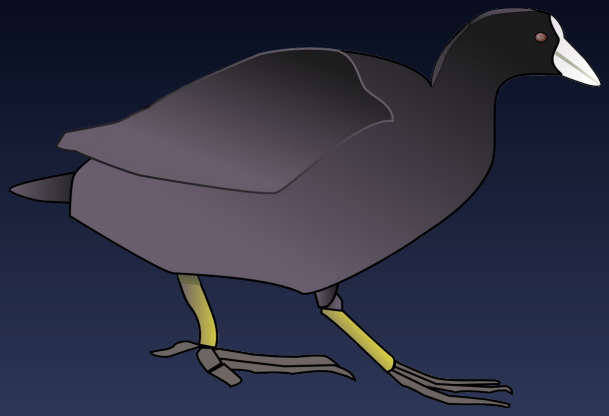


June 2010 Chicago

Ligand fitting and Validation with Coot

(Paul Emsley)
(University of Oxford)

Bernhard Lohkamp
Karolinska Institutet

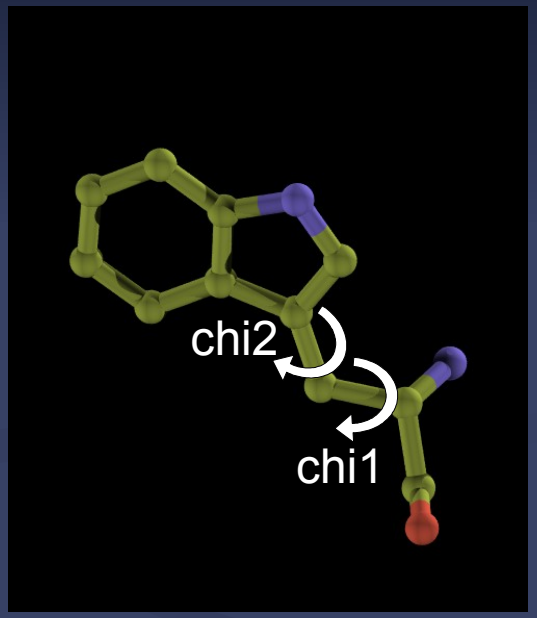
Fitting Ligands

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    0.033
 ALA      CB   HB1    single    0.960    0.020
 ALA      CB   HB2    single    0.960    0.020
 ALA      CB   HB3    single    0.960    0.020
 ALA      CA   C      single    1.525    0.021
 ALA      C    O      double    1.231    0.020
```

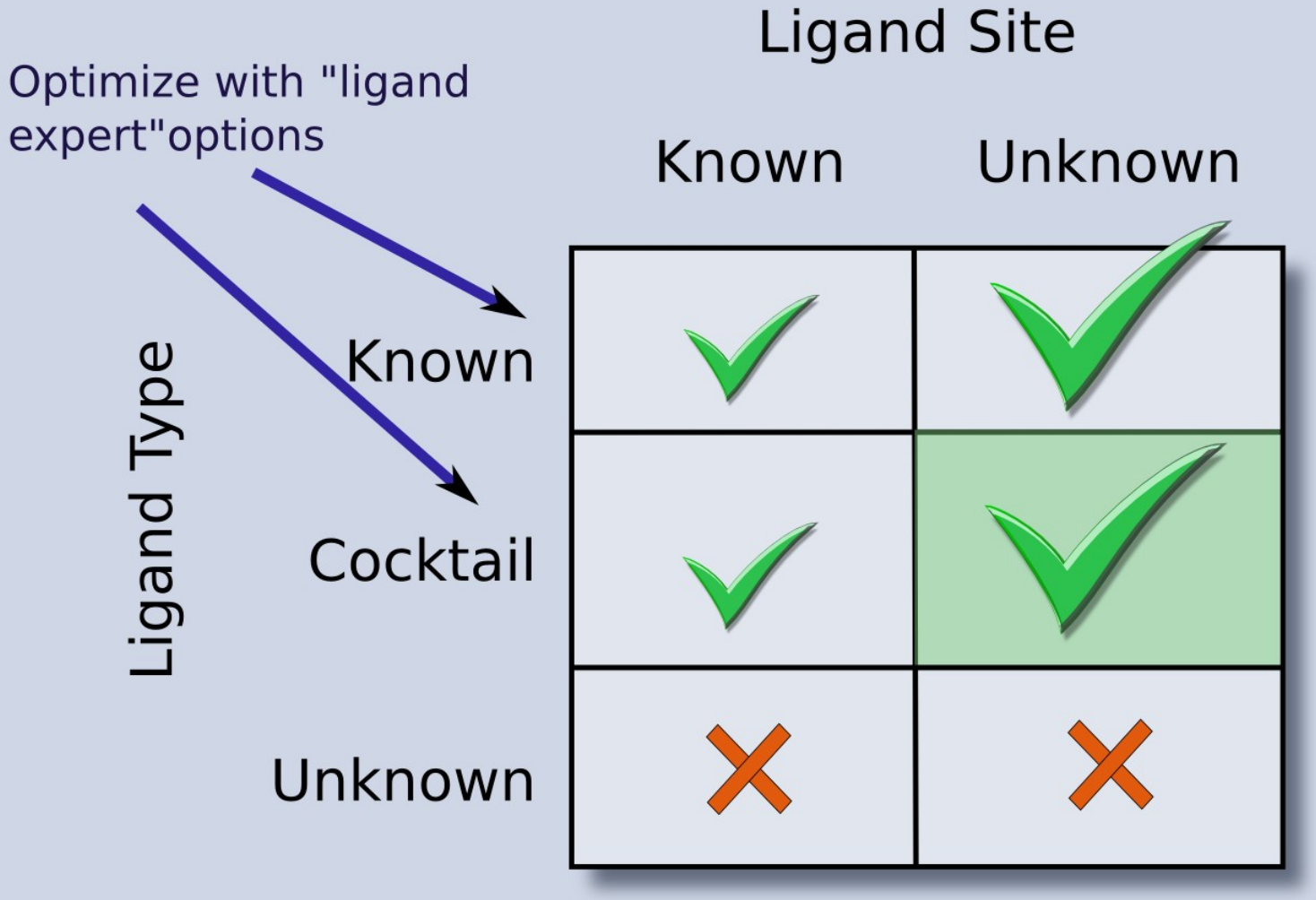
REFMAC Monomer Library chem_comp_tor

```
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
 TRP   chi1  N   CA  CB  CG  180.000   15.000   3
 TRP   chi2  CA  CB  CG  CD1  90.000   20.000   2
```

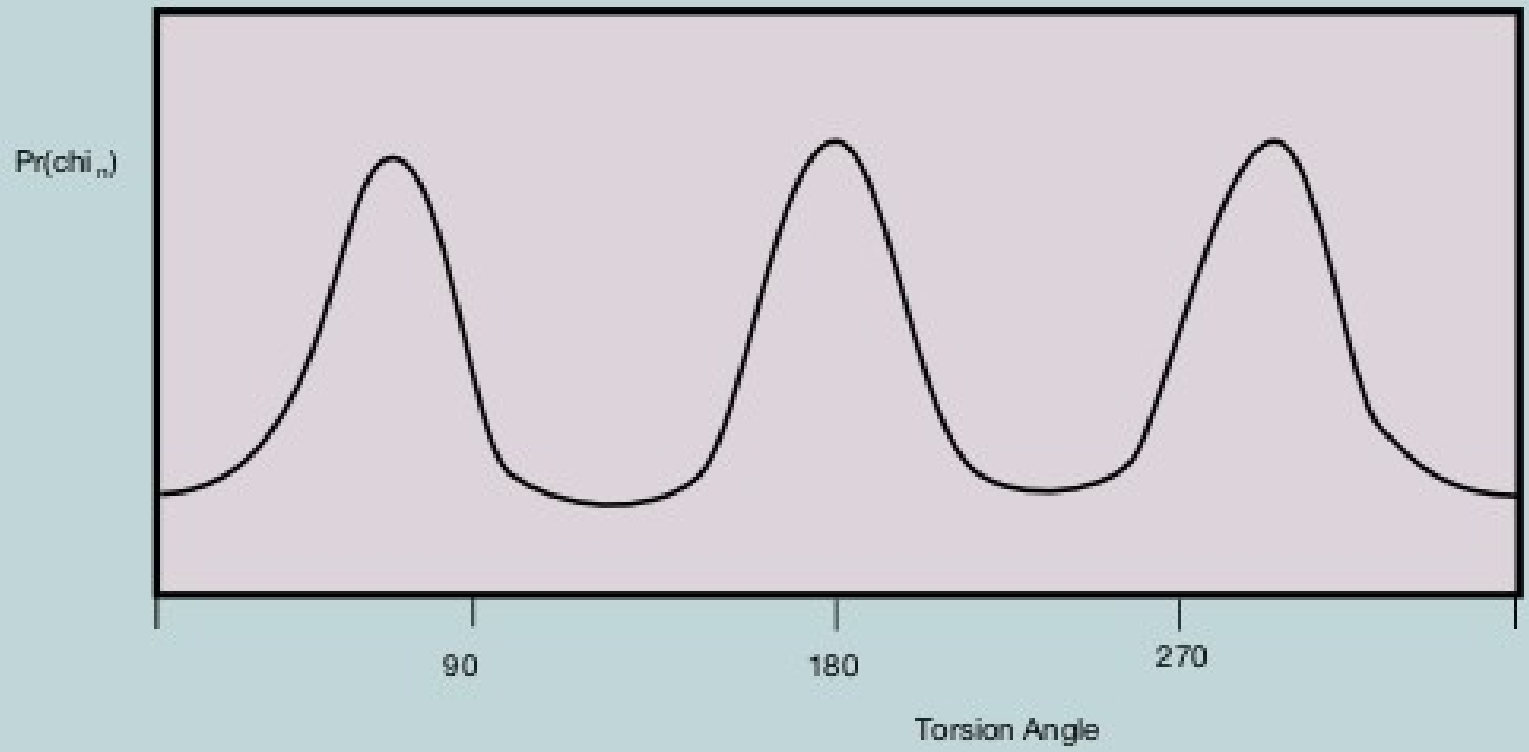


Ligand Fitting

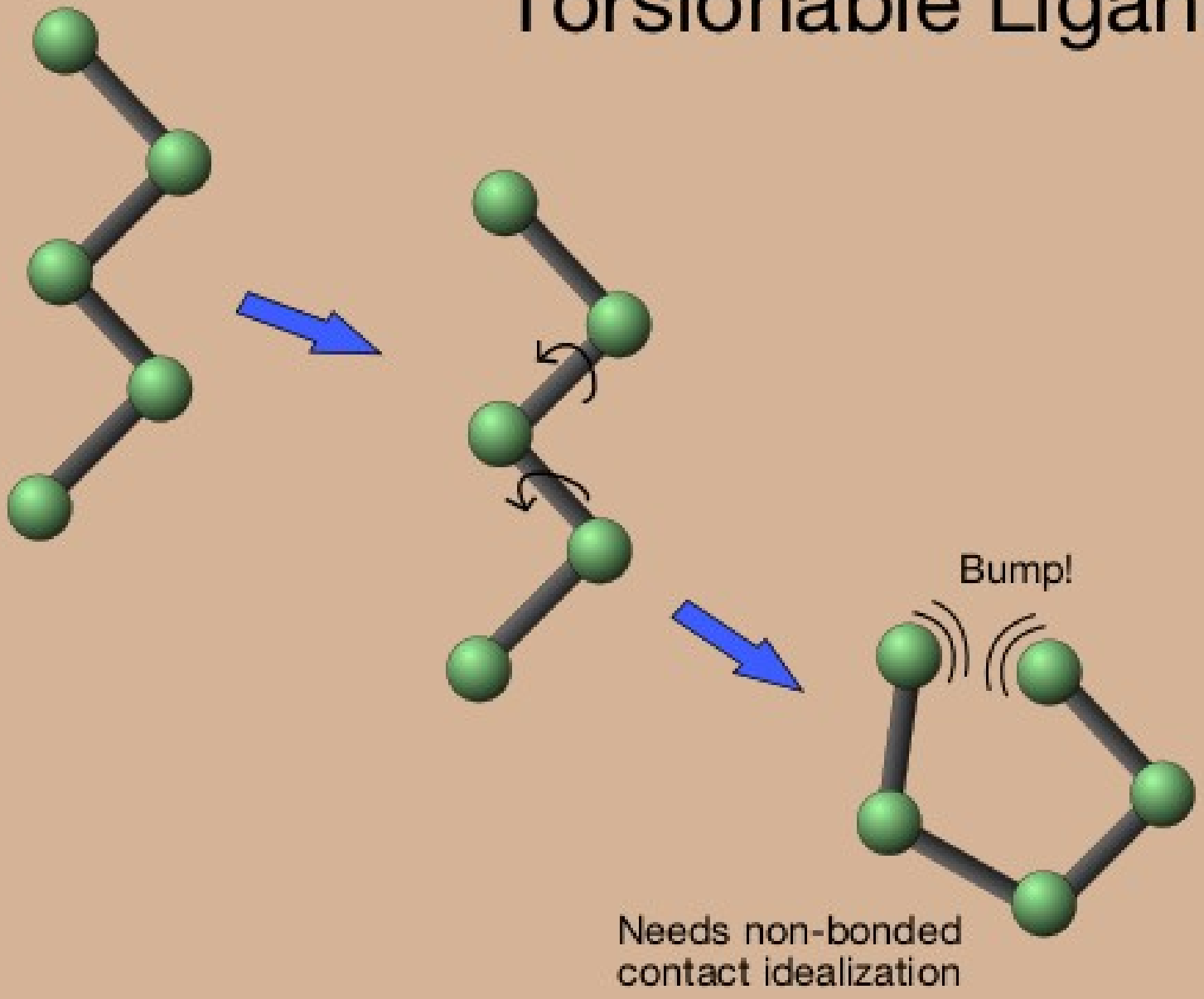
- *c.f.* Oldfield (2001) *Acta Cryst. D* X-LIGAND
- Somewhat different torsion search algorithm
- Build in crystal-space


Optimize with "ligand expert"options
Ligand Site
Known
Unknown
Ligand Type
Known
Cocktail
Unknown

Ligand Torsionable Angle Probability from CIF file



Torsionable Ligands

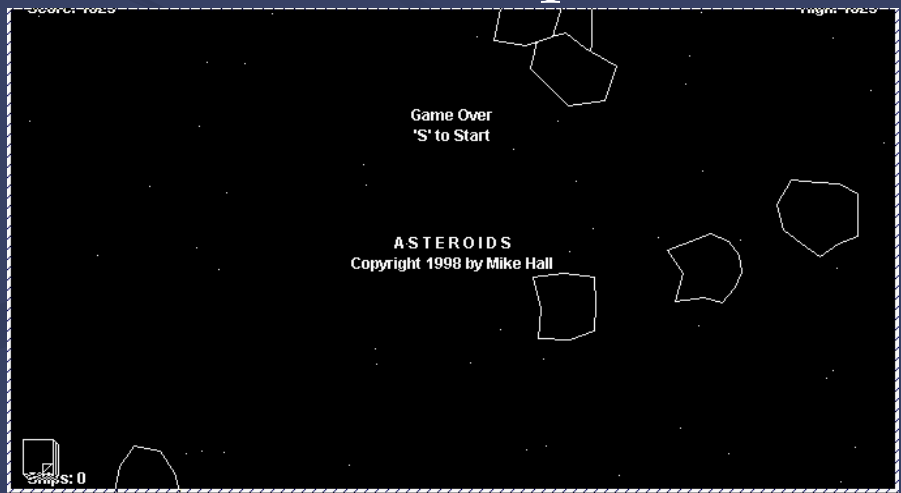


Crystal Space

- Build in “crystal space”
  - Like real-space, but wrapped by crystal symmetry
    - Like “Asteroids”
  - Assures only one real-space representation of map features
    - Build everything only once,
    - No symmetry clashing
  - However, more difficult to calculate real space geometries
    - …such as bonds, torsions

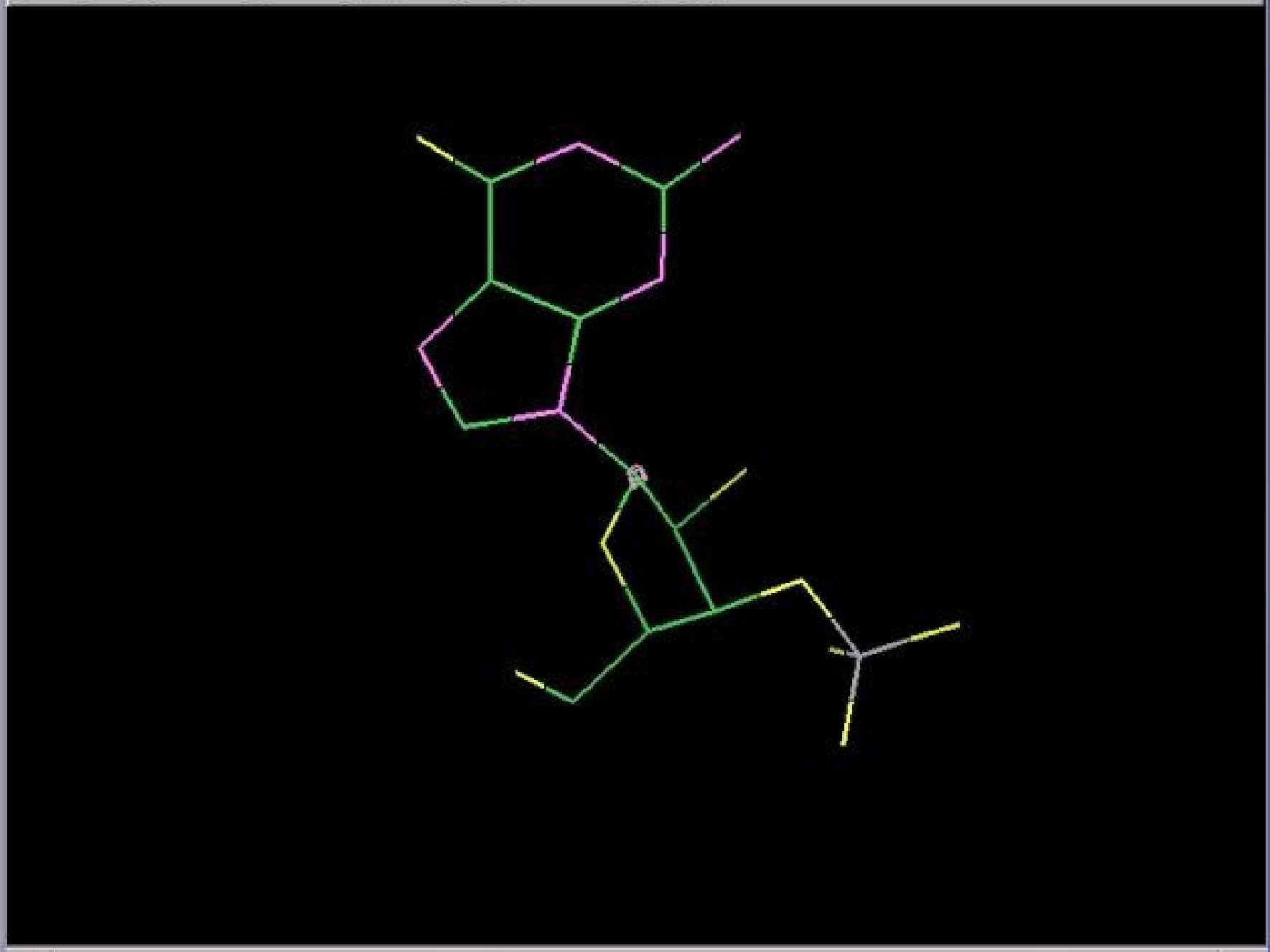
Crystal Space

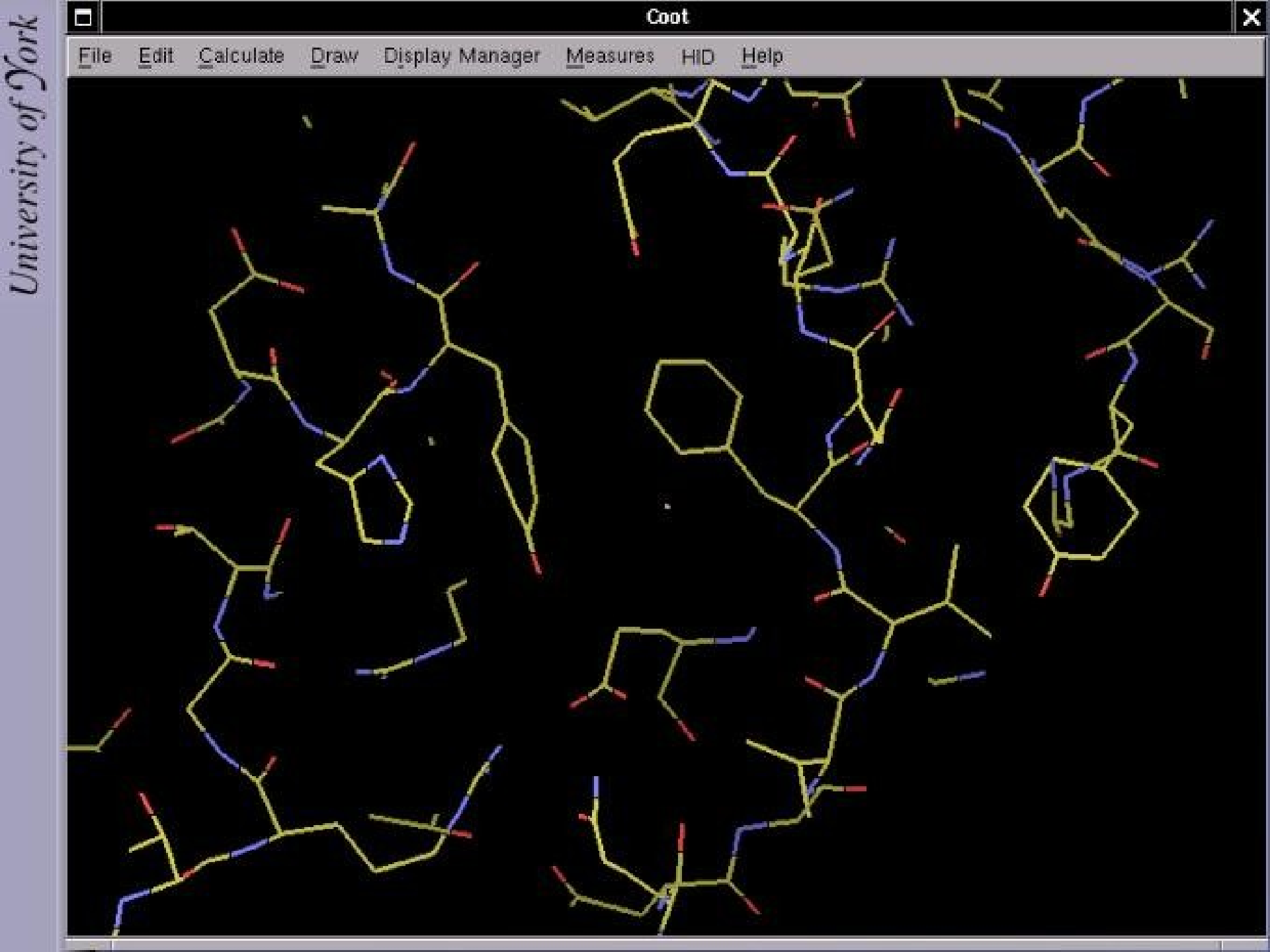
- Build in “crystal space”
  - Like real-space, but wrapped by crystal symmetry
    - Like “Asteroids”
  - Assures only one real-space representation of map features
    - Build everything only once
    - No symmetry clashing
  - However, more difficult to calculate real space geometries
    - …such as bonds, torsions

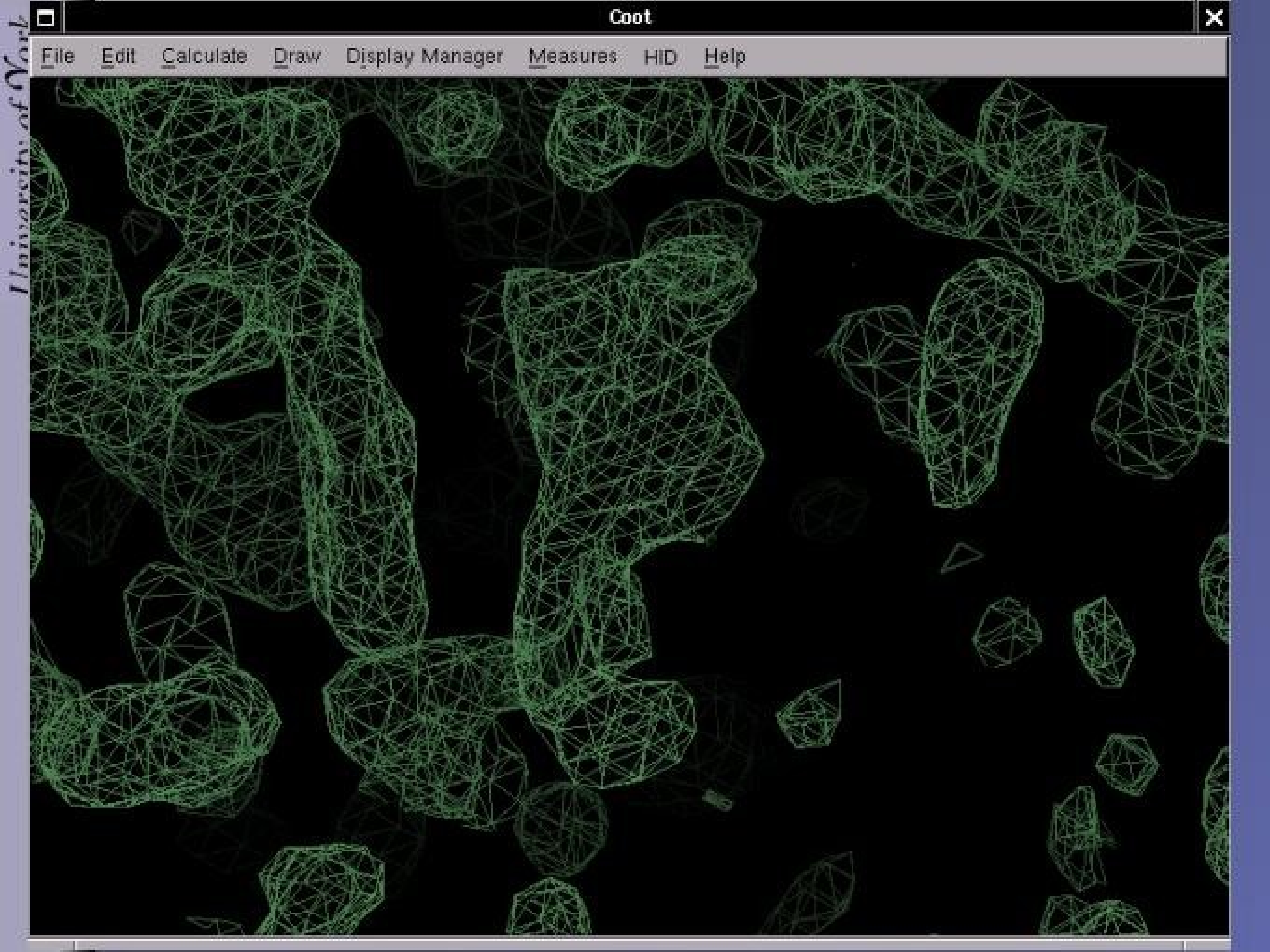


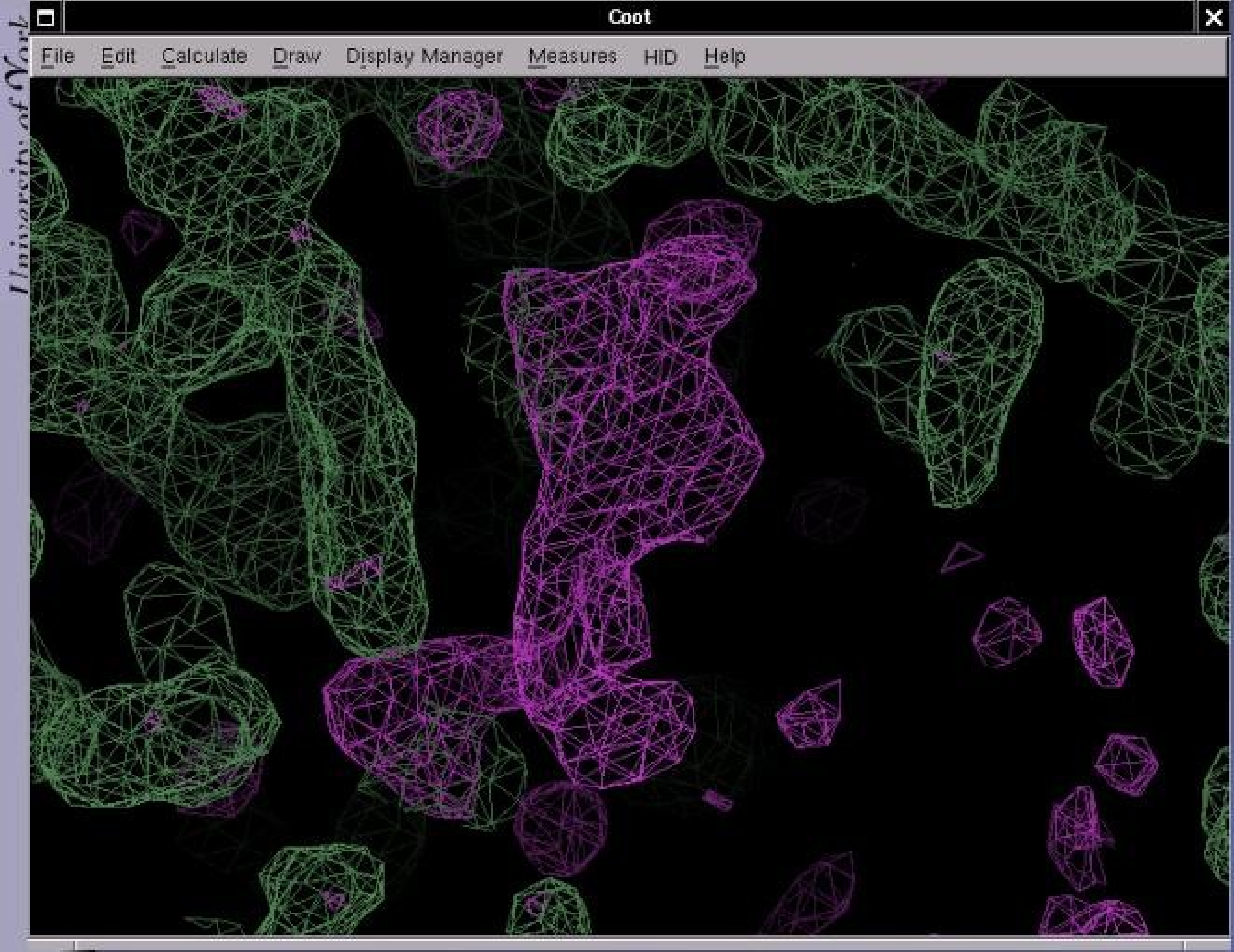
Coot

File Edit Calculate Draw Display Manager Measures HID Help



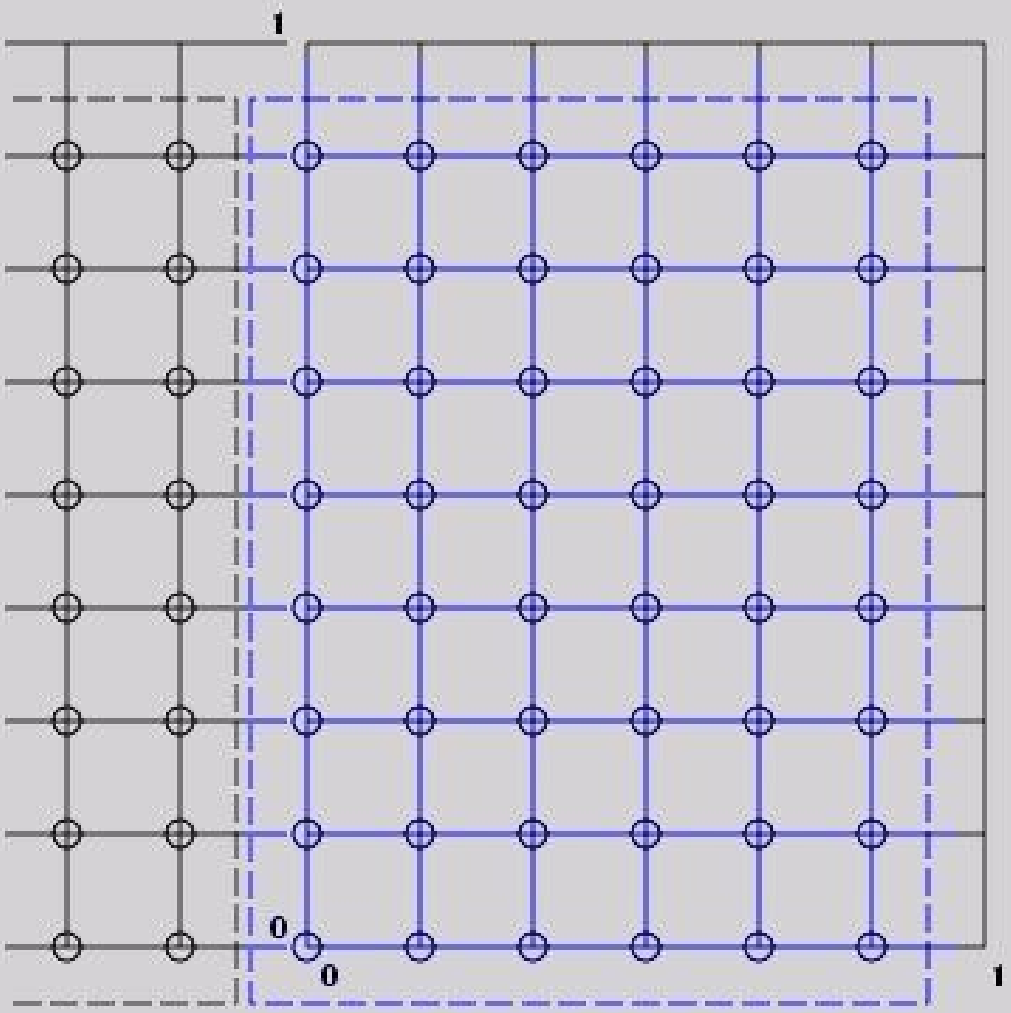

University of York
Coot
File
Edit
Calculate
Draw
Display Manager
Measures
HID
Help

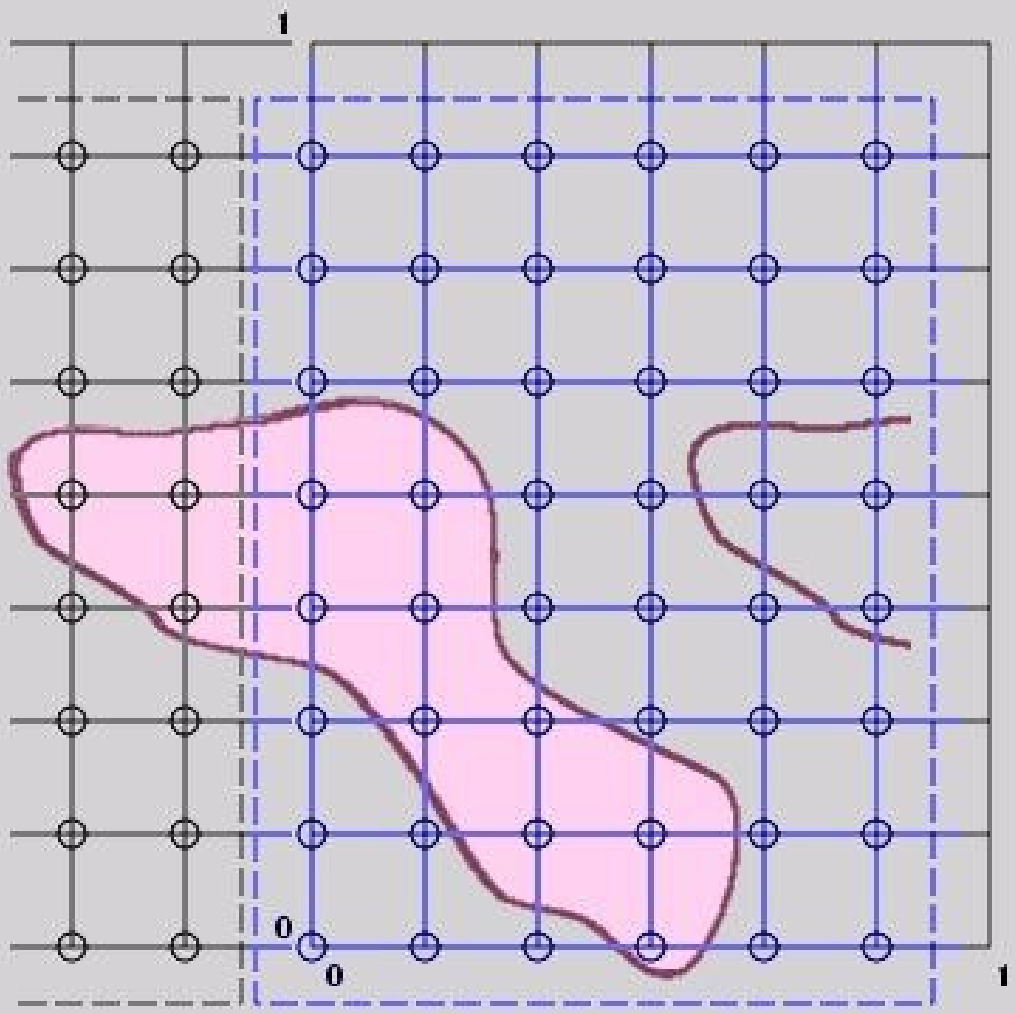

Coot
File
Edit
Calculate
Draw
Display Manager
Measures
HID
Help
University of York

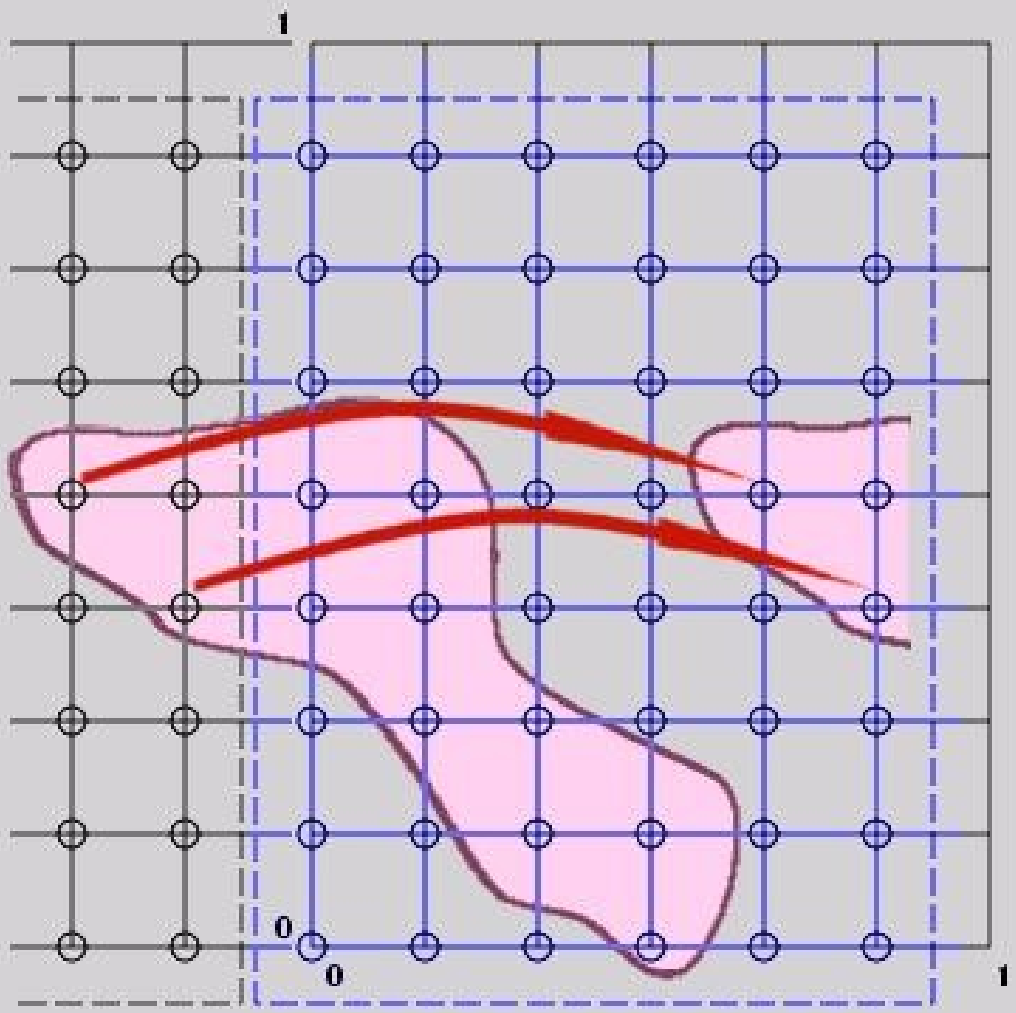

Coot
File
Edit
Calculate
Draw
Display Manager
Measures
HID
Help
University of York

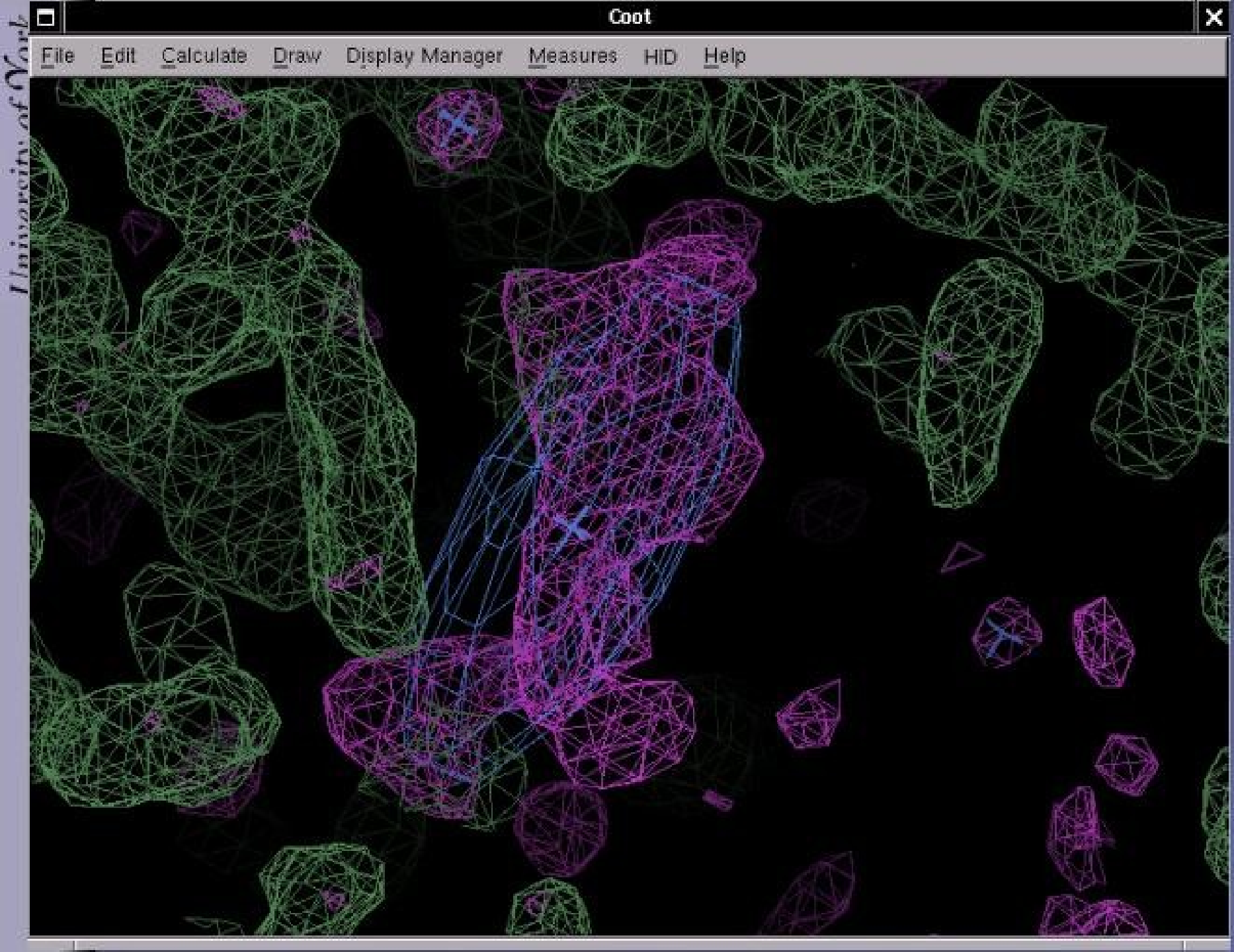
Clipper Map Mapping

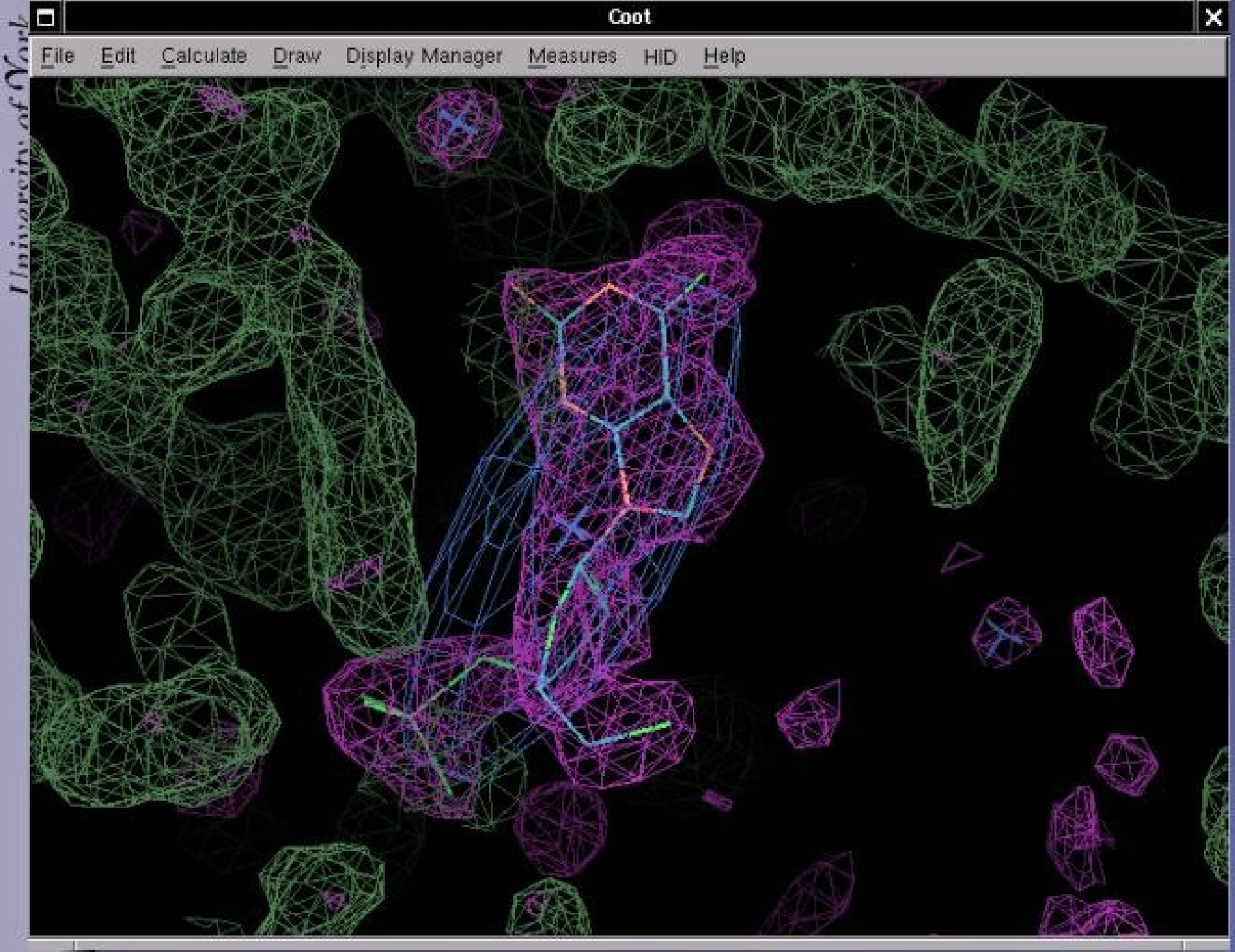
- Clipper maps
  - Appear to be “infinite”
  - Density value can be queried anywhere in space

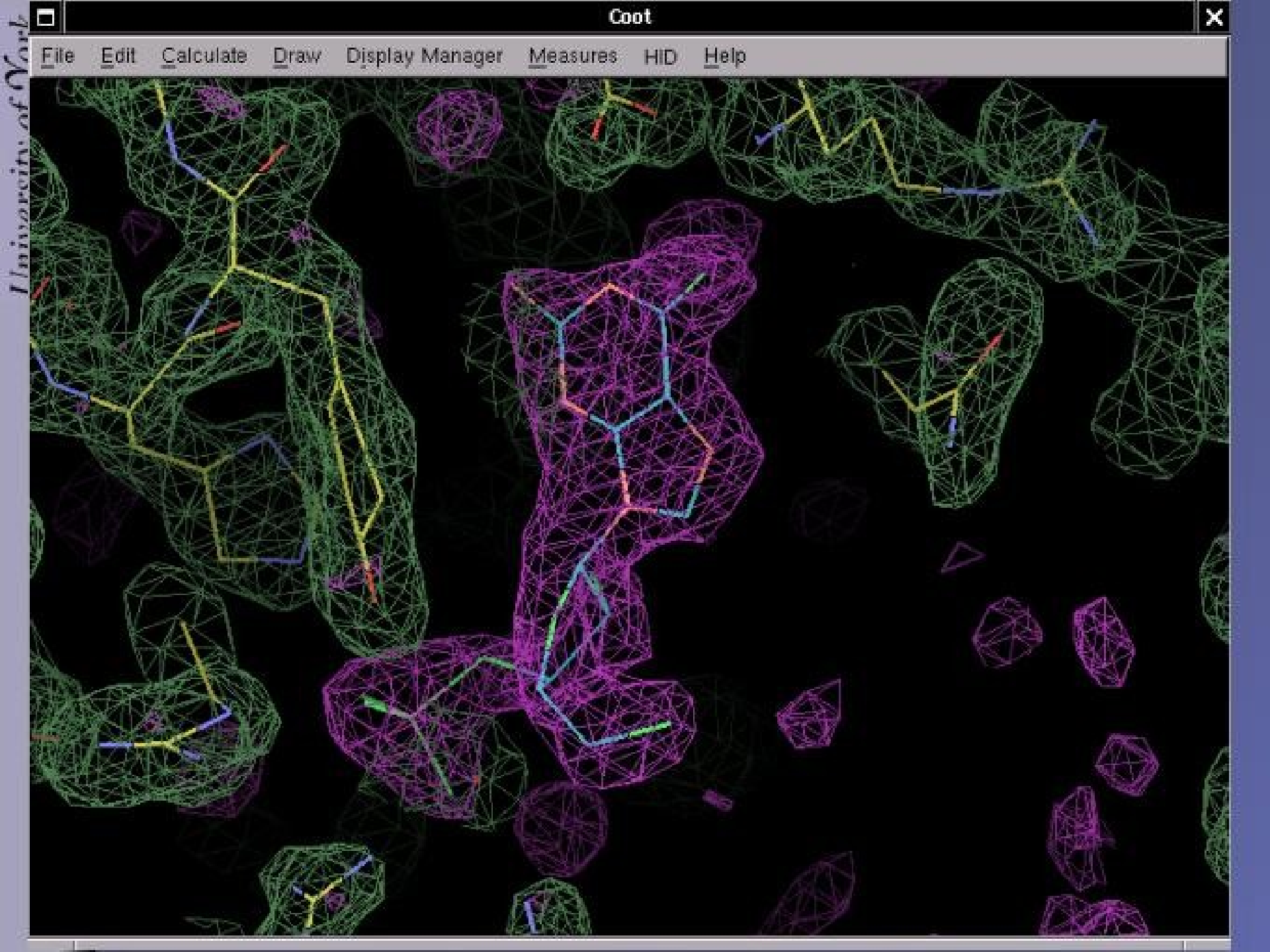

1
0
0
1


1
0
0
1


1
0
0
1


Coot
File
Edit
Calculate
Draw
Display Manager
Measures
HID
Help
University of York


Coot
File
Edit
Calculate
Draw
Display Manager
Measures
HID
Help
University of York


Coot
File
Edit
Calculate
Draw
Display Manager
Measures
HID
Help
University of York

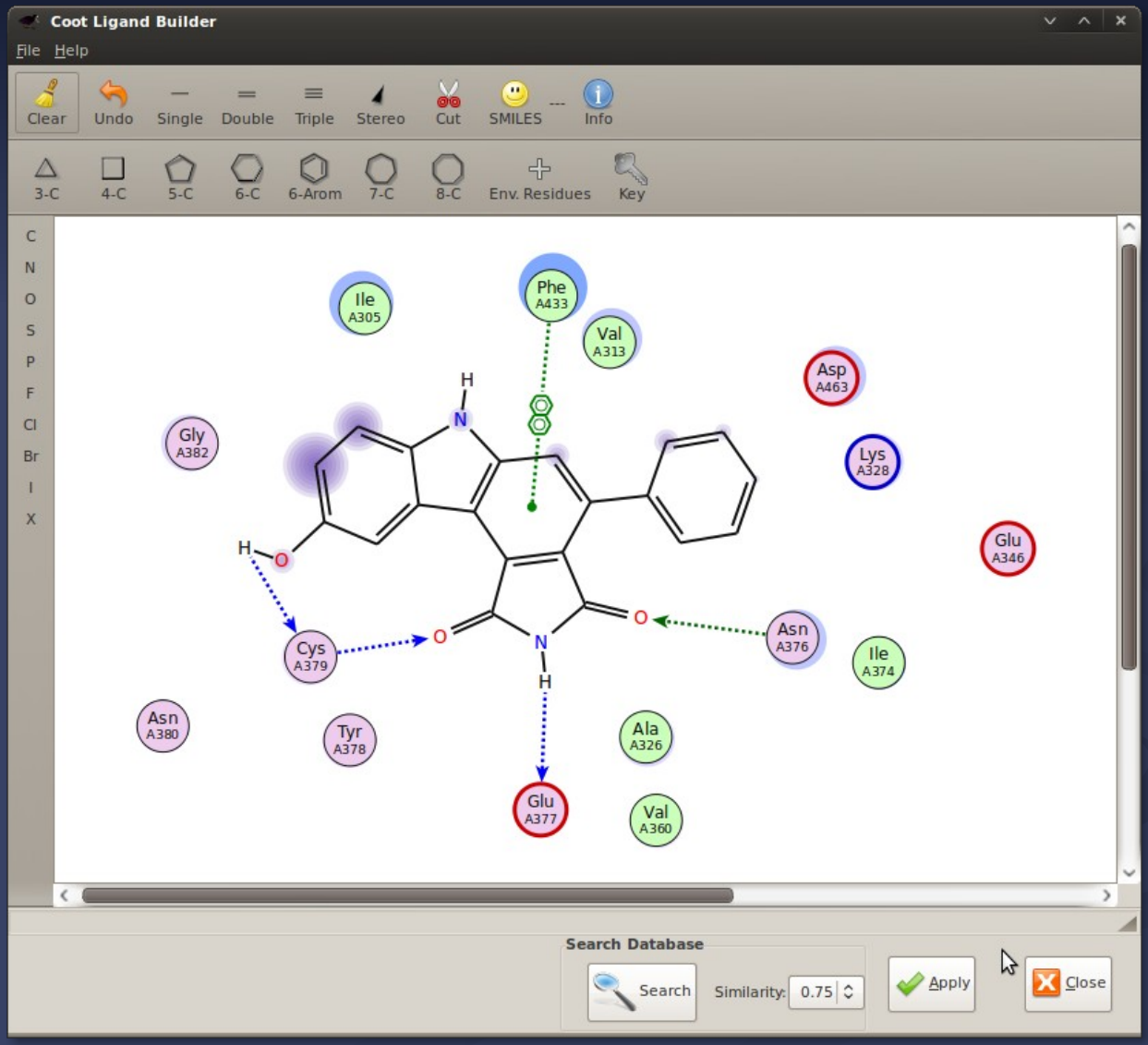
Conformation Idealization

- Each conformer is passed through the "Regularization" function of Coot
  - Non-bonded terms included
- Better to have hydrogen atoms on the model
- Slows things down a good deal…
  - May not be the best method to explore conformational variability for many rotatable bonds

Ligand Overlay

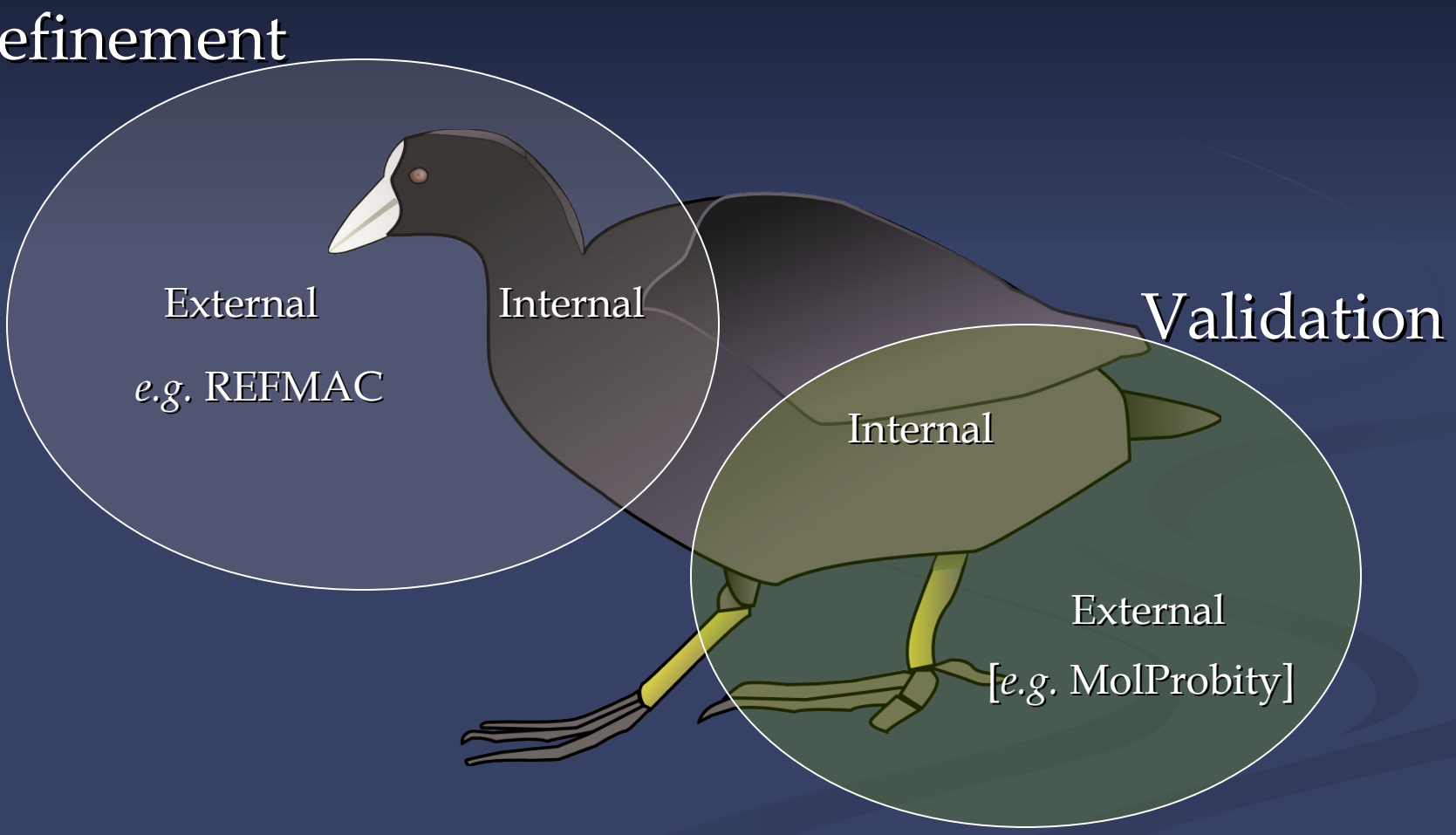
- Algorithm and Code by Eugene Krissinel
- Tries to overlay different ligands/monomers by graph matching
- Useful for “database” ligands where atom names are not selected by hand
- Has been used as the basis of the function which “mutates” residues to alternative monomer types
  - e.g. phosphorylation

New ligand view and builder

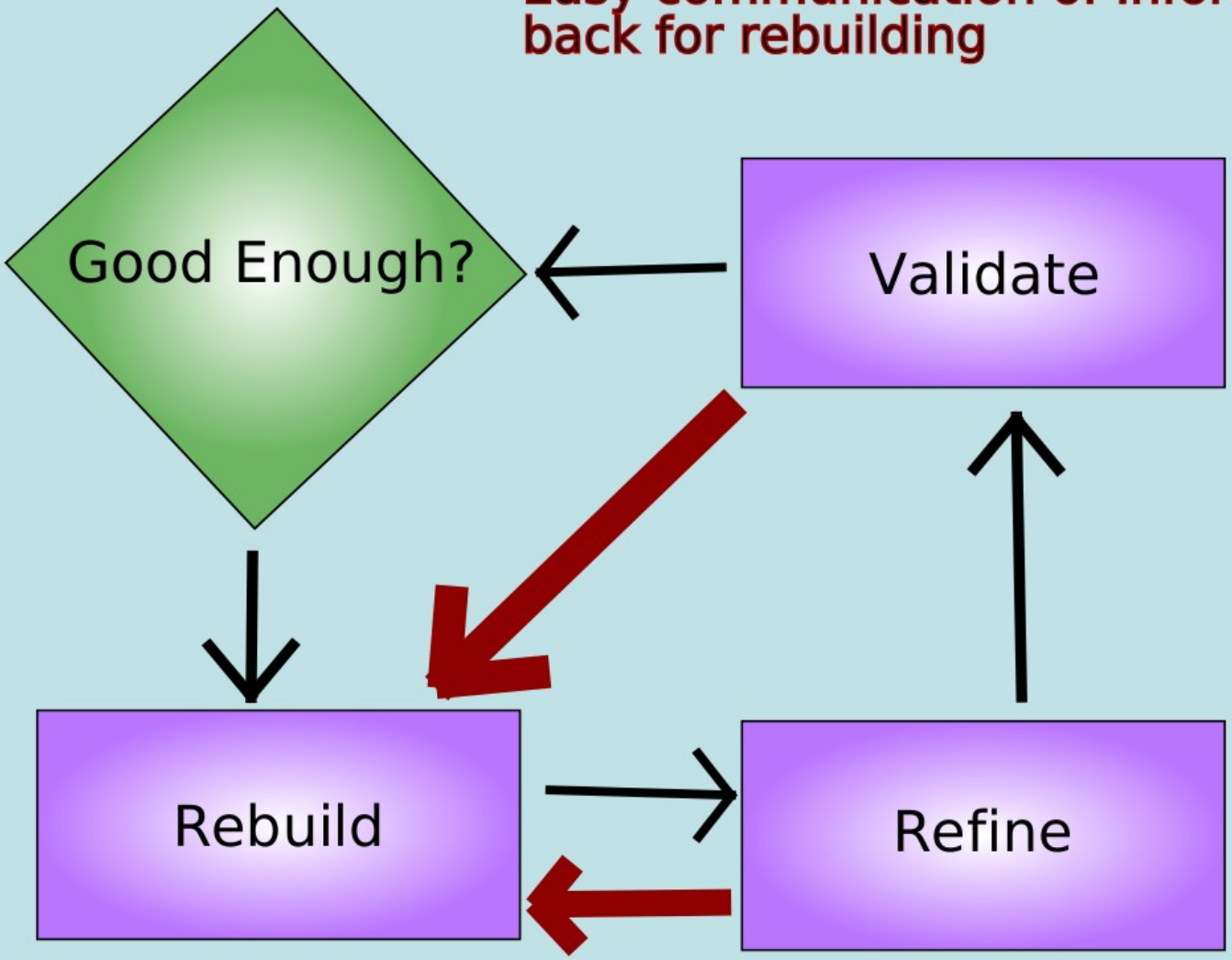


Coming soon...

Feature Integration



Easy communication of Information back for rebuilding



Validation...

What is Validation?

- Comparison of Various aspects of the model with pre-conceived notions of "good quality"
  - Includes **unrestrained** and **restrained** criteria
  - Many aspects of validation overlap with refinement and model-building

Why Validate?

- Model-building is error-prone
  - (although automated methods seem to do better)
- Someone else did the model-building
- The model was built several years ago
  - and the notion of "good quality" has changed
- Deposition requires validation

Observations to Parameters Ratio

- Some typical numbers
  - to 2Å, 22000 reflections
  - 200 residues x 10 (atoms/residue) x 4 params/atom
  - -> about 2.6
- To 3Å:
  - Ratio is about 1:1
- As statisticians, we prefer our models to be parsimonious

Depending on solvent content and the manner in which NCS is handled

A “good” model

- Makes statistical sense
  - The reciprocal space representation agrees tolerably well with the observations (R-factor)
  - No meaningful difference map peaks
- Makes Chemical sense
  - Model Geometry is consistent with the restraints
  - Ramachandran Plot has less than 1% outliers
  - A good **clashscore**
- Makes Biological sense
  - Residues in chemically sensible environment
  - Is consistent (on the whole) with external biochemistry observations (active site residues)

Quick Bayes

- Bayes Eq:
- $\Pr(\text{model} \mid \text{data}) \propto \Pr(\text{data} \mid \text{model}) * \Pr(\text{model})$
- $\Pr(\text{data} \mid \text{model})$ is also called the Likelihood, $L(\text{model} \mid \text{data})$

Validation Tools - Pr(model)

- Ramachandran Plot
  - Kleywegt Plot (NCS differences)
- Geometry Analysis
- Peptide ω Analysis
- Temperature Factor Analysis
- Rotamer Analysis
- [Clashes]

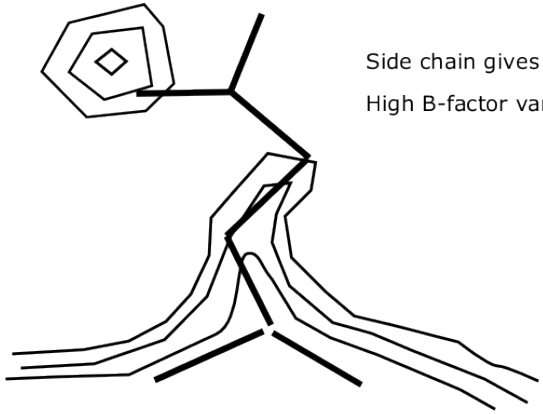
Rotamers

- Side-chains have certain preferred combinations of torsions round their rotatable bonds
- An analysis (batched around the staggered conformations) will give rotamer occurrence

Validation Tools - Pr(data|model)

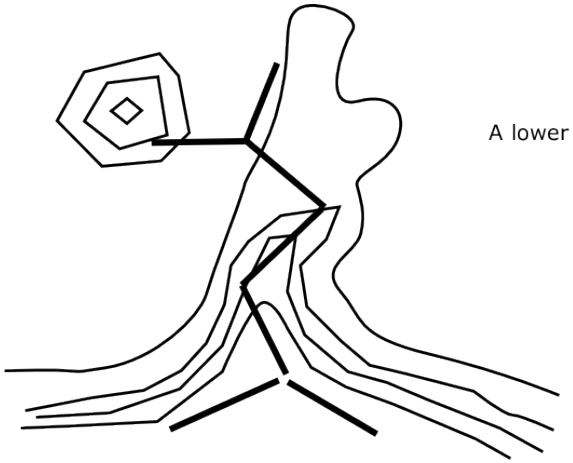
- Density Fit Analysis
- Difference Map Peaks
  - Variance analysis at Water Positions
- Unmodelled blobs

B-factor variance

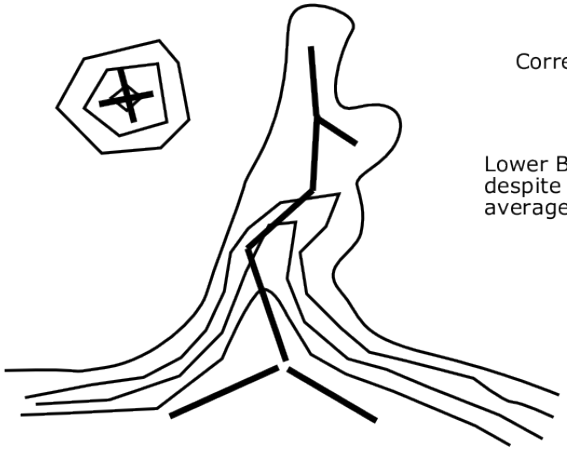


Side chain gives us concern

High B-factor variance



A lower contour level

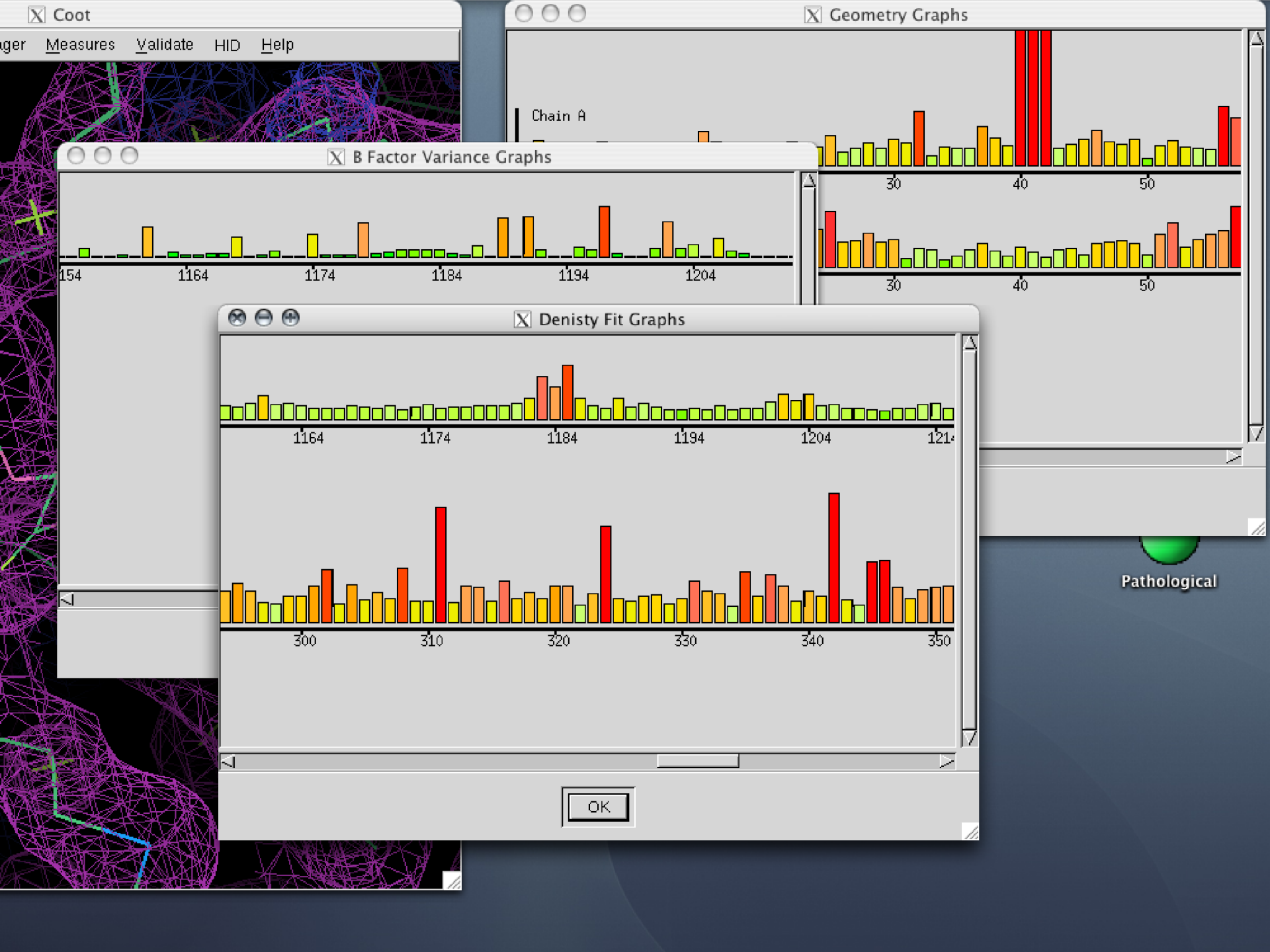


Correctly modelled side-chain

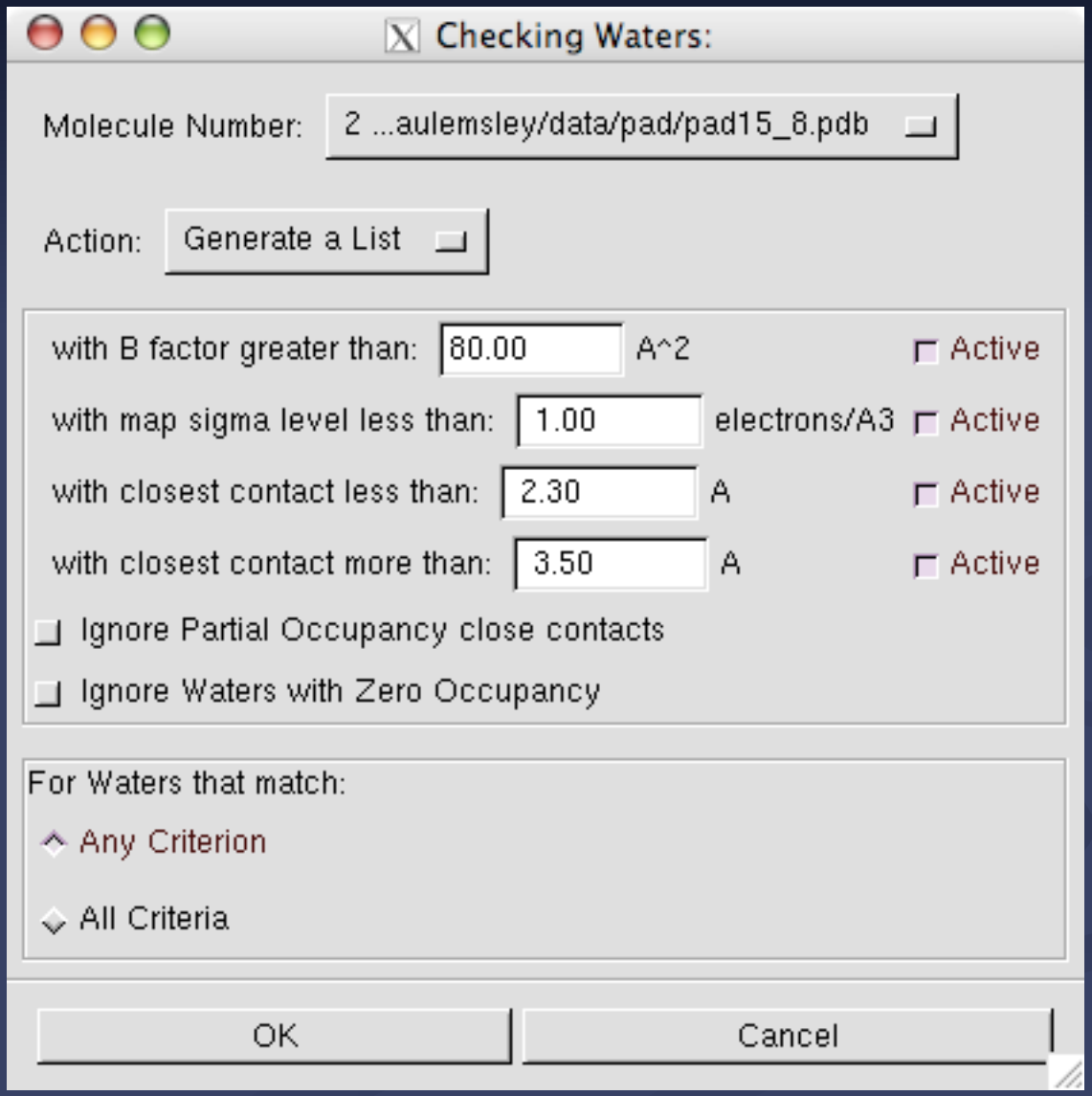
Lower B-factor variance
despite potentially higher
average B-factor

Chiral Volume Analysis

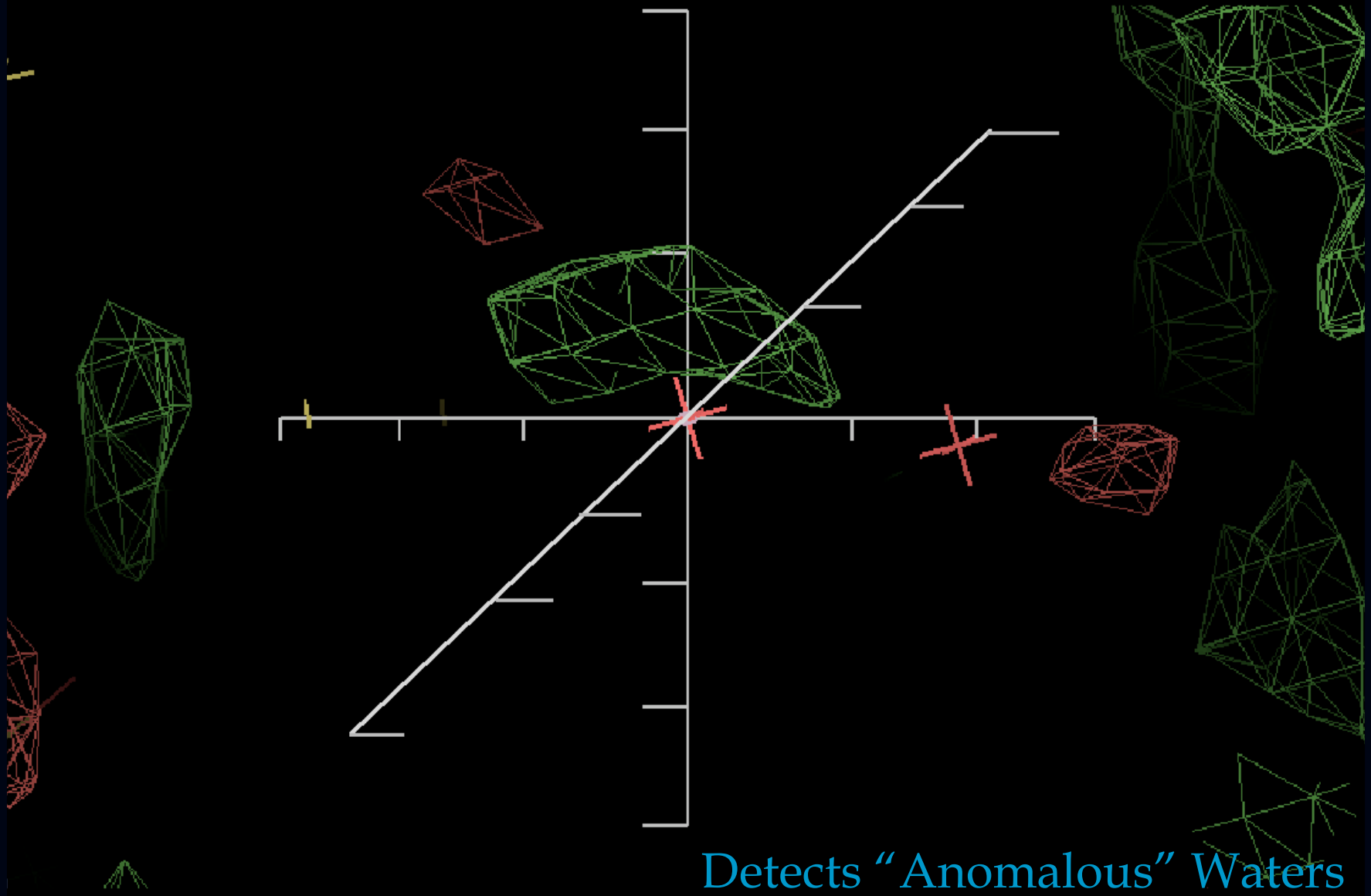
- Based on data in the Refmac dictionary
- …was needed because it was possible with Coot to accidentally invert Chiral centres
  - *e.g.* Cαs, Cβ (THR)
  - (Easily corrected with the Mutate & Autofit tool)
  - These days we have chiral volume restraints


Coot
ger Measures Validate HID Help
Geometry Graphs
Chain A
30
40
50
30
40
50
B Factor Variance Graphs
154
1164
1174
1184
1194
1204
Denisty Fit Graphs
1164
1174
1184
1194
1204
300
310
320
330
340
350
OK
Pathological

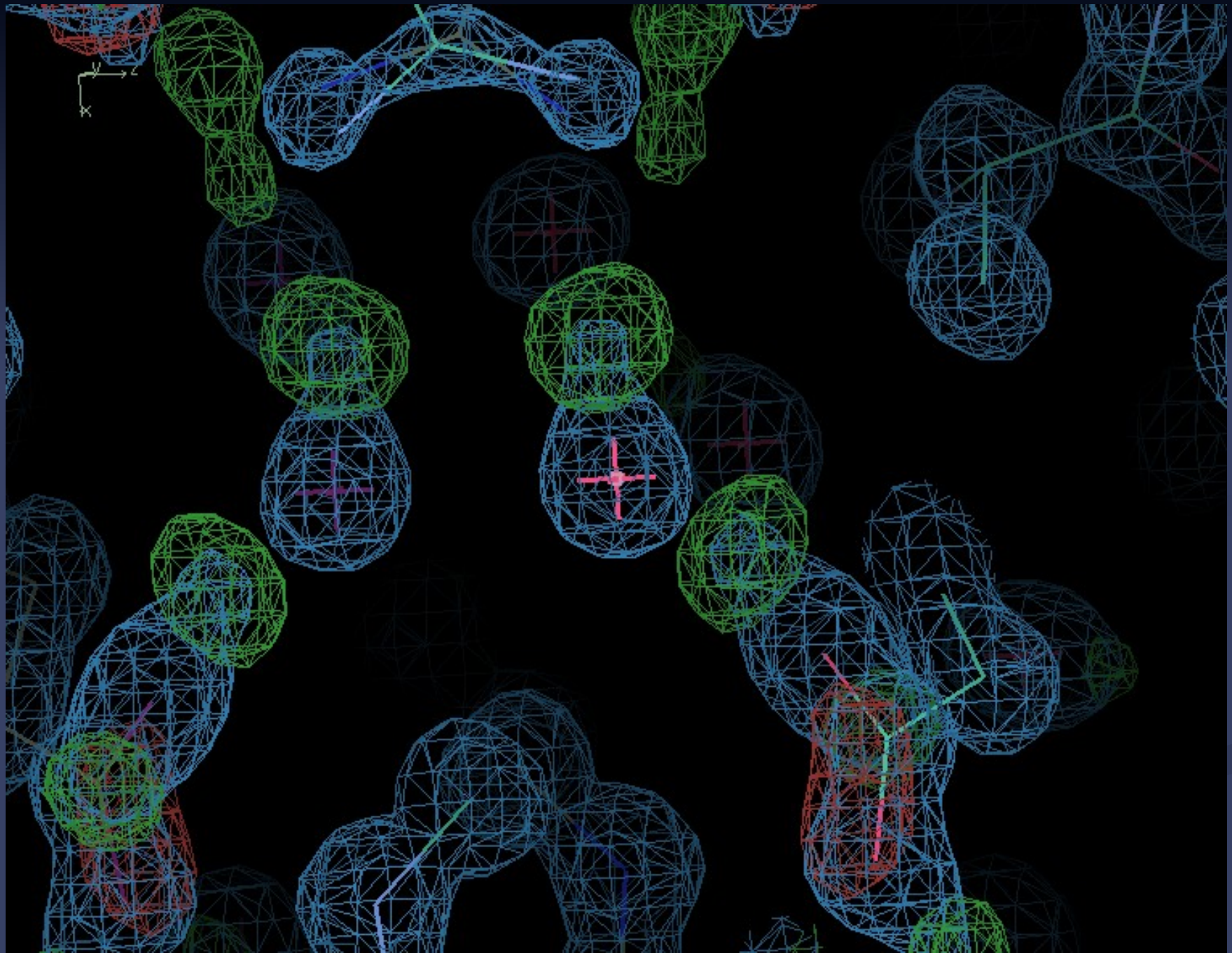
Check/Delete Waters

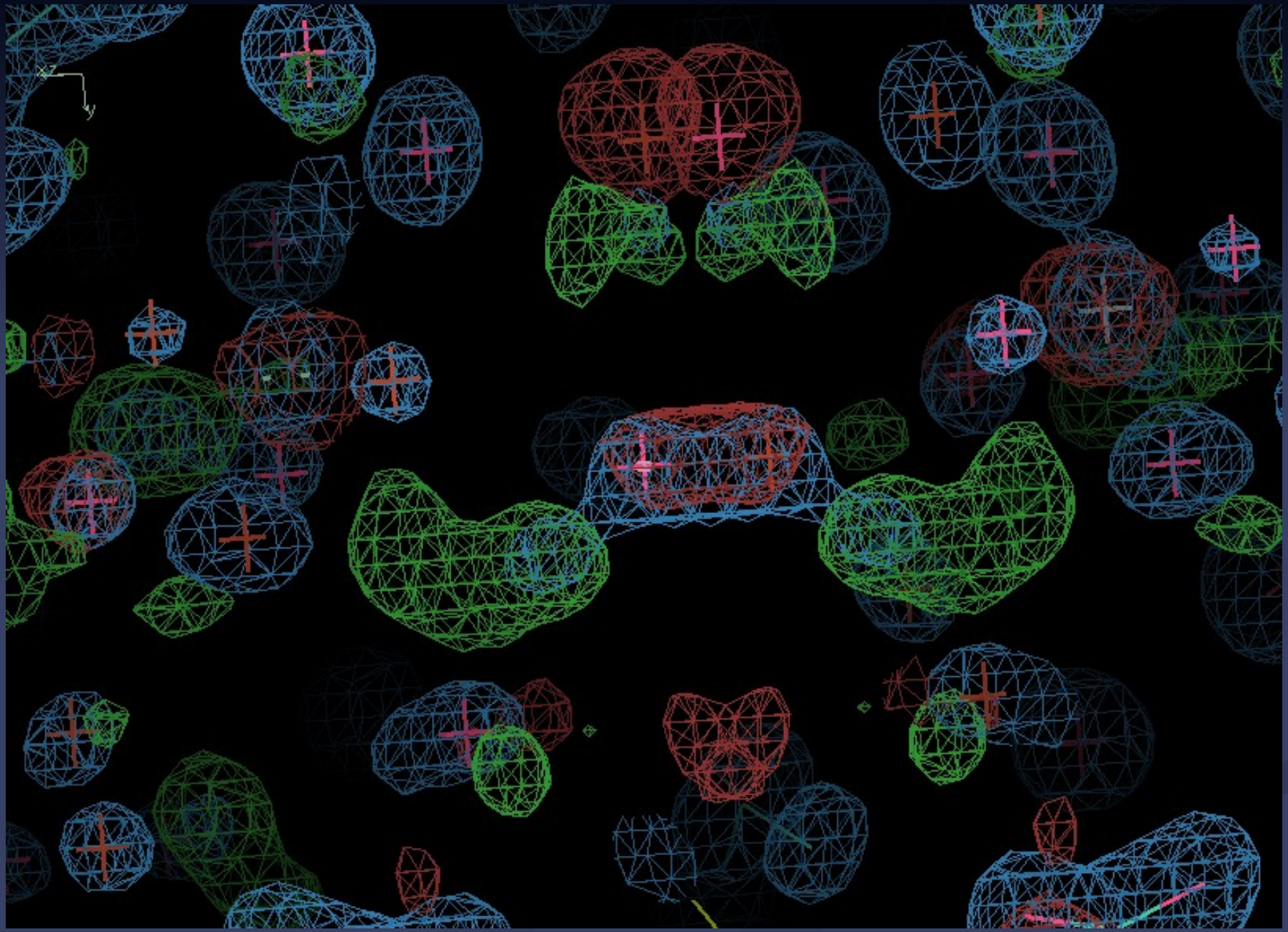


Difference Map Sampling

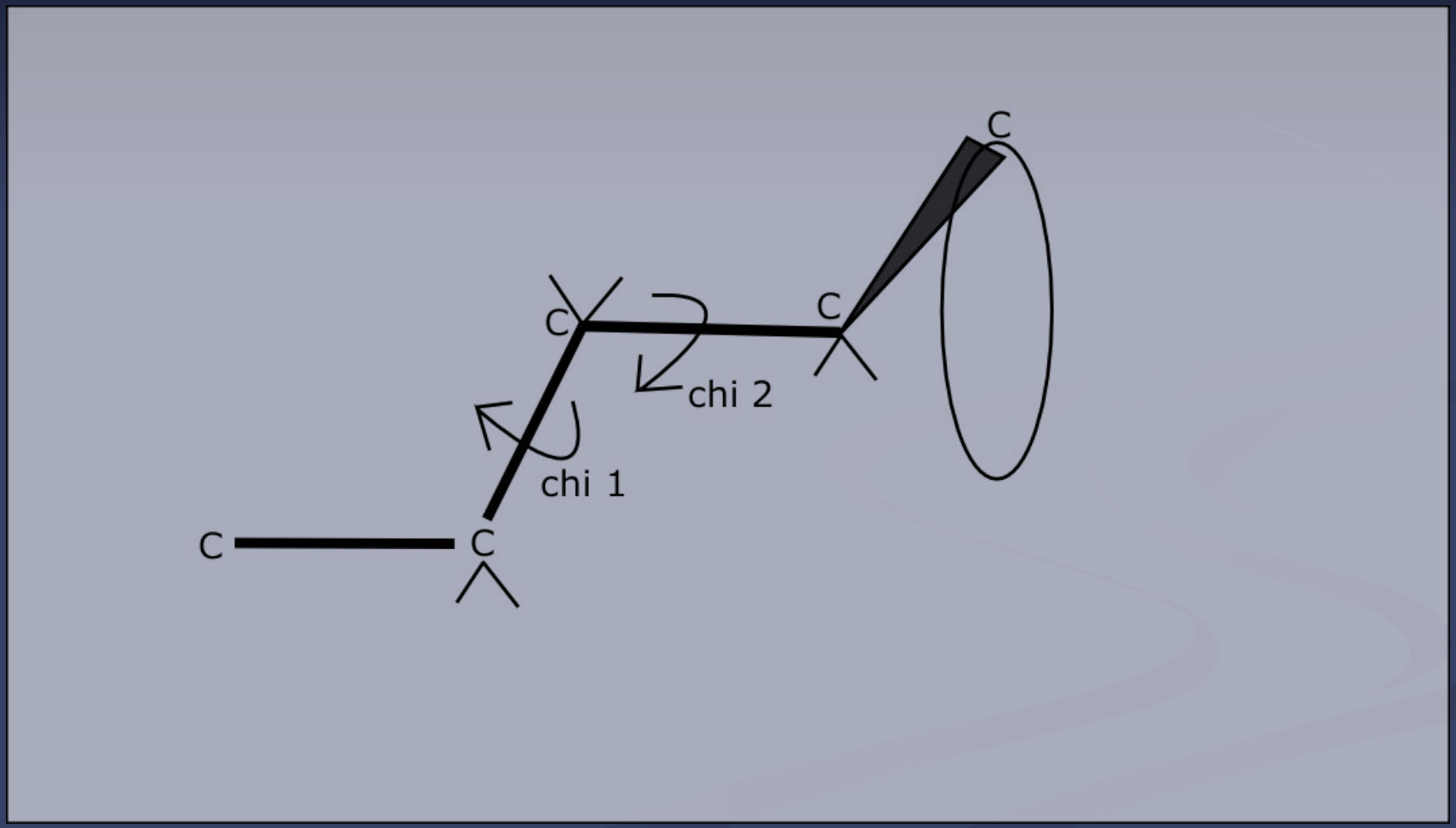


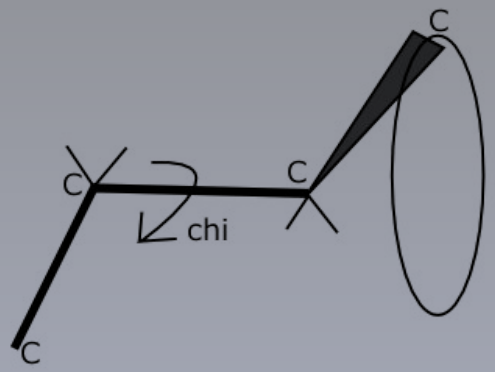
Detects “Anomalous” Waters


Y
Z
X



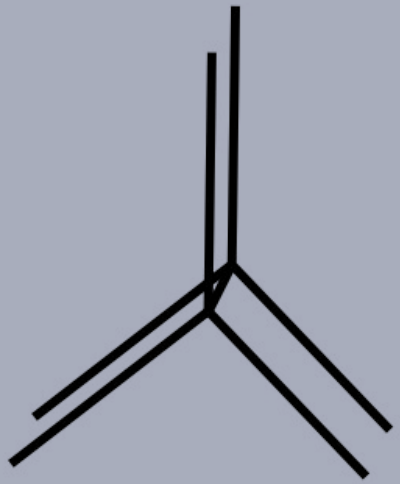
Torsion-based Validation





In principle, there is free rotation

In practice, staggered is energetically more favoured

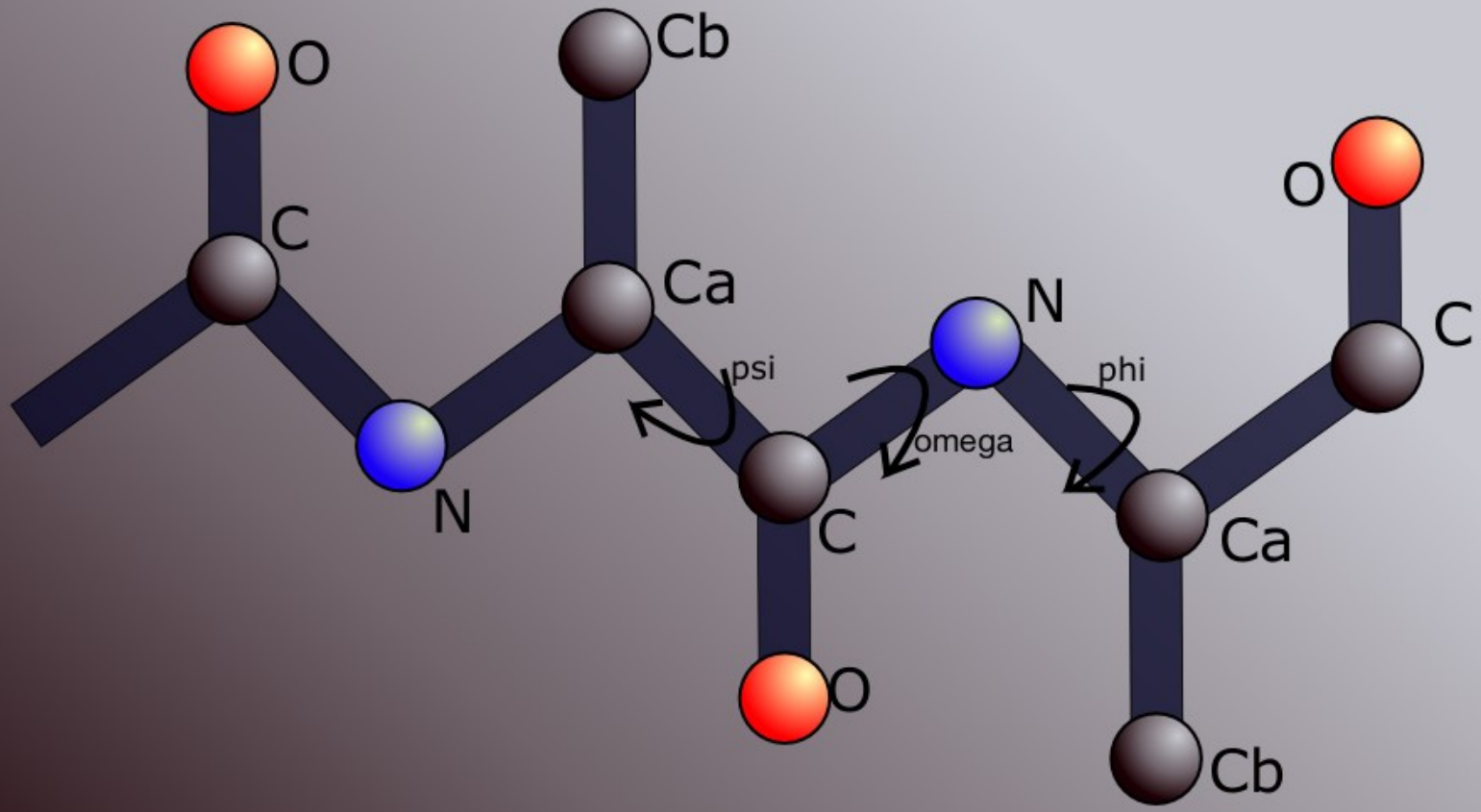


Eclipsed

Staggered

Most favoured staggering angles 60, 180, -60 degrees

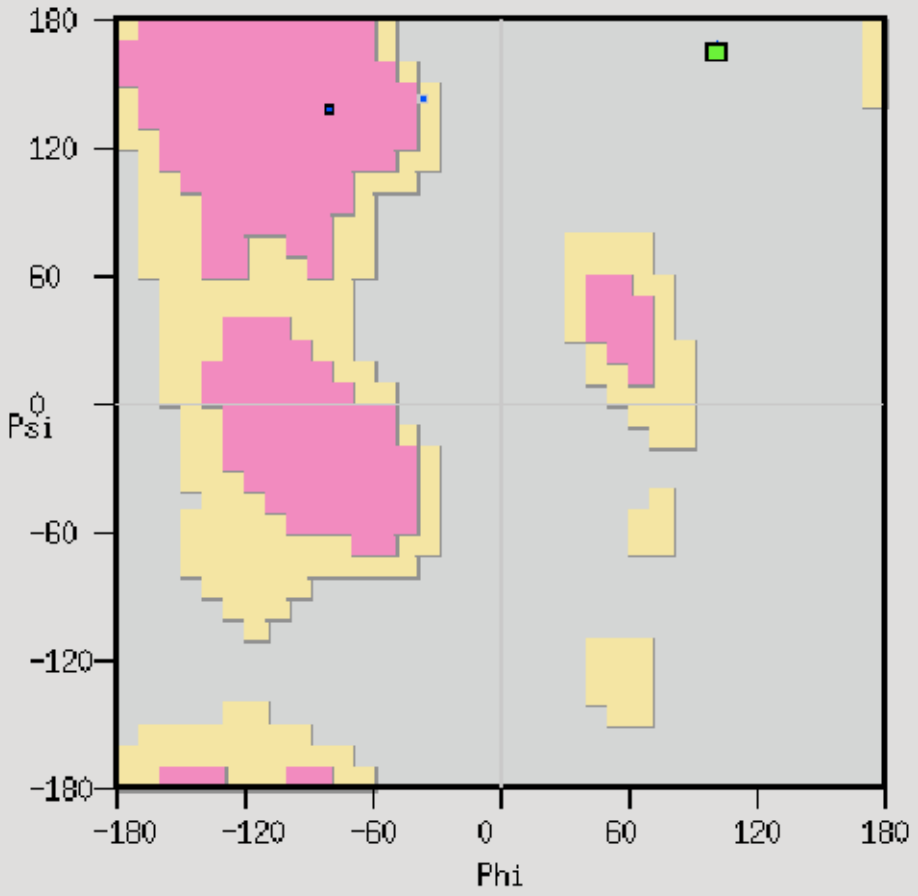
Peptide Torsion Angles



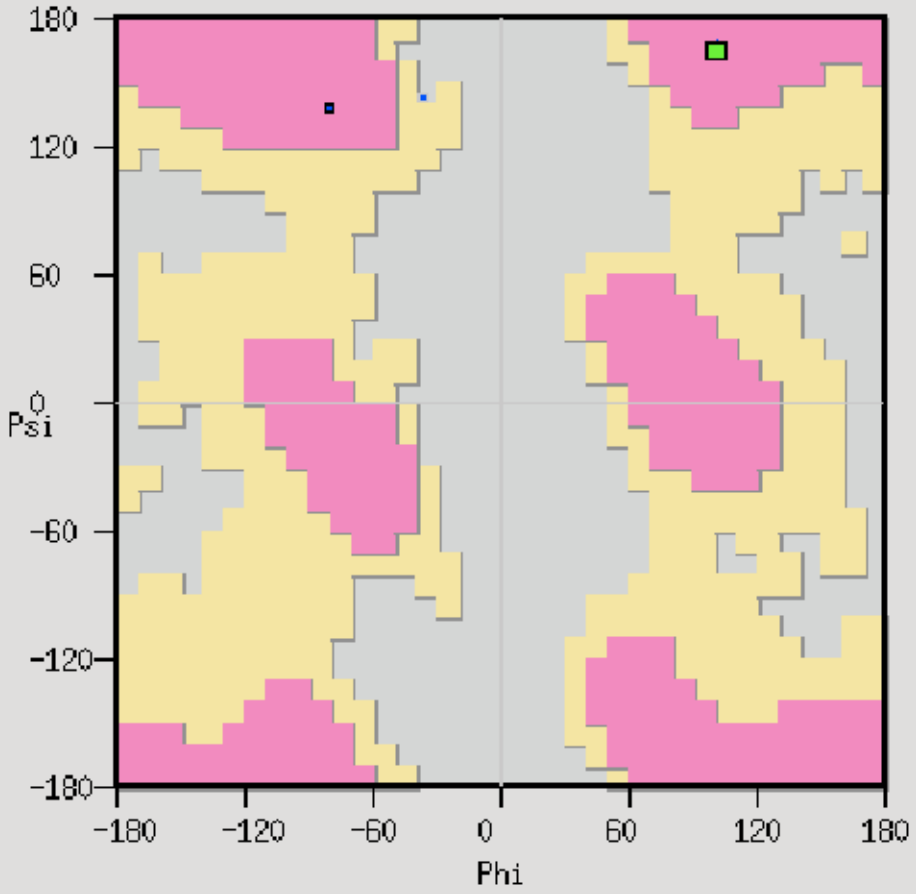
Peptide ω

- Needed to check the planarity of the peptide link
  - At low resolutions it is possible to give the protein lots of (too much) freedom to optimize the fit to the density
- Can accidentally create CIS peptides
  - When discovered they are easily reconverted using the CIS<->TRANS peptide tool
- Less accidents happen when peptide plane restraints are applied

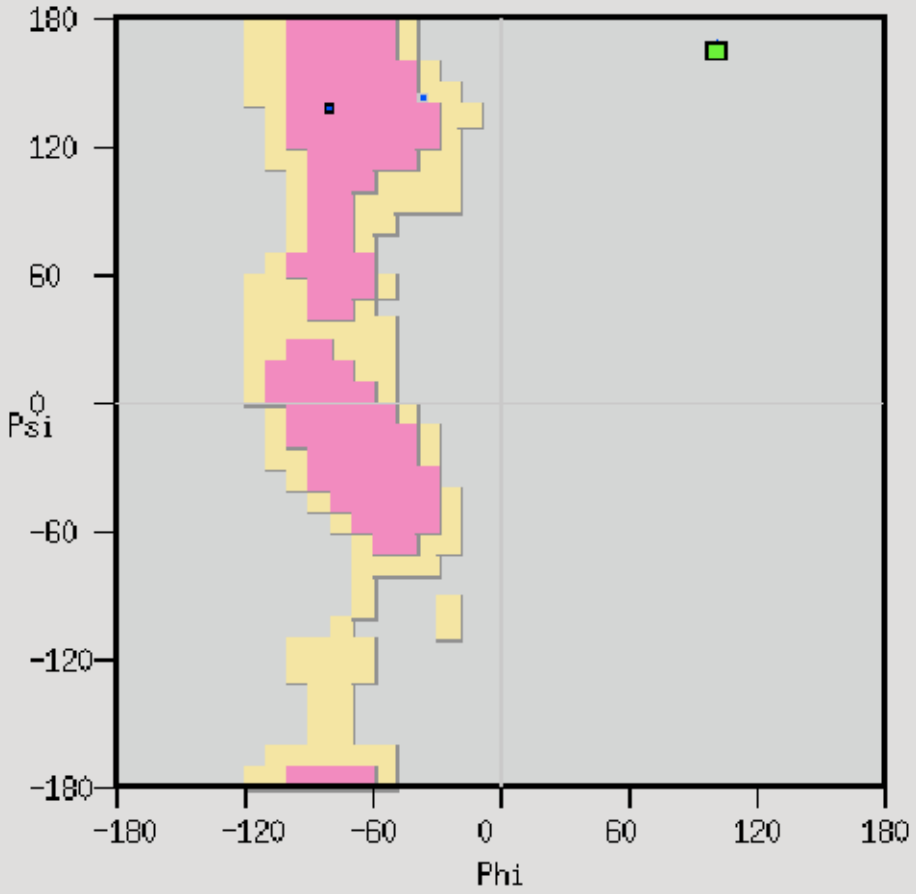
Ramachandran Plot for residues with CB

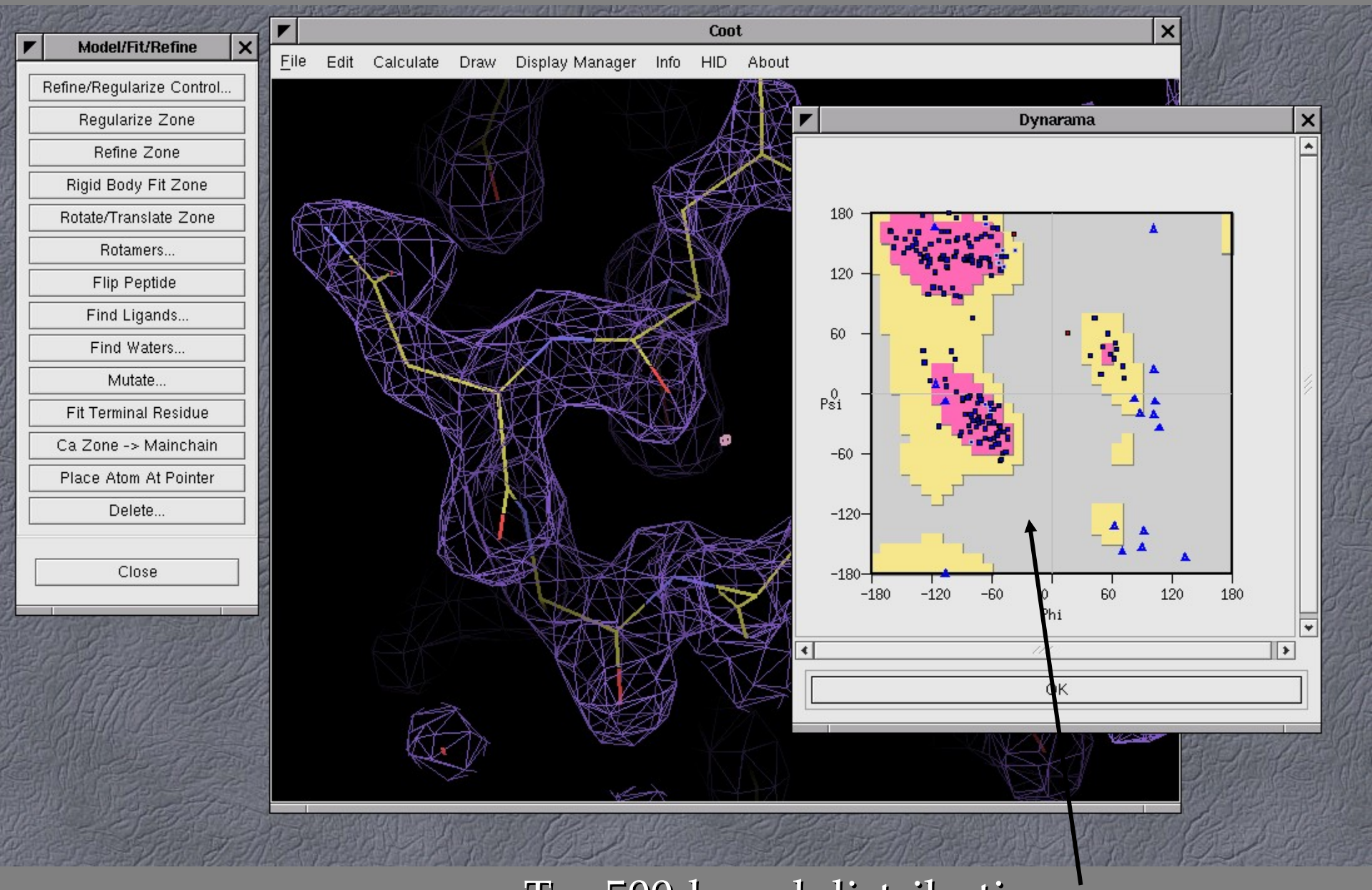


Ramachandran Plot for GLY



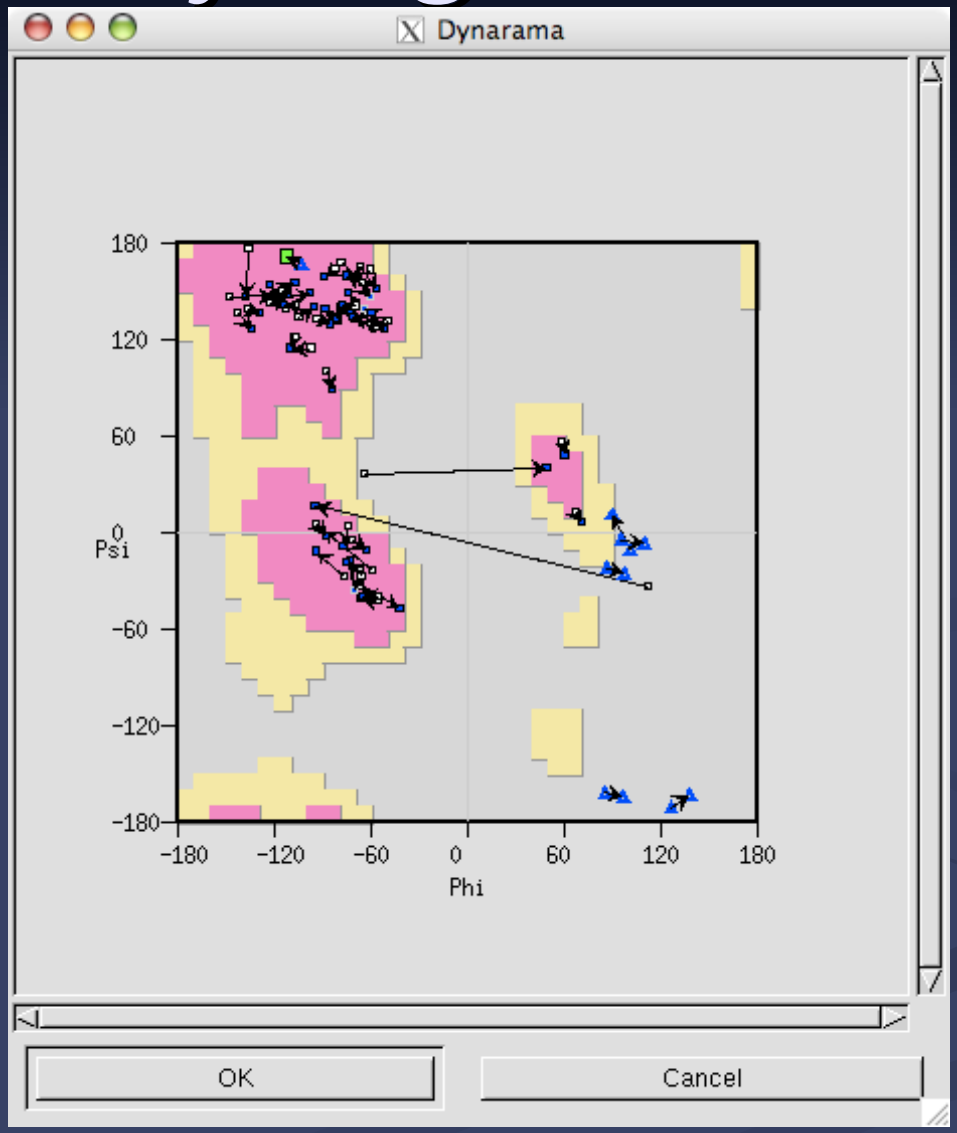
Ramachandran Plot for PRO





Top500-based distribution

Kleywegt Plots[*]



[*] Named by George Sheldrick

More Validation Pr(model)

- Coot has interface to Molprobity
  - (Molprobity is the widely regarded as the best model validation suite)
- Uses identical Ramachandran plot
- Uses identical Rotamer library
- Coot reads probe dots directly

Analyzed all-atom contacts and geometry for 1sarH.pdb

Analyzed all-atom contacts and geometry for 1sarH.pdb

Entry begun: Today at 1:49pm EST

Last modified: Today at 1:49pm EST

Summary statistics

All-Atom Contacts	Clashscore, all atoms:	12.49	60th percentile* (N=837, 1.55Å - 2.05Å)
	Clashscore, B<40:	10.76	41st percentile* (N=837, 1.55Å - 2.05Å)
Protein Geometry	Rotamer outliers	1.83%	Goal: <1%
	Ramachandran outliers	0.00%	Goal: <0.2%
	Ramachandran favored	99.47%	Goal: >98%
	Cβ deviations >0.25Å	11	Goal: 0
	MER [ALPHA TEST - don't ask]	1.81	78th percentile* (N=11444, 1.55Å - 2.05Å)

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

Multi-criterion visualizations

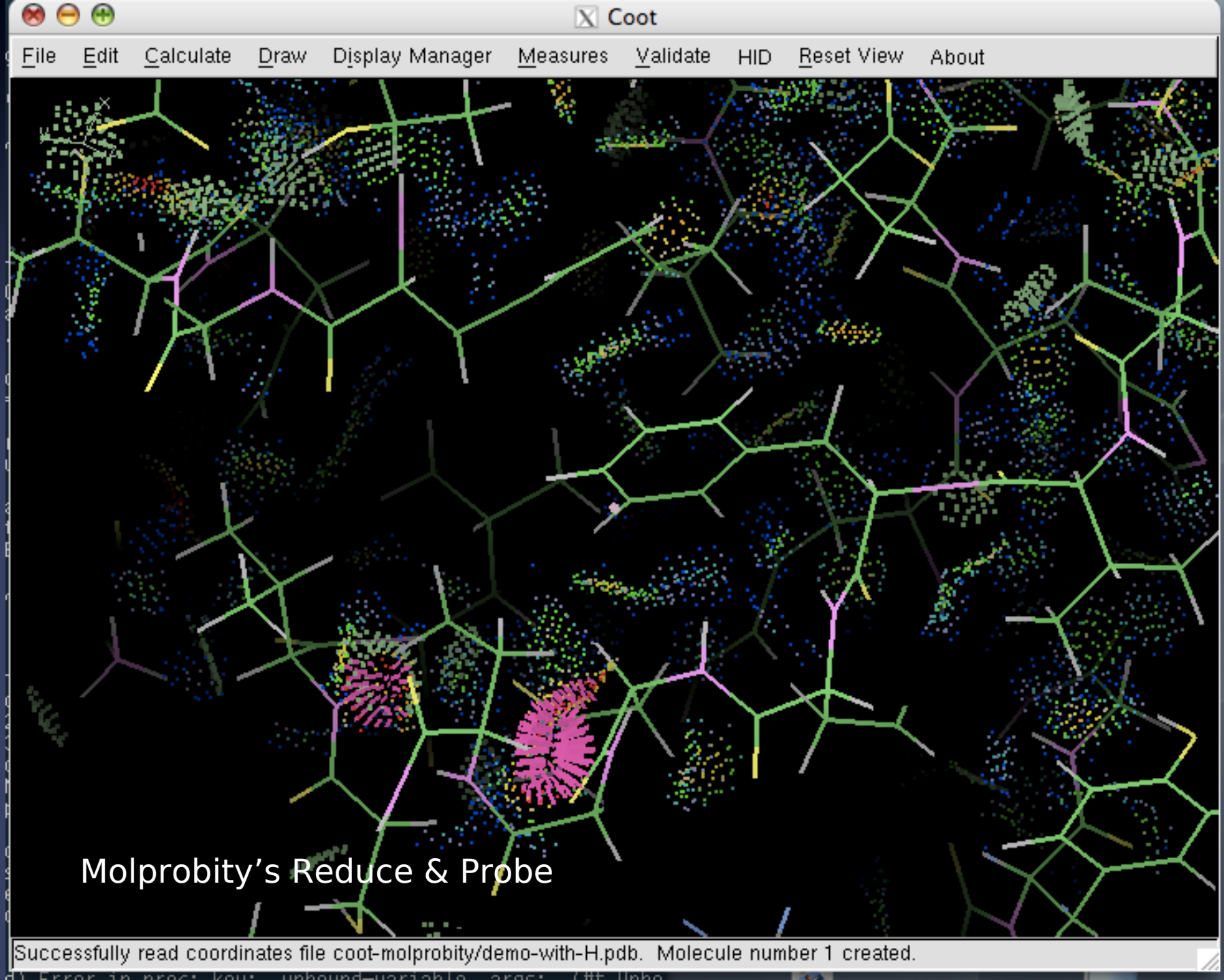
Multi-criterion kinemage (970 Kb): View in KiNG | Download

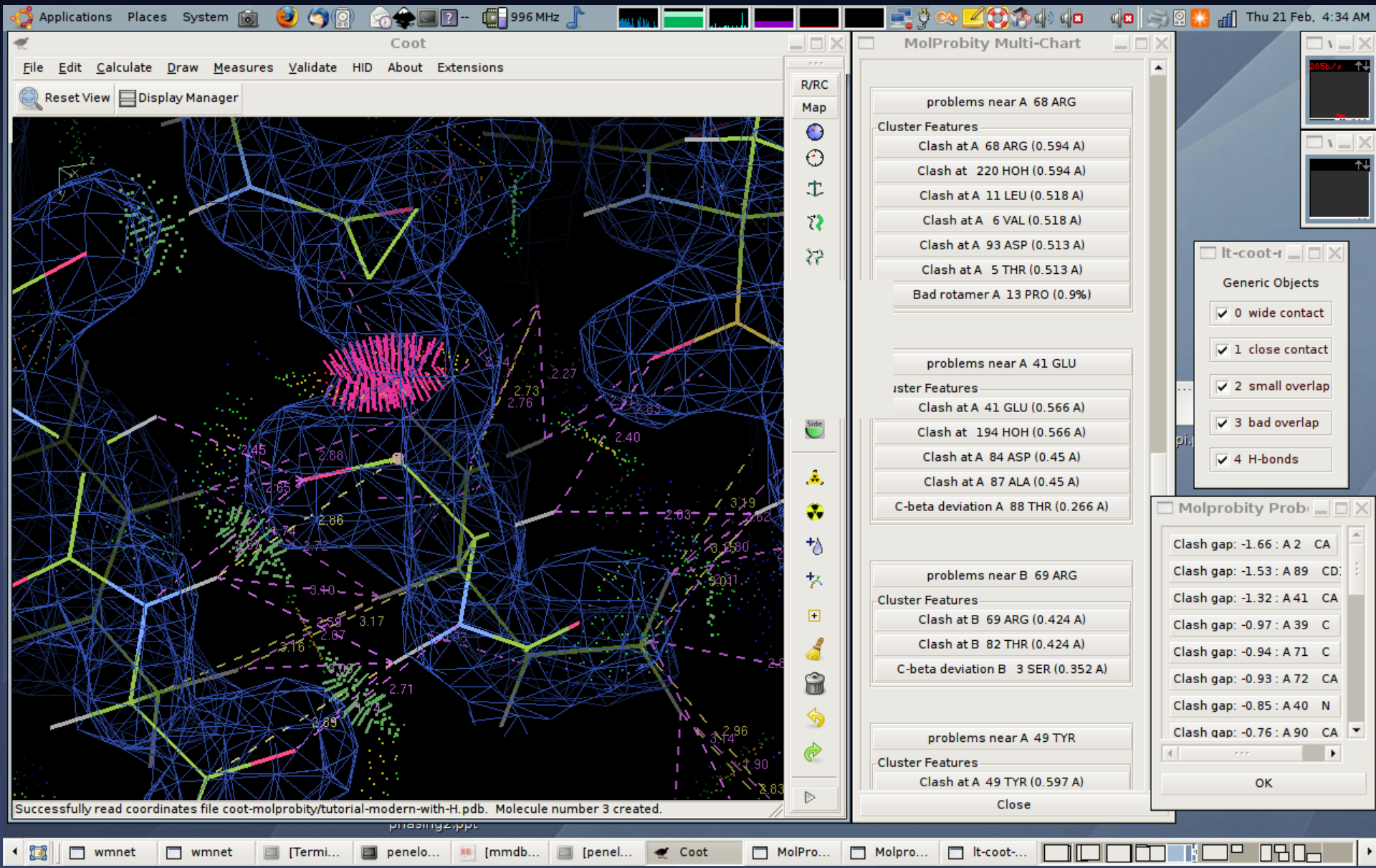
Multi-criterion chart

Download multi-criterion to-do list for Coot [ALPHA TEST]
Open this in Coot using Calculate | Run Script...

Single-criterion visualizations

- Clash list
- **Ramachandran plot kinemage** (344 Kb): View in KiNG | Download
- Ramachandran plot PDF
- **Cβ deviation scatter plot (2D)** (19 Kb): View in KiNG | Download


Coot
File
Edit
Calculate
Draw
Display Manager
Measures
Validate
HID
Reset View
About
Molprobity’s Reduce & Probe
Successfully read coordinates file coot-molprobity/demo-with-H.pdb. Molecule number 1 created.


Applications Places System 996 MHz Thu 21 Feb, 4:34 AM
Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View Display Manager
R/RC
Map
Side
MolProbity Multi-Chart
problems near A 68 ARG
Cluster Features
Clash at A 68 ARG (0.594 A)
Clash at 220 HOH (0.594 A)
Clash at A 11 LEU (0.518 A)
Clash at A 6 VAL (0.518 A)
Clash at A 93 ASP (0.513 A)
Clash at A 5 THR (0.513 A)
Bad rotamer A 13 PRO (0.9%)
problems near A 41 GLU
ıster Features
Clash at A 41 GLU (0.566 A)
Clash at 194 HOH (0.566 A)
Clash at A 84 ASP (0.45 A)
Clash at A 87 ALA (0.45 A)
C-beta deviation A 88 THR (0.266 A)
problems near B 69 ARG
Cluster Features
Clash at B 69 ARG (0.424 A)
Clash at B 82 THR (0.424 A)
C-beta deviation B 3 SER (0.352 A)
problems near A 49 TYR
Cluster Features
Clash at A 49 TYR (0.597 A)
Close
lt-coot-r
Generic Objects
0 wide contact
1 close contact
2 small overlap
3 bad overlap
4 H-bonds
Molprobity Prob
Clash gap: -1.66 : A 2 CA
Clash gap: -1.53 : A 89 CD
Clash gap: -1.32 : A 41 CA
Clash gap: -0.97 : A 39 C
Clash gap: -0.94 : A 71 C
Clash gap: -0.93 : A 72 CA
Clash gap: -0.85 : A 40 N
Clash gap: -0.76 : A 90 CA
OK
Successfully read coordinates file coot-molprobity/tutorial-modern-with-H.pdb. Molecule number 3 created.
wmnet wmnet [Termi... penelo... [mmdb... [penel... Coot MolPro... Molpro... lt-coot-...

Other Programs

- Moprobity Suite
  - molprobity.biochem.duke.edu
- WHATCHECK
- VERIFY-3D

Acknowledgements

- Paul Emsley
- Kevin Cowtan
- Eleanor Dodson
- Keith Wilson

http://www.biop.ox.ac.uk/coot/

or

Google: Coot

or for WinCoot

http://www.ysbl.ac.uk/~lohkamp/coot

- Libraries, dictionaries
  - Alexei Vagin, Eugene Krissinel, Stuart McNicholas
  - Dunbrack, Richardsons
- Coot Builders and Testers
  - William Scott, Ezra Peisach
  - York YSBL, Dundee, Glasgow (early adopters)
  - Coot Mailing List subscribers