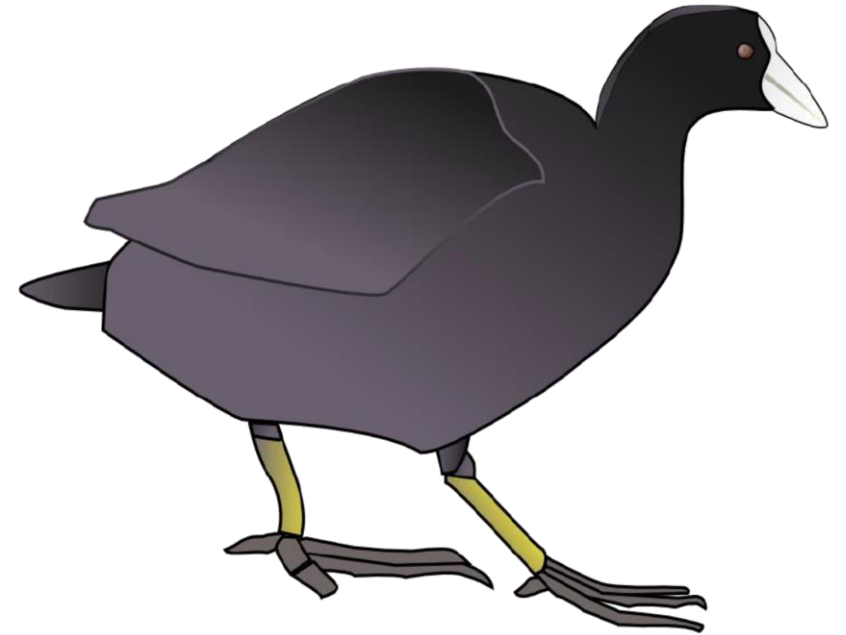


Ligand fitting with Coot

Daren Fearon
XChem Beamline Scientist
Diamond Light Source

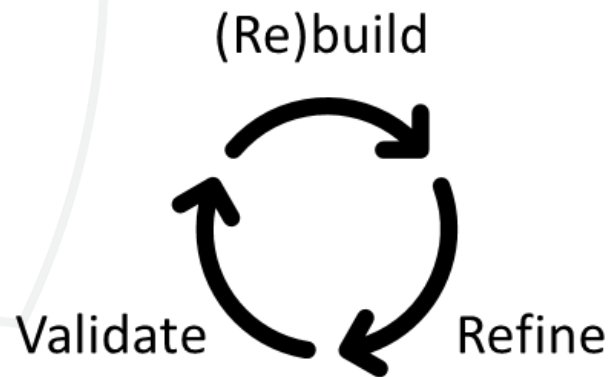
What is Coot?

- Crystallographic Object-Orientated Toolkit
- Developed by Paul Emsley and others
- Interface for model building and validation
- Integrates with other programmes
 - REFMAC/Phenix/Buster, AceDRG/Grade/eLBOW , Molprobability/Mogul
- Critical part of crystallography and cryoEM toolkit



Ligand fitting aims

- Model small molecule bound to protein of interest
- Good correlation with experimental data (electron density, EM maps)
- Sensible molecular interactions with protein and surrounding solvent
- Realistic ligand restraints (bond angles & lengths, planarity etc)

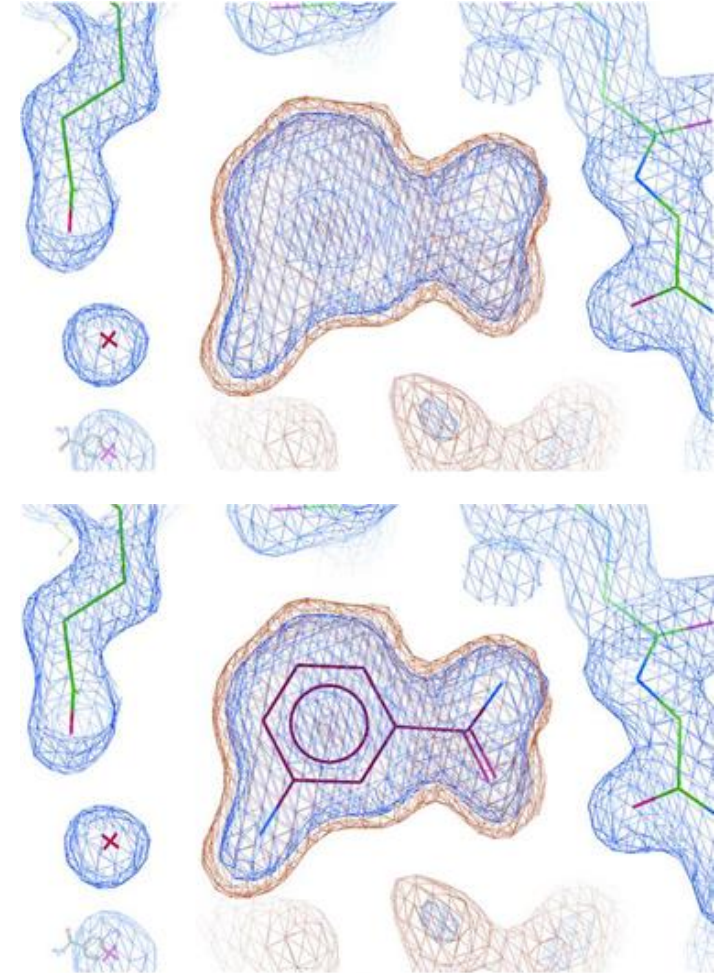


Ligand fitting scenarios with Coot

		Ligand Site	
		Known	Unknown
Ligand Structure	Known	✓	✓
	Cocktails	✓	✓
	Unknown	✗	✗

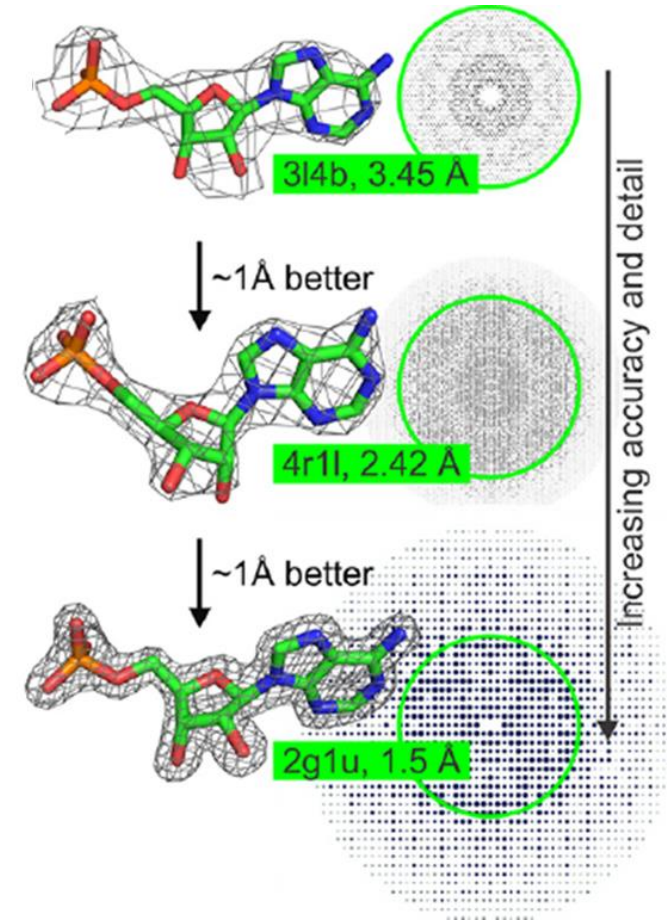
Ligand fitting process

- Identification of ligand electron density
- Generation of ligand restraints
- Ligand fitting
- Optimisation of ligand conformation and model refinement
- Validation of model
- Analysis and visualisation



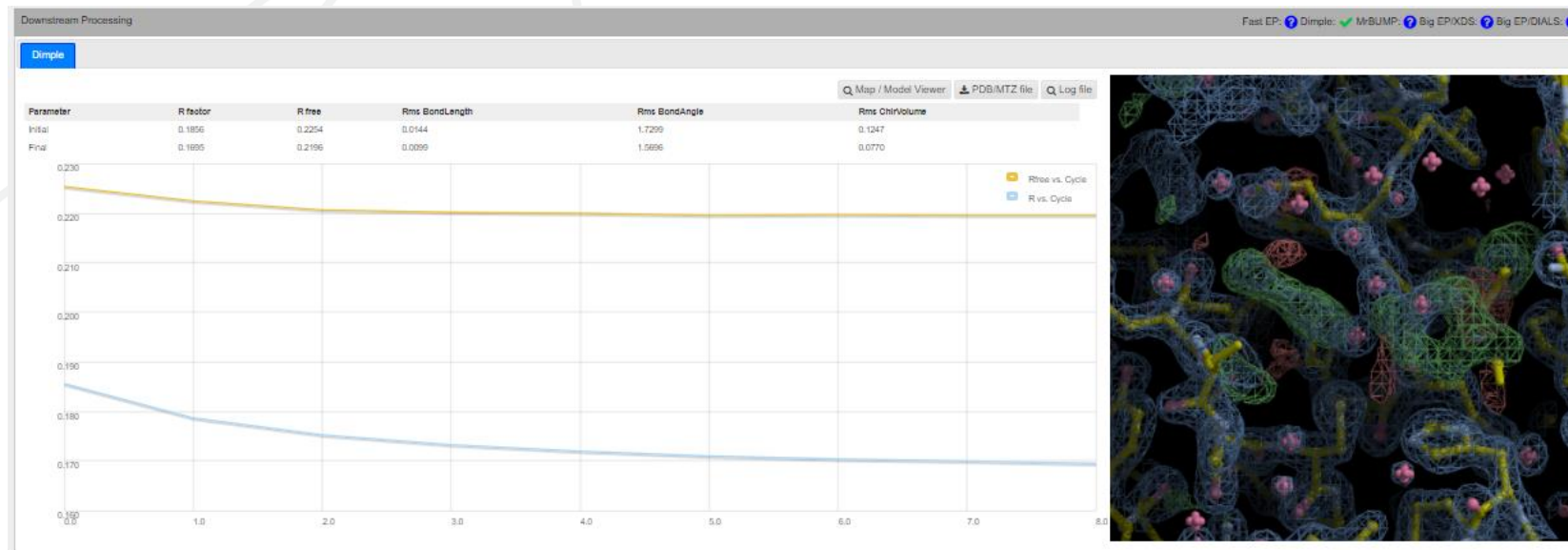
Assumptions

- A good starting model is available for the protein
 - If not, see slides from Judit Debreczeni last week for how to refine & validate your protein model with Coot
- Ligand description (2D structure/SMILES string) available or ligand is contained in CCP4/Refmac ligand dictionary (e.g. ATP)
- Data is of sufficient quality to model ligand
 - Ideally 2.5 Å or better

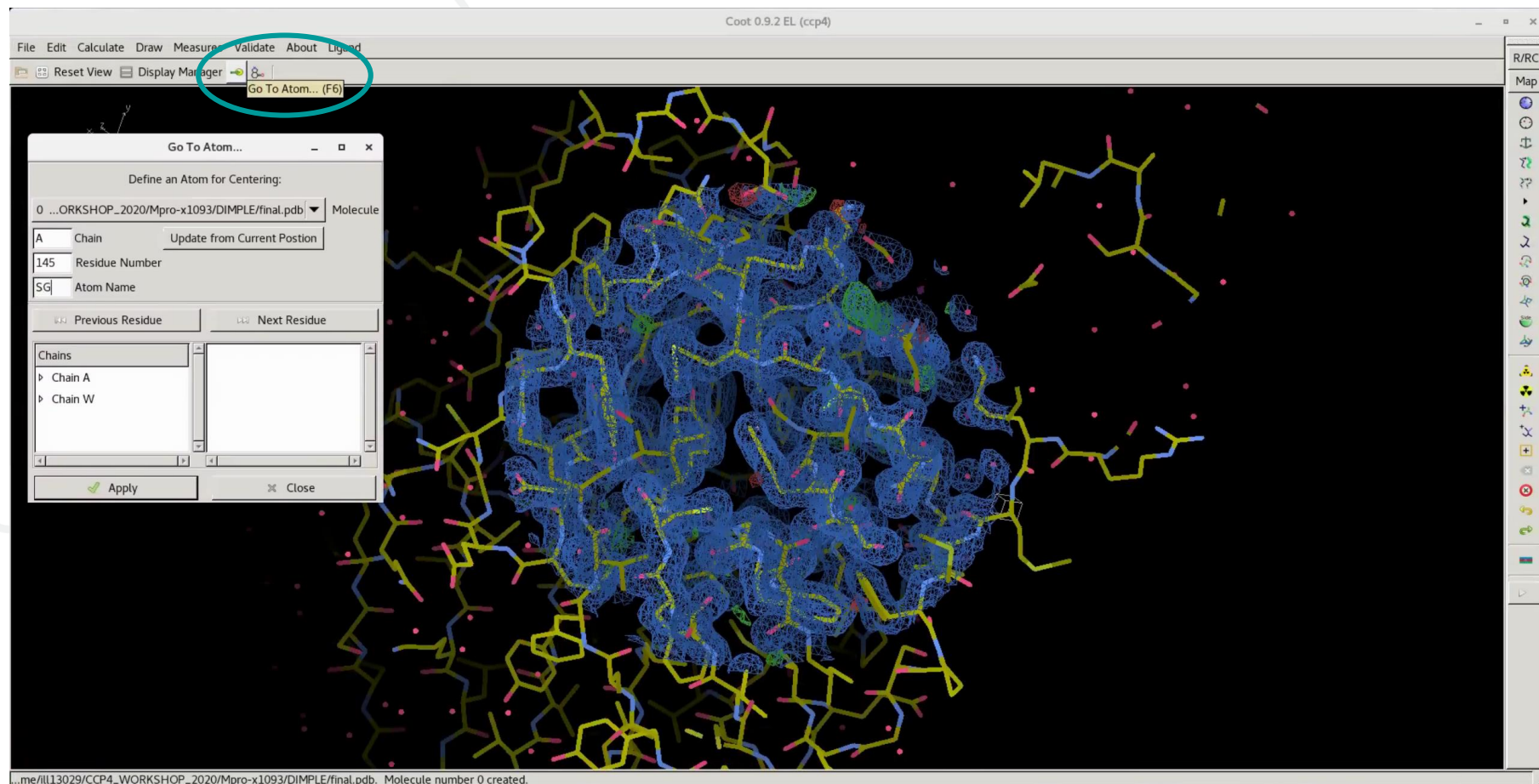


Identifying ligand density

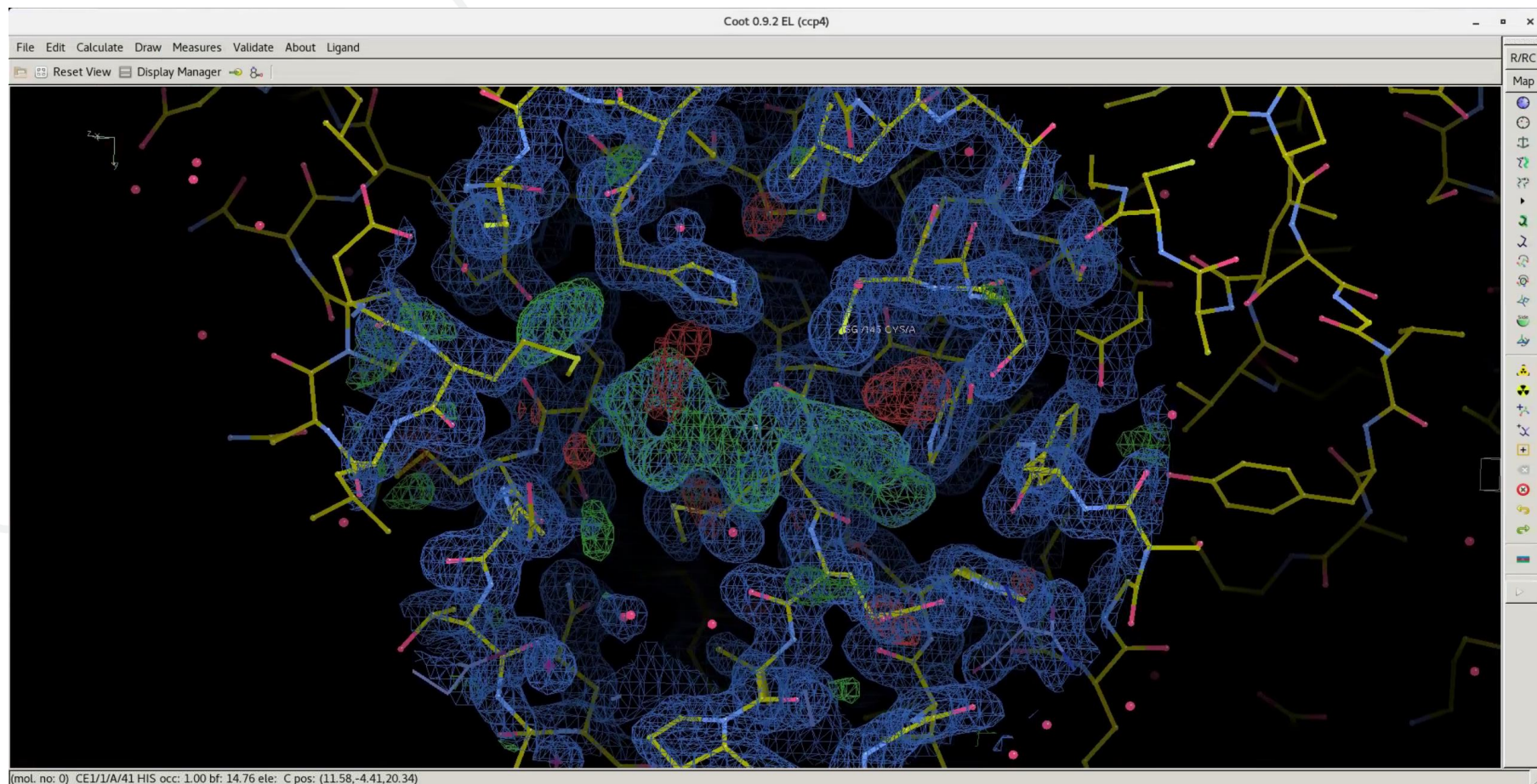
- DIMPLE
 - Runs molecular replacement and provides model and difference map
 - Checks difference map for unmodelled blobs that could correspond to ligands
 - At Diamond, automatically runs after auto processing if model provided



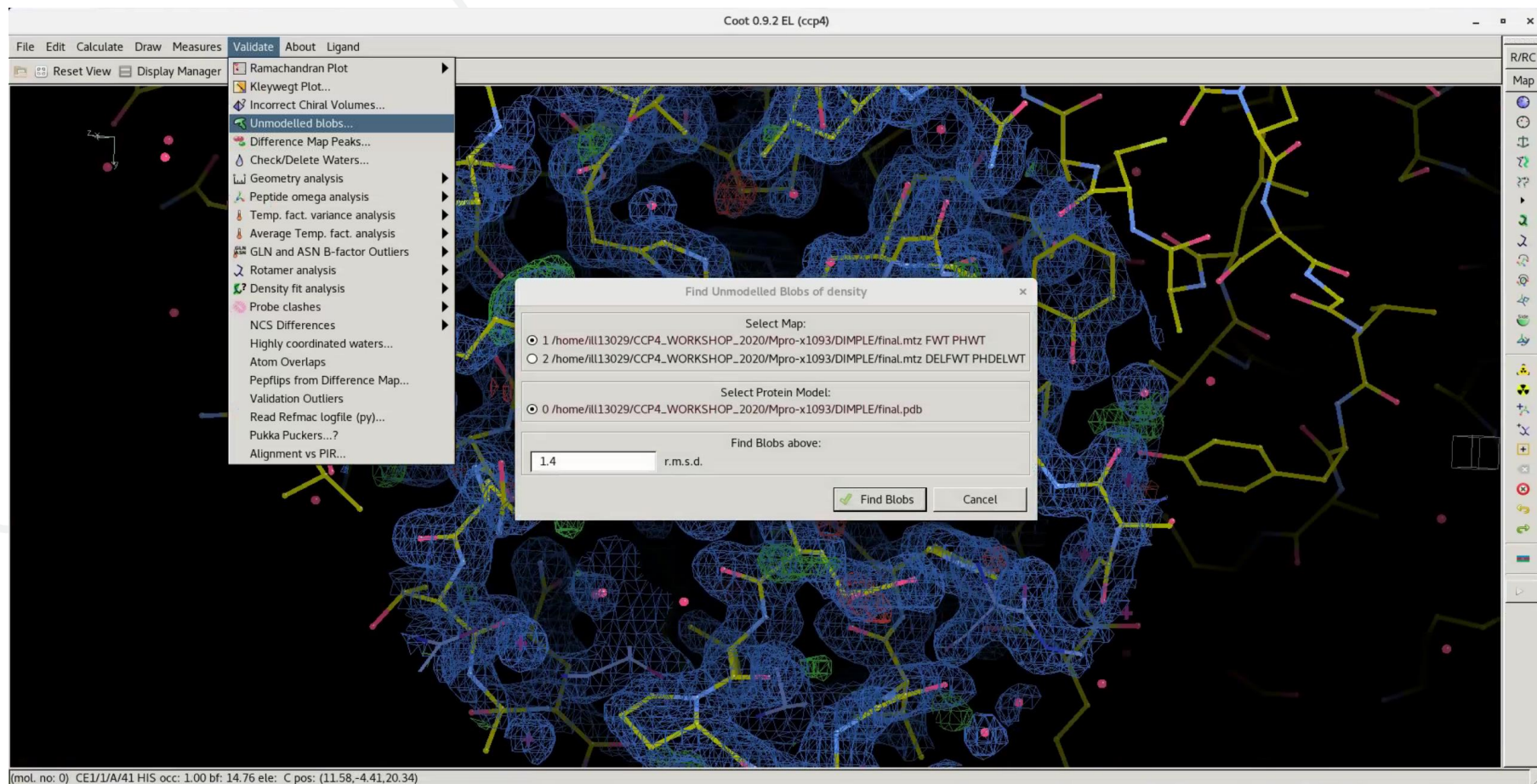
Identifying ligand density – known ligand site



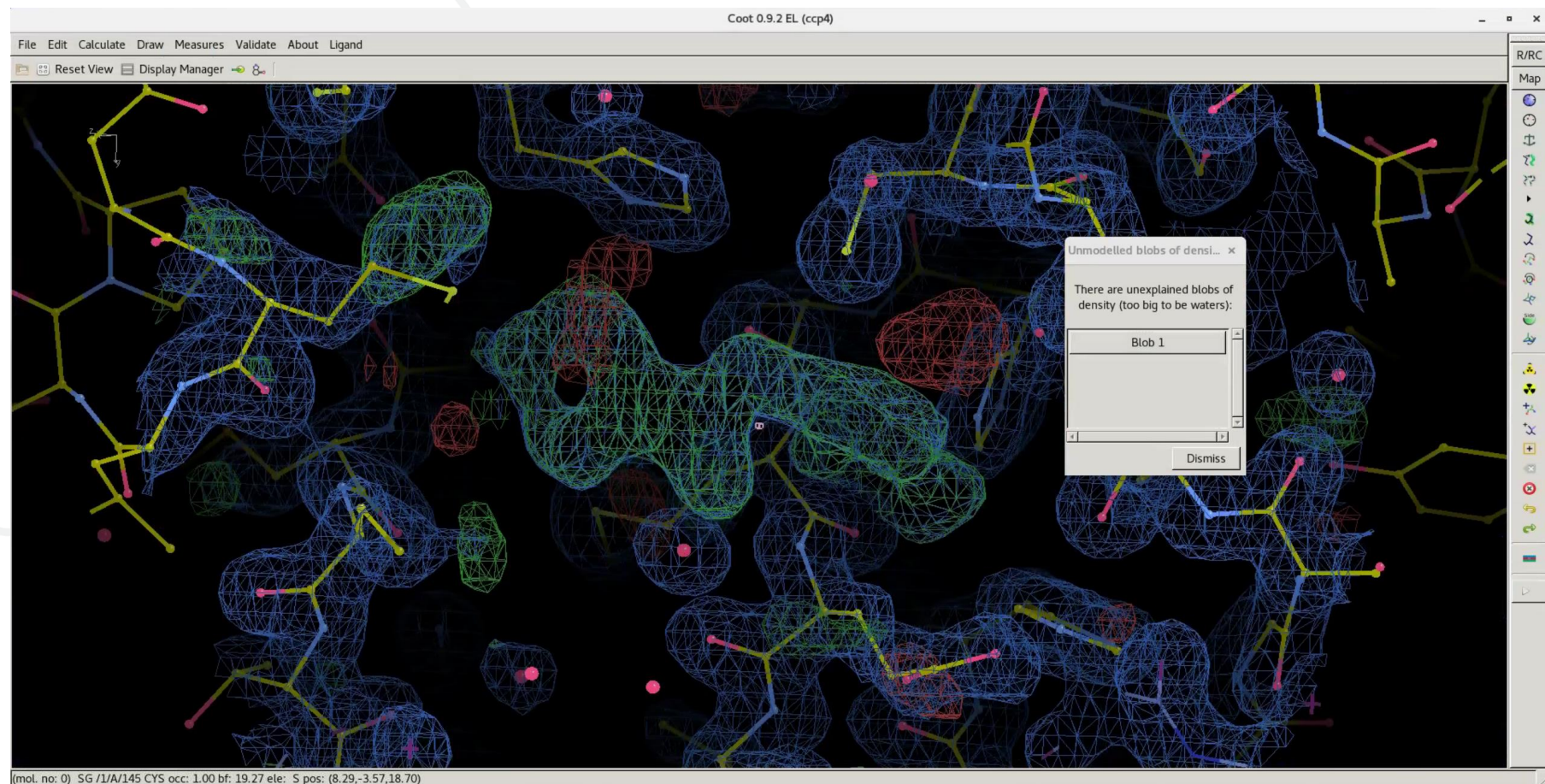
Identifying ligand density – known ligand site



Identifying ligand density – unknown site



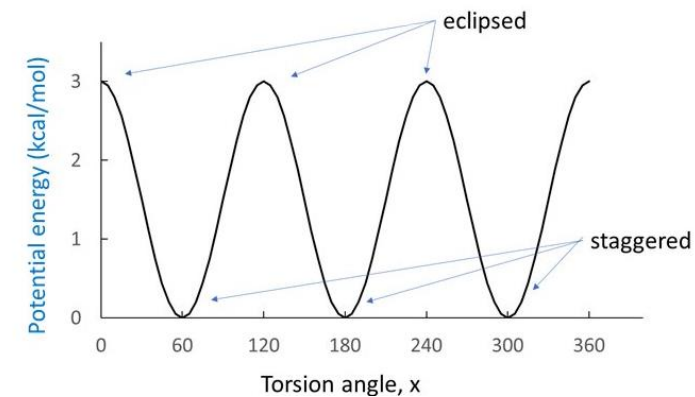
Identifying ligand density – unknown site



Ligand restraints

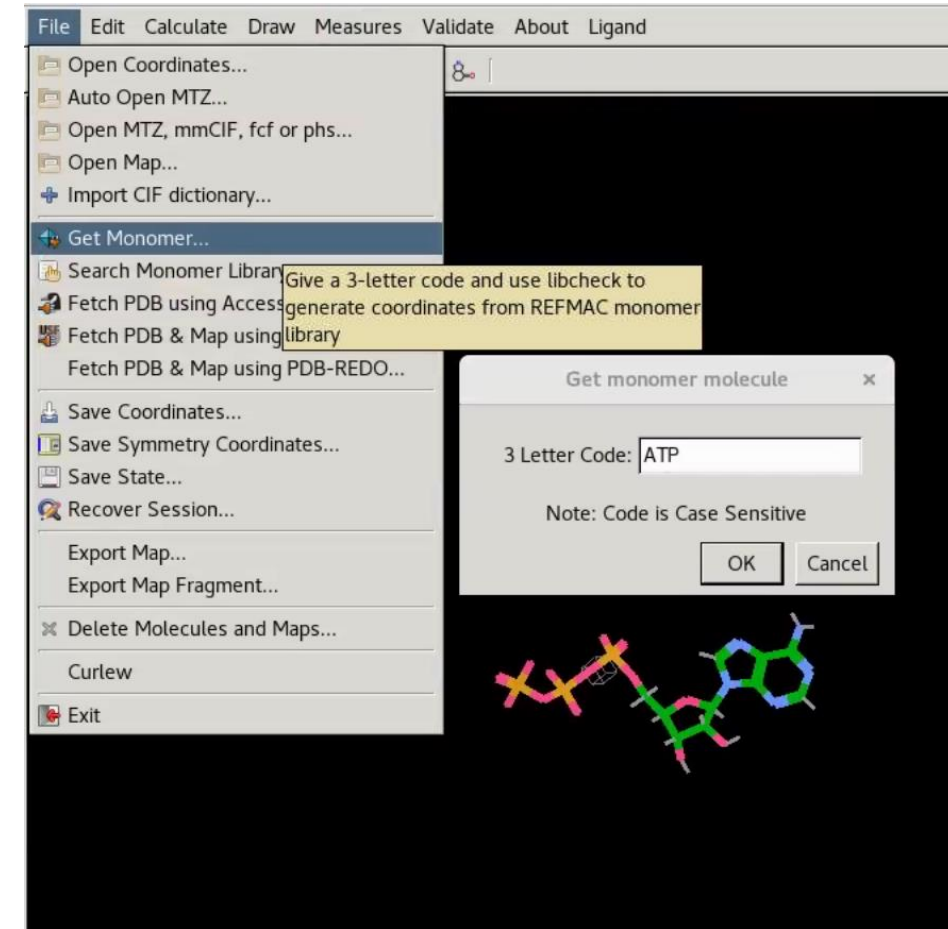
- In macromolecular crystallography experimental data alone insufficient to accurately define 3D structure
- Refinement packages utilize both geometry and experimental data
- For small molecules, a ligand description is used to generate a dictionary used for modelling
- Restraints describe ideal values and estimated standard deviations for bond lengths/angles

```
#
data_comp_list
loop_
  chem_comp.id
  chem_comp.three_letter_code
  chem_comp.name
  chem_comp.group
  chem_comp.number_atoms_all
  chem_comp.number_atoms_nh
  chem_comp.desc_level
LIG      LIG      .      non-polymer      34      21
#
data_comp_LIG
#
loop_
  chem_comp_atom.comp_id
  chem_comp_atom.atom_id
  chem_comp_atom.type_symbol
  chem_comp_atom.type_energy
  chem_comp_atom.charge
  chem_comp_atom.x
  chem_comp_atom.y
  chem_comp_atom.z
LIG      CL1      CL      CL      0      0.933      -0.811      3.593
LIG      C1      C      CR6      0      1.581      -0.608      1.989
LIG      C2      C      CR16     0      2.567      0.333      1.772
LIG      C3      C      CR16     0      3.071      0.481      0.492
LIG      C4      C      CR16     0      2.597      -0.299     -0.551
LIG      C5      C      CR6      0      1.604      -1.248     -0.334
LIG      C6      C      CH2      0      1.078      -2.101     -1.467
LIG      C7      C      C        0      0.531      -1.279     -2.618
LIG      O1      O      O        0      1.108      -1.270     -3.697
LIG      N1      N      NH1      0     -0.609     -0.569     -2.399
LIG      C8      C      CR6      0     -1.415     -0.505     -1.239
LIG      C9      C      CR16     0     -2.300     -1.538     -0.925
LIG      N2      N      NRD6     0     -3.081     -1.514     0.169
LIG      C10     C      CR16     0     -3.011     -0.474     0.984
LIG      C11     C      CR66     0     -2.152     0.648     0.787
LIG      C12     C      CR16     0     -2.091     1.751     1.669
LIG      C13     C      CR16     0     -1.241     2.788     1.411
LIG      C14     C      CR16     0     -0.415     2.779     0.268
LIG      C15     C      CR16     0     -0.451     1.726     -0.605
LIG      C16     C      CR66     0     -1.322     0.628     -0.373
LIG      C17     C      CR16     0      1.100     -1.393     0.955
```

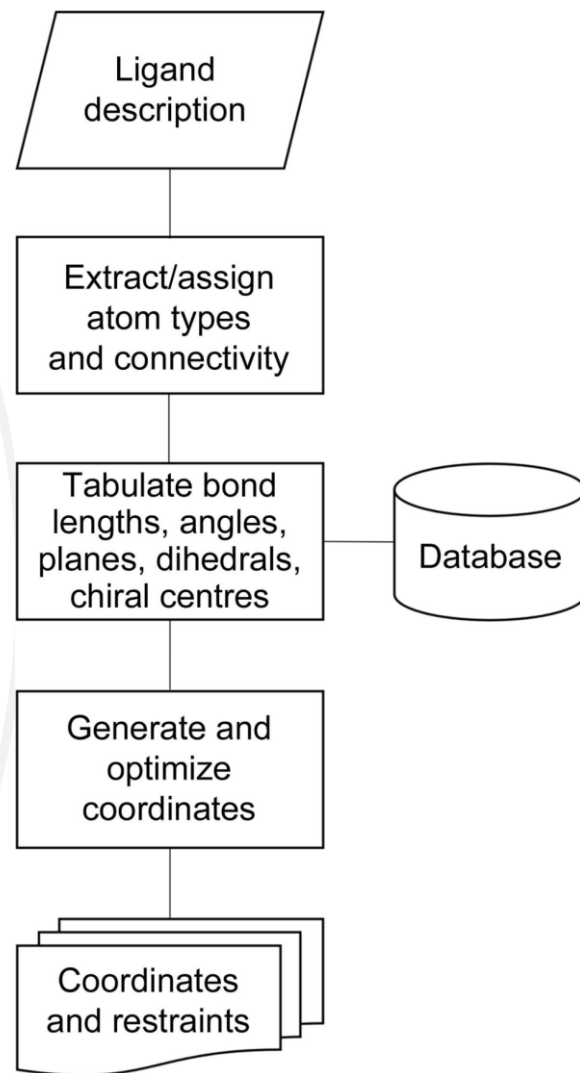


Fitting common ligands

- CCP4/REFMAC library contains dictionaries for over 2,000 common monomers including amino acids, nucleic acids and saccharides
- Import ligand using “Get Monomer” function and three letter code for ligand e.g. ATP



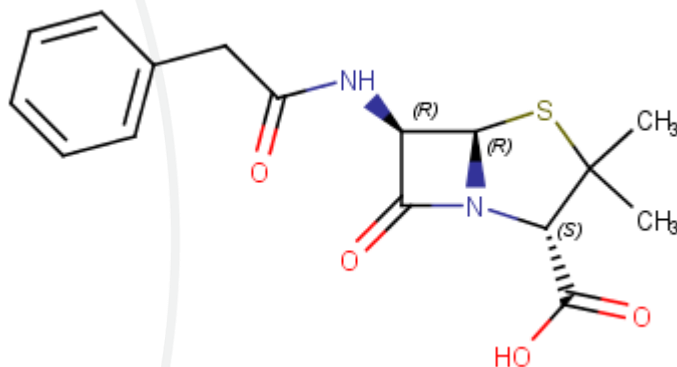
Generating dictionary for novel ligands



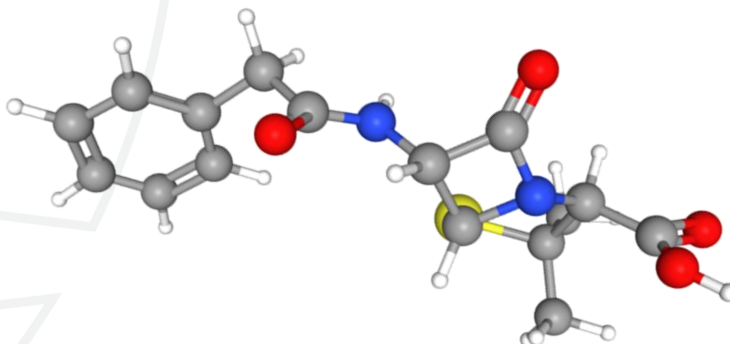
Ligand description – common input formats

- Simplified molecular-input line-entry system (SMILES) string
 - CC1([C@@H](N2[C@H](S1)[C@@H](C2=O)NC(=O)CC3=CC=CC=C3)C(=O)O)C

- 2D structure:



- 3D coordinates:
 - PDB/mol file



Atom energy types

- First step in generating ligand dictionary is defining atom type
 - Determined by chemical element and connectivity

Atom name	Atom energy type	Atom energy type description
C3	CR5	Carbon without hydrogen in five-atom ring
C7	C1	Carbon connected to one hydrogen
C8	CR6	Carbon without hydrogen in six-atom ring

Deriving bond distances, angles and torsion restraints

- Can be obtained from experimental sources (wwPDB CCD, CSD, COD)
- Where experimental data lacking, molecular simulations can fill the gaps

Force field	Full name	Citation	Parameterization	Usage
MMFF94	Merck Molecular Force Field 94	Halgren (1996)	Electronic structure calculations	Pyrogen, eLBOW, writedict
AM1	Austin Model 1	Dewar et al. (1985)	Semi-empirical method	eLBOW, grade
RM1	Recife Model 1	Rocha et al. (2006)	Semi-empirical method	eLBOW, grade
PM3	Parametrized Model No. 3	Stewart (1989)	Semi-empirical method	eLBOW, grade

- Ideal values and associated deviations are defined

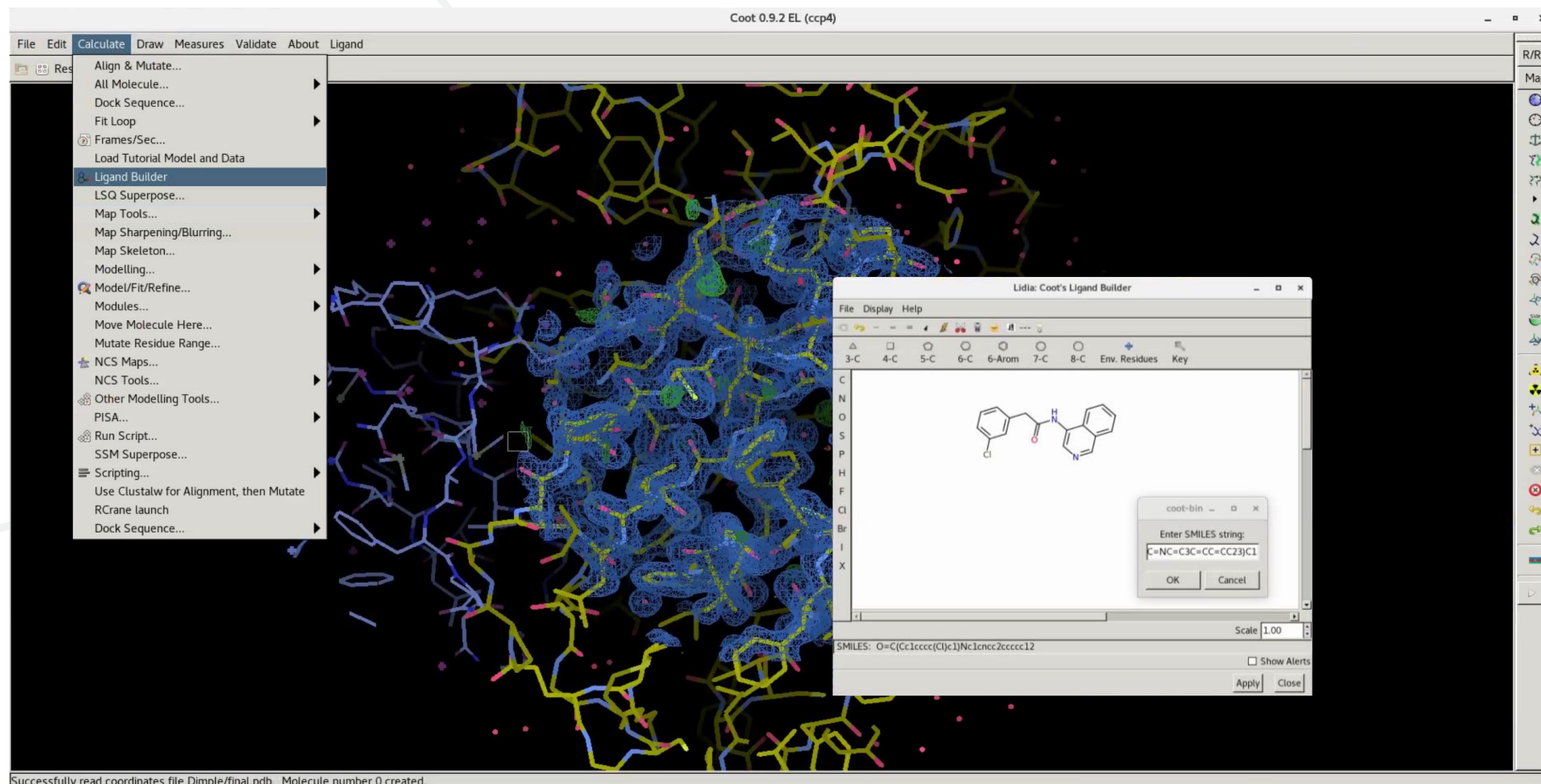
Generating and optimizing coordinates

- Various programs available, mostly free for academic users

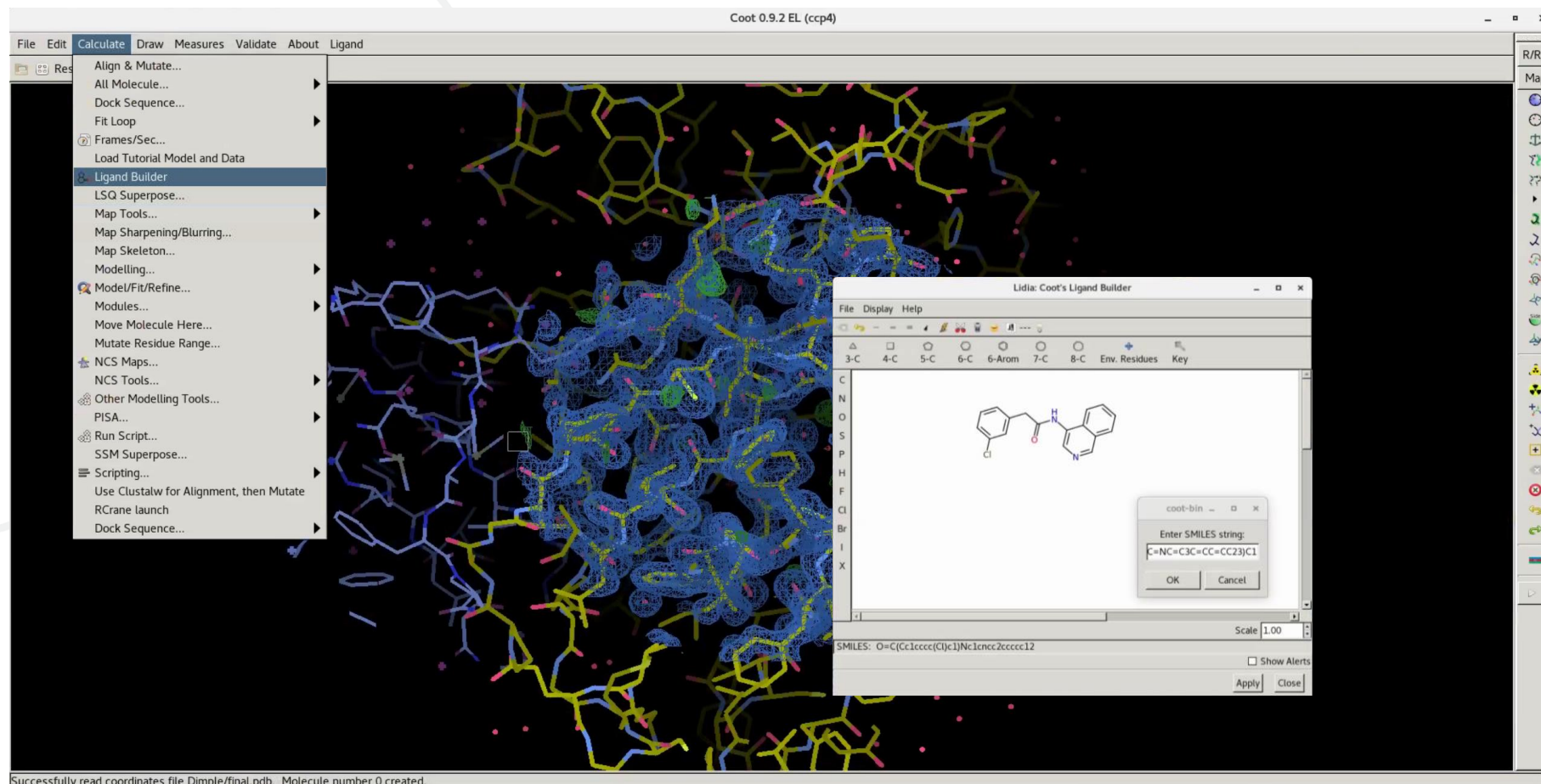
Program name	AceDRG	Grade	eLBOW
Distributor	CCP4	Global Phasing	PHENIX
Input formats	SMILES, PDB, CIF	SMILES, Molfile, CIF	SMILES, PDB, CIF
Output formats	PDB, CIF	PDB, CIF, SHELX	Multiple, including PDB, CIF
Experimental data source(s)	COD (curated)	CSD	CSD
Force field(s)	None	AM1/RM1/PM3	Multiple including AM1, MMFF94
Standard deviation source(s)	COD (curated)	CSD	Multiple including CSD
Restraints editor	JLigand	Edit REFMAC	REEL
Other features and limitations	Hierarchical atom typing	Flexible planar definitions. Available through web server.	Atom name preservation. Metal coordination.
Citation	Long et al. (2017)	Smart et al. (2011)	Moriarty et al. (2009)

- Dictionaries serve as starting points for refinement and model building

Lidia: Coot's ligand builder (AceDRG)

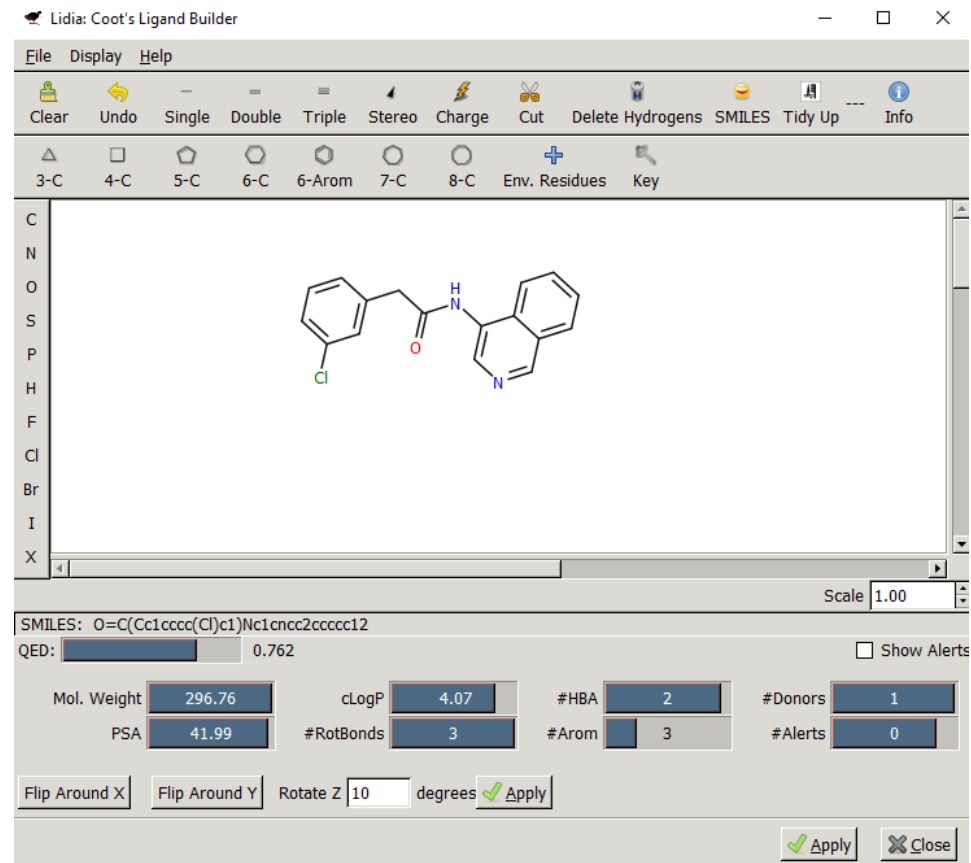


Lidia: Coot's ligand builder (AceDRG)



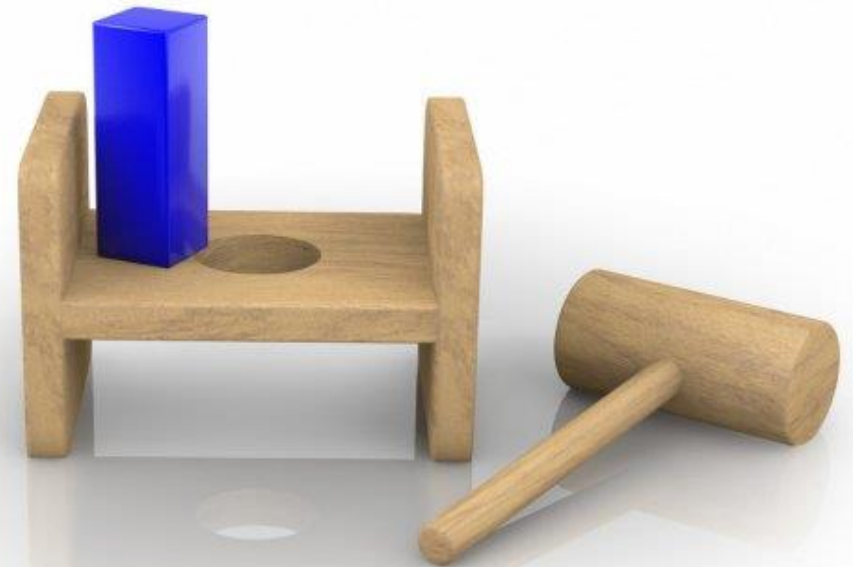
Quantitative Evaluation of Drug-Likeness

- QED score - measure of drug likeness
- Lidia displays various physical chemical properties
 - Molecular weight
 - cLogP
 - Polar Surface Area
 - Hydrogen bond donors/acceptors
 - Rotatable bonds



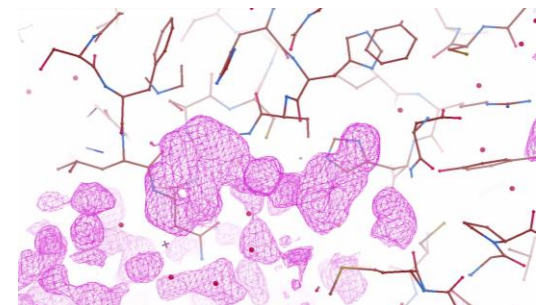
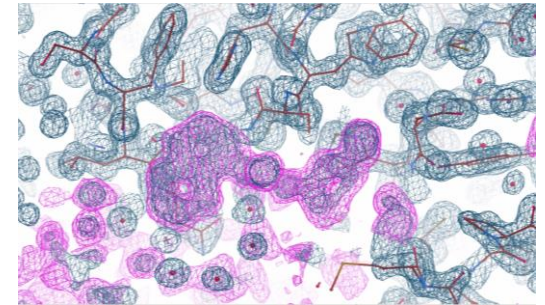
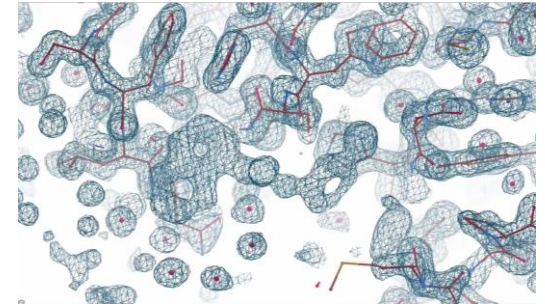
Ligand fitting with Coot

- Involves selection of the correct ligand conformation and positioning/orientating the ligand in density
- What files?
 - Protein model (pdb/mmCIF)
 - Reflection data/maps (mtz)
 - Ligand coordinates (pdb) and restraints (cif)
- Can manually fit ligand to density in Coot using real-space/rigid body refinement
 - Remember to merge ligand and protein model



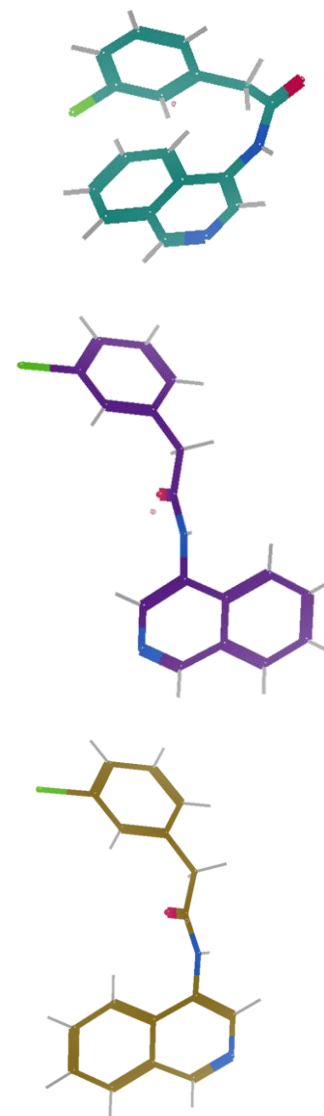
Ligand fitting with Coot – “Find Ligands”

- Automated procedure:
 - Mask the map – protein coordinates used to identify unmodelled density
 - Find blobs – masked map searched for regions with substantial density
 - Cut out blobs and rank them according to density volume
 - Fit ligand into each blob
 - Match centres
 - Orientate ligand
 - Refine ligand
 - Score fit of ligand to blob (density correlation)
- Must merge selected ligand with protein model



Conformation generation

- Ligand from restraint generation may not adopt the same conformation as that of the true structure in the crystal
- Large molecules with greater complexity can adopt a greater number of conformations
- Conformer generation procedure in Coot:
 - Generate torsion angles
 - Generate conformers
 - Optimize coordinates to avoid clashes and high-energy conformations



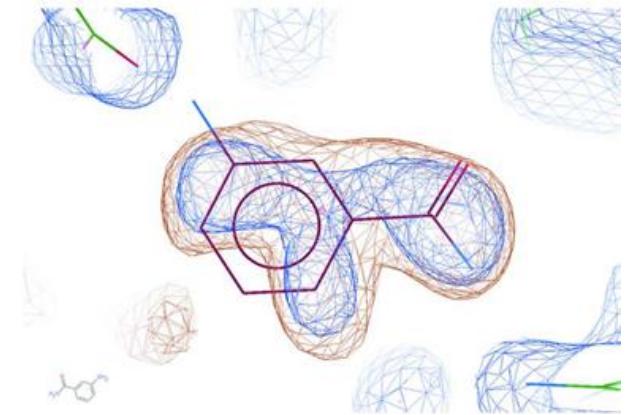
Optimizing ligand fit

- Real-space refinement
 - Sphere refinement (hot key “R”) refines ligand and surrounding residues/solvent molecules
- Rigid body fit zone
 - Constrained refinement of ligand as single body
- Jiggle fit
 - Uses a combination of trialling positions and orientations, rigid-body fitting and real-space refinement
 - Hot key “J”



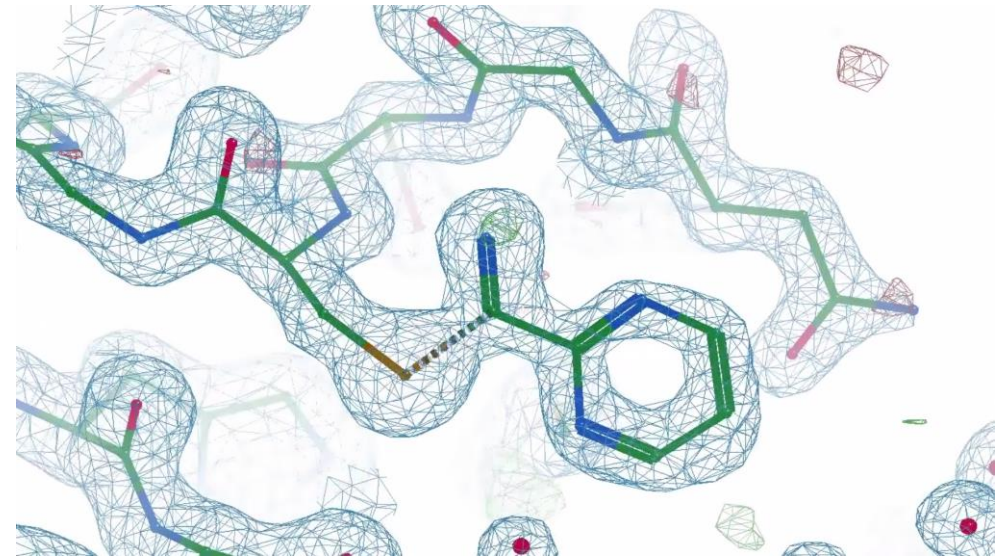
Fitting unknown ligand

- Sometimes additional density observed for unexpected ligand
 - Crystallisation buffer components (HEPES, MES, salts)
 - Cryoprotectants (PEG, glycerol)
 - Compounds from purification (co-factors, substrates)
 - Solvents (DMSO, ethylene glycol)
- Possible to search for some common ligands using Phenix
- May require further experimental work to deconvolute



Fitting covalent ligand

- Initial steps for ligand fitting identical to non-covalent ligand
- Need to create link between ligand and protein and modify restraints
 - Calculate -> Modelling -> Make link (click 2 atoms)
 - Creates link record for PDB file
 - Use JLigand to update restraints
 - Calculate -> Modelling -> JLigand launch

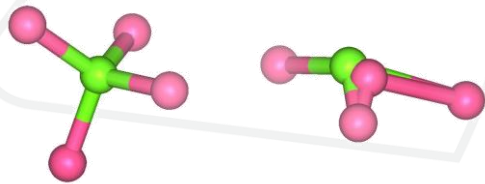


Non-Crystallographic Symmetry and ligands

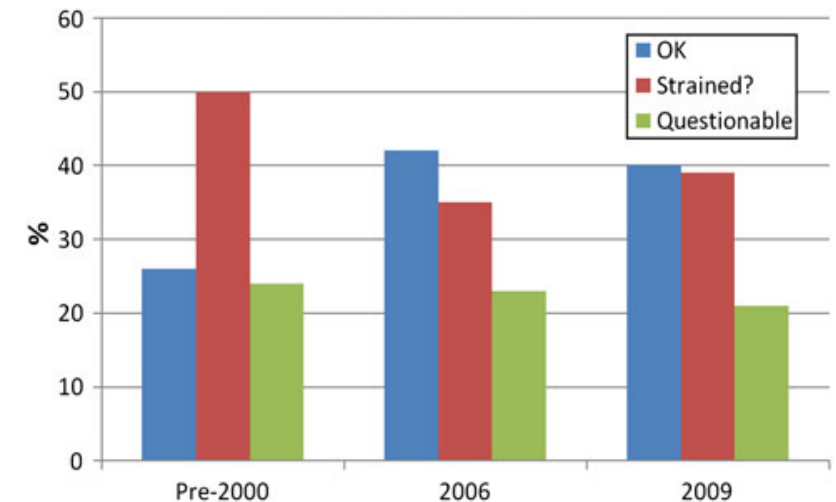
- For structures with 2 or more protein chains related by NCS, it can be exploited to help fit ligands
- Place first copy of ligand as before
- Calculate -> NCS Tools -> NCS ligands
 - Provide protein with NCS and molecule containing ligand (chain ID and res no)
 - Find candidate positions
 - Refine as necessary and merge ligand again
- Show NCS ghosts to check surrounding residues
 - Calculate -> NCS Tools -> NCS ghosts by residue

Ligand validation – why?

- Quality of small molecules in protein-ligand complexes in the PDB is varied - many examples exist of either incorrectly fitted ligands and/or ligands with poor geometry
- Ligand quality not necessarily reflected by overall quality of structure
 - 1DW9
 - 1.65 Å resolution
 - R/Rfree = 0.15/0.19
 - Sulfate geometry clearly incorrect

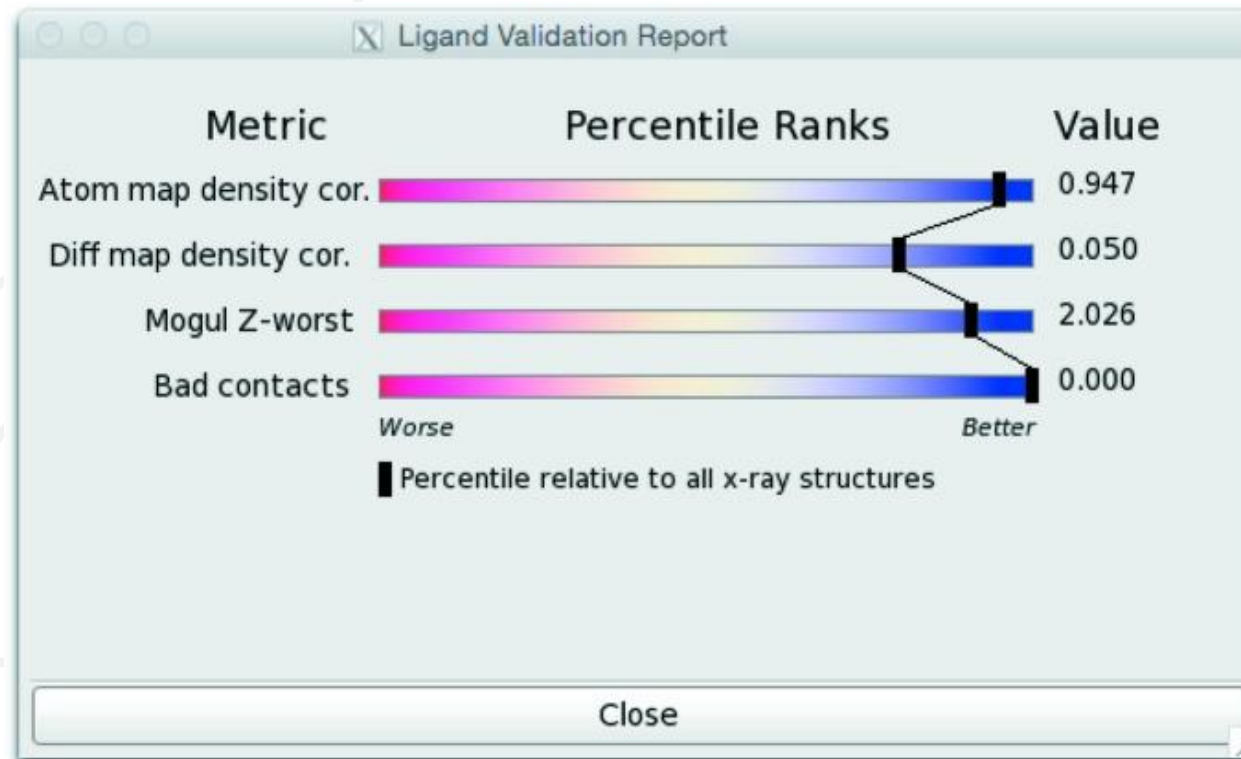


<https://doi.org/10.1107/S0907444912044423>
<https://doi.org/10.1021/ml500220a>
<https://doi.org/10.1517/17460441.2011.585154>



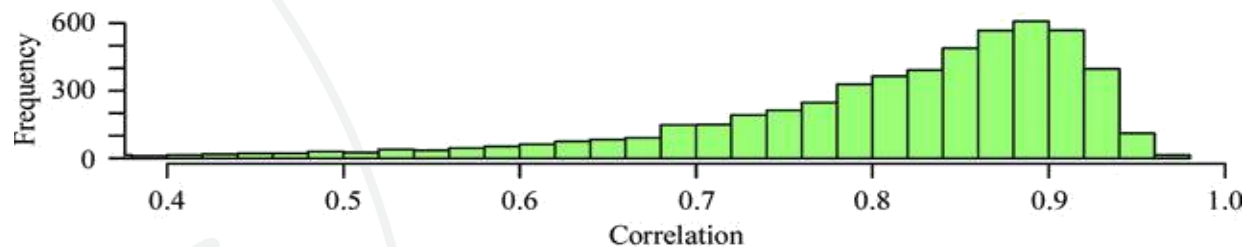
Ligand validation in Coot

- Ligand Metric Sliders

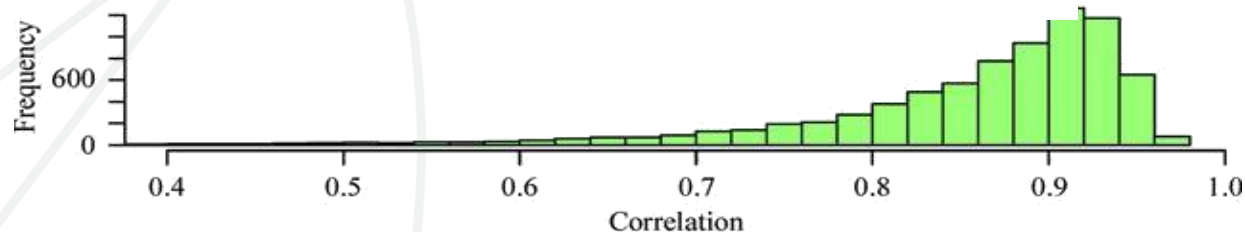


Ligand validation - density correlation

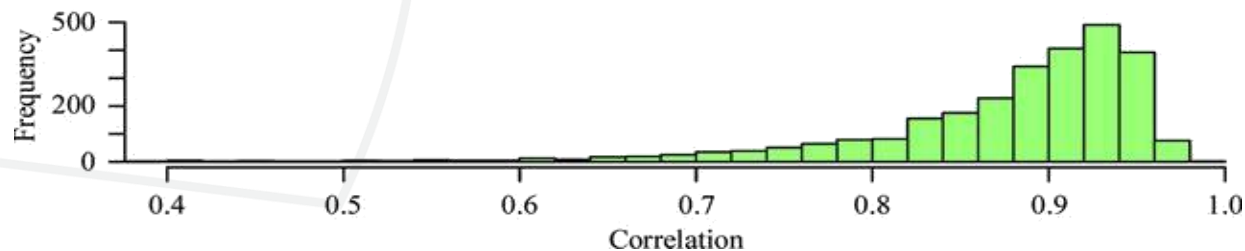
Low-2.3 Å resolution



2.3 – 1.7 Å resolution

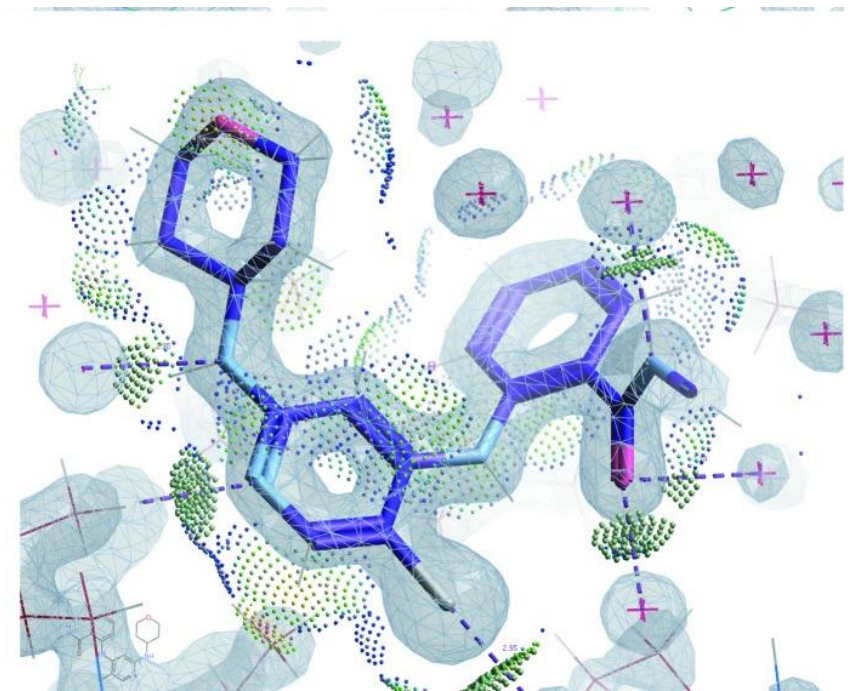


1.7-1.2 Å resolution



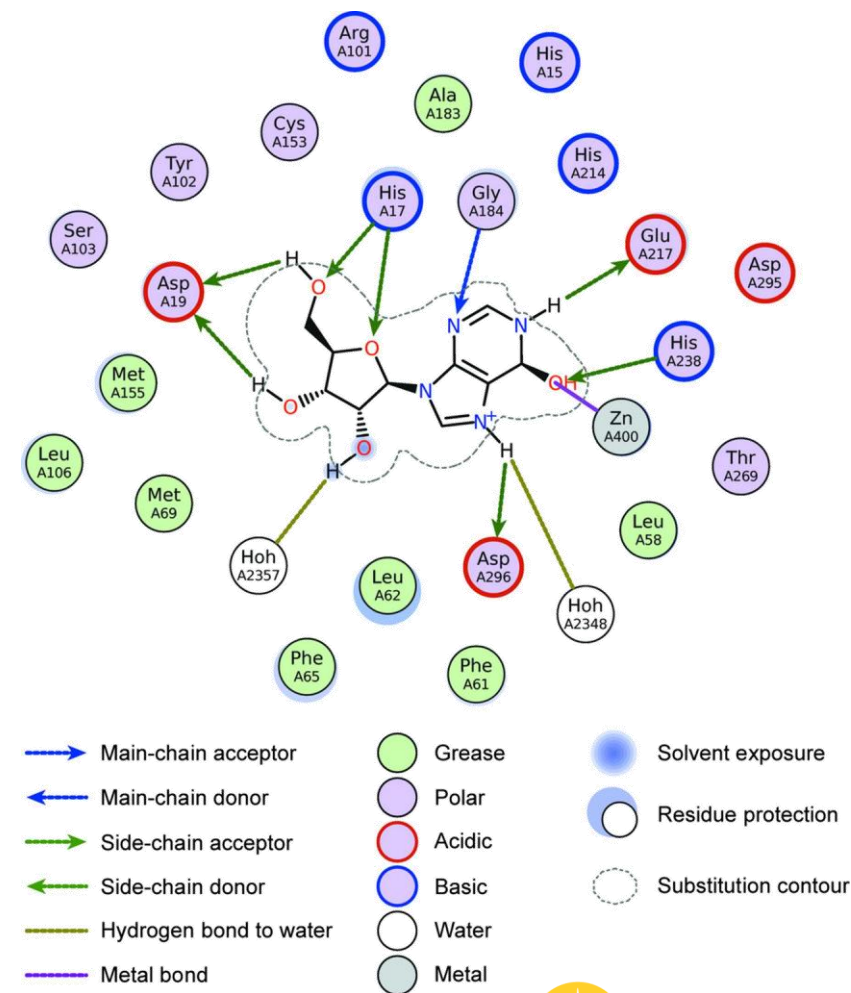
Ligand validation - interactions and clashes

- Check molecular interactions
 - Measures -> Environment Distances -> Show Residues Environments
 - Show H-bonds and other interactions
- Bad contacts
 - Ligand -> Isolated Molprobity dots for this ligand
 - Runs *Reduce*, then *Probe*
 - Shows clashes as coloured dots



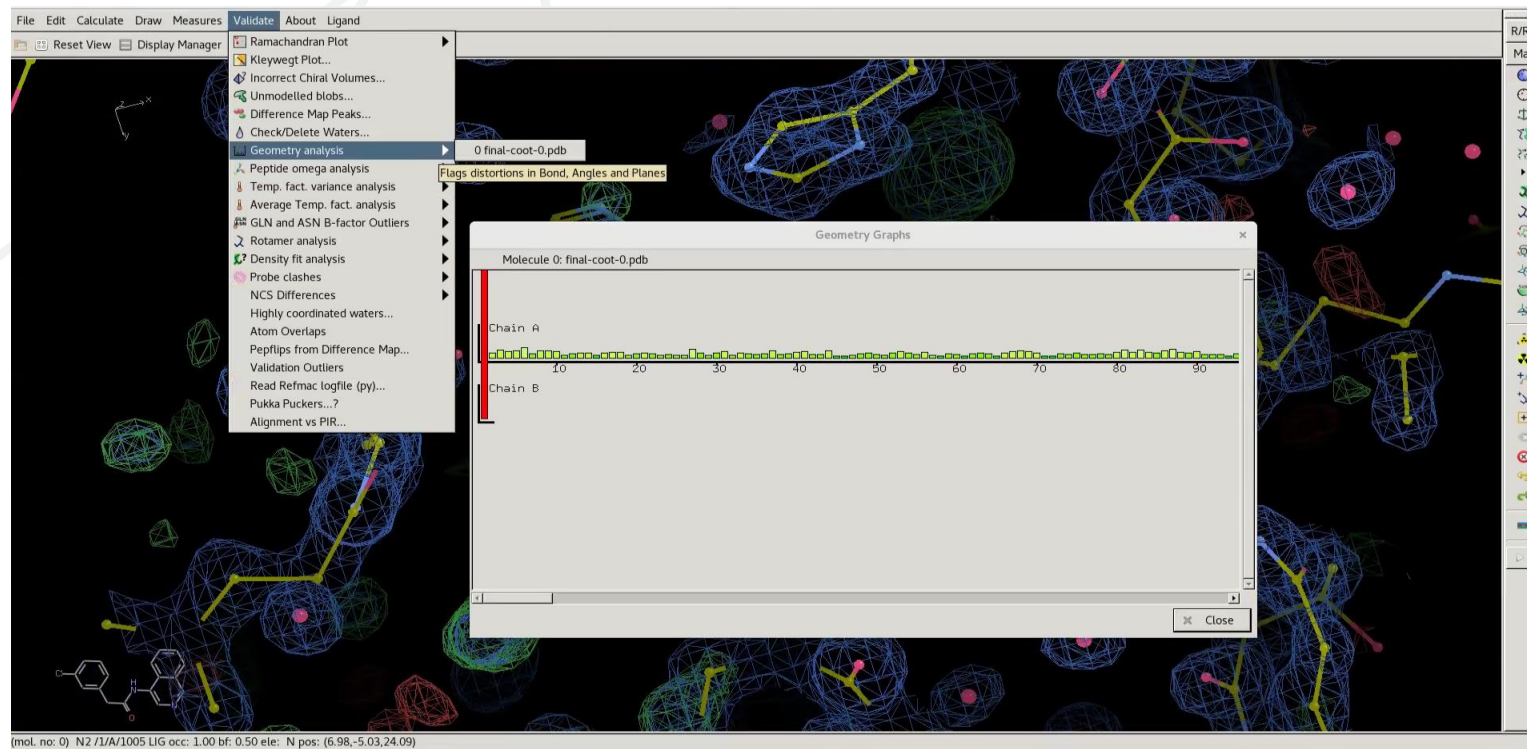
Ligand validation – Ligand environment (FLEV)

- Flatland Ligand Environment View
 - Ligand -> Toggle FLEV Ligand Interactions
- 2D representation of ligand environment
 - Binding pocket residues
 - Intermolecular interactions
 - Solvent accessible regions
 - Substitution contour



Ligand validation - geometry

- Compare modelled structure to restraints
 - Validate -> Geometry analysis
 - Only shows agreement, doesn't tell you if errors in restraints



Ligand validation - geometry

- Quantum Mechanical Methods
 - CPU intensive
 - Lowest energy conformation may not represent bound-state (carried out in vacuum)
- Comparison with PDB ligands:
 - May be useful for common ligands (nucleotides, cofactors, amino acids)
 - Less useful for novel ligands – few examples exist, if any
 - Need to confirm quality of PDB entry first

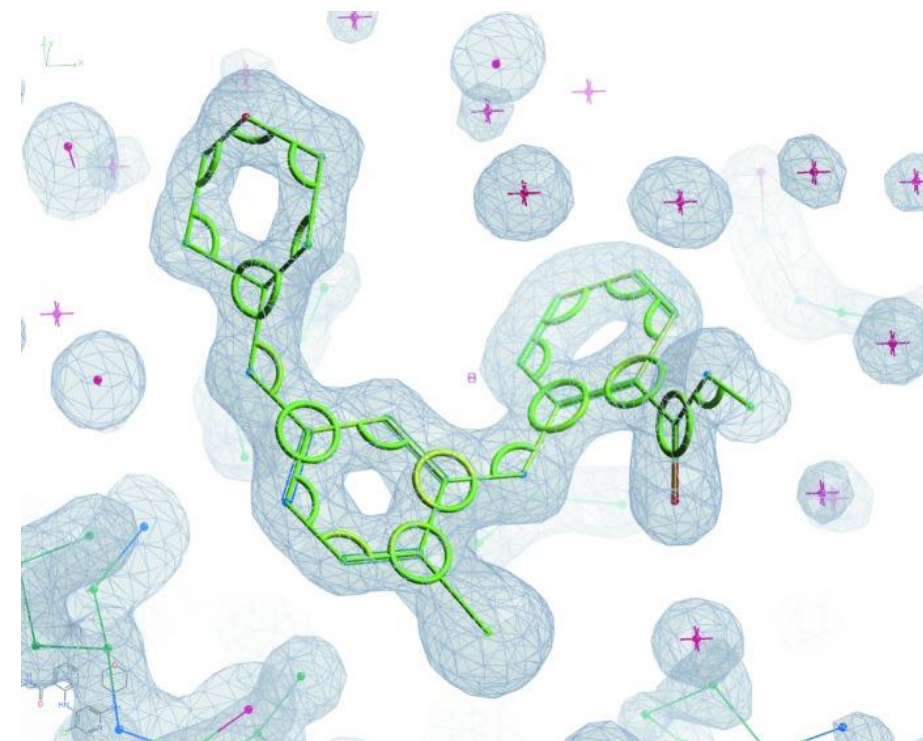
Ligand validation - geometry

- Mogul (CCDC)
 - Compares geometries against CSD - database of small-molecule crystal structures
 - Can be run as Coot plugin:
 - Ligand -> Display Ligand Distortions
 - Shows how well the bond lengths and angles correspond to expect values (Z-score)
 - Tabulated values also be provided in Coot
 - Can also be run as stand alone programme
 - Useful cross-validation if restraints not generated using CSD data (e.g. AceDRG)

<https://doi.org/10.1107/S2059798317003382>

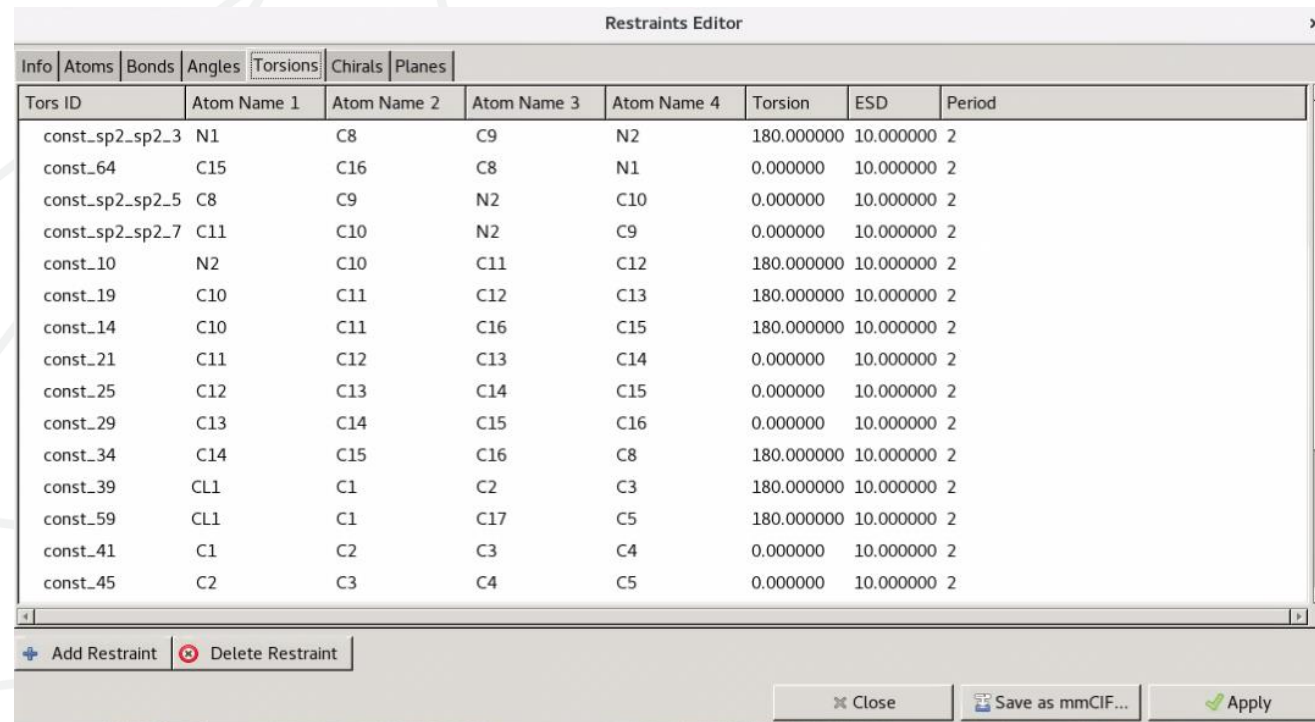
<https://doi.org/10.1021/ci049780b>

<https://doi.org/10.1021/ci200439d>



Ligand validation – sources of errors

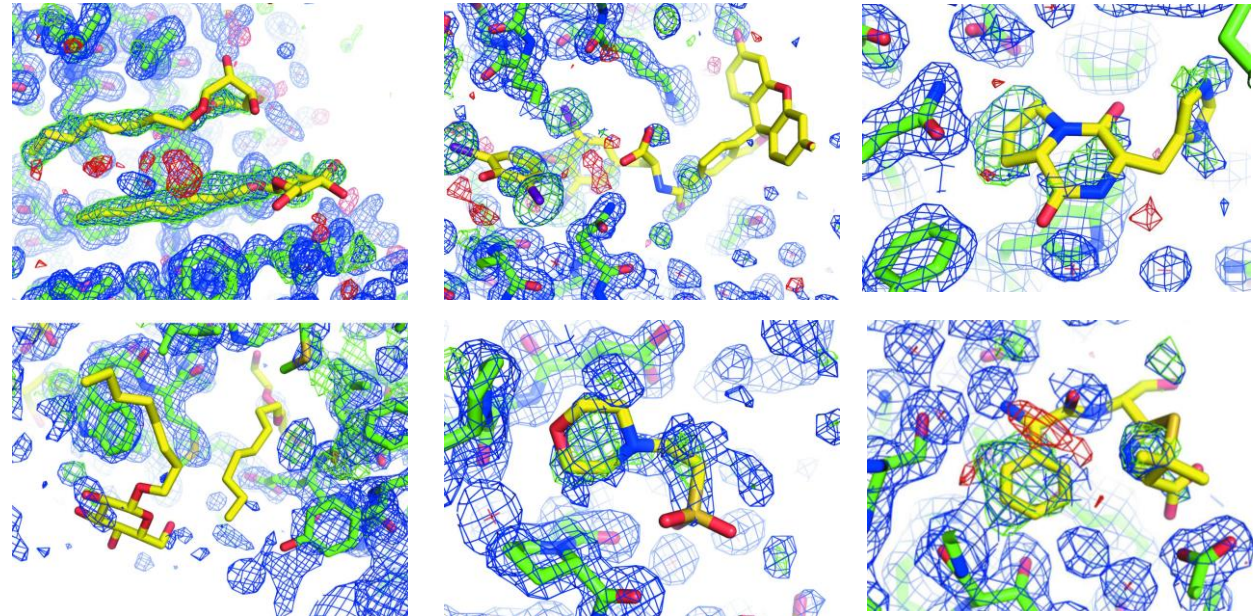
- Poor ligand restraints
 - Use geometry validation to identify regions for improvement
 - Can edit in Coot with Restraints Editor



Tors ID	Atom Name 1	Atom Name 2	Atom Name 3	Atom Name 4	Torsion	ESD	Period
const_sp2_sp2_3	N1	C8	C9	N2	180.000000	10.000000	2
const_64	C15	C16	C8	N1	0.000000	10.000000	2
const_sp2_sp2_5	C8	C9	N2	C10	0.000000	10.000000	2
const_sp2_sp2_7	C11	C10	N2	C9	0.000000	10.000000	2
const_10	N2	C10	C11	C12	180.000000	10.000000	2
const_19	C10	C11	C12	C13	180.000000	10.000000	2
const_14	C10	C11	C16	C15	180.000000	10.000000	2
const_21	C11	C12	C13	C14	0.000000	10.000000	2
const_25	C12	C13	C14	C15	0.000000	10.000000	2
const_29	C13	C14	C15	C16	0.000000	10.000000	2
const_34	C14	C15	C16	C8	180.000000	10.000000	2
const_39	CL1	C1	C2	C3	180.000000	10.000000	2
const_59	CL1	C1	C17	C5	180.000000	10.000000	2
const_41	C1	C2	C3	C4	0.000000	10.000000	2
const_45	C2	C3	C4	C5	0.000000	10.000000	2

Ligand fitting – common sources of errors

- Map interpretation is subjective and commonly encountered problems can make unambiguous ligand placement challenging
 - Poor quality/low resolution data
 - Incomplete density
 - Flexible ligands
 - Partial occupancy
 - Multiple ligand conformations
 - Wrong ligands
- Important to validate structure!



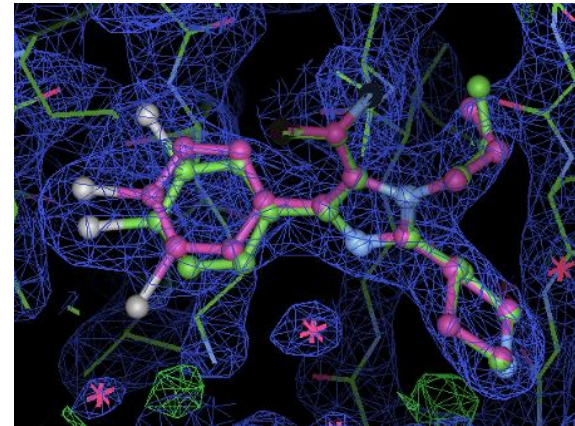
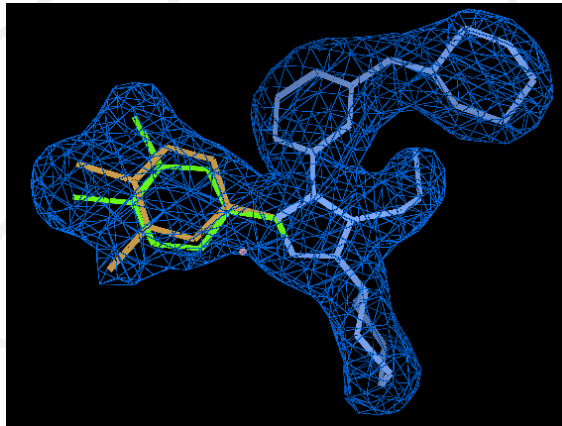
<https://doi.org/10.1107/S0907444912044423>

<https://doi.org/10.1021/ml500220a>

<https://doi.org/10.1517/17460441.2011.585154>

Alternate ligand conformations

- Some ligands capable of binding in multiple conformations
- Add Alternative Conformation to a Residue
 - Refine occupancy/B-factors
 - If ligands overlap, total occupancy should not exceed 1



Heavy Atoms

- Presence of heavy atoms can help with ligand identification and placement
- Fraglites - exploit the anomalous scattering of a halogen substituent

Journal of
**Medicinal
Chemistry**

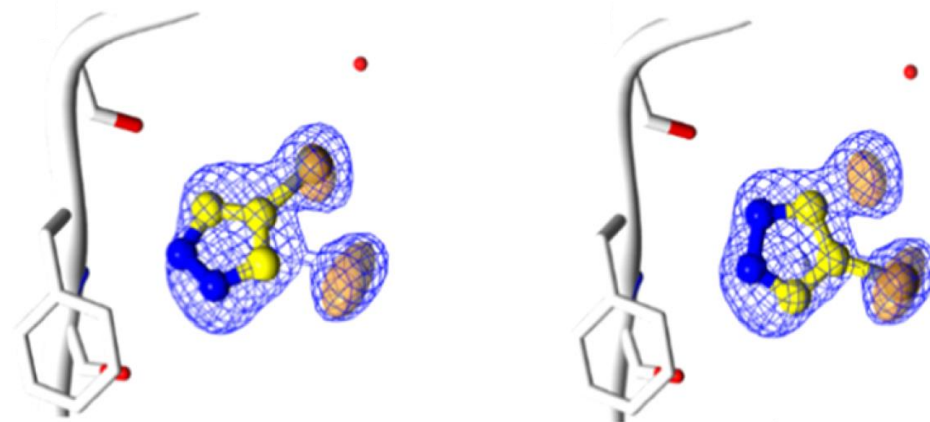
Cite This: *J. Med. Chem.* 2019, 62, 3741–3752

Article

pubs.acs.org/jmc

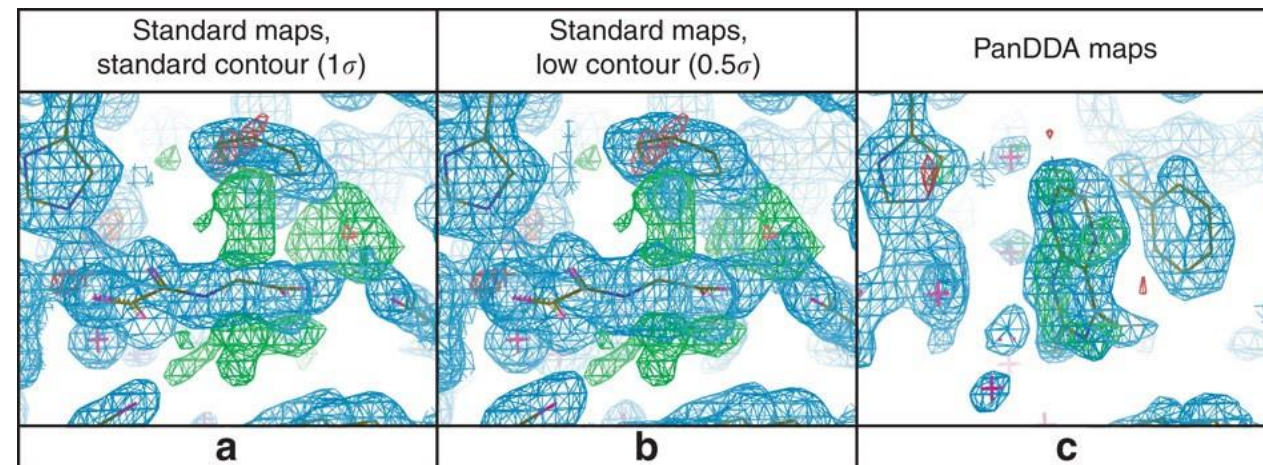
FragLites—Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation

Daniel J. Wood,[†] J. Daniel Lopez-Fernandez,[‡] Leanne E. Knight,[‡] Islam Al-Khawaldeh,[‡] Conghao Gai,[‡] Shengying Lin,[‡] Mathew P. Martin,[†] Duncan C. Miller,[‡] Céline Cano,[‡] Jane A. Endicott,[†] Ian R. Hardcastle,[‡] Martin E. M. Noble,^{*,†} and Michael J. Waring^{*,‡}



Low occupancy ligands

- Poor correlation with difference maps
- B-factor significantly higher for ligand than protein when modelled at full occupancy
- Adjust ligand occupancy and refine
 - Edit -> Residue info
- Particularly common with fragments
 - Use specialist software (PanDDA)
 - Need large number of datasets

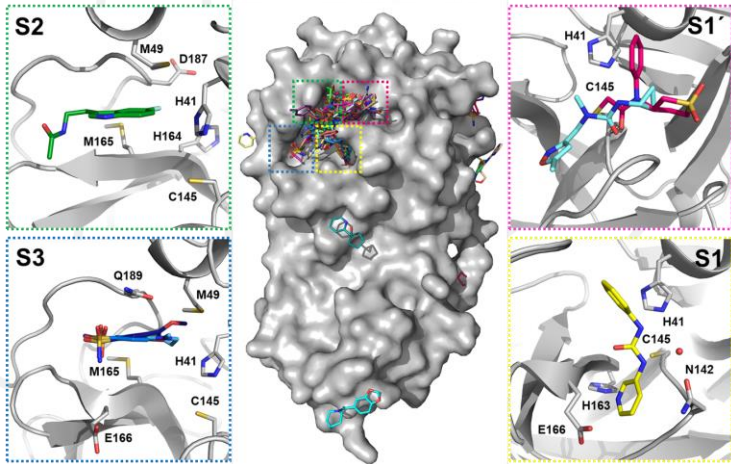


Refinement

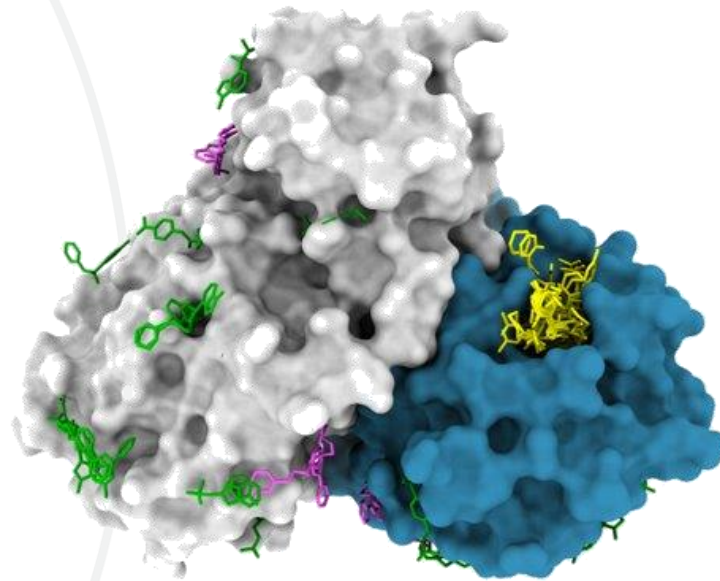
- Following ligand fitting, full refinement should be carried out with ligand restraints you generated
 - REFMAC/Buster/Phenix
- Allows protein and ligand to co-refine, optimizing agreement between your model and the experimental data
- As ligand now contributes to model phases, further density interpretation should be done with caution
 - Refer to omit maps

Analysis and validation

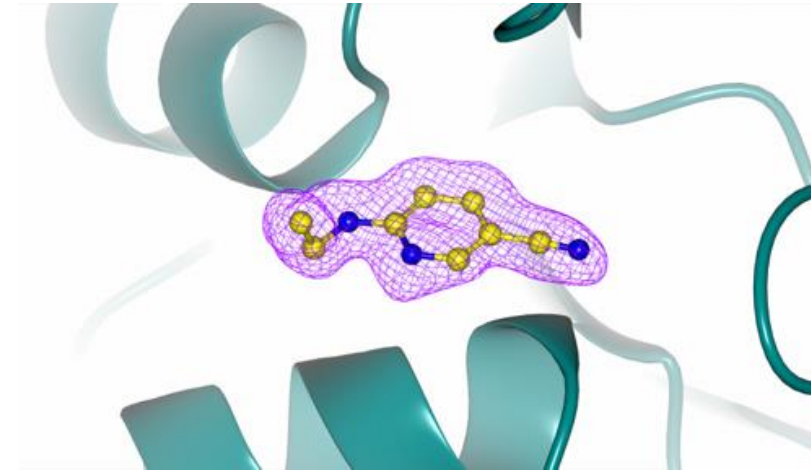
- Can use Coot to display surfaces etc but better tools for visualisation available



Pymol



Chimera

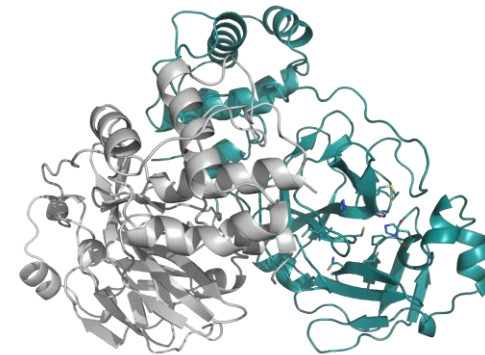


CCP4MG

Other useful Coot tools

- Inverting chiral centre
 - Calculate -> Modelling -> Invert This Chiral Centre
- Aligning ligands
 - Calculate -> Modelling -> Superpose ligands
- Renaming/renumbering residues
 - Edit -> Change Chain IDs
 - Edit -> Renumber Residues
- Display chemical features
 - Ligand -> Show chemical features
- And more...

Tutorial – Introduction



- SARS-CoV-2 Main protease (Mpro) is one of two cysteine proteases that are essential for viral replication, making it an attractive drug discovery target for the treatment of COVID-19.
- We have provided X-ray diffraction data collected from crystals soaked with inhibitors from the COVID Moonshot – a crowdsourced open drug discovery initiative
- In this tutorial we will identify the ligand binding site, generate ligand restraints, model the ligand, refine and validate our model
- Files:
 - DIMPLE output (pdb/mtz)
 - Reflection data (mtz)
 - SMILES string

Tutorial - Part 1: Generating restraints

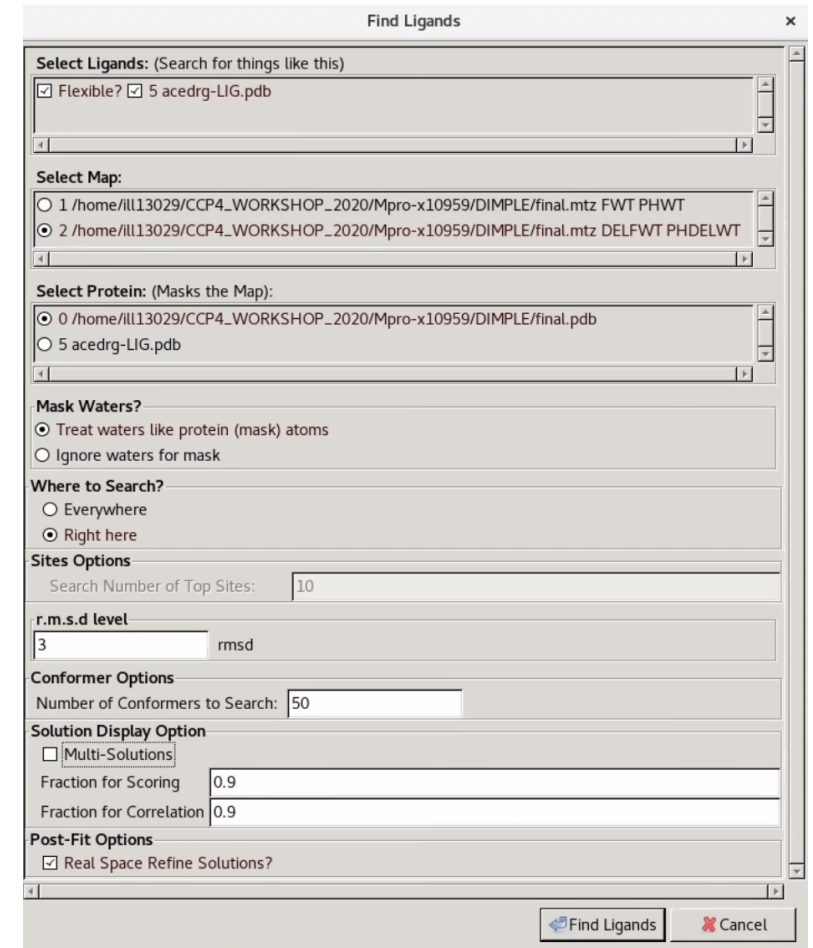
- Select a dataset Mpro_1/2/3 (ignore 4 for now)
- Start Coot and load protein model and reflection data
 - Open Coordinates: dimple.pdb
 - Auto open MTZ: dimple.mtz
- We will begin by using AceDRG via Coot to generate the coordinates and description for the ligand
 - Select SMILES -> 2D from Ligand menu
 - Enter provided SMILES string and click “Send to 2D viewer”
 - Check structure in Lidia then click “Apply”
 - This will generate the ligand coordinates and restraints file
 - acedrg-LIG.cif and acedrg-LIG.pdb

Tutorial – Part 2: Identify binding site

- Validate -> Unmodelled blobs...
- Select difference map and protein model
- Click “Find blobs” and search through identified blobs
- Did you find a blob near Cys145?

Tutorial – Part 3: Fitting ligand

- We now have the ligand restraints and know where to place the ligand
- *Run Ligand -> Find Ligands*
 - Select Ligand (Flexible), difference map and protein model
 - Search “Right here” with r.m.s.d. level set to 3
 - Click Find Ligands – do you find a hit?
- Use real-space refinement/Jiggle fit to optimize ligand fit and then merge molecules
- Tidy up any surrounding residues/solvent molecules



The screenshot shows the 'Find Ligands' dialog box with the following settings:

- Select Ligands:** (Search for things like this)
 - ☒ Flexible? ☒ 5 acedrg-LIG.pdb
- Select Map:**
 - ☐ 1 /home/ill13029/CCP4_WORKSHOP_2020/Mpro-x10959/DIMPLE/final.mtz FWT PHWT
 - ☒ 2 /home/ill13029/CCP4_WORKSHOP_2020/Mpro-x10959/DIMPLE/final.mtz DELFWT PHDELWT
- Select Protein:** (Masks the Map):
 - ☒ 0 /home/ill13029/CCP4_WORKSHOP_2020/Mpro-x10959/DIMPLE/final.pdb
 - ☐ 5 acedrg-LIG.pdb
- Mask Waters?**
 - ☒ Treat waters like protein (mask) atoms
 - ☐ Ignore waters for mask
- Where to Search?**
 - ☐ Everywhere
 - ☒ Right here
- Sites Options**
 - Search Number of Top Sites: 10
- r.m.s.d level**
 - 3 rmsd
- Conformer Options**
 - Number of Conformers to Search: 50
- Solution Display Option**
 - ☐ Multi-Solutions
 - Fraction for Scoring: 0.9
 - Fraction for Correlation: 0.9
- Post-Fit Options**
 - ☒ Real Space Refine Solutions?

Buttons at the bottom: Find Ligands, Cancel

Tutorial – Part 4: Validation

- Measures -> Environment Distances -> Show Residue Environment?
 - Are molecular interactions sensible?
- Ligand -> FLEV this residue -> Show Env. Residues
- Ligand -> Isolated Molprobit dots for this ligand
 - Do you see any significant clashes?
- Ligand -> Display Ligand Distortions
 - How does the geometry look?

Tutorial – Part 5: Refinement and validation

- Take saved model and refine using favoured software package externally or refine in Coot using REFMAC
 - Use Mpro_*.free.mtz, your merged molecule (pdb) and the generated cif file for refinement
- When refinement is complete, inspect ligand and surrounding residues/solvent molecules again
 - Good fit to density?
 - Sensible interactions with surrounding residues?
 - Any clashes?
- Fix any problems, tidy up structure if necessary and repeat refinement



Tutorial – Part 6

- Dataset Mpro_4 contains DIMPLE files and SMILES strings for 3 ligands
- Using the methods described previously, identify which ligand is bound to the protein

