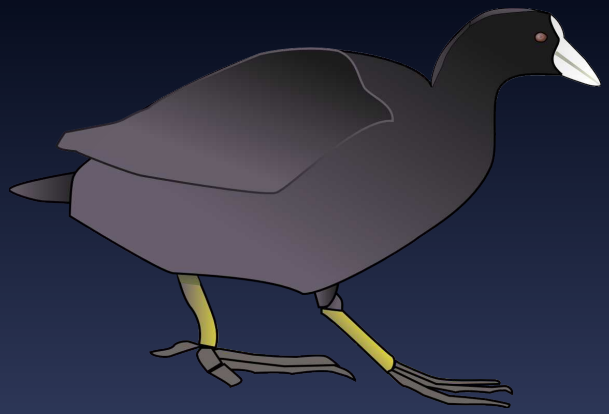


December 2011 OIST

Ligands and validation with Coot

Bernhard Lohkamp

Karolinska Institutet

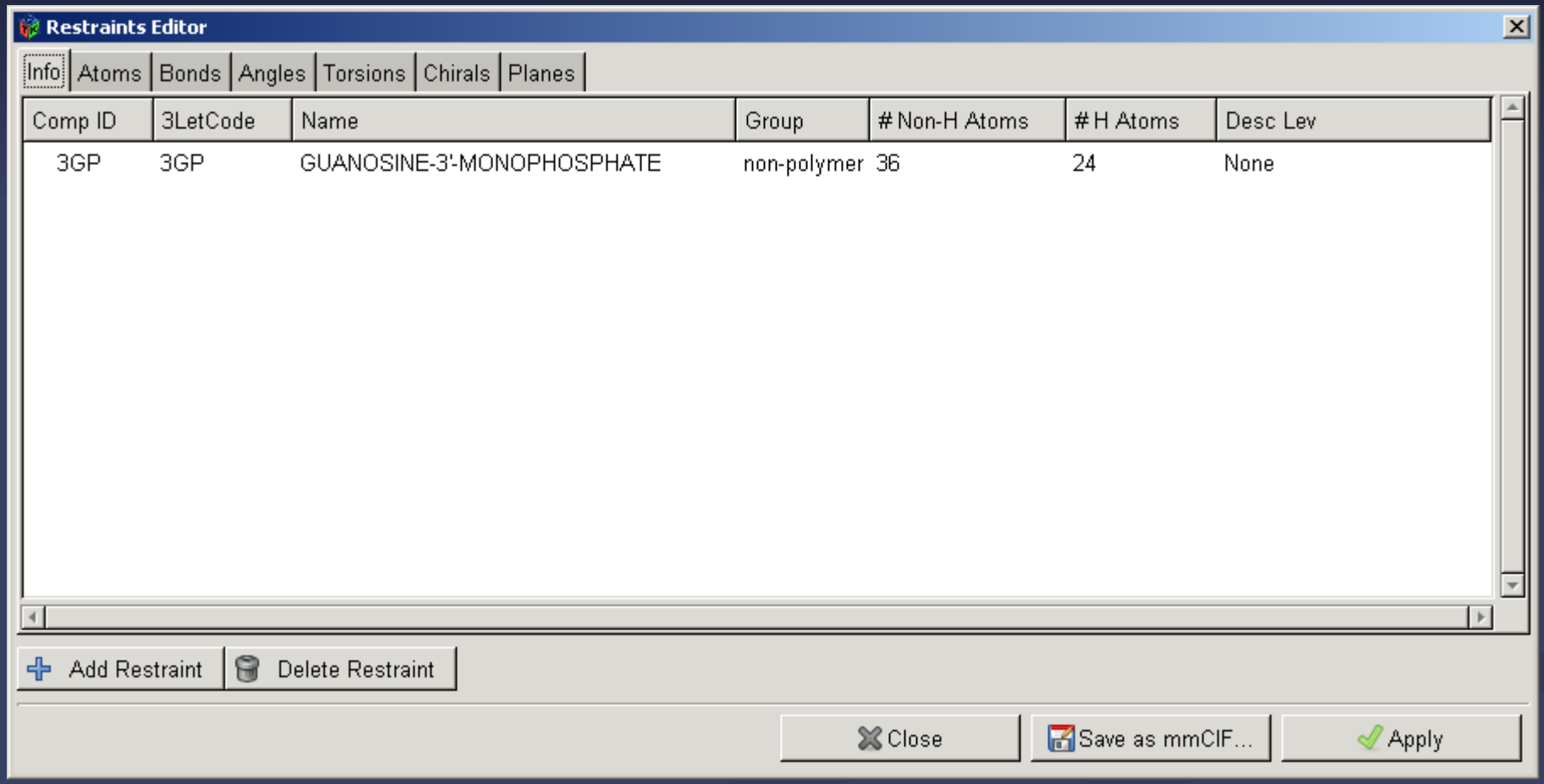
Ligands in Coot

- Fitting ligand
- Ligand validation
- Ligand representation
- Ligand builder

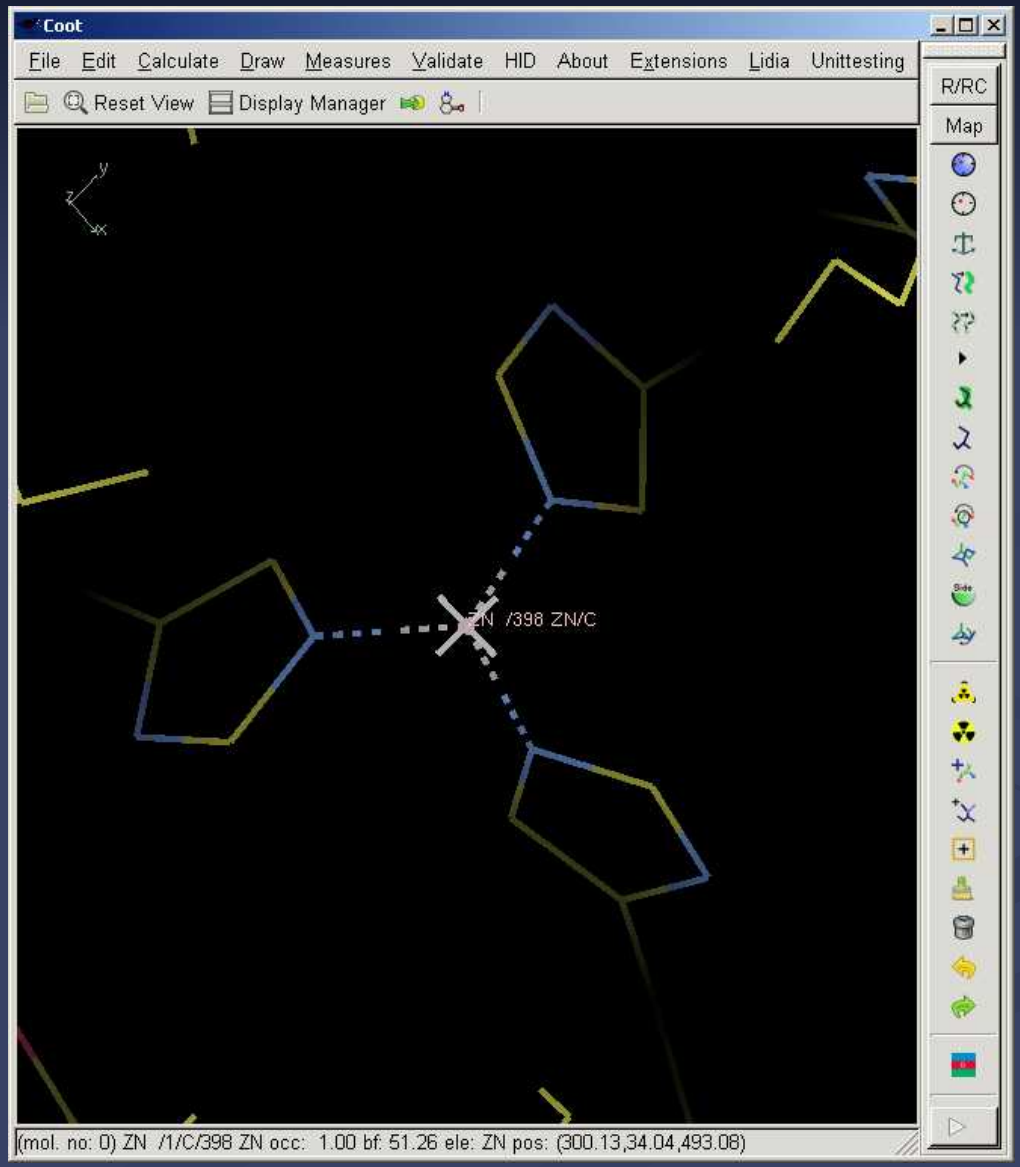
Ligand Overlay

- 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

Restraint Editor



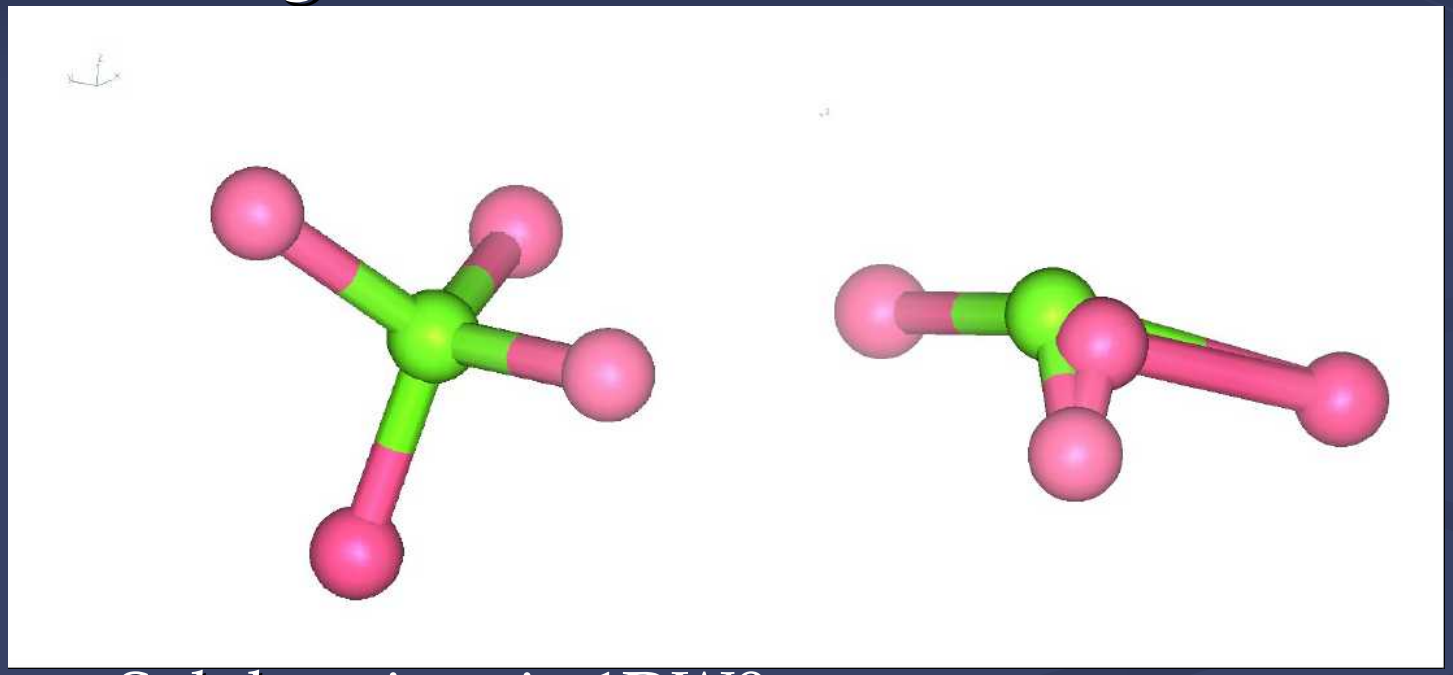
Displaying LINKs



Ligand validation

Ligand validation
Why validate?

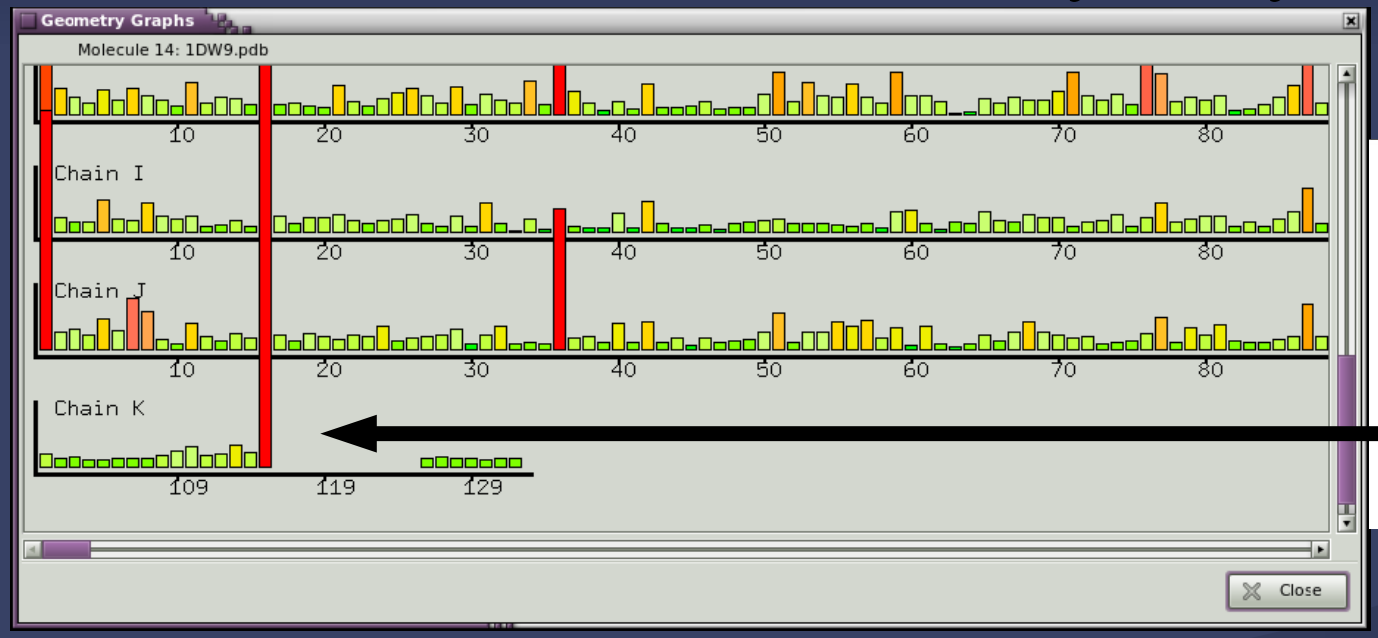
- 70.000 structures in PDB
- 11.000 'ligands' in 50.000 structures

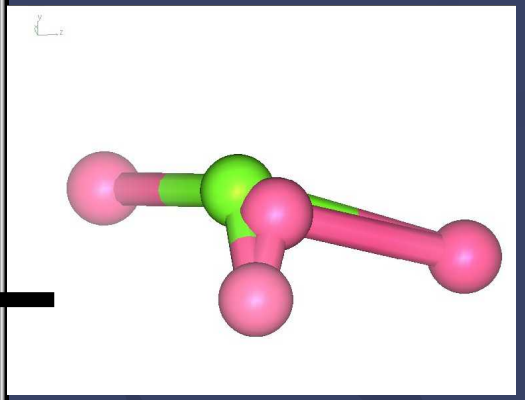


  - Sulphate ions in 1DW9
  - 1.65 A resolution
  - R/Rfree 0.15/0.19

Ligand validation

- How to validate ligand geometry?
  - Compare observed structure to restraints
    Coot: Validate->Geometry analysis





  - But what if parametrization (dictionary) is wrong?

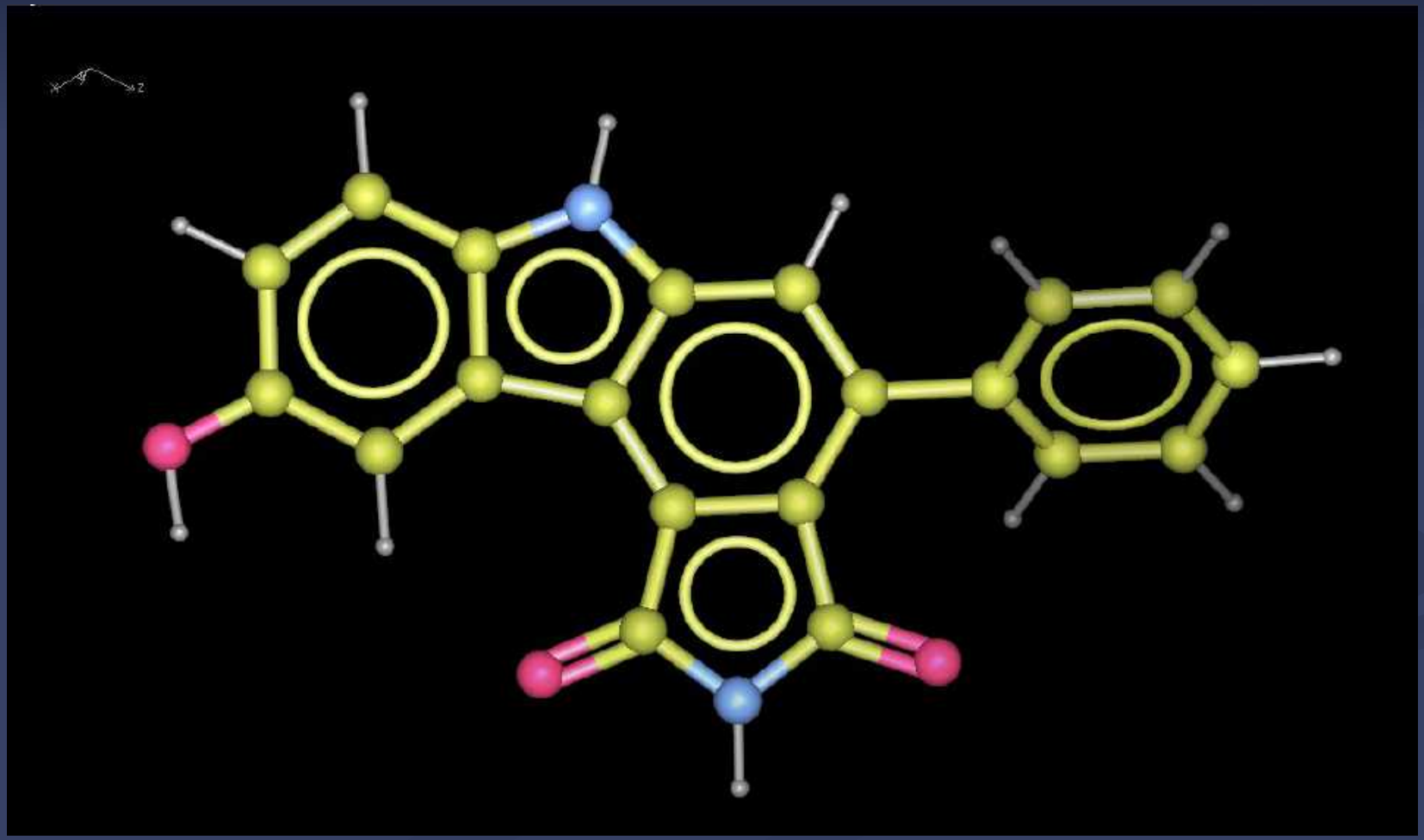
How to validate ligand geometry?

- QM
  - CPU hungry
  - In vaccuo
  - Low energy ≠ (?) bound ligand
- PDB (e.g. ValLigURL)
  - Good cofactor structures
  - Less useful for novel ligands
  - (occ.) questionable quality
- Cambridge Structural Database (CSD)
  - Using small molecule geometries
  - e.g. Mogul

Ligand representation

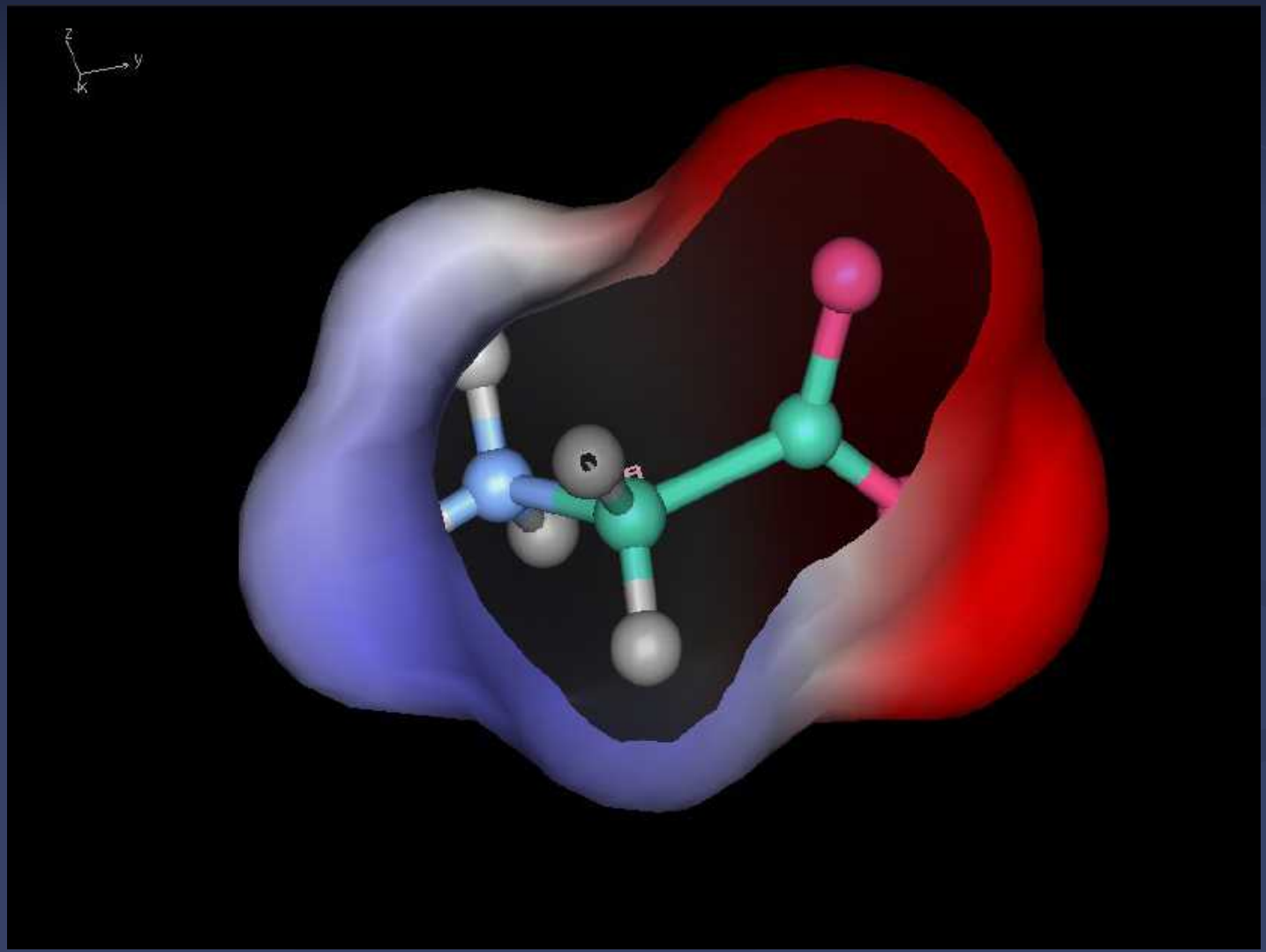
Ligand representation

- Bond order representation from dictionary files



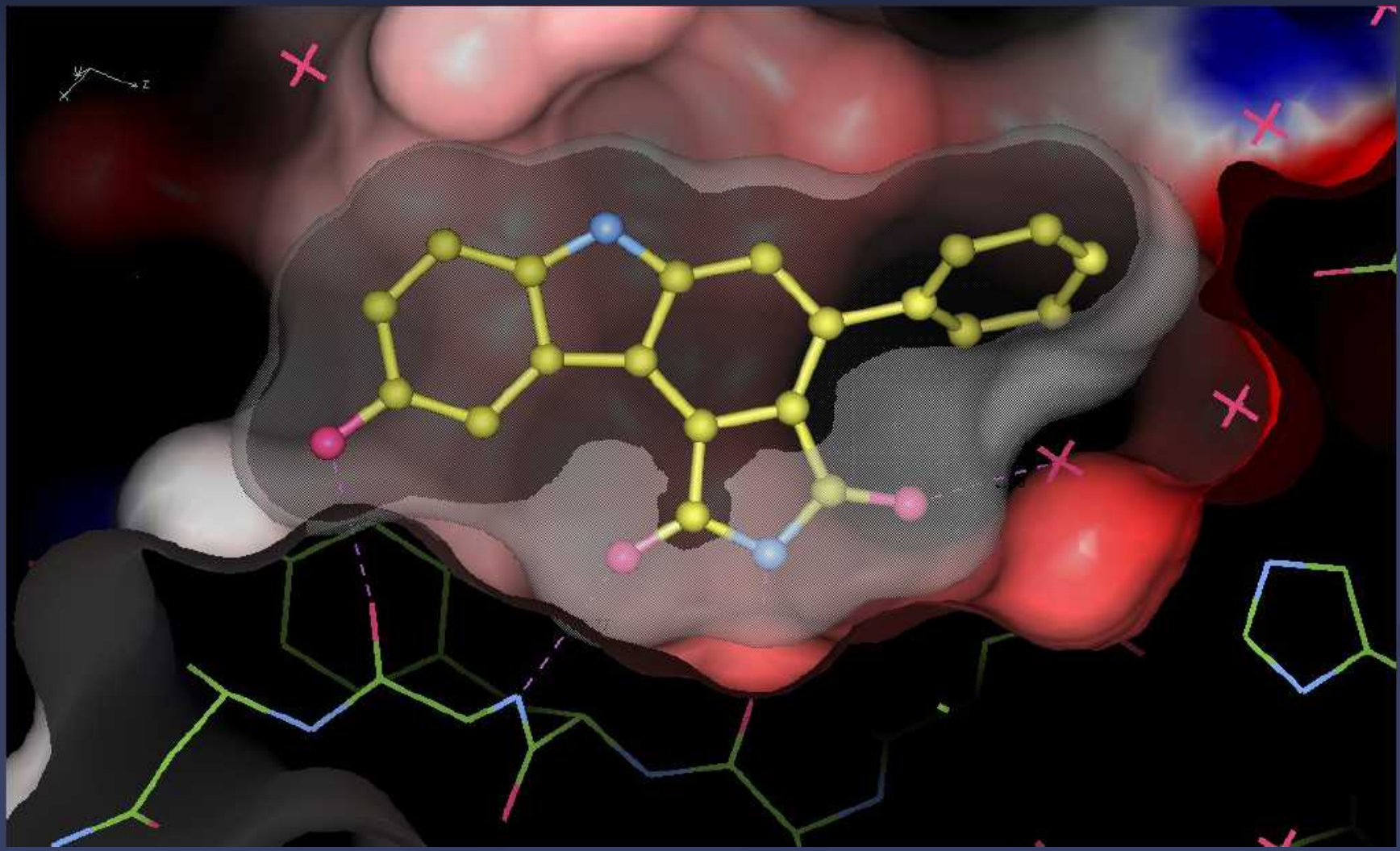
Surfaces

- Partial charges



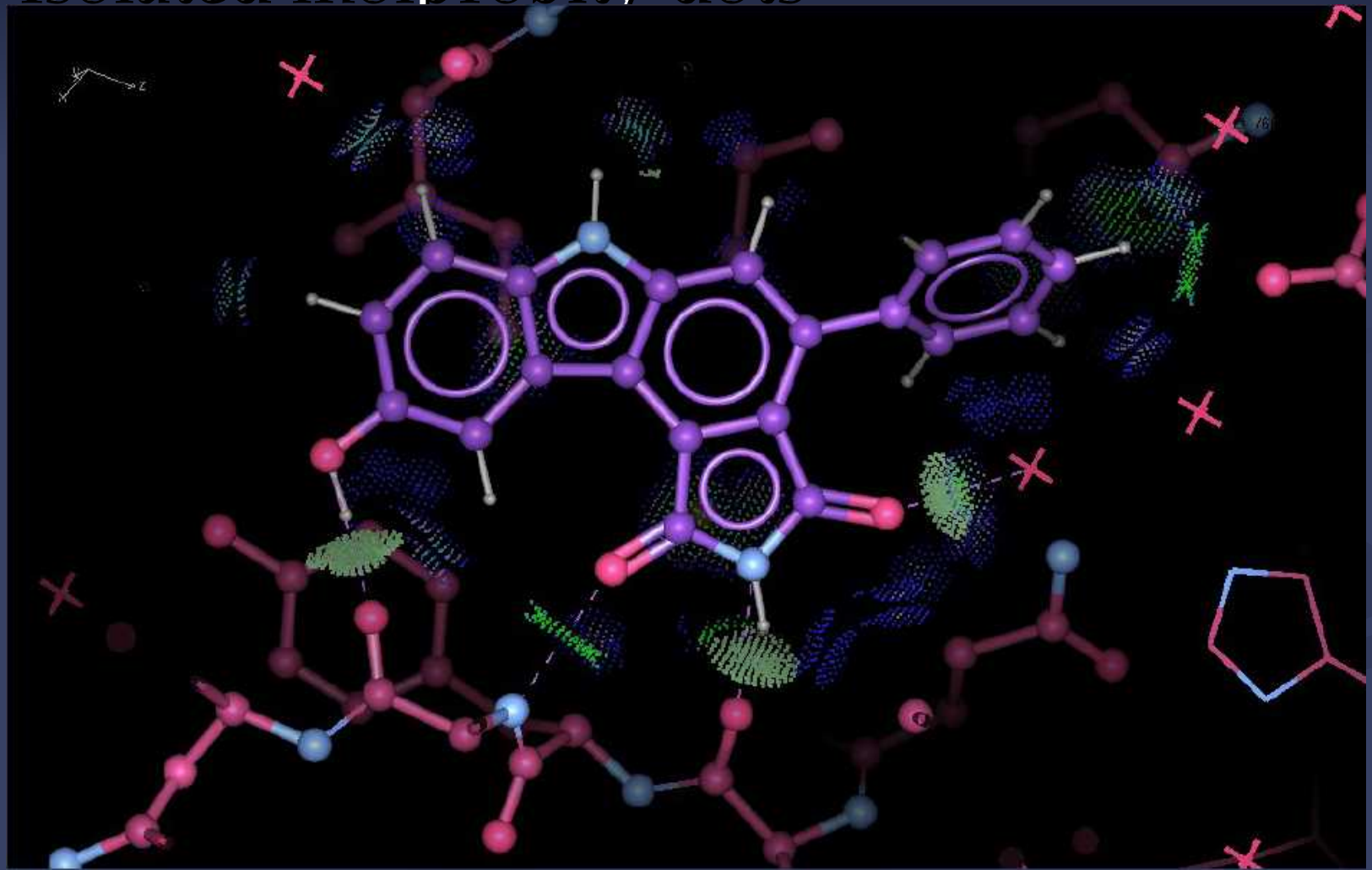
Surfaces

- Transparent surfaces – surface complementarity



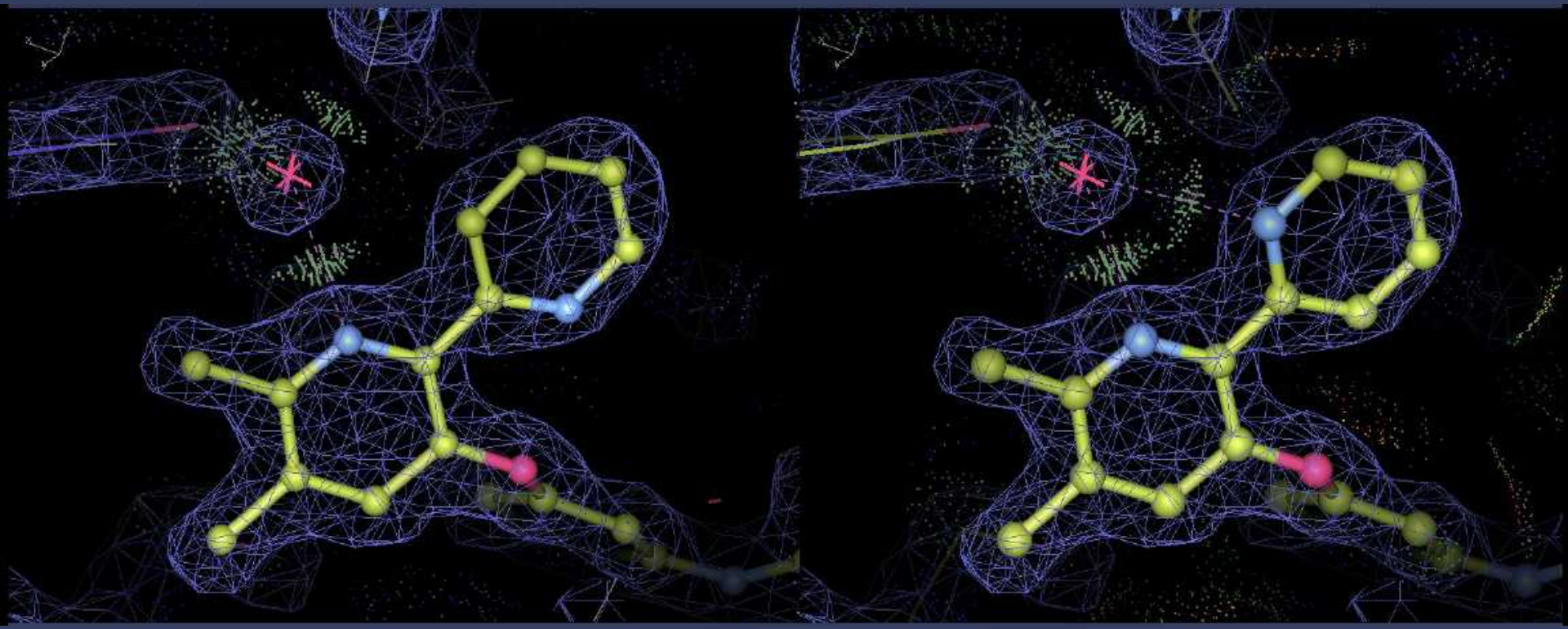
Binding mode analysis

- Binding site highlighting
- Isolated molprobity dots



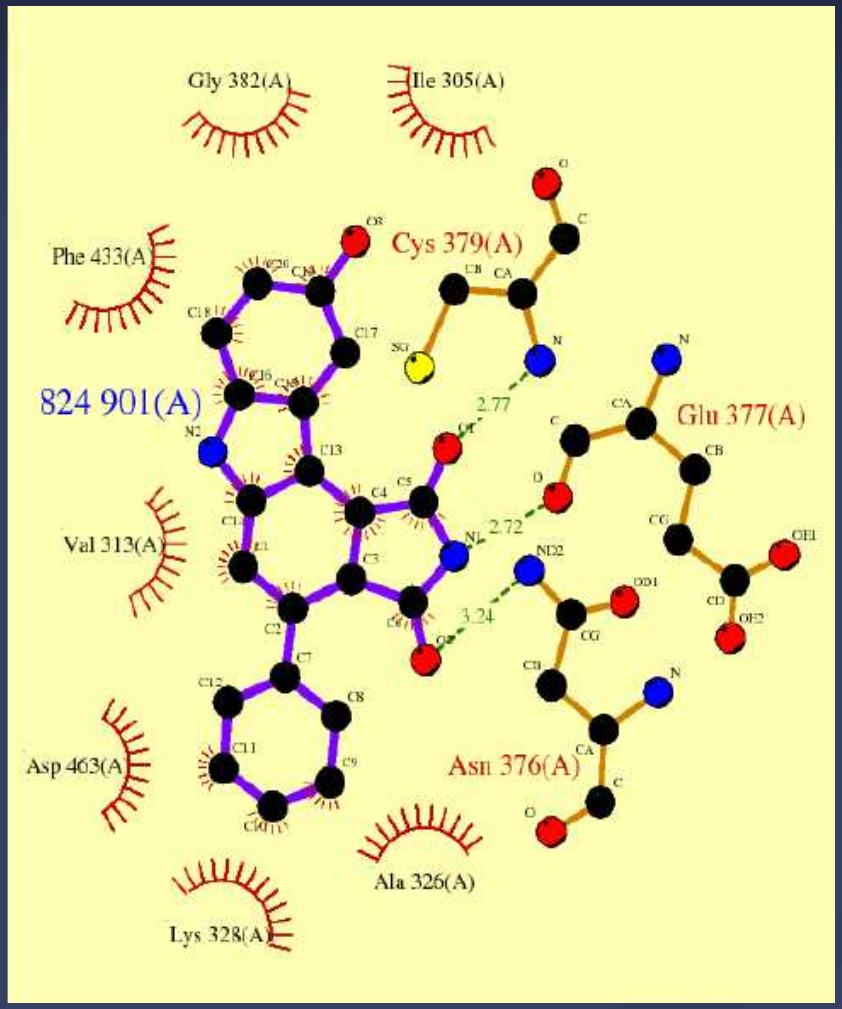
Binding mode analysis

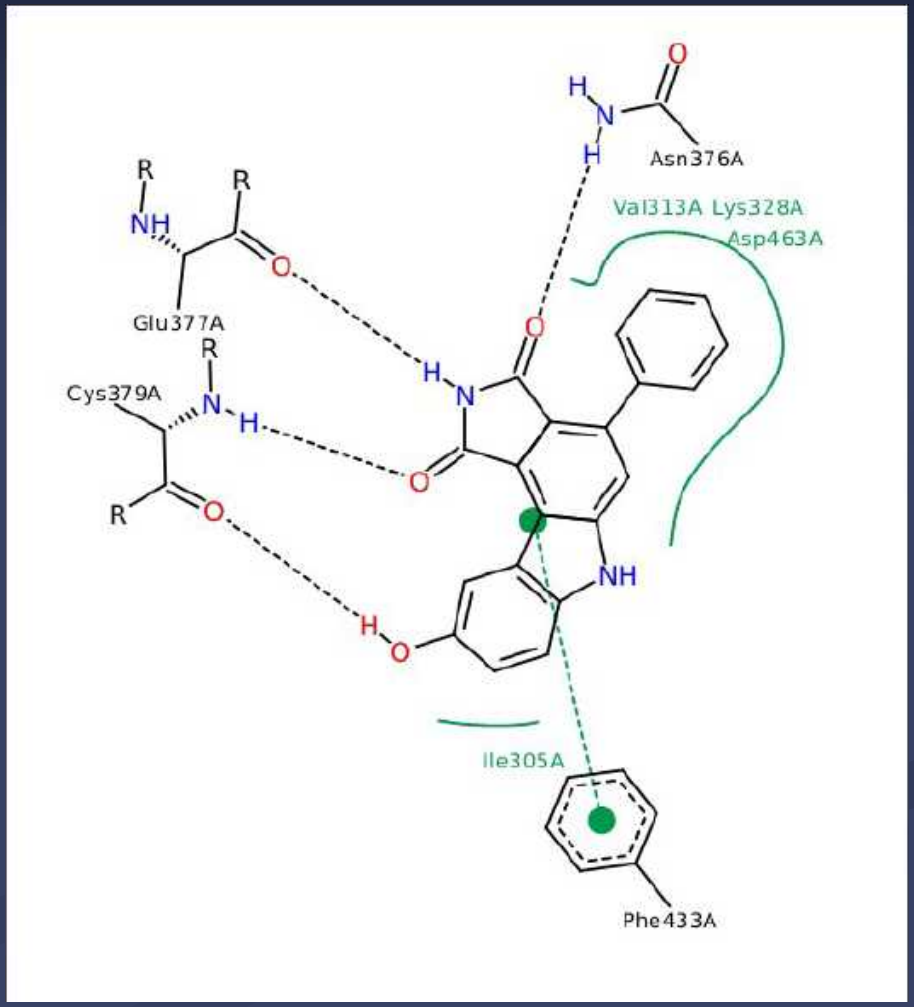
- Binding site highlighting
- Isolated molprobity dots



Ligand environment

- 2D ligand pocket layout (ligplot, poseview)





Can we do better?

Ligand environment layout

- Binding pocket residues
- Interactions
- Substitution contour
- Solvent accessible halos

Work in progress!!!!!

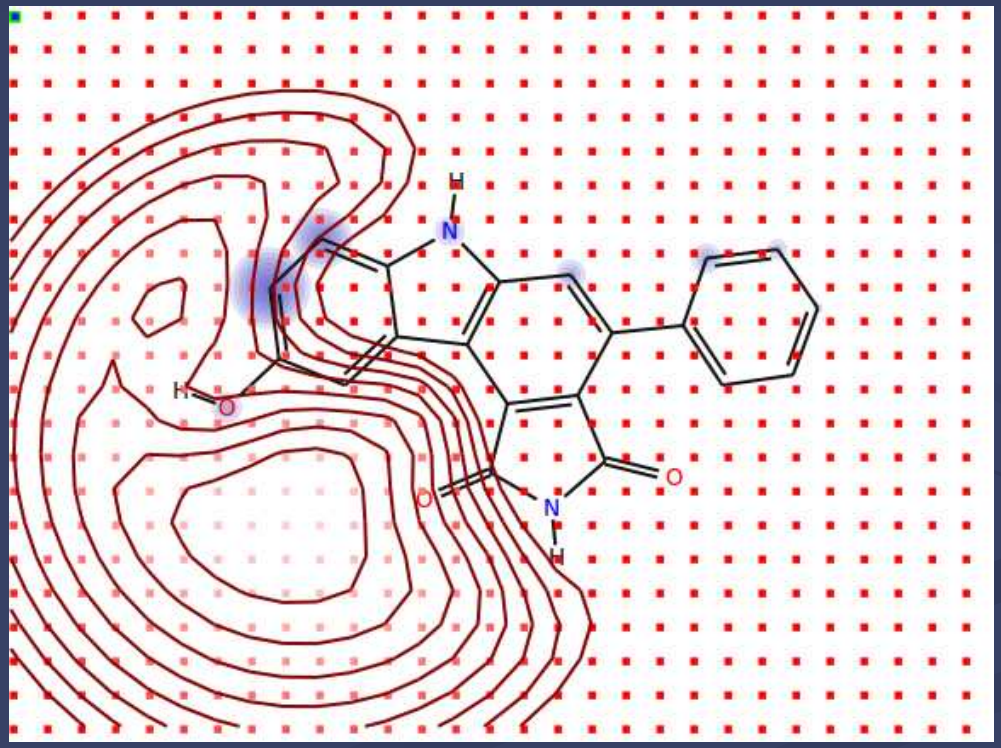
Ligand environment layout

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

Work in progress

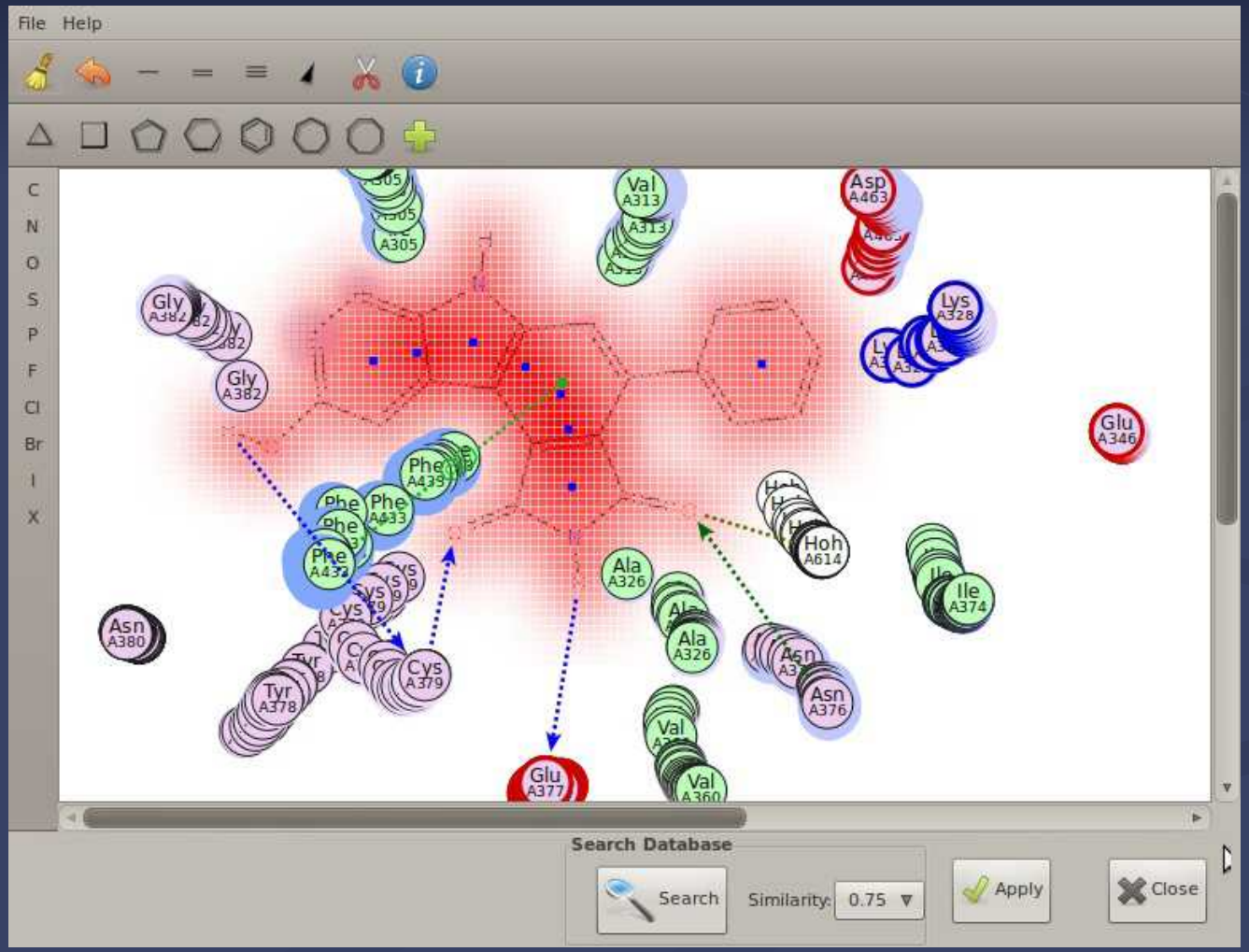
Ligand environment layout

- Initial residue placement



Ligand environment layout

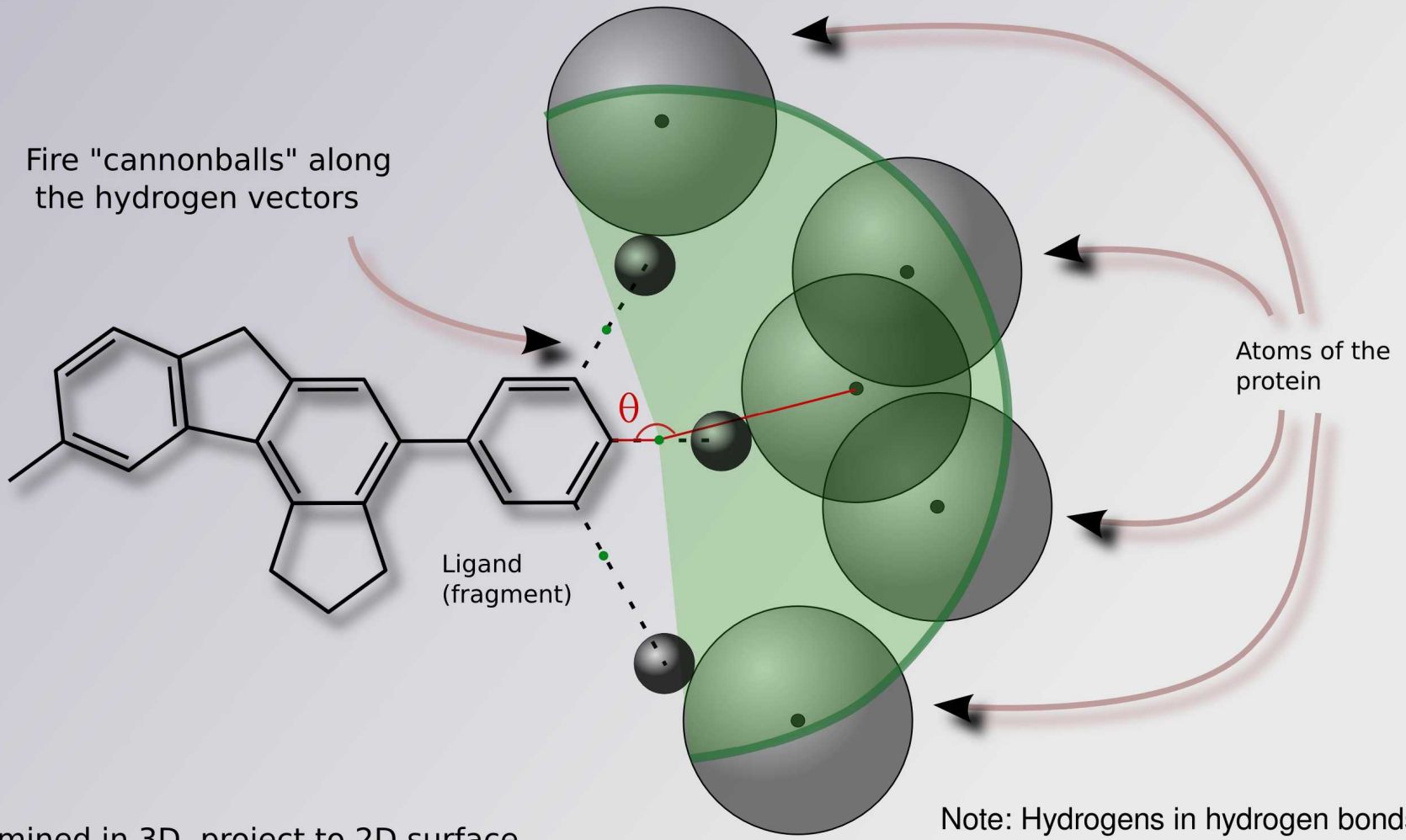
- Residue position minimization



Determination of the Substition Contour

c.f. Clarke & Labute (2007)

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

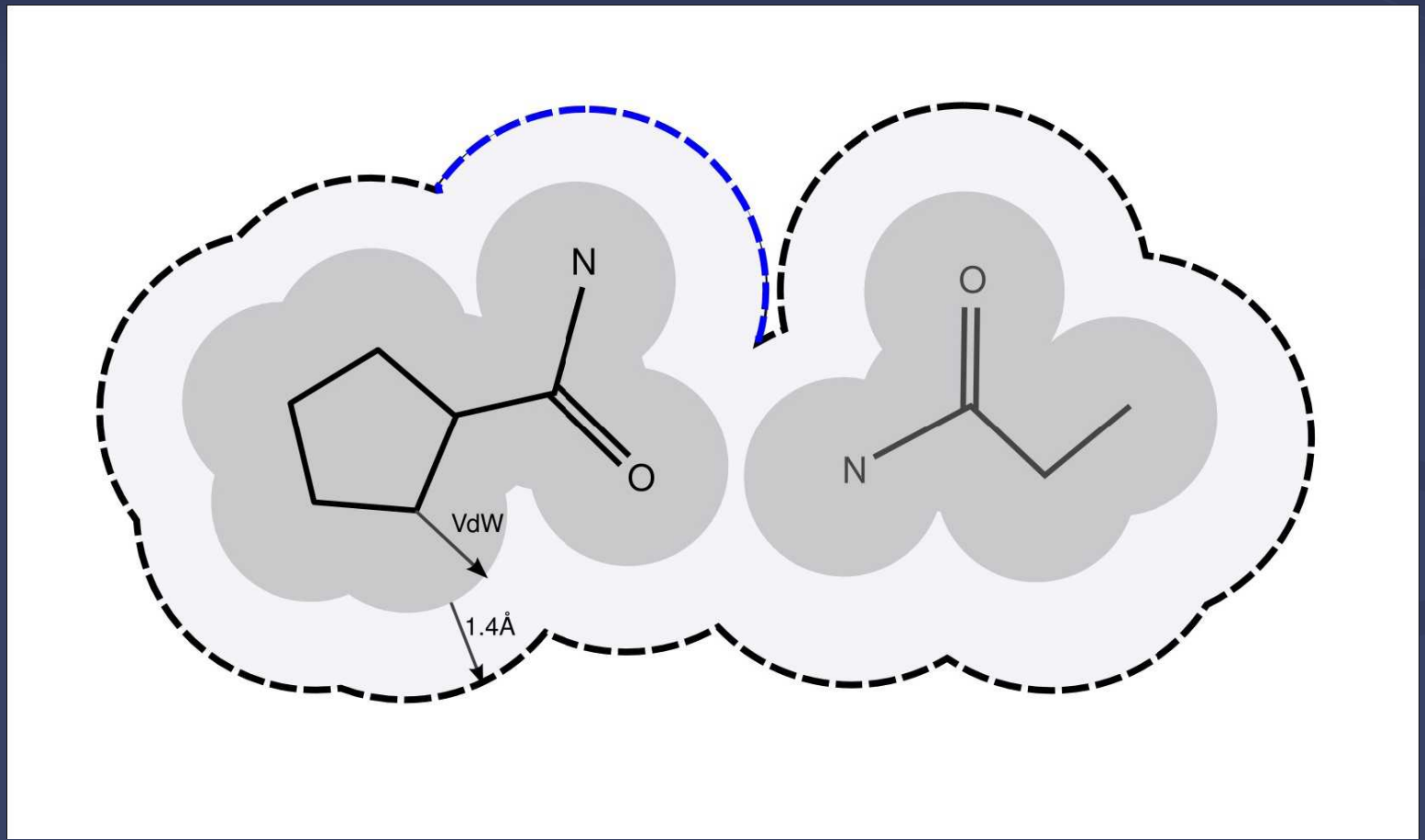


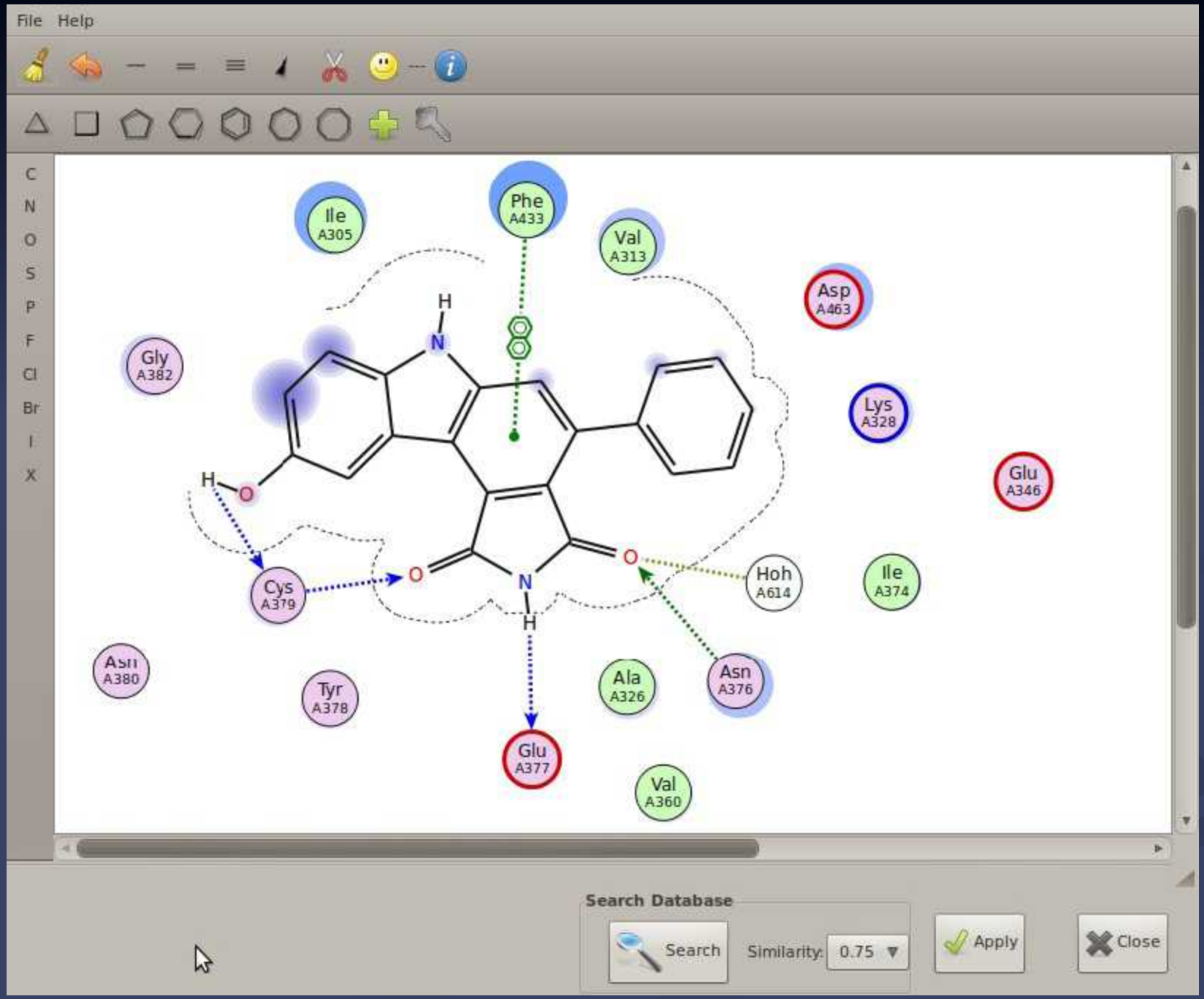
Determined in 3D, project to 2D surface

Note: Hydrogens in hydrogen bonds are a confounding factor

Solvent exposure calculation

- Identification of solvent accessible atoms
- Different from substitution contour



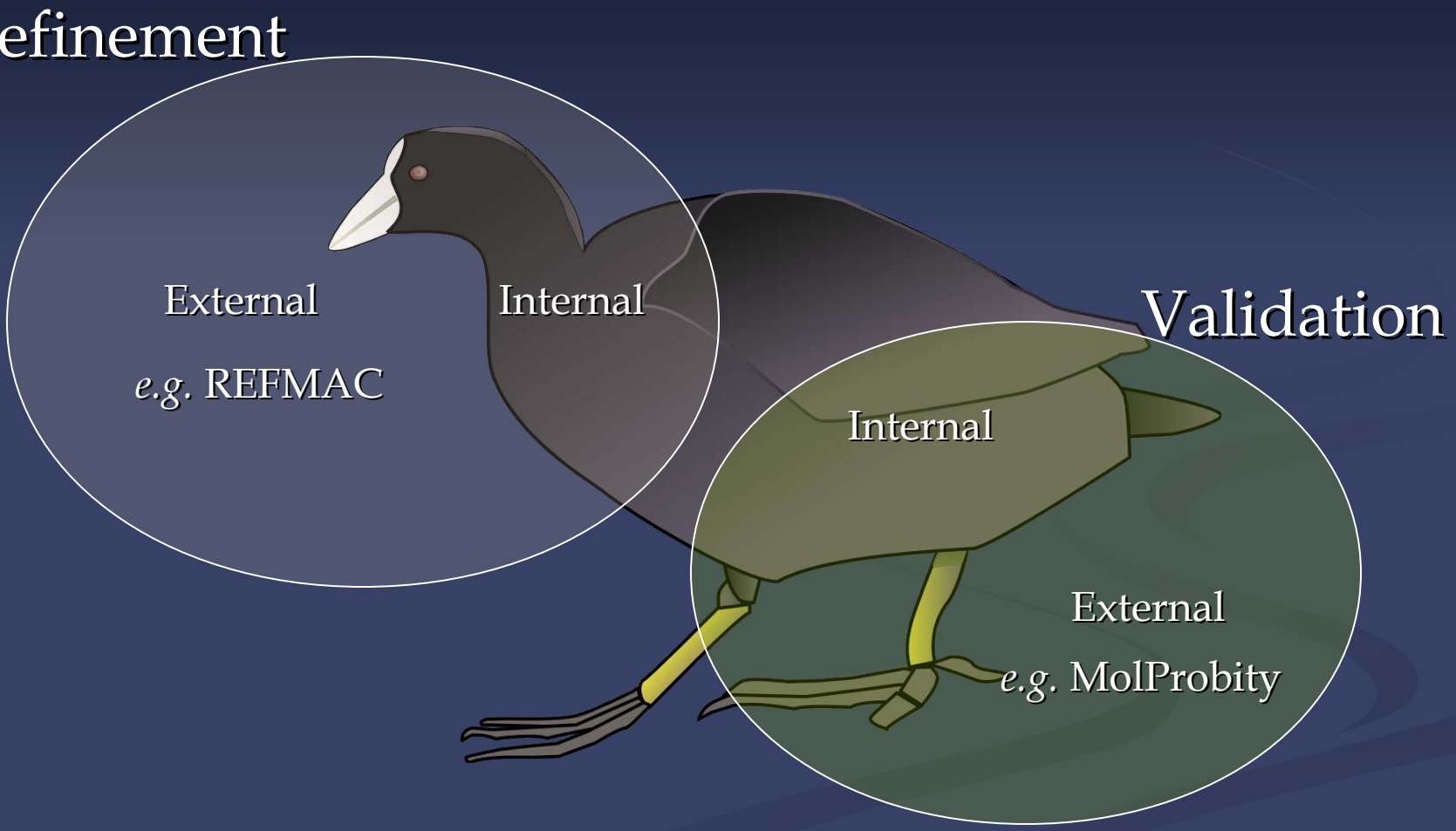

File Help
C
N
O
S
P
F
Cl
Br
I
X
Ile A305
Phe A433
Val A313
Asp A463
Gly A382
Lys A328
Glu A346
Cys A379
Hoh A614
Ile A374
Asn A380
Tyr A378
Ala A326
Asn A376
Glu A377
Val A360
Search Database
Search
Similarity: 0.75
Apply
Close

Ligand futures

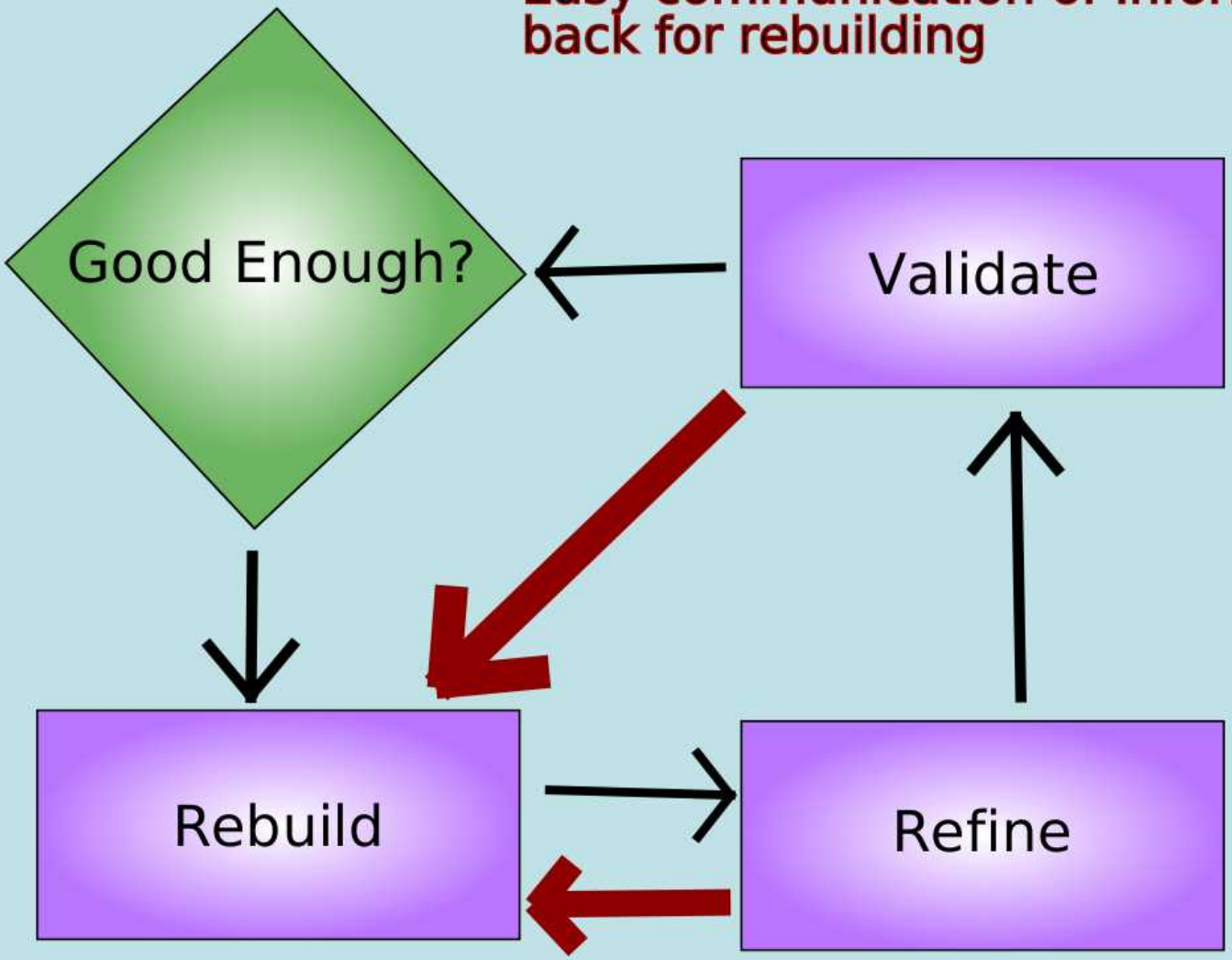
- Make ligand builder (and environment layout) available
- Interface to JLigand
- Ligand validation with CSD (Mogul or similar)

Validation

Feature Integration



Easy communication of Information back for rebuilding



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
  - Global and local indicators

Why Validate?

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

A “good” model

- Makes statistical sense
  - The reciprocal space representation agrees tolerably well with the observations (R-factor)
  - No meaningful difference map peaks (no under- or over-fitting)
- Makes chemical sense
  - Model geometry is consistent with the restraints
  - Ramachandran Plot has less than 1% outliers
  - A good **clashscore**
- Make physical sense
  - Crystal packing and contacts
- Makes biological sense
  - Residues in chemically sensible environment
  - Is consistent (on the whole) with external biochemistry observations (active site residues)

Quick Bayes

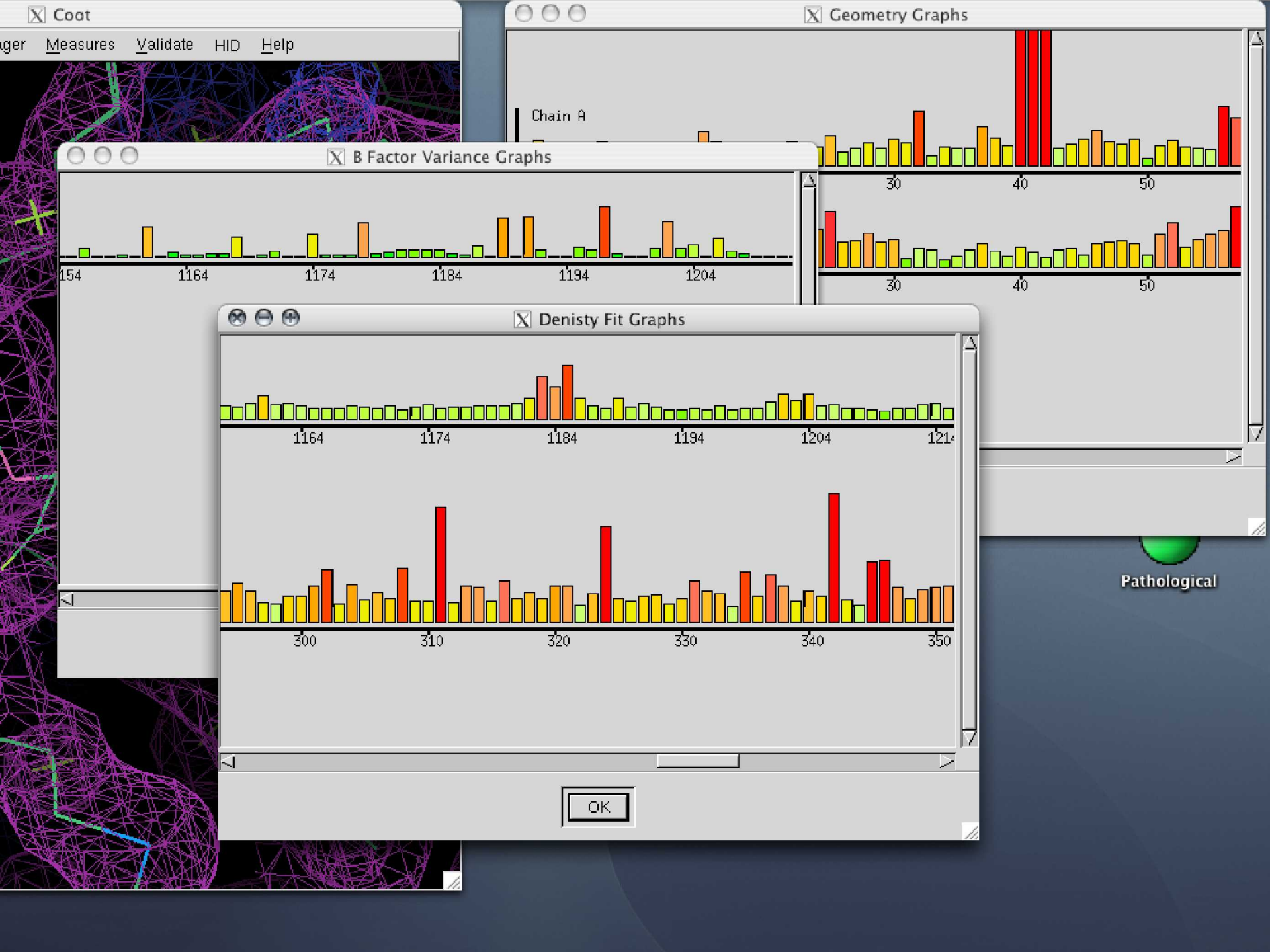
- Bayes Eq:
- $\Pr(\text{model} \mid \text{data}) \propto \Pr(\text{data} \mid \text{model}) * \Pr(\text{model})$
- $\Pr(\text{data} \mid \text{model})$ is also called the model likelihood, $L_M(\text{model} \mid \text{data})$

Validation Tools - Pr(model)

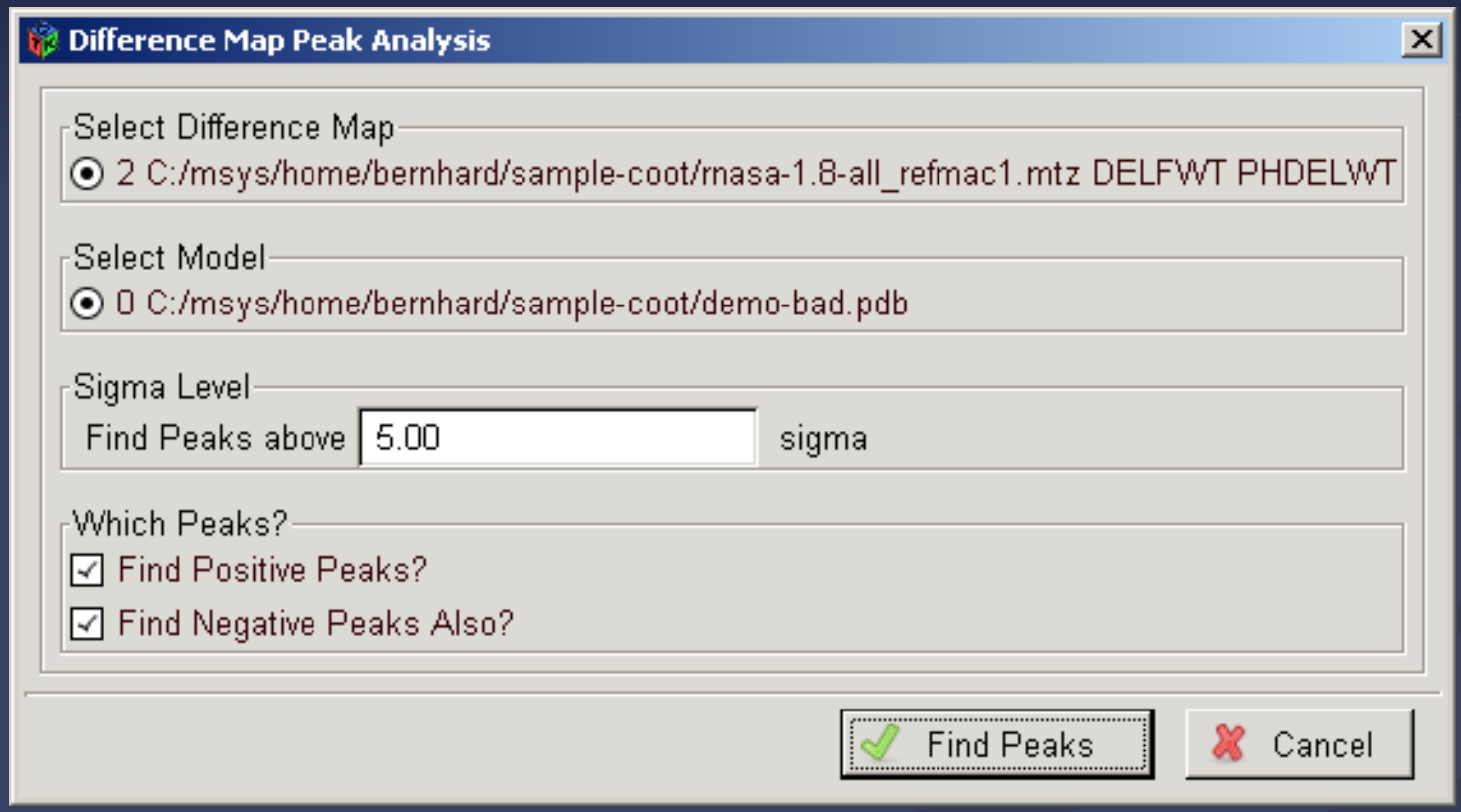
- Ramachandran Plot
  - Kleywegt Plot (NCS differences)
- Geometry Analysis
- Peptide ω Analysis
- Temperature Factor Analysis
- Rotamer Analysis
- Clashes

Validation Tools - Pr(data|model)

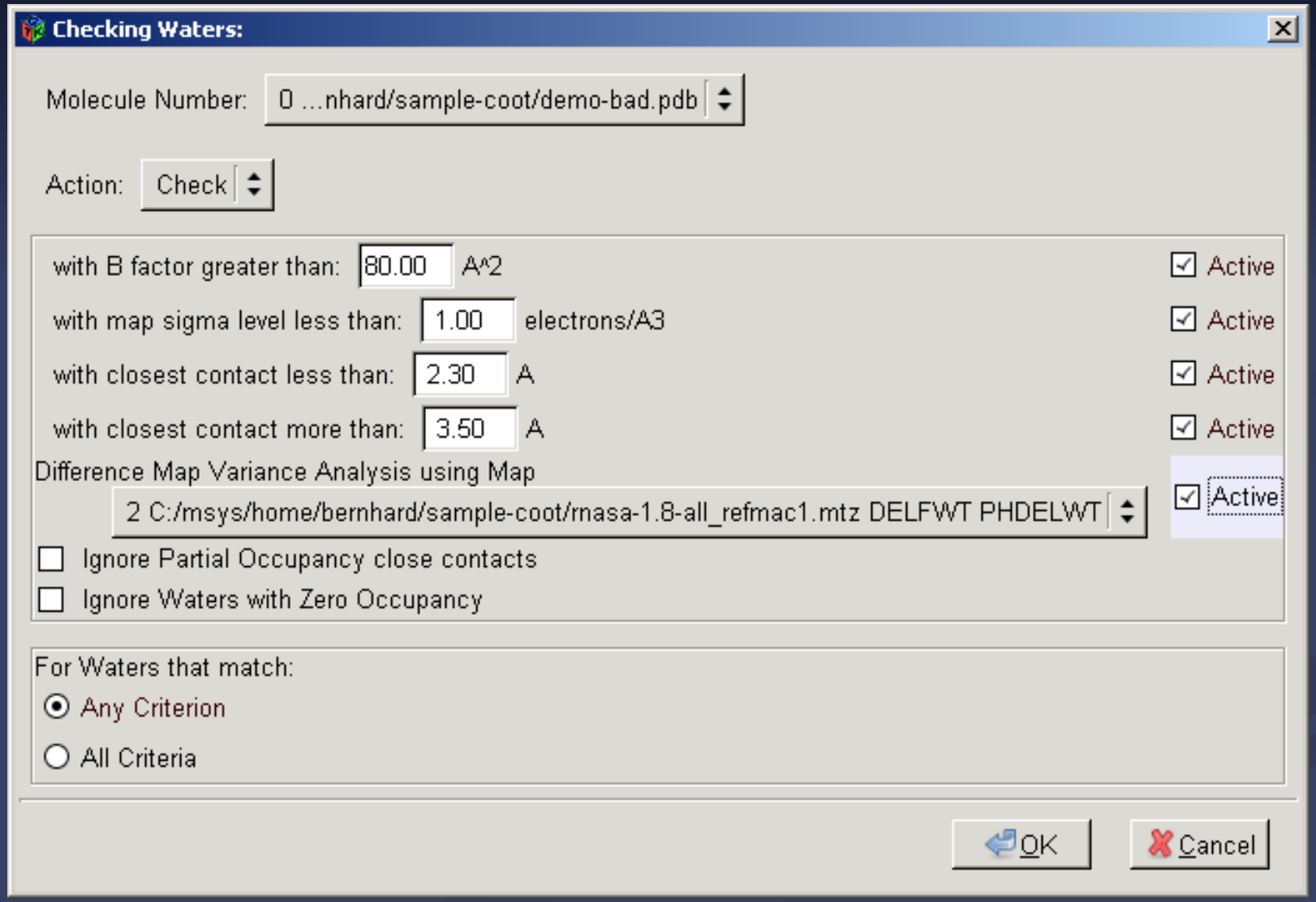
- Density Fit Analysis
- Difference Map Peaks
  - Variance analysis at water positions
- Unmodelled blobs


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

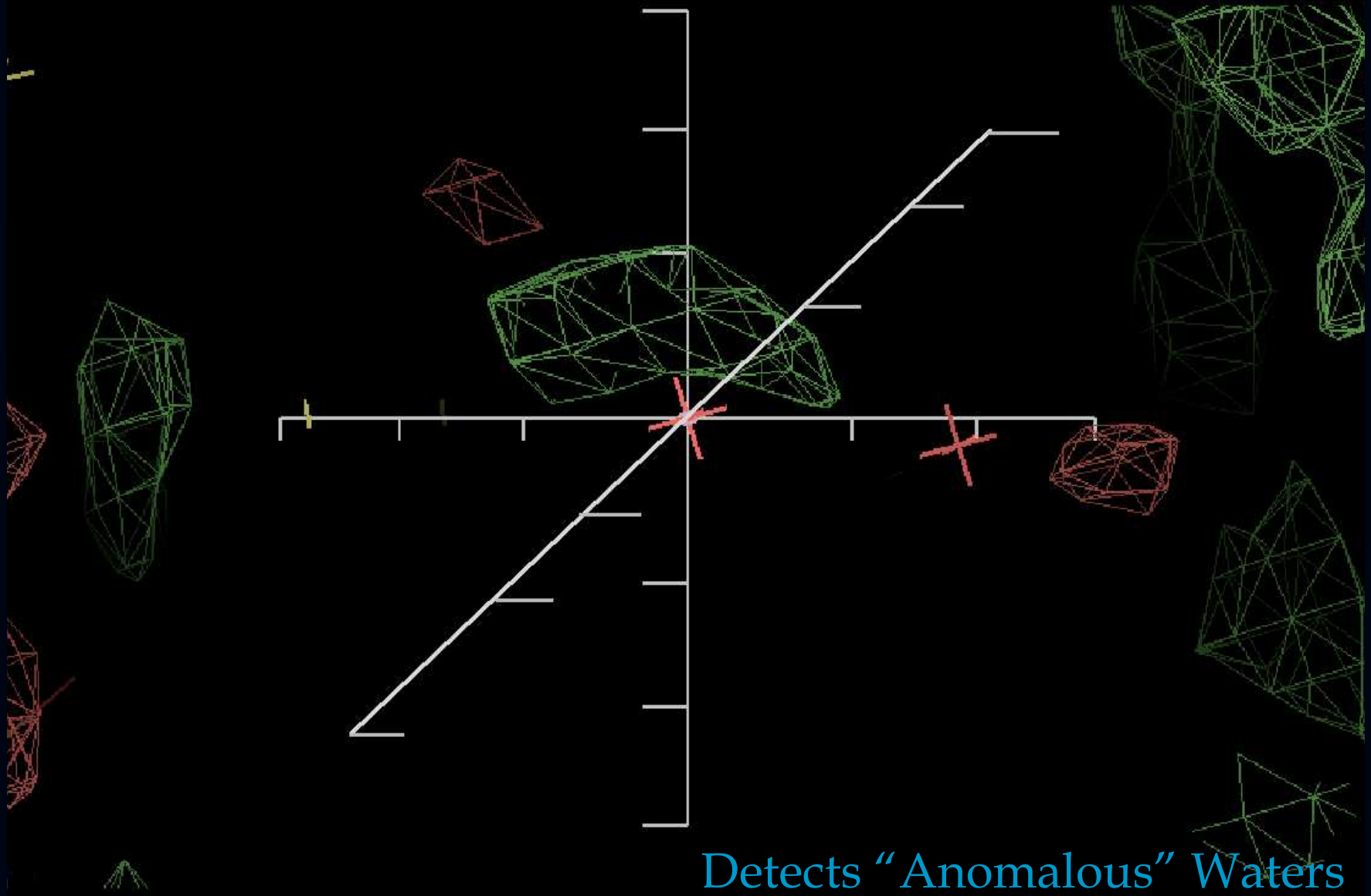
Find difference map peaks



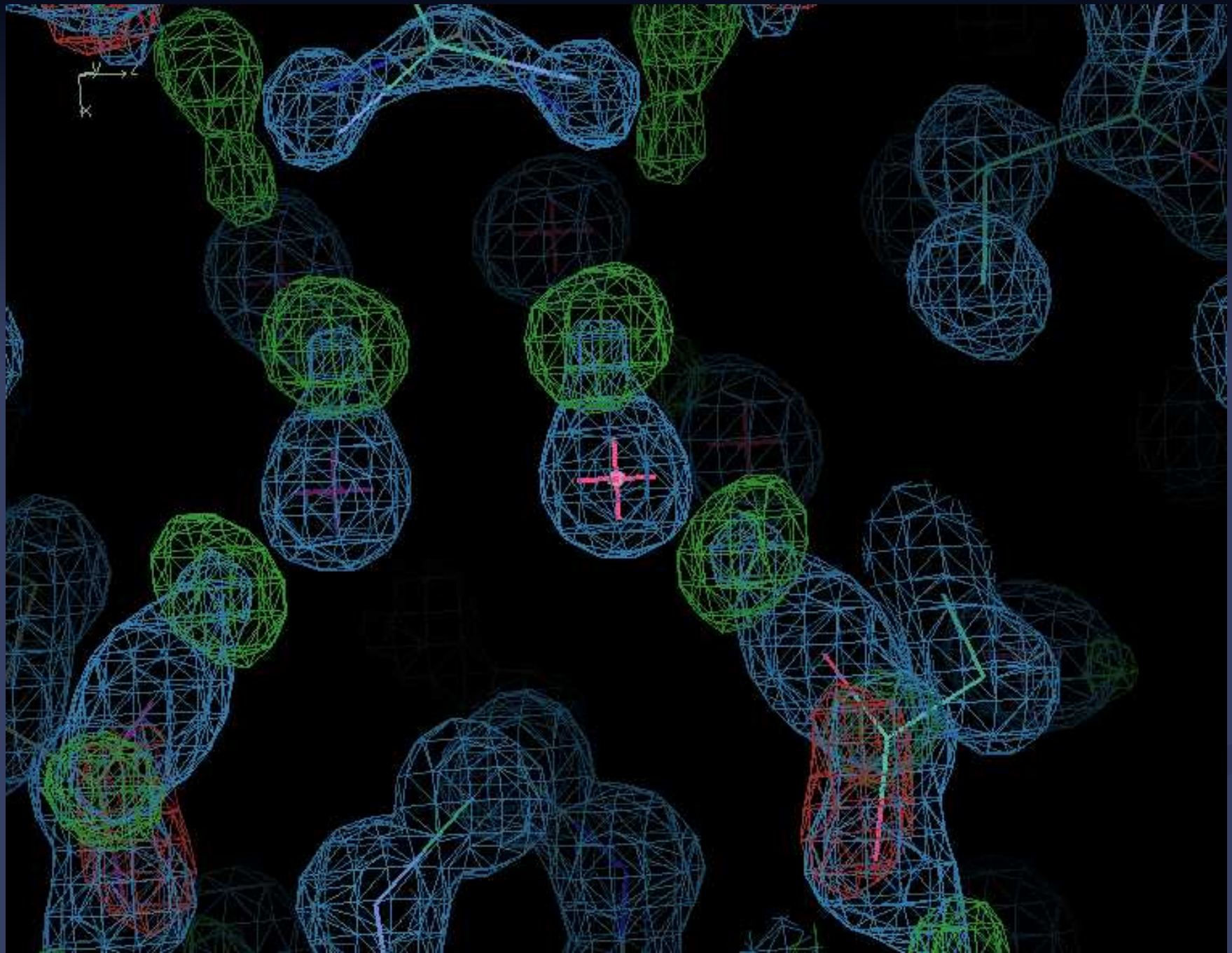
Check/Delete Waters

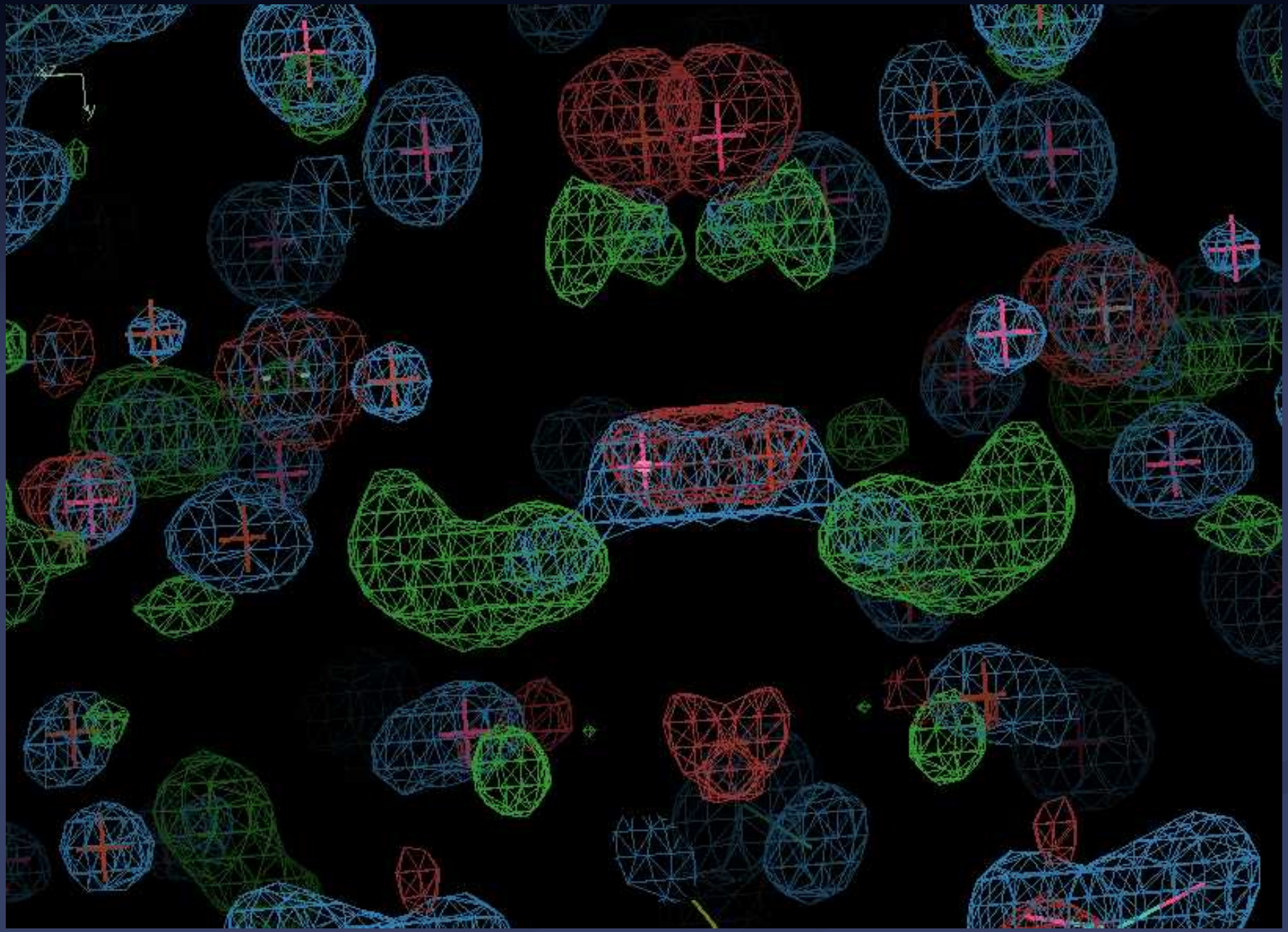


Difference Map Sampling

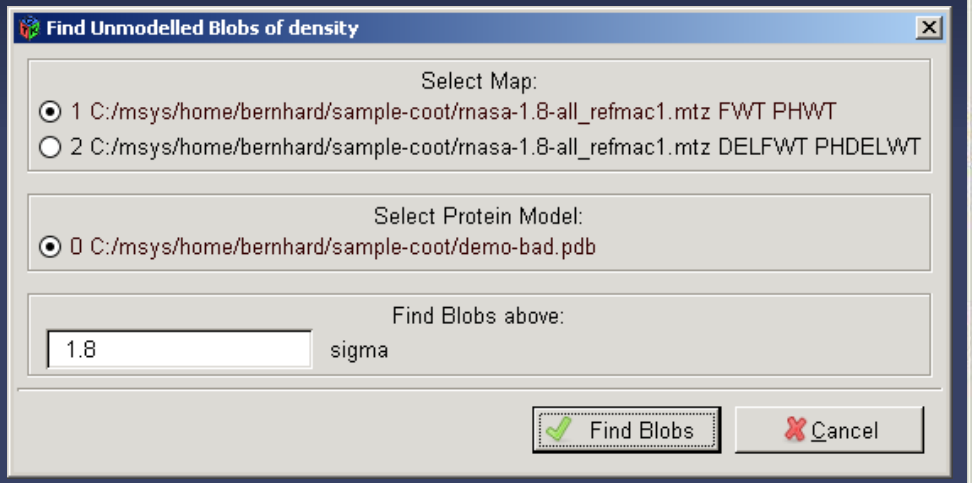


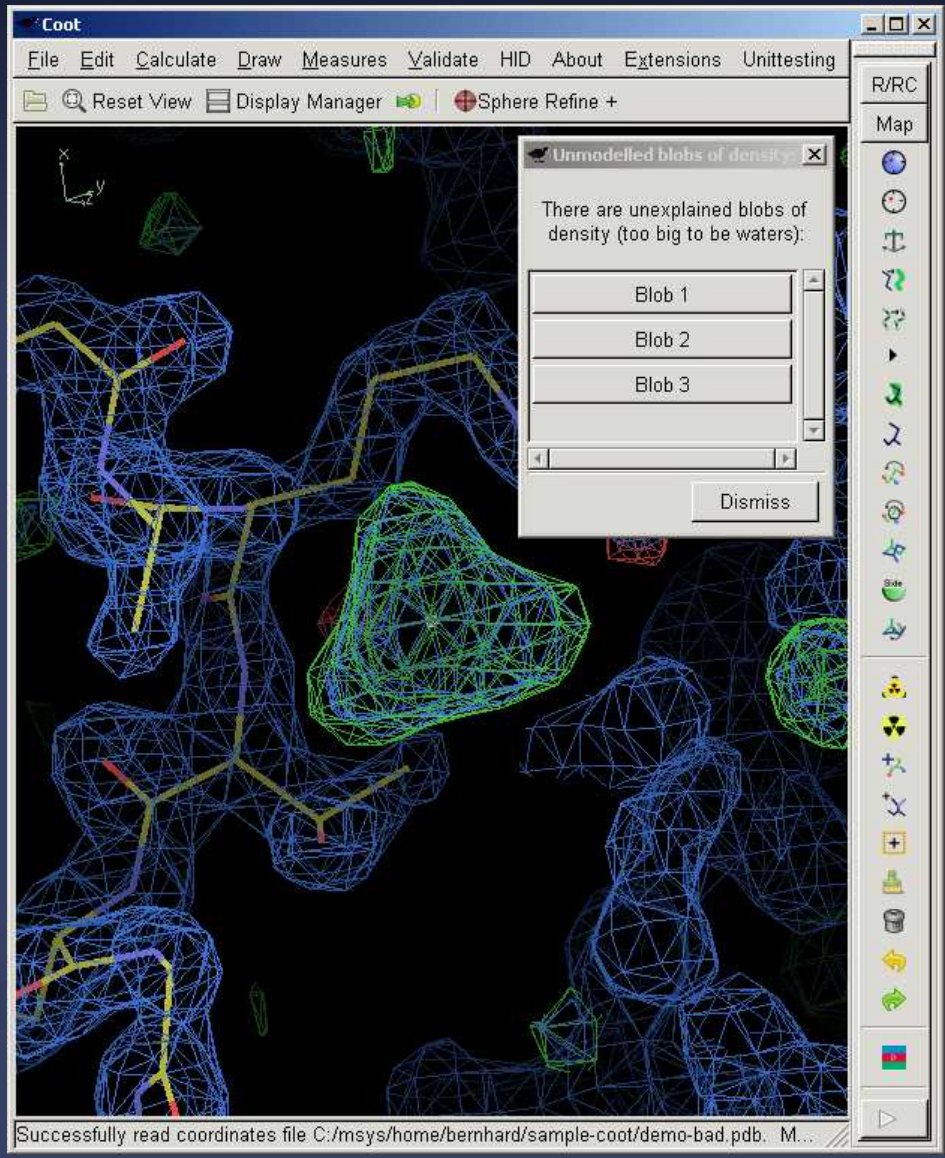
Detects “Anomalous” Waters





Find unmodelled blobs

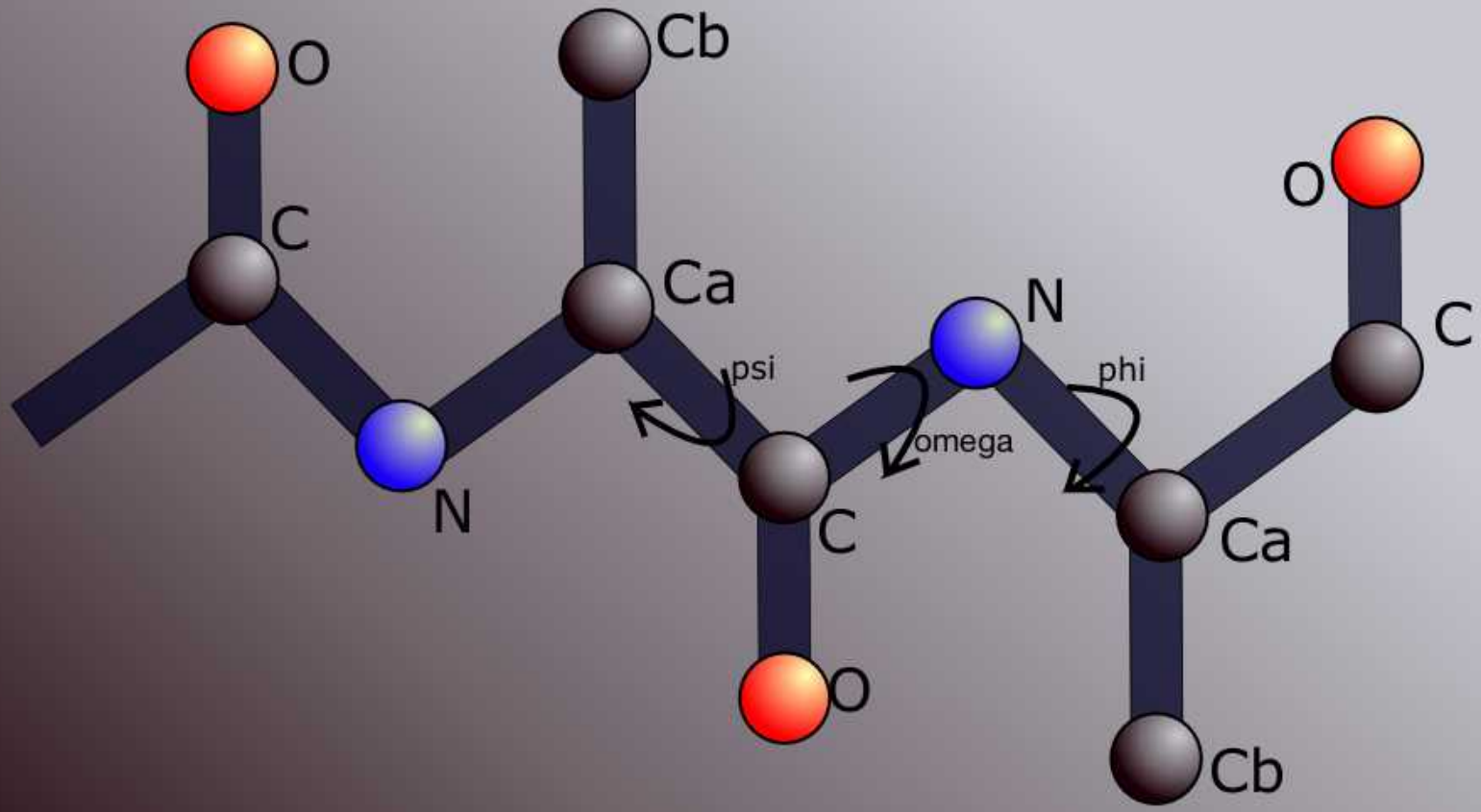




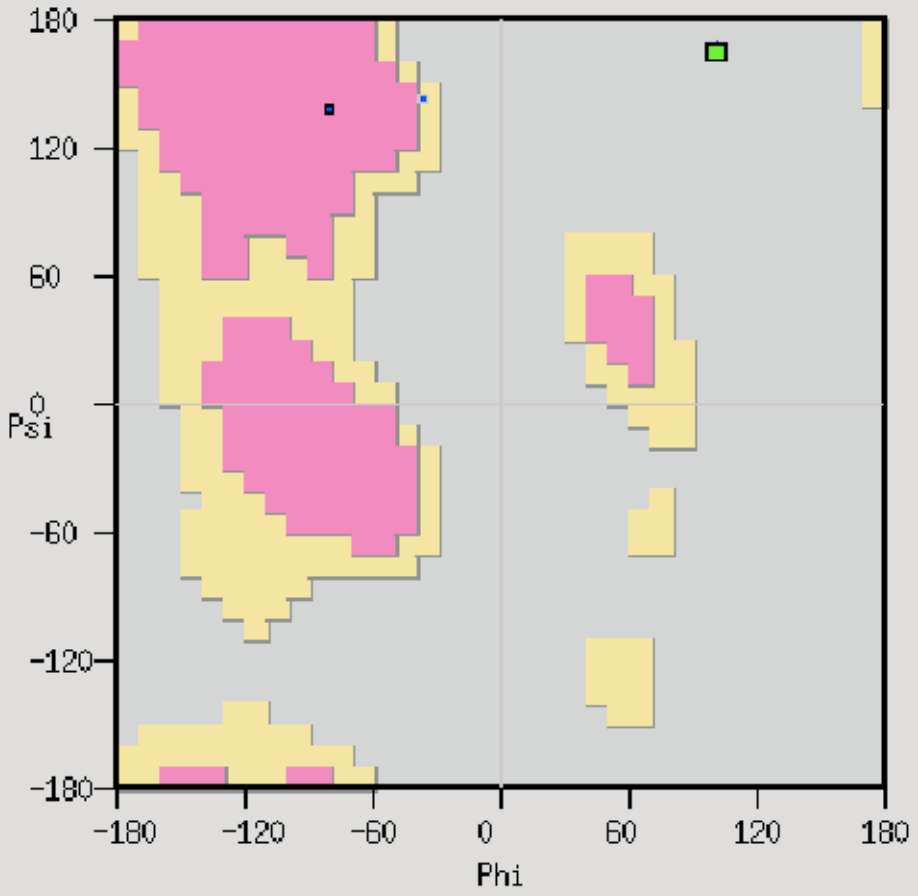
Pr(model)...

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

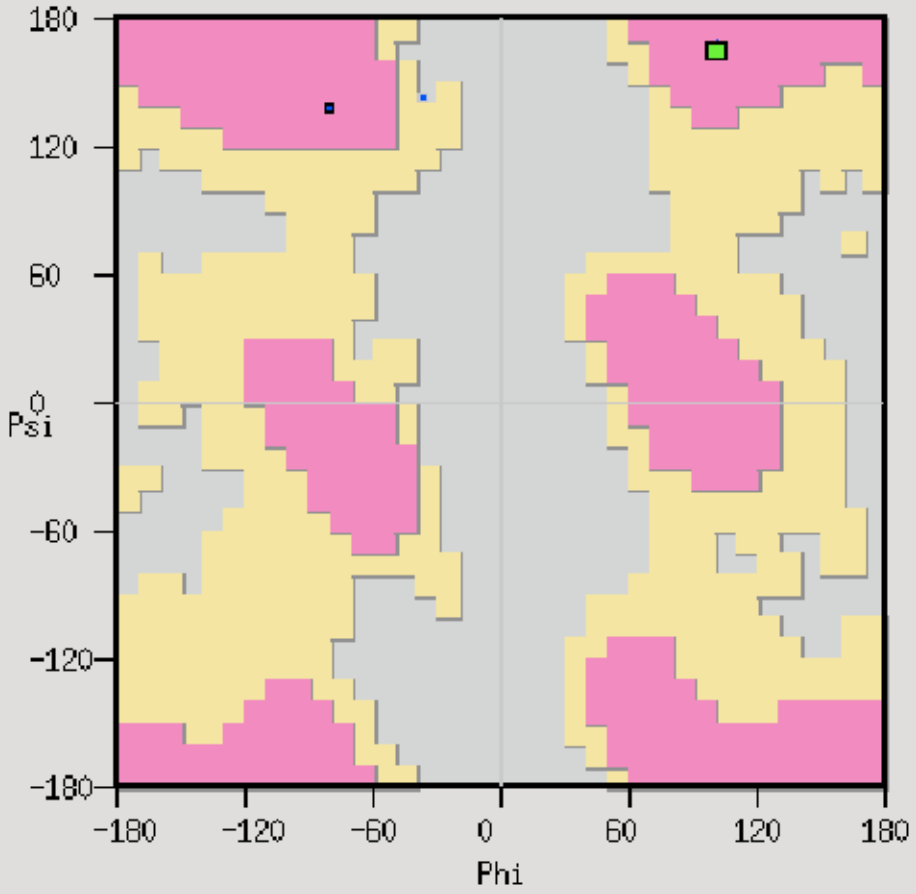
Peptide Torsion Angles



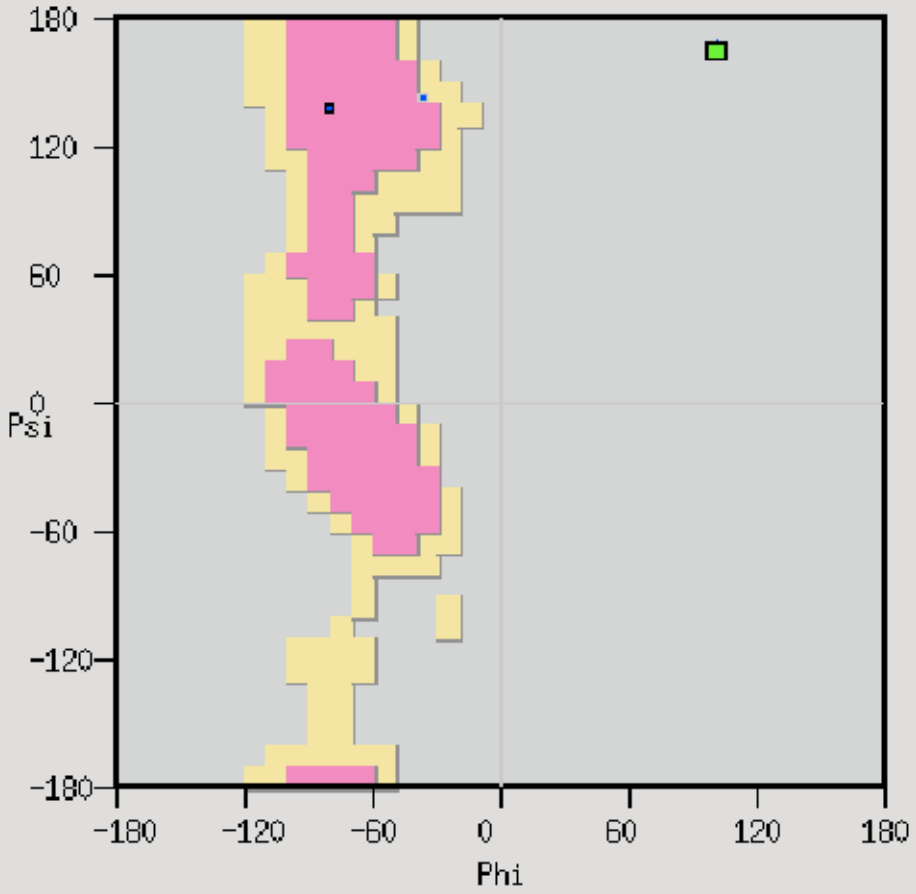
Ramachandran Plot for residues with CB

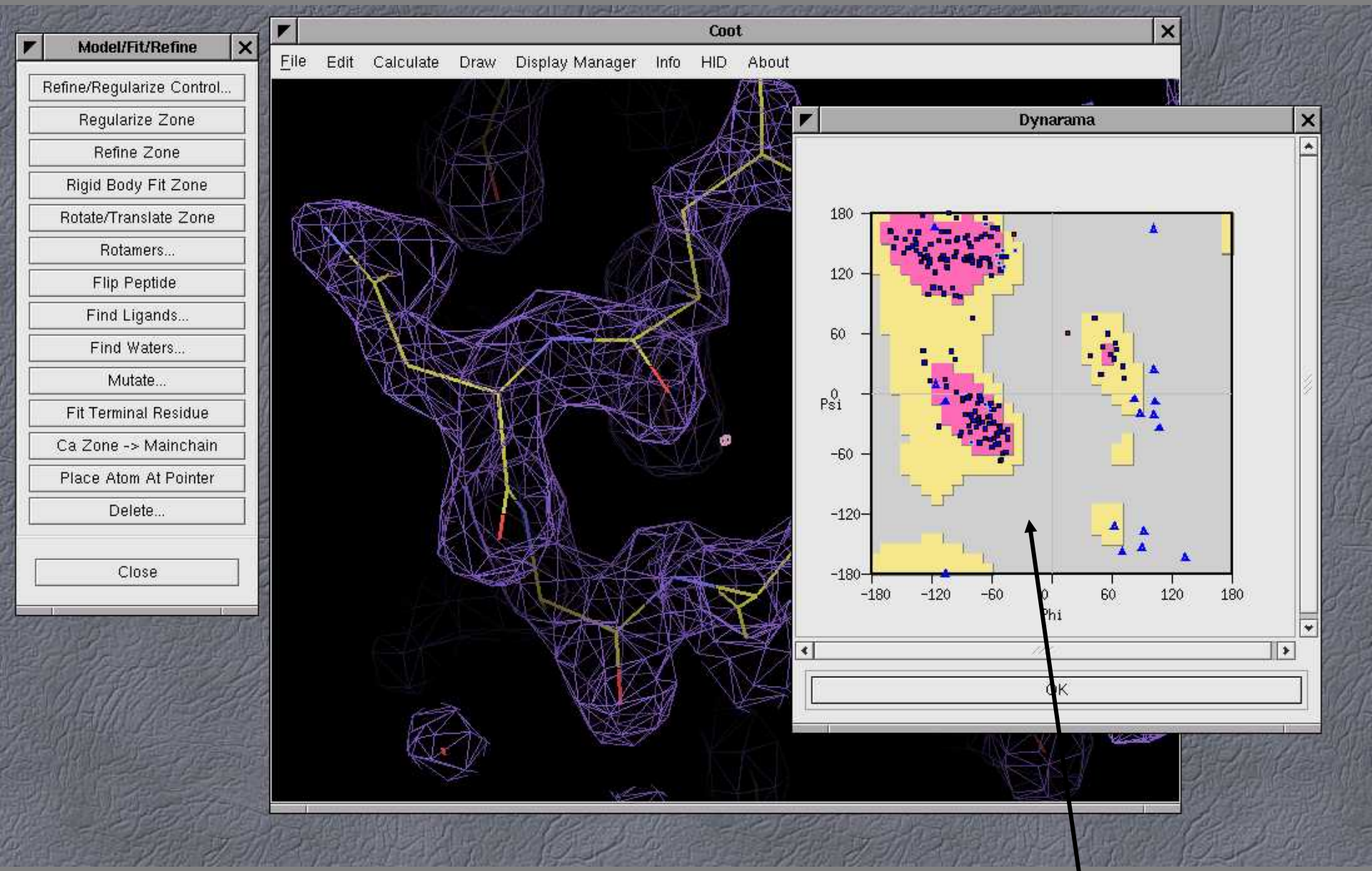


Ramachandran Plot for GLY



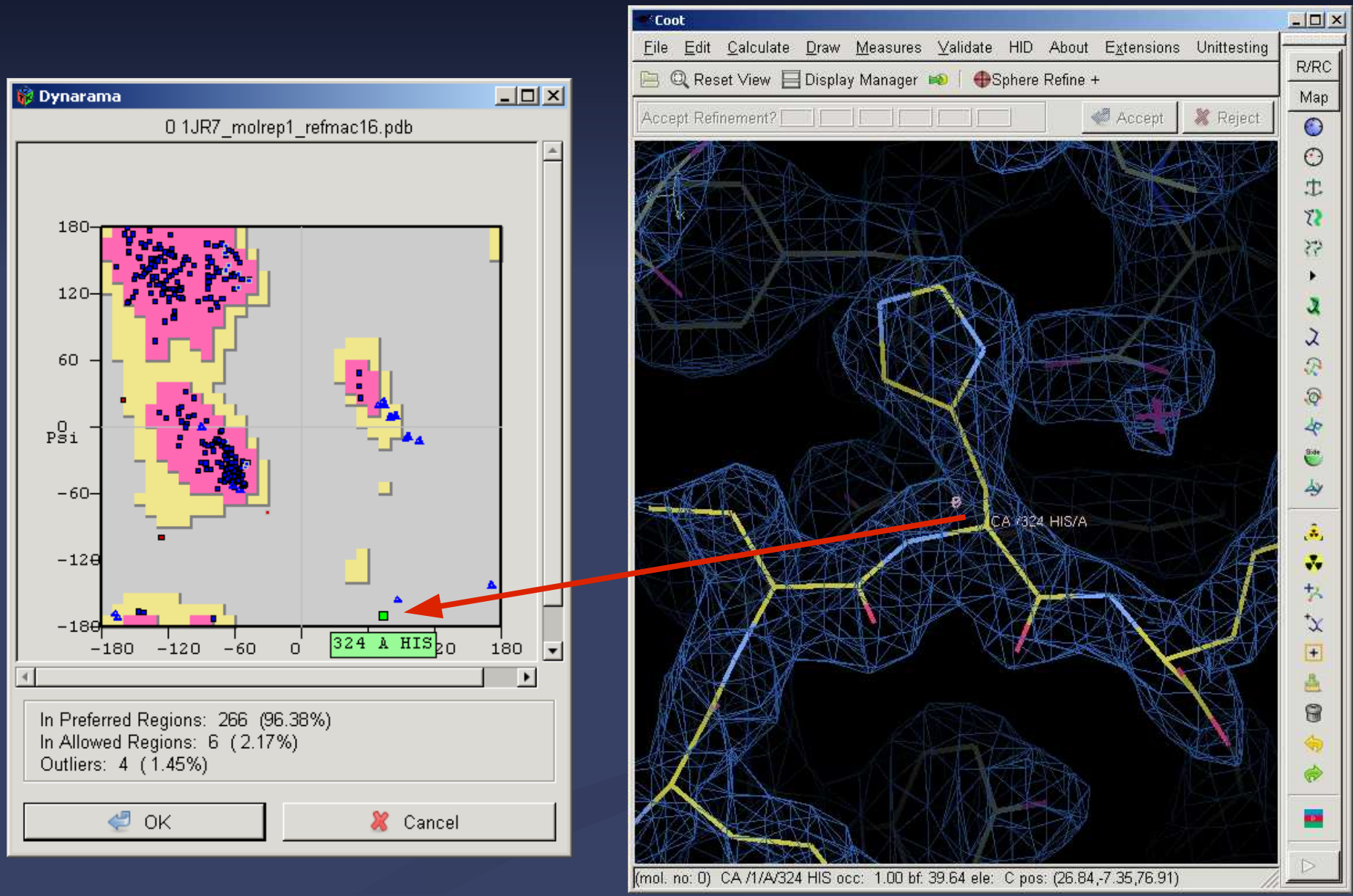
Ramachandran Plot for PRO



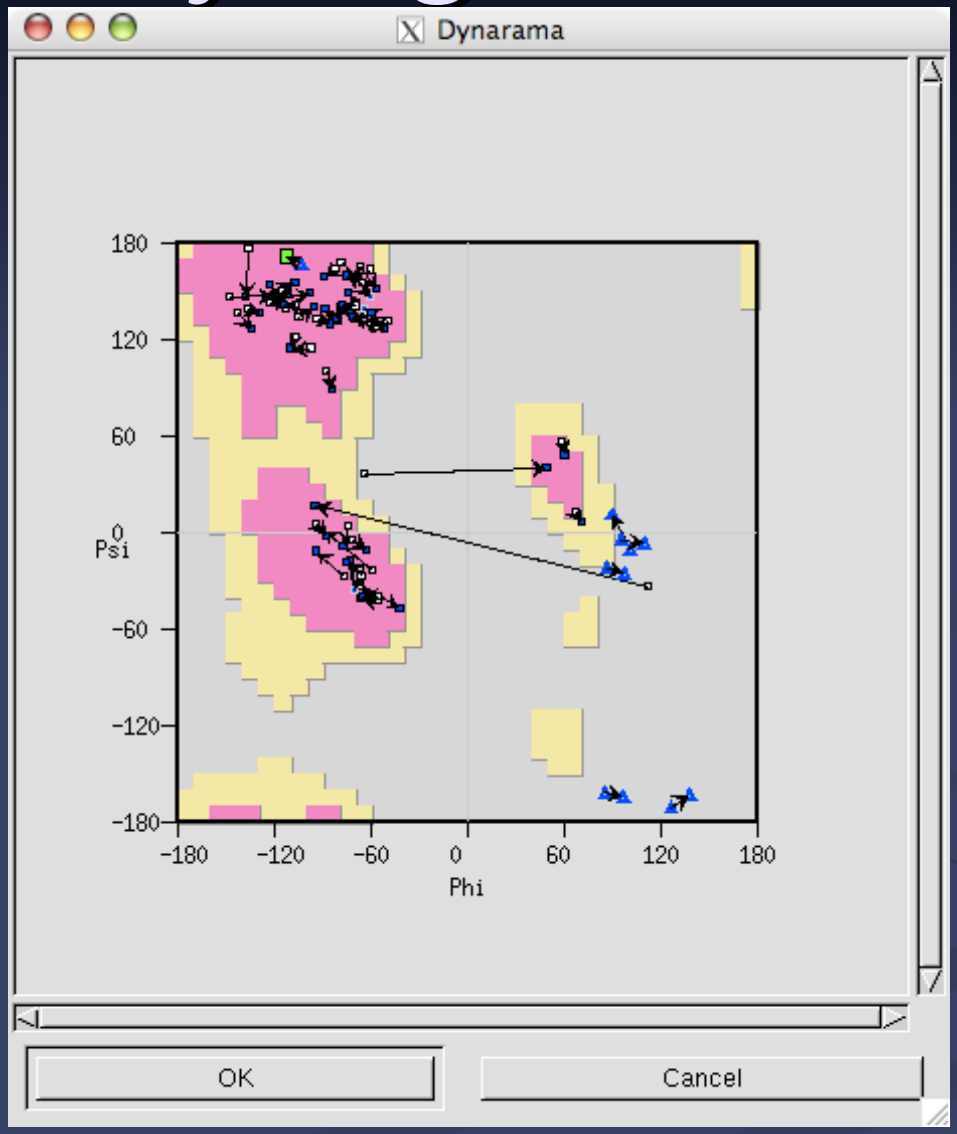


Top500-based distribution

Not all outliers are bad

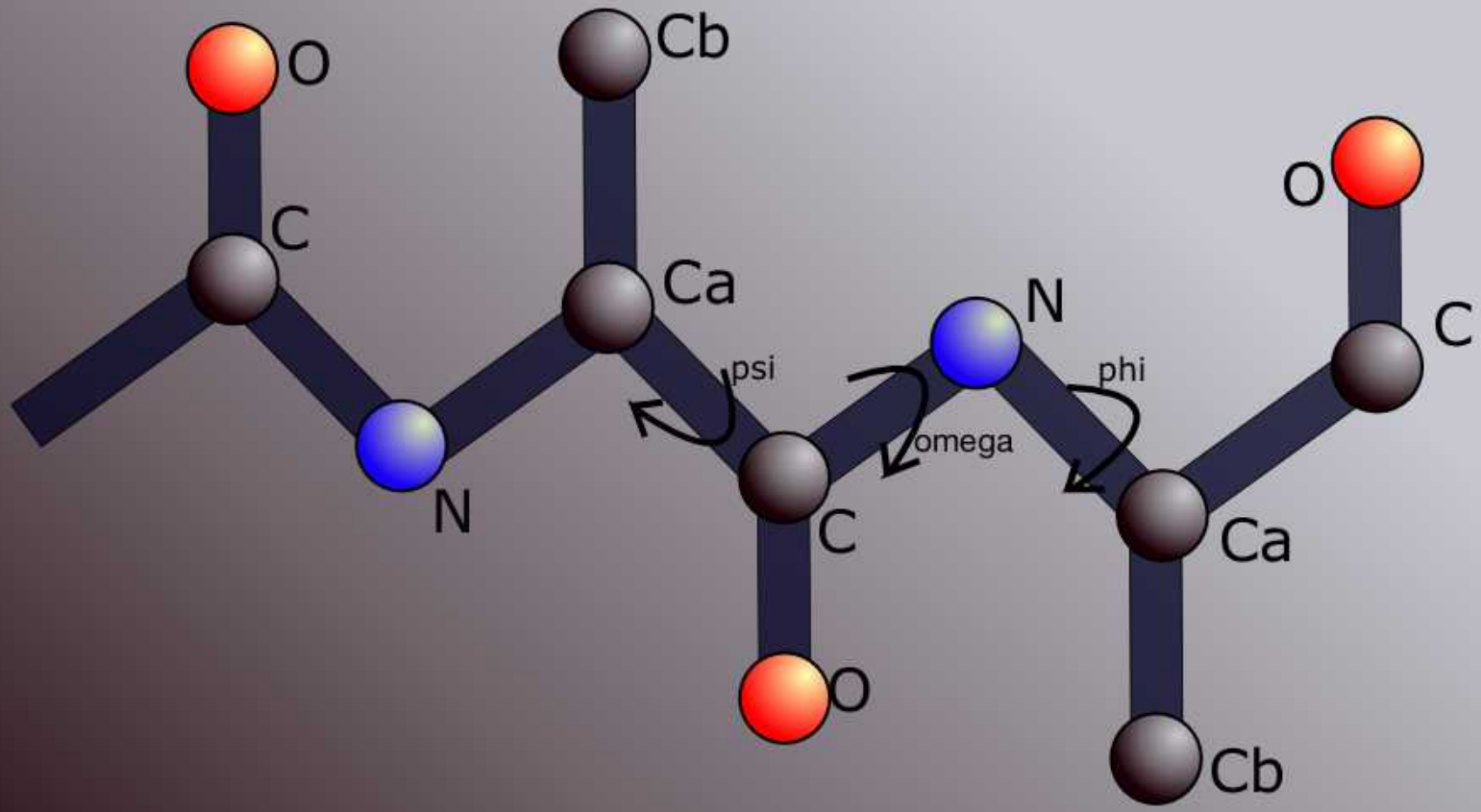


Kleywegt Plots[*]



[*] Named by George Sheldrick

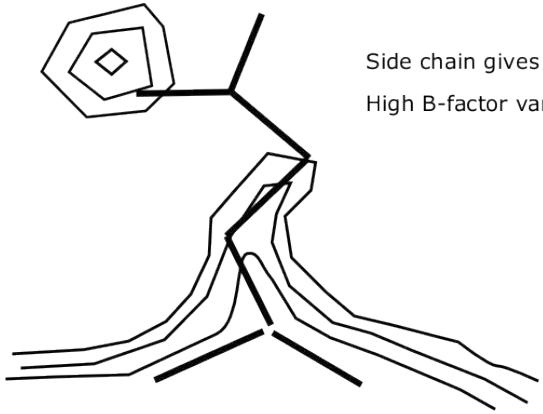
Peptide Torsion Angles



Peptide ω

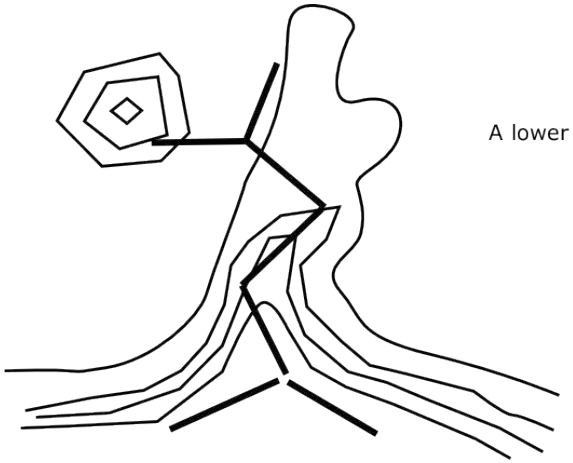
- Needed to check the planarity of the peptide link
  - At low resolutions it is possible to give the protein lots of (too much) freedom to optimize the fit to the density
- Can accidentally create CIS peptides
  - When discovered they are easily reconverted using the CIS<->TRANS peptide tool
- Less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

B-factor variance

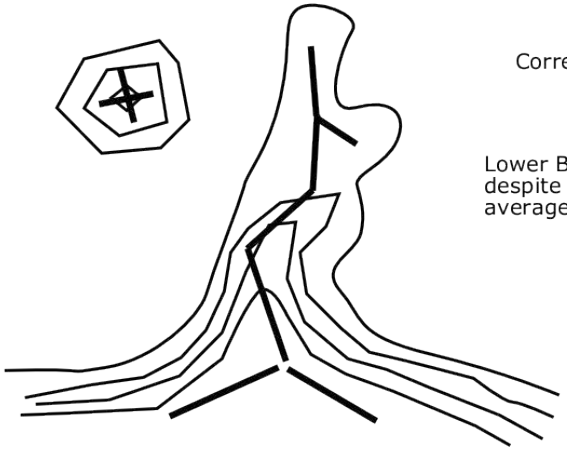


Side chain gives us concern

High B-factor variance



A lower contour level



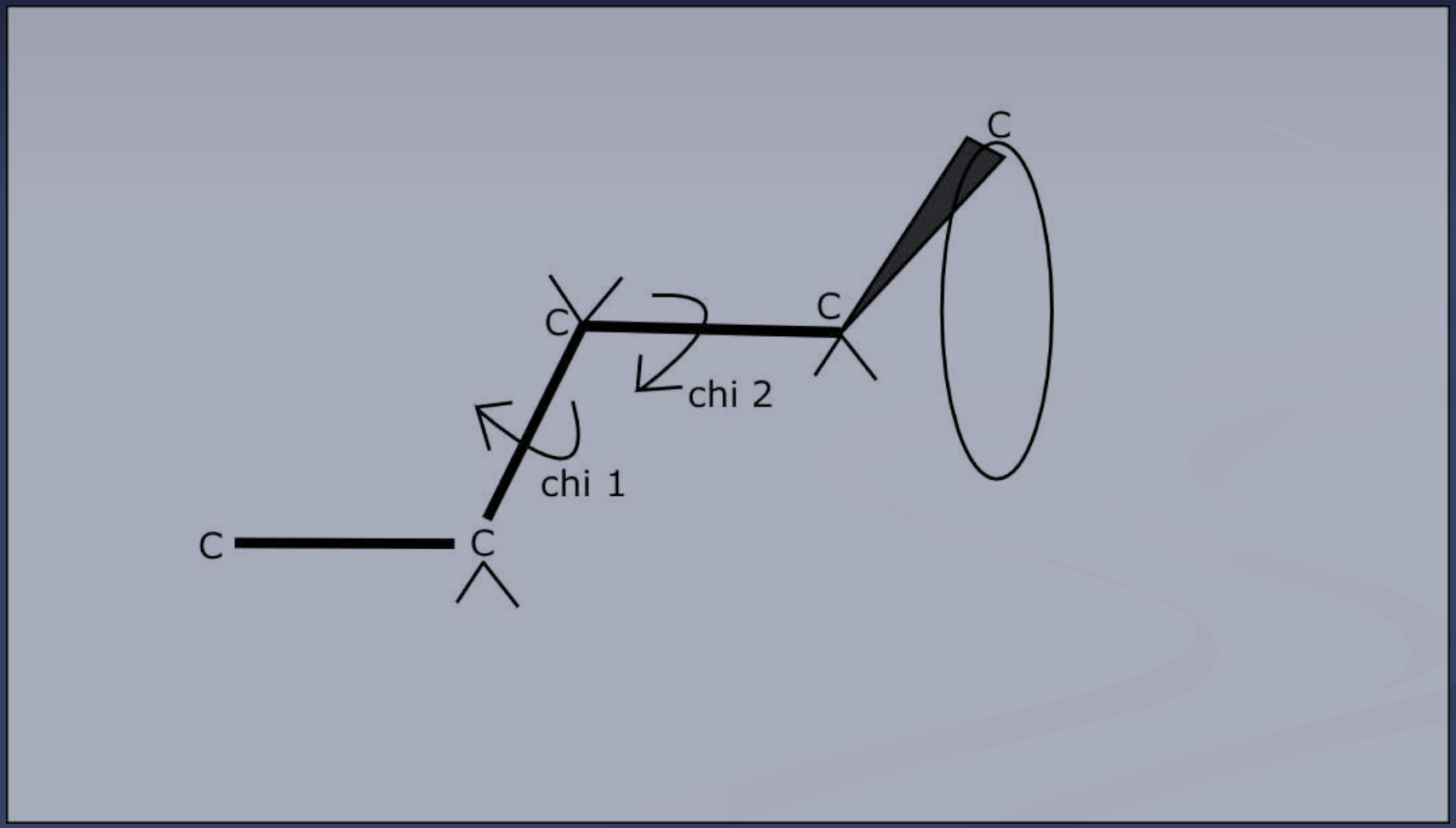
Correctly modelled side-chain

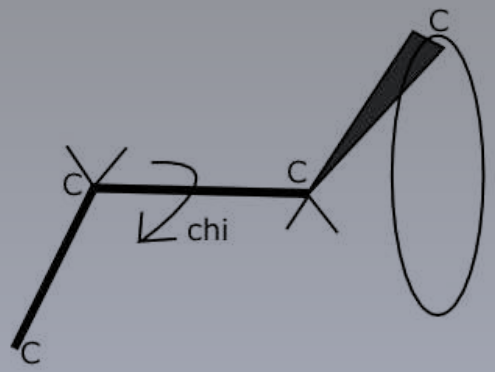
Lower B-factor variance
despite potentially higher
average B-factor

Rotamers

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

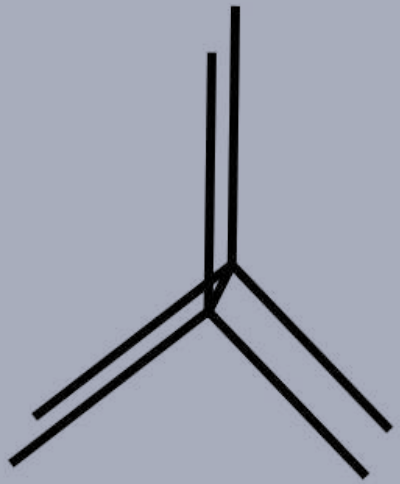
Torsion-based Validation





In principle, there is free rotation

In practice, staggered is energetically more favoured

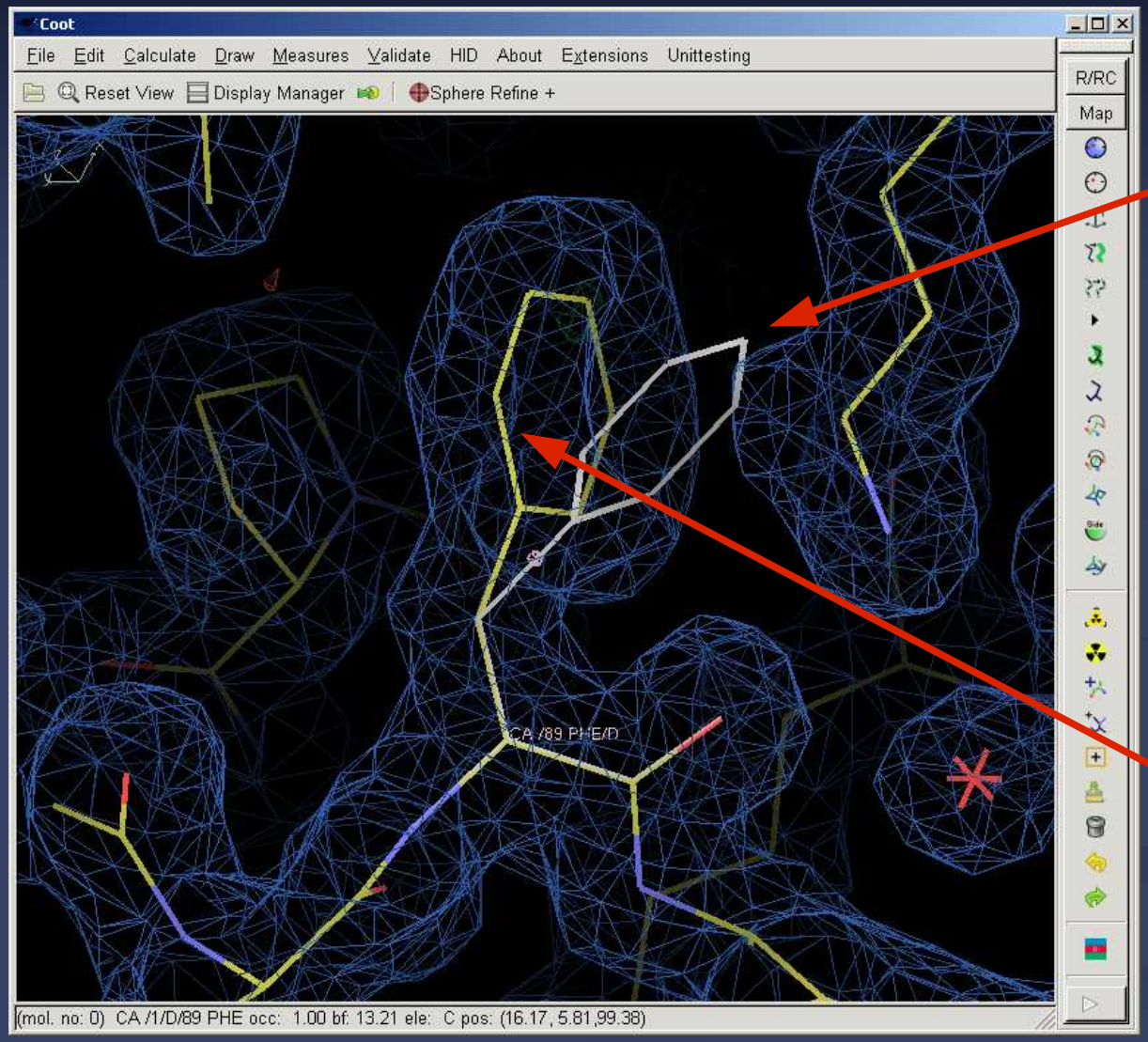


Eclipsed

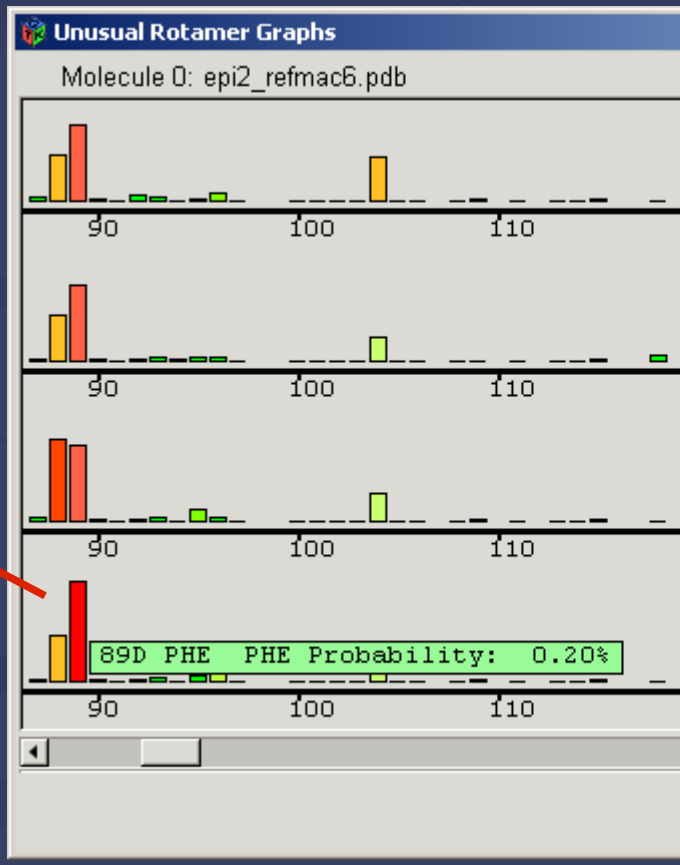
Staggered

Most favoured staggering angles 60, 180, -60 degrees

Not all outliers are bad

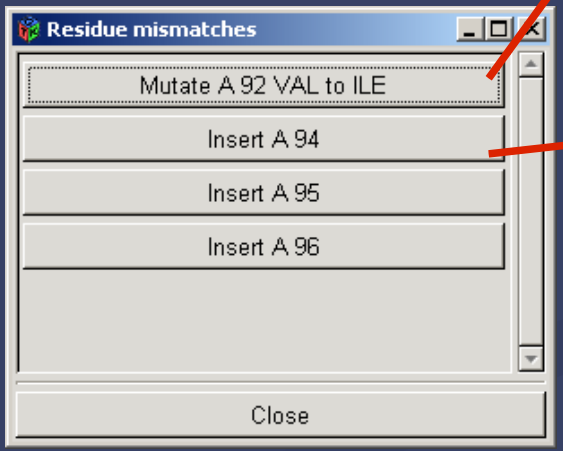


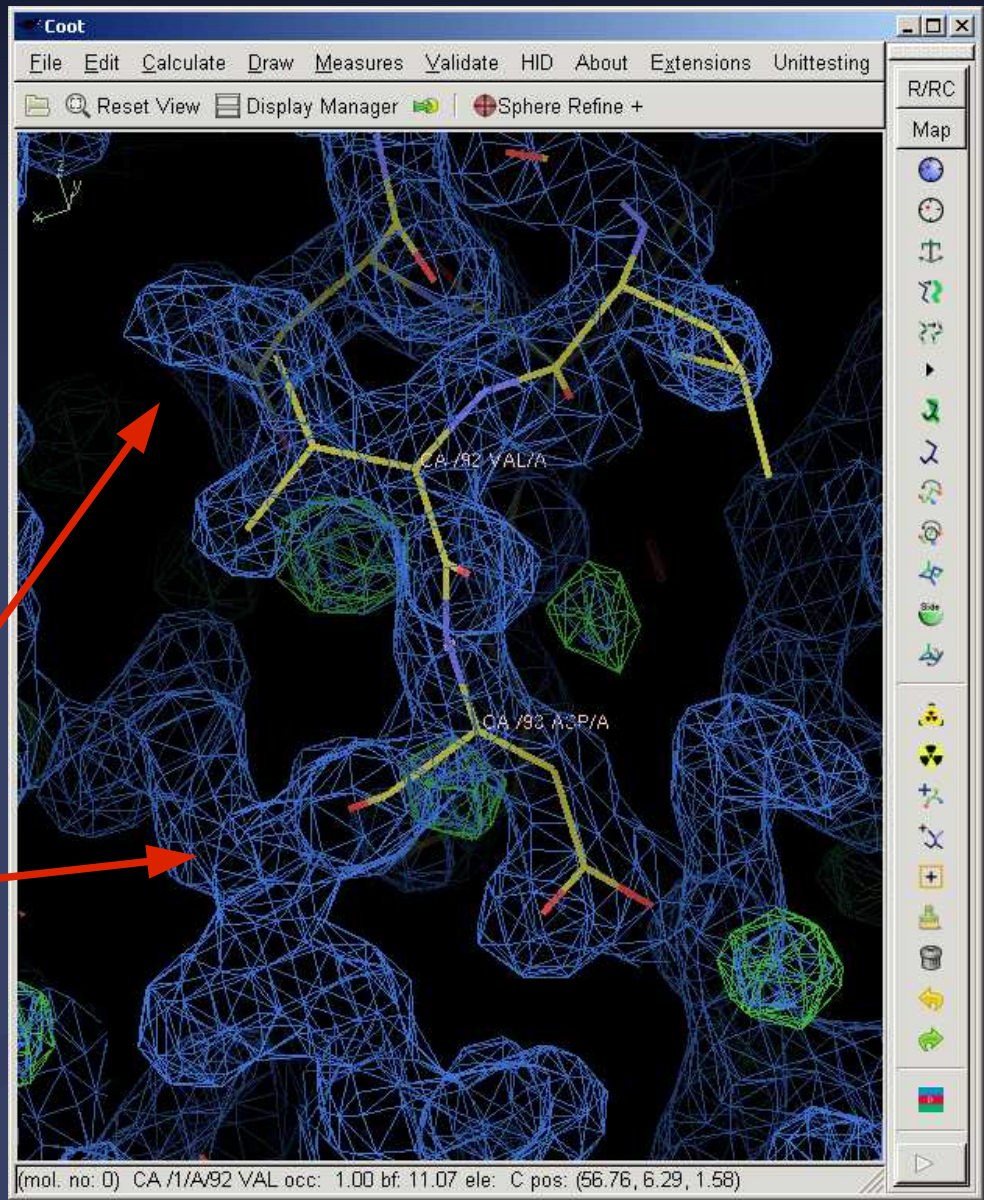
Correct rotamer(?!)



Sequence validation

- Assign sequence (Extensions → Dock sequence → Associate sequence)
- Sequence alignment of model to given sequence (Validate →Alignment vs PIR)





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

Analysis output: all-atom contacts and geometry for 1bkrFH.pdb

Summary statistics

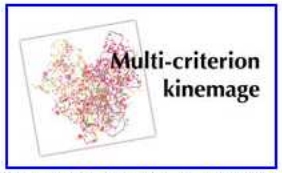
All-Atom Contacts	Clashscore, all atoms:	8.76	38th percentile* (N=222, 1.10Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	6.19%	Goal: <1%
	Ramachandran outliers	0.00%	Goal: <0.2%
	Ramachandran favored	98.11%	Goal: >98%
	Cβ deviations >0.25Å	2	Goal: 0
	MolProbity score^	2.07	18th percentile* (N=932, 1.10Å ± 0.25Å)
	Residues with bad bonds:	0.00%	Goal: 0%
	Residues with bad angles:	0.00%	Goal: <0.1%

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

^ MolProbity score is defined as the following: 0.42574*log(1+clashscore) + 0.32996*log(1+max(0,pctRotOut-1)) + 0.24979*log(1+max(0,100-pctRamaFavored-2)) + 0.5

By adding H to this model and allowing Asn/Gln/His flips, you have already improved your clashscore by 3.51 points. **Make sure you download the modified PDB to take advantage of these improvements!**

Multi-criterion visualizations



View in KiNG | Download (633 Kb)

View (232 Kb)

Scheme Script Download (8.6 Kb) | Python Script Download (6.2 Kb)

Open this in Coot 0.1.2 or later using Calculate | Run Script...

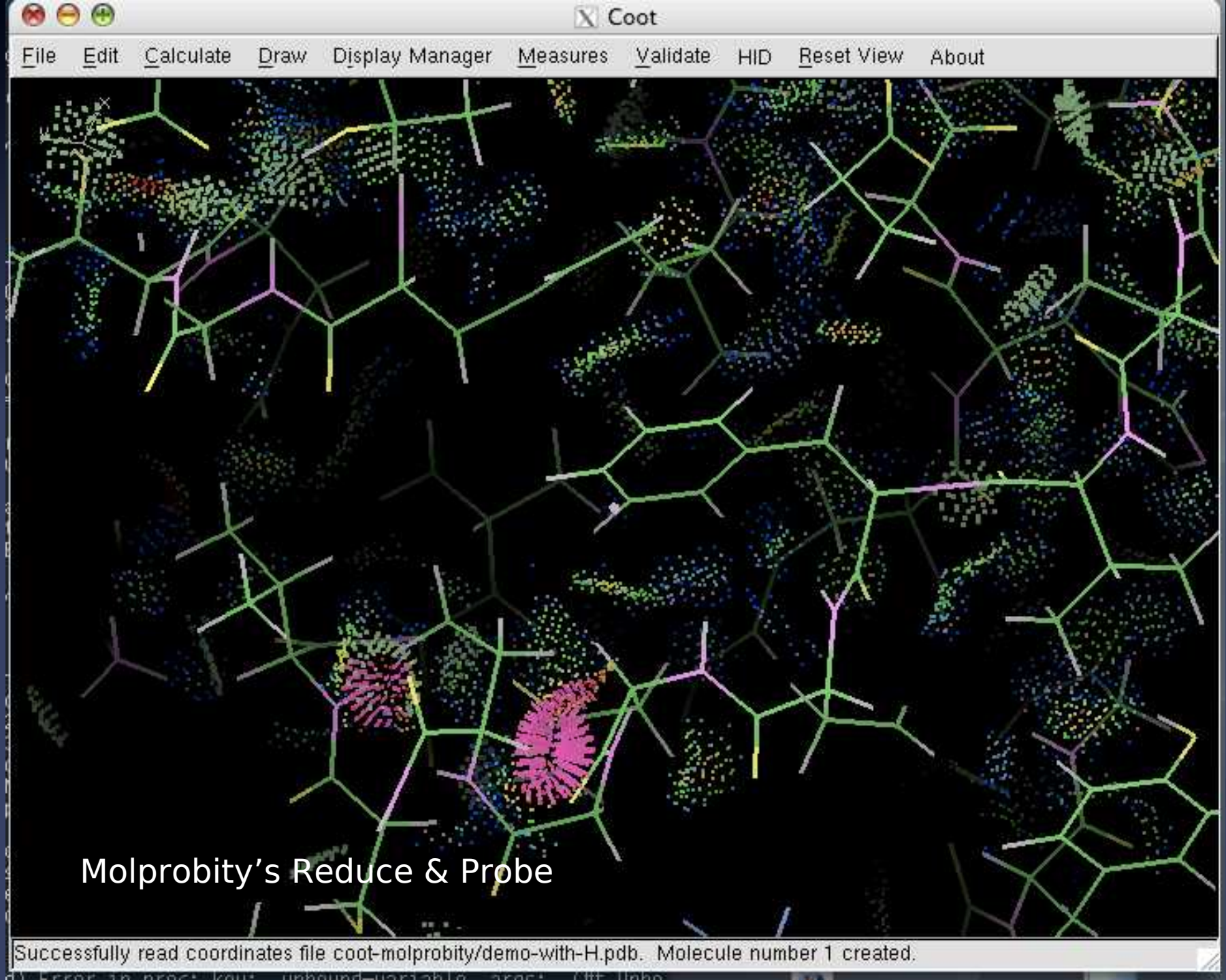
Single-criterion visualizations

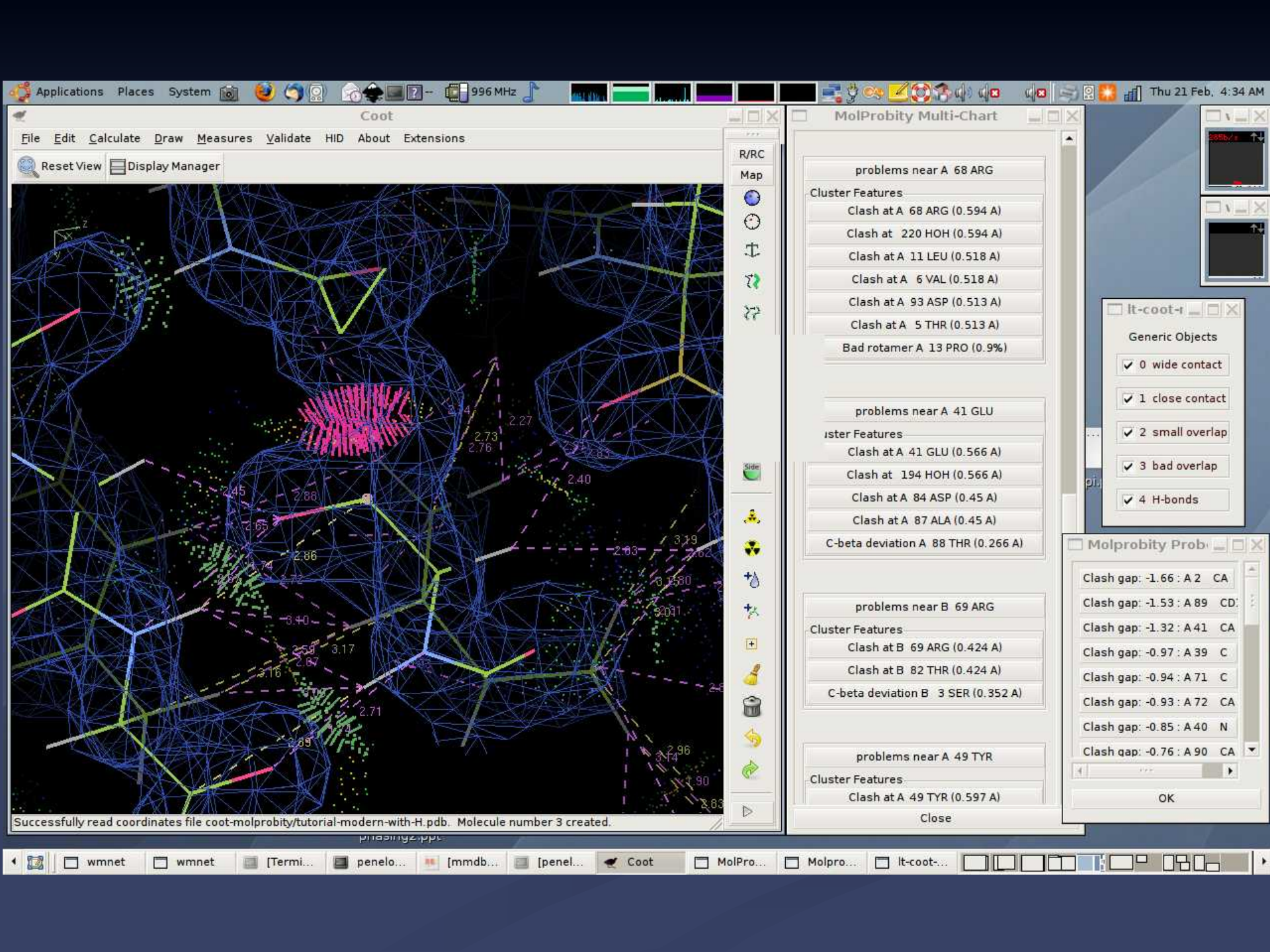
- **Clash list** (1.1 Kb): View
- **Ramachandran plot kinemage** (339 Kb): View in KiNG | Download
- **Ramachandran plot PDF** (187 Kb): View
- **Cβ deviation scatter plot** (16 Kb): View in KiNG | Download

REMARK 40

You can download your PDB file with REMARK 40 inserted, or the same without hydrogens.

```
REMARK  40
REMARK  40 MOLPROBITY STRUCTURE VALIDATION
REMARK  40  PROGRAMS     : MOLPROBITY  (KING, REDUCE, AND PROBE)
REMARK  40  AUTHORS      : I.W.DAVIS,V.B.CHEN,
REMARK  40               : R.M.IMMORMINO,J.J.HEADD,W.B.ARENDALL,J.M.WORD
REMARK  40  URL          : HTTP://KINEMAGE.BIOCHEM.DUKE.EDU/MOLPROBITY/
REMARK  40  AUTHORS      : I.W.DAVIS,A.LEAVER-FAY,V.B.CHEN,J.N.BLOCK,
```


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
Successfully read coordinates file coot-molprobity/tutorial-modern-with-H.pdb. Molecule number 3 created.
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
wmnet wmnet [Termi... penelo... [mmdb... [penel... Coot MolPro... Molpro... lt-coot-...

Other Programs

- Molprobity 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