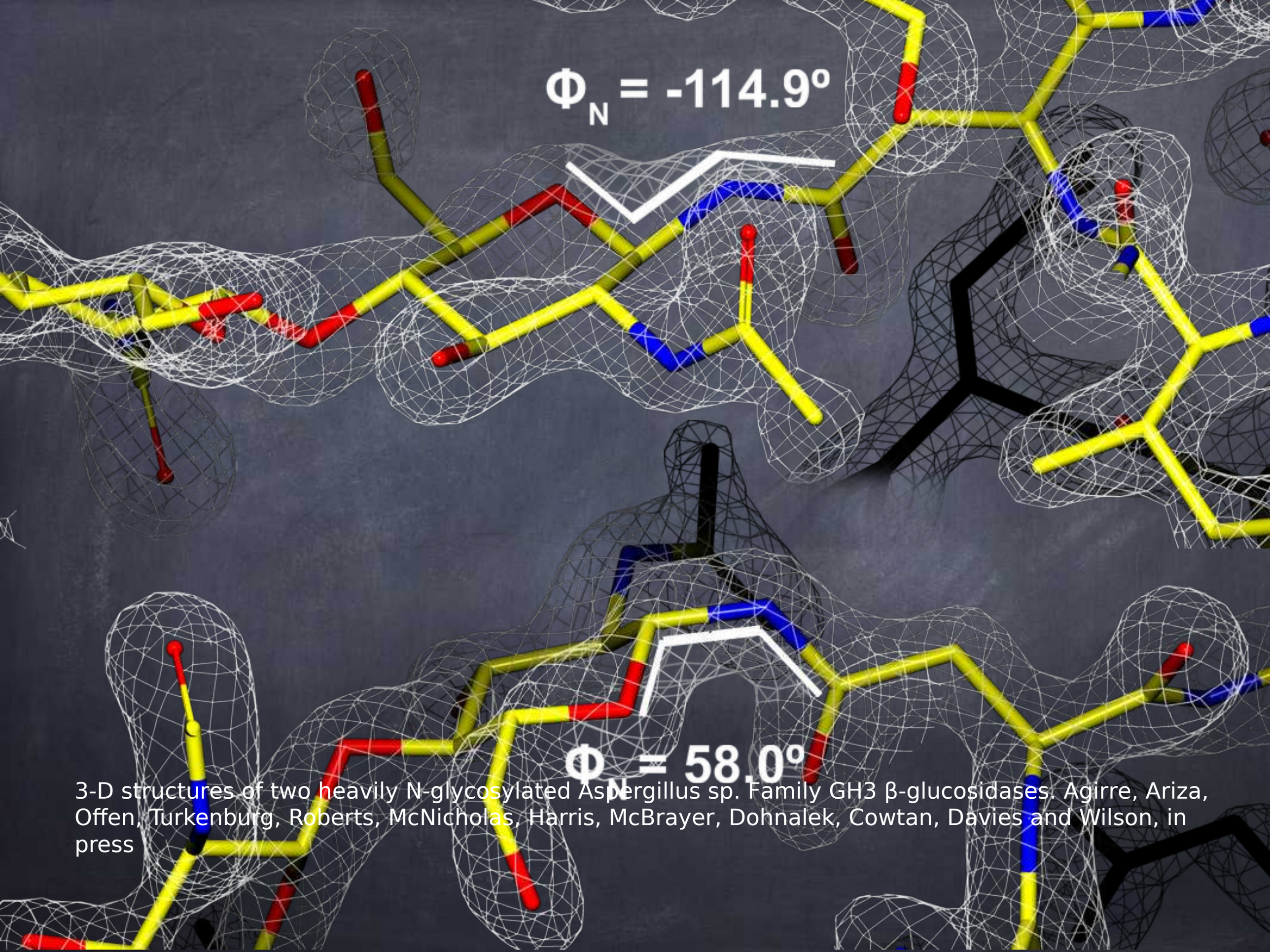


CCP4mg’s Glycoblocks view
McNicholas and Agirre, in preparation

PDB 3SGK

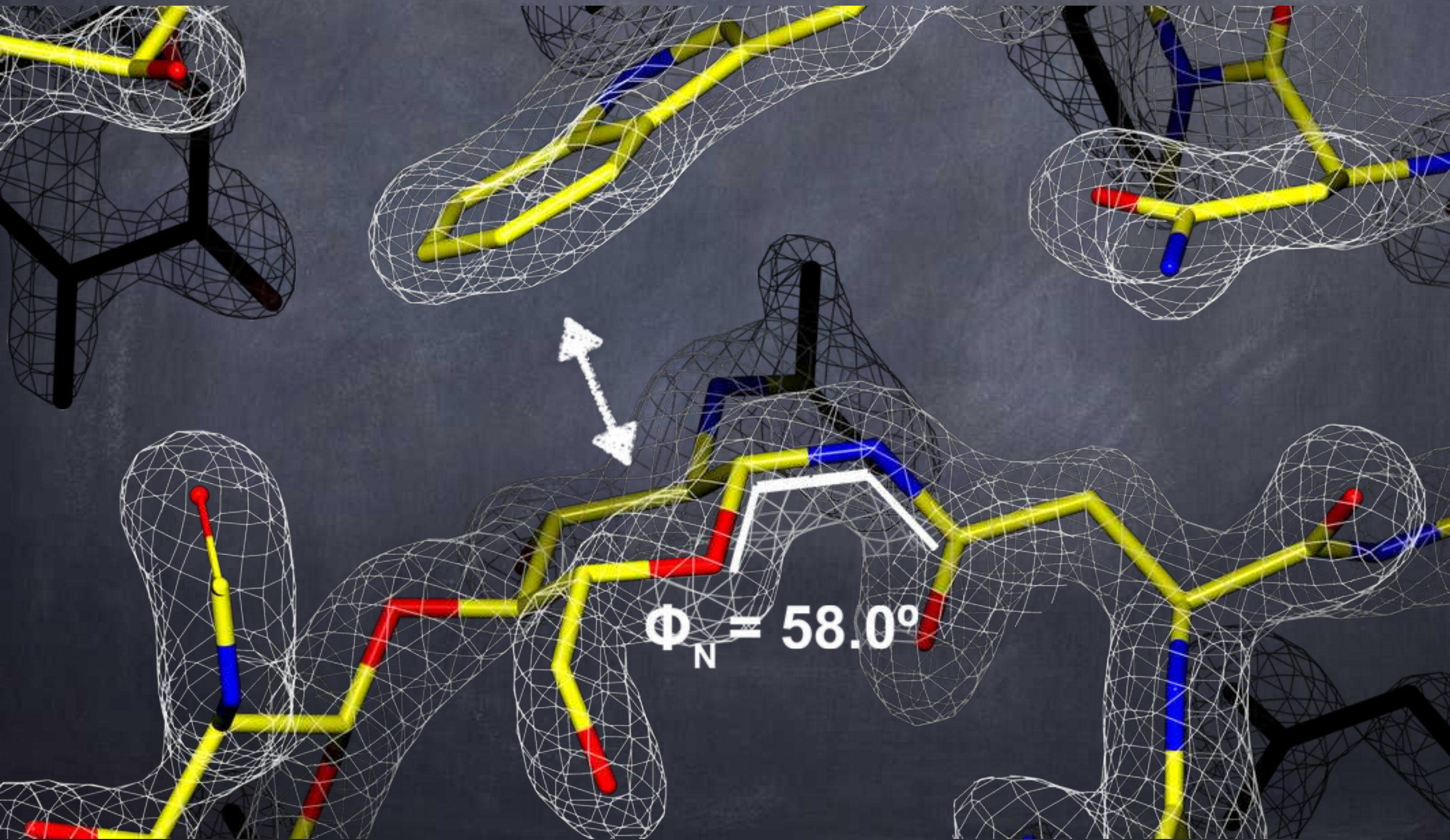
N- and O-glycans

- Glycans provide important contacts (H-bonds, stacking interactions) that are key to the stability of glycoproteins, including antibodies
- Therefore, the specificity of their 3D structure is critical. Incorrect glycans usually means incorrect folding!



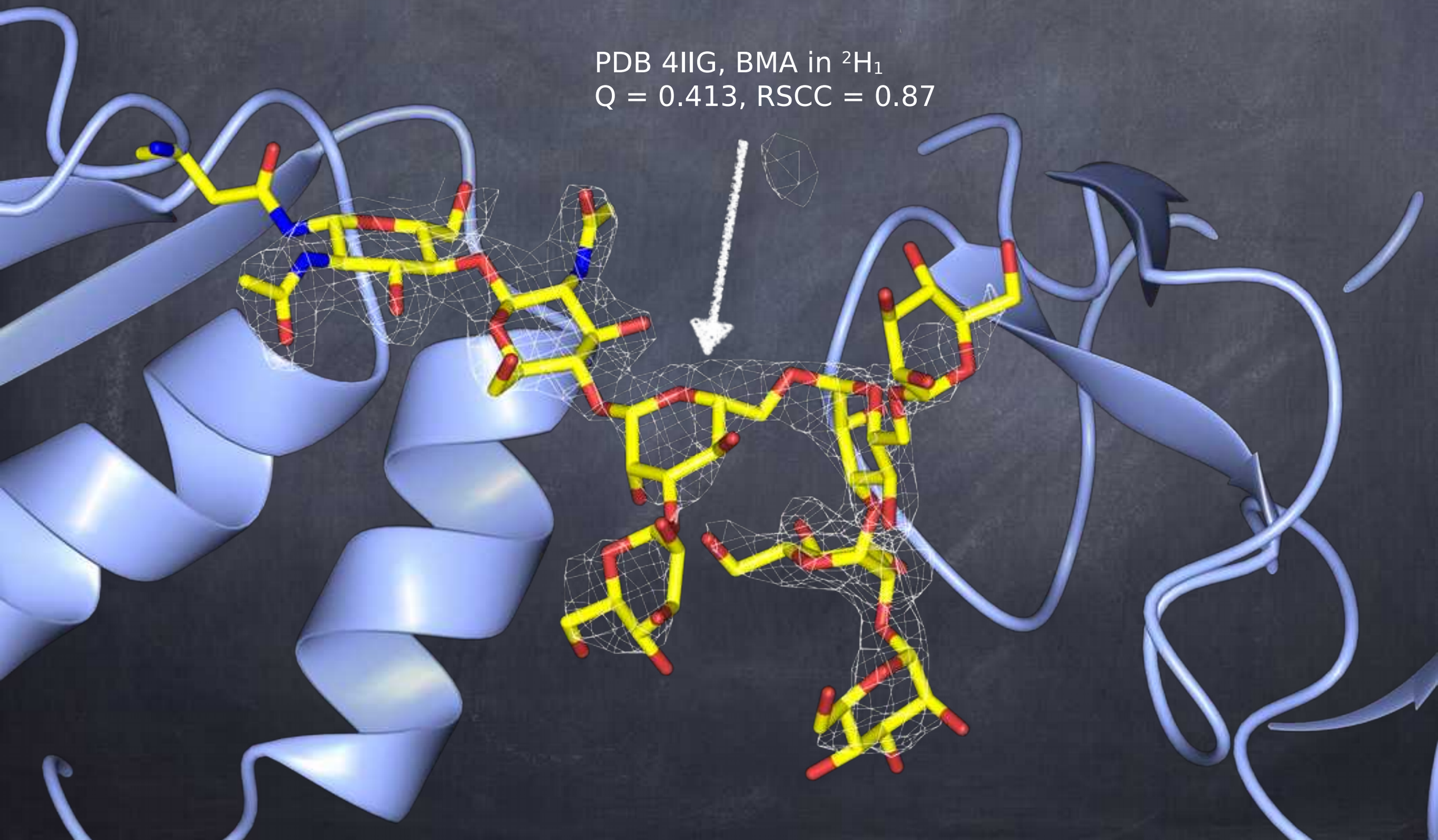
3-D structures of two heavily N-glycosylated Aspergillus sp. Family GH3 β-glucosidases. Agirre, Ariza, Offen, Turkenburg, Roberts, McNicholas, Harris, McBrayer, Dohnalek, Cowtan, Davies and Wilson, in press

Stacking interactions can favour other link conformations!



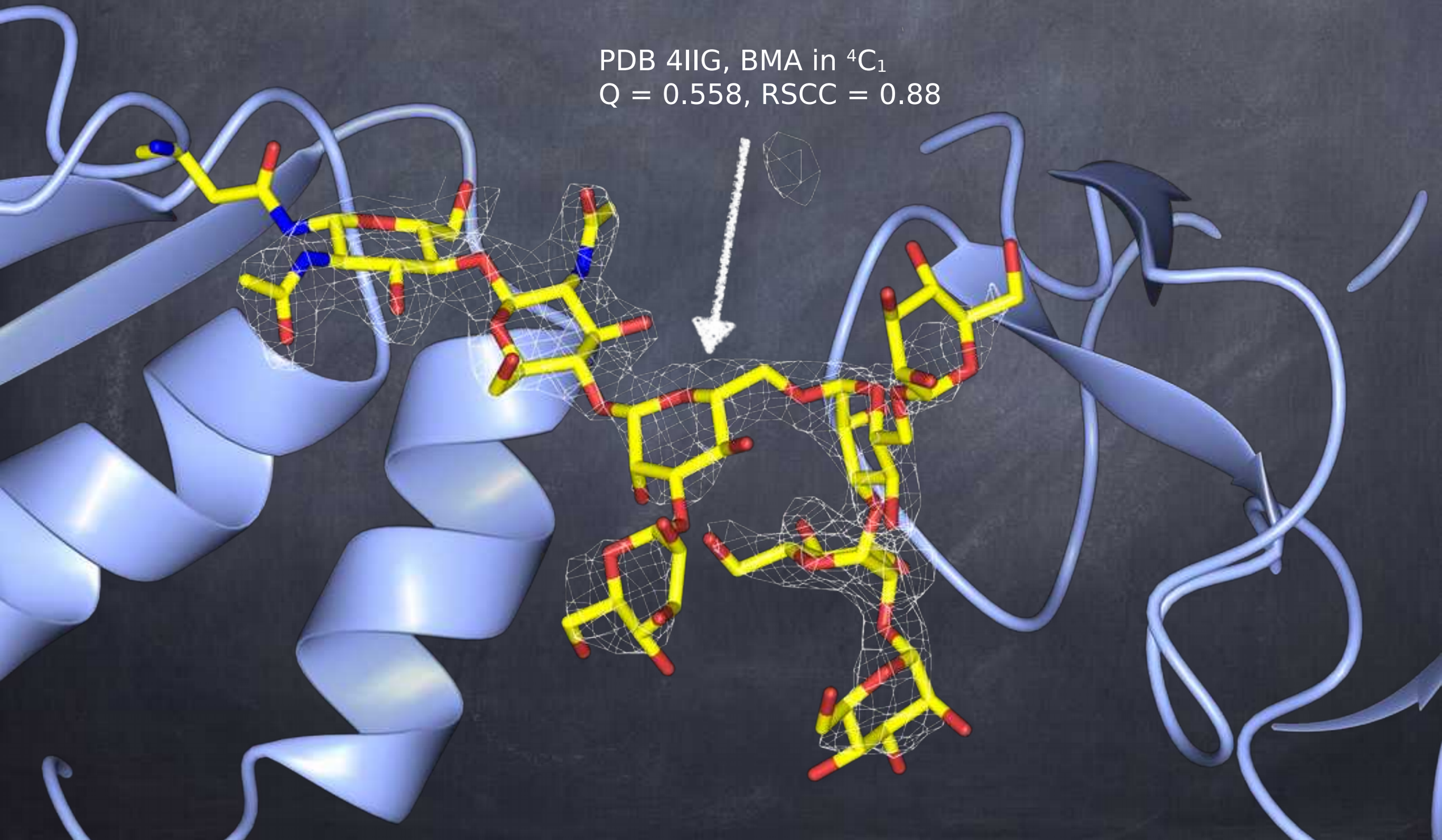
Wrong conformation

PDB 4IIG, BMA in $^{2}H_{1}$
Q = 0.413, RSCC = 0.87



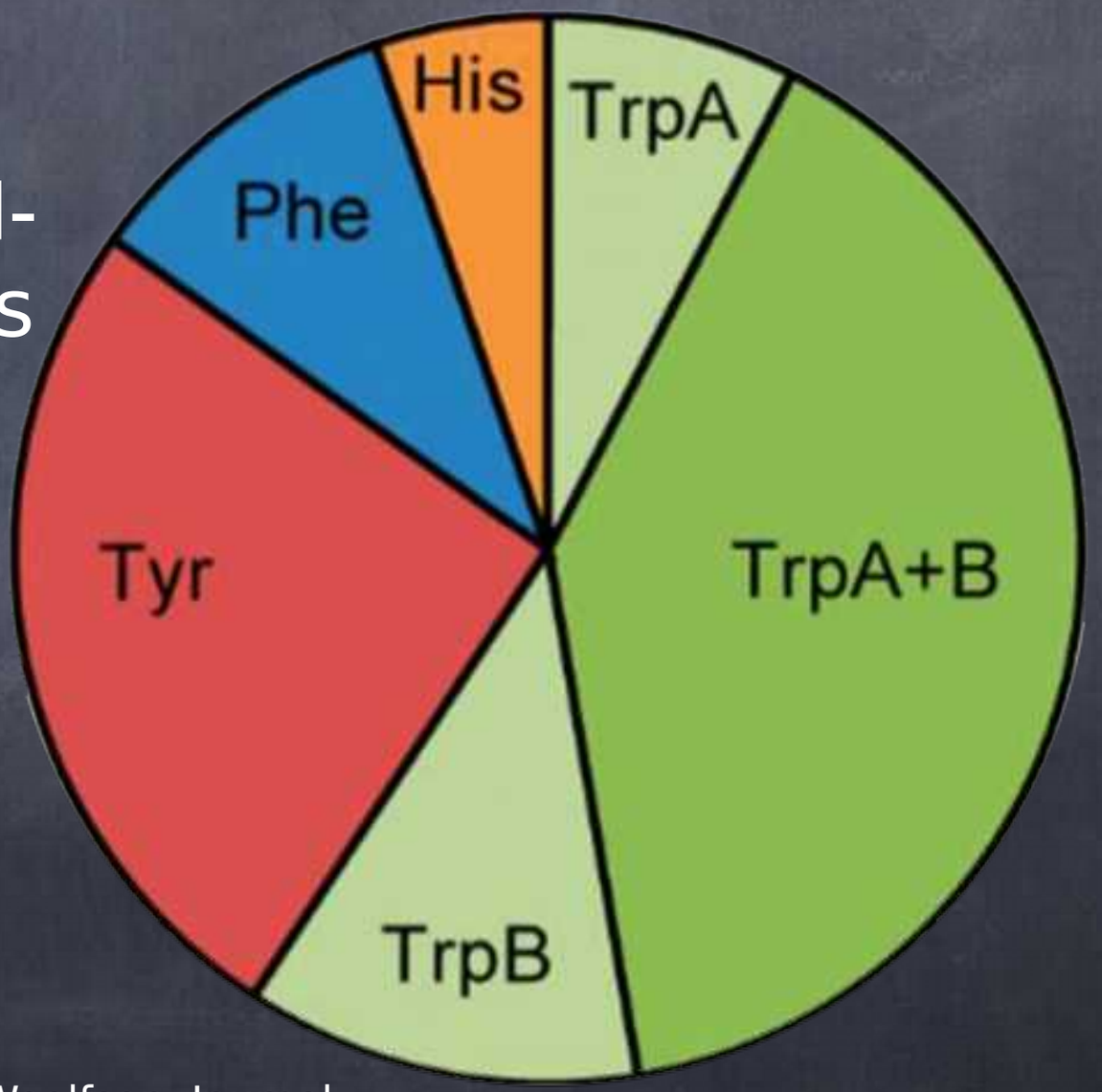
Wrong conformation

PDB 4IIG, BMA in $^{4}C_{1}$
Q = 0.558, RSCC = 0.88



Ligand sugars

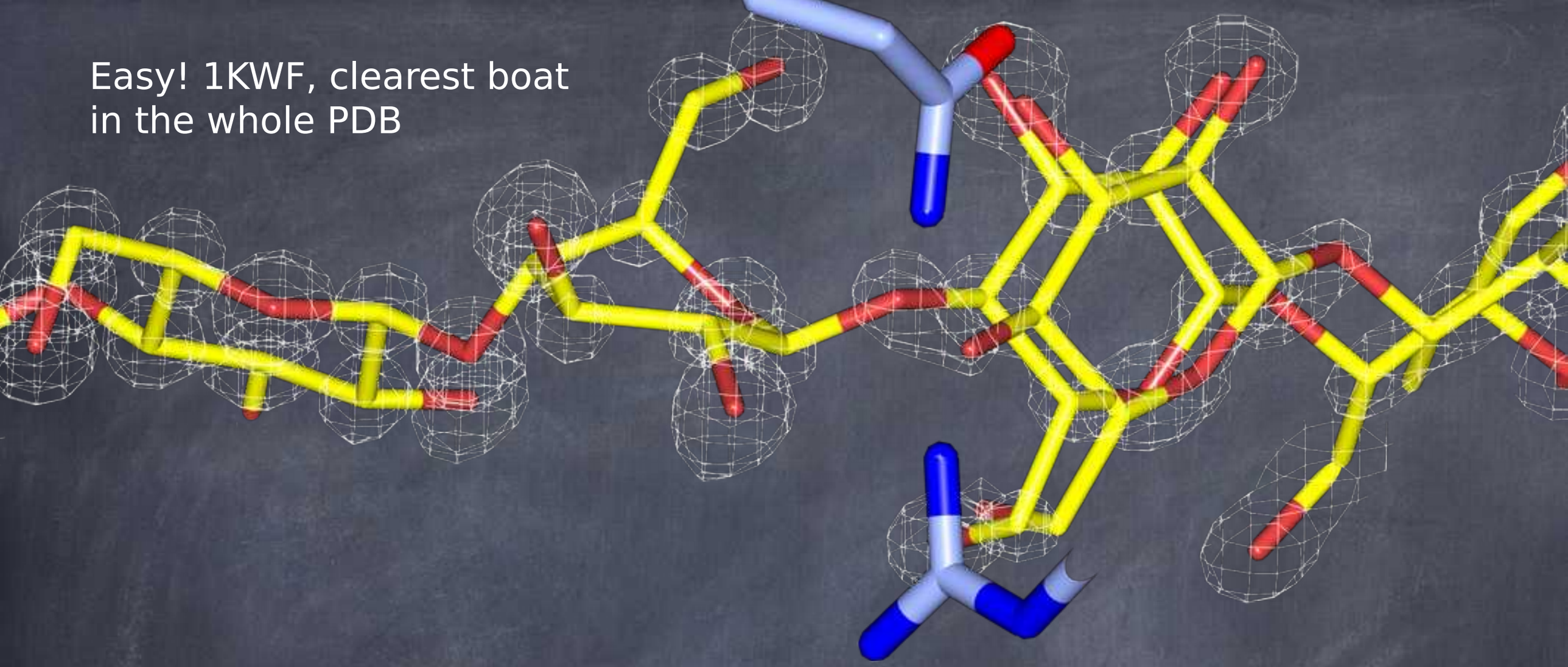
- Typically bound via CH-Pi stacking interactions



TRP accounts for ~60% of them

Carbohydrate-aromatic interactions in proteins.
Hudson, Bartlett, Diehl, Agirre, Gallagher, Kiessling, Woolfson. Journal of the American Chemical Society, in press.

Easy! 1KWF, clearest boat
in the whole PDB



Easy!

Low resolution
Harder, prone to
errors like this

Easy!

It is in a higher-energy conformation

It shouldn’t be

active site miles away

Easy!

Summary | 3D View | Sequence | Annotations | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | **Methods** | Geometry | Links

X-RAY DIFFRACTION

2QZ3

Display Files
Download Files
Share this Page

Materials and Methods page

Crystallization — Hide

Crystallization Experiments

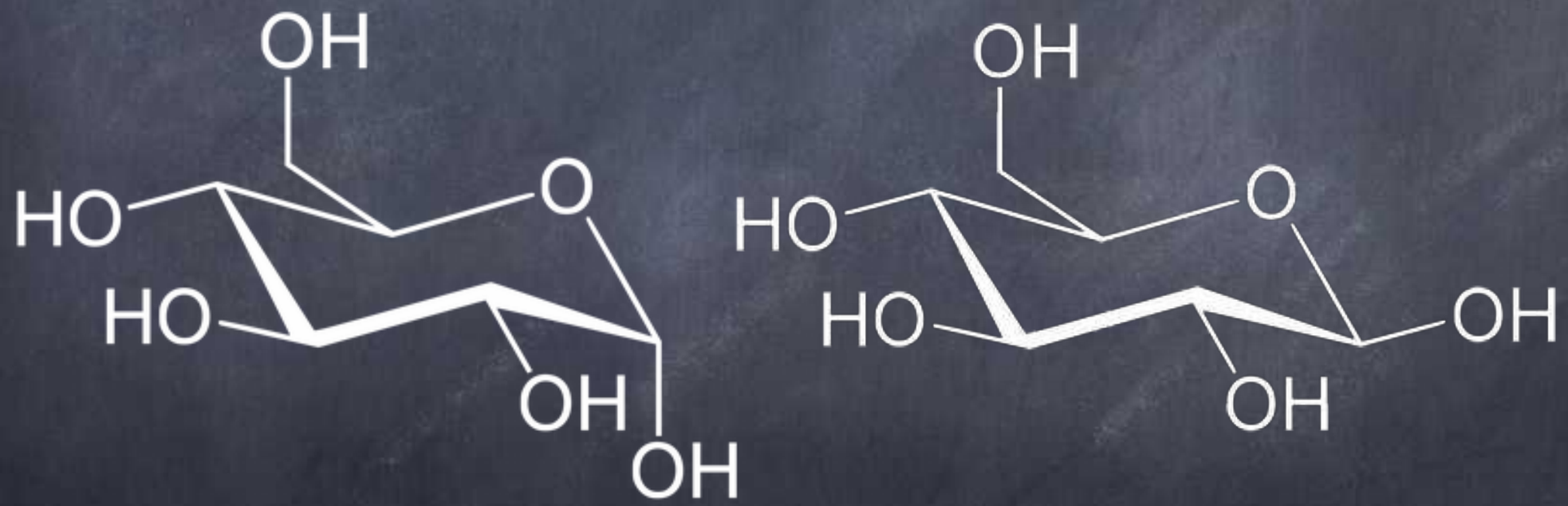
Method	VAPOR DIFFUSION, HANGING DROP
pH	8.5
Temperature	277.0
Details	0.2M ammonium sulphate, 0.1M Tris-HCl pH 8.5, 30% isopropanol, VAPOR DIFFUSION, HANGING DROP, temperature 277K

A practical application

- Say we want to build software for detecting and modelling sugar moieties in electron density maps
- Method: matching a fingerprint across the map
- Fingerprint: a synthesis of superposed sugar moieties and their environment

“About 30% of all carbohydrate containing PDB entries comprise one or several errors”

Lütteke et al, Carb Res 339(5) 2004

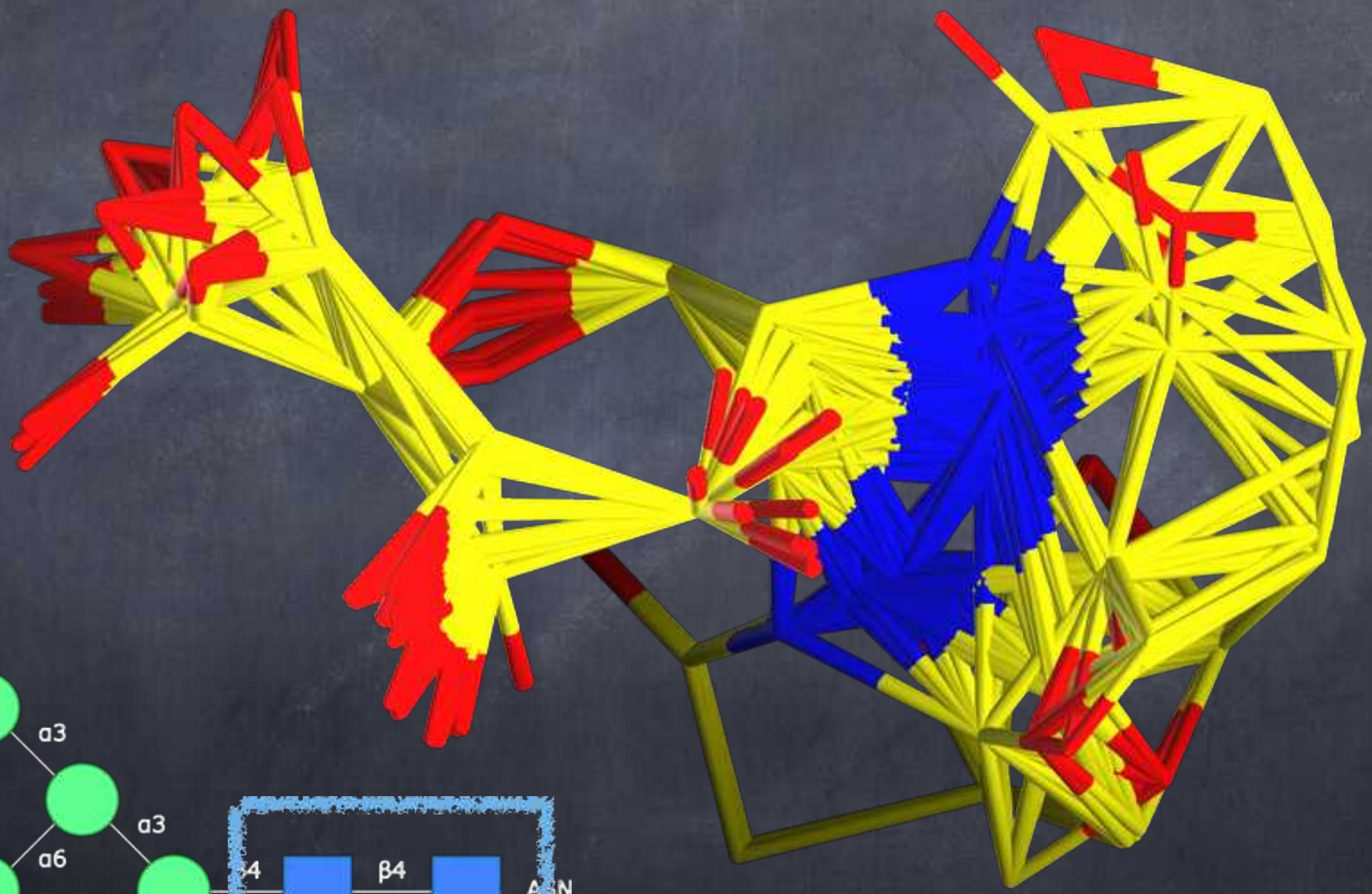


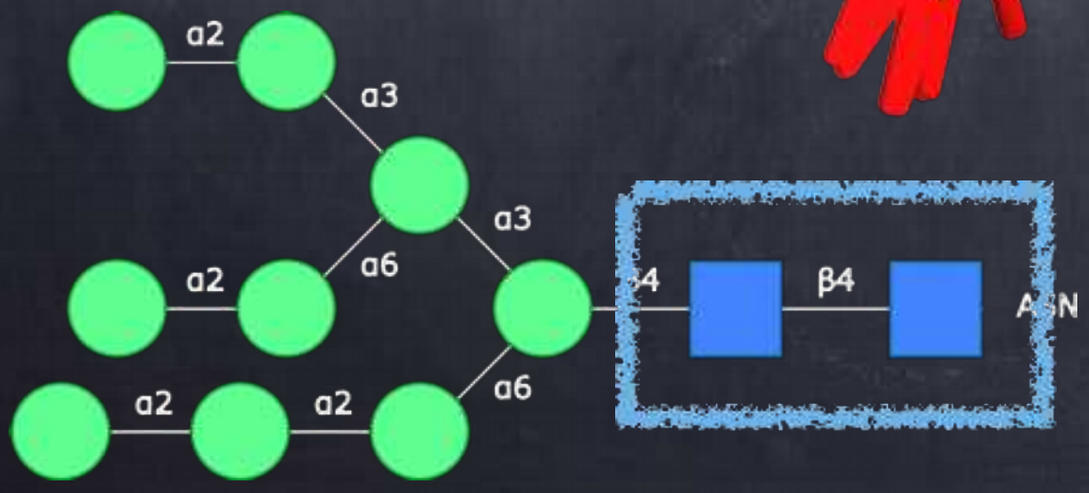
GLC
(alpha anomer)

BGC
(beta anomer)

Initial attempt

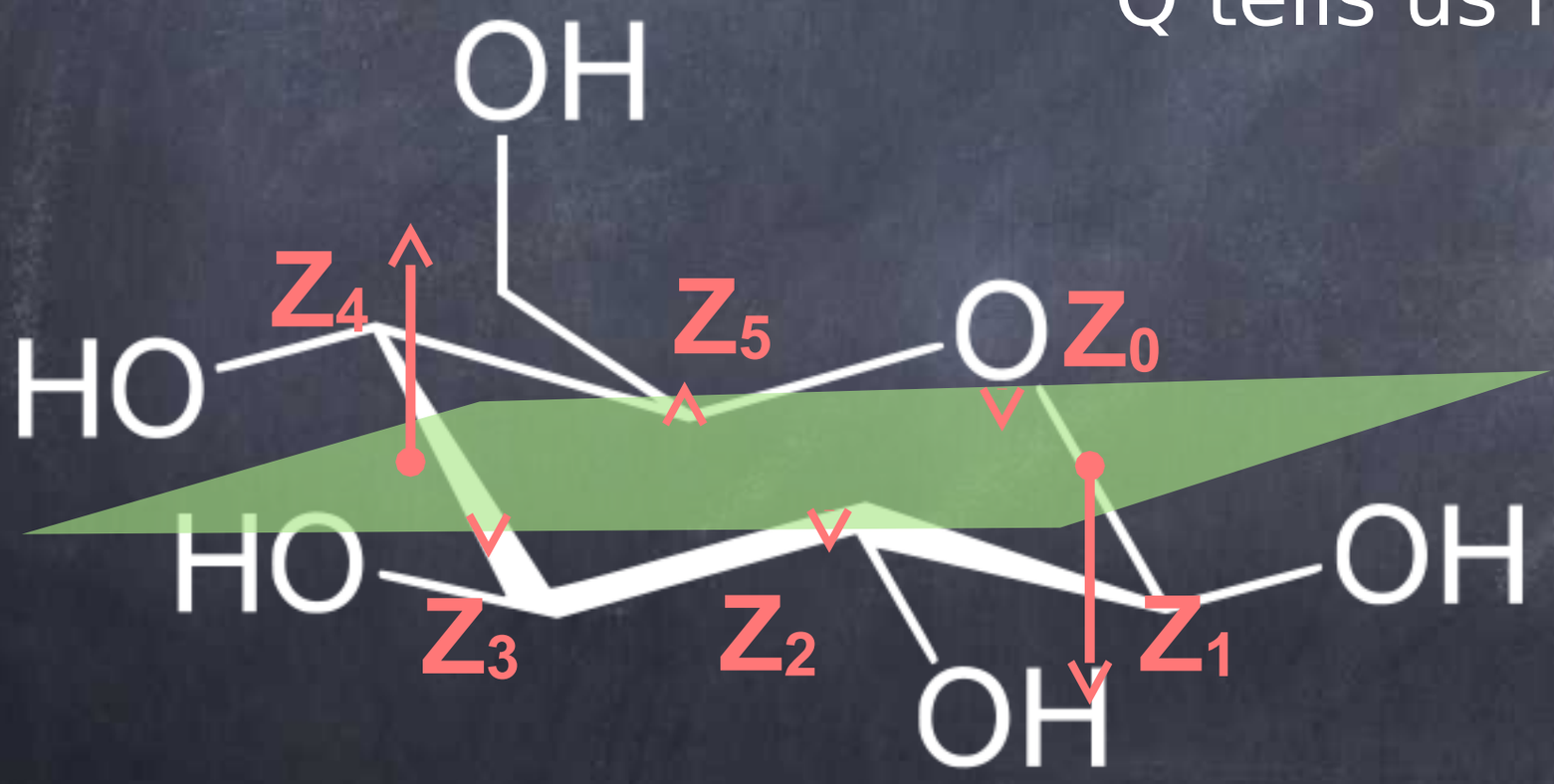
all GlcNac’s in one glycoprotein





Cremer-Pople analysis

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



Q tells us how much:

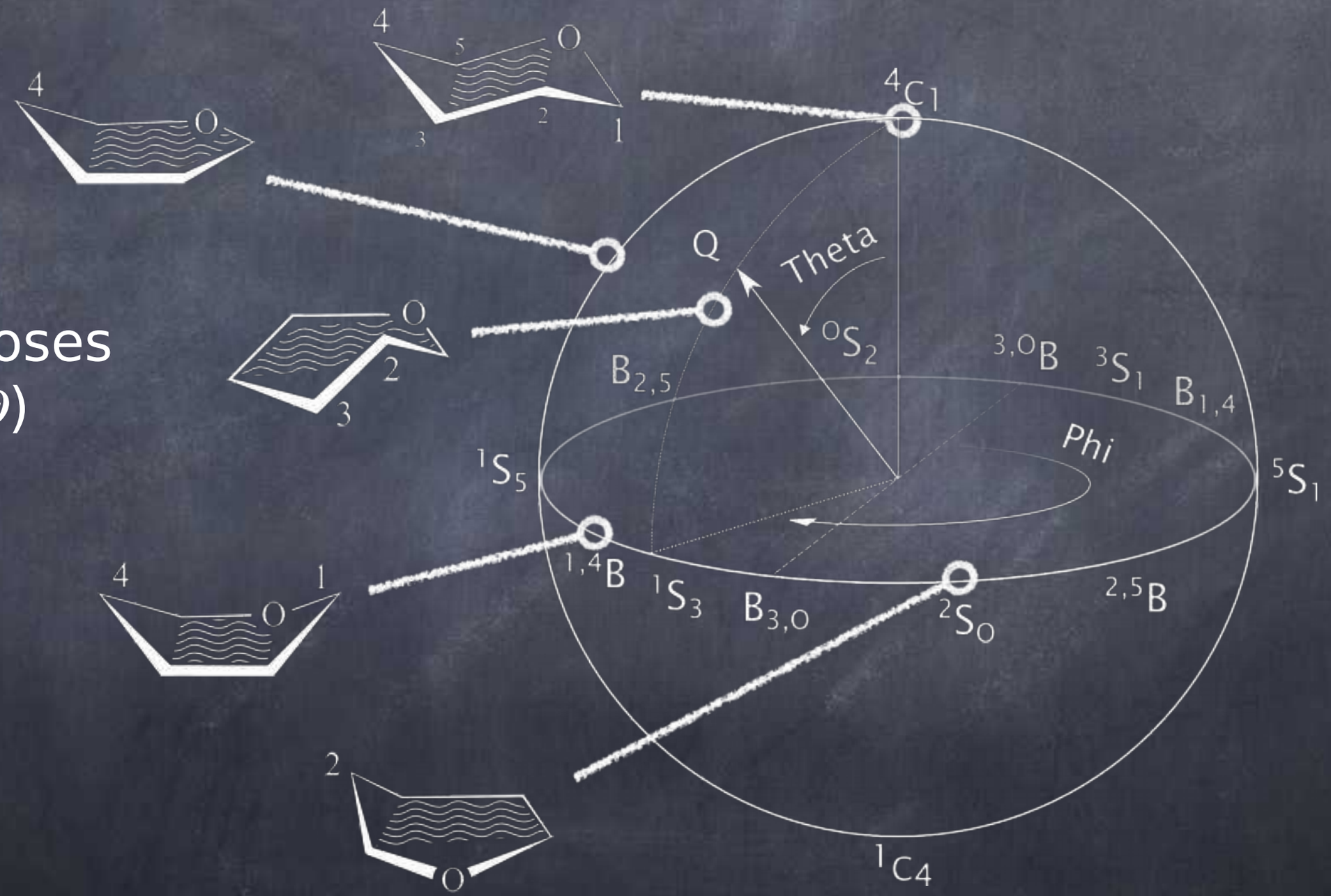
$$Q = \sqrt{\sum_{i=0}^{5} Z_i^2}$$

“total puckering amplitude”

Cremer and Pople, JACS 97(6) (1975)

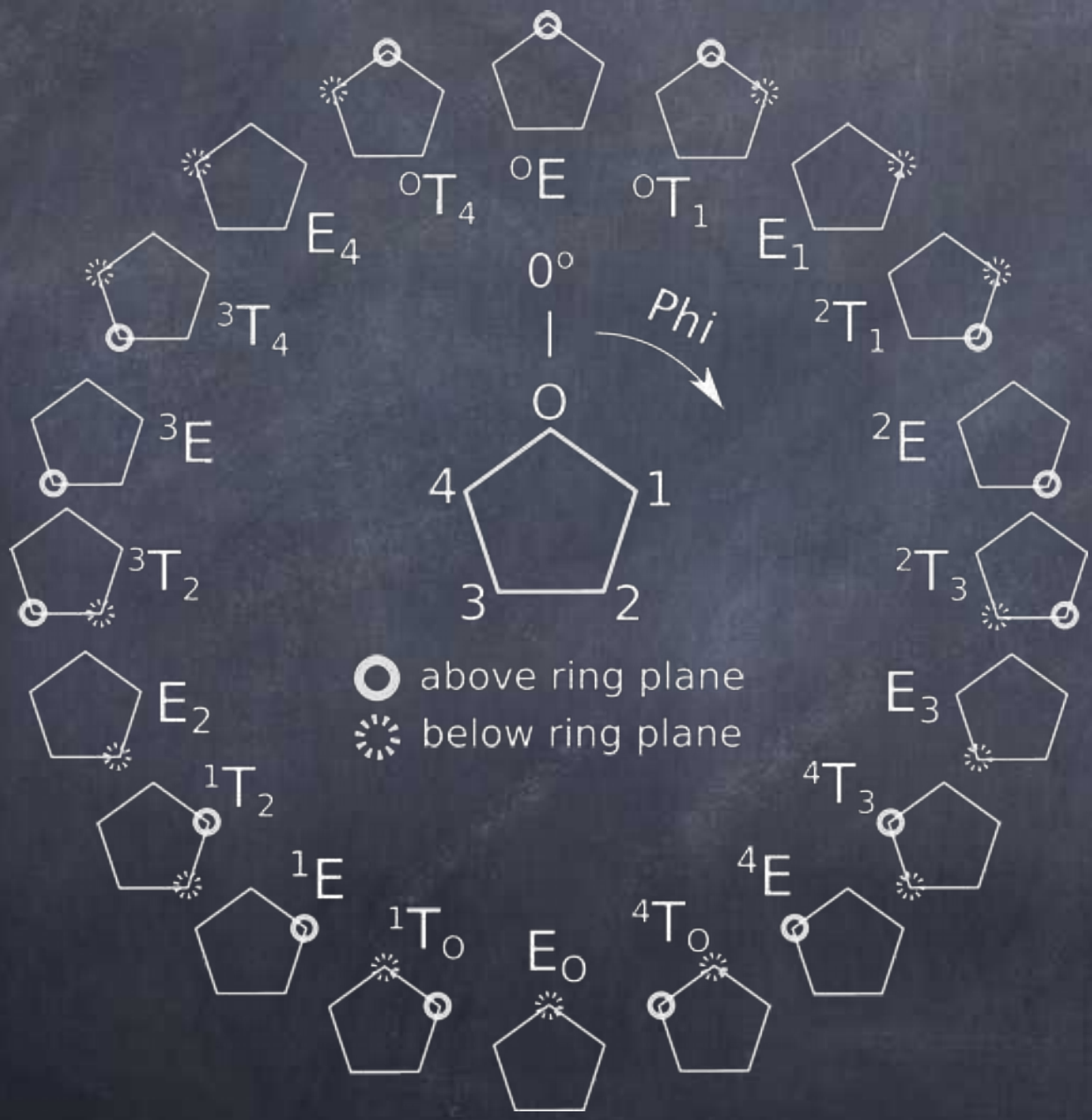
Cremer-Popple analysis

Pyranoses
(Q,Φ,Θ)



Agirre and Cowtan, Computational Crystallography Newsletter Jan:10-12 (2015)

Cremer-Popple analysis



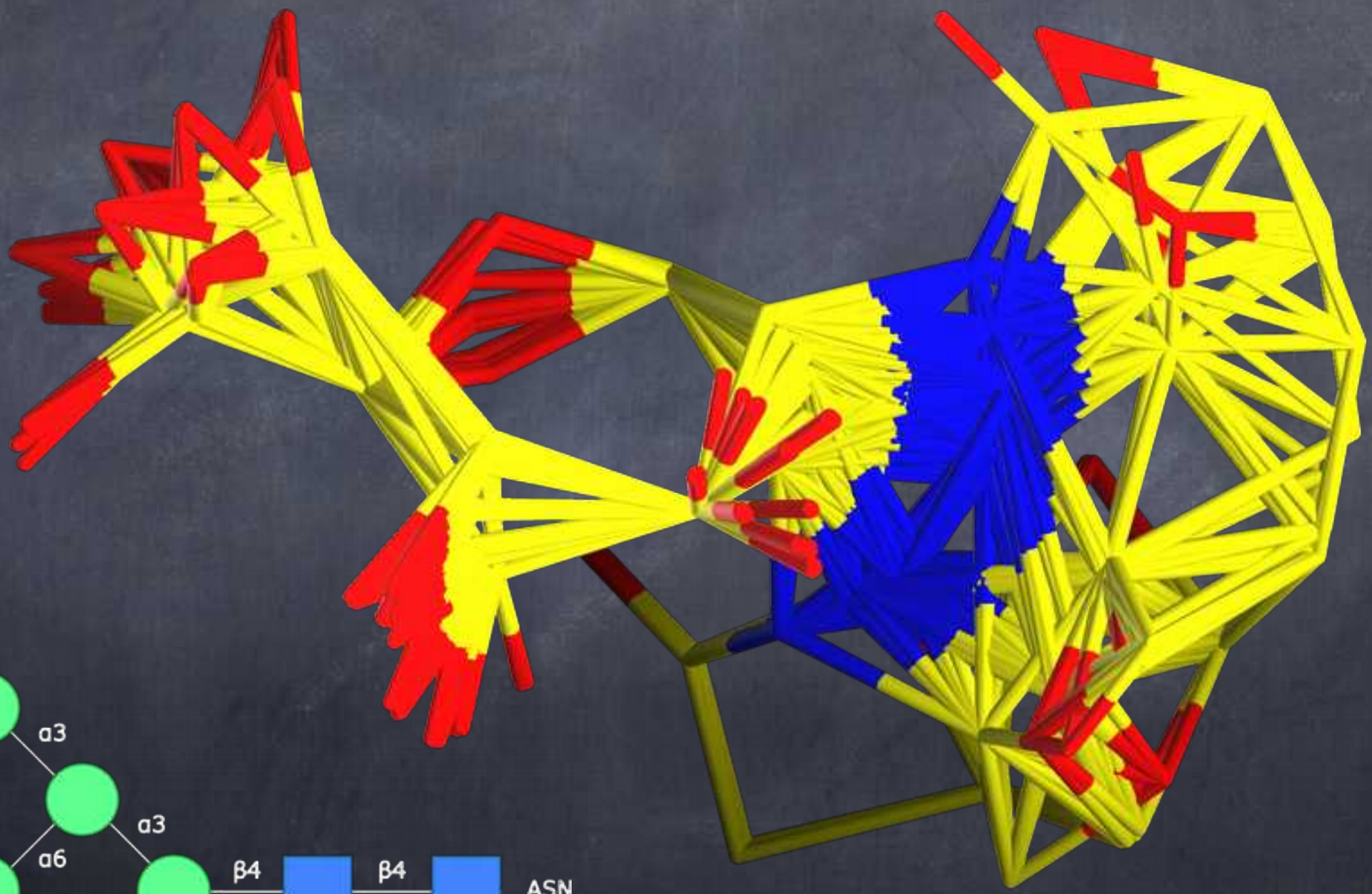
Furanoses (Q,Φ)

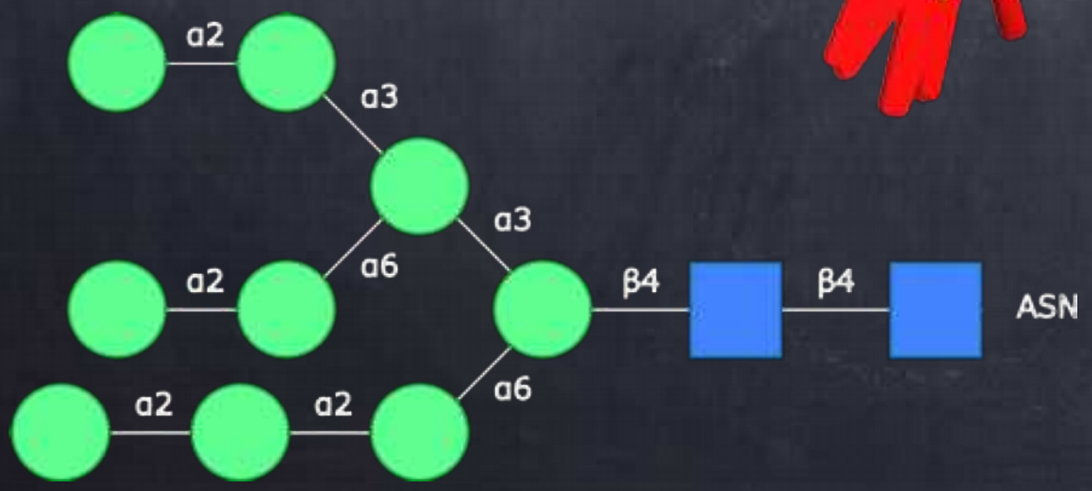
Agirre and Cowtan, Computational Crystallography Newsletter Jan:10-12 (2015)

To make your life easier - and ours too - we have developed a software tool for carbohydrate structure validation called Privateer

Initial attempt

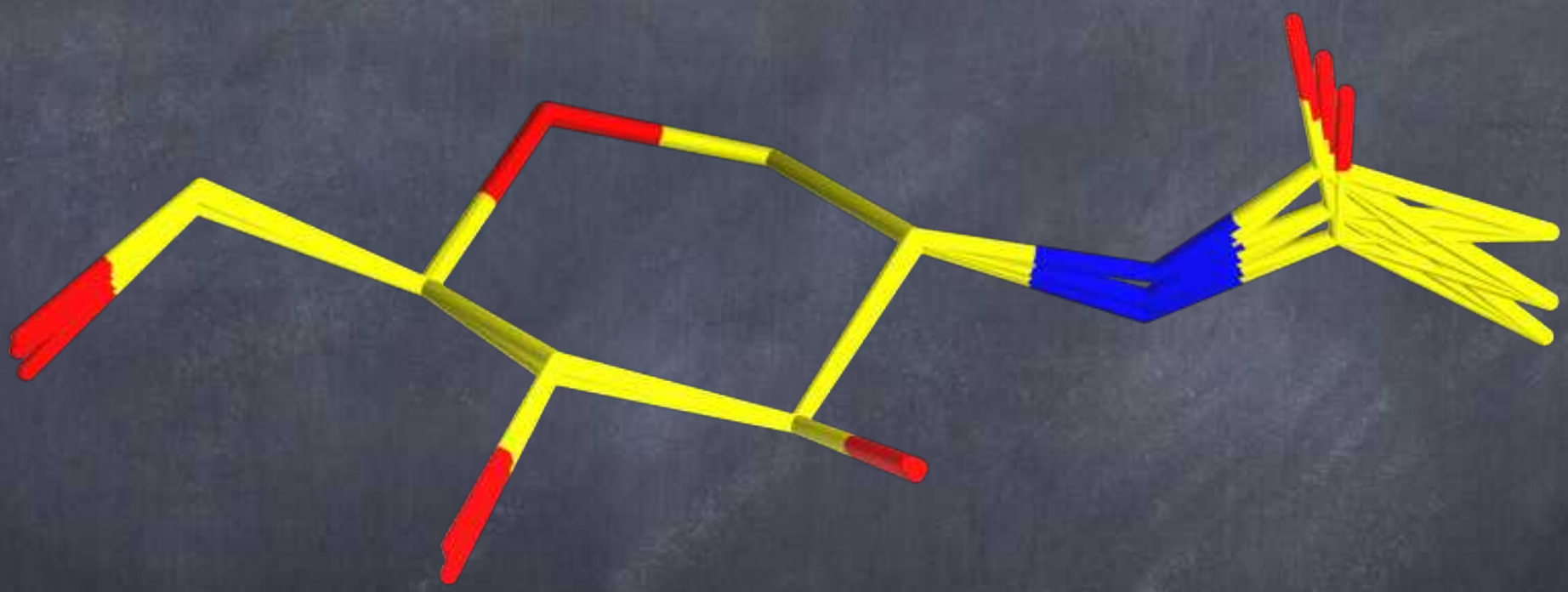
all GlcNac’s in one glycoprotein

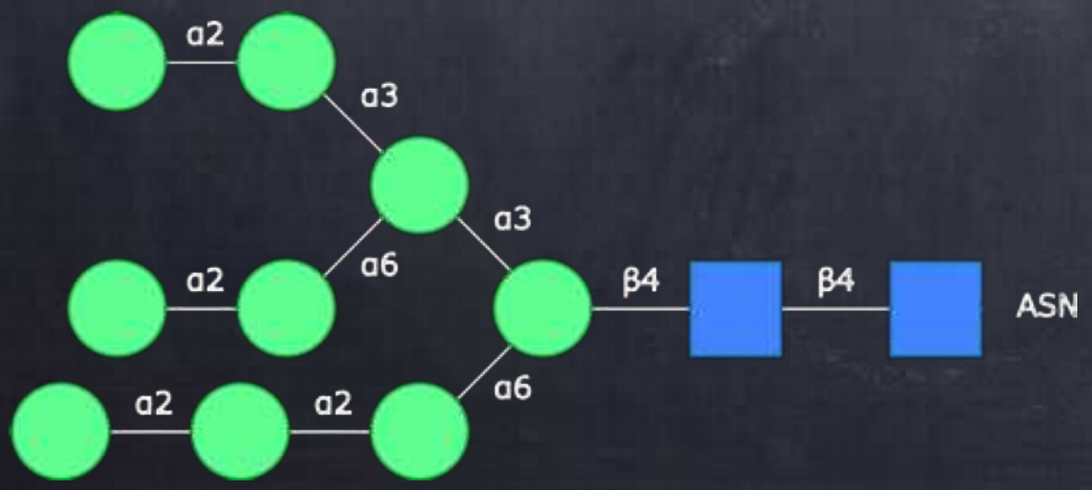




Initial attempt

all GlcNac’s in one glycoprotein



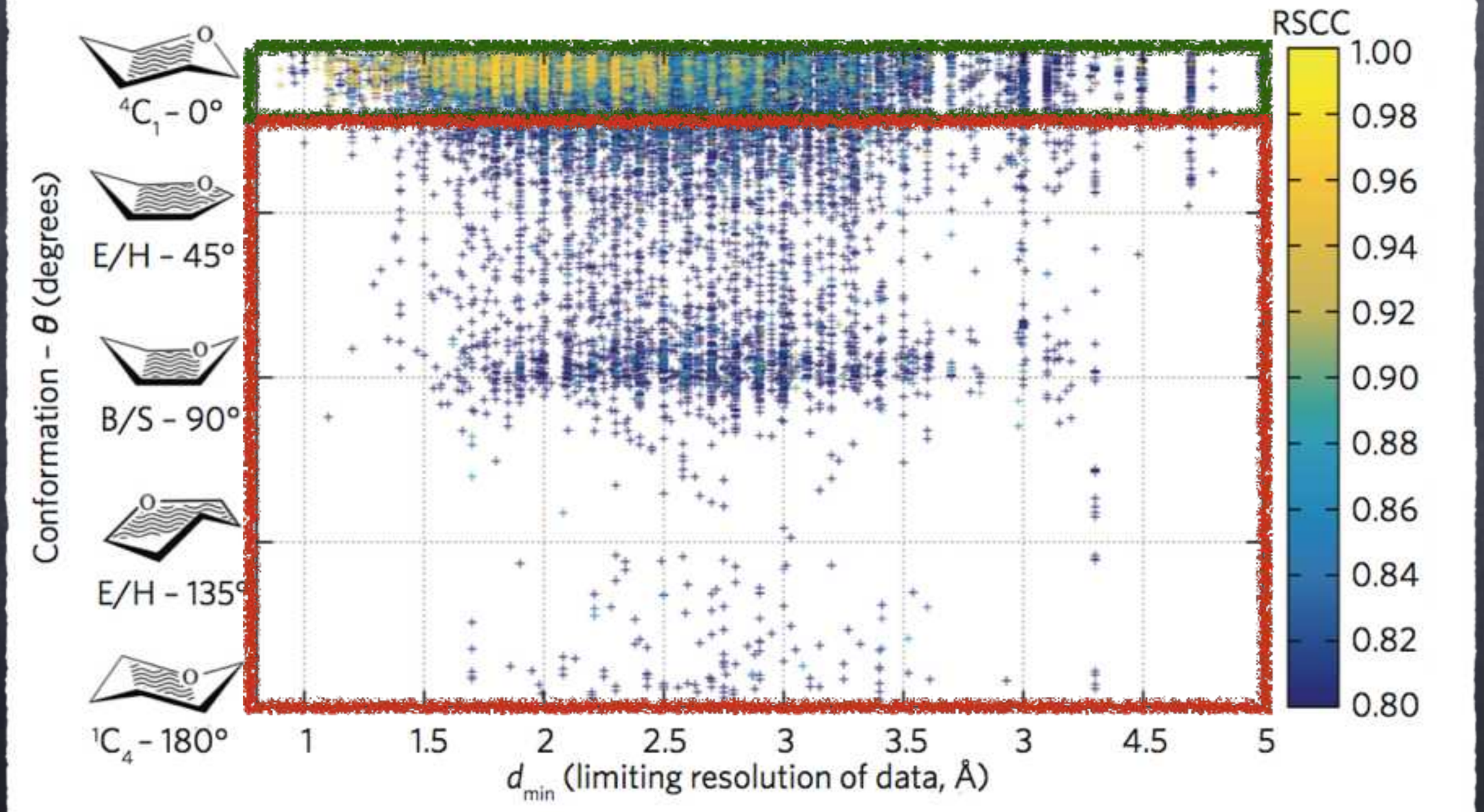


Privateer has filtered out the wrong structures

Have we chosen the worst structure in the whole PDB?

Let’s use Privateer on all N-glycan sugars in the databank...

Nope



this is a specialized event; of all the sugars deposited in the PDB, 65% sit undisturbed on N-glycosylation trees.

Using the Privateer[4] software, we computed the real-space correlation coefficient (RSCC) against positive omit electron density (as a bias-minimized fit to observed density)

higher-energy conformations start to appear. Most of these models also show low RSCC (<0.8; blue entries).

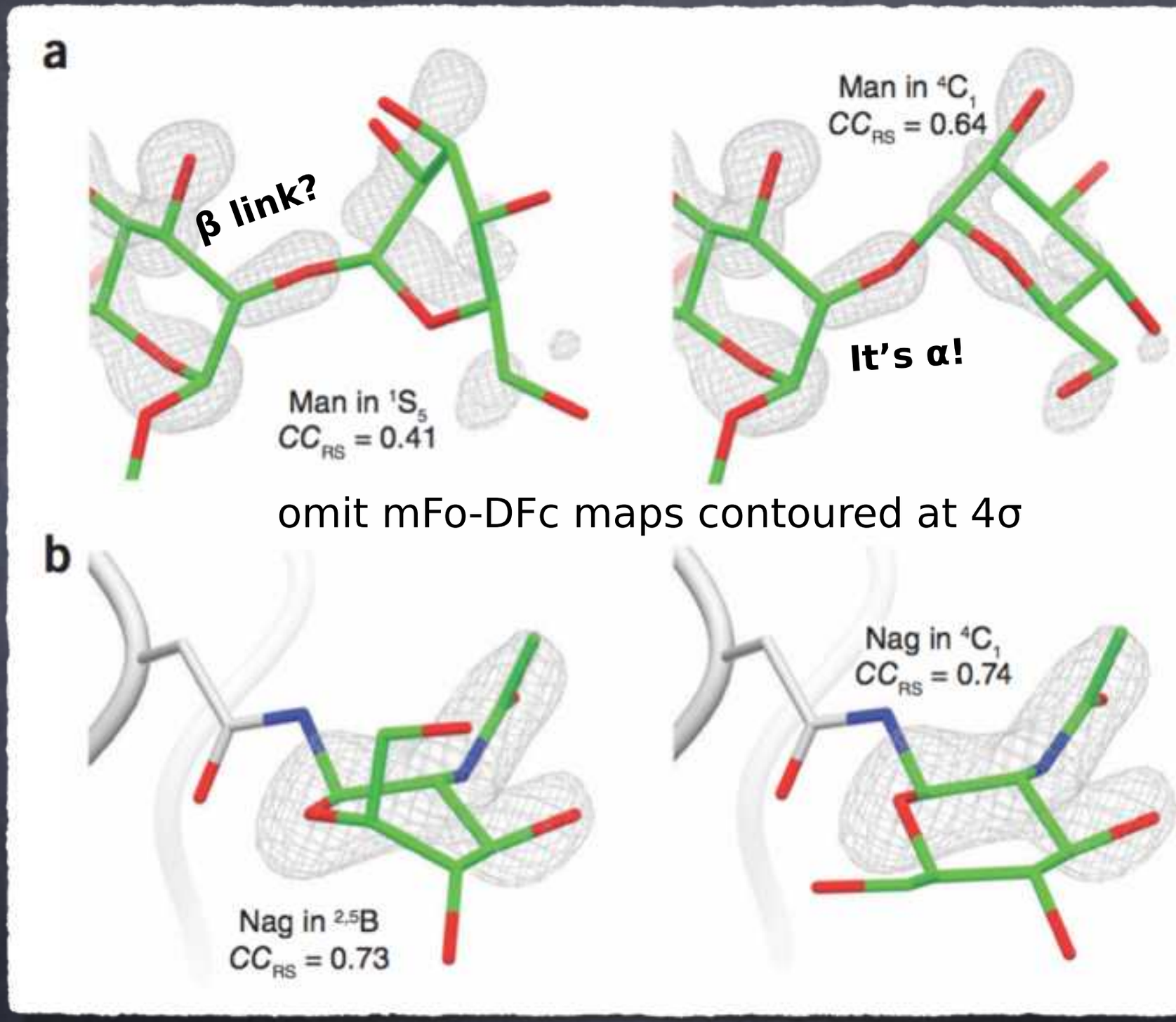
Although energetically unfavorable models may reflect a poor knowledge of glycochemistry and 'optimistic' density interpretation (reflected in low RSCCs)

got left behind. The fundamental roles of carbohydrates in cell biology and medicine, the extraordinary experimental advances in carbohydrate synthesis and the large increase in eukaryotic expression systems now demand improved refinement protocols for these key biological species, too.

2015)

We can now use Privateer to detect and prevent these and other anomalies

Built wrong



omit mFo-DFc maps contoured at 4σ

Built OK! but…

Agirre, Iglesias-Fernandez, Rovira, Davies, Wilson and Cowtan, Nature Structural and Molecular Biology, 22(11): 1-2 (2015)

Geometric restraints

- With weak/incomplete density, conformation should be imposed

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

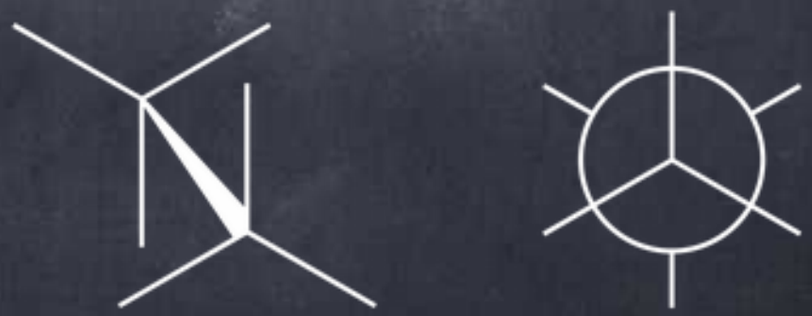
bond lengths ✓

bond angles ✓

harmonic torsions ?

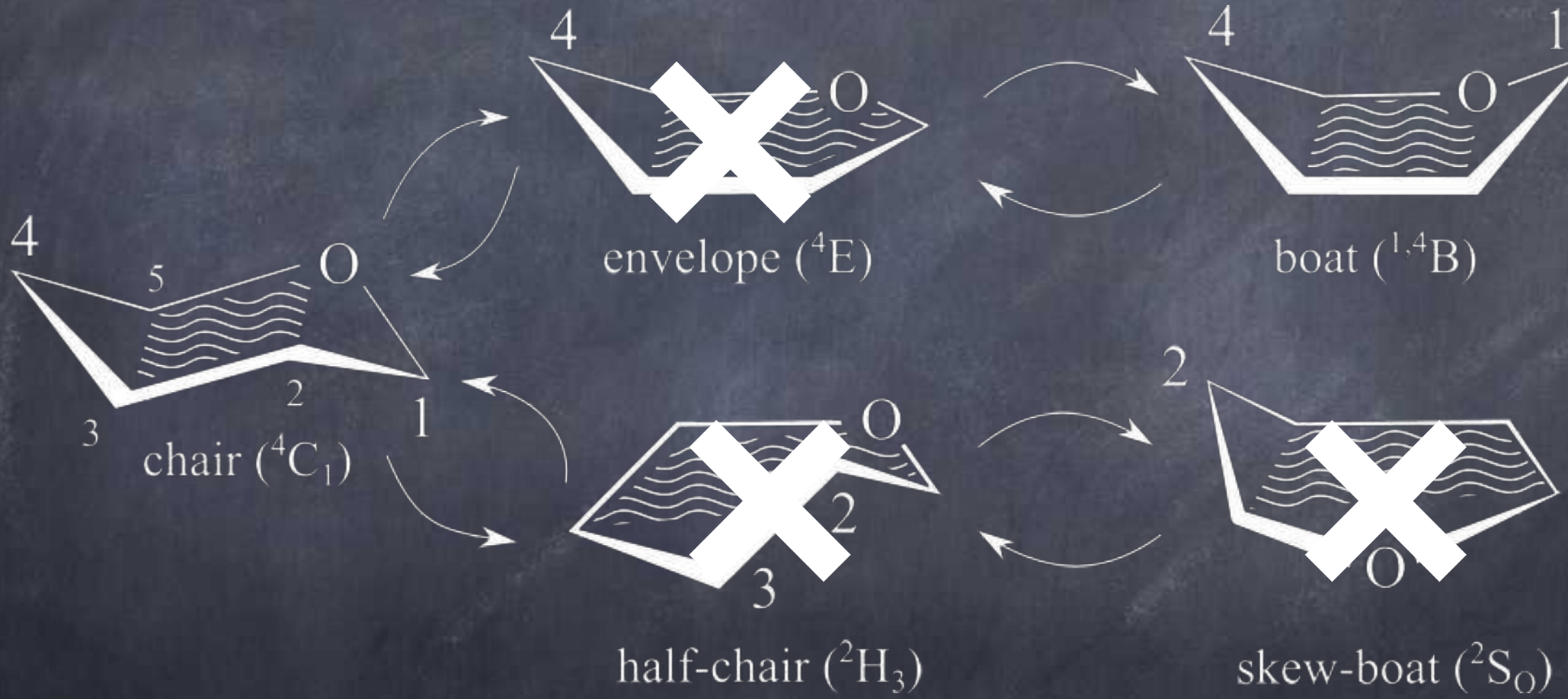
NCS, etc ?

Eclipsed substituents



Staggered substituents

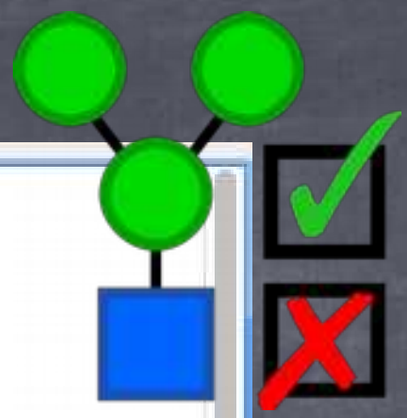
Eclipsed VS staggered



Monoperiodic torsions

- Dictionaries contain idealised coordinates
- These show the minimal energy conformer
- We can derive monoperiodic torsions from that
  - That’s what Privateer does from CCP4 7.0
    - Uses Eugene Krissinel’s CCP4srs & MMDB

How to validate and correct structures with Privateer, Coot and Refmac5



- Import merged data, sequences, alignments or coordinates
- Integrate X-ray images
- X-ray data reduction and analysis
- Experimental phasing
- Bioinformatics and model preparation for Molecular Replacement
- Molecular Replacement
- Model building and Graphics
- Refinement

Refinement - REFMAC5
Refine (Refmac5) with optional restraints (Prosmart)

Import and/or edit TLS set definitions
Enter TLS information to be used later in the project

Rigid body refinement - PHASER
Define rigid bodies for refinement (Phaser), fill partial residues (Coot) and refine (Refmac)

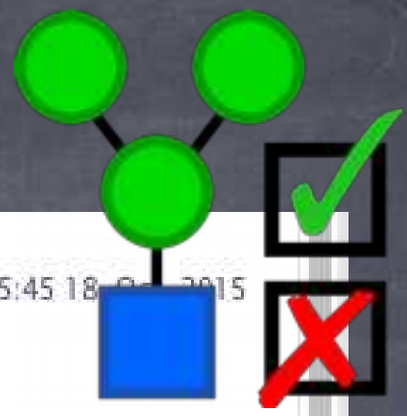
- Ligands

Make Ligand
Generate a PDB file and dictionary (acedrg) from MOL file, SMILES, or sketch (lidia)

Automated solution of isomorphous ligand complex
A ligand workflow, starting from merged or unmerged reflections, SMILES, and an isomorphous parent struc

- Validation and analysis

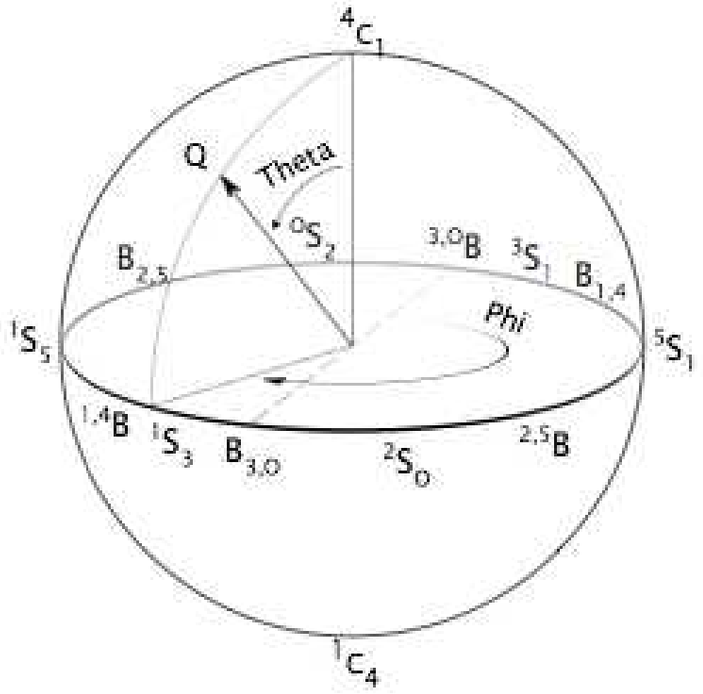
Analyse agreement between model and density - EDSTATS
Calculates real-space metrics for evaluating the agreement between model and density (Edstats.pdf)

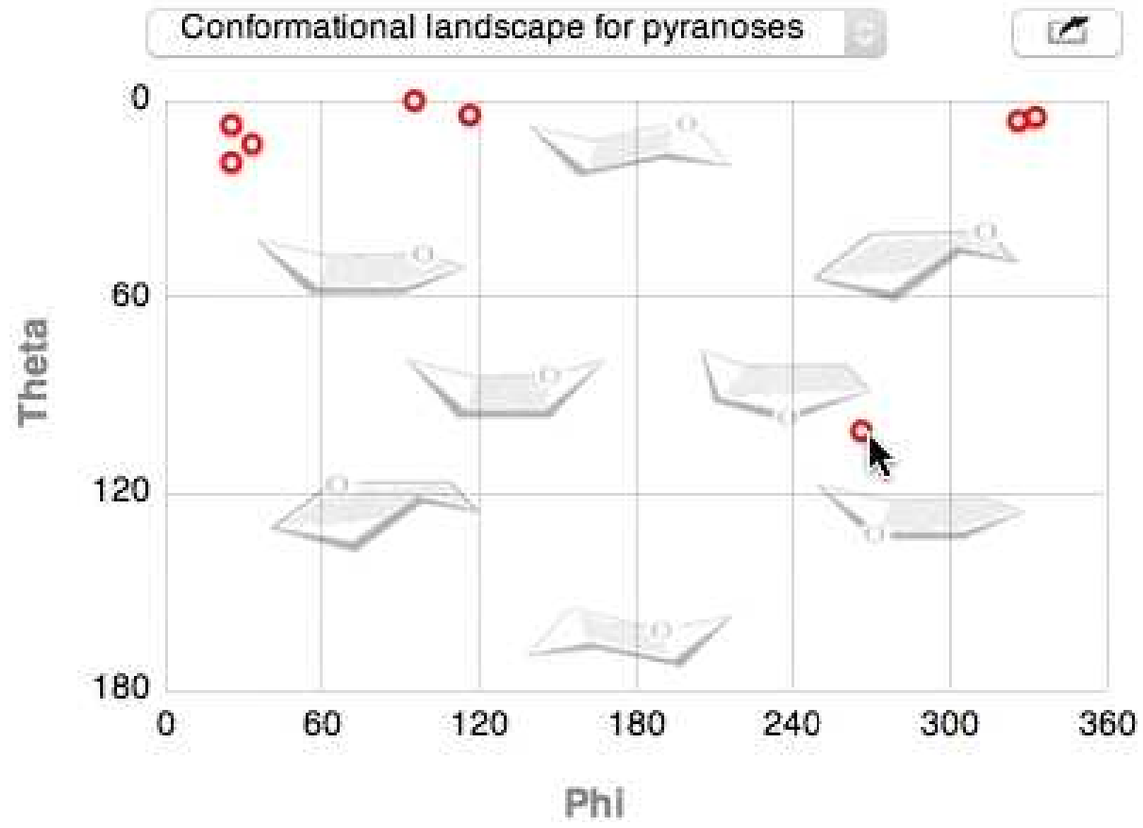


15:45 18 [illegible] 2015

Results

The Cremer-Pople analysis (Cremer and Pople, 1975, JACS 97:1354-58) is used to determine sugar ring conformation. Below is a 2D plot of the conformational parameters (Q, Phi, Theta for pyranoses; Q and Theta for furanoses) along with a depiction of the conformational sphere for pyranoses:



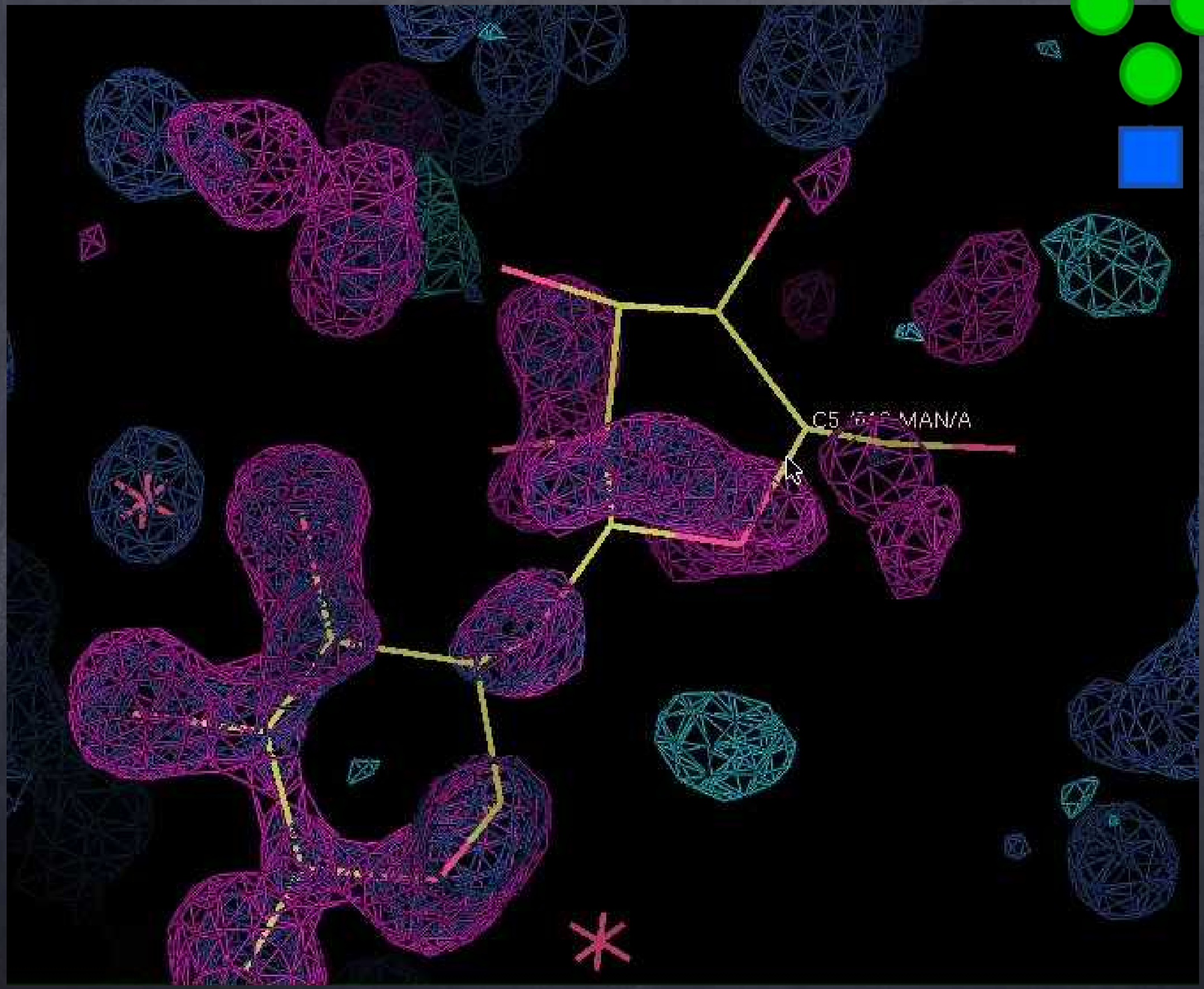


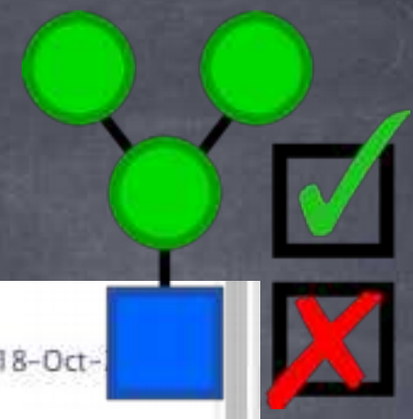
▸ N- and O-glycan structure 2D descriptions

▾ Detailed monosaccharide validation data

Validation results for pyranose sugars:

Manual model building - COOT

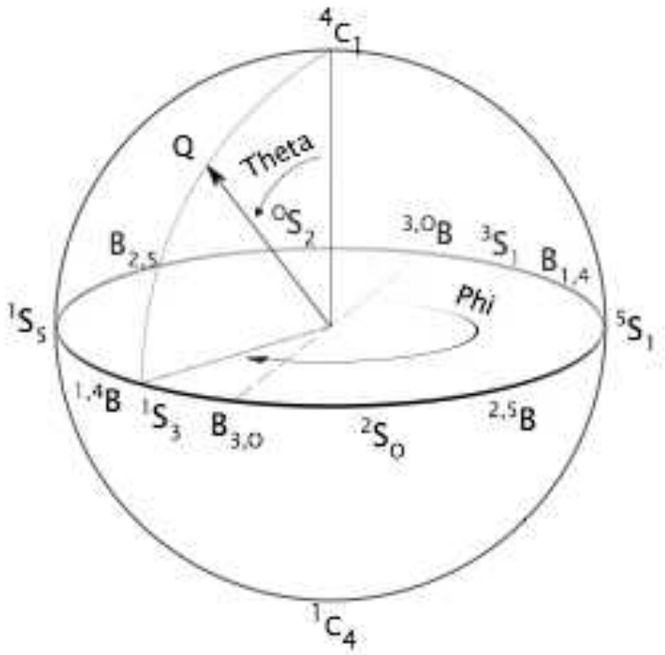

C5 MAN/A

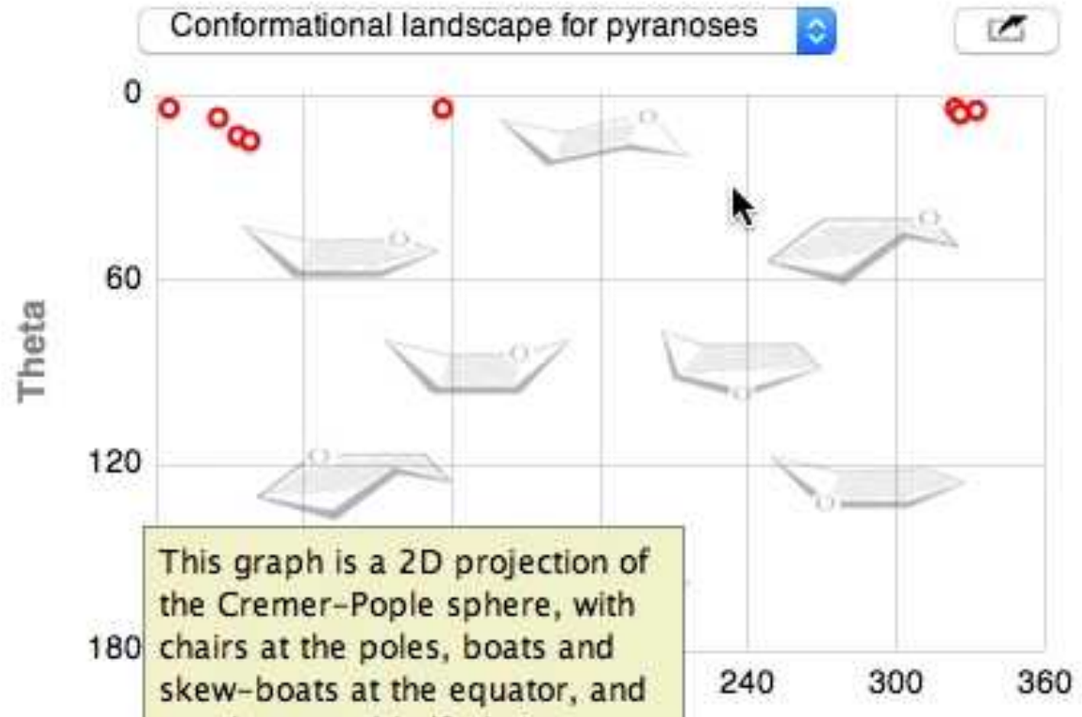


16:19 18-Oct-

Results

The Cremer-Pople analysis (Cremer and Pople, 1975, JACS 97:1354-58) is used to determine sugar ring conformation. Below is a 2D plot of the conformational parameters (Q, Phi, Theta for pyranoses; Q and Theta for furanoses) along with a depiction of the conformational sphere for pyranoses:





This graph is a 2D projection of the Cremer-Pople sphere, with chairs at the poles, boats and skew-boats at the equator, and envelopes and half-chairs elsewhere.

► N- and O-glycan structure 2D descriptions

▼ Detailed monosaccharide validation data

Validation results for pyranose sugars:

Manual model building - COOT

Things can go very wrong if you don’t know your chemistry!

ARTICLE
PUBLISHED ONLINE: 10 APRIL 2011 | DOI: 10.1038/NCHEMBIO.556

nature chemical biology

Interception of teicoplanin oxidation intermediates yields new antimicrobial scaffolds

Yu-Chen Liu[1-3], Yi-Shan Li[1], Syue-Yi Lyu[1], Li-Jen Hsu[1], Yu-Hou Chen[1], Yu-Ting Huang[1], Hsiu-Chien Chan[1], Chuen-Jiuan Huang[1], Gan-Hong Chen[1],

In the search for new efficacious antibiotics, alterations to antibiotic structures that may ove oxidation of a vancomycin-like glycopeptide to insights into residues that govern flavinylati the serendipitous discovery of a reaction inte to intercept the normal enzyme mechanism at synthesized families of antibiotic analogs with efficacy against multidrug resistant pathogens combat antibacterial resistance.

The rise of bacterial resistance against long-st last resort, such as vancomycin, has created a

NATURE CHEMICAL BIOLOGY CORRECTIONS

ADDENDUM

Interception of teicoplanin oxidation intermediates yields new antimicrobial scaffolds

Yu-Chen Liu, Yi-Shan Li, Syue-Yi Lyu, Li-Jen Hsu, Yu-Hou Chen, Yu-Ting Huang, Hsiu-Chien Chan, Chuen-Jiuan Huang, Gan-Hong Chen, Chia-Cheng Chou, Ming-Daw Tsai & Tsung-Lin Li

Nat. Chem. Biol. 7, 304–309 (2011); published online 10 April 2011; addendum published after print 17 April 2015

We have determined that the original crystal structure in our paper (Protein Data Bank (PDB) code 2WDX) was not of sufficient quality to provide strong support for a diol intermediate, although the mechanistic conclusions of the paper, which were also based on mutational and biochemical analysis, still stand. At the urging of our colleagues, we have replaced this structure with a second structure (PDB code 4K3T), solved as follows: Dbv29 protein in the buffer of 20 mM Tris-HCl, pH 8.0, and 100 mM NaCl was concentrated to 10 mg ml^{-1}. Dbv29 ligand-free crystals were grown under crystallization conditions of 0.2 M diammonium hydrogen citrate and 17% (w/v) PEG3350, optimized through thousands of screening conditions using Mosquito HTS liquid handler (TTP Labtech Ltd, Hertfordshire, UK). Bigger integral crystals were singled out and soaked with different concentrations of teicoplanin versus various soaking time courses (10 s–24 h). Diffraction data of teicoplanin-soaked crystals were collected at beamlines 13B1, 13C1 and 15A1 of NSRRC (Taiwan) and beamlines 12B2 and 44XU of SPring-8 (Japan). Corresponding data sets were indexed and scaled by HKL2000 package (http://www.hkl-xray.com/hkl-package/). Soaking the protein crystals with 40.5 mM teicoplanin for 1.2 h was determined to be the best condition for the formation of ternary complexes. the coordinates of native Dbv29 (PDB code 2WDW) as the search PHENIX (http://www.phenix-online.org/). Ligands were identified by ram Coot (http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/), the occupancy of ligands. The best-complexed structure was obtained d by the Uppsala Electron Density Server (EDS; http://eds.bmc.uu.se/) b.org/twilight/). The detailed data collection and refinement statistics

editorial

Structures under scrutiny

Structural biology advances have democratized access to biomolecular snapshots, but high standards must be maintained to ensure their utility.

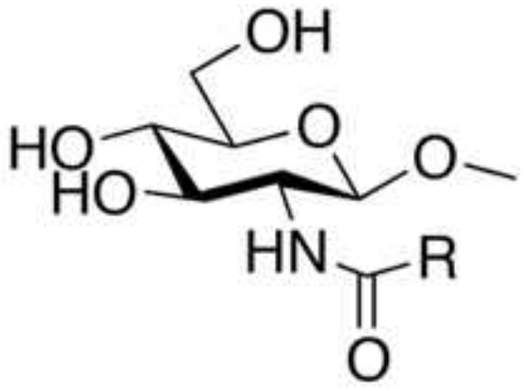
Protein crystallography has been pivotal in defining the landscape of modern chemical biology. Visualizing the detailed architecture of proteins, nucleic acids and other biomolecules has shaped our molecular understanding of life and inspired countless mechanistic investigations. Given the sometimes extensive engineering required to generate crystallizable constructs, as well

can lead to errors. For example, incorrect timing of ligand placement in a structural refinement can result in misleading models. Further, because many current refinement tools do not adequately handle certain post-translational modifications or small molecules, structural problems may be missed. Even when appropriate refinement and validation steps could be used, Cowtan

whether they should be required at *Nature Chemical Biology* as well.

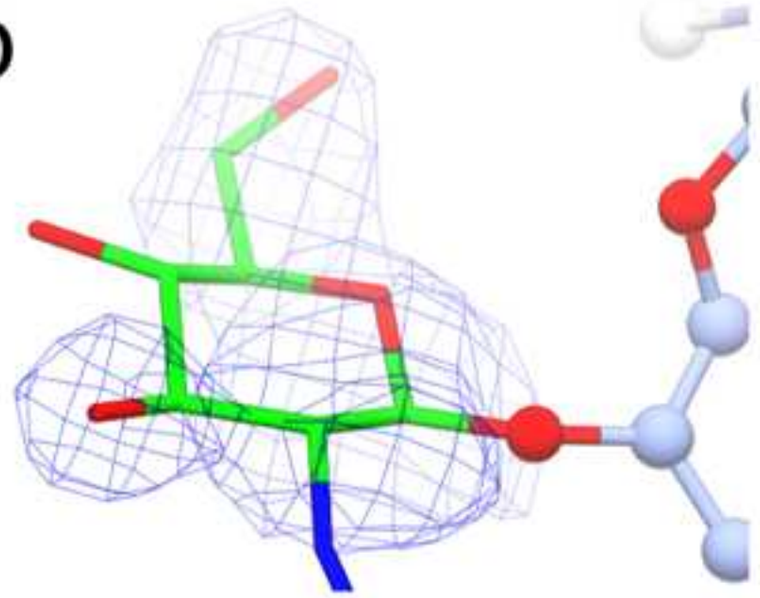
The community also has a role to play. As Brändén and Jones advised, "…the readership should be sophisticated enough to judge the quality of the data." Chemical biologists draw extensively on crystal structures to study enzyme active sites, protein-protein interactions, protein- or

a

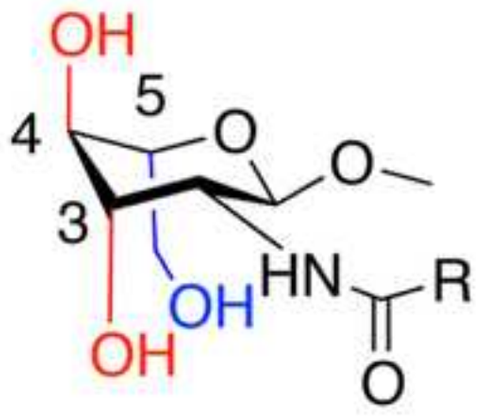


N acetyl β-D glucosamine derivative

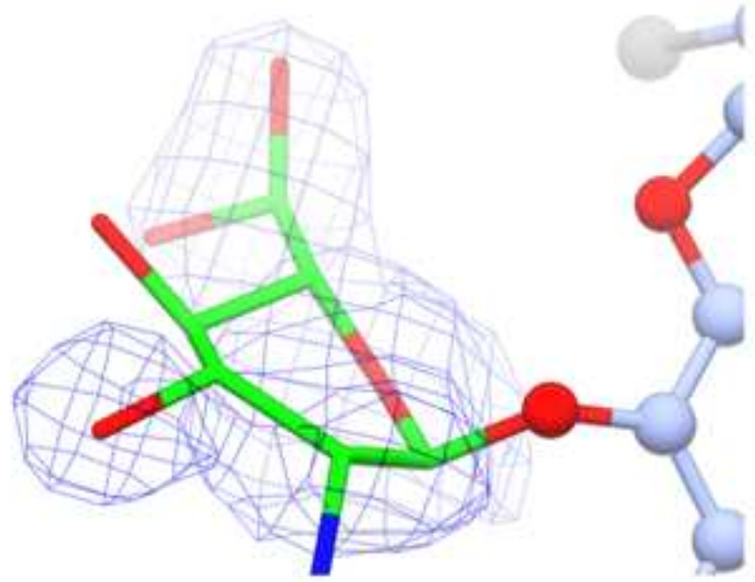
b



GCS (chain I), $^4C_1$ conformation



N acetyl α-L mannosamine derivative



TM9 (chain I), ^5E conformation

Conclusions

- Higher-energy conformations should be backed by clear density, otherwise treated as outliers
- With weak density, refinement software may distort sugars. Tighter geometry should keep them in place
- The already distorted ones may not refine to chairs without torsion restraints
- Linkages are not implied. Have a care how you link your sugars!

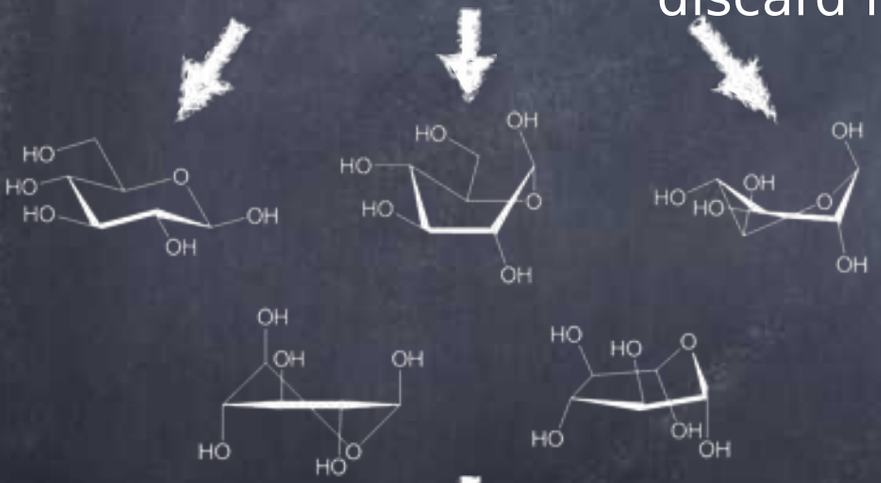
Motivation revisited

- We want to build software for detecting and modelling sugar moieties in electron density maps
- Method: matching a fingerprint across the map
- Fingerprint: a synthesis of superposed sugar moieties and their environment

Building a fingerprint

<u>input</u>
X-ray structures containing furanose and pyranose rings

<u>filter</u>
discard models with errors



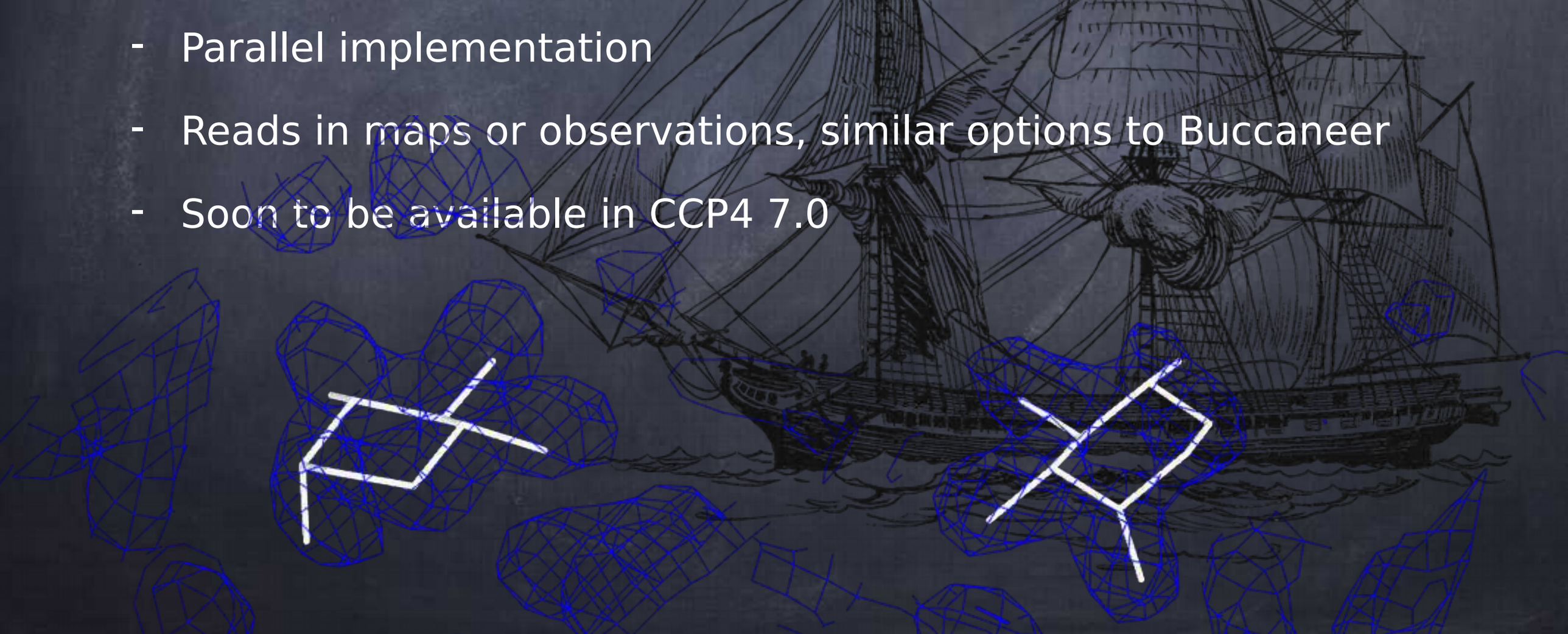
<u>classify</u> and <u>superpose</u>
by conformational analysis (Cremer-Pople)

compute <u>maps</u> and generate <u>fingerprints</u>

Features

Distributed by CCP4

- Detects N- and O-glycosylation trees, ligand sugars...
- Fast enough to be used interactively in Coot
- Parallel implementation
- Reads in maps or observations, similar options to Buccaneer
- Soon to be available in CCP4 7.0



Acknowledgements

UNIVERSITY of York

Kevin Cowtan
Jon Agirre*
Eleanor Dodson
Gideon Davies
Keith Wilson*
Stuart McNicholas*
Christian Roth
Wendy Offen
Saioa Urresti

NETHERLANDS CANCER INSTITUTE
ANTONI VAN LEEUWENHOEK

Robbie Joosten

JUSTUS-LIEBIG-UNIVERSITAT GIESSEN

Thomas Lütteke

iqtc
Institut de Química Teòrica i Computacional

Carme Rovira
Javier Iglesias-Fernandez

MRC | Laboratory of Molecular Biology

Paul Emsley*
Garib Murshudov*
Rob Nicholls*
Fei Long*

Well supported by:

BBSRC
bioscience for the future

CCP4

*** *Ligands* initiative**