Model-Building with *Coot*

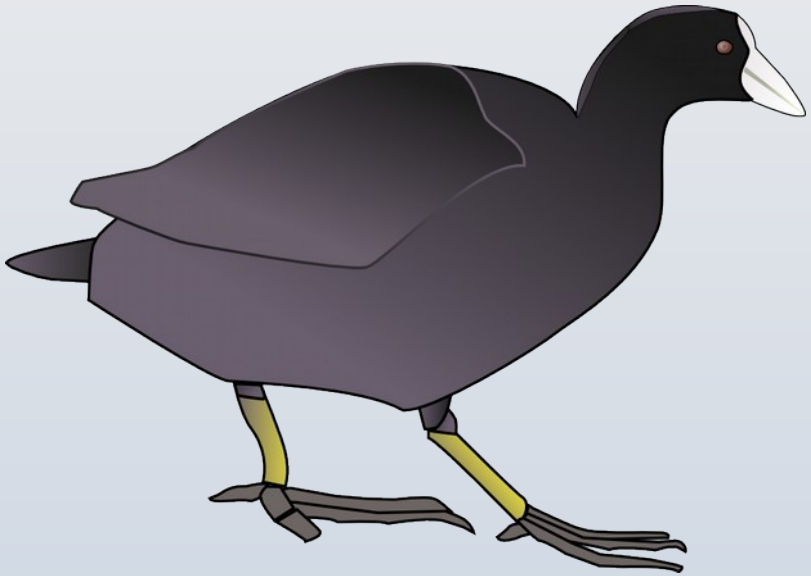
Adding N-linked Glycosylation
&
Demo

Paul Emsley
MRC Laboratory of Molecular Biology
May 2017

About this Presentation

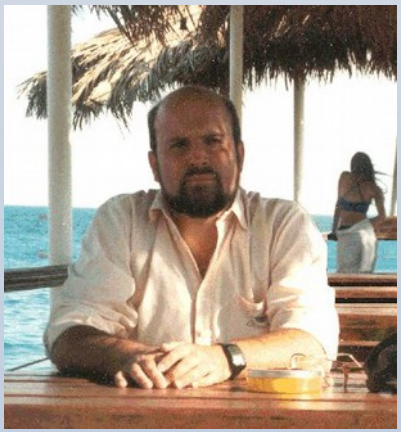
- Modelling N-linked Carbohydrate
- Demo: Carbohydrate Fitting
- Demo: Protein-fitting
  - [*Coot* & Key-bindings]
- When is the available?
  - Available from the *Coot* Web Site now
    - Get the latest 0.8.9-pre pre-release
  - much is in 0.8.8
  - And CCP4
- pdf

Coot Collaborators



Bernhard Lohkamp

Kevin Cowtan



Eugene Krissinel

Stuart McNicholas

Martin Noble

Alexei Vagin

Richardsons & co-workers, Duke

A Brief History of *Coot*

- Released in 2004, Coot was designed primarily for model-building protein models into maps from x-ray data
  - Torsions: Rotamers, Ramachandran plots
  - Several optimisers, including Real Space Refinement
- Used typically after automated model-building or refinement
- Since:
  - Nucleic Acids, Ligands & Cryo-EM
- It's never been pretty...

  - Not the best tool for presentation graphics and animations

Coot Key-bindings

- Many hundreds of functions available in *Coot*'s API
  - available via scheme or python
- *Coot*'s gui doesn't help much to learn key-bindings
  - they are "off" by default
  - so that you can program your own
- If you are more than a casual/occasional users of *Coot*... are probably worth learning

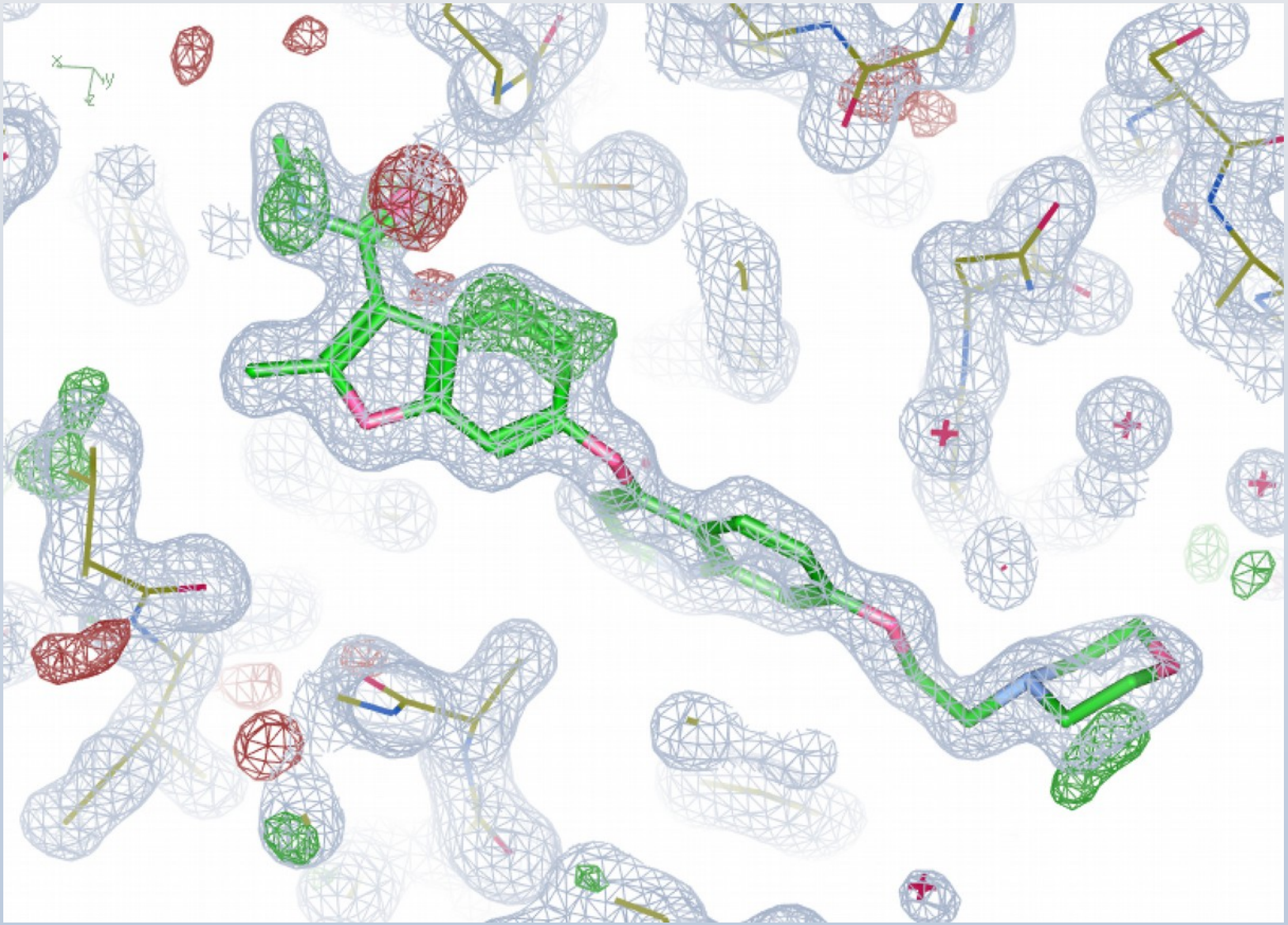
Using *Coot* on a Mac

- Coot is an X11-based application
- Xquartz ⟶ Preferences
  - Input
    - Emulate 3-button Mouse
  - Windows
    - Focus Follows Mouse

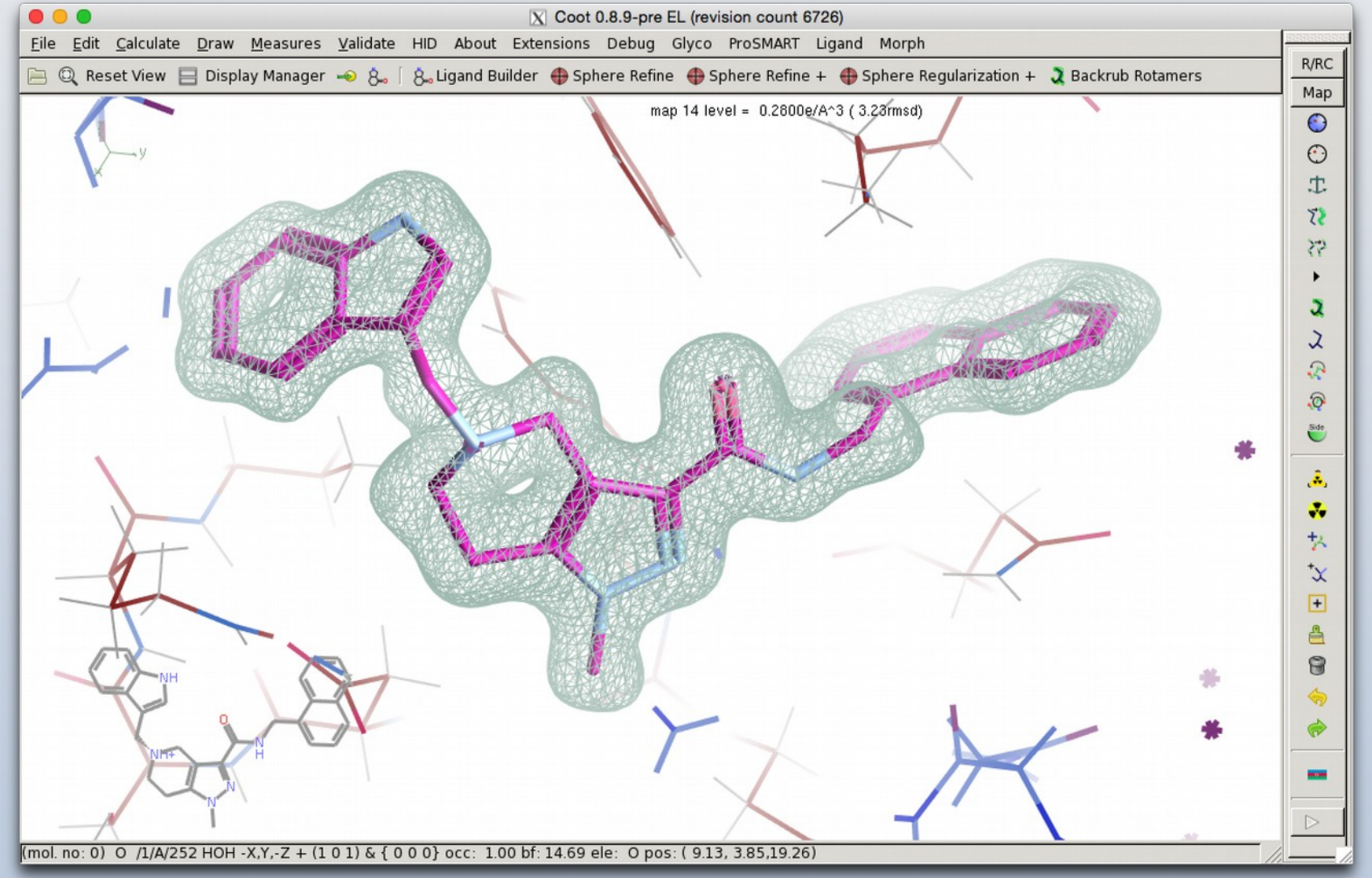
Making Density Slides with *Coot*

- White background
- “High” Oversampling (2.3x)
- Pale gray (or very pastel) density colour
- Enable Cut-glass mode 5-10%
- Licorice Bonds (Shift-4)
  - Highlight Interesting Site
- Anti-aliased Coot
  - `$ setenv __GL_FSAA_MODE 5`

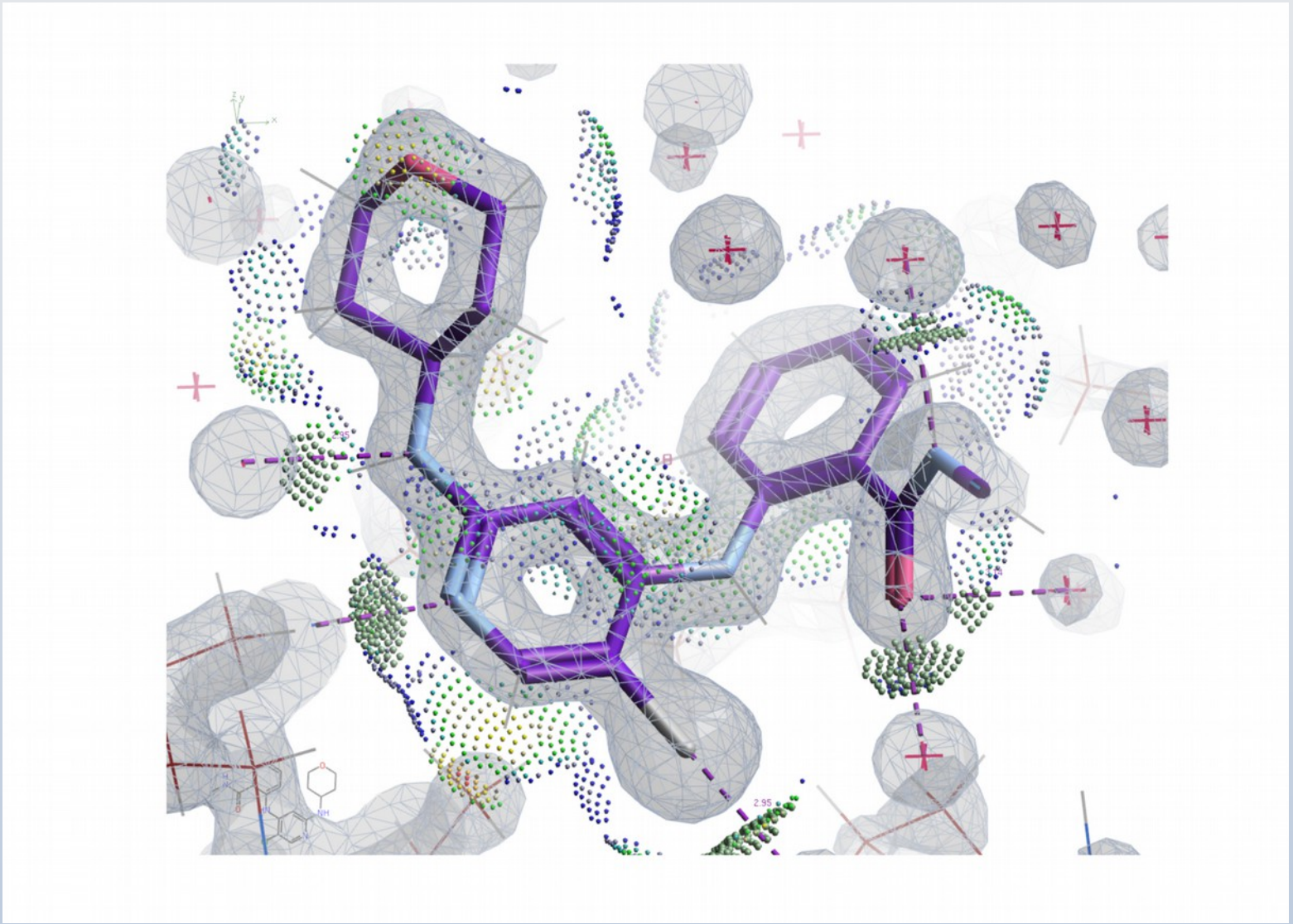
Example Density Slide



Ligands Representation in *Coot*



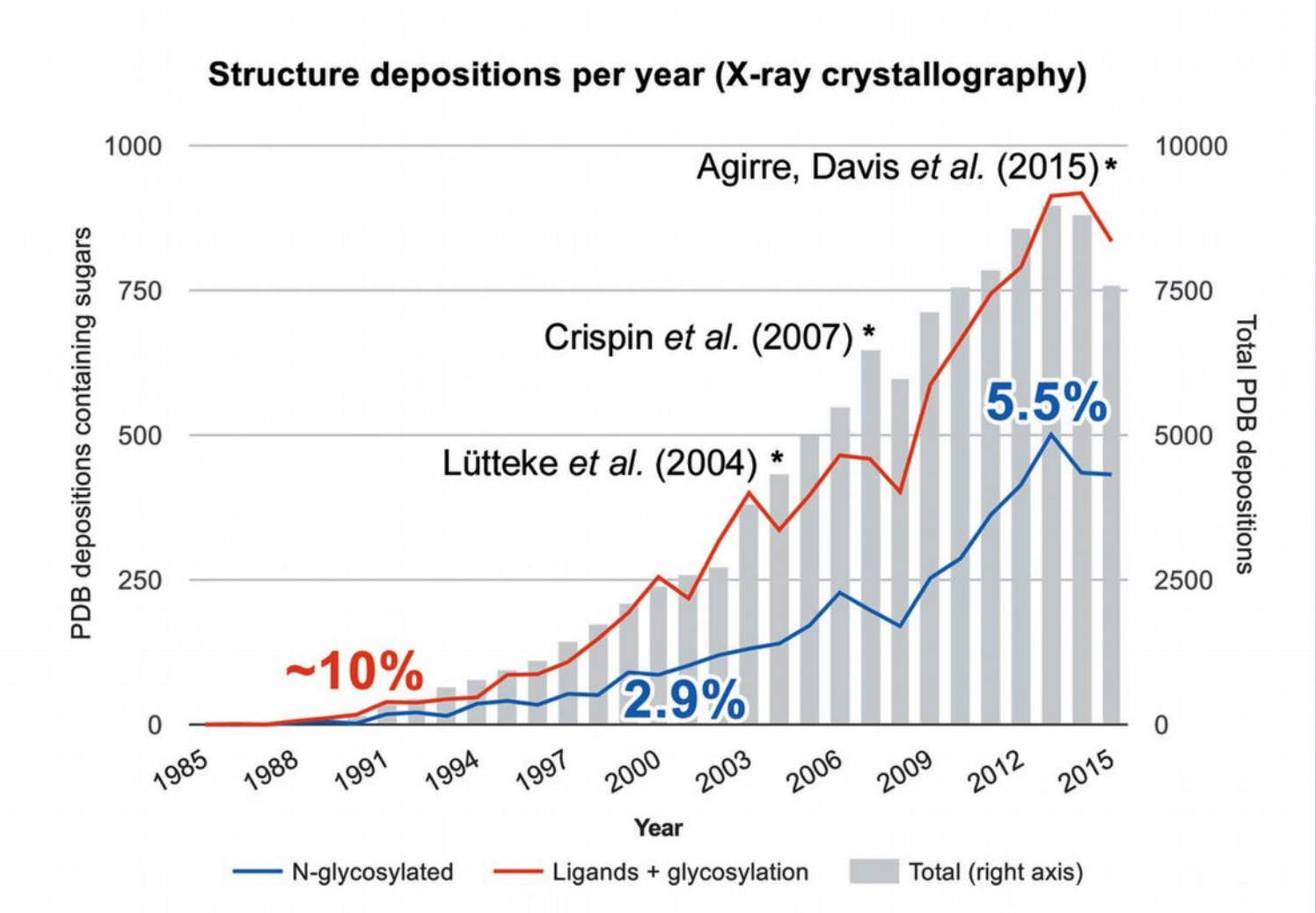
Ligand Representation in *Coot*



Glycoproteins

- Glycosylation is the most frequent and most complex protein modification
- Unlike protein, the structure is glycans is only indirectly encoded in the genome
  - (glycosyltransferases)
- Even in the same cell, a particular type of protein can be differently glycosylated
  - microheterogeneity
- Although N-linked glycans are relatively conserved
- Glycans stabilize the protein and shield it from attack by proteases
- Play a role in cell-cell signalling:
  - fertilization, differentiation, host-pathogen, immune response

Glycoproteins

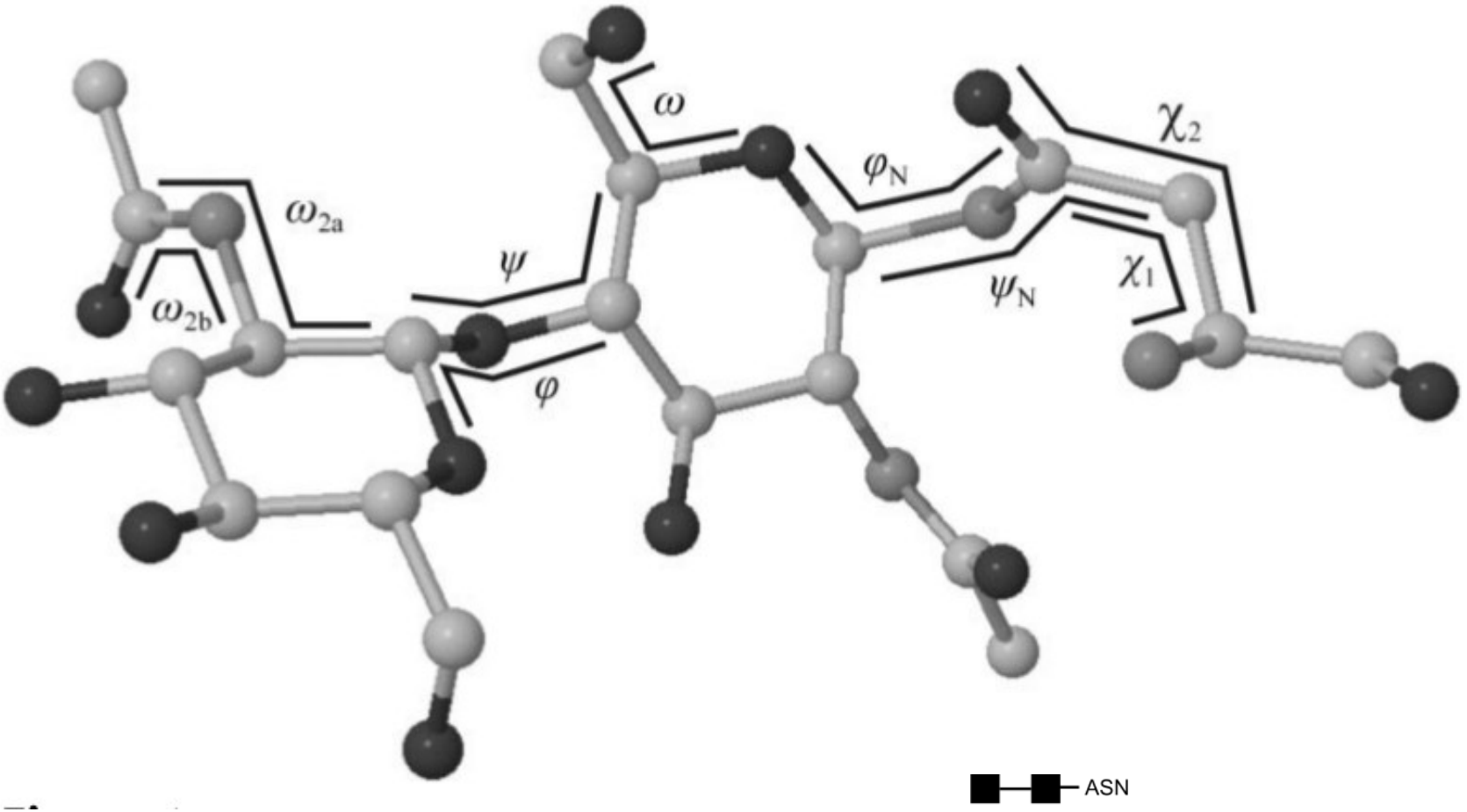


Agirre (2017)

Problematic Glycoproteins

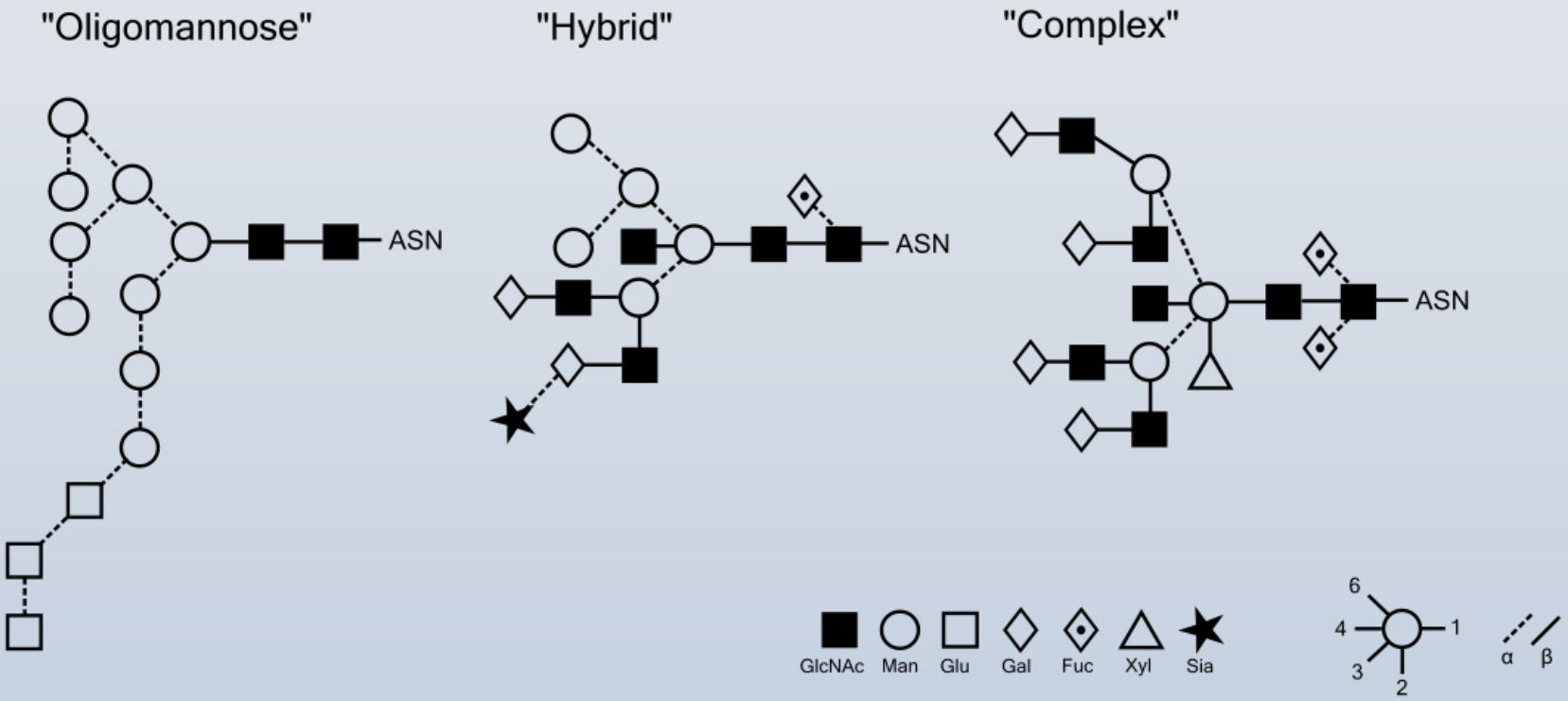
- Crispin, Stuart & Jones (2007)
  - NSB Correspondence
  - “one third of entries contain significant errors in carbohydrate stereochemistry…”
  - “carbohydrate-specific building and validation tools capable of guiding and construction of biologically relevant stereochemically accurate models should be integrated into popular crystallographic software. Rigorous treatment of the structural biology of glycosylation can only enhance the analysis of glycoproteins and our understanding of their function”
  - PDB curators concur
  - More recently Joosten & Lűtteke (2017), Agirre *et al.* (2017)

Carbohydrate Links



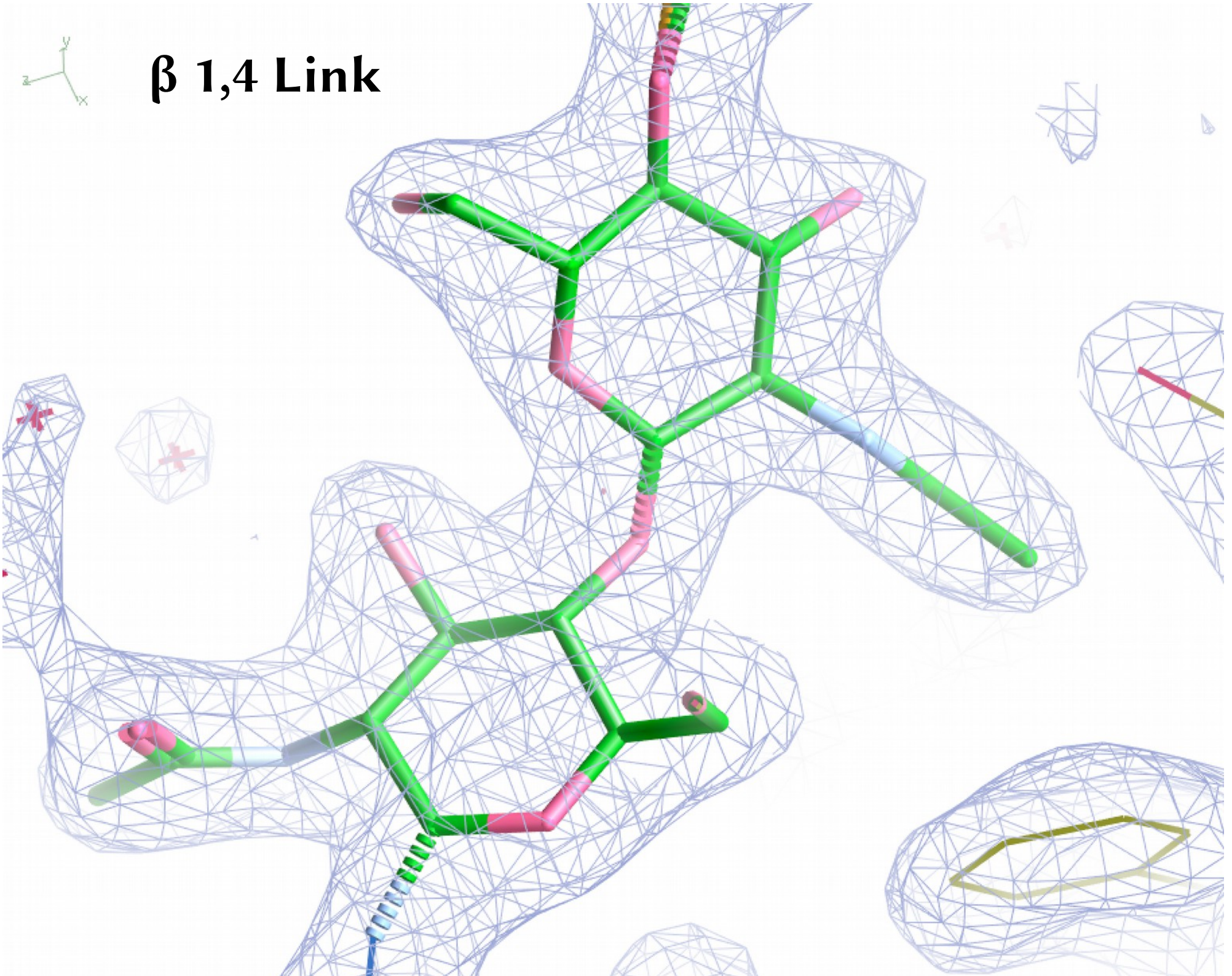
Thomas Lütteke (2007)

N-linked carbohydrates

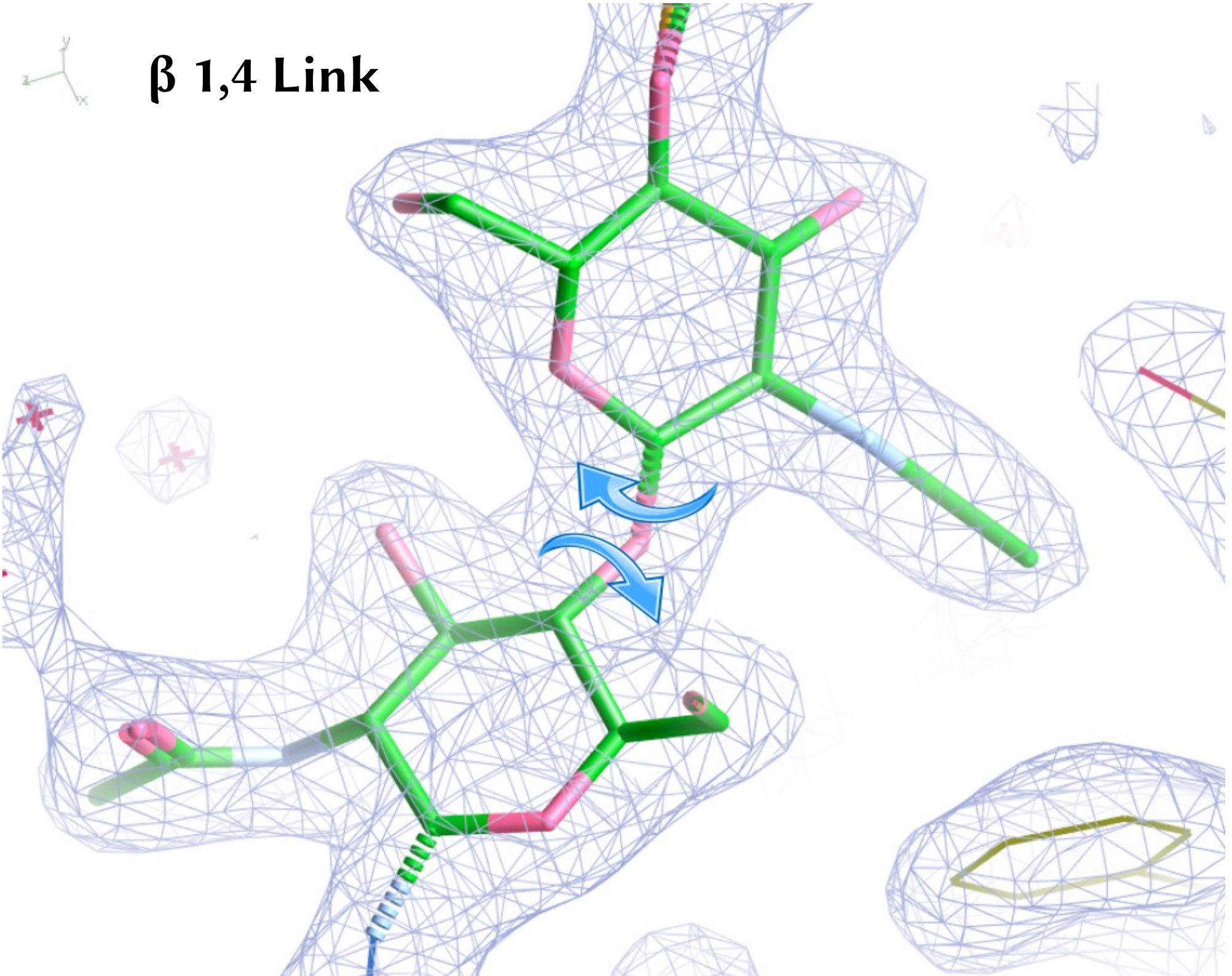


Linking Oligosaccharides/Carbohydrates: LO/Carb

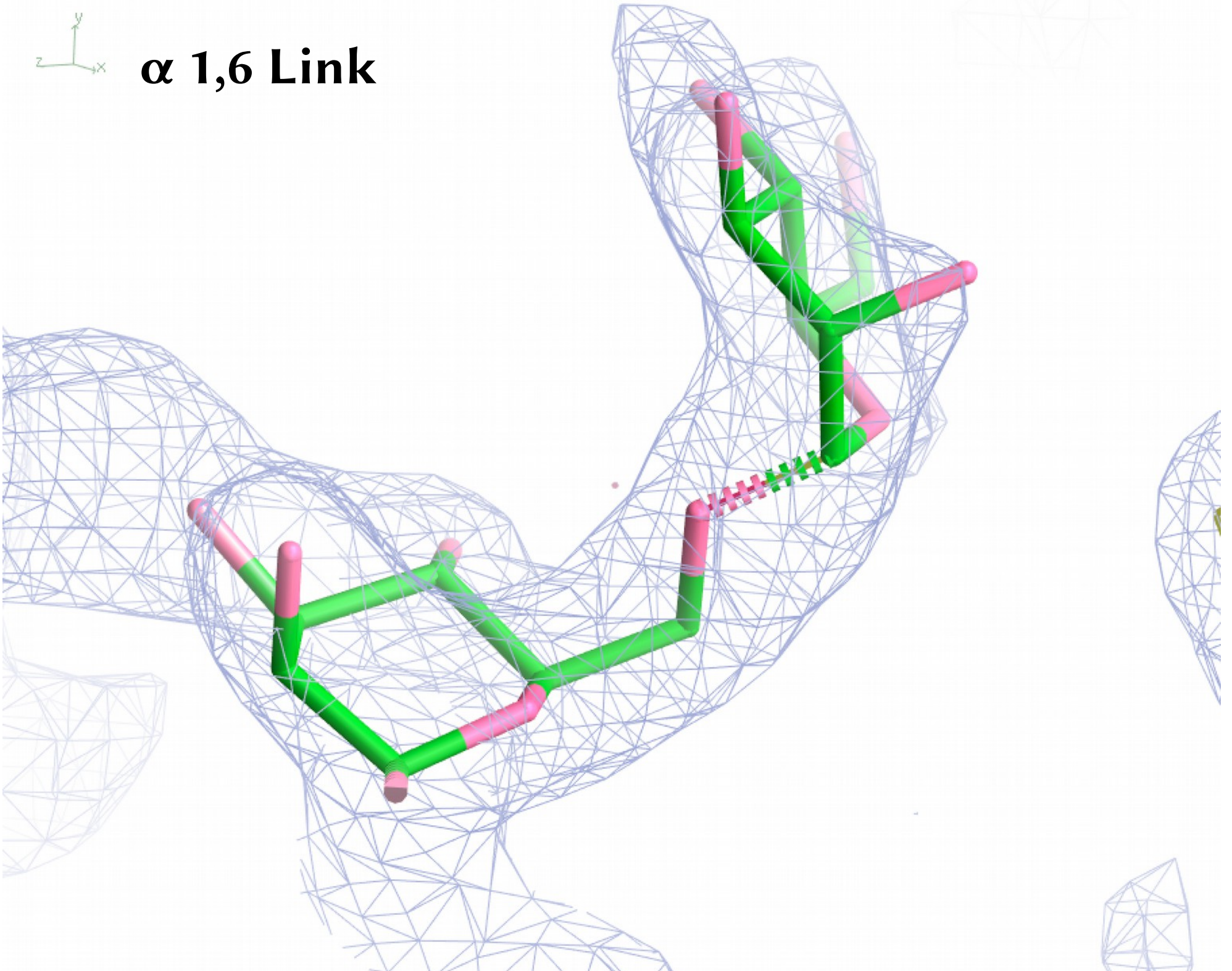
- One can fully define carbohydrate structure by the primary structure and a set of torsion angles
- Build complex carbohydrate structure
  - from a dictionary of standard links
  - and monomers
  - torsion-angle refinement
    - by simulated annealing


β 1,4 Link

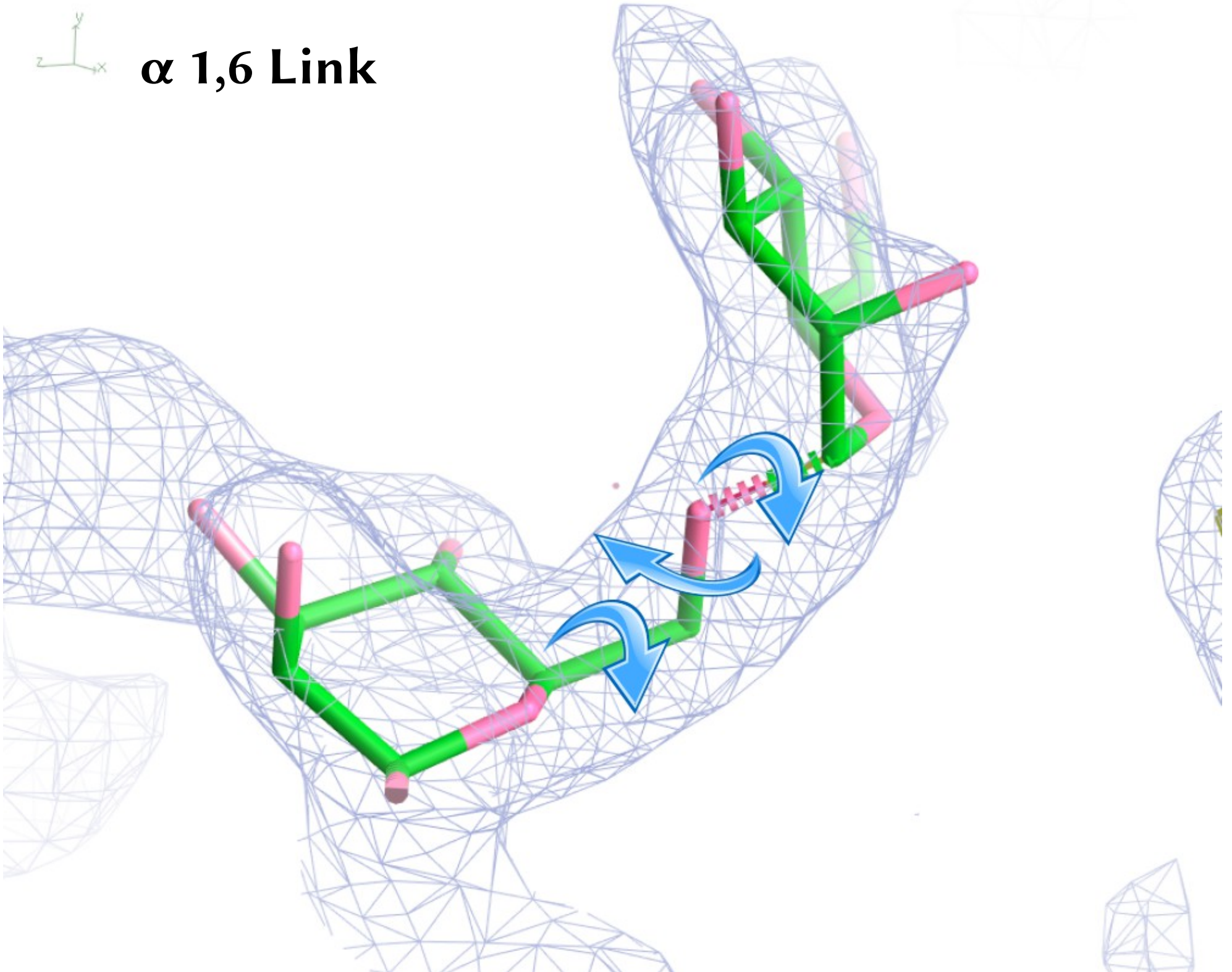
β 1,4 Link



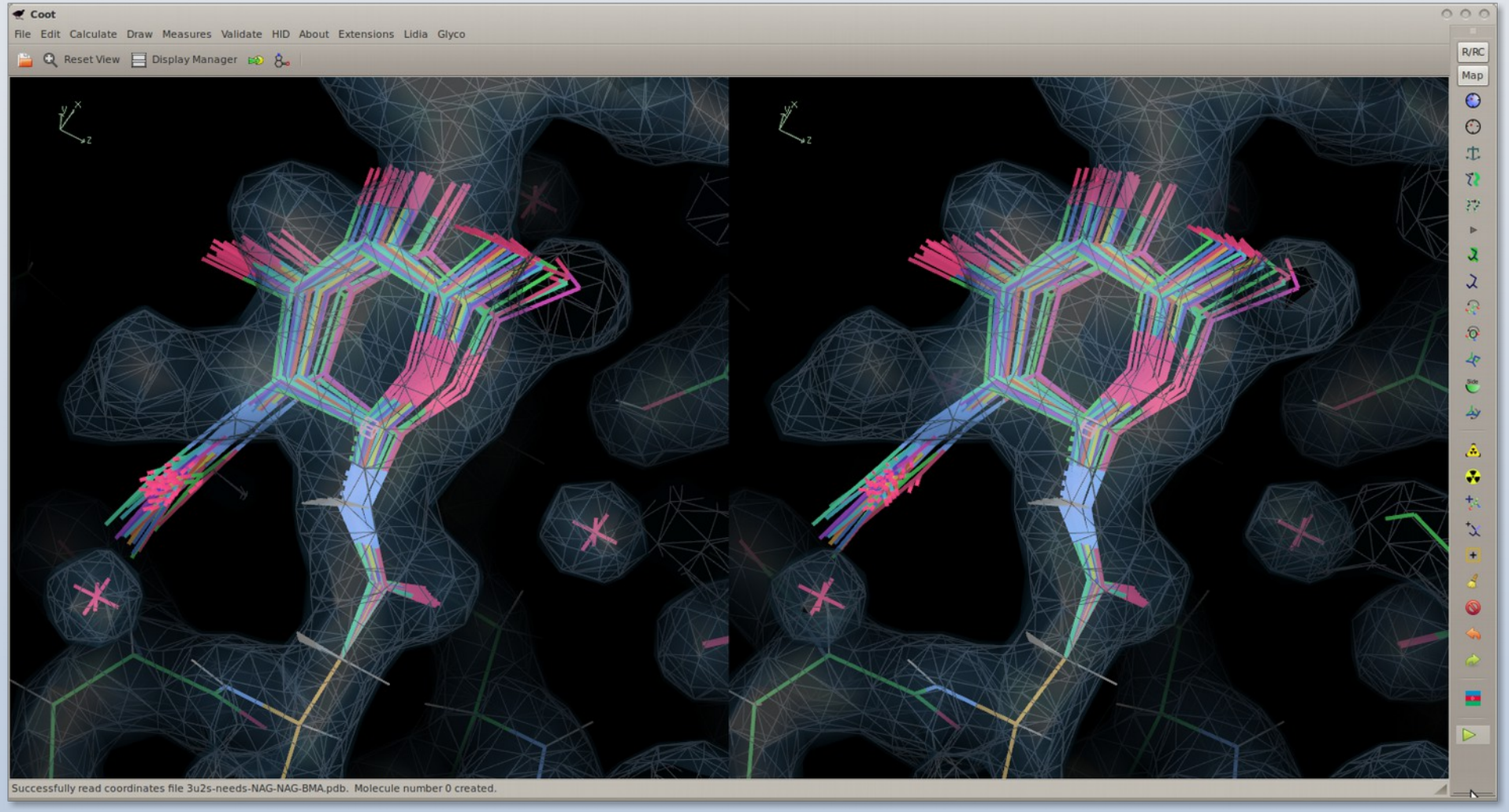
α 1,6 Link



α 1,6 Link

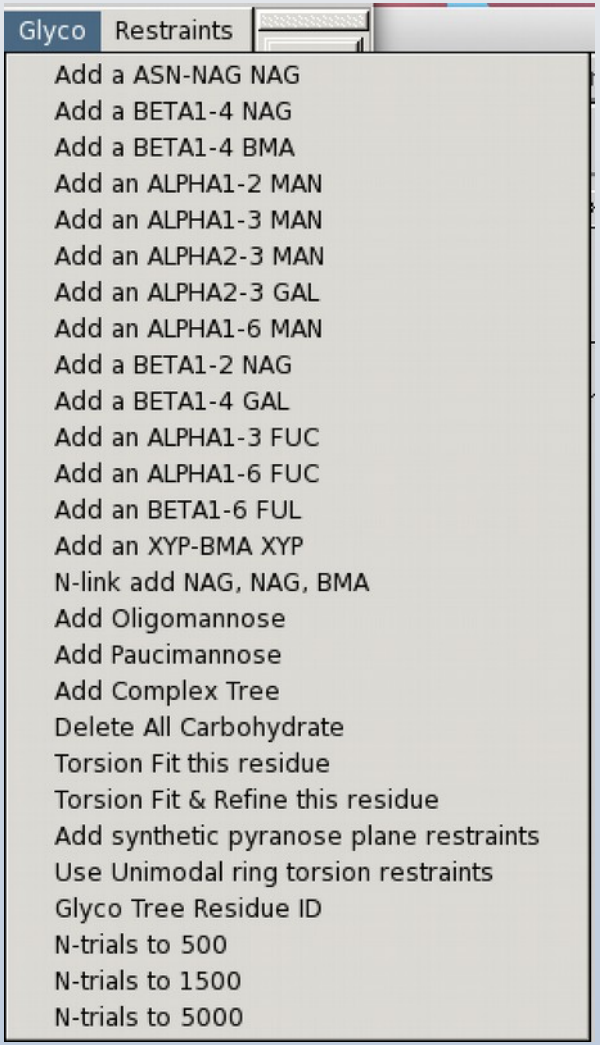


Refinement Progress (NAG-ASN example)



N-linked Glycan Modelling in *Coot*

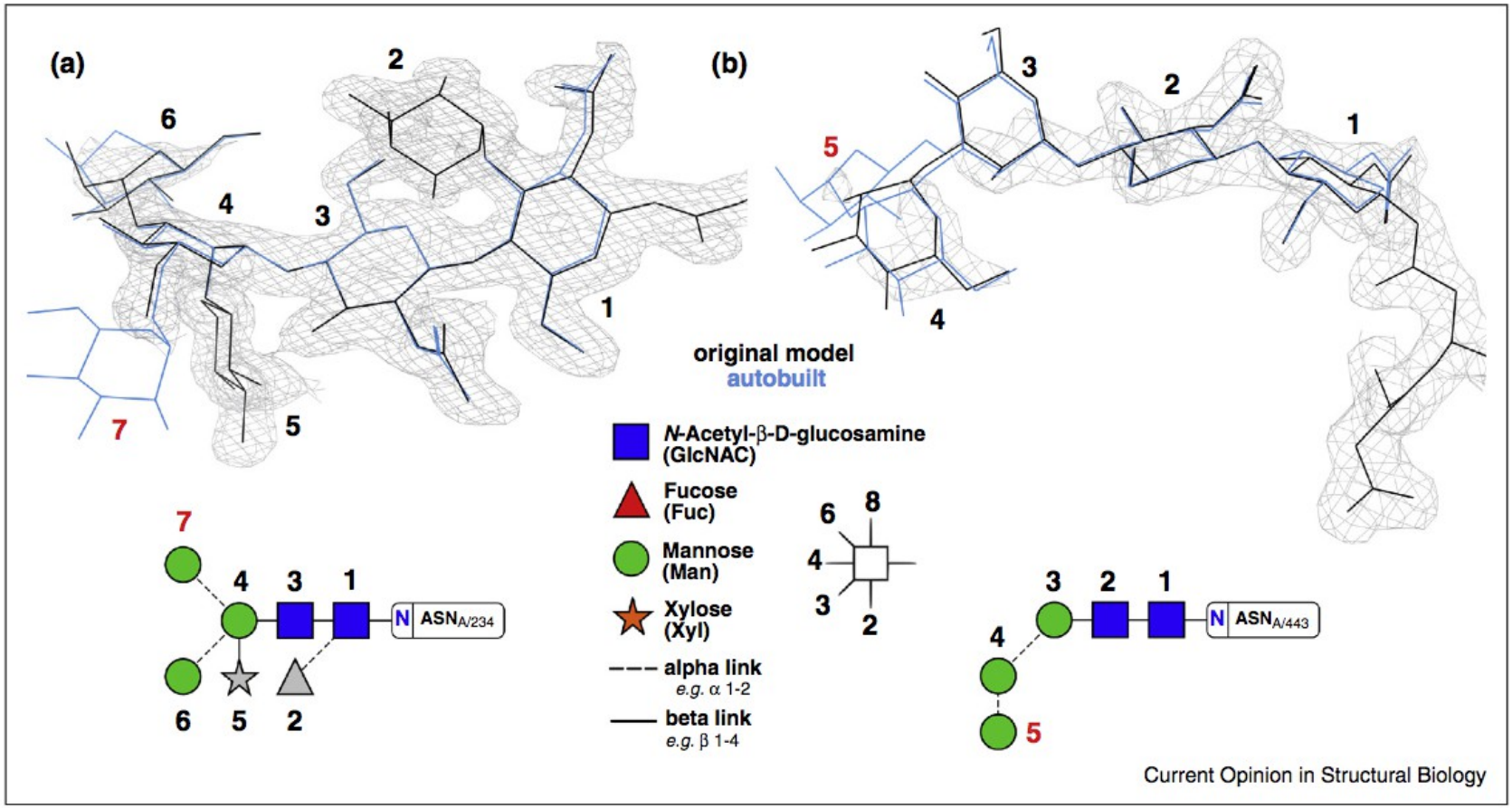
- Two Modes:
- Residue by Residue
  - User Control
- Whole Tree Addition
  - Automated
  - Embedded decision-making



Problematic Glycosylation

Agirre *et al.* (2017) The Rocky Road to Automation

Figure 2



Linking Fucose: Fuc-α1,3

- Add a menu item to wrap the command
  - `add_linked_residue("FUC", "ALPHA1-3")`

Added into a new N-linked tree:

- paucimannose

Xyl-β1,2

- Xyl - β1,2 – Man
  - using XYP (beta D xylosepyranose)
  - was not in the Refmac Monomer Library list of links
  - It has been added and will be available to CCP4 shortly

Refinement Stabilizers

- Using Coot's Real Space Refinement
  - (in default mode)
  - allowed saccharides to result in twisted or boat ring conformations
    - even without user intervention
- Unimodal Torsion Restraints
- Pseudo-plane restraints
- Inter-residue Geman-McClure external distance restraints

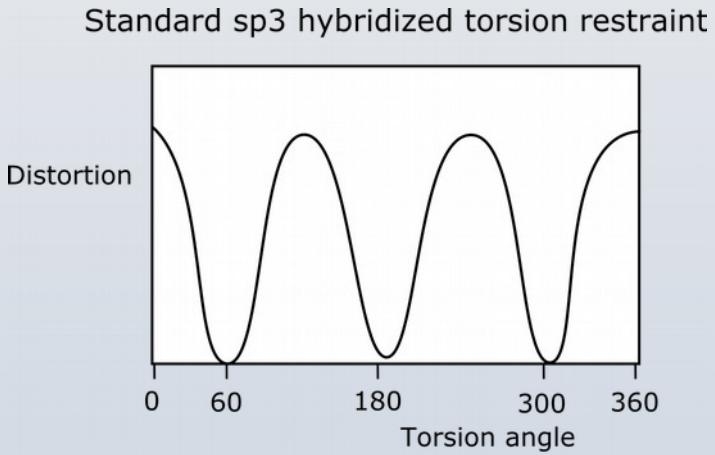
REFMAC Monomer Library
chem_comp_bond

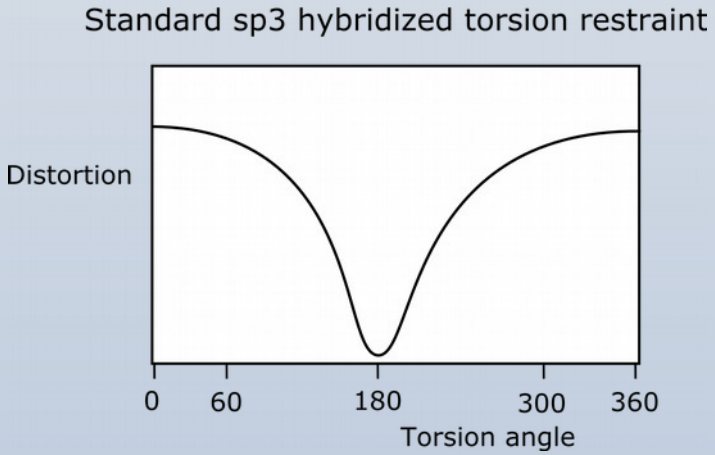
```
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
 ALA      N      H         single      0.860    0.020
 ALA      N      CA        single      1.458    0.019
 ALA      CA     HA        single      0.980    0.020
 ALA      CA     CB        single      1.521    0.020
 ALA      CB     HB1       single      0.960    0.020
 ALA      CB     HB2       single      0.960    0.020
```

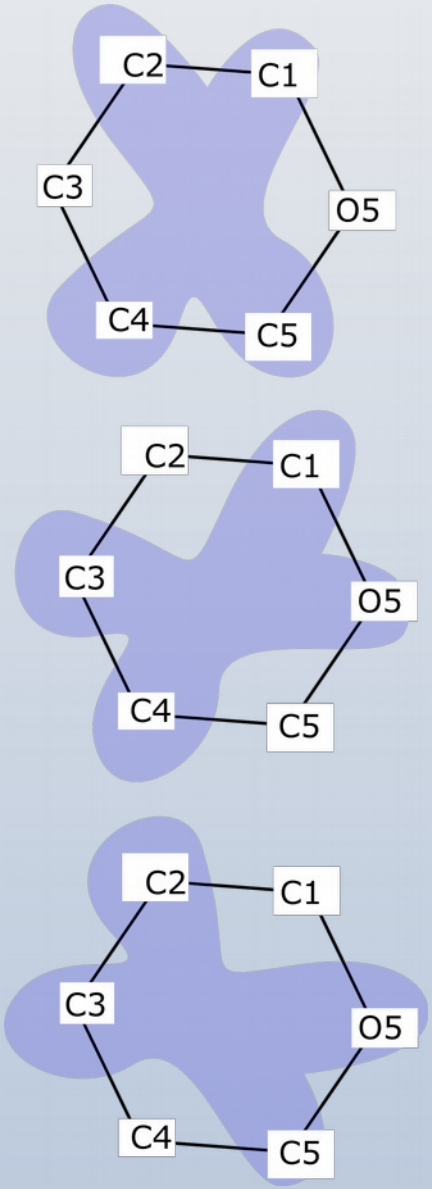
REFMAC Monomer Library `chem_comp_tor`

```
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 ADP      var_1    O2A    PA     O3A    PB       60.005   20.000   1
 ADP      var_2    PA     O3A    PB     O1B      59.979   20.000   1
 ADP      var_3    O2A    PA     "O5'"  "C5'"   -59.942   20.000   1
 ADP      var_4    PA     "O5'"  "C5'"  "C4'"   179.996   20.000   1
 ADP      var_5    "O5'"  "C5'"  "C4'"  "C3'"   176.858   20.000   3
 ADP      var_6    "C5'"  "C4'"  "O4'"  "C1'"   150.000   20.000   1
 ADP      var_7    "C5'"  "C4'"  "C3'"  "C2'"  -150.000   20.000   3
```

Unimodal Torsions and Pseudo Planes





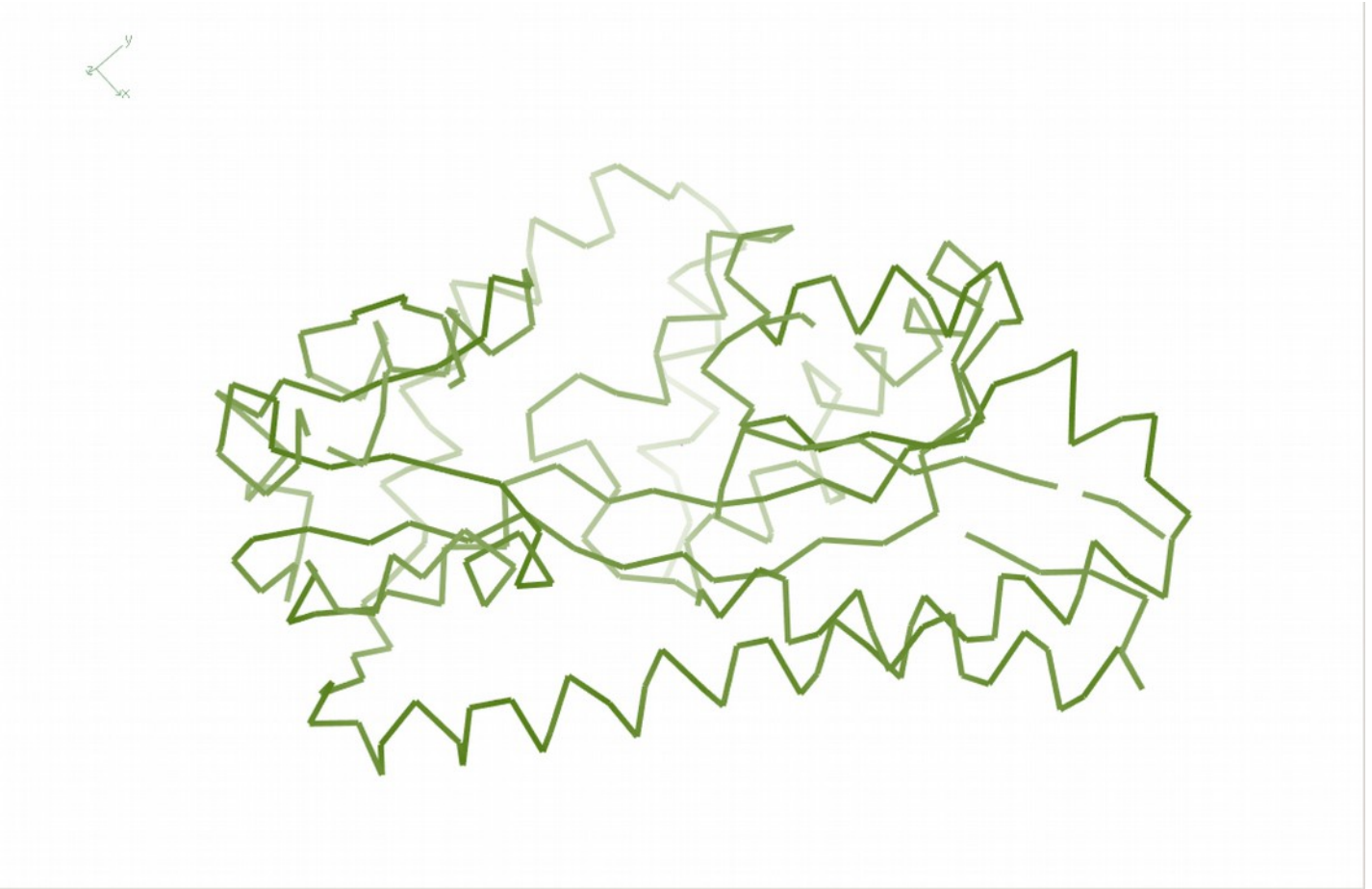


External Distance Restraints

- Using distances from prior-know glycan crystal structures
  - c.f. ProSMART for protein models
  - here we use intent to use a consensus model rather than a particular model

ProSMART Restraints

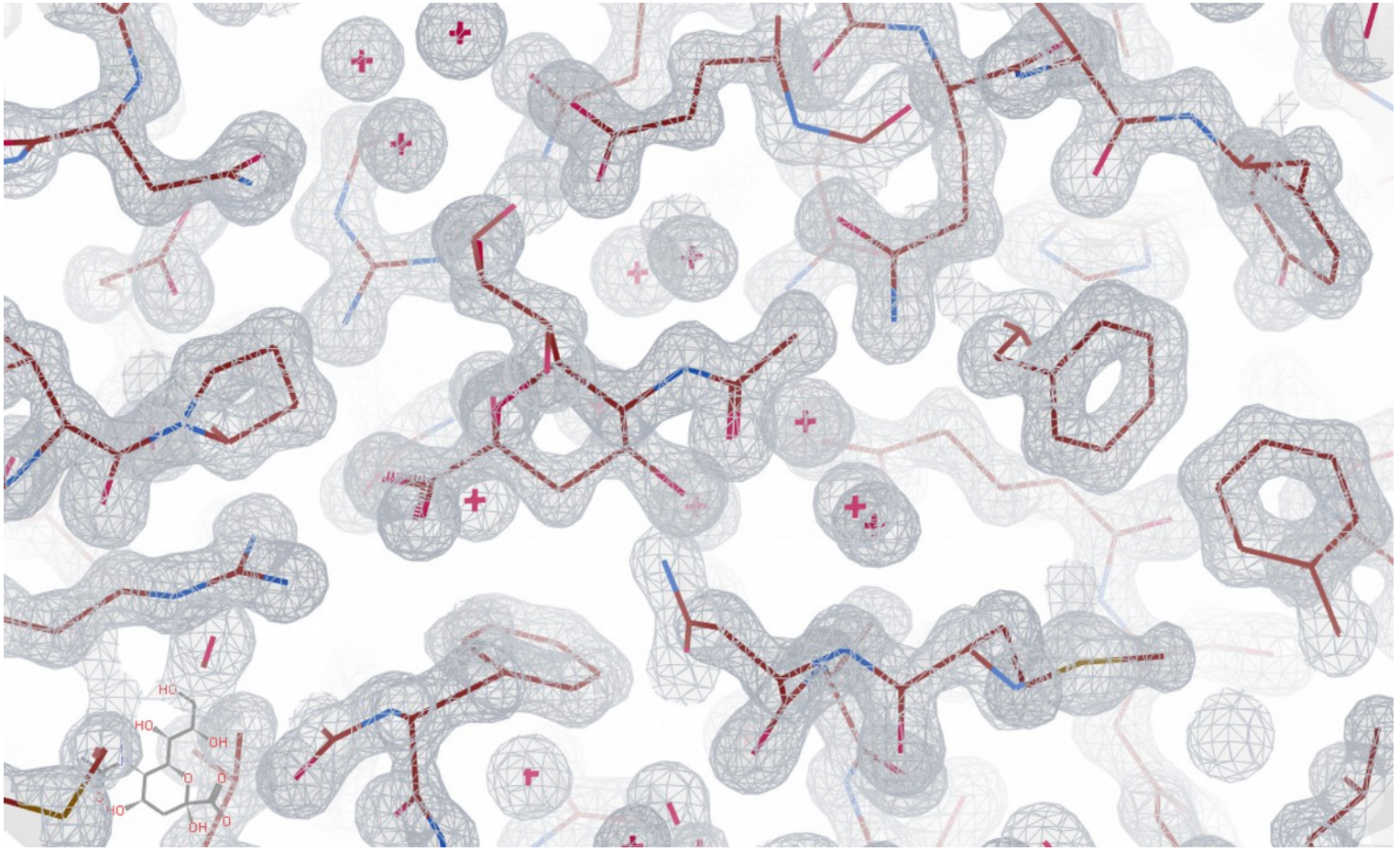
using Distance Restraints for Refinement Stabilization



Low resolution current structure

ProSMART Restraints

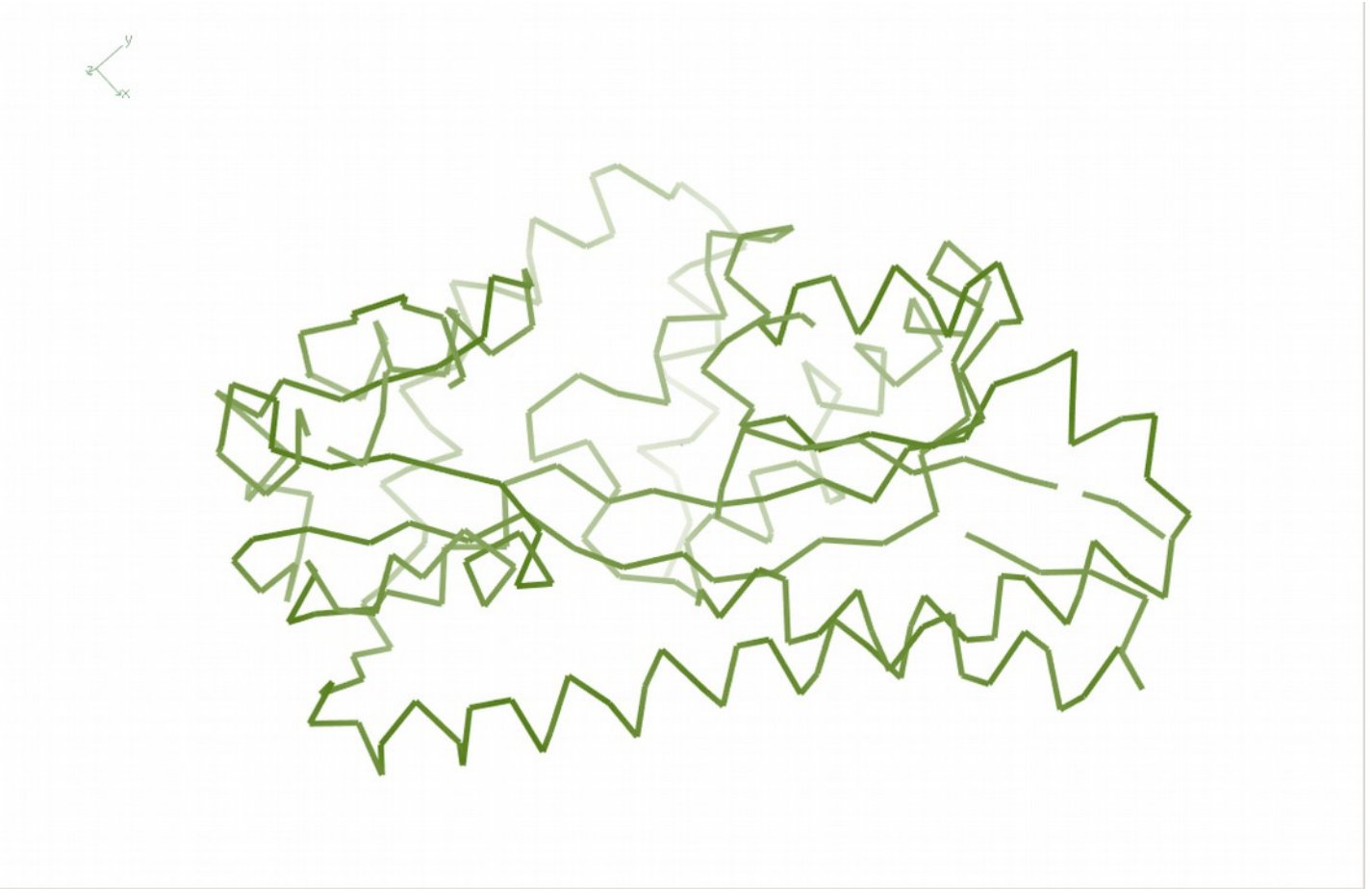
using Distance Restraints for Refinement Stabilization



Previously Known High-resolution Reference

ProSMART Restraints

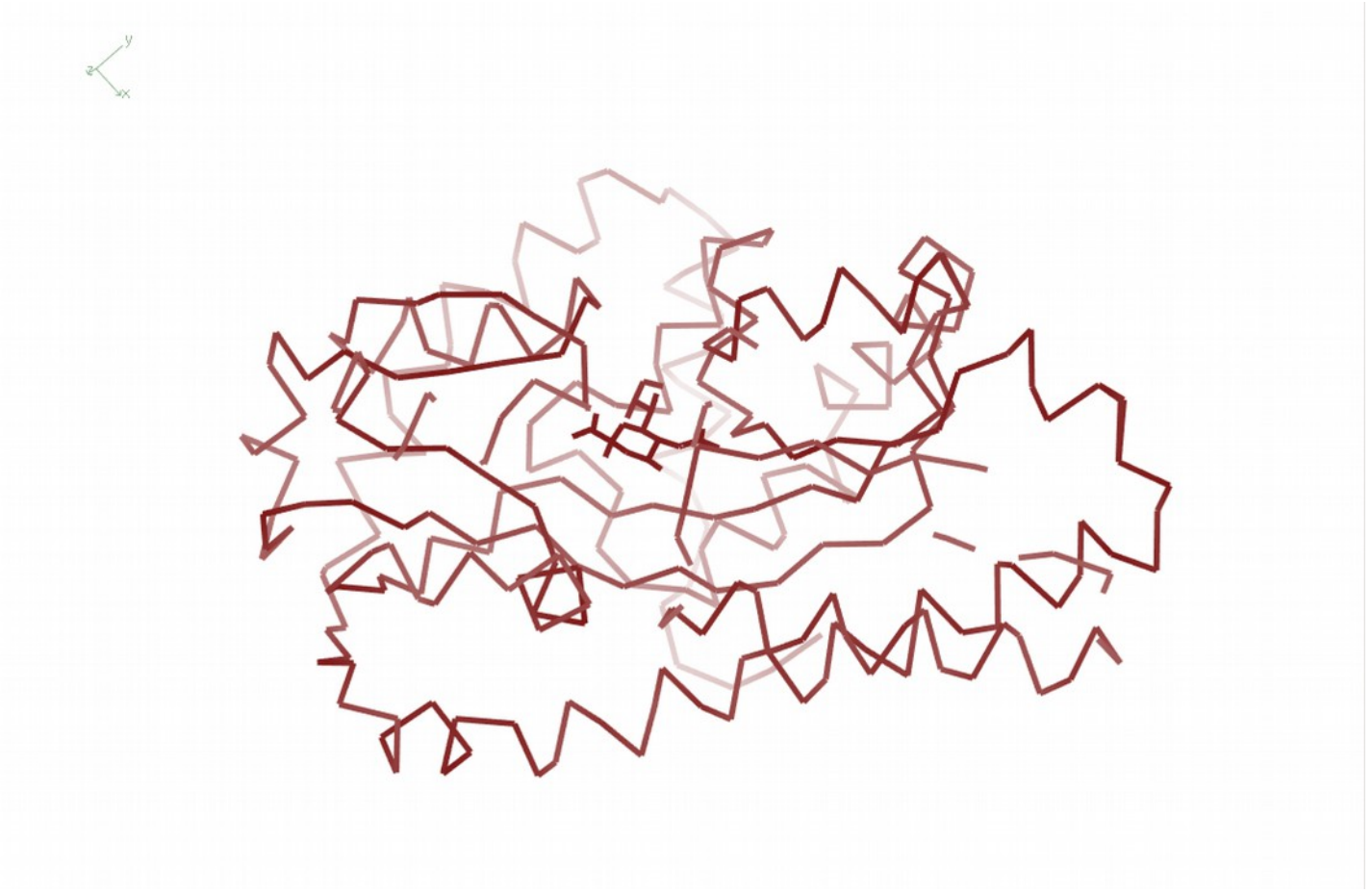
using Distance Restraints for Refinement Stabilization



Low resolution current structure

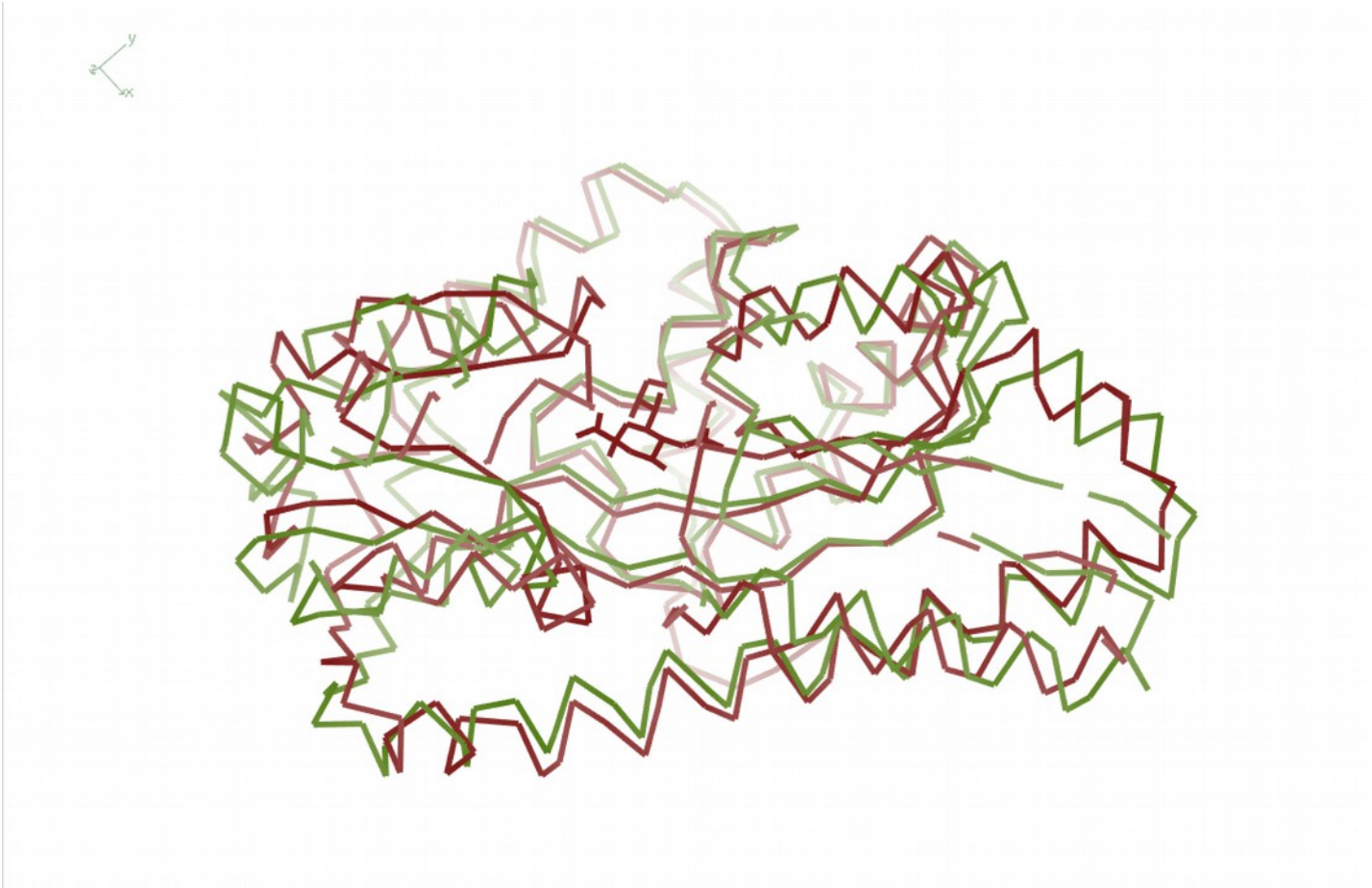
ProSMART Restraints

using Distance Restraints for Refinement Stabilization



High resolution reference structure

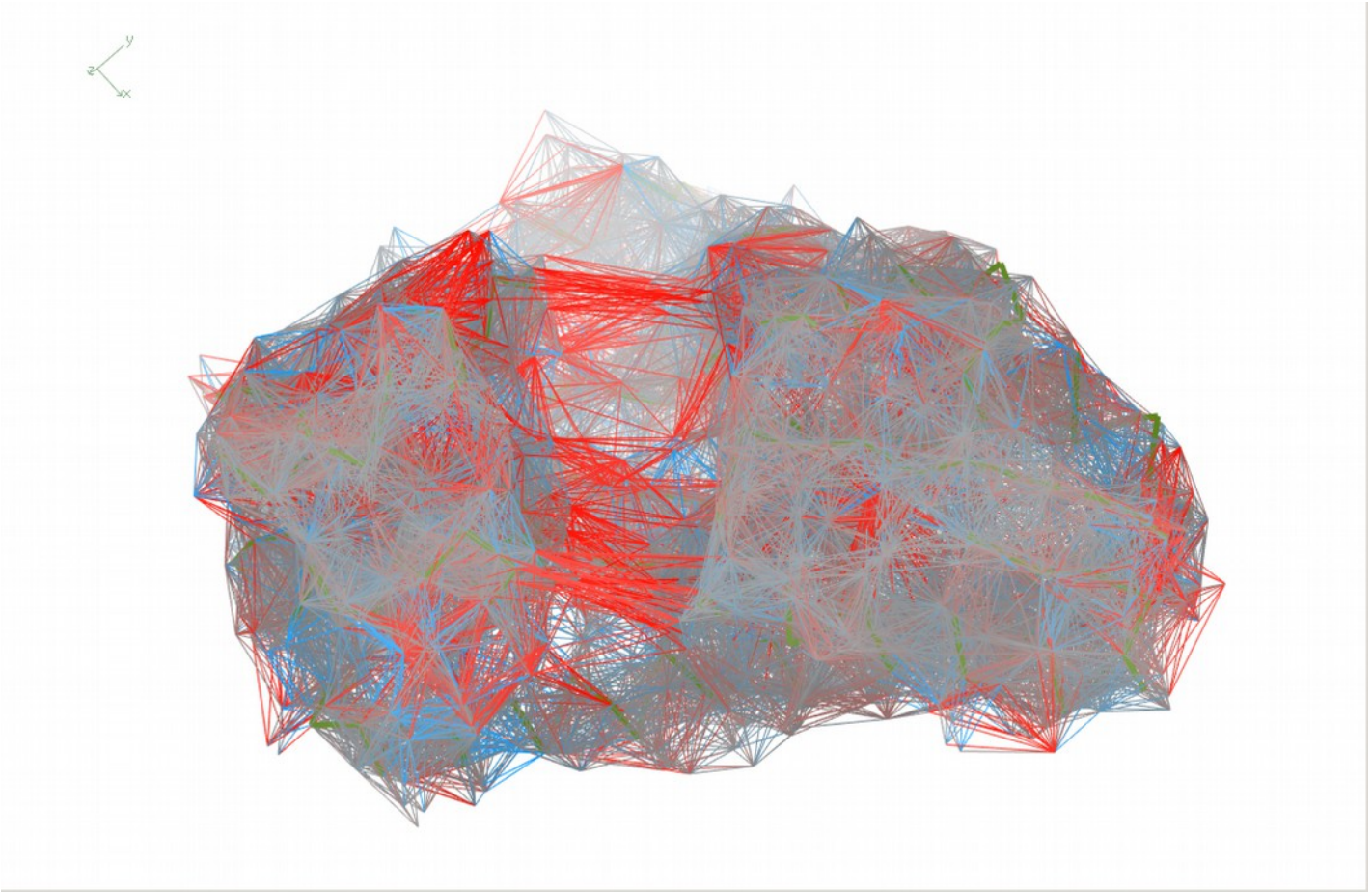
ProSMART Restrains



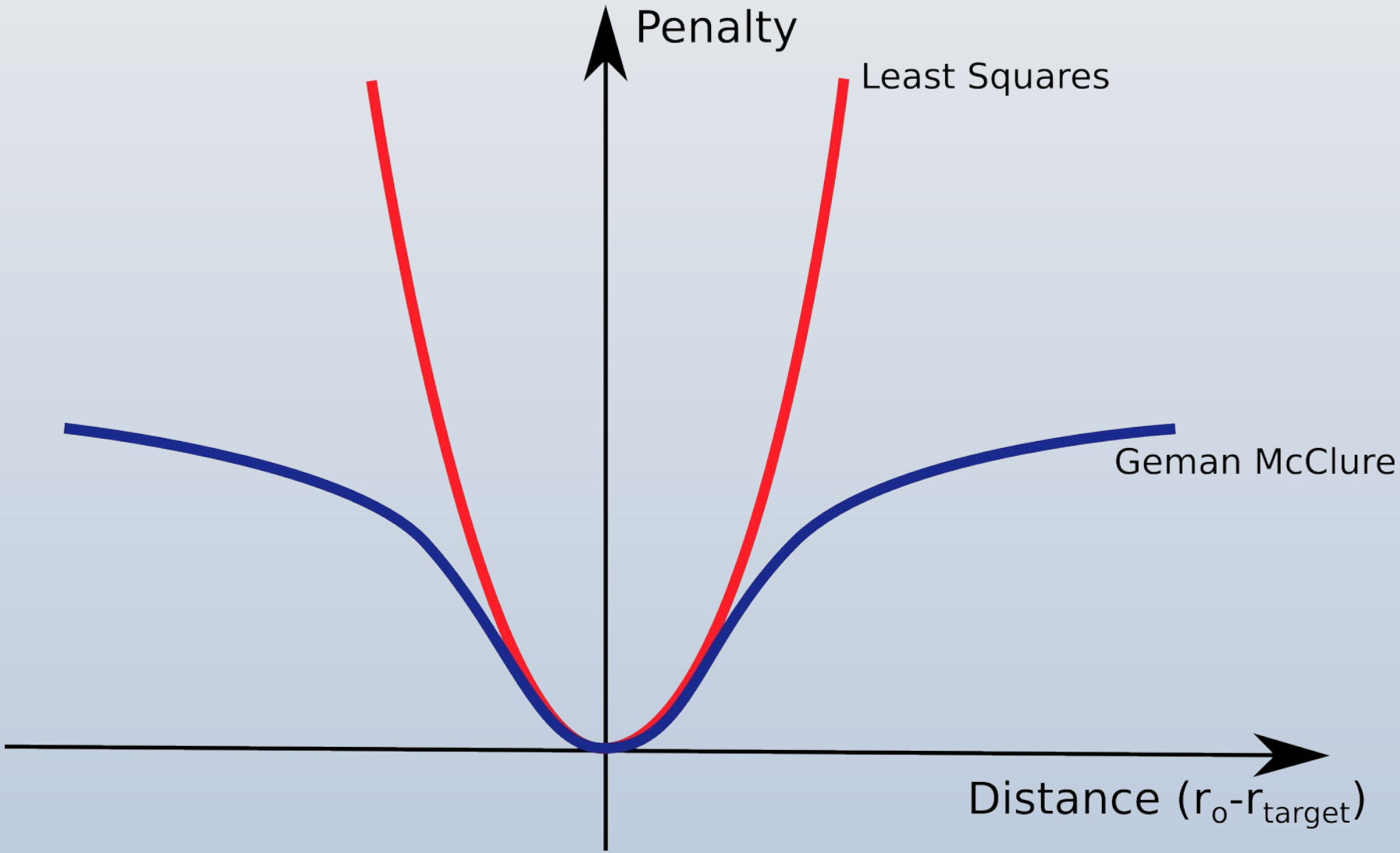
Similar but different

ProSMART Restraints

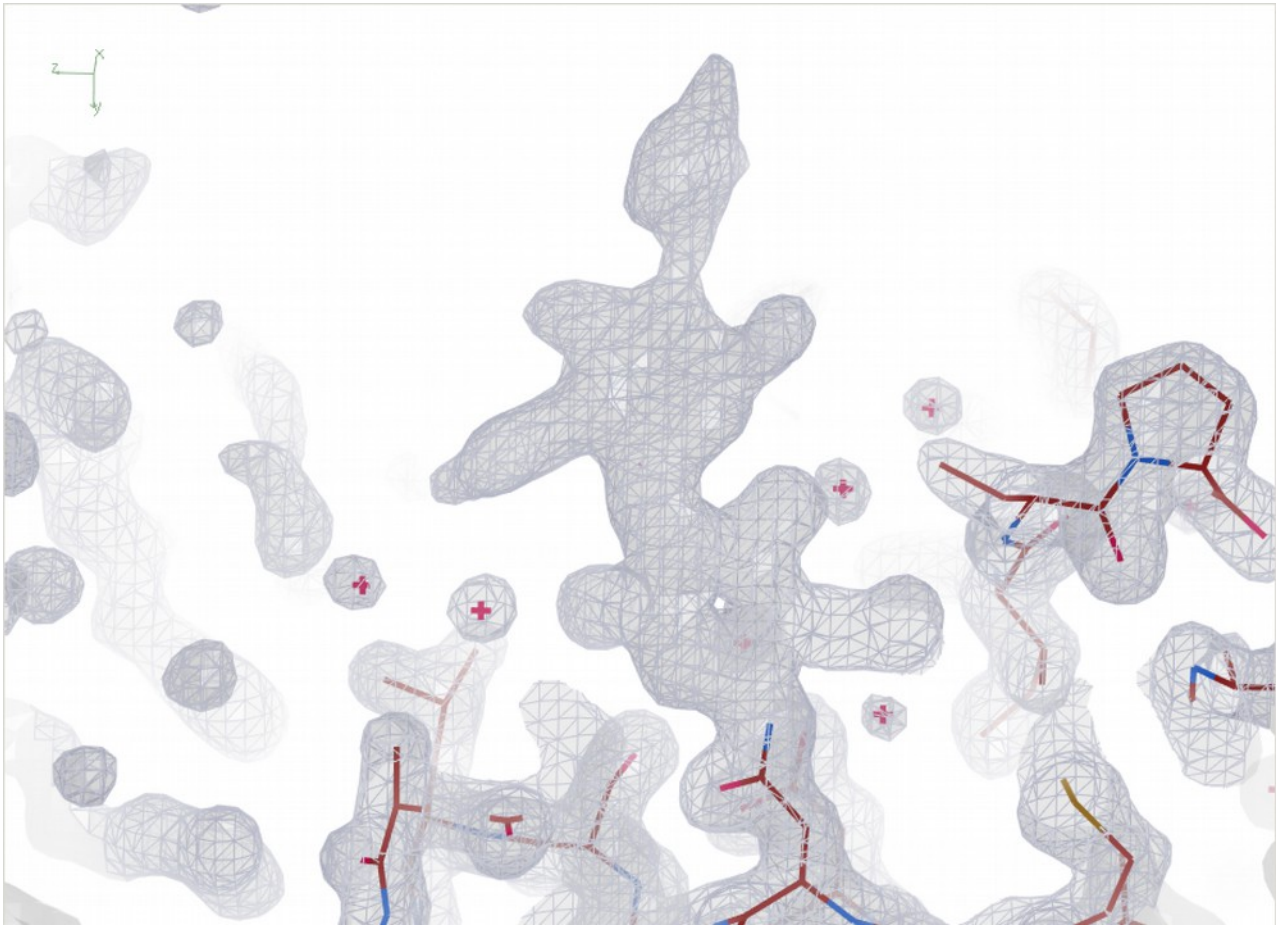
using Distance Restraints for Refinement Stabilization



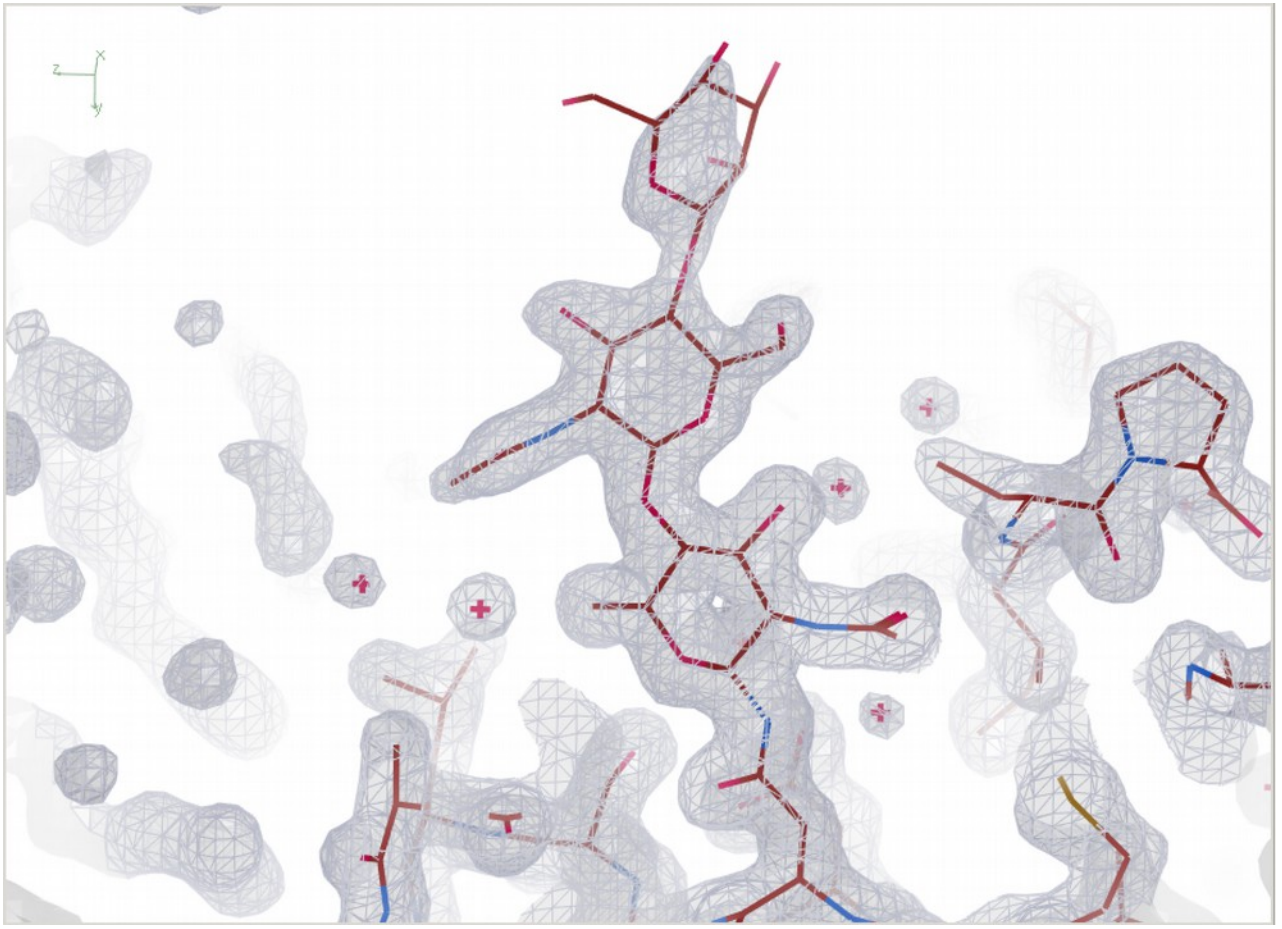
Modified Target Function



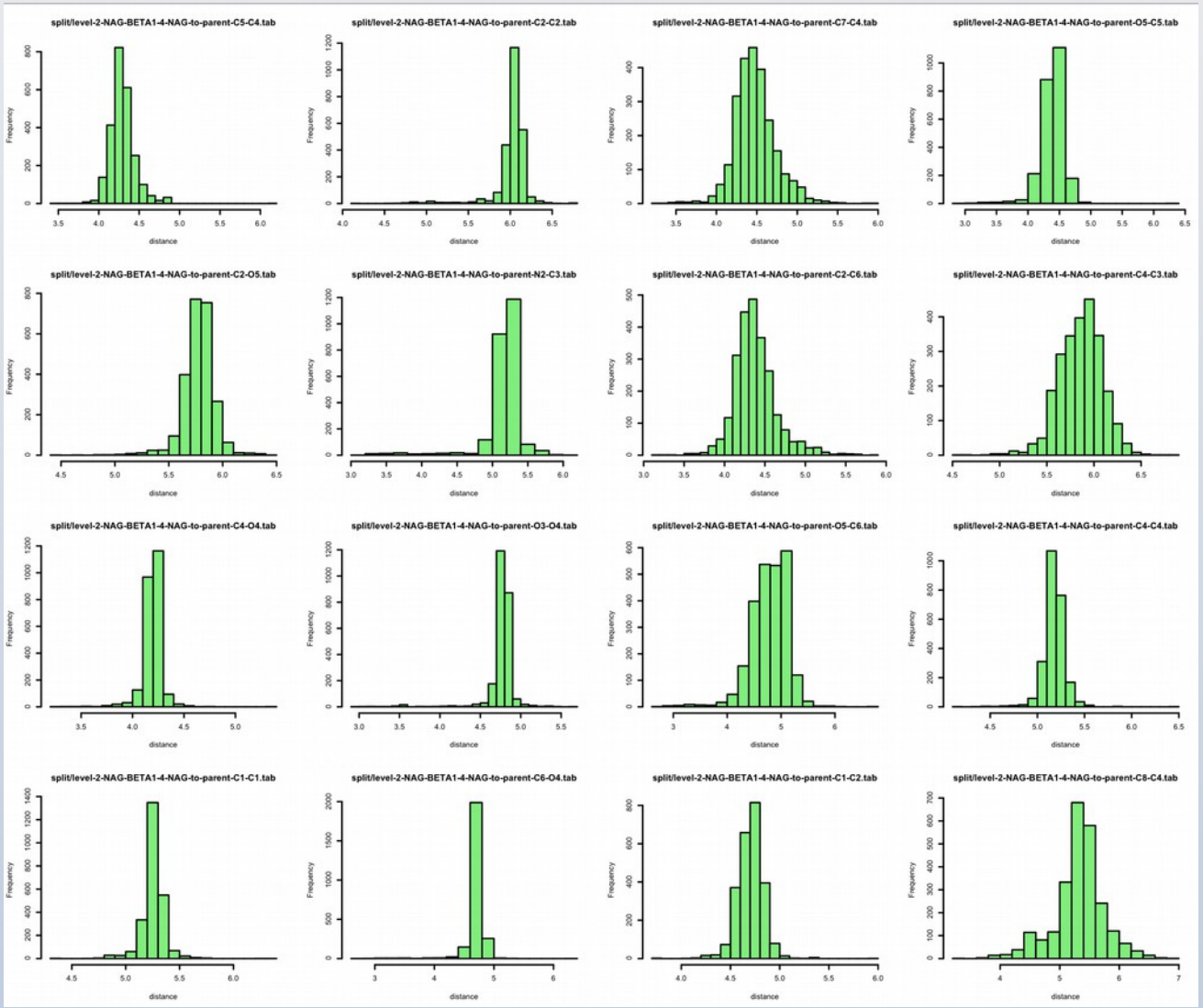
N-linked Carbohydrate



N-linked Carbohydrate



Ligand Tools: N-linked Carbohydrate



→ consensus restraints (no user-defined prior)

Adding PRIVATEER for Model Validation

- Until 2016 *Coot* had no validation for carbohydrate geometry
  - (only fit to density was used)
- Now the model is validated (and filtered) by tree
  - using the output of PRIVATEER
  - both GUI interface and built into the auto-builder
- New Interface

Video/Demo

-

Building Models "Wrongly" (judging by density)

	Good Density	Poor/Bad density
Model built	✓	False Positive
No Model	False Negative	✓