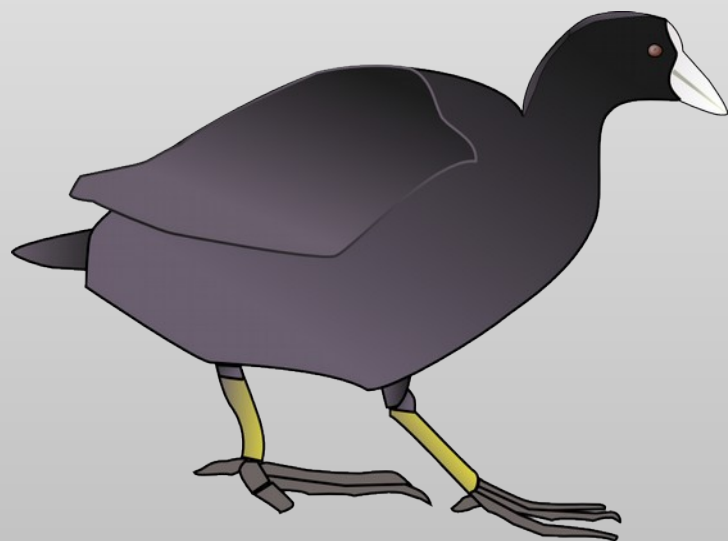


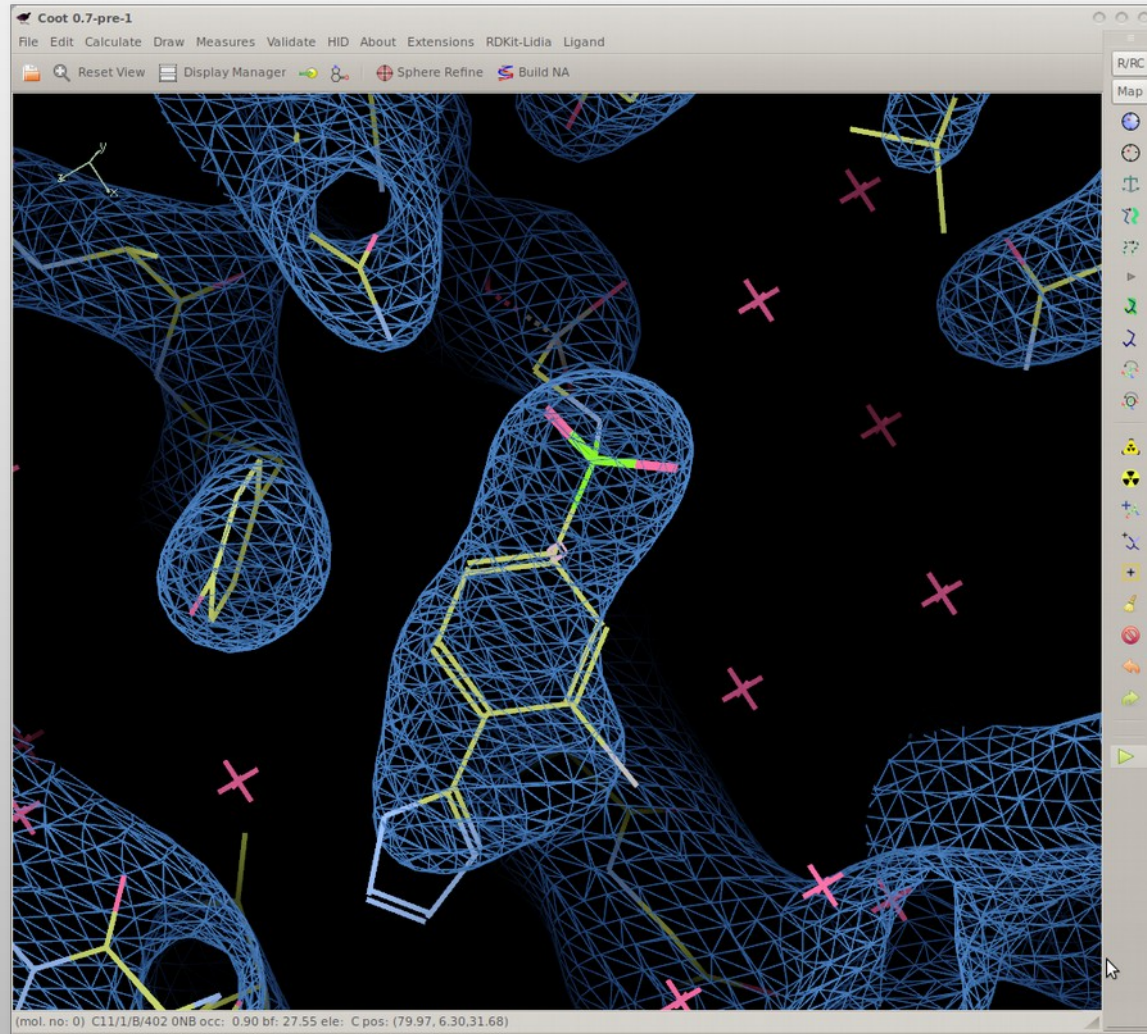
Manipulating Ligands Using *Coot*

with a bit of carbohydrates



Paul Emsley
May 2017

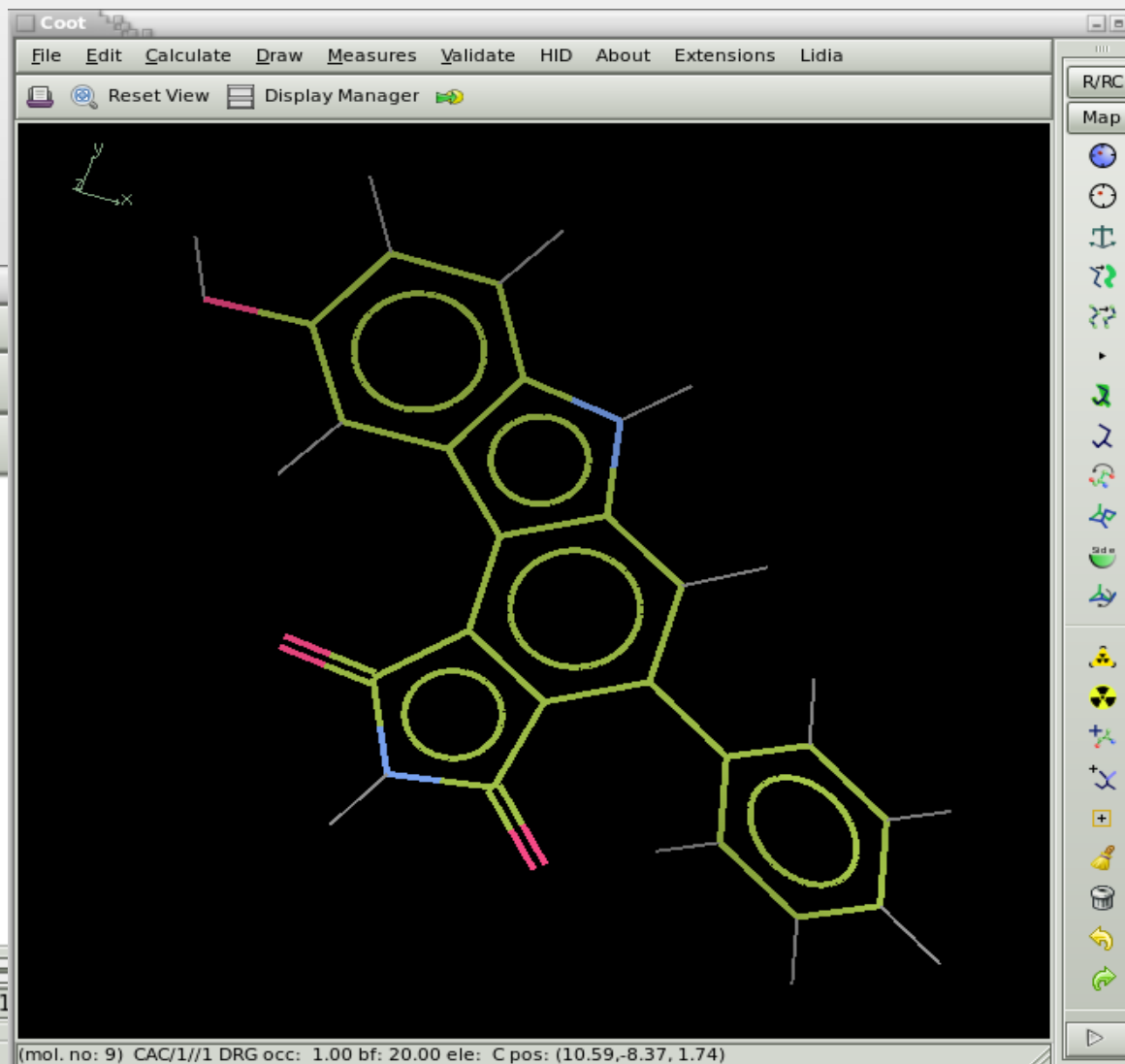
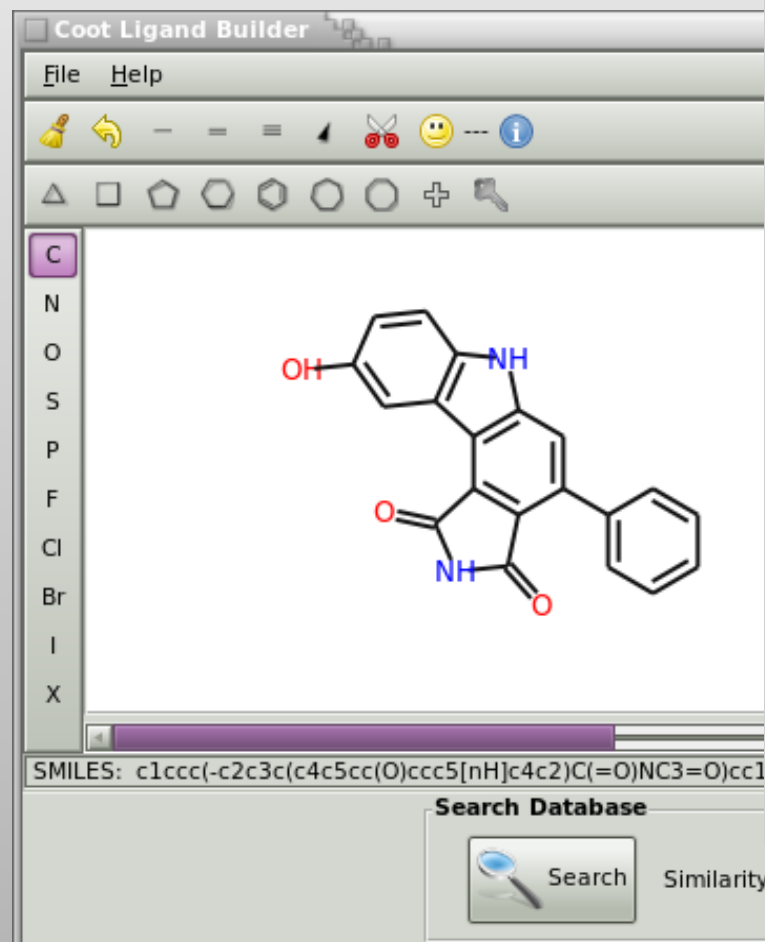
Ligand and Density...



Protein-ligand complex models are often a result of subjective interpretation

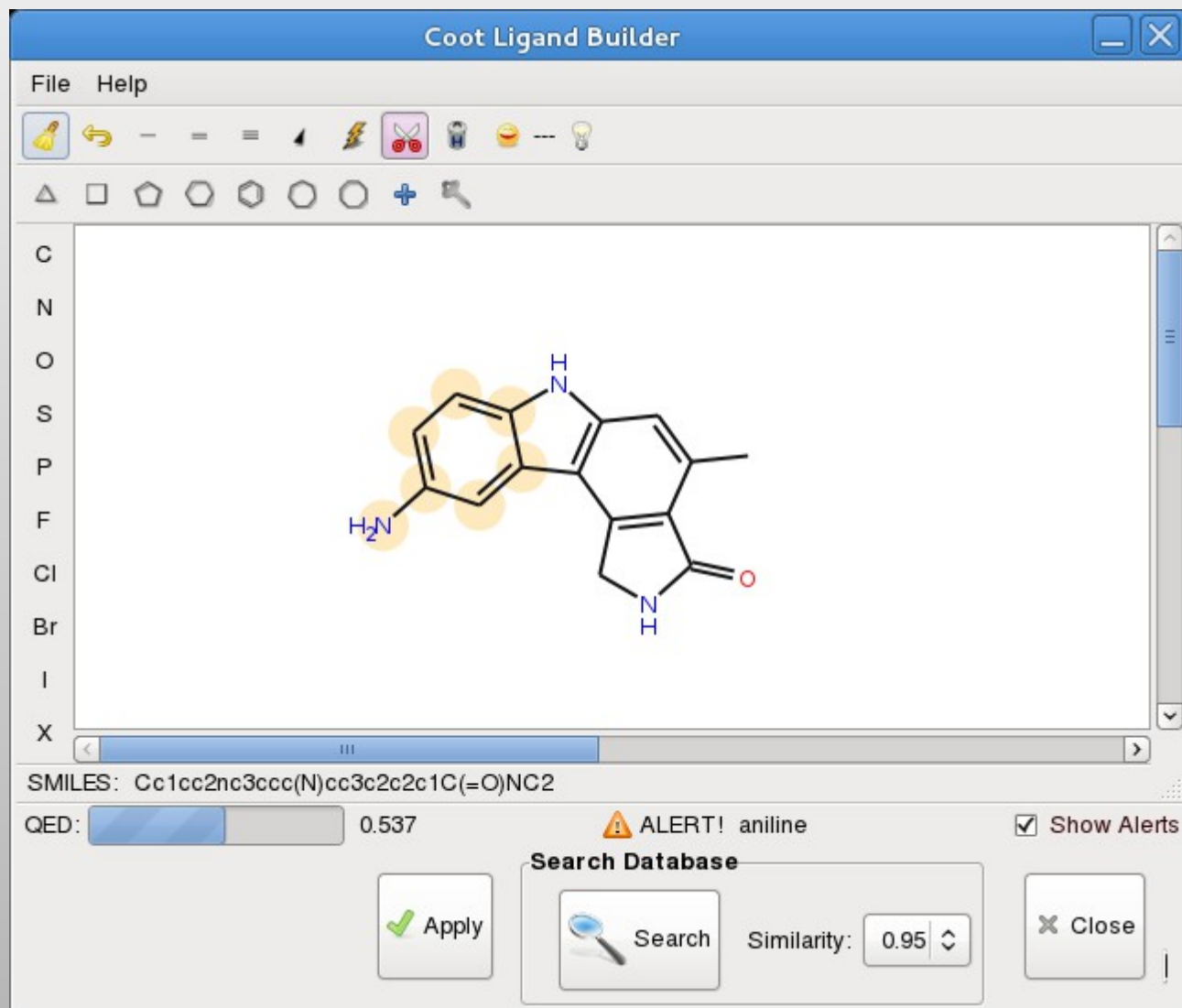
2D Ligand Builder

- Free sketch
- SBase search



2D Sketcher

- Structural Alerts



- On the fly ROMol creation
- Check vs. vector of SMARTS
 - (from Biscu-it)
 - And user-defined (python variable) list

QED Score

Quantitative Evaluation of Drug-likeness

ARTICLES

PUBLISHED ONLINE: 24 JANUARY 2012 | DOI: 10.1038/NCHEM.1243

nature
chemistry

Quantifying the chemical beauty of drugs

G. Richard Bickerton¹, Gaia V. Paolini², Jérémy Besnard¹, Sorel Muresan³ and Andrew L. Hopkins^{1*}

Drug-likeness is a key consideration when selecting compounds during the early stages of drug discovery. However, evaluation of drug-likeness in absolute terms does not reflect adequately the whole spectrum of compound quality. More worryingly, widely used rules may inadvertently foster undesirable molecular property inflation as they permit the encroachment of rule-compliant compounds towards their boundaries. We propose a measure of drug-likeness based on the concept of desirability called the quantitative estimate of drug-likeness (QED). The empirical rationale of QED reflects the underlying distribution of molecular properties. QED is intuitive, transparent, straightforward to implement in many practical settings and allows compounds to be ranked by their relative merit. We extended the utility of QED by applying it to the problem of molecular target druggability assessment by prioritizing a large set of published bioactive compounds. The measure may also capture the abstract notion of aesthetics in medicinal chemistry.

The concept of drug-likeness provides useful guidelines for early-stage drug discovery^{1,2}. Analysis of the observed distribution of some key physicochemical properties of approved drugs, including molecular mass (M_r), hydrophobicity and polarity, reveals that they occupy preferentially a relatively narrow range of possible values³. Compounds that fall within this range are described as 'drug-like'. This definition holds in the absence of any obvious structural similarity to an approved drug. It has been shown that the preferential selection of drug-like compounds increases the likelihood of surviving the well-documented high rates of attrition in drug discovery⁴.

Drug-likeness can be rationalized by considering how simple physicochemical properties impact molecular behaviour *in vivo*, with particular respect to solubility, permeability, metabolic stability and transporter effects. Indeed, drug-likeness is often used as a proxy for oral bioavailability. However, drug-likeness provides a broad composite descriptor that implicitly captures several criteria,

Paradoxically, since the publication of the seminal paper by Lipinski *et al.*⁵ there appears to be a growing epidemic, which Hann has termed 'molecular obesity'⁶, among new pharmacological compounds (Supplementary Fig. S1). Compounds with higher relative M_r and lipophilicity have a higher probability of attrition at each stage of clinical development^{4,7-11}. Thus, the inflation of physicochemical properties that increases the risks associated with clinical development may explain, in part, the decline in productivity of small-molecule drug discovery over the past two decades⁴. However, the mean molecular properties of new pharmacological compounds are still considered Lipinski compliant, even though their property distributions are far from historical norms.

Although the Ro5 is predictive of oral bioavailability, 16% of oral drugs violate at least one of the criteria and 6% fail two or more (although this does include natural products and substrates of transporters) (Supplementary Fig. S2a and Supplementary Table S1). High-profile drugs, such as atorvastatin (Lipitor) and montelukast

NATURE CHEMISTRY DOI: 10.1038/NCHEM.1243

ARTICLES

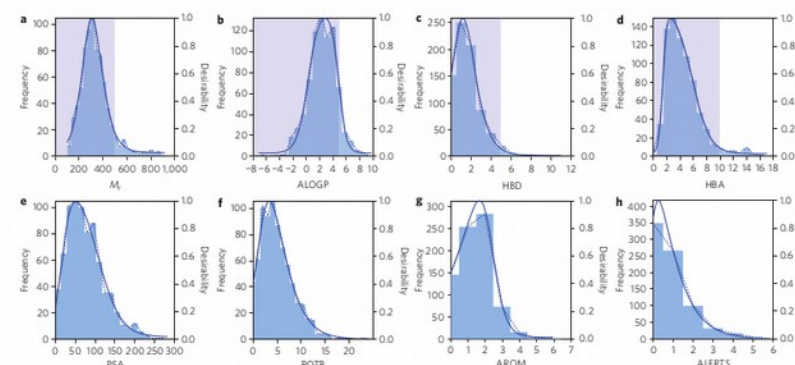


Figure 1 | Histograms of eight selected molecular properties for a set of 771 orally absorbed small molecule drugs. **a–h**, Molecular properties M_r (**a**), lipophilicity estimated by atom-based prediction of ALOGP (**b**), number of HBDs (**c**), number of HBAs (**d**), PSA (**e**), number of ROTBs (**f**), number of AROMs (**g**) and number of ALERTS (**h**). The Lipinski-compliant areas are shown in pale blue in (**a**), (**b**), (**c**) and (**d**). The solid blue lines describe the ADS functions (equation (2)) used to model the histograms. The parameters for each function are given in Supplementary Table S1.

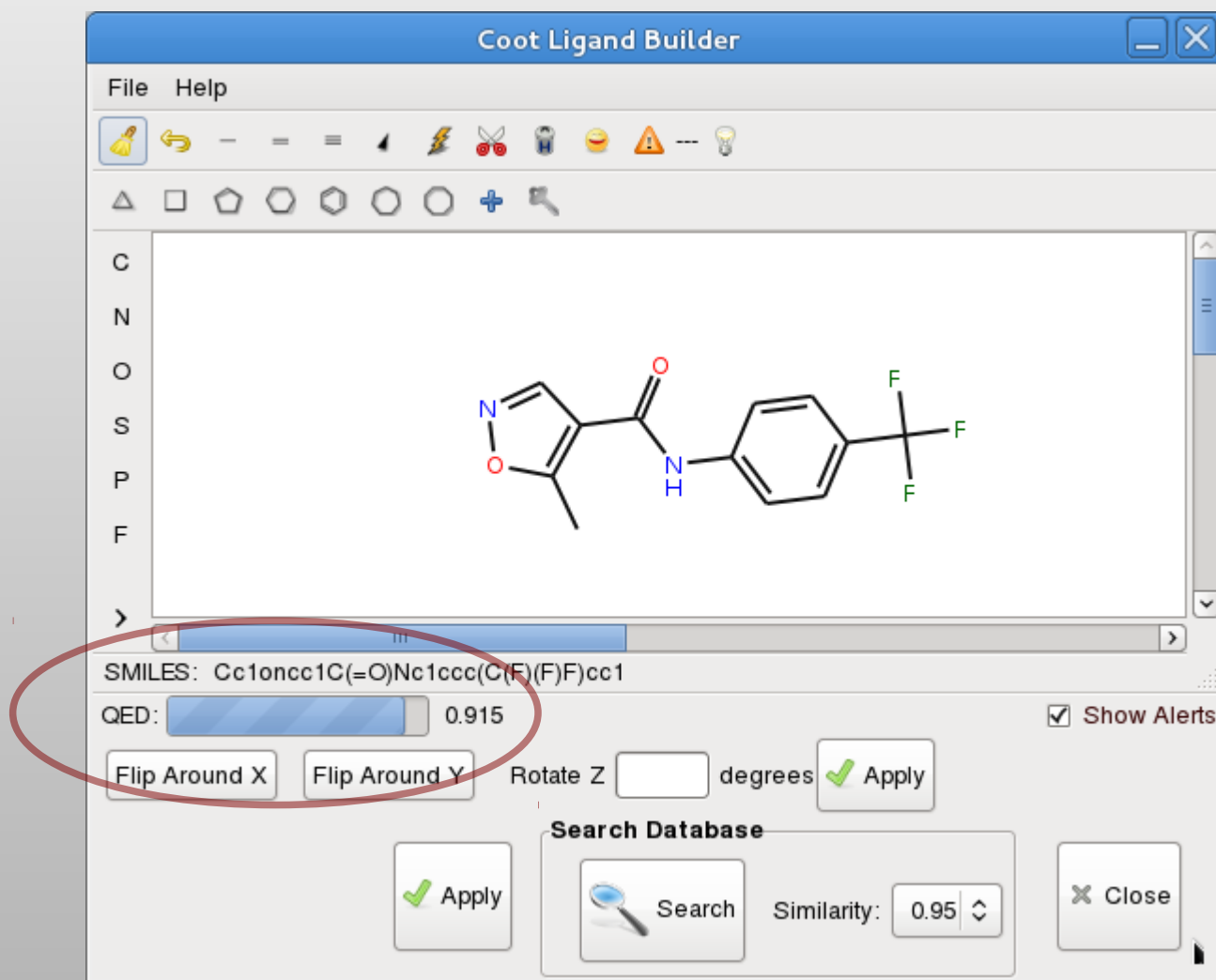
design^{17,18}, prioritization of molecular targets, penetration of the central nervous system¹⁹ and estimating the reliability of screening data²⁰. The concept was introduced originally by Harrington¹⁵ in the area of process engineering and further refined by Derringer and Suich²¹. Desirability takes multiple numerical or categorical parameters measured on different scales and describes each by an individual desirability function. These are then integrated into a single dimensionless score. In the case of compounds, a series of desirability functions (d) are derived, each of which corresponds to a different molecular descriptor. Combining the individual desirability functions into the QED is achieved by taking the geometric

asymmetric double sigmoidal (ADS) functions, which are also shown in Fig. 1 over the same range. The general ADS function is shown in equation (2), where $d(x)$ is the desirability function for molecular descriptor x :

$$d(x) = a \left[\frac{b}{1 + \exp\left(\frac{x - c + \frac{d}{2}}{\epsilon}\right)} \right] \left[1 - \frac{1}{1 + \exp\left(\frac{x - c - \frac{d}{2}}{\epsilon}\right)} \right]$$

2D Sketcher

- QED score



Silicos-it's
Biscu-it™

Look up the function with
PyModule_GetDict()
and
PyModule_GetItem()

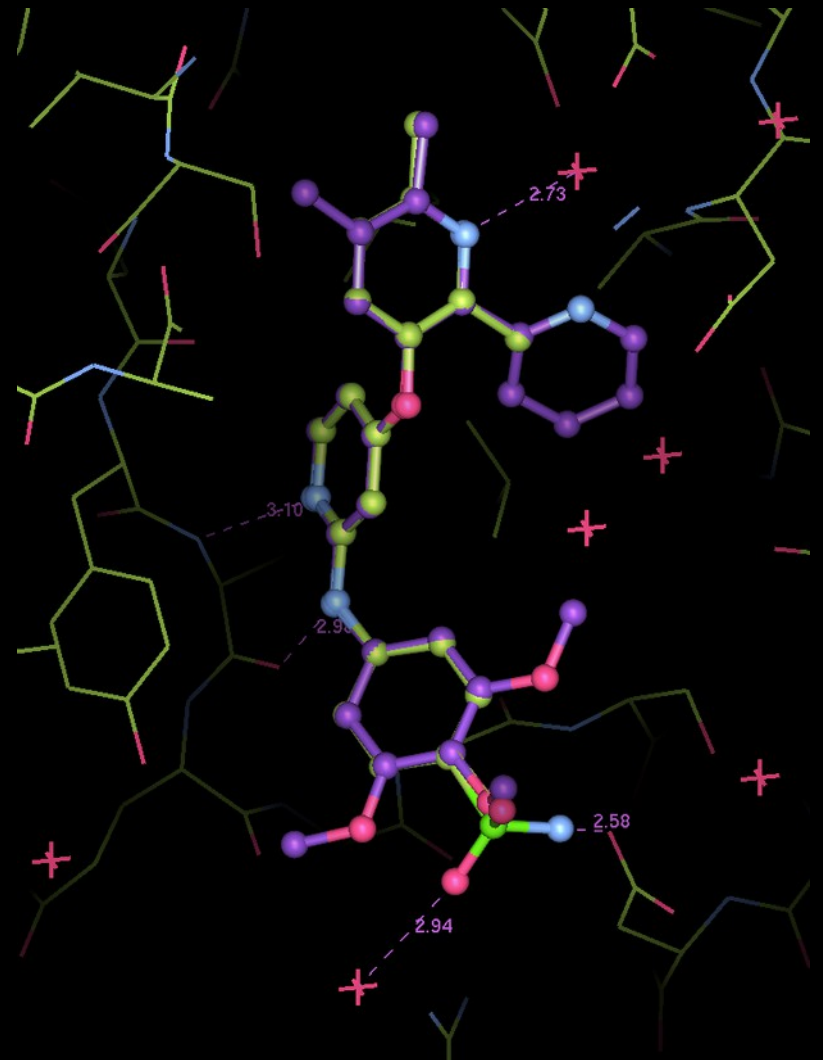
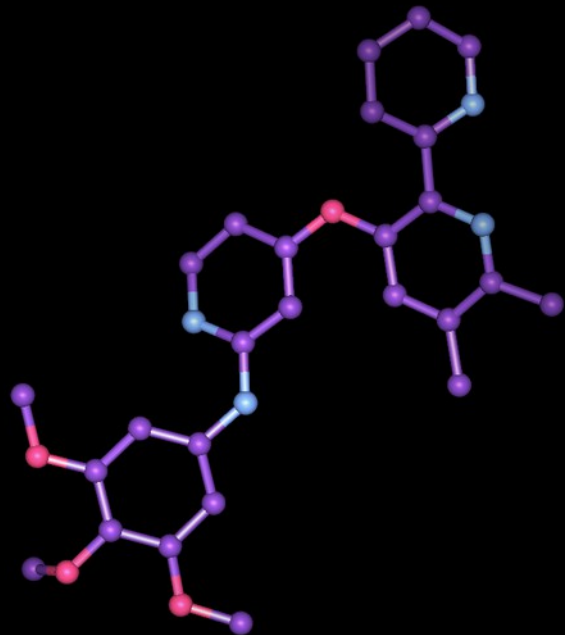
Ligand Utils

- “Get Molecule”
 - Uses network connection to Wikipedia
- Get *comp-id* ligand-description from PDBe
 - downloads and reads (e.g.) AAA.cif
 - (extracted from chemical component library)
- Drag and drop
 - Uses network connection to get URLs
 - or file-system files
- pyrogen
 - restraints generation

Manipulating Ligands

Using "Yesterday's" Ligand

Common subgraph isomorphism, Krissinel & Henrick (2004)



- Atom name matching
- Torsion matching
- Ligand overlay

Generating Conformers

- Using restraint information...

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

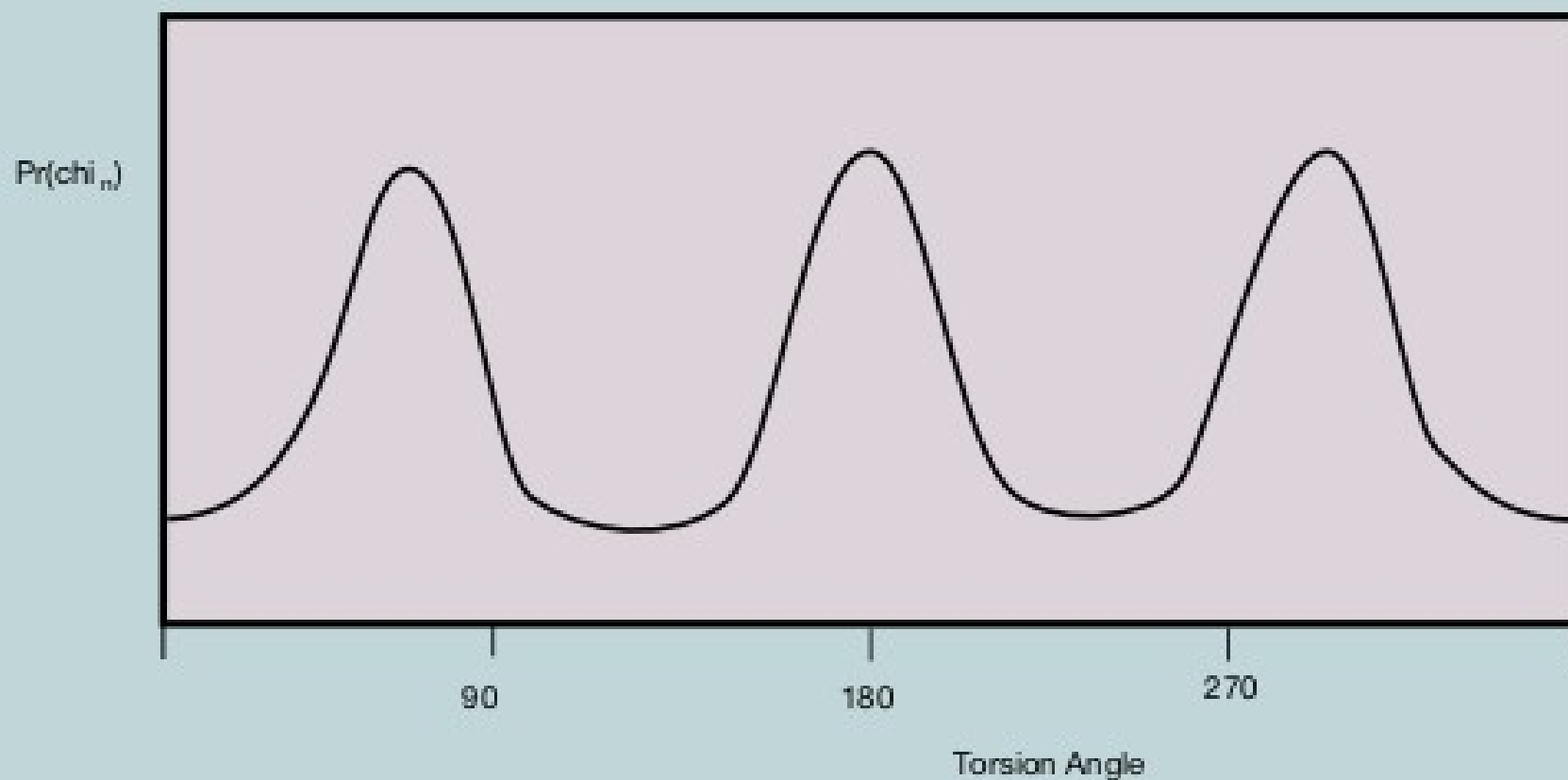
_chem_comp_tor.value_angle

_chem_comp_tor.value_angle_esd

_chem_comp_tor.period

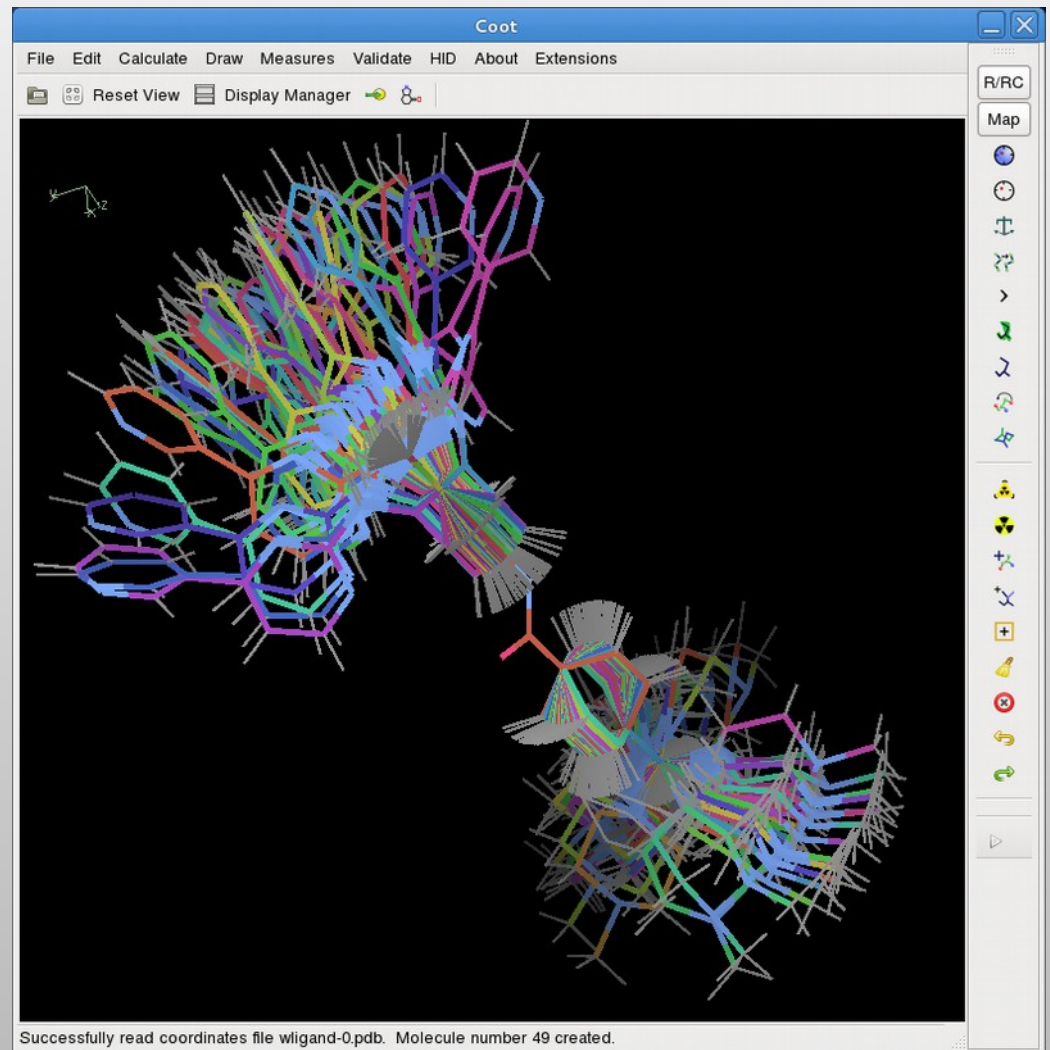
ADP	var_1	02A	PA	03A	PB	60.005	20.000	1
ADP	var_2	PA	03A	PB	01B	59.979	20.000	1
ADP	var_3	02A	PA	"05' "	"C5' "	-59.942	20.000	1
ADP	var_4	PA	"05' "	"C5' "	"C4' "	179.996	20.000	1
ADP	var_5	"05' "	"C5' "	"C4' "	"C3' "	176.858	20.000	3
ADP	var_6	"C5' "	"C4' "	"04' "	"C1' "	150.000	20.000	1
ADP	var_7	"C5' "	"C4' "	"C3' "	"C2' "	-150.000	20.000	3

Ligand Torsionable Angle Probability from CIF file



Conformer Generation

Non-Hydrogen
Non-CONST
Non-Ring



Fitting Ligands

Ligand Type

Ligand Site

Known

Unknown

Known



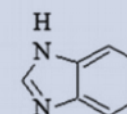
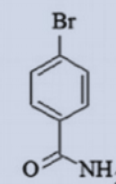
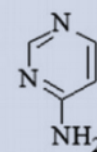
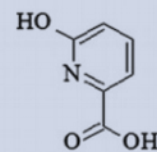
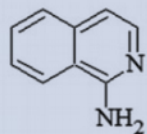
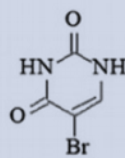
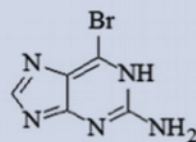
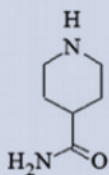
Cocktail



Unknown

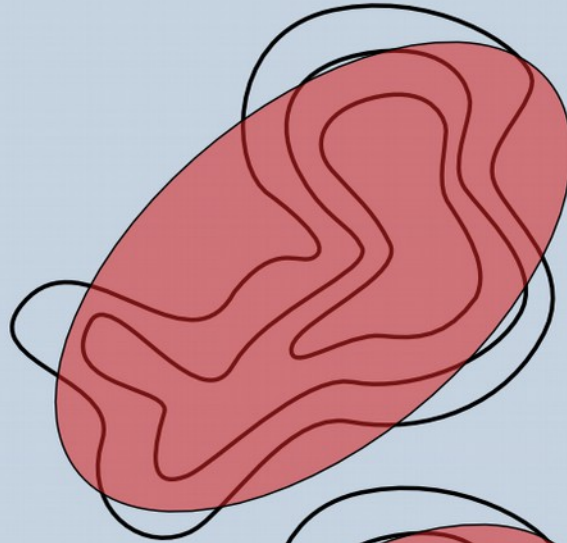


Cocktail Examples

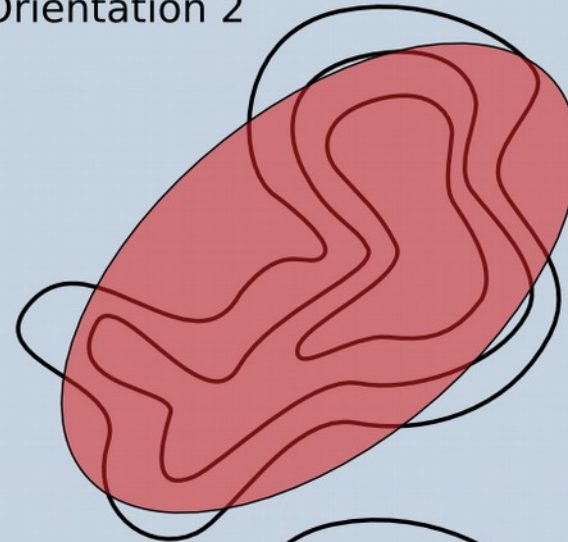


Orienting the Ligand

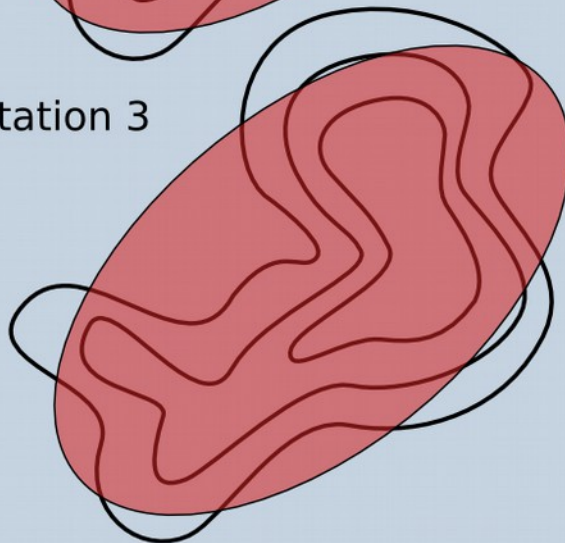
Orientation 1



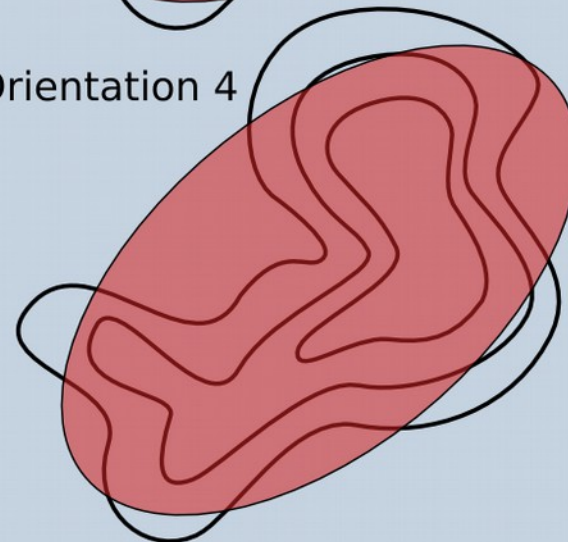
Orientation 2



Orientation 3

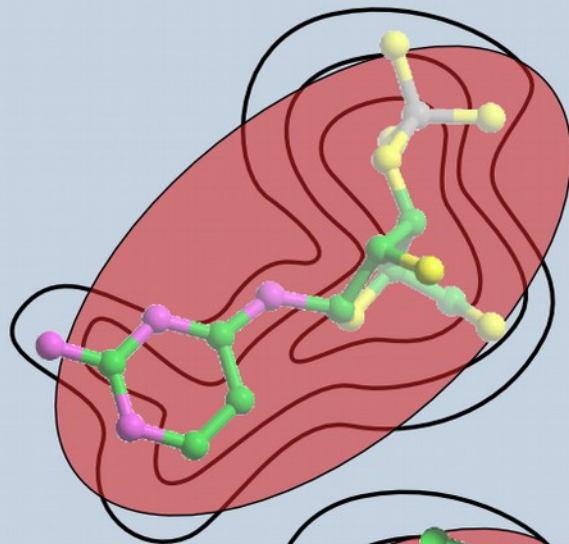


Orientation 4

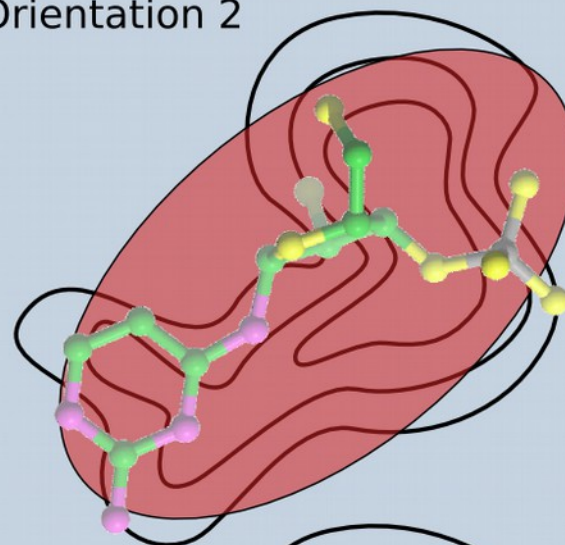


Orienting the Ligand

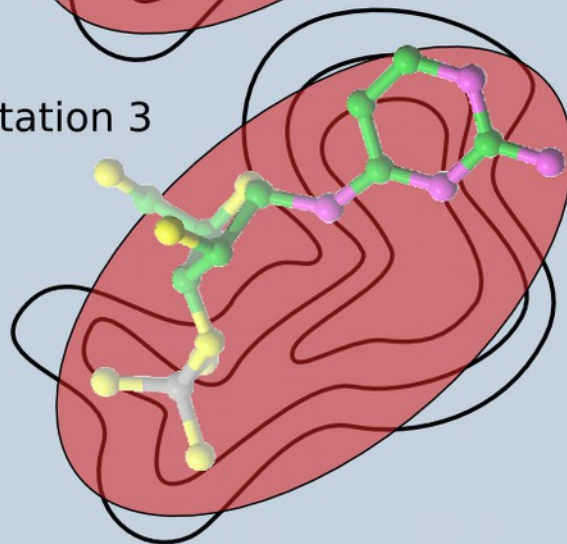
Orientation 1



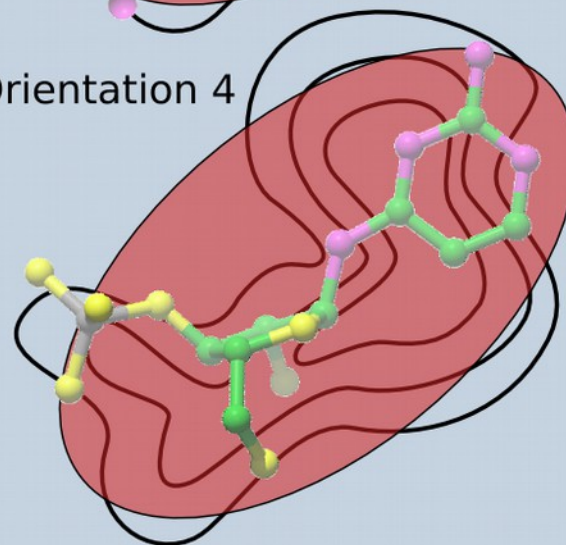
Orientation 2



Orientation 3



Orientation 4

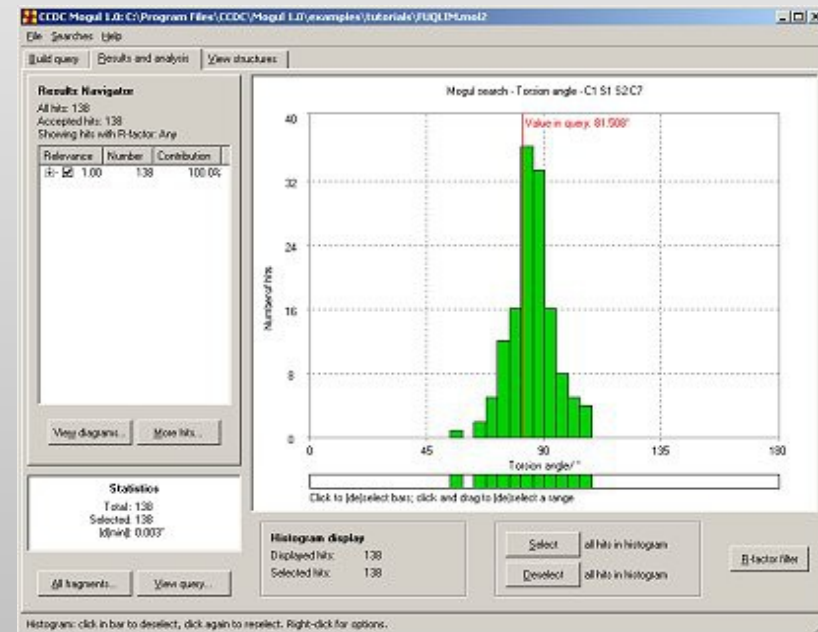


Ligand Validation

- Mogul plugin in Coot
 - Run mogul, graphical display of results
 - Update restraints (target and esds for bonds and angles)
 - CSD data not so great for plane, chiral and torsion restraints
 - (not by me, anyway)

Parmatisation issues... (what if they are wrong?)

- Perfect refinement with incorrect parameters → distorted structure
- CSD's Mogul
- This time:
 - Display and Interactive



Example Coot Ligand Distortion Score

Residue Distortion List:

plane	O3	C19	C20	C18	C16	C15	C17	C13	C14	N2	C4	C5	O1	C3	C6	O2	penalty-score:	36.51
plane	C2	C7	C8	C9	C10	C11	C12										penalty-score:	8.82
bond	C13	to	C4	target_value:	1.490	d:	1.432	sigma:	0.020	length-devi	-0.058						penalty-score:	8.44
bond	C4	to	C3	target_value:	1.490	d:	1.436	sigma:	0.020	length-devi	-0.054						penalty-score:	7.21
bond	O3	to	C19	target_value:	1.362	d:	1.318	sigma:	0.020	length-devi	-0.044						penalty-score:	4.75
bond	C19	to	C20	target_value:	1.390	d:	1.433	sigma:	0.020	length-devi	0.043						penalty-score:	4.67
bond	C1	to	C2	target_value:	1.390	d:	1.428	sigma:	0.020	length-devi	0.038						penalty-score:	3.70
bond	C4	to	C5	target_value:	1.490	d:	1.454	sigma:	0.020	length-devi	-0.036						penalty-score:	3.26
bond	C13	to	C14	target_value:	1.490	d:	1.456	sigma:	0.020	length-devi	-0.034						penalty-score:	2.91
bond	C15	to	C13	target_value:	1.490	d:	1.458	sigma:	0.020	length-devi	-0.032						penalty-score:	2.57
bond	C16	to	C15	target_value:	1.490	d:	1.459	sigma:	0.020	length-devi	-0.031						penalty-score:	2.45
angle	C13	-	C4	-	C5	target:	108.00	model_angle:	133.80	sigma:	3.00	angle-devi	25.80				penalty-score:	73.93
angle	O1	-	C5	-	C4	target:	108.00	model_angle:	126.59	sigma:	3.00	angle-devi	18.59				penalty-score:	38.38
angle	C13	-	C15	-	C16	target:	120.00	model_angle:	102.30	sigma:	3.00	angle-devi	17.70				penalty-score:	34.83
angle	O2	-	C6	-	N1	target:	108.00	model_angle:	122.80	sigma:	3.00	angle-devi	14.80				penalty-score:	24.34
angle	O2	-	C6	-	C3	target:	108.00	model_angle:	122.76	sigma:	3.00	angle-devi	14.76				penalty-score:	24.19
angle	C13	-	C15	-	C17	target:	120.00	model_angle:	133.33	sigma:	3.00	angle-devi	13.33				penalty-score:	19.76
angle	C4	-	C13	-	C15	target:	120.00	model_angle:	132.99	sigma:	3.00	angle-devi	12.99				penalty-score:	18.76
angle	N1	-	C5	-	O1	target:	108.00	model_angle:	120.48	sigma:	3.00	angle-devi	12.48				penalty-score:	17.32
angle	C15	-	C13	-	C14	target:	120.00	model_angle:	110.43	sigma:	3.00	angle-devi	-9.57				penalty-score:	10.18
angle	N1	-	C6	-	C3	target:	108.00	model_angle:	114.28	sigma:	3.00	angle-devi	6.28				penalty-score:	4.38
angle	C6	-	C3	-	C4	target:	108.00	model_angle:	101.75	sigma:	3.00	angle-devi	-6.25				penalty-score:	4.34

Residue Distortion Summary:

29 bond restraints

44 angle restraints

sum of bond distortions penalties: 59.5697

sum of angle distortions penalties: 300.405

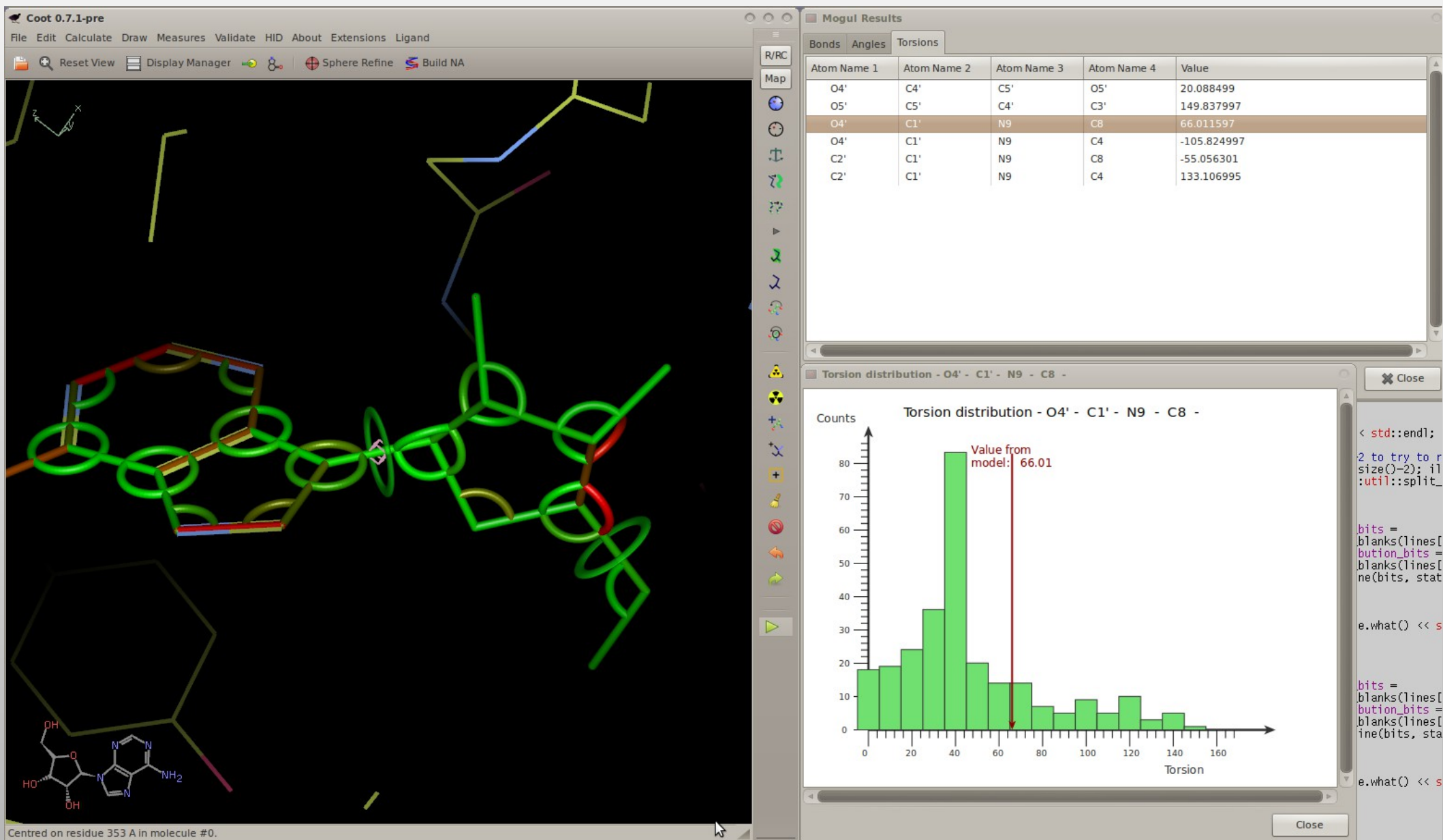
average bond distortion penalty: 2.05413

average angle distortion penalty: 6.82739

total distortion penalty: 405.304

average distortion penalty: 4.93116

Mogul Results Representation

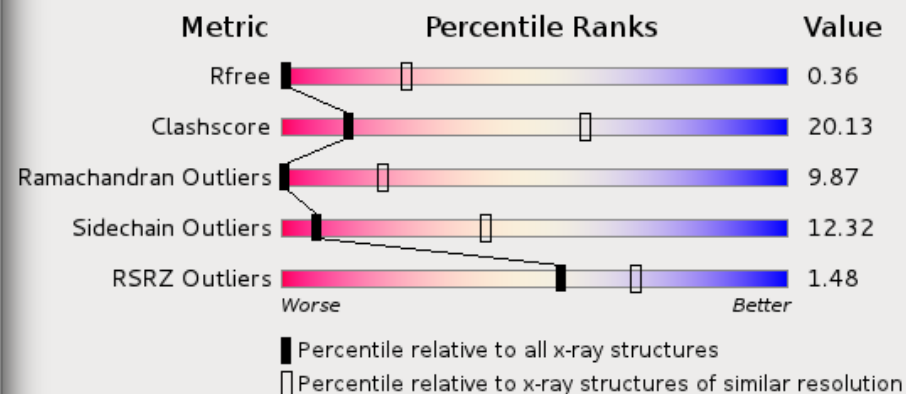


File Edit Calculate Draw Measures Validate HID About Extensions Ligand

Reset View Display Manager Ligand Builder Sphere Refine

R/R/C

Map



Bad RSRZ 0.573

Residue A 676 XNM:

Mogul-based Bond Outlier CAG,CAH, z = -5.11
Mogul-based Bond Outlier CAL,NAK, z = -2.45
Mogul-based Bond Outlier CAV,NAW, z = 2.64
Mogul-based Bond Outlier CBC,NBB, z = -16.67
Mogul-based Angle Outlier CAF,CAG,CAD, z = 2.16
Mogul-based Angle Outlier CAG,CAH,NAI, z = 2.97
Mogul-based Angle Outlier CAH,NAI,CAJ, z = 7.12
Mogul-based Angle Outlier NAR,CAJ,NAI, z = -9.85
Mogul-based Angle Outlier CAP,CAQ,NAR, z = -4.47
Mogul-based Angle Outlier CAQ,NAR,CAJ, z = 10.16
Mogul-based Angle Outlier OAO,CAV,NAW, z = -2.68
Mogul-based Angle Outlier CAU,CAV,NAW, z = 2.96
Mogul-based Angle Outlier CBC,NBB,CAY, z = 2.70
Mogul-based Angle Outlier CBC,NBB,CBA, z = 4.48
Clash atom HAQ score: 1.10
Clash atom HAQ score: 0.53
Clash atom CAZ score: 0.88
Clash atom CAJ score: 0.56
Clash atom CAN score: 0.92
Clash atom HAN score: 1.08

Close

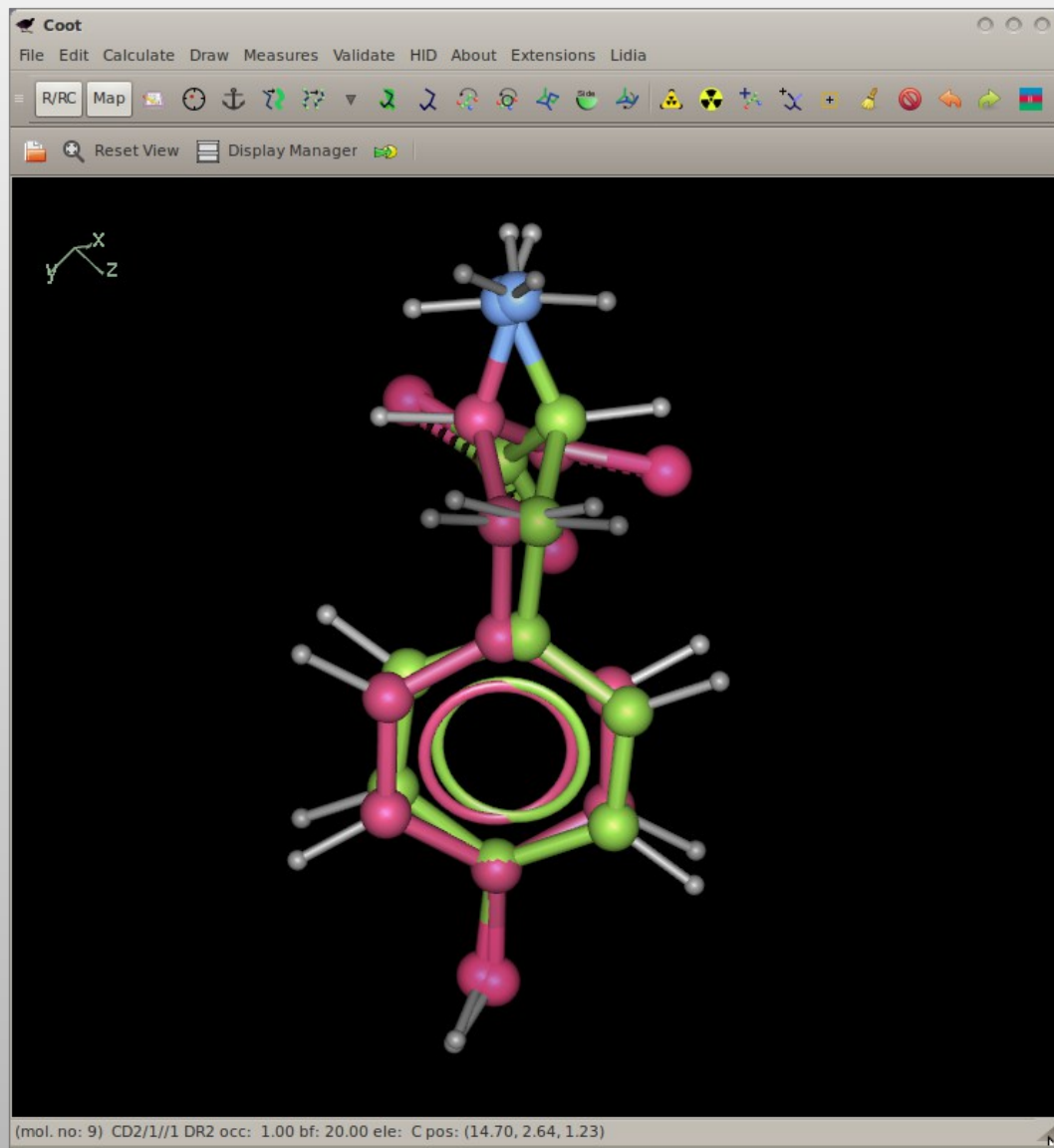
(mol. no: 0) NAK/1/A/676 XNM occ: 1.00 bf: 299.85 ele: N pos: (89.50,-24.64,51.94)

Ligand Representation

- Bond orders (from dictionary restraints)



Chiral Centre Inversion



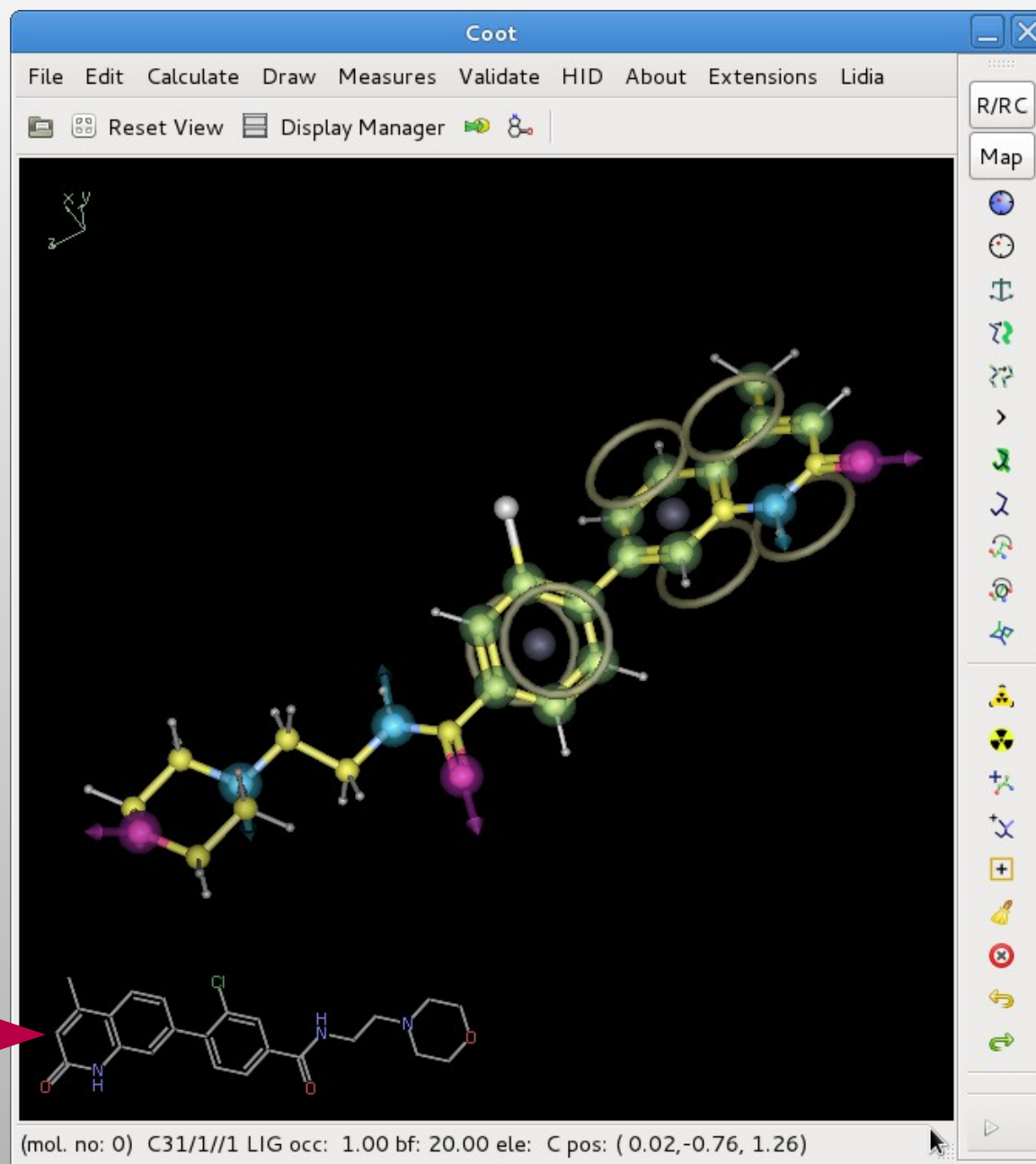
Inverted chiral centre
refinement pathology
detection

Hydrogen tunnelling

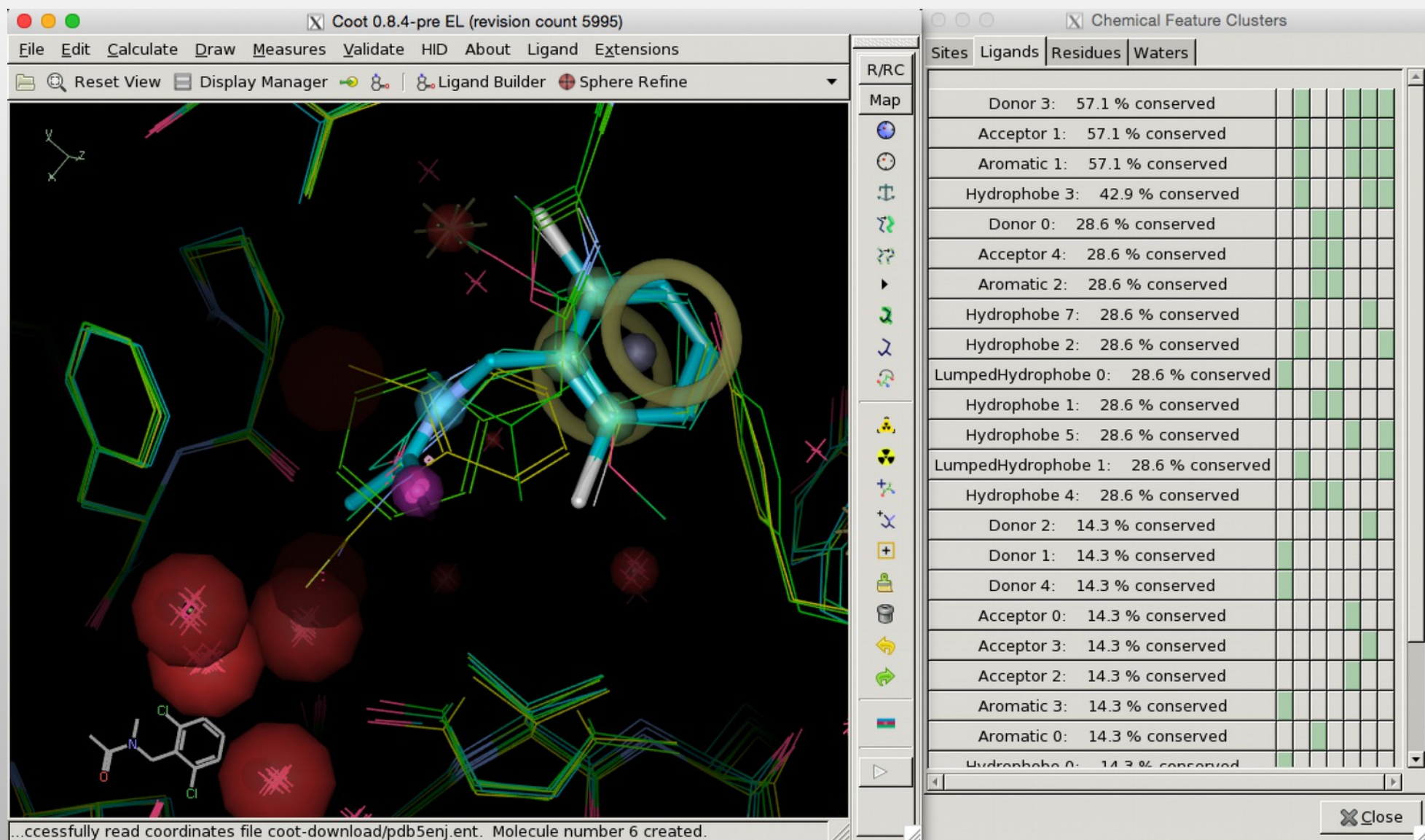
Chemical Features

Uses built-in
FeatureFactory

...and on the fly
thumbnailing

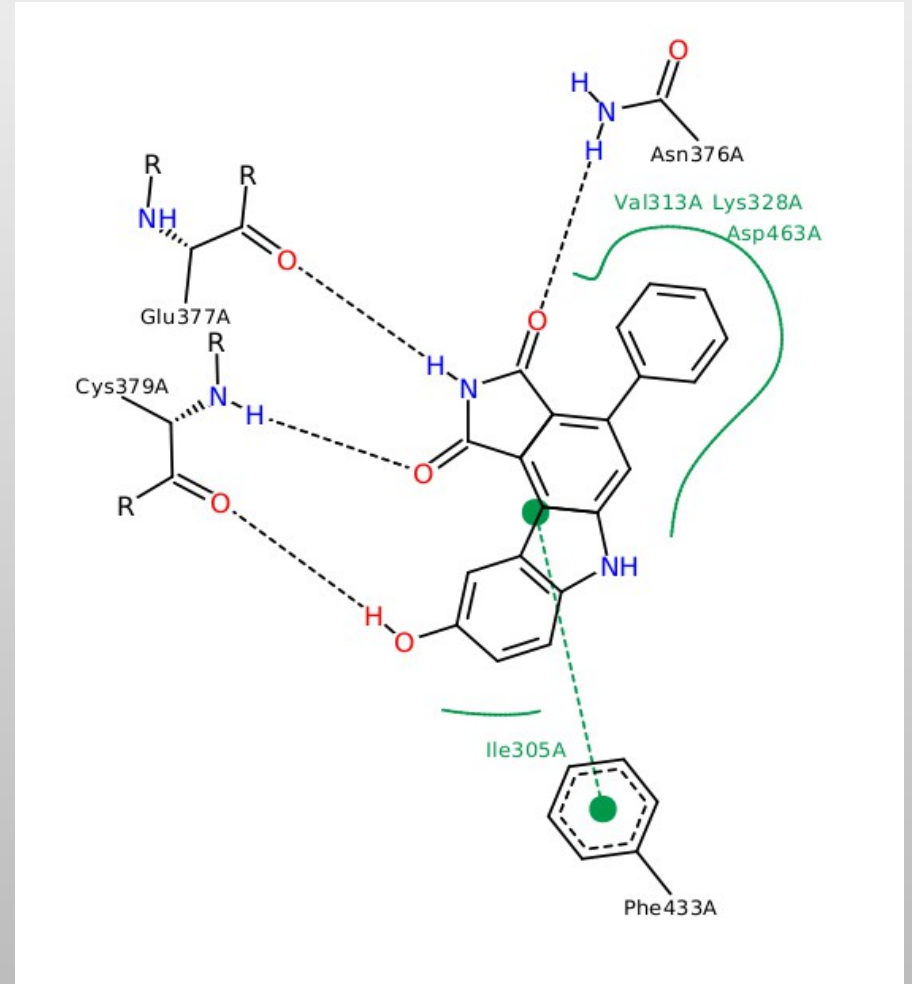
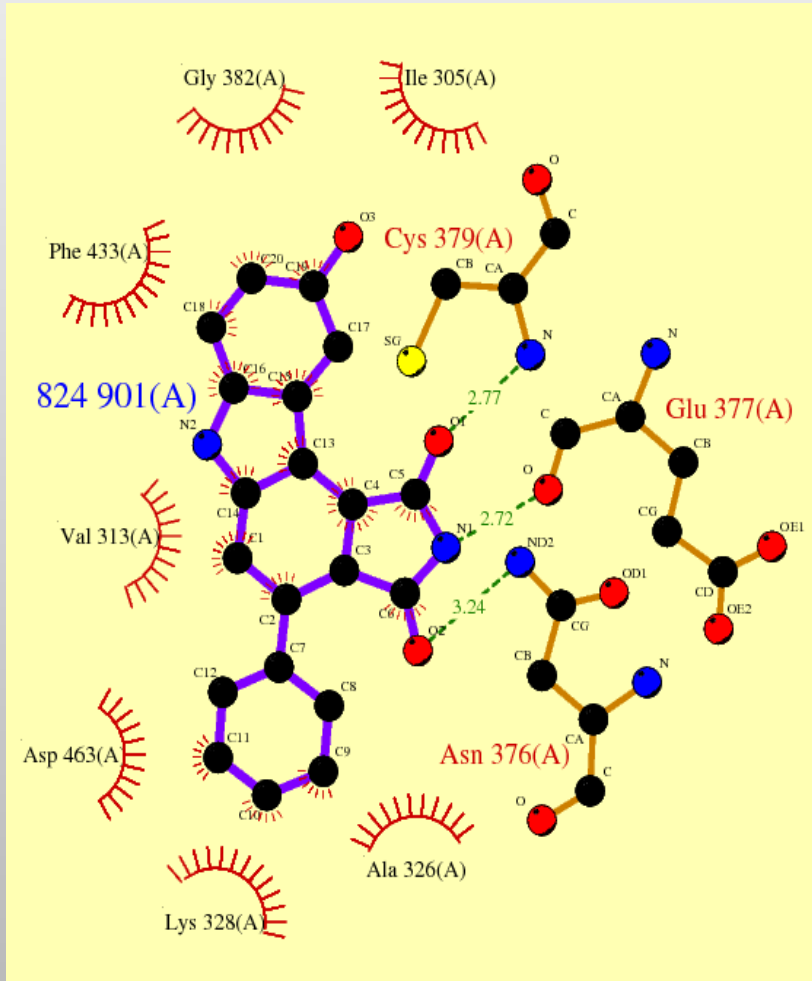


Conserved Pharmacophores



Ligand Environment Layout

- 2d Ligand pocket layout (ligplot, poseview)



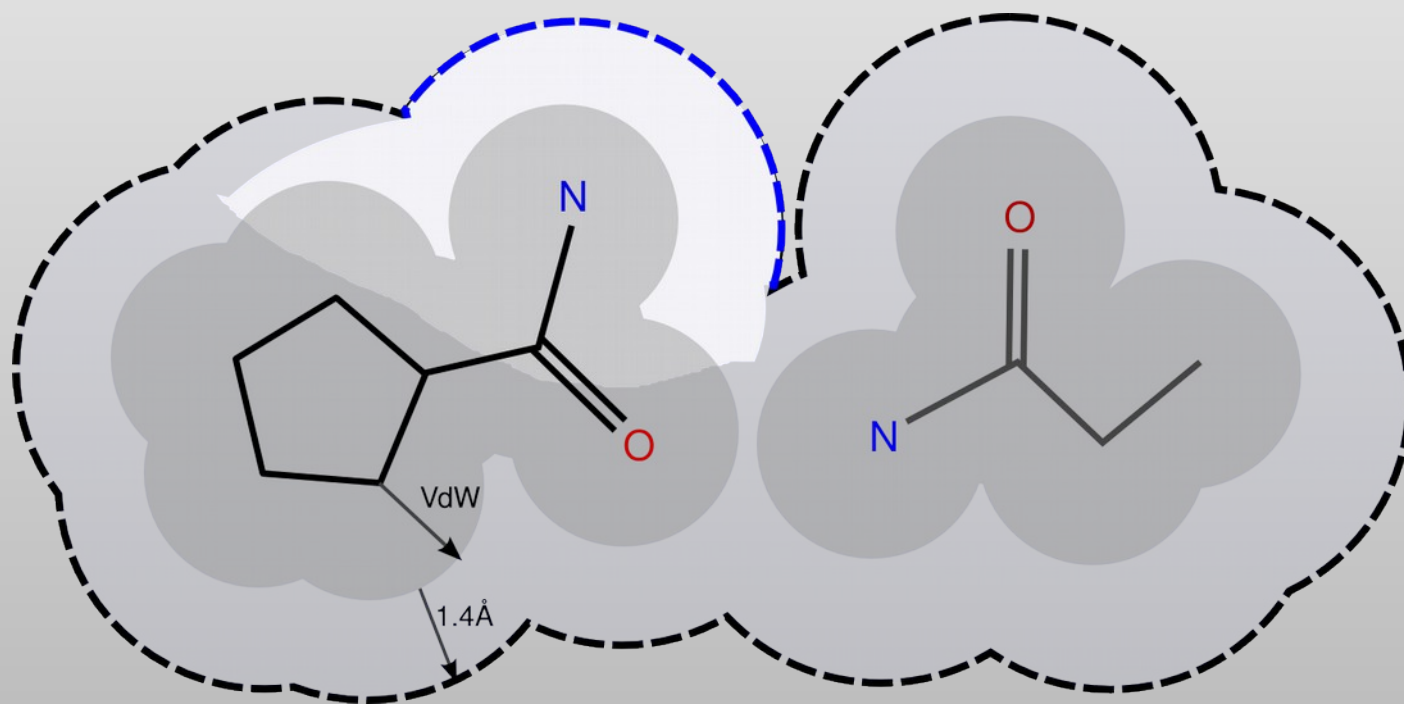
Can we do better? - Interactivity?

Ligand Environment Layout

- Binding pocket residues
- Interactions
- Substitution contour
- Solvent accessibility halos
- Solvent exclusion by ligand

Solvent Exposure

- Identification of solvent accessible atoms



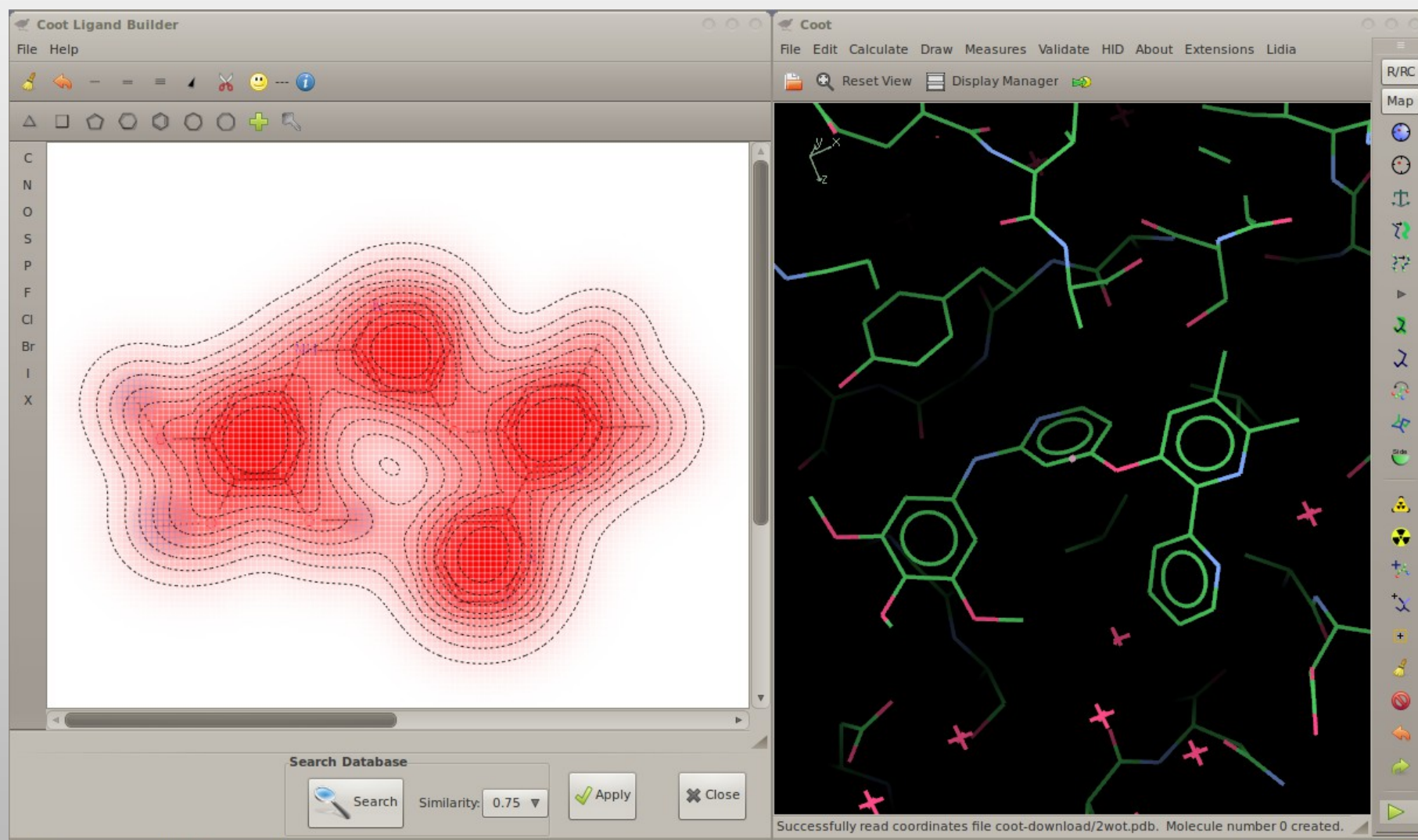
Ligand Environment Layout

- Considerations
 - 2D placement and distances should reflect 3D metrics (as much as possible)
 - H-bonded residues should be close the atoms to which they are bonded
 - Residues should not overlap the ligand
 - Residues should not overlap each other
 - *c.f.* Clark & Labute (2007)

Layout Energy Terms

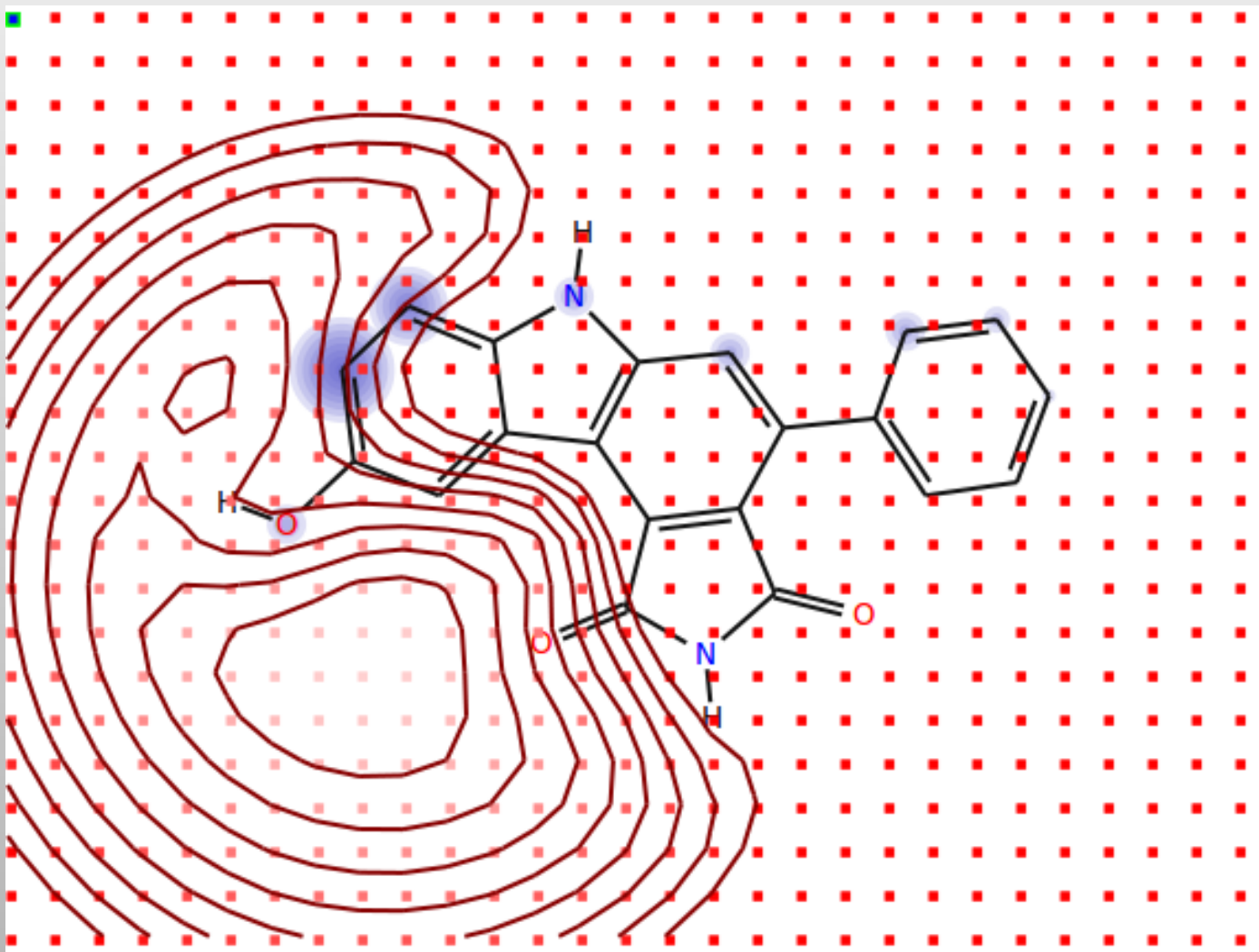
$$\begin{aligned} E = & \sum \sum w_{ij} (d_{ij}^2 - D_{ij}^2) + && \text{Residues match 3D Distances} \\ & \sum \sum \exp\left(-\frac{1}{2}d_{ij}^2\right) + && \text{Residues don't overlay each other} \\ & \sum \sum (d_{ik}^2 - D_{ik}^2) + && \text{Residues are close to H-bonding ligand atoms} \\ & \sum \sum \exp\left(-\frac{1}{2}d_{ik}^2\right) && \text{Residues don't overlap ligand} \end{aligned}$$

"Don't overlap the ligand"



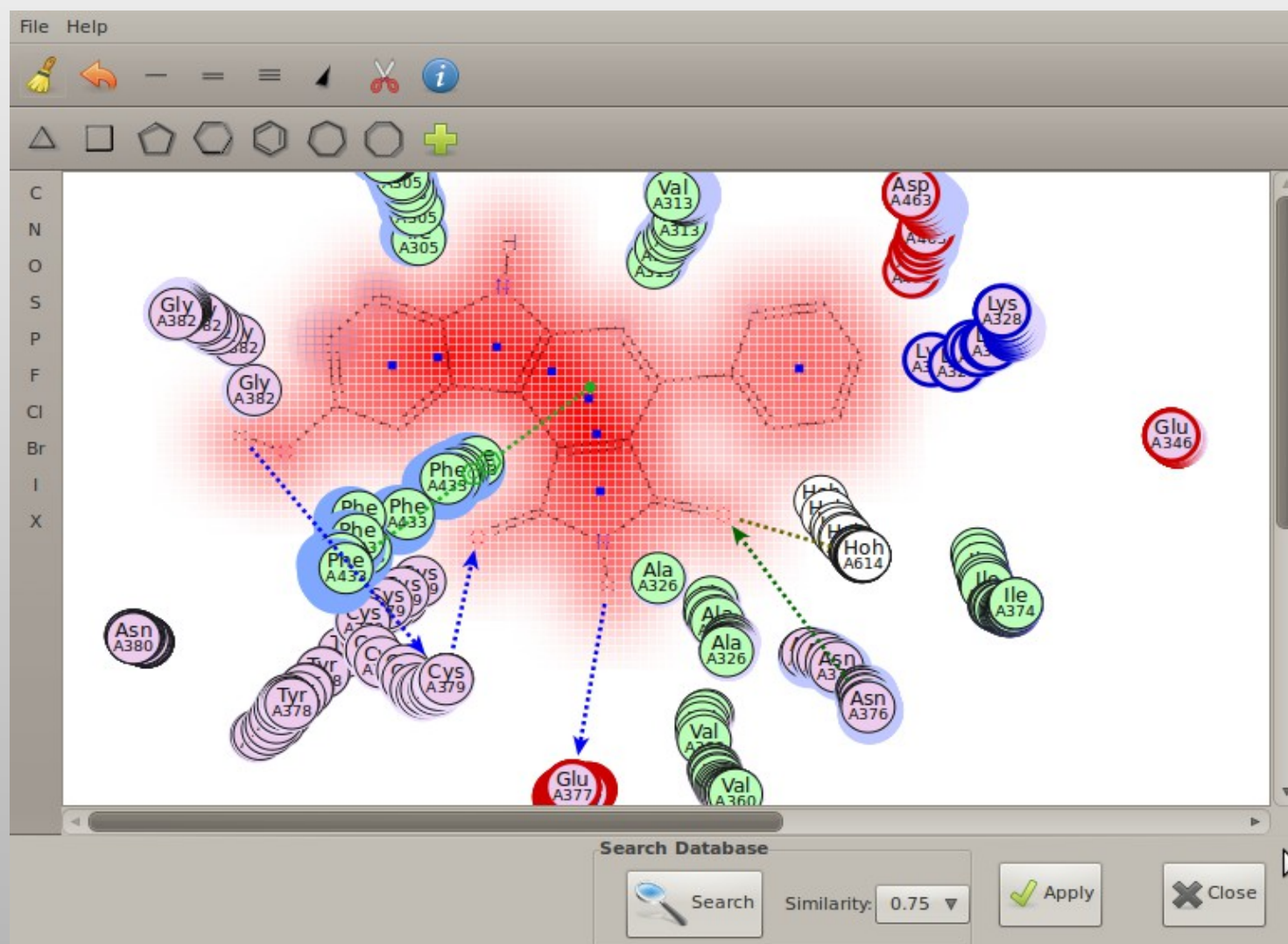
Ligand Environment Layout

- Initial residue placement



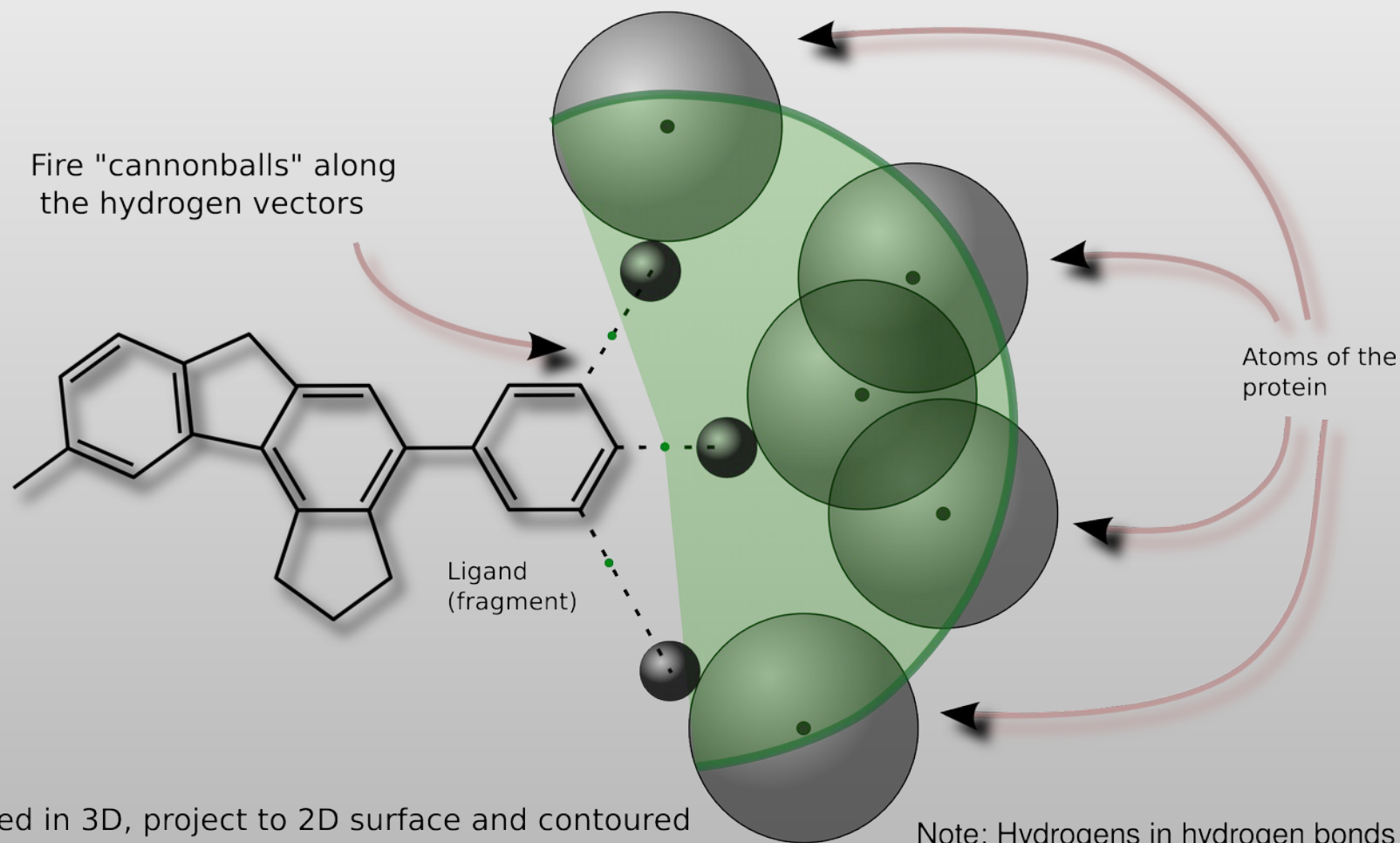
Ligand Environment Layout

- Residue position minimisation



Determination of the Substitution Contour

How far can we go (in the direction of the hydrogens) before hitting atoms of the protein?

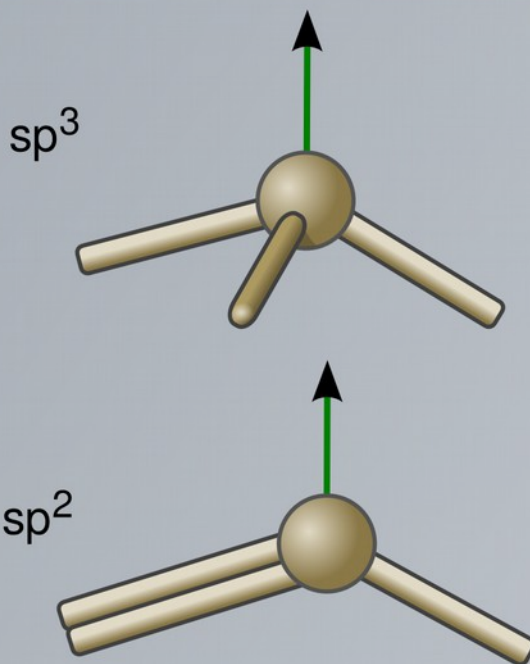


Determined in 3D, project to 2D surface and contoured

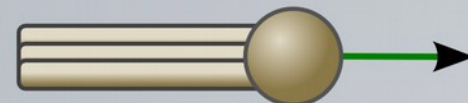
c.f. Clarke & Labute (2007)

Substitution Contour: Extending along Hydrogens

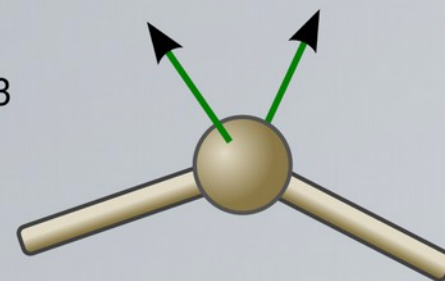
Riding Hydrogens



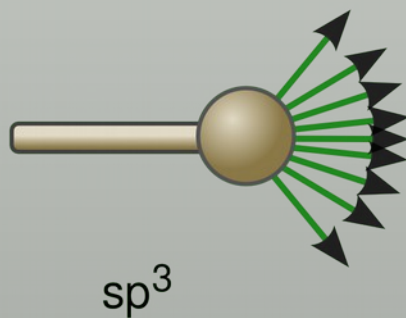
sp



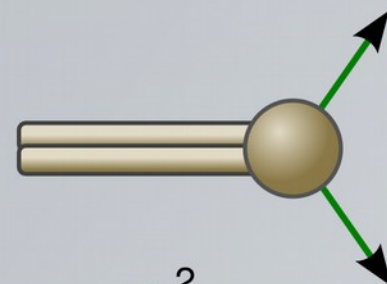
sp^3

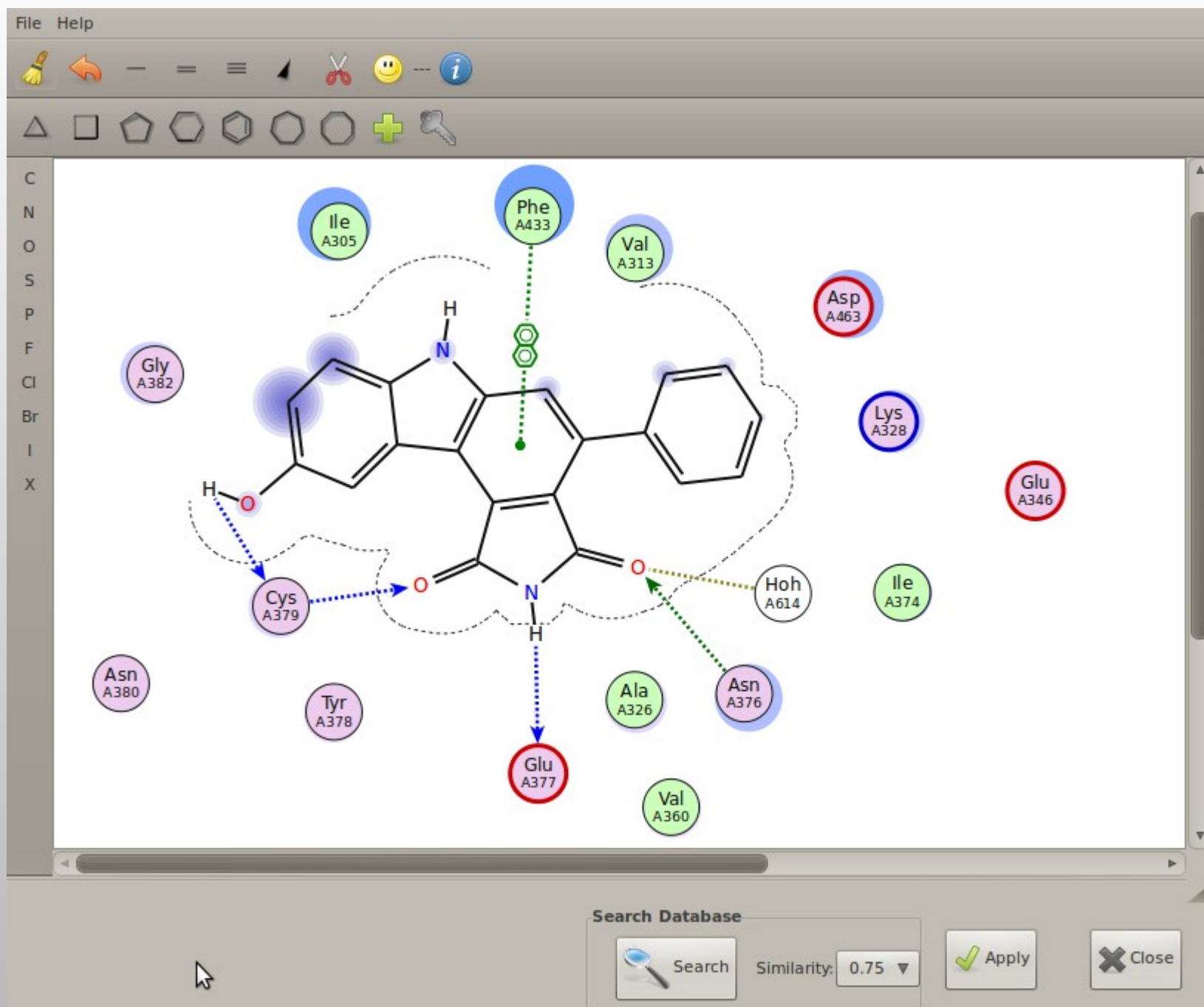


Torsionable Hydrogens
(test multiple directions)

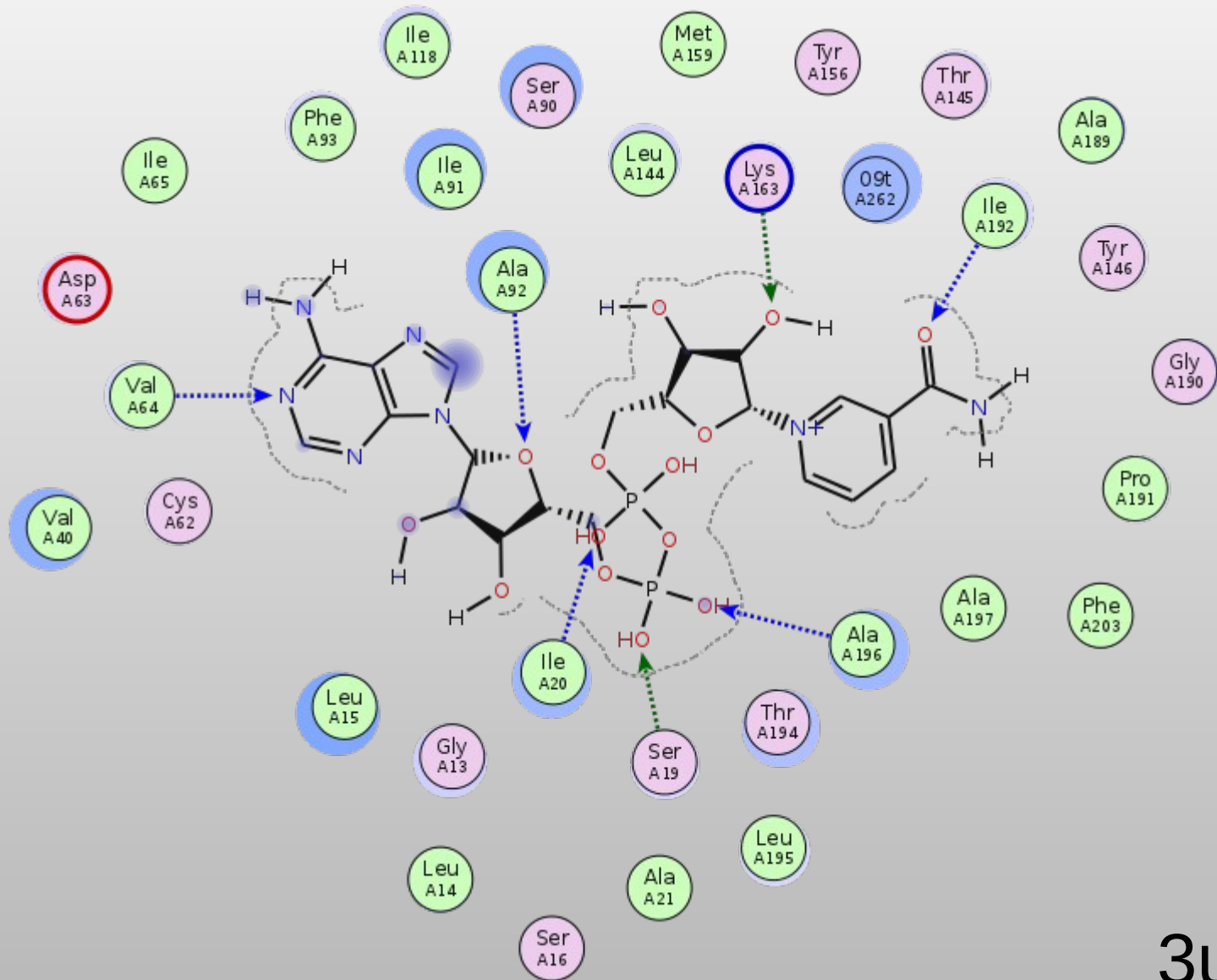


sp^2

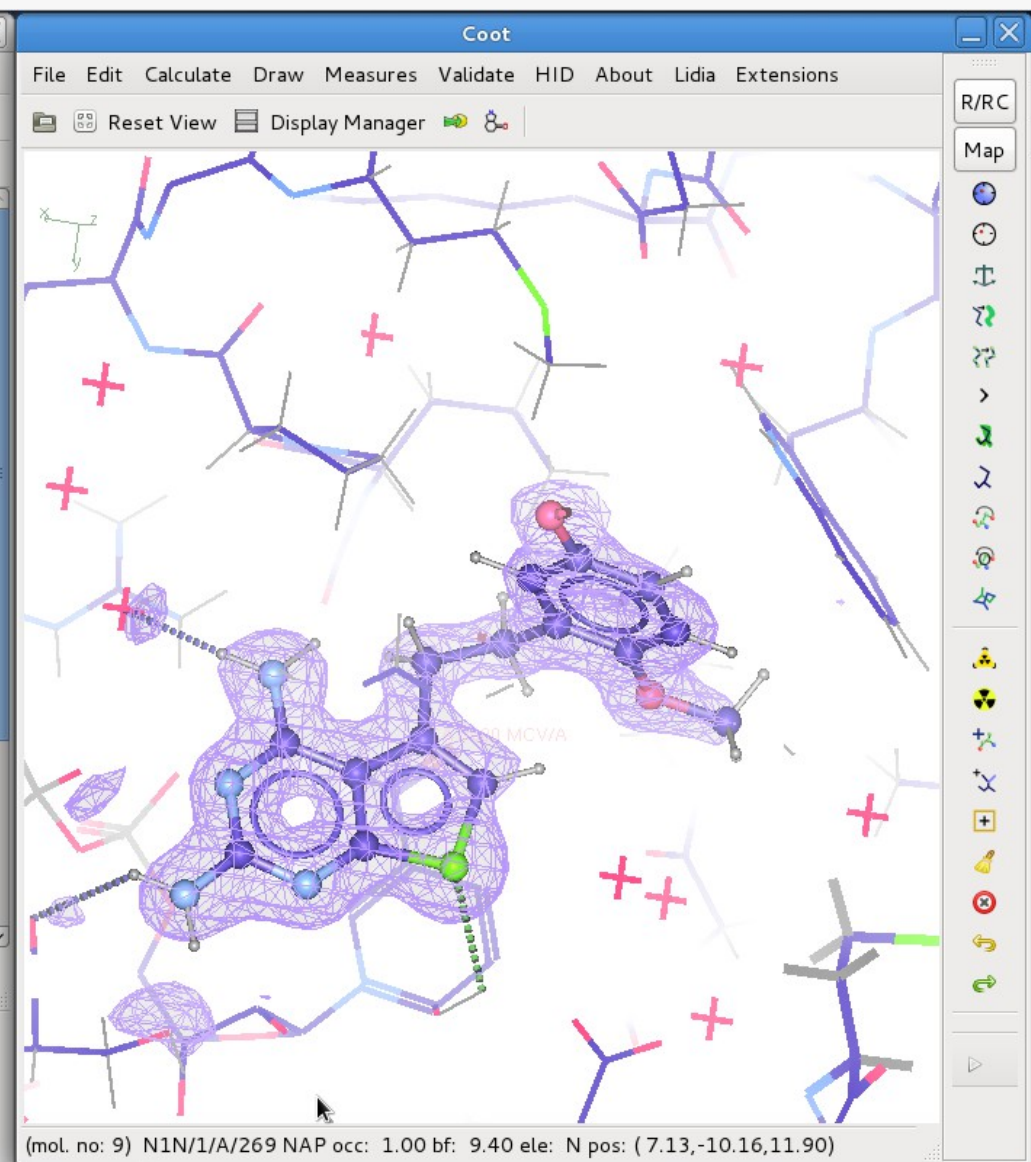
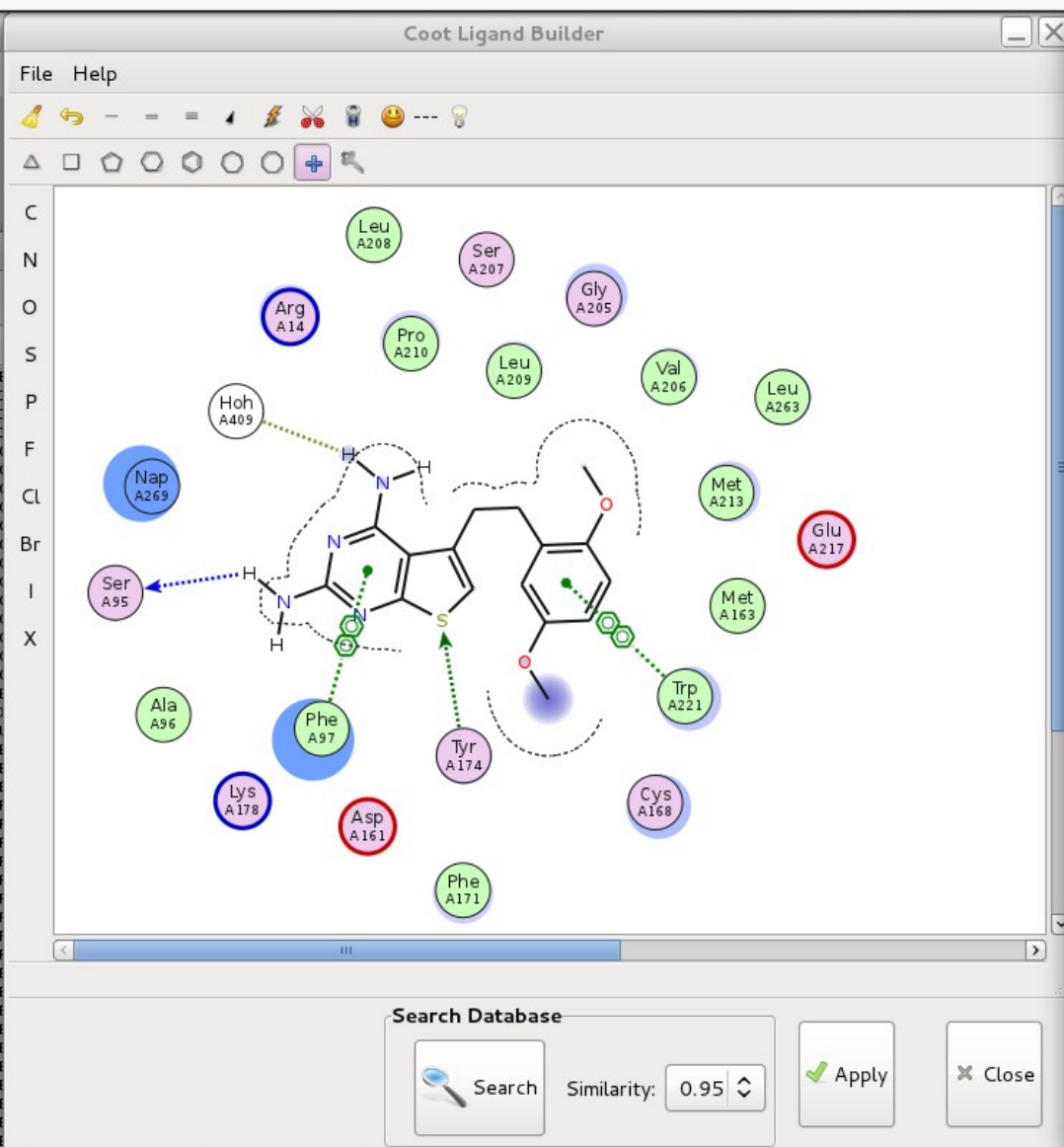




Layout Examples



3uic

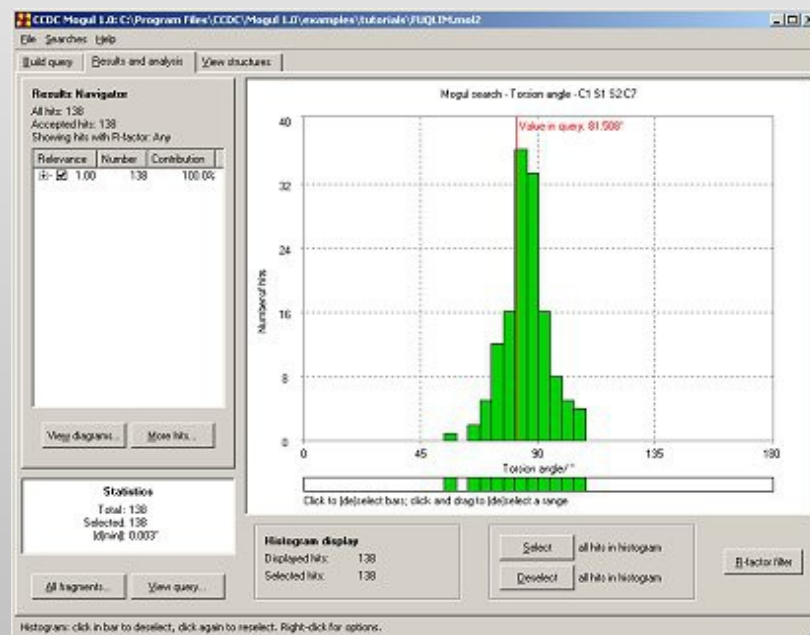


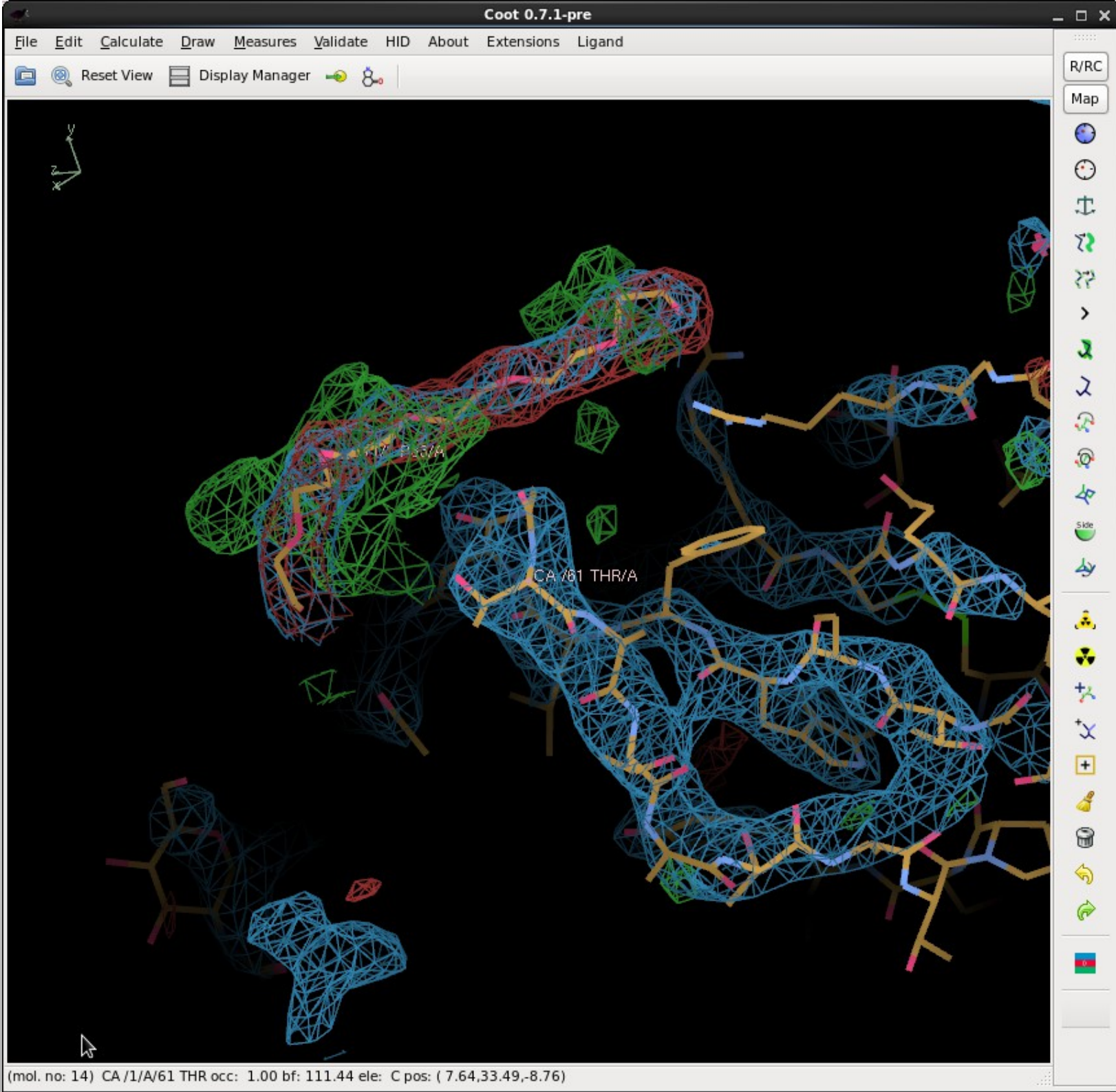
Scoring Protein-Ligand Complexes

- Score all PDB protein-ligand complexes
 - No covalent link to protein
 - No alt confs
 - Hetgroups with more than 6 atoms
- Score:
 - Correlation of maps: omit vs calculated
 - around the ligand
 - Mogul distortion
 - z-worst
 - Clash-score
 - c.f. Molprobability tool

Assessing Ligand Geometry Accuracy

- CSD's Mogul
- Knowledge-base of geometric parameters based on the CSD
- Can be run as a “batch job”
- Mean, median, mode, quartiles, Z-scores.

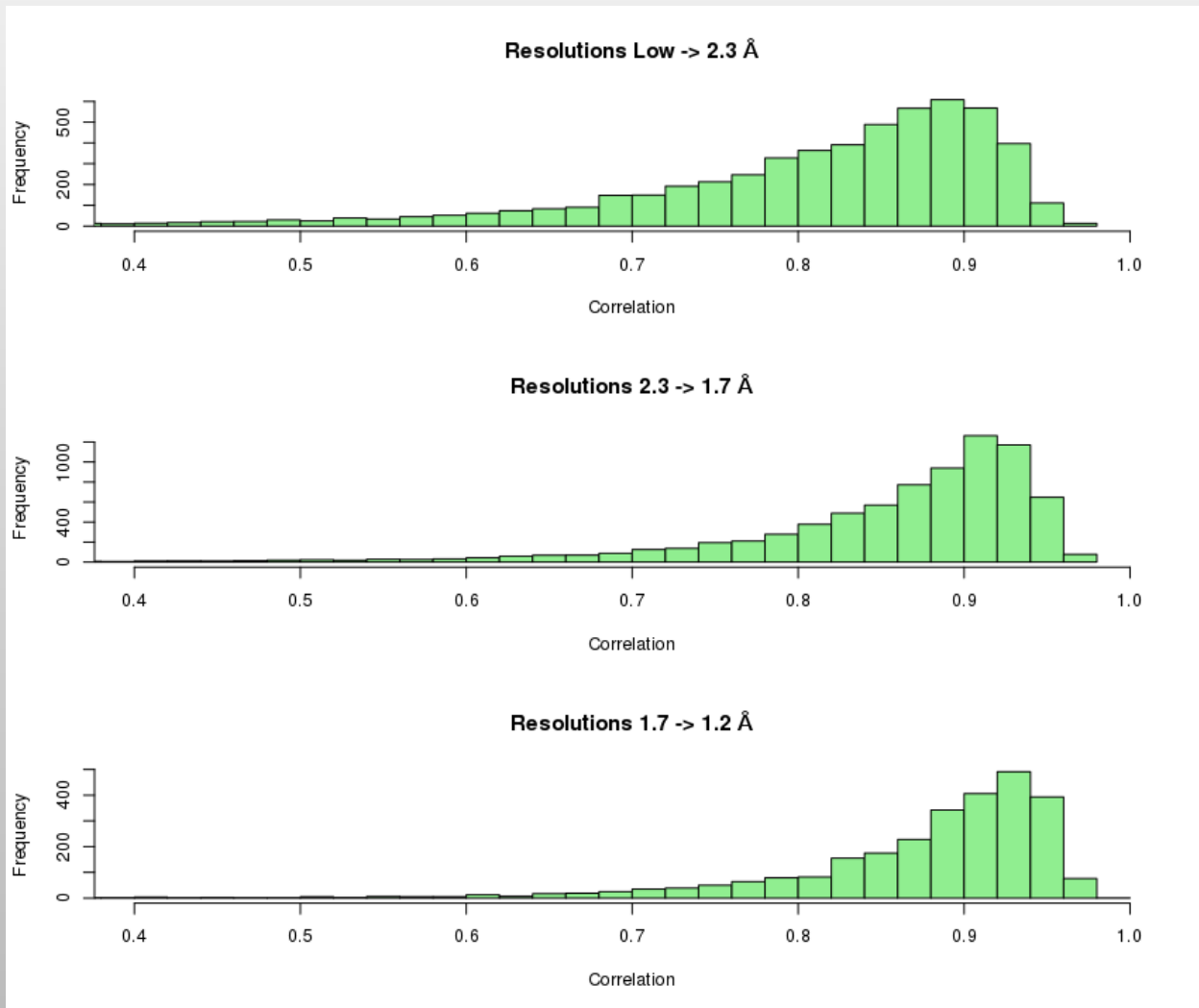




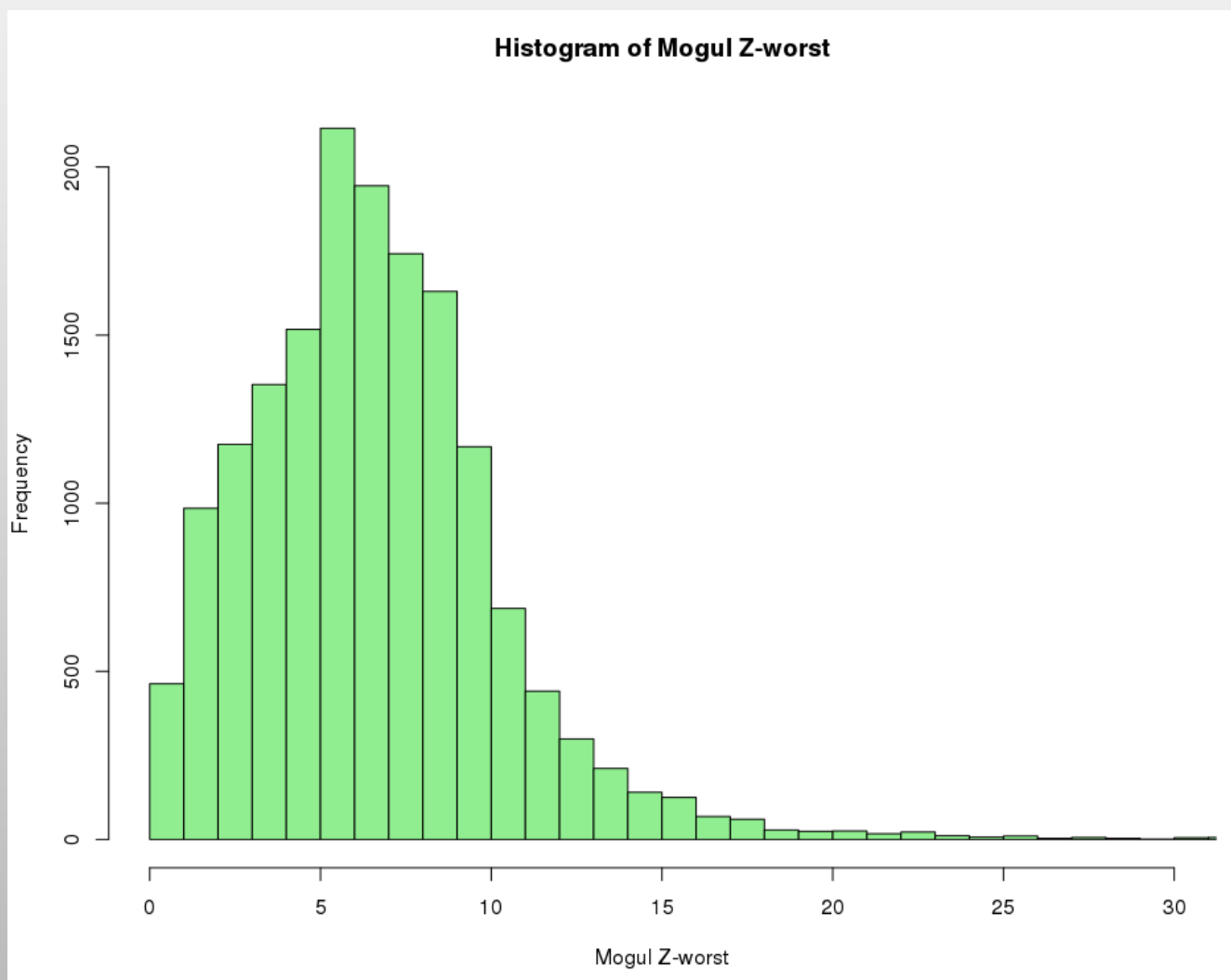
Score Histograms

- Density Correlations
- Mogul z-score
- # Bumps/ligand

Resolution dependence of Density Correlation



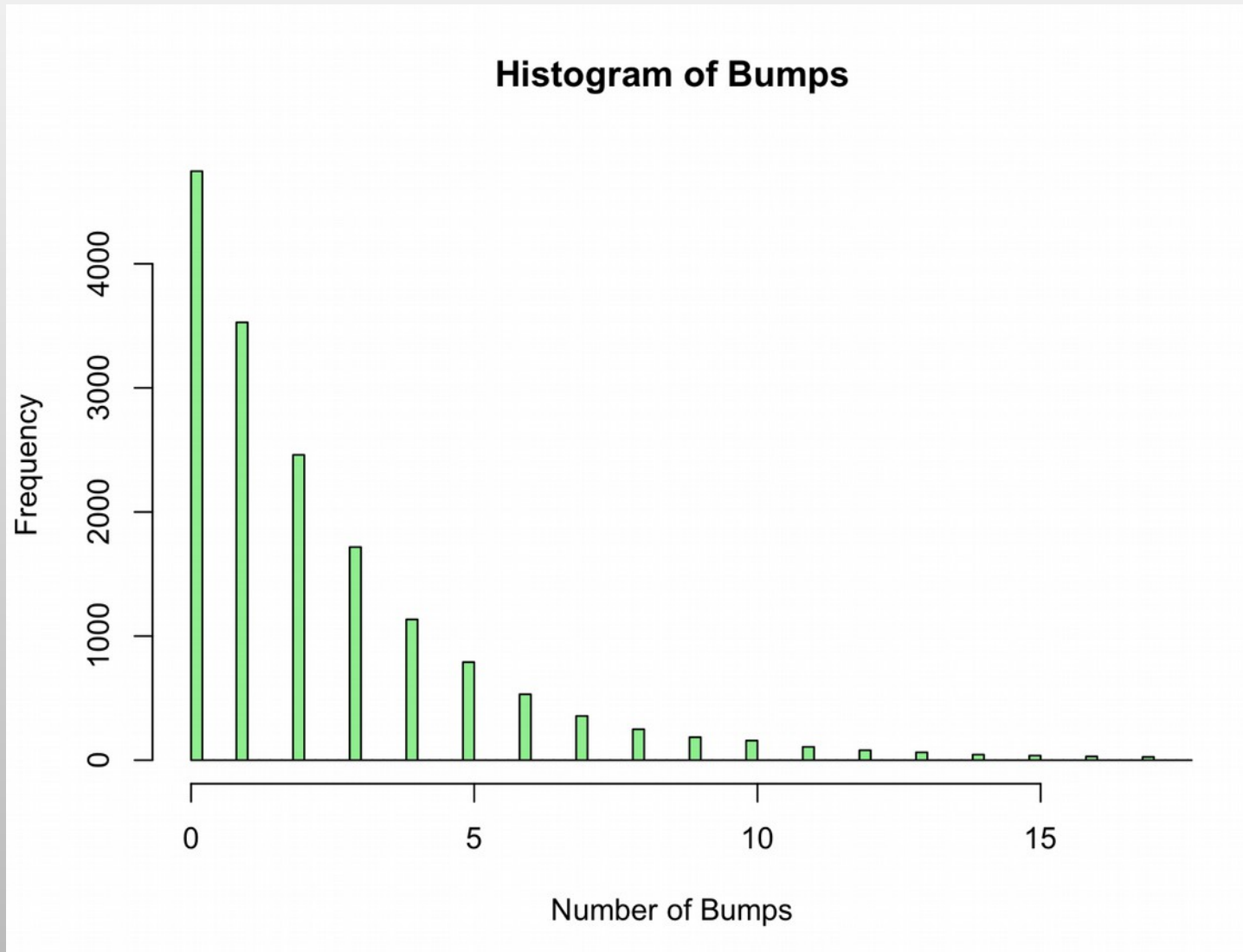
Overall Histogram of Mogul Z-worst of wwPDB Ligands



Resolution Dependence of Mogul Z-worst

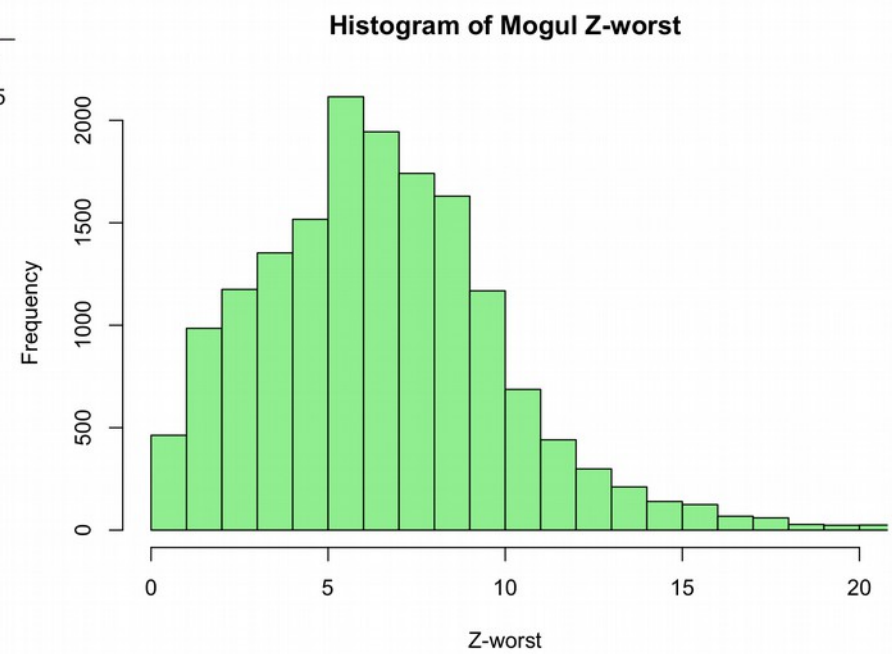
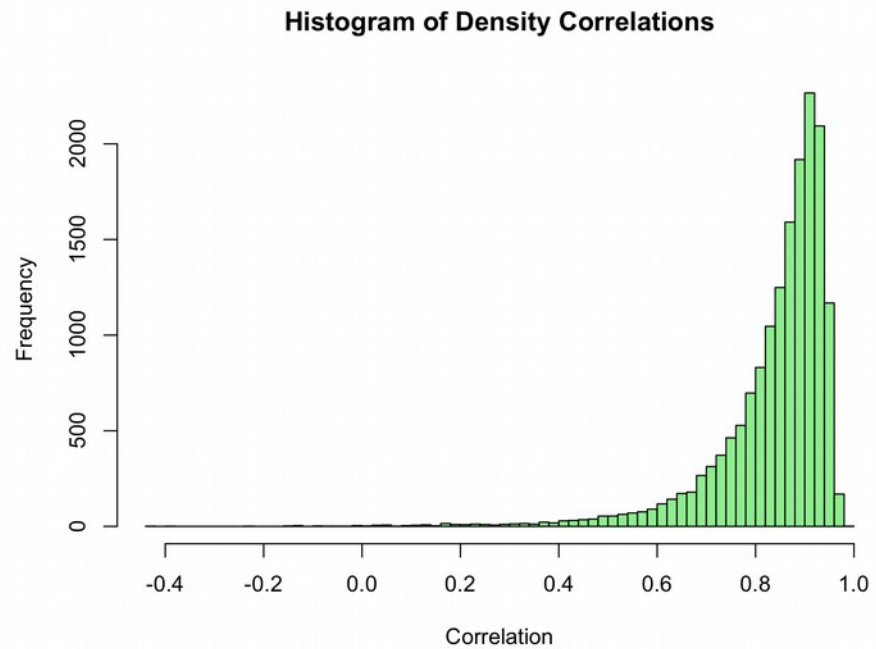
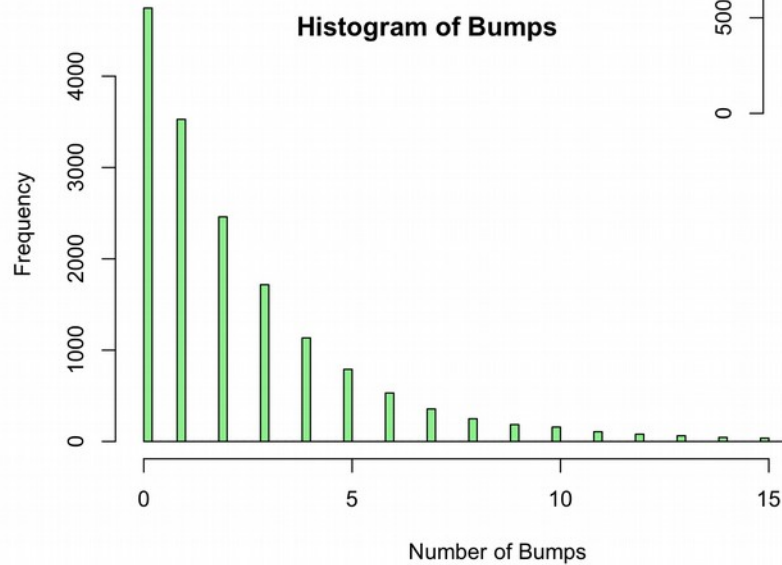


Histogram of Bad Contacts



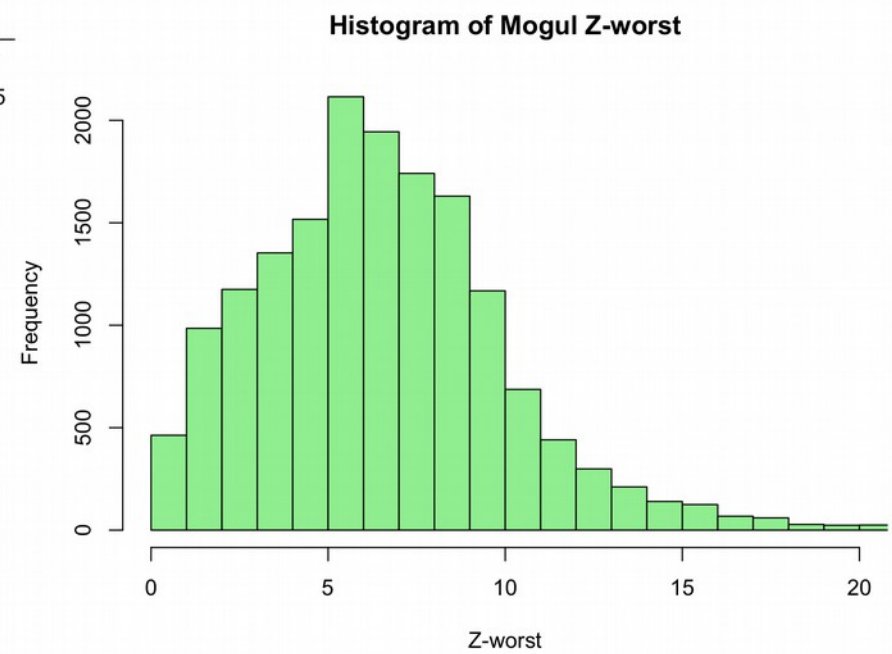
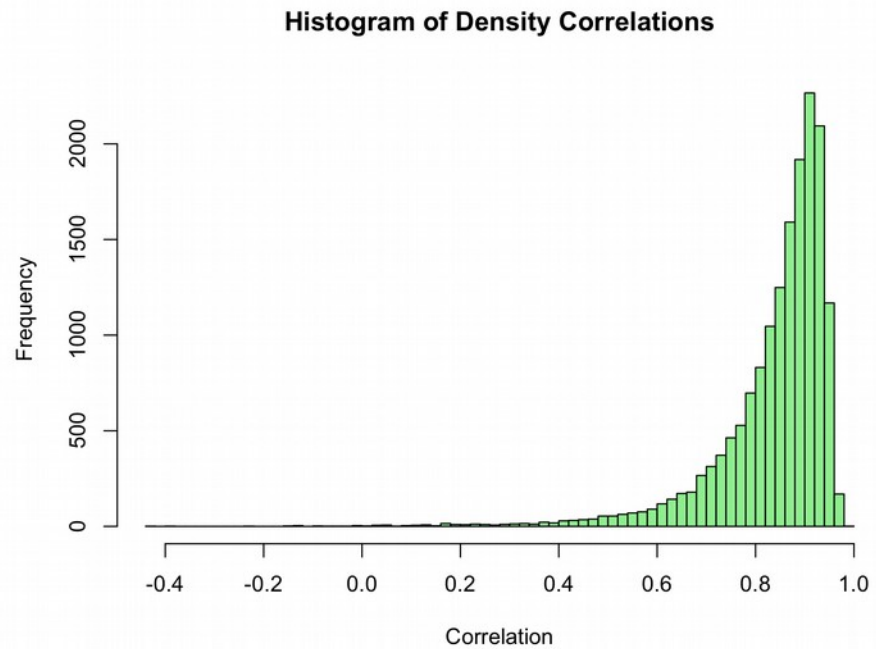
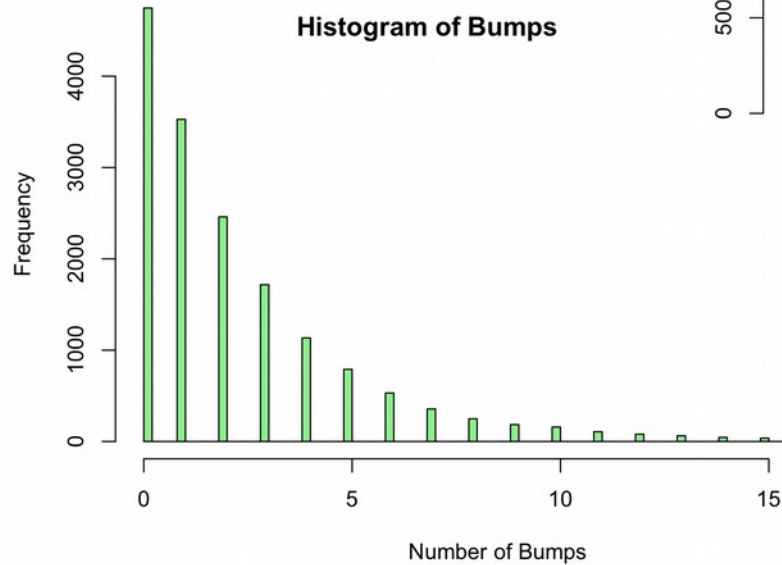
Ligand Scoring

recommendations...

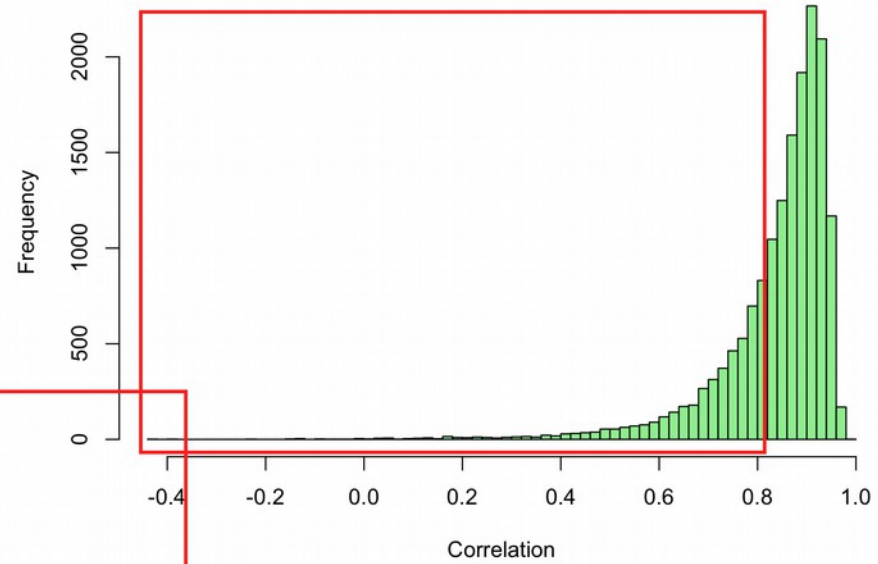


Scoring Ligands: To Be Better Than The Median:

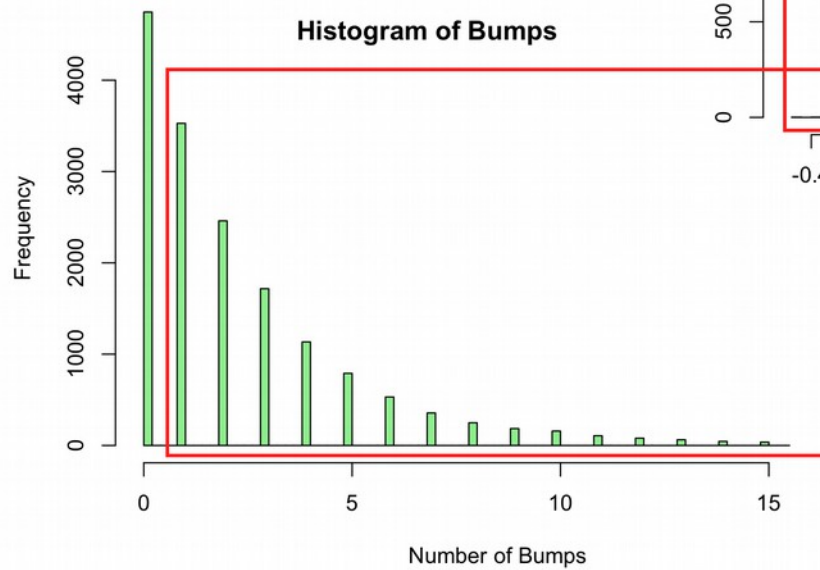
- 1 or 0 bumps
- Mogul $z(\text{worst}) < 6.3$
- Density correlation > 0.88



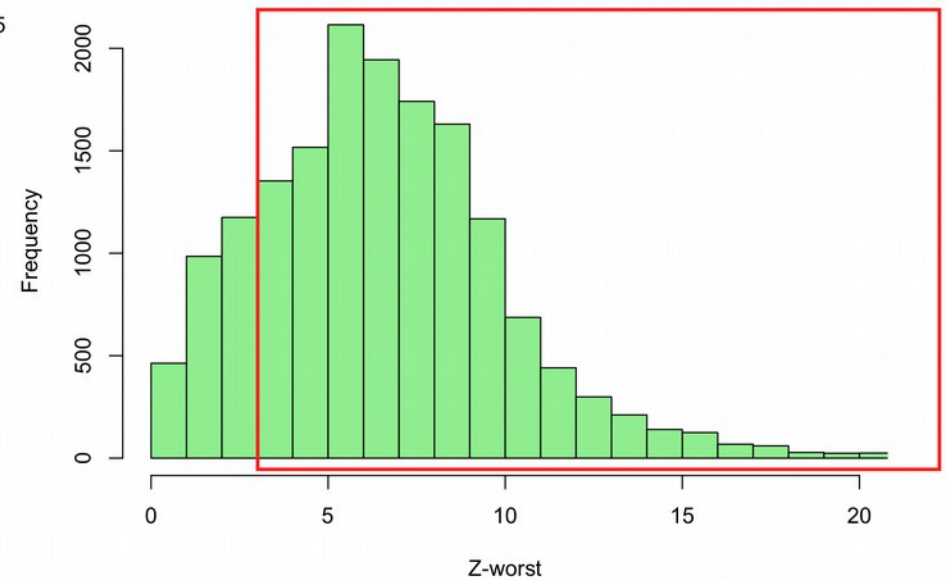
Histogram of Density Correlations



Histogram of Bumps



Histogram of Mogul Z-worst

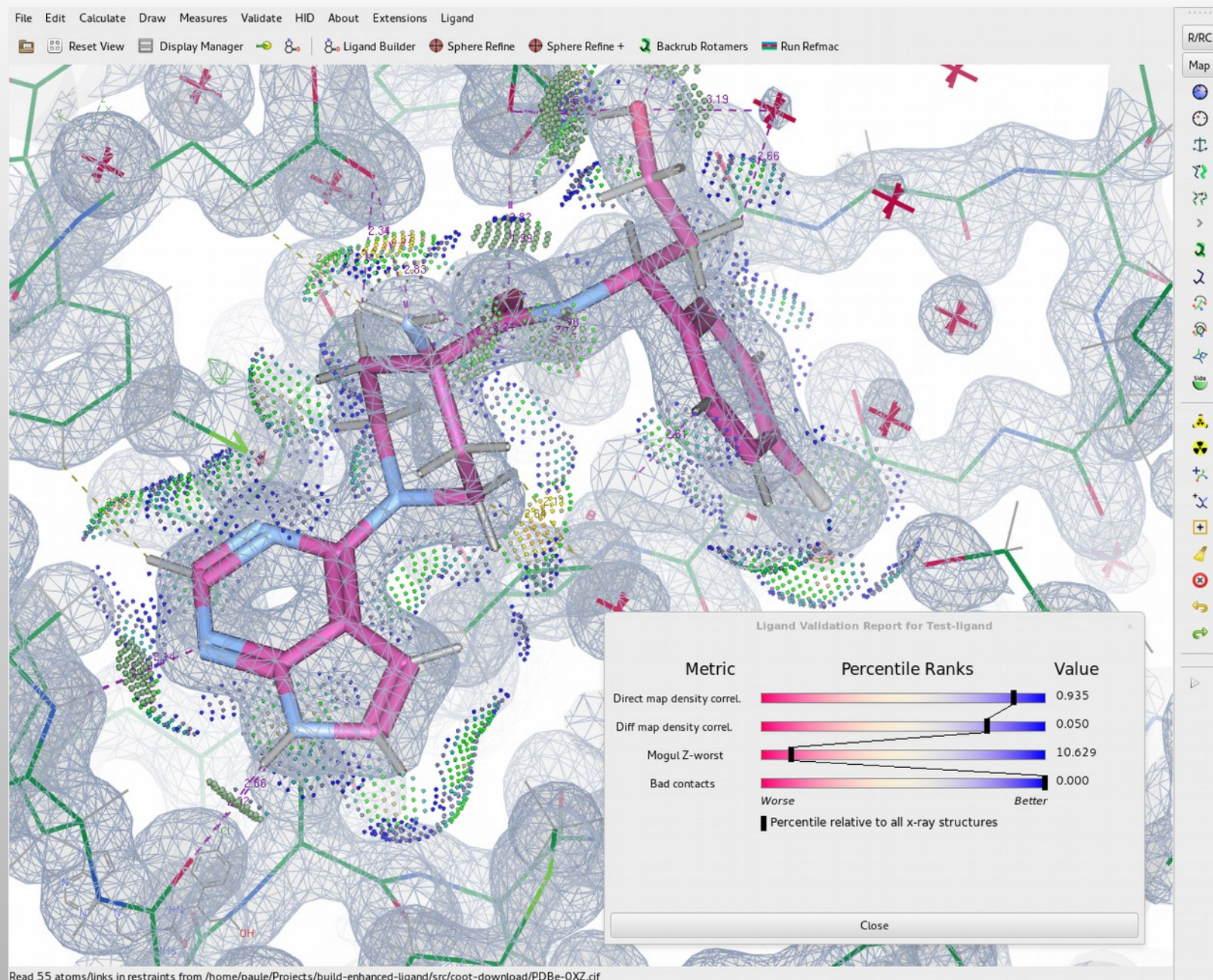


Sliders

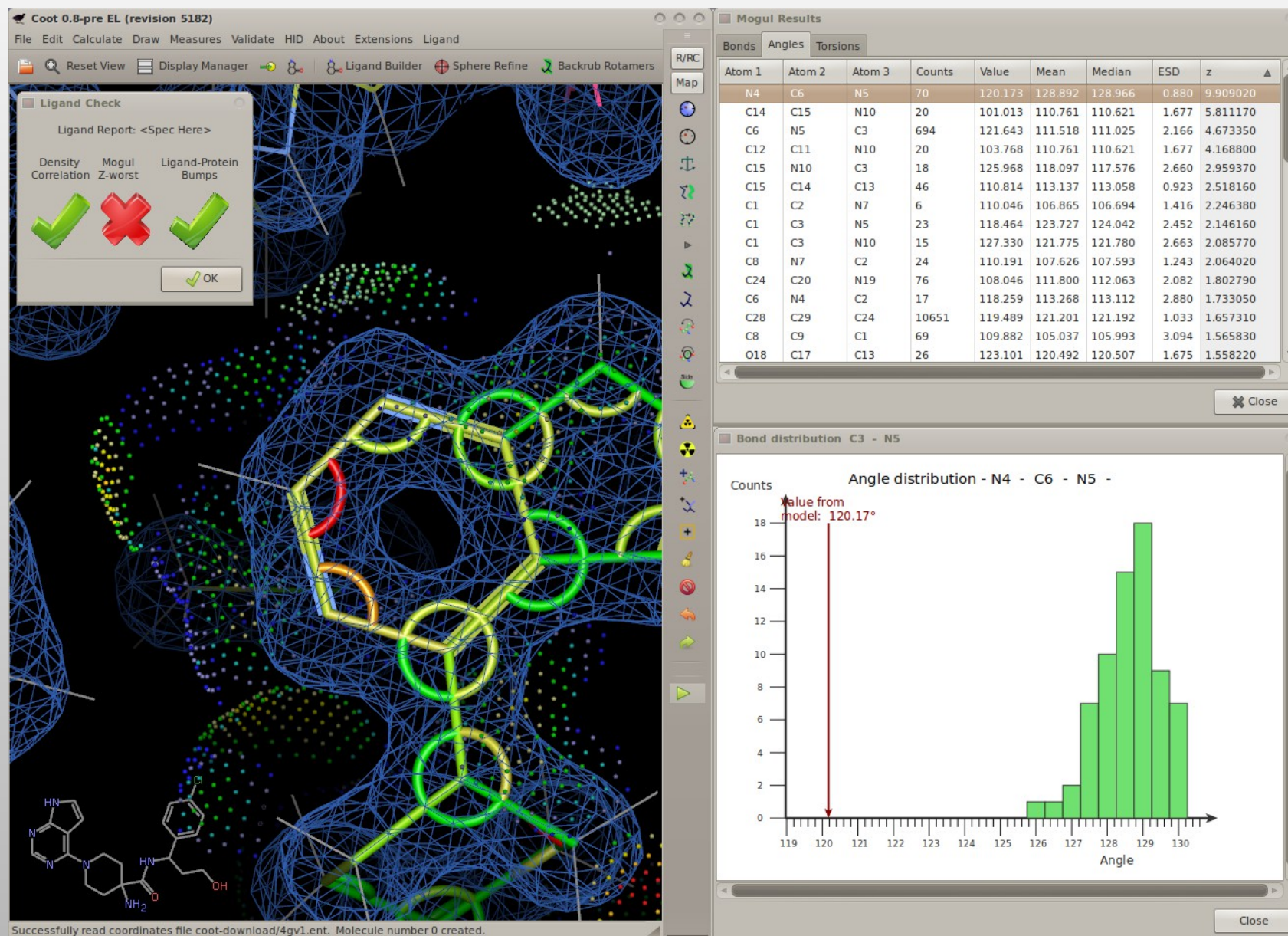
or

Yes/No?

Ligand Validation Sliders



Coot Ligand Validation Metrics Screenshot



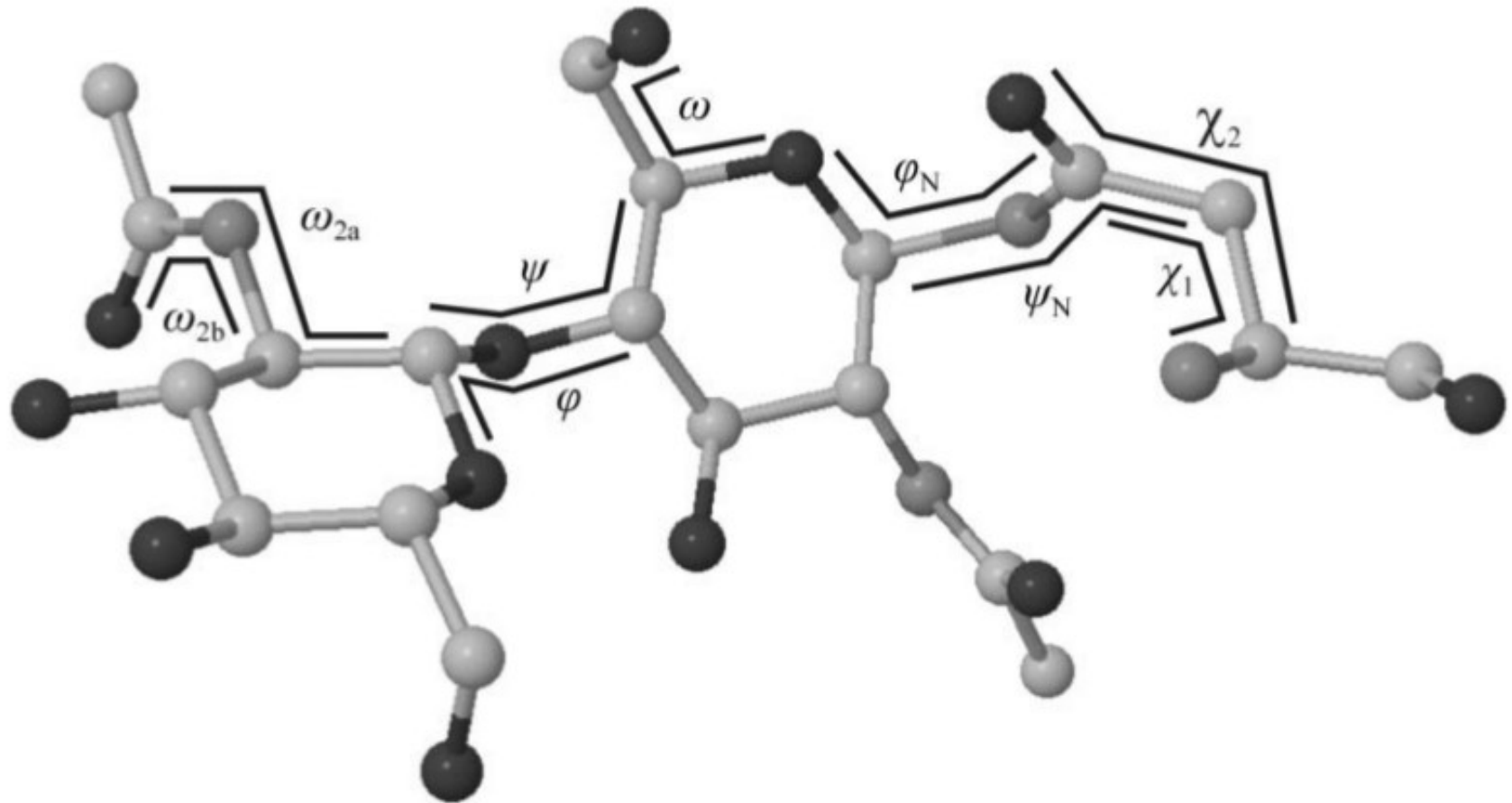
Modelling Carbohydrates

- Validation,
- Model-building,
- Refinement

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

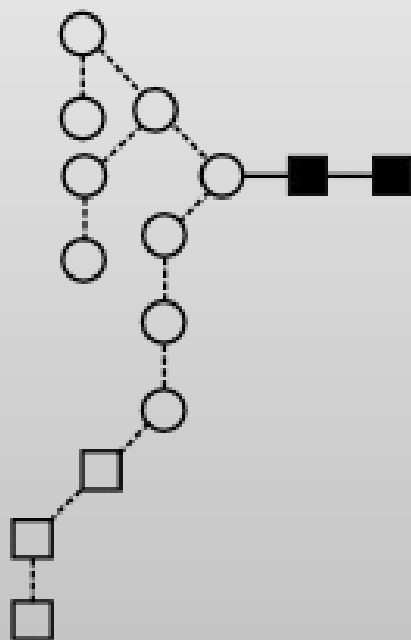
Carbohydrate Links



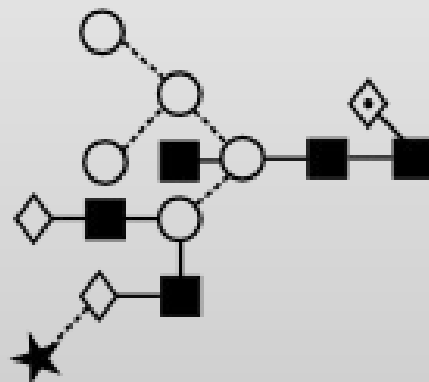
Thomas Lütteke (2007)

Validate the Tree: N-linked carbohydrates

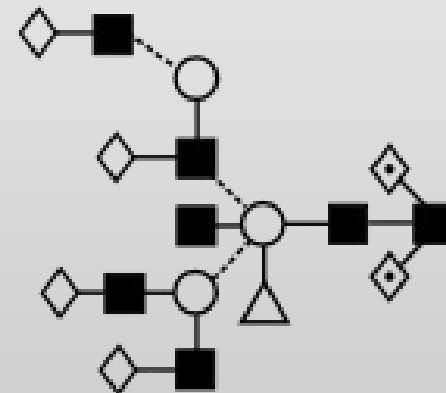
"Oligomannose"



"Hybrid"

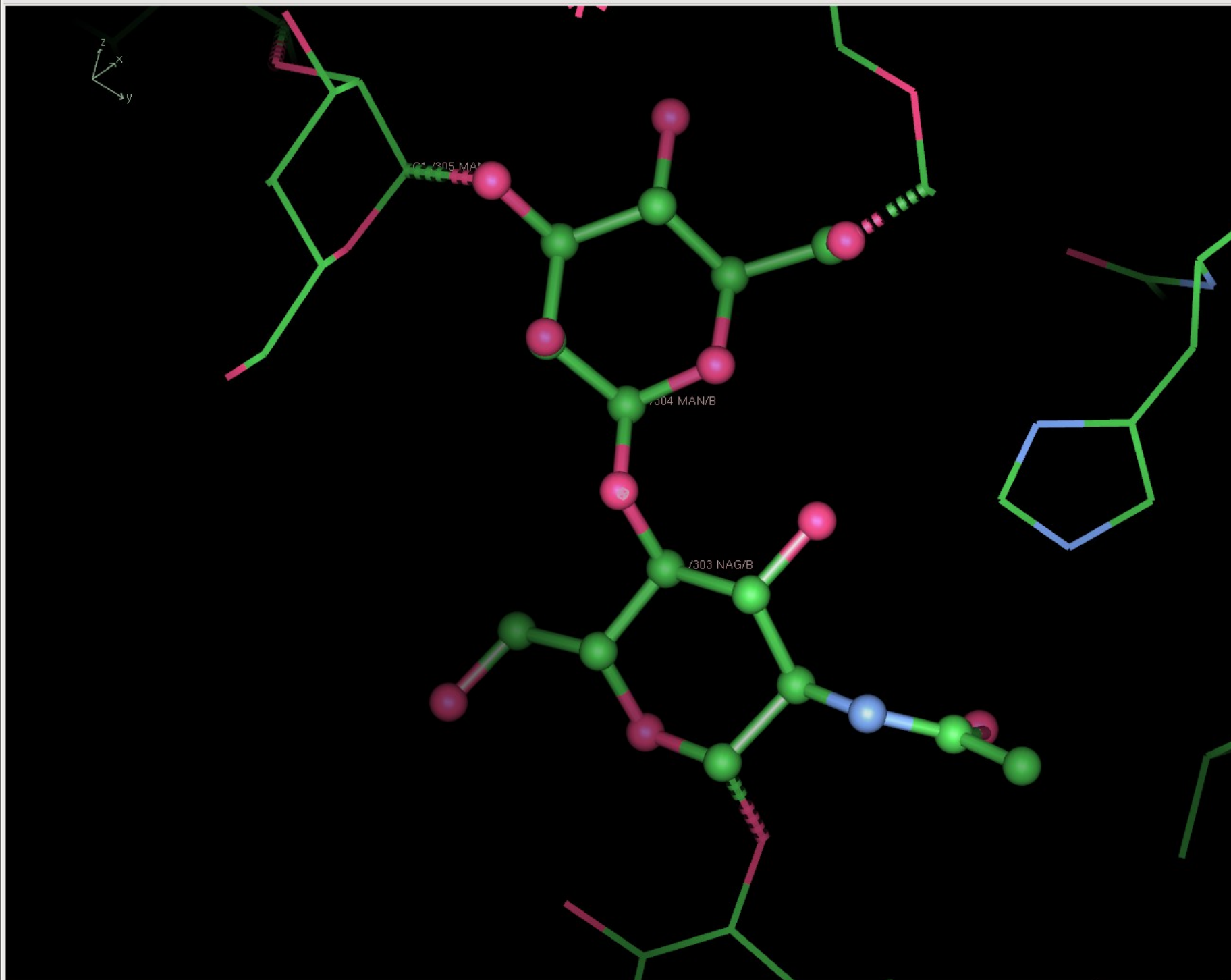


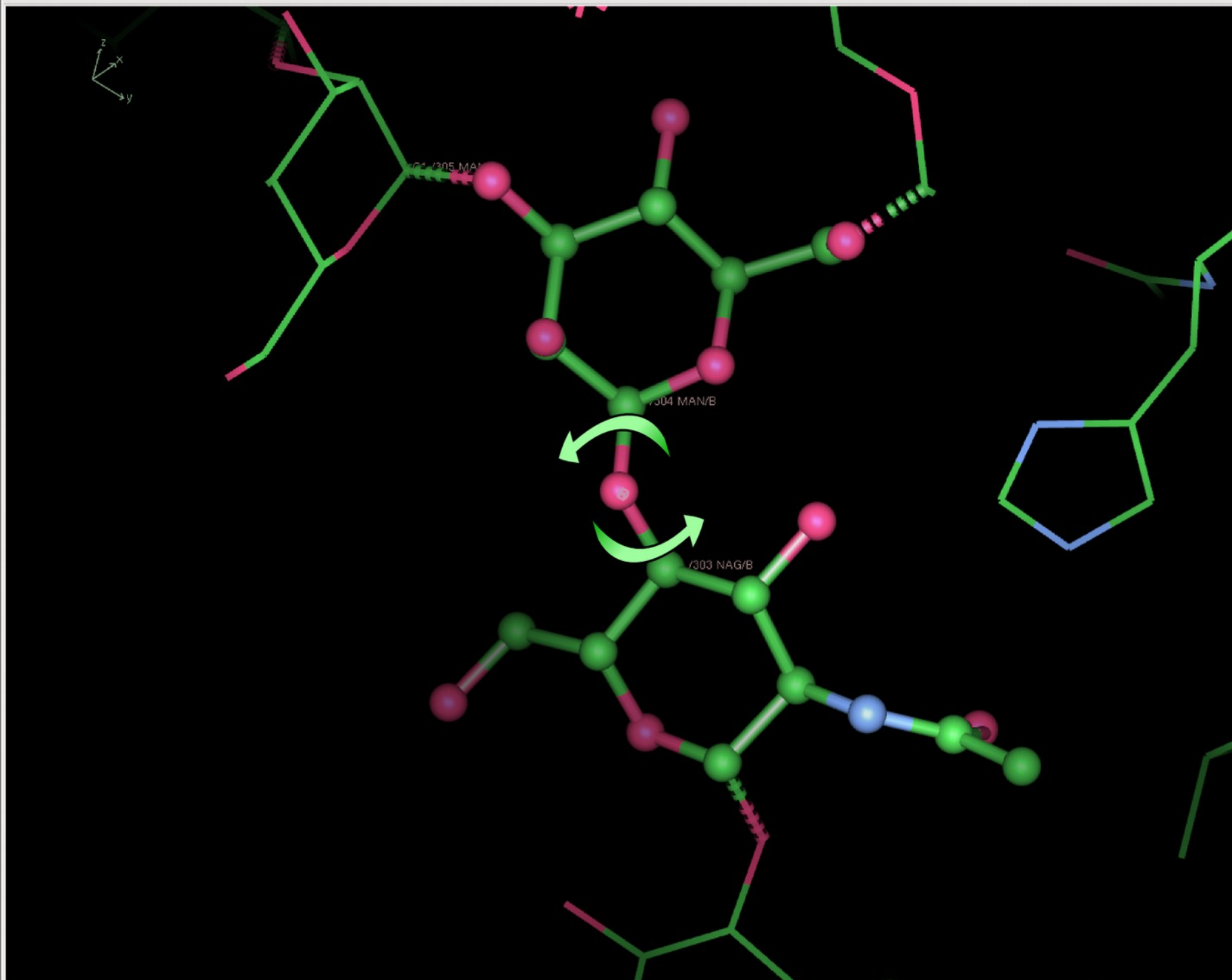
"Complex"



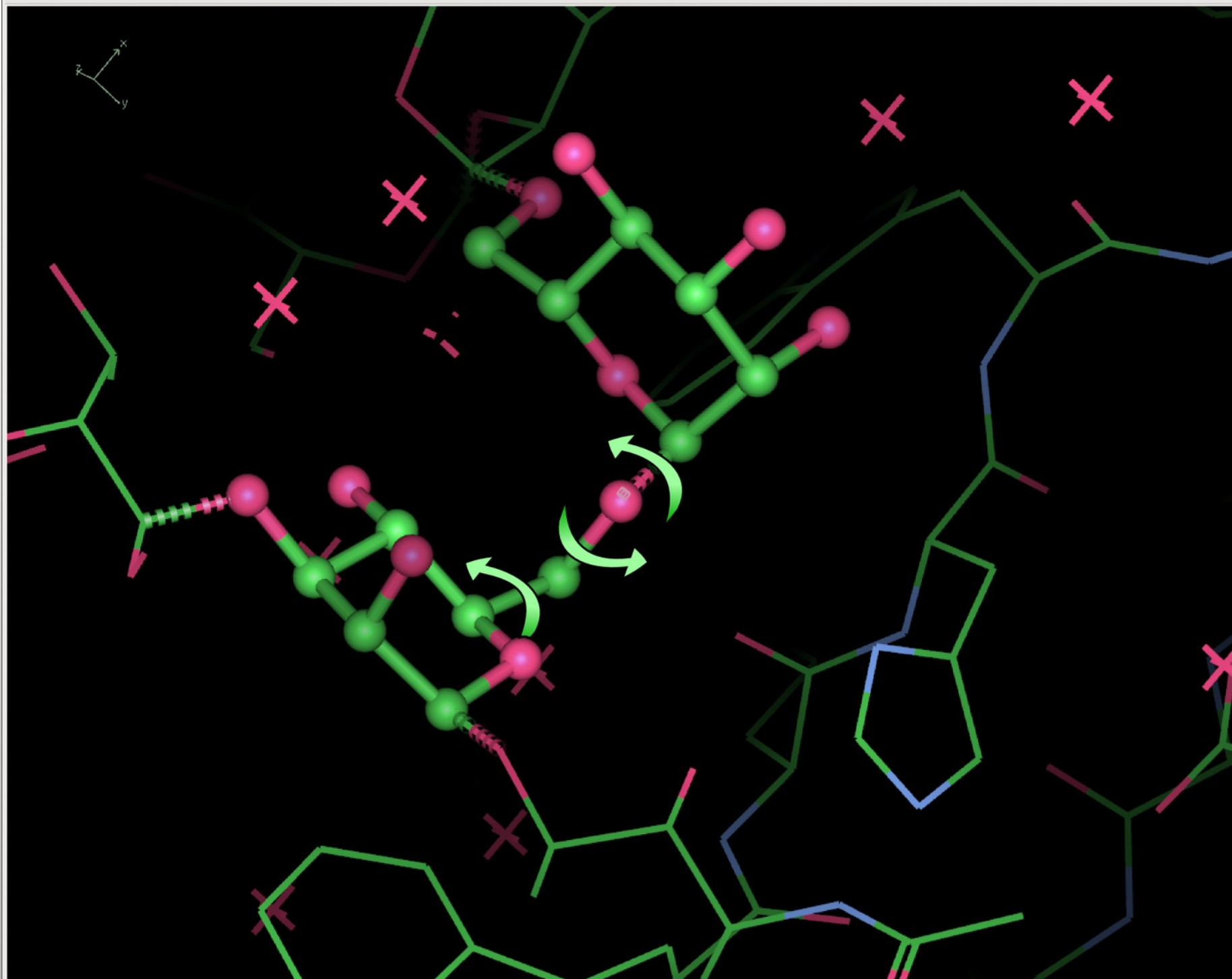
Linking Oligosaccharides/Carbohydrates: LO/Carb

- Complex carbohydrate structure
 - from a dictionary of standard links
 - and monomers
 - torsion-angle refinement

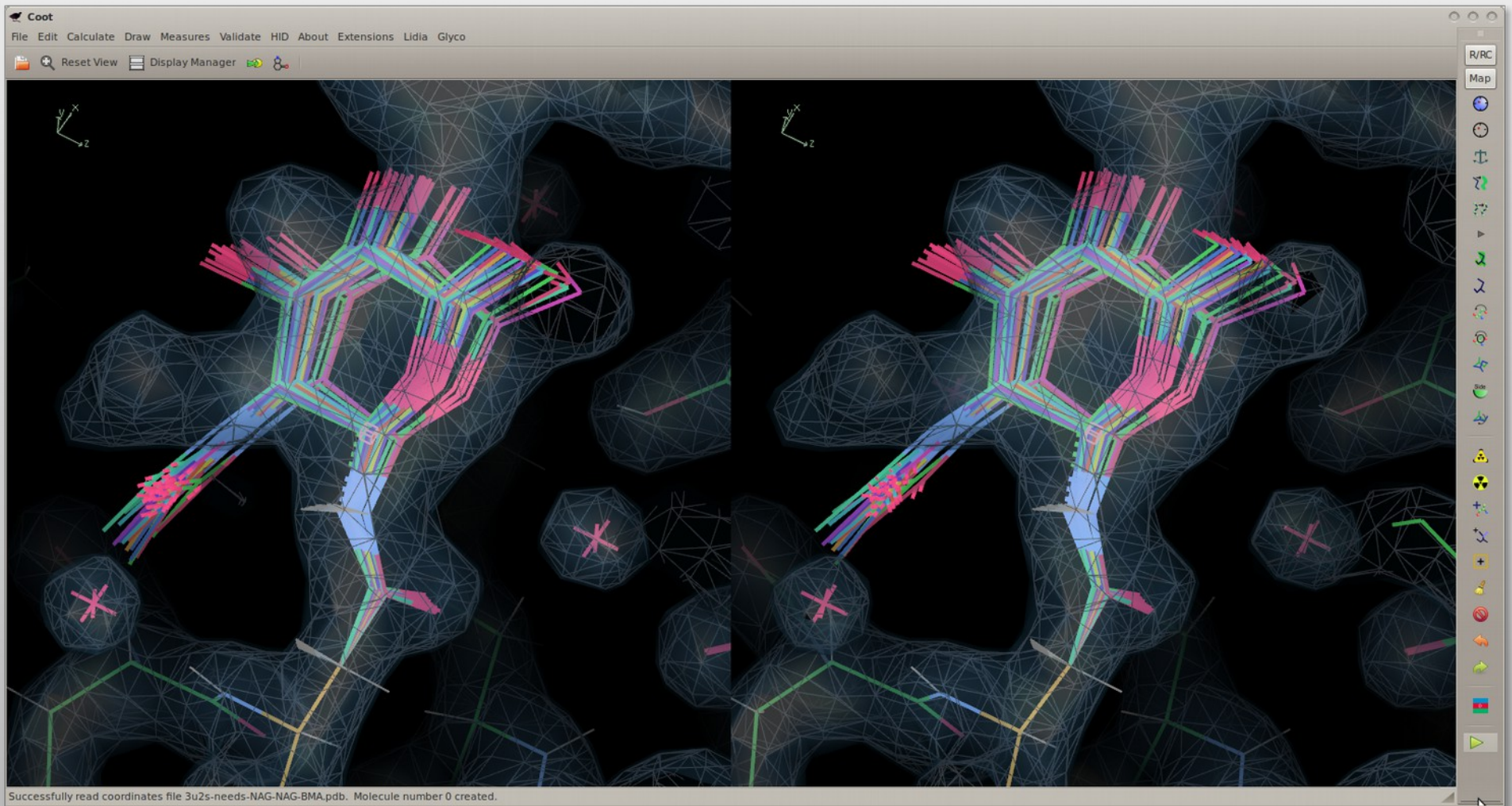




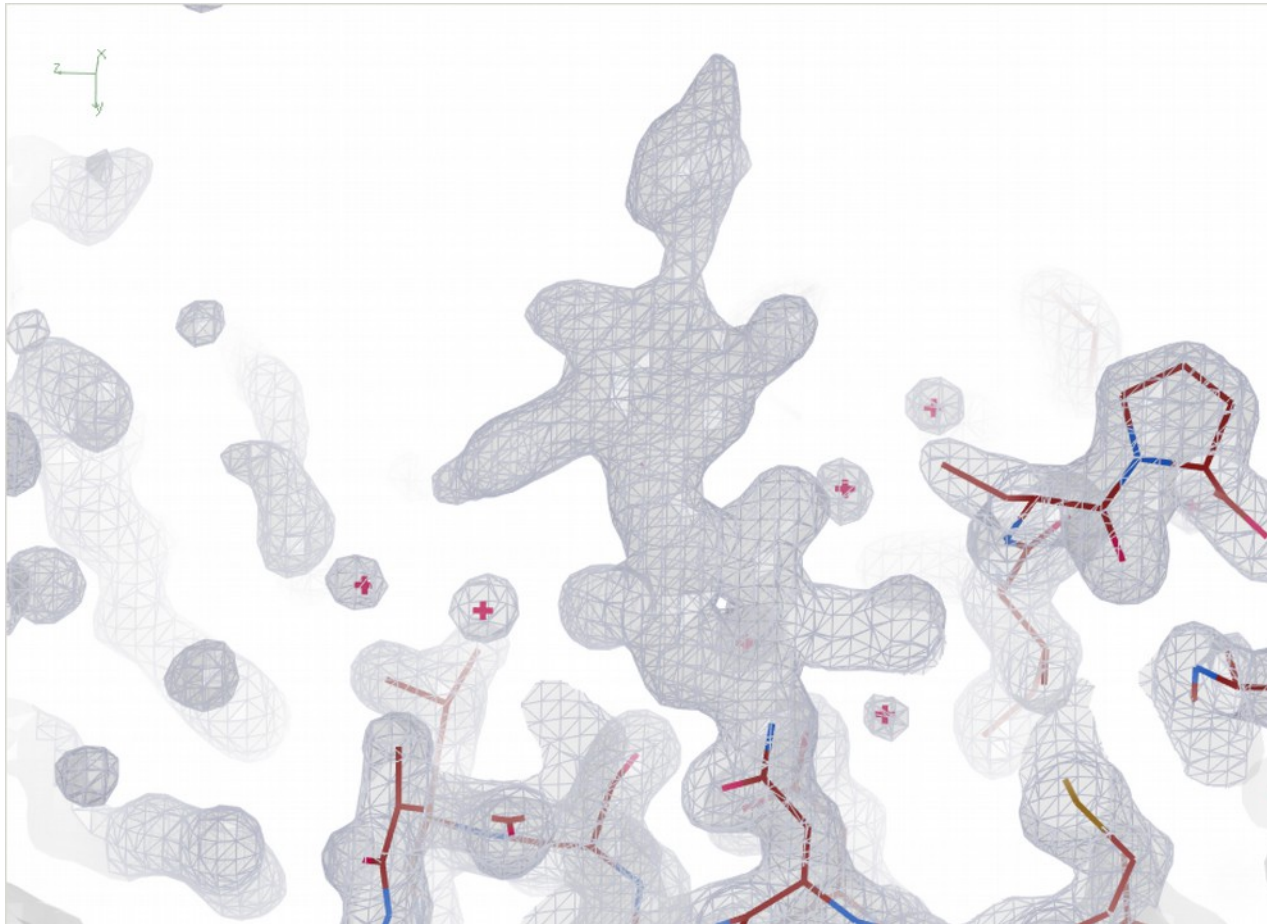




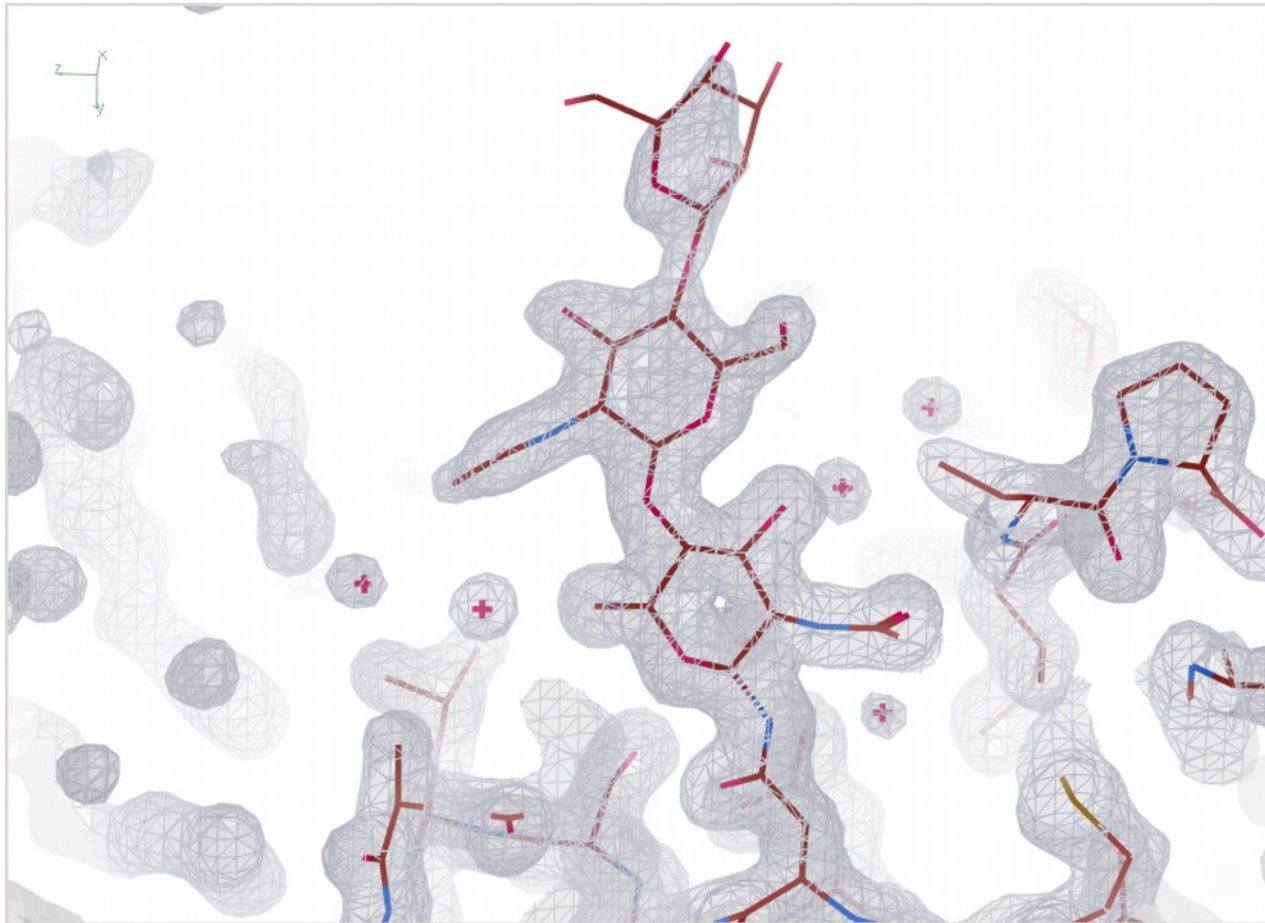
Refinement Trials (NAG-ASN example)



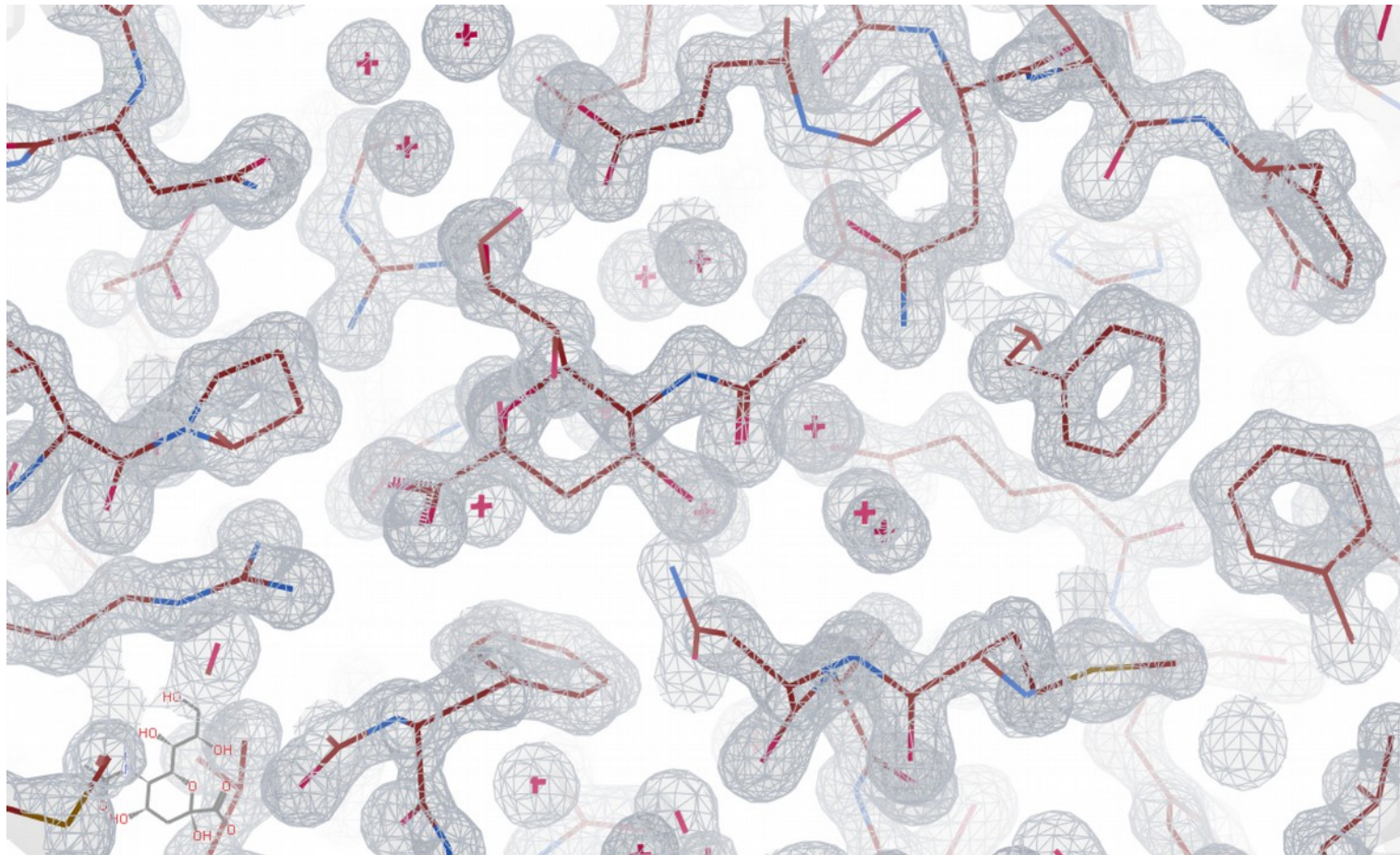
N-linked Carbohydrate



N-linked Carbohydrate

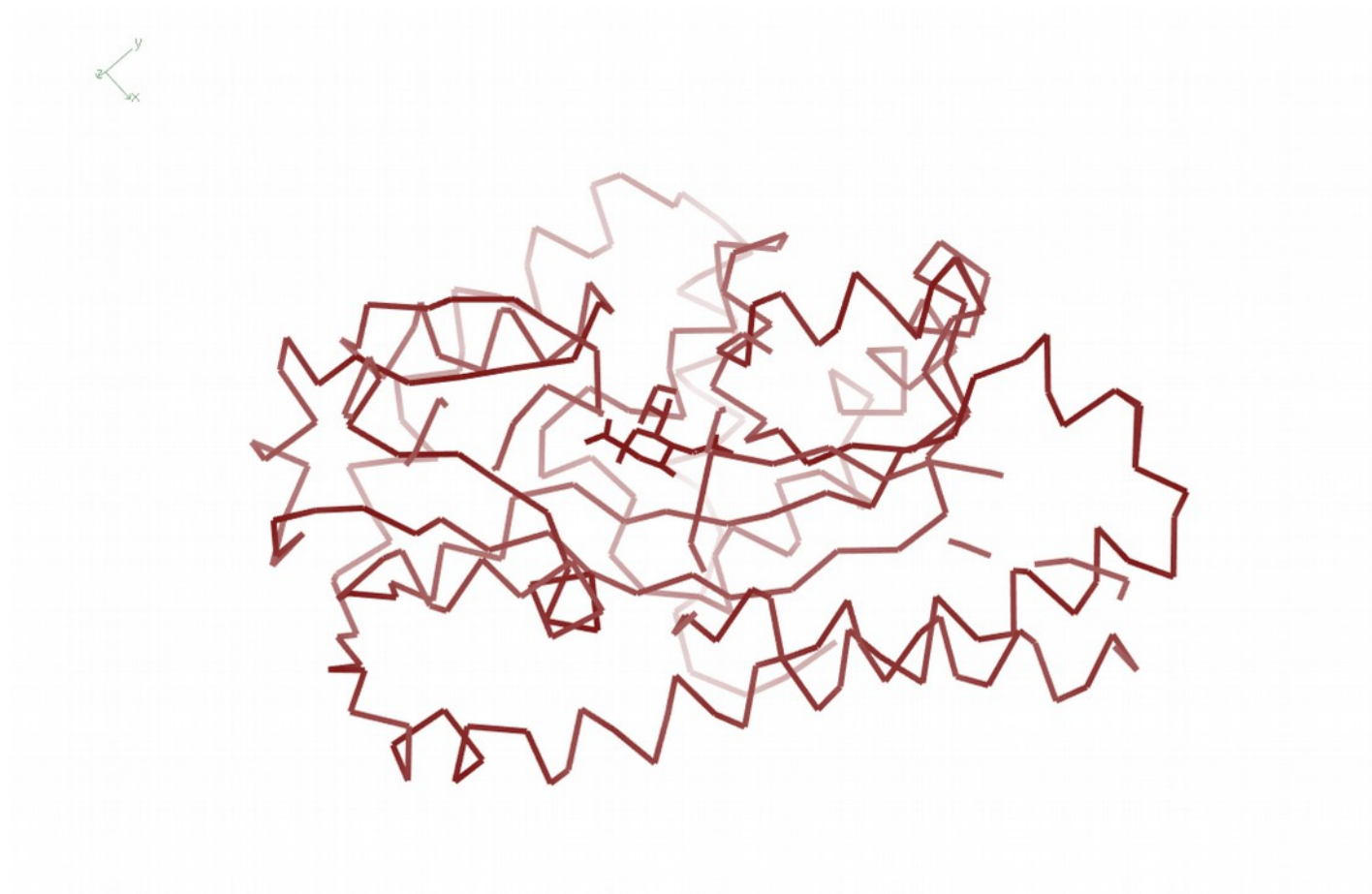


ProSMART restraints

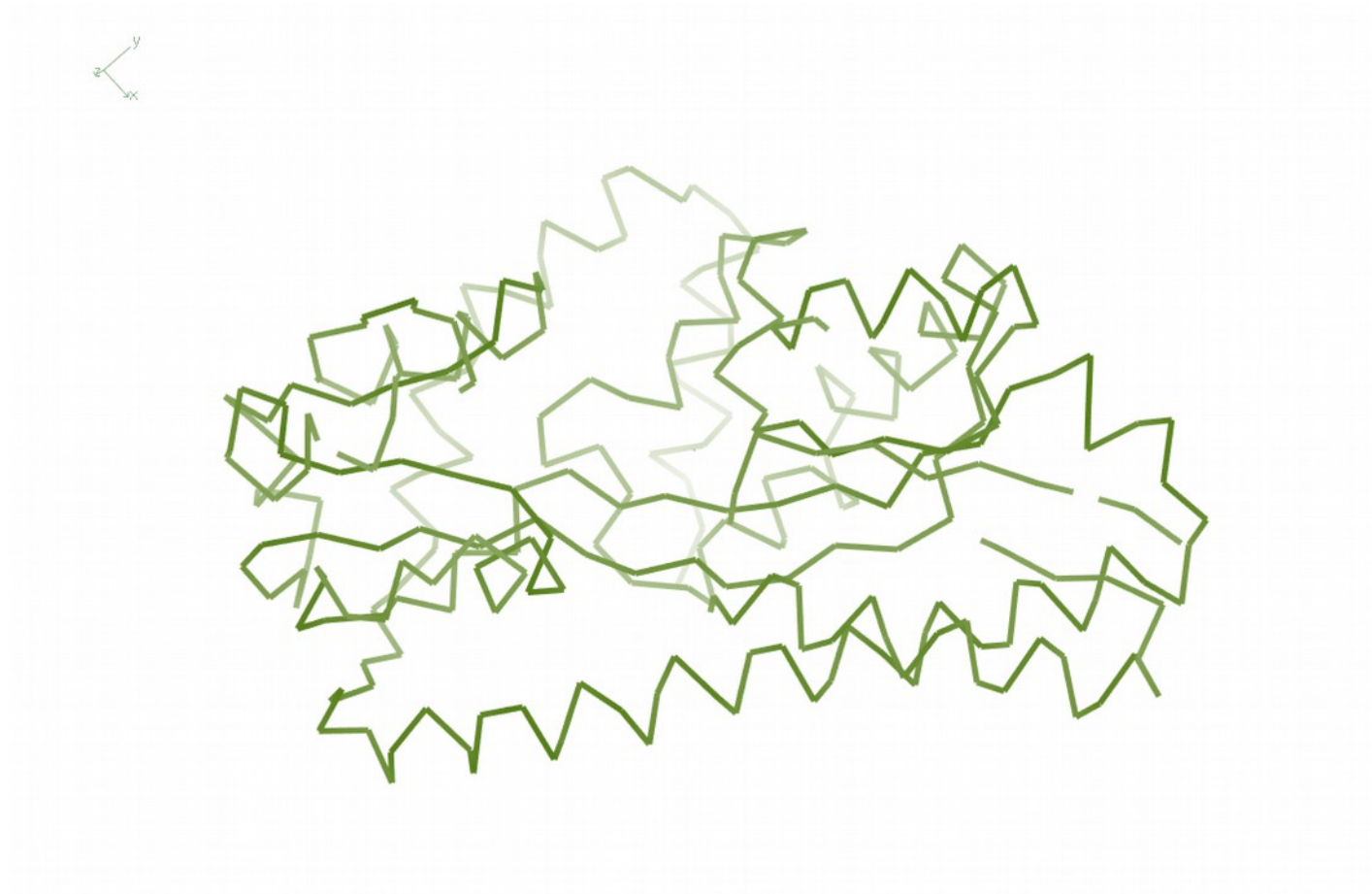


Previously Known High-resolution Reference

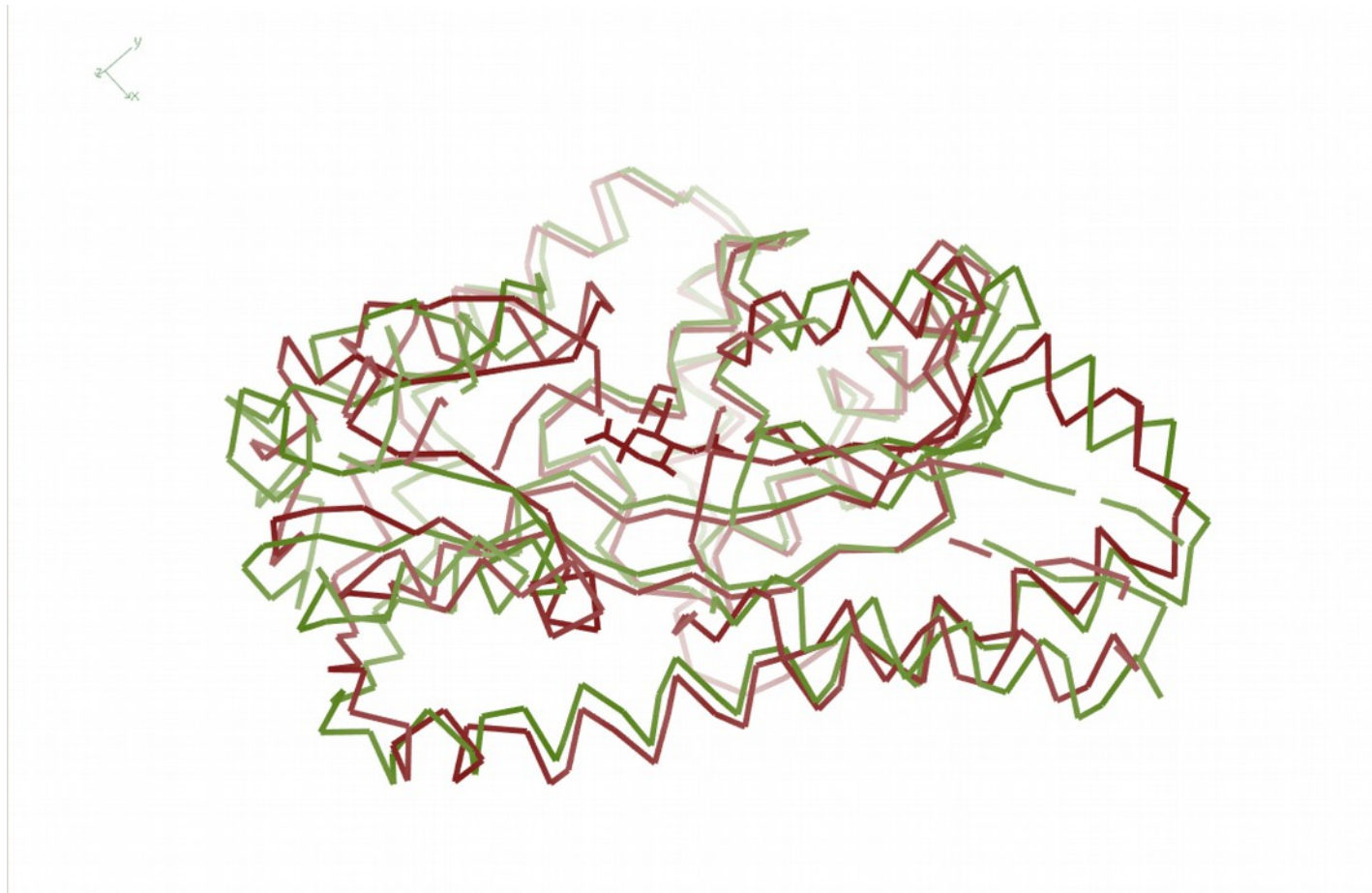
ProSMART Restraints



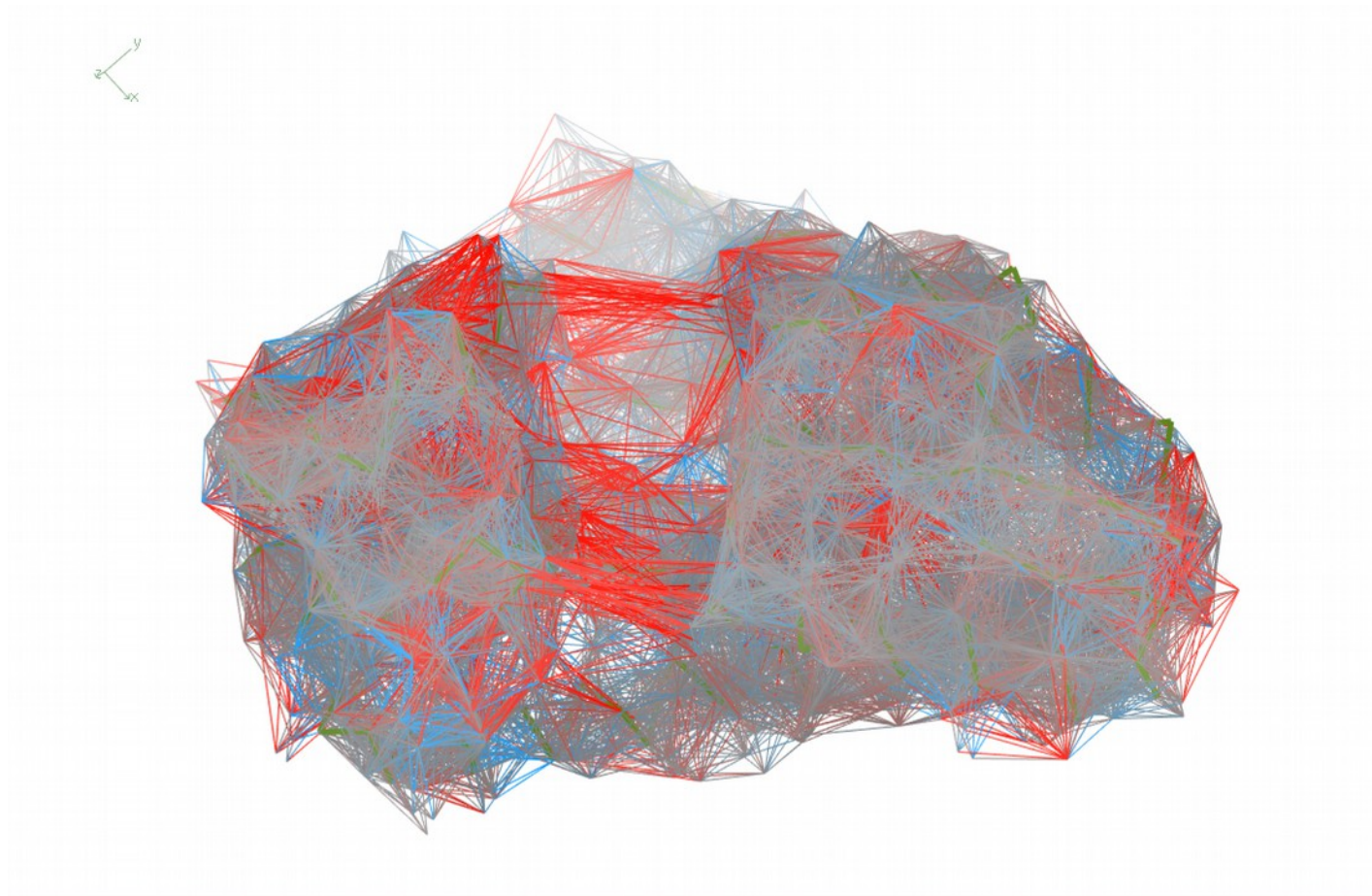
ProSMART Restraints



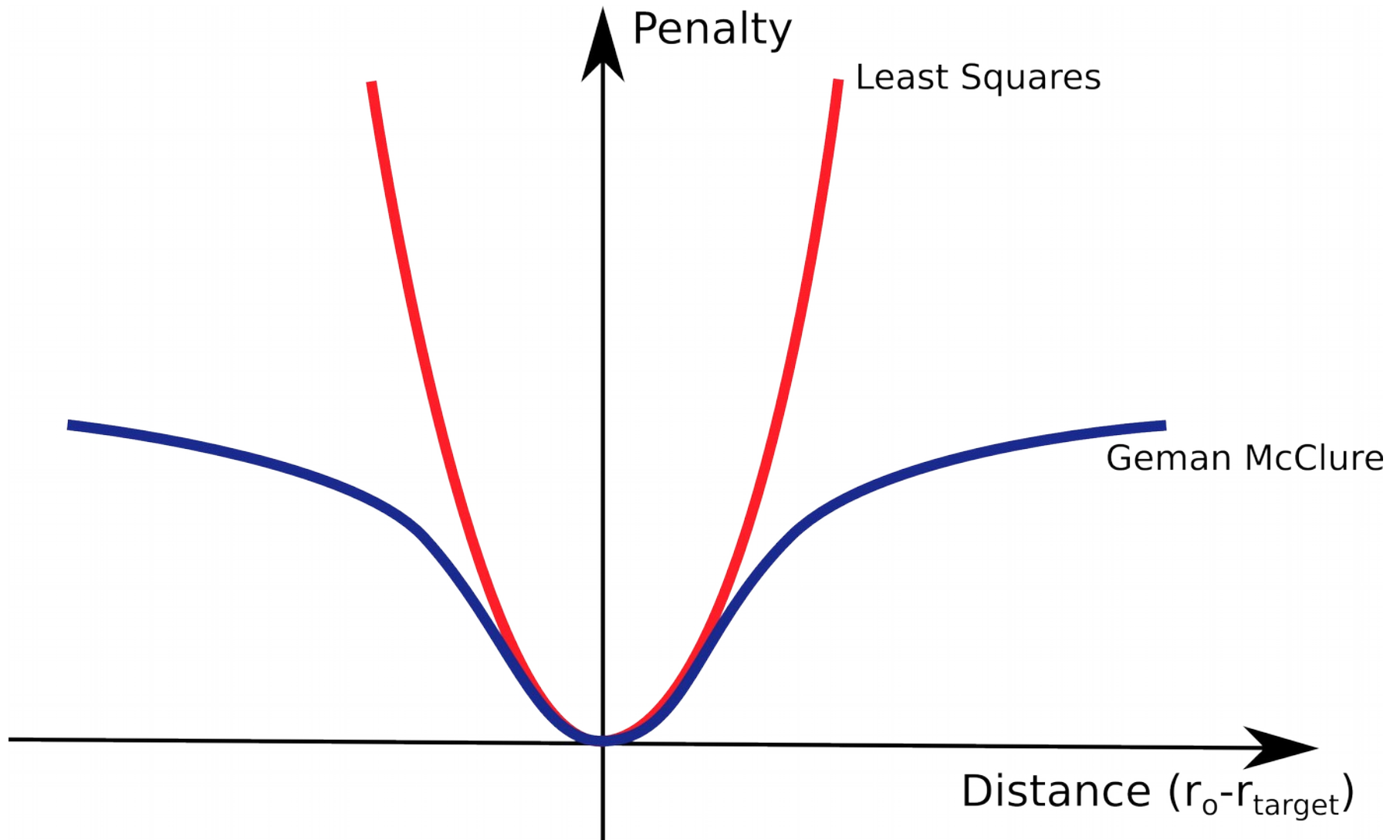
ProSMART Restraints



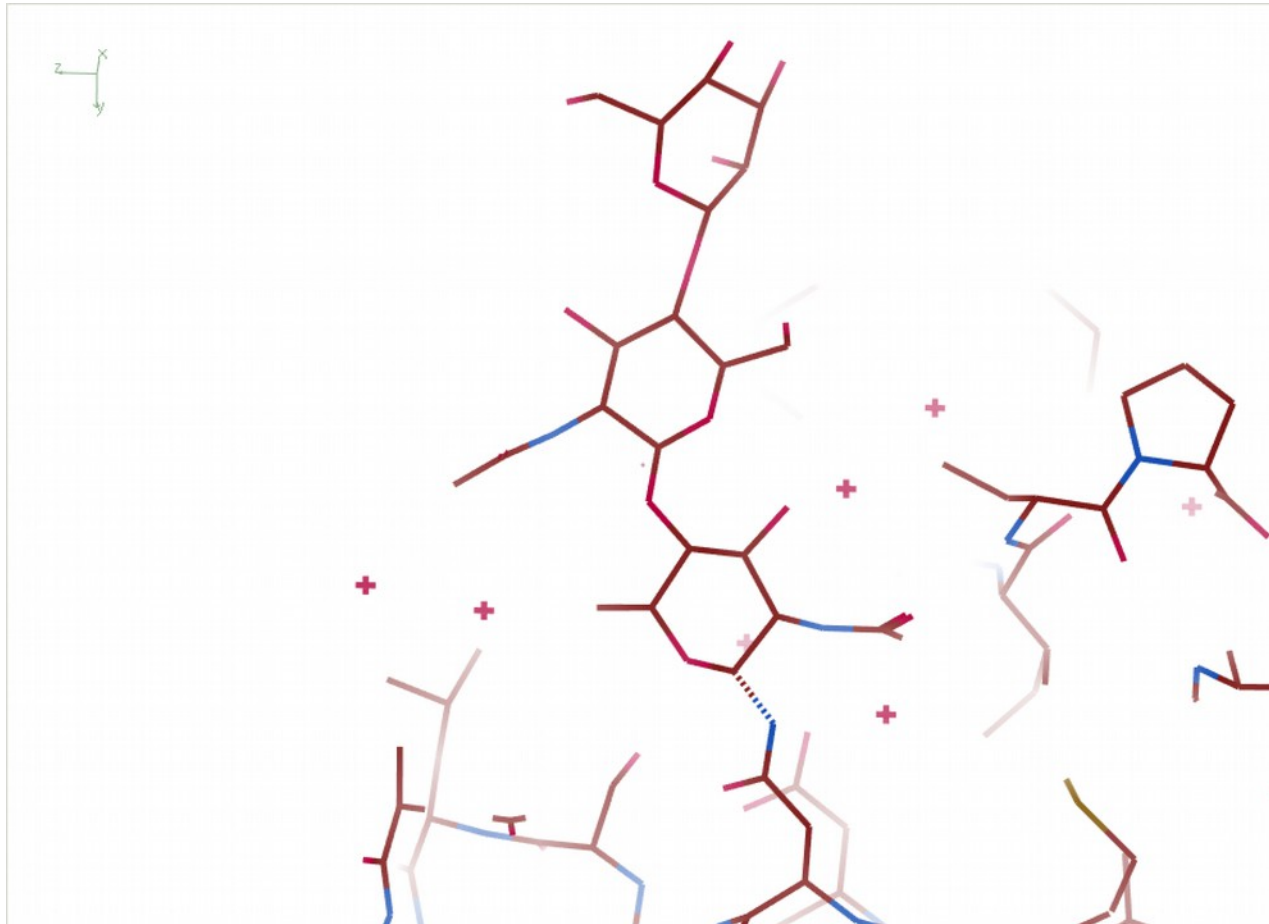
ProSMART Restraints



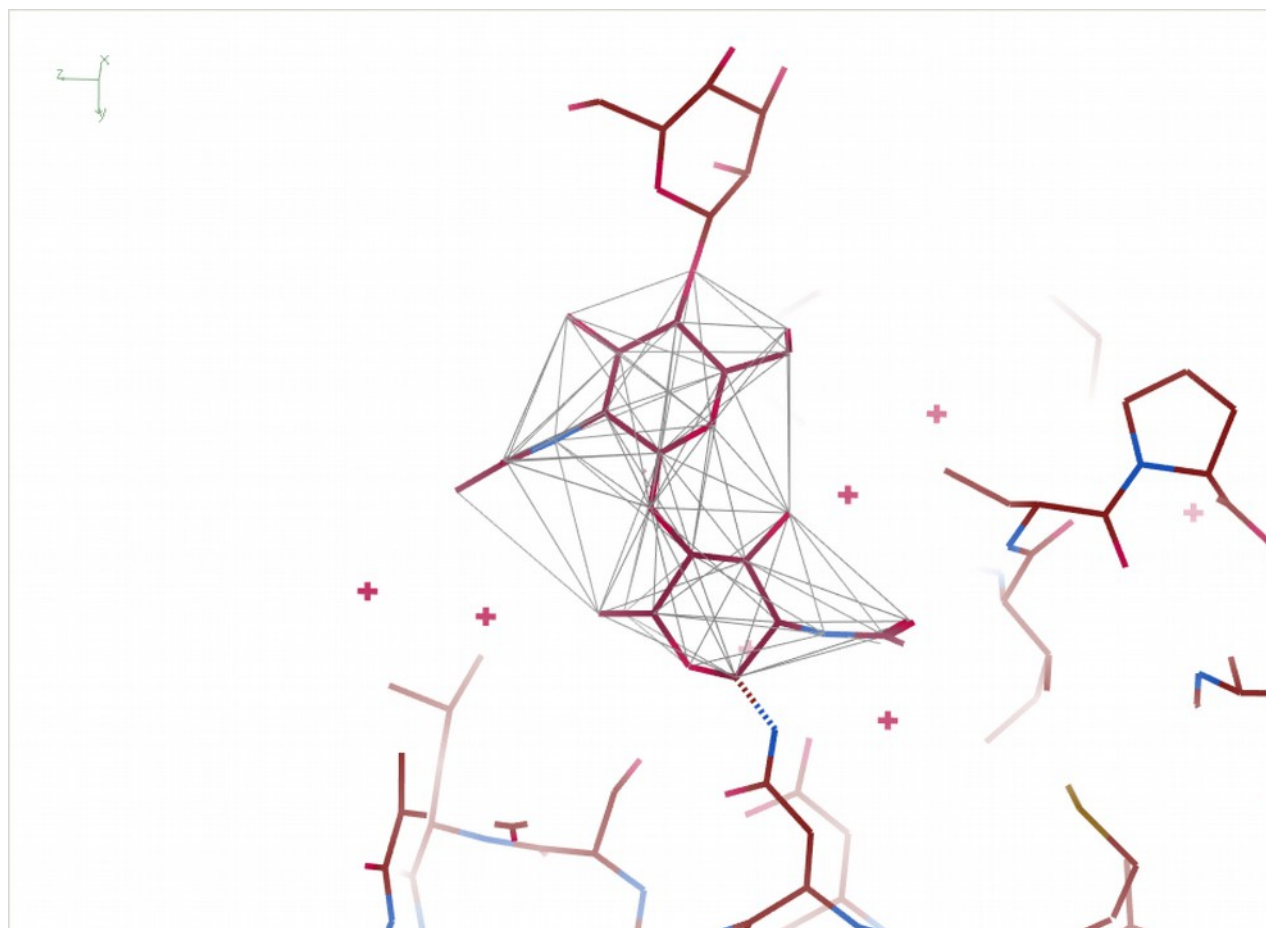
Modified Target Function



N-linked Carbohydrate

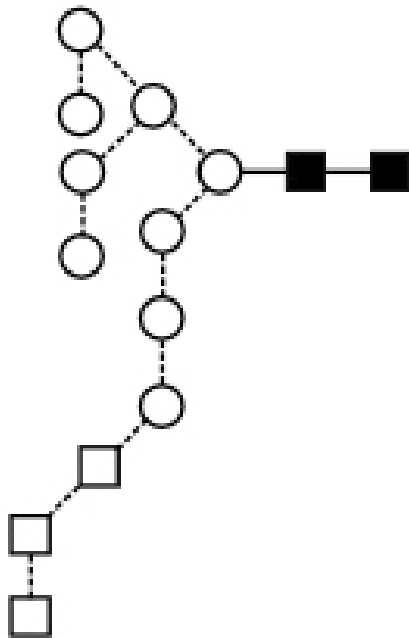


Ligand Tools: N-linked Carbohydrate

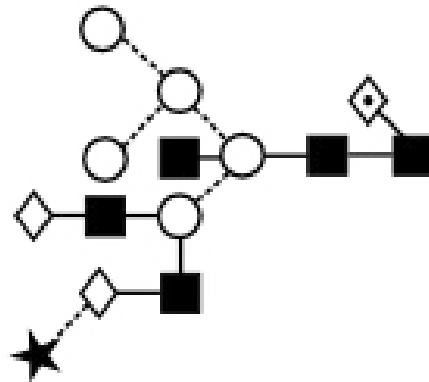


Know N-linked glycosylation schemes

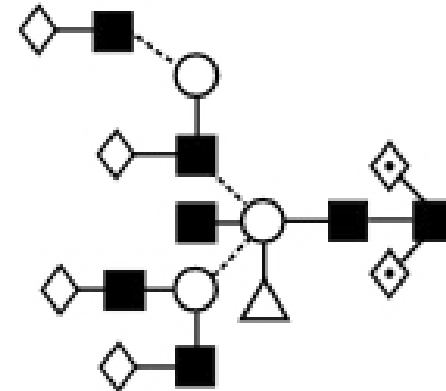
"Oligomannose"



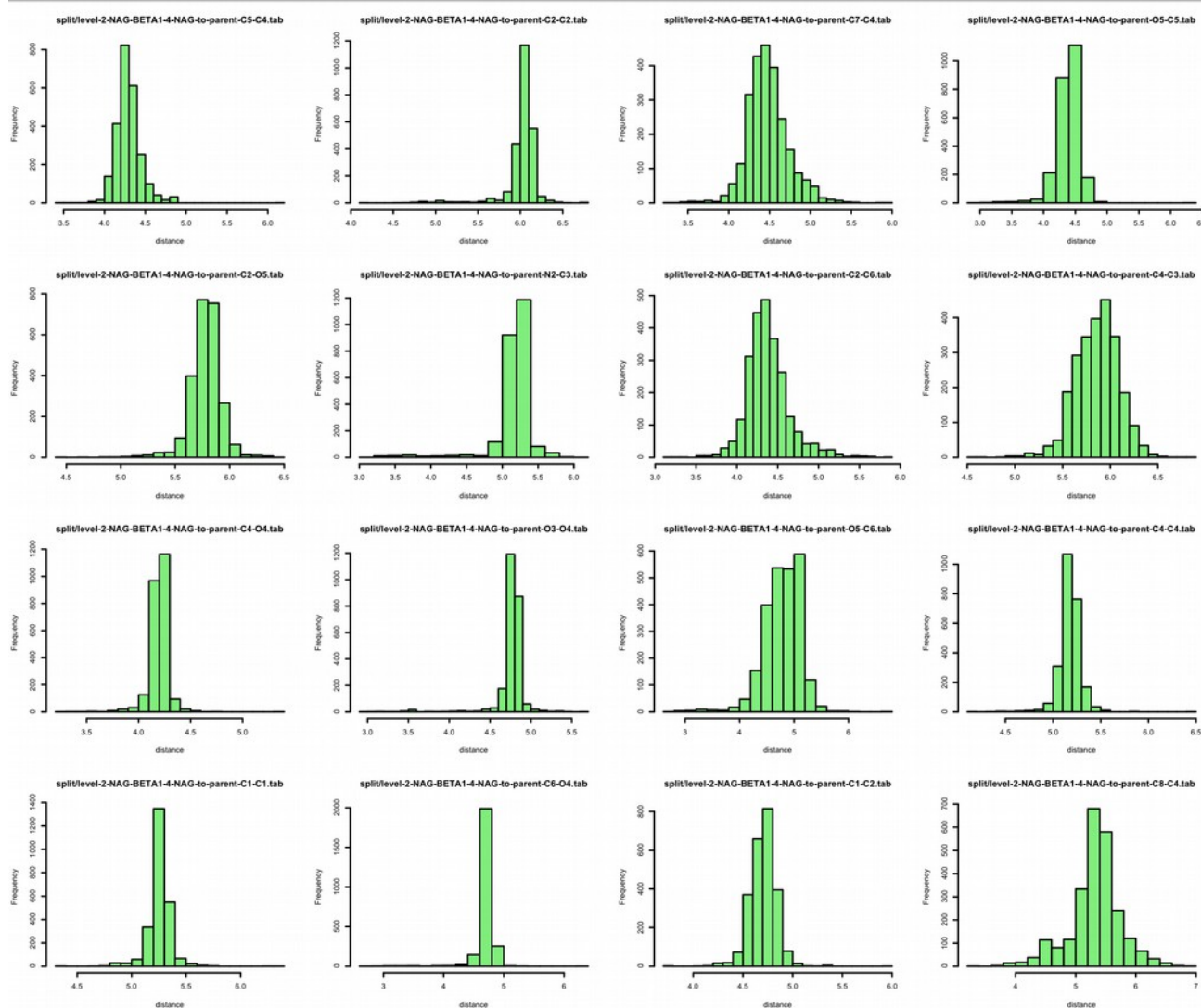
"Hybrid"



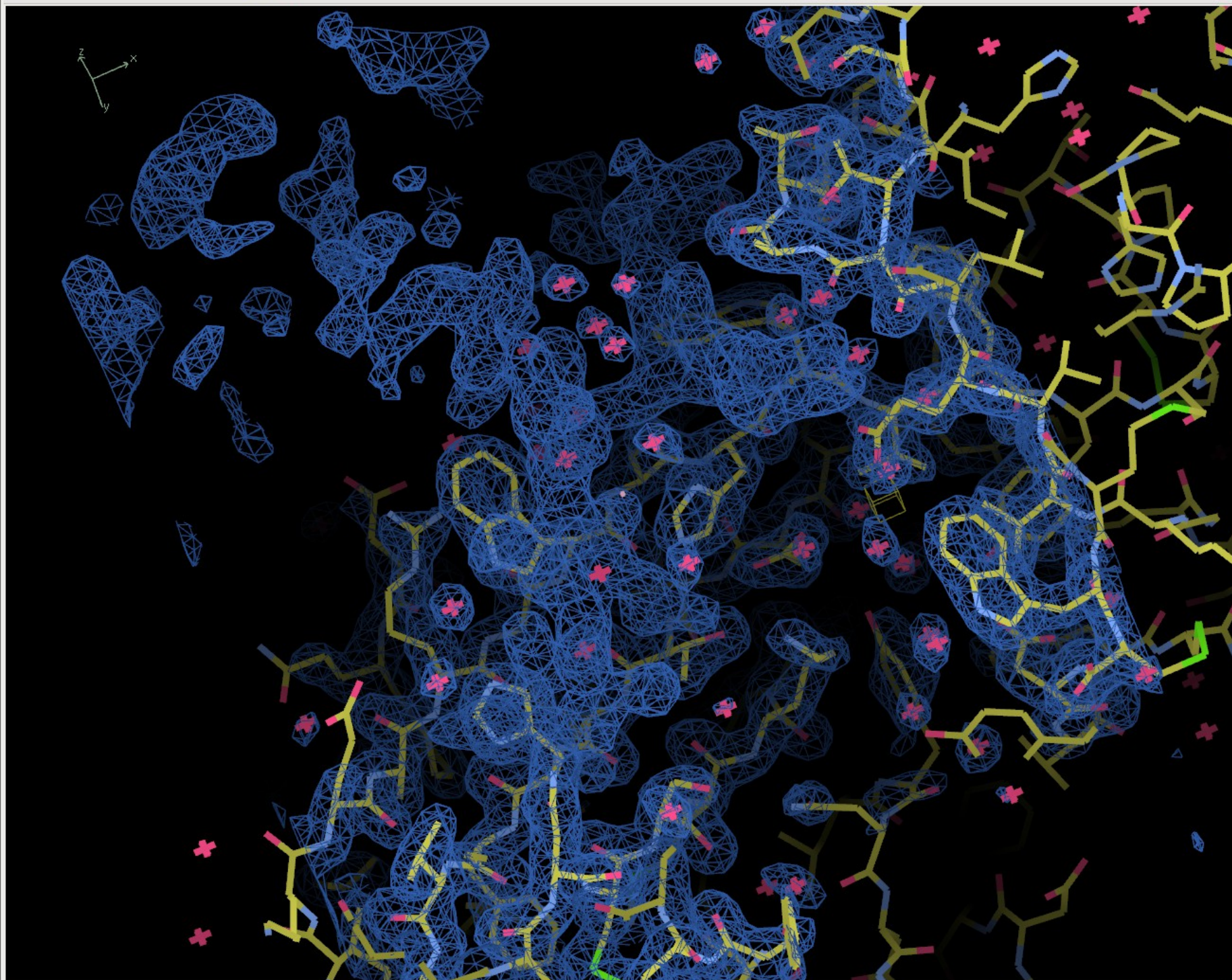
"Complex"

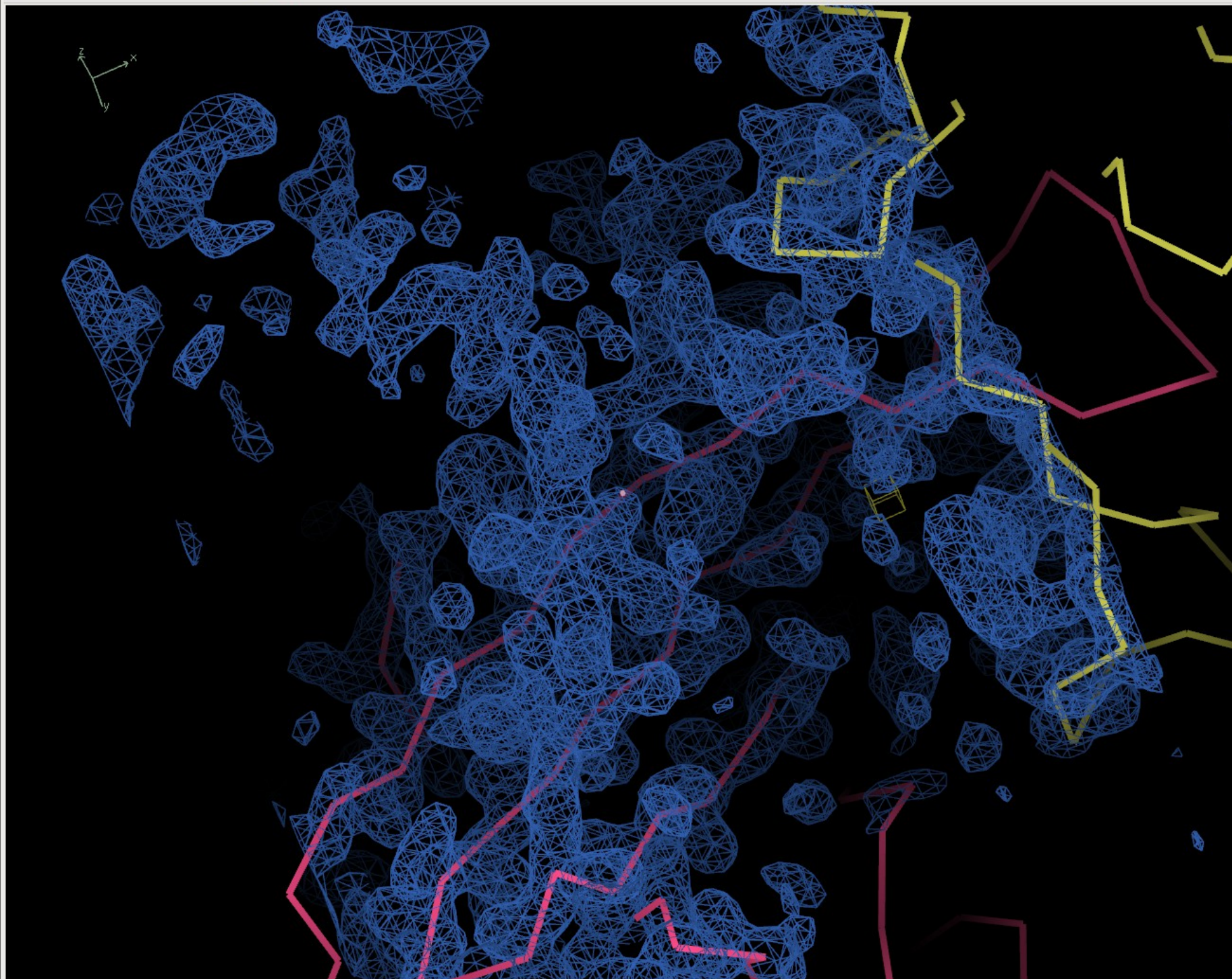


Ligand Tools: N-linked Carbohydrate



→ consensus restraints (no user-defined prior)





R/RC

Map



Acknowledgements

- Kevin Cowtan
- Bernhard Lohkamp
- Libraries, Dictionaries
 - Alexei Vagin, Garib Murshudov
 - Eugene Krissinel
 - Greg Landrum
- Funding:
 - BBSRC & CCP4