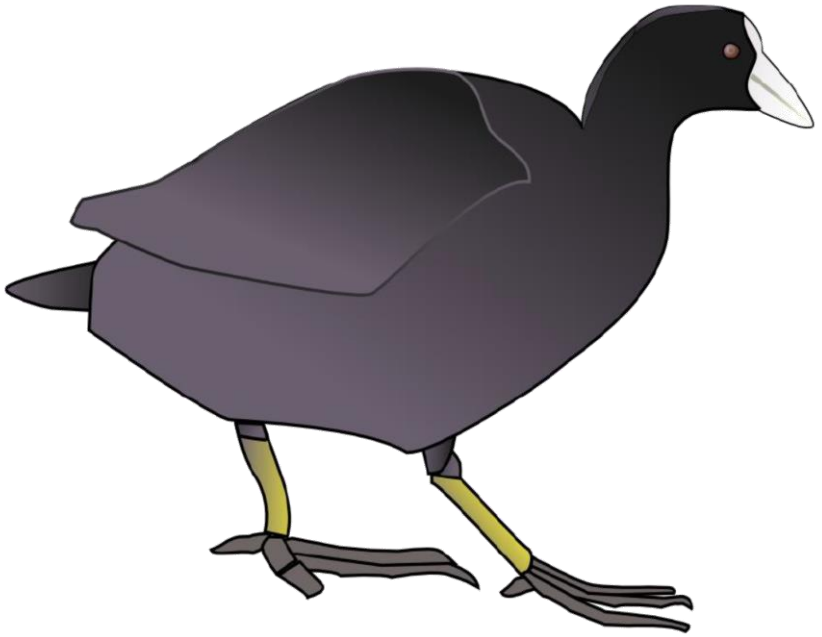
Ligand fitting with Coot

Daren Fearon
Senior Beamline Scientist XChem
Diamond Light Source

Daren.Fearon@diamond.ac.uk

What is Coot?



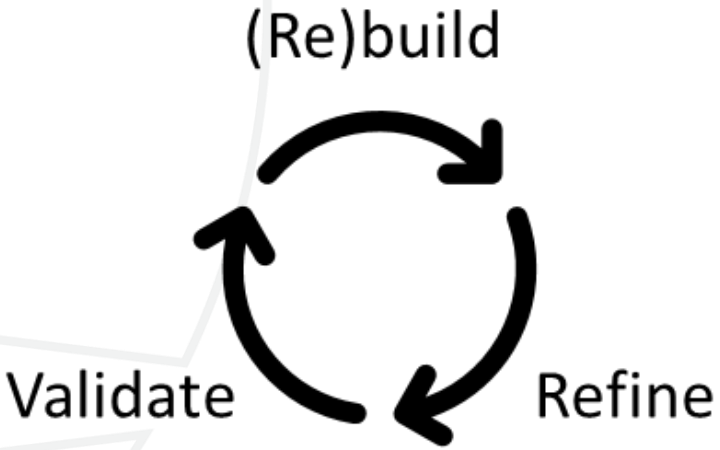
- Crystallographic Object-Orientated Toolkit
- Developed by Paul Emsley and others
- Interface for model building and validation
- Integrates with other programmes
  - REFMAC/Phenix/Buster, AceDRG/Grade/eLBOW, Molprobity/Mogul
- Critical part of crystallography toolkit

https://doi.org/10.1107/S0907444904019158

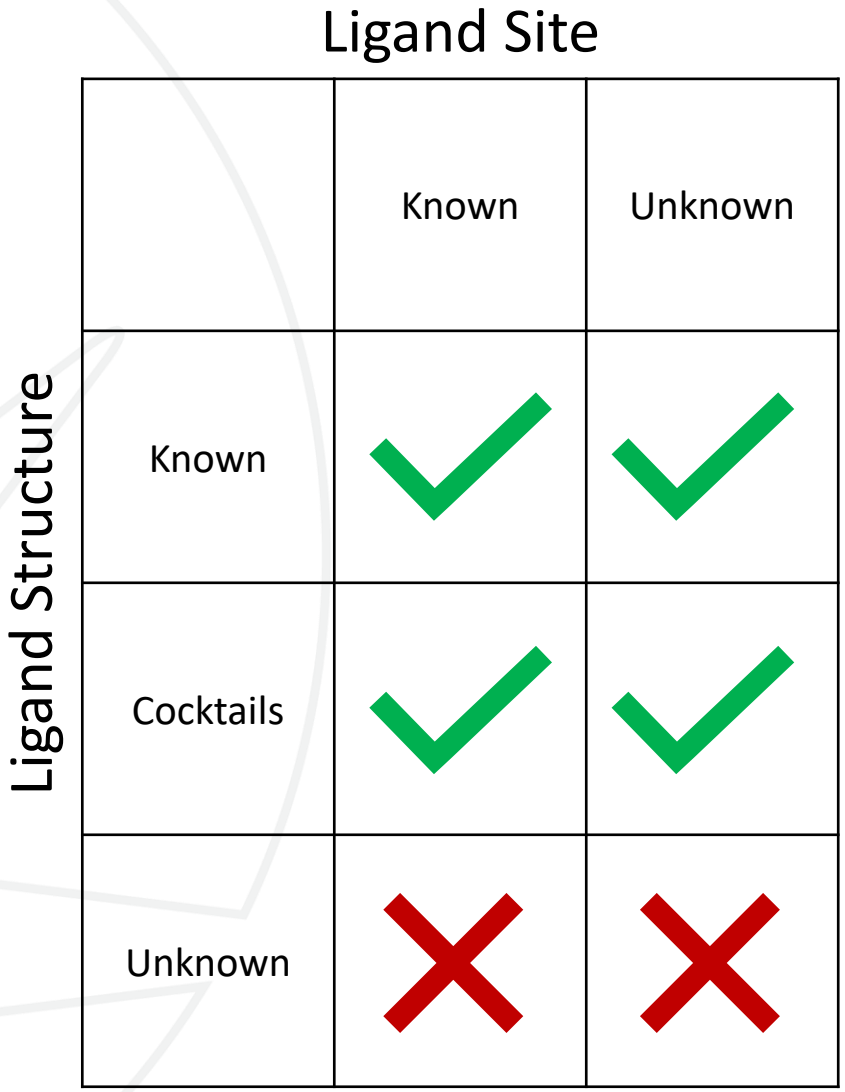
diamond

Ligand fitting aims

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

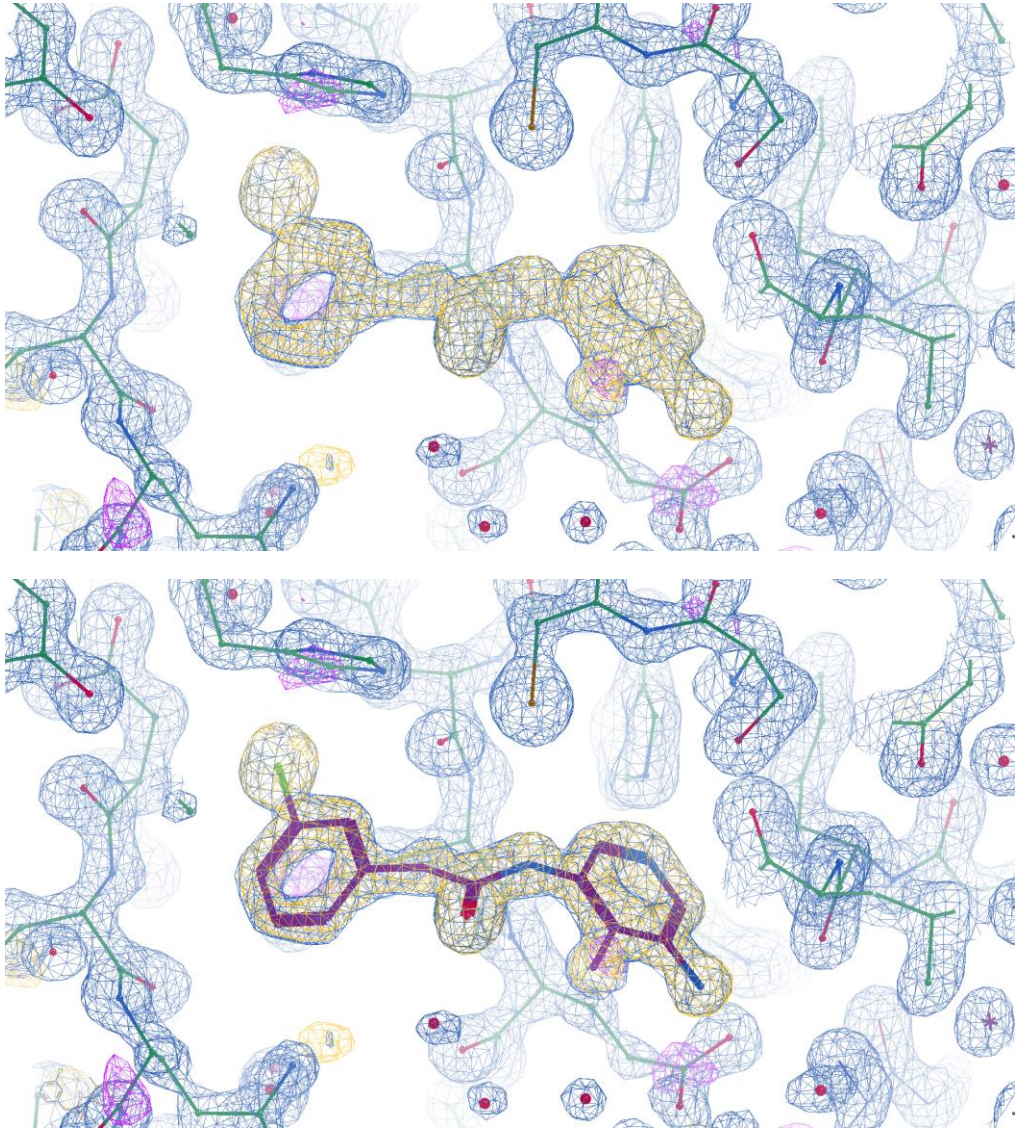


Ligand fitting scenarios with Coot



Ligand fitting process

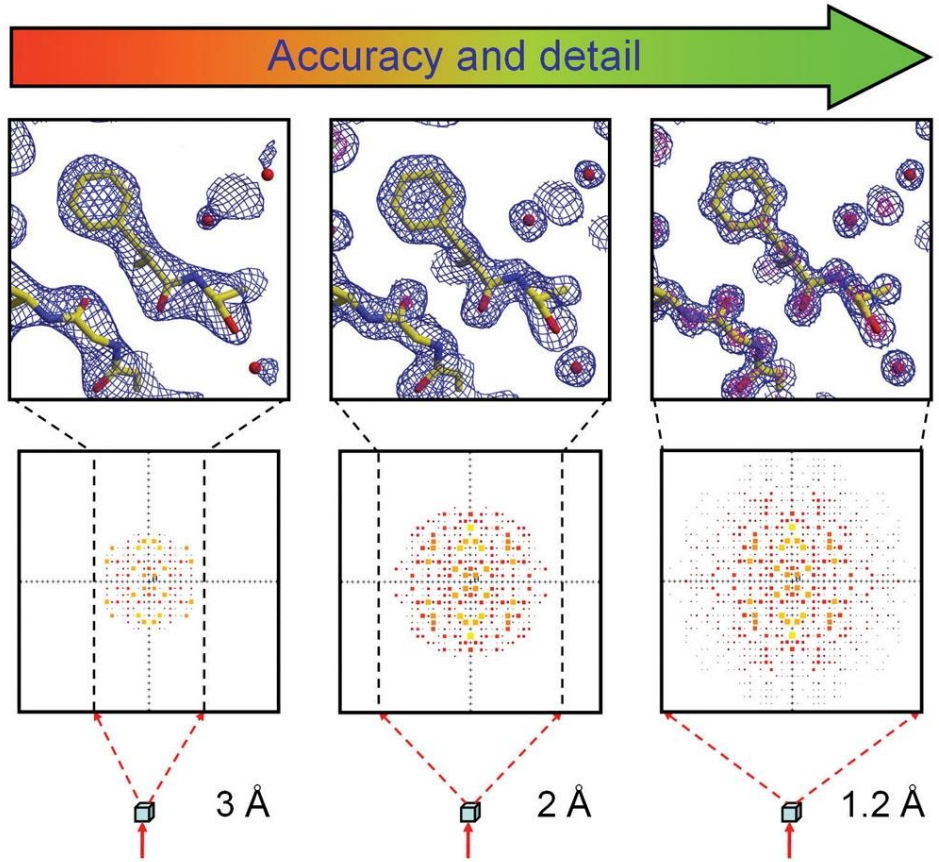
- Identification of ligand electron density
- Generation of ligand restraints
- Ligand fitting
- Optimisation of ligand conformation
- Model refinement
- Validation of model
- Analysis and visualisation



https://doi.org/10.1107/S2059798316020143

Ligand fitting assumptions

- Data is of sufficient quality to model ligand
  - Ideally 2.5 Å or better
- A good starting model is available for the protein
- Ligand is known
  - Description (2D structure/SMILES string) available or ligand is contained in ligand dictionary (e.g. ATP)
- Ligand is actually bound

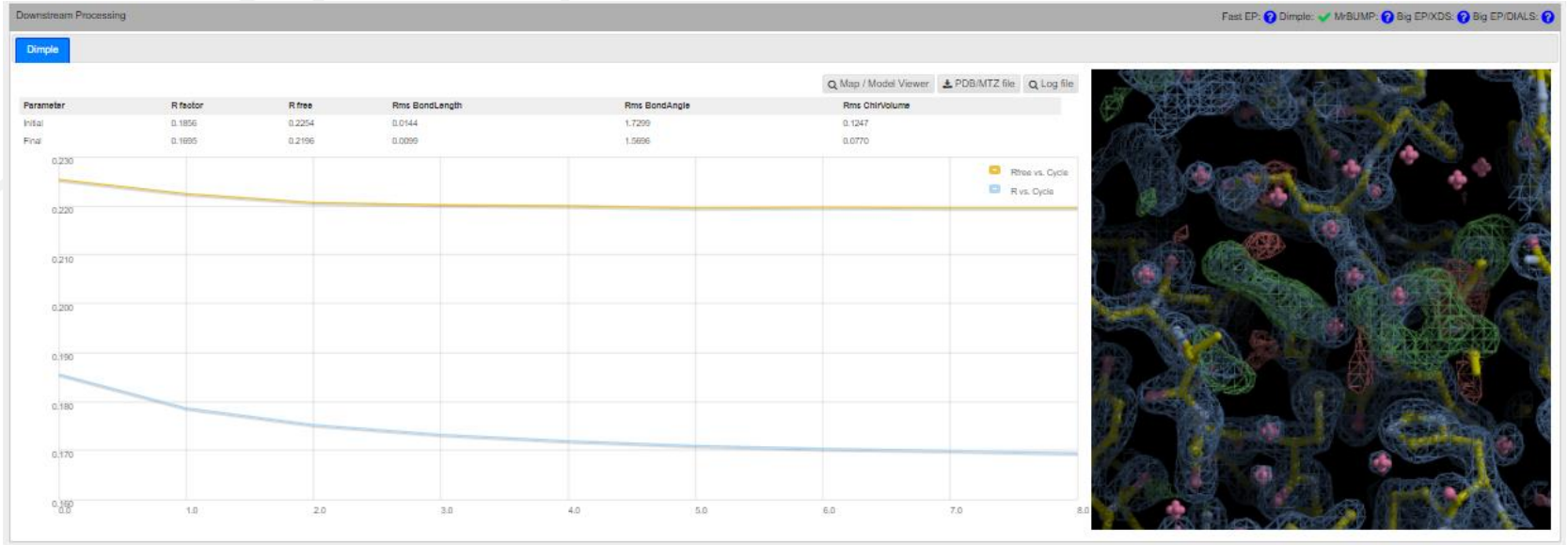


Reproduced from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

https://doi.org/10.1007/s10822-015-9833-8

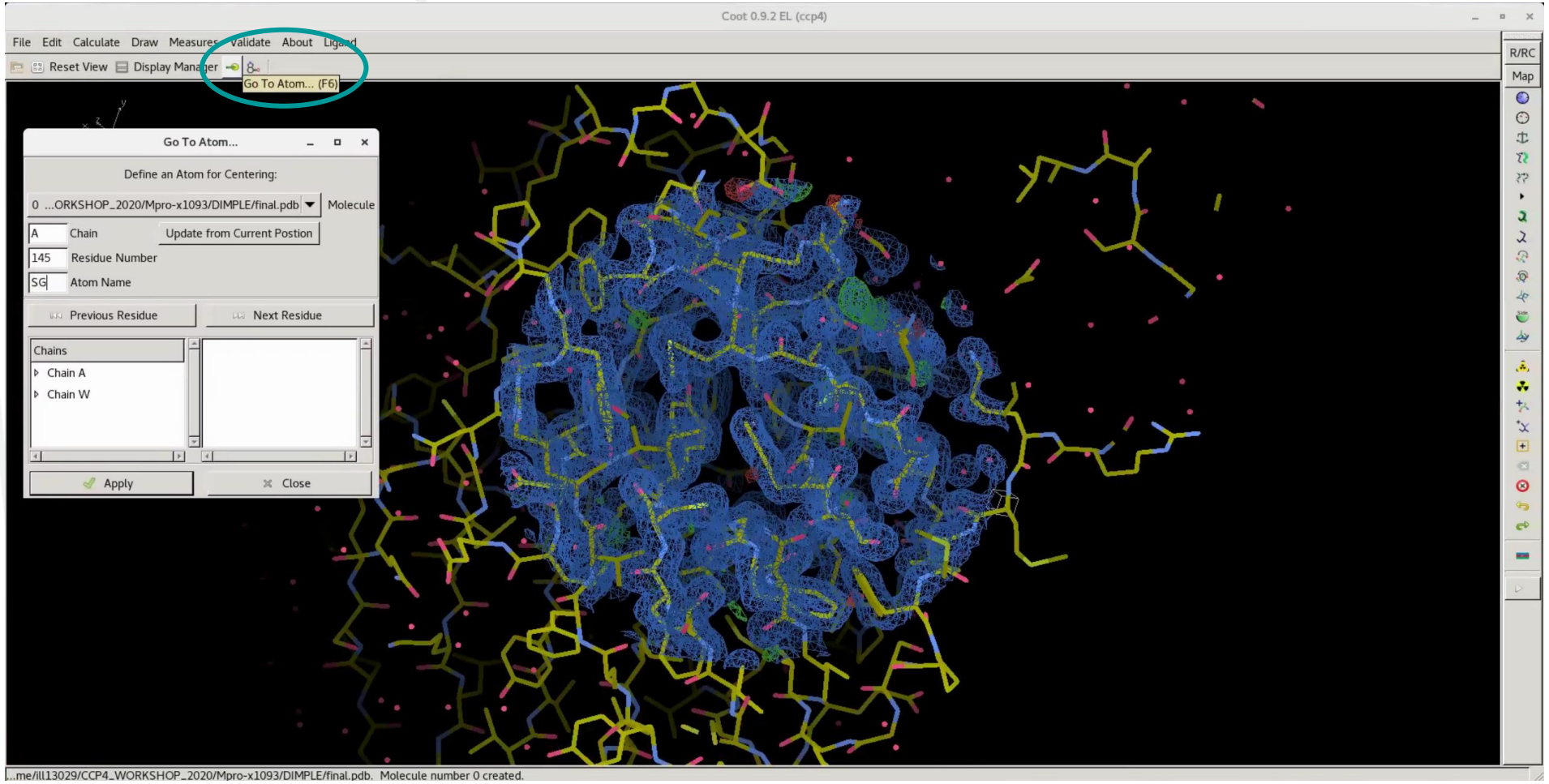
Identifying ligand density

- 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)

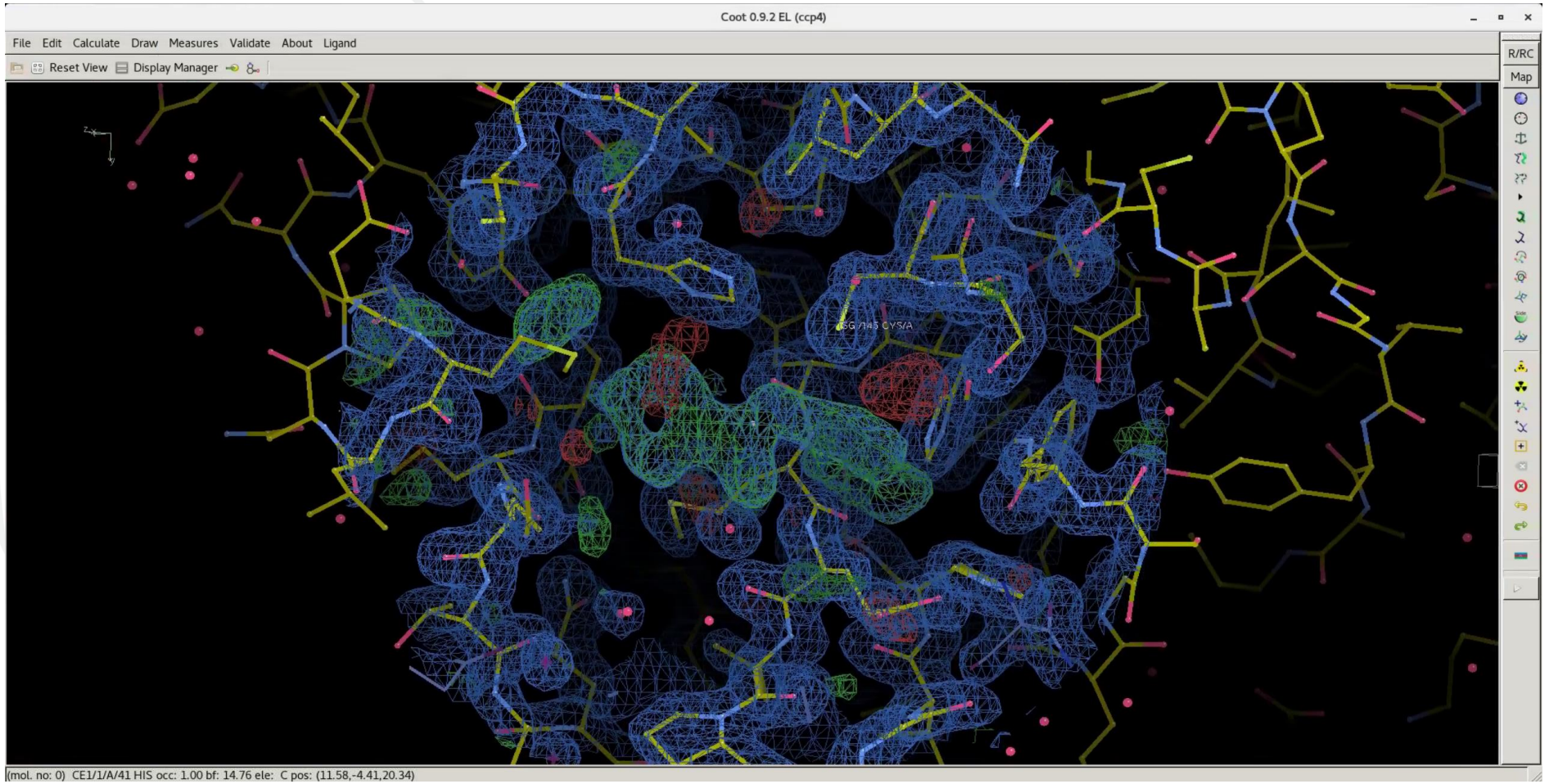


http://ccp4.github.io/dimple/

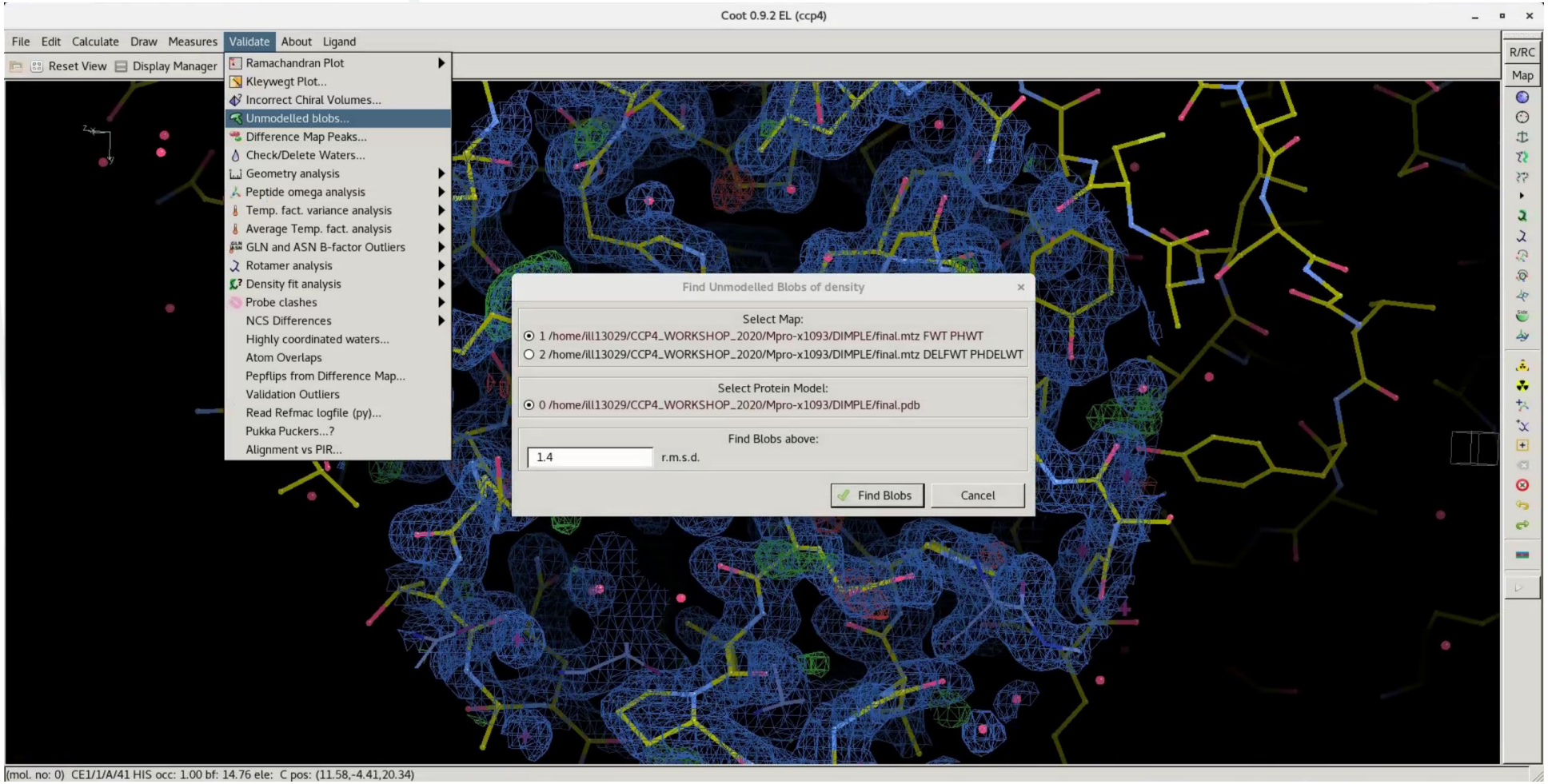
Identifying ligand density – known ligand site



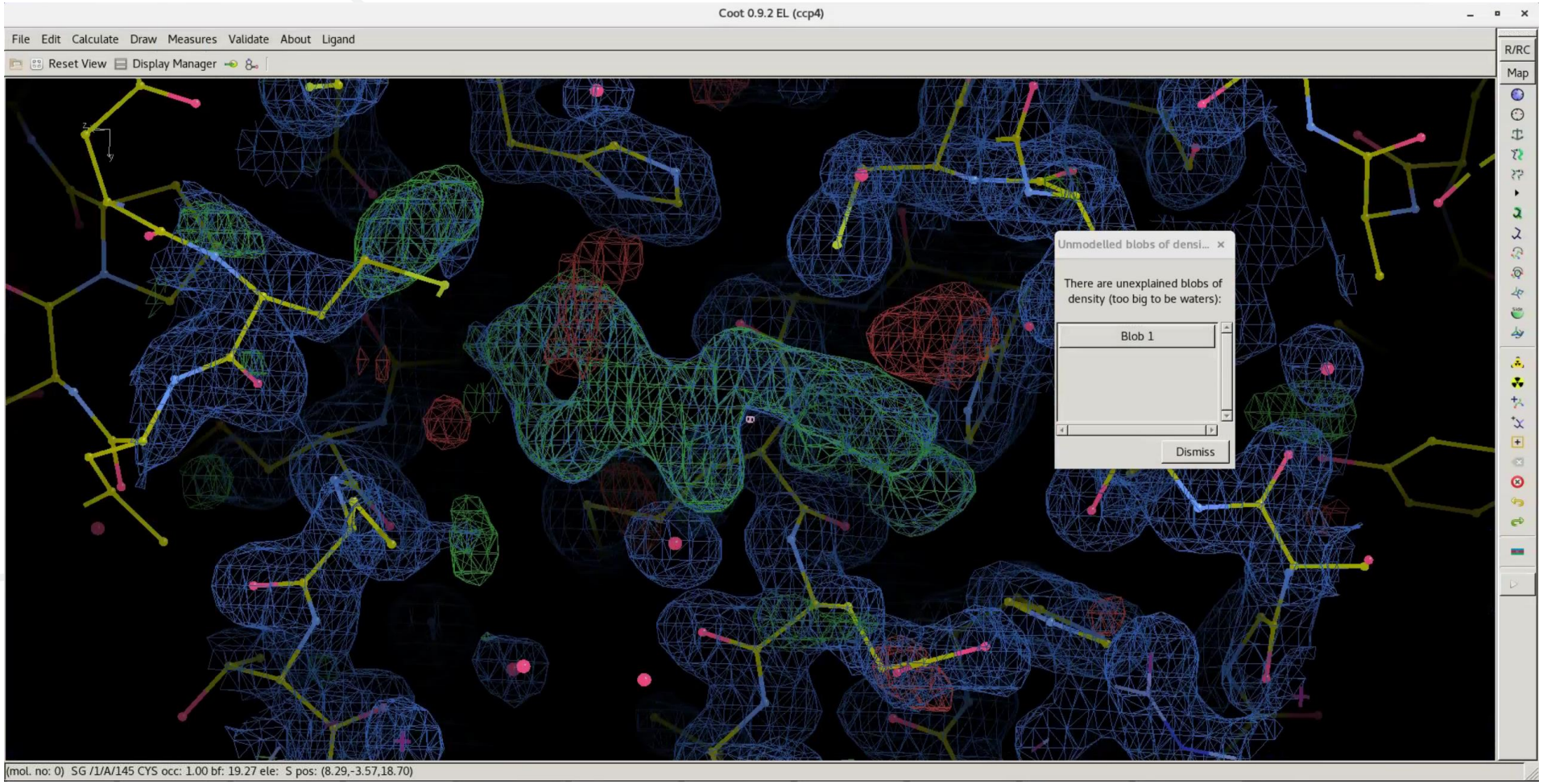
Identifying ligand density – known ligand site



Identifying ligand density – unknown site



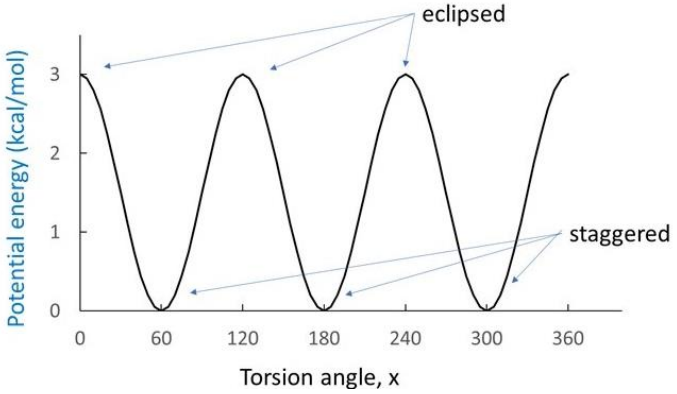
Identifying ligand density – unknown site



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 the relative positions of atoms and their connectivity plus 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
```

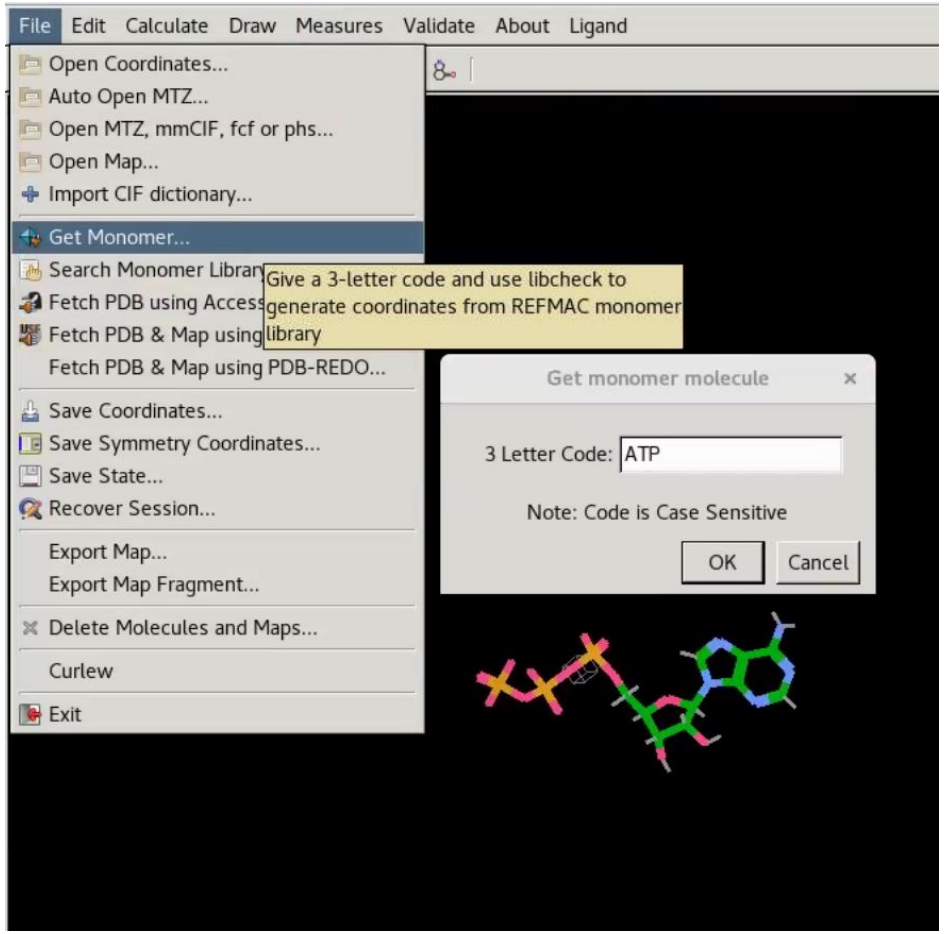


https://doi.org/10.1107/S2059798316017964
https://doi.org/10.1107/S2059798316020143

diamond

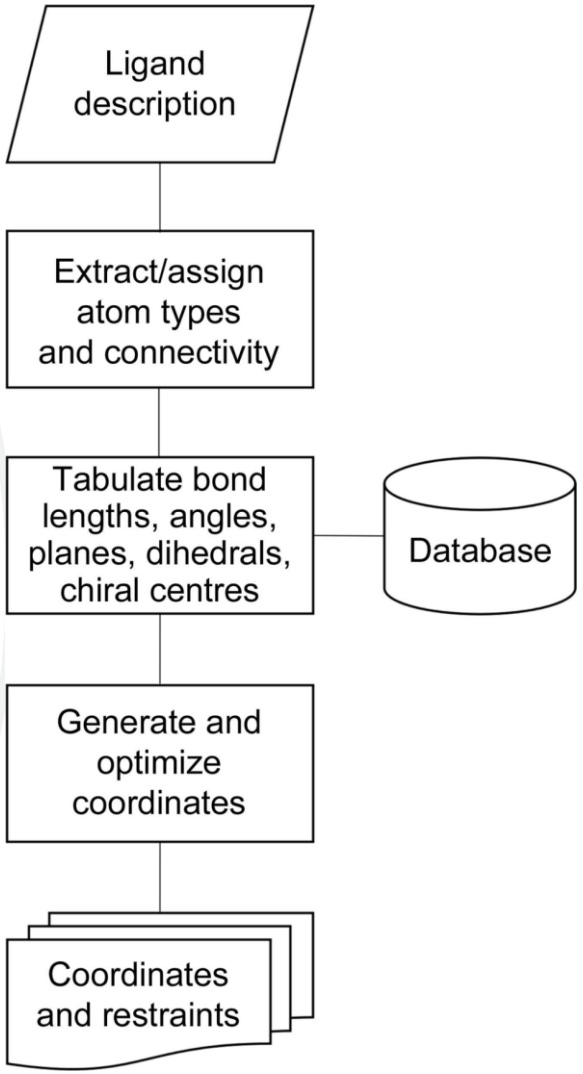
Fitting common ligands

- 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, DMS



https://doi.org/10.1107/S0907444904023510
http://legacy.ccp4.ac.uk/html/lib_list.html

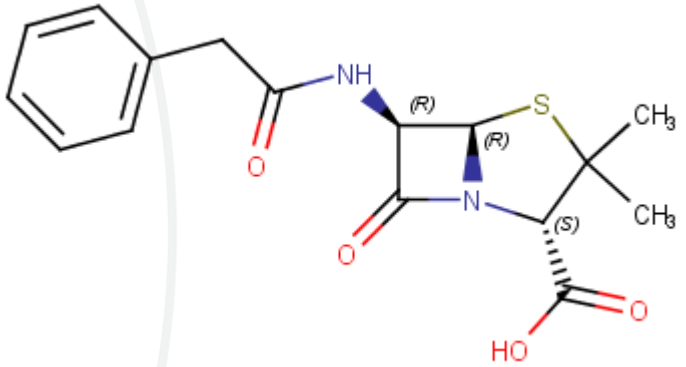
Generating dictionary for novel ligands



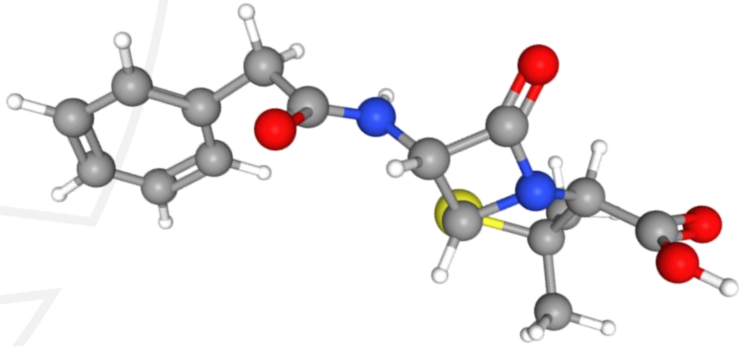
https://doi.org/10.1107/S2059798316017964

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

https://doi.org/10.1107/S2059798316017964

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

https://doi.org/10.1107/S2059798316017964

Generating and optimizing coordinates

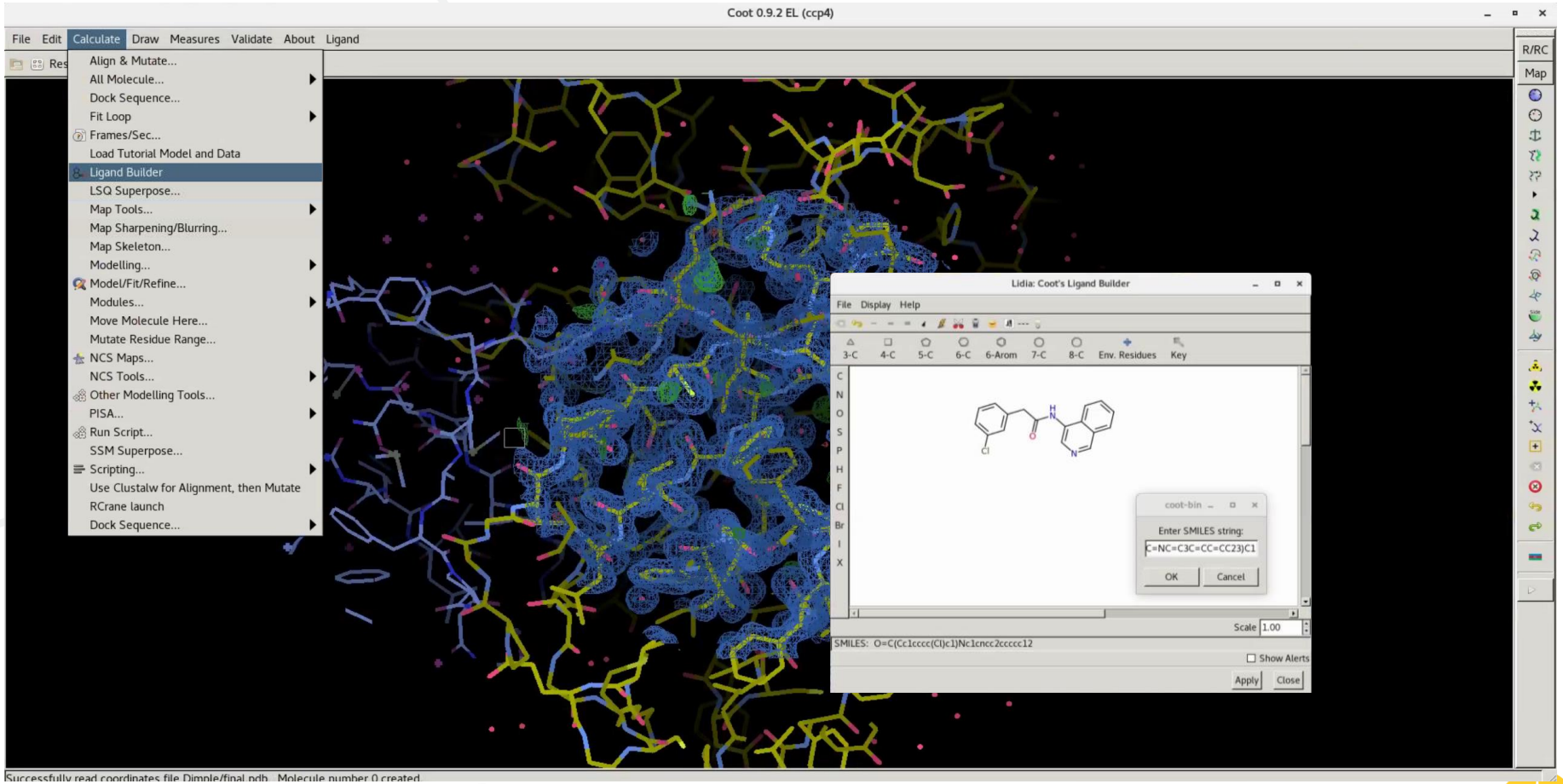
- Various programs available, mostly free for academic users

Program name	AceDRG	Grade	eLBOW	Pyrogen
Distributor	CCP4	Global Phasing	PHENIX	CCP4
Input formats	SMILES, PDB, CIF	SMILES, Molfile, CIF	SMILES, PDB, CIF	SMILES, CIF, sketch
Output formats	PDB, CIF	PDB, CIF, SHELX	Multiple, including PDB, CIF	PDB, CIF
Experimental data source(s)	COD (curated)	CSD	CSD	CSD, ener_lib.cif
Force field(s)	None	AM1/RM1/PM3	Multiple including AM1, MMFF94	MMFF94
Standard deviation source(s)	COD (curated)	CSD	Multiple including CSD	CSD
Restraints editor	JLigand	Edit REFMAC	REEL	Coot restraints editor
Other features and limitations	Hierarchical atom typing	Flexible planar definitions. Available through web server.	Atom name preservation. Metal coordination.	Atom name preservations. Tautomer enumeration.
Citation	Long et al. (2017)	Smart et al. (2011)	Moriarty et al. (2009)	Debreczeni & Emsley (2012), Emsley & Debreczeni (2012)

- Dictionaries serve as starting points for refinement and model building

https://doi.org/10.1107/S2059798316017964

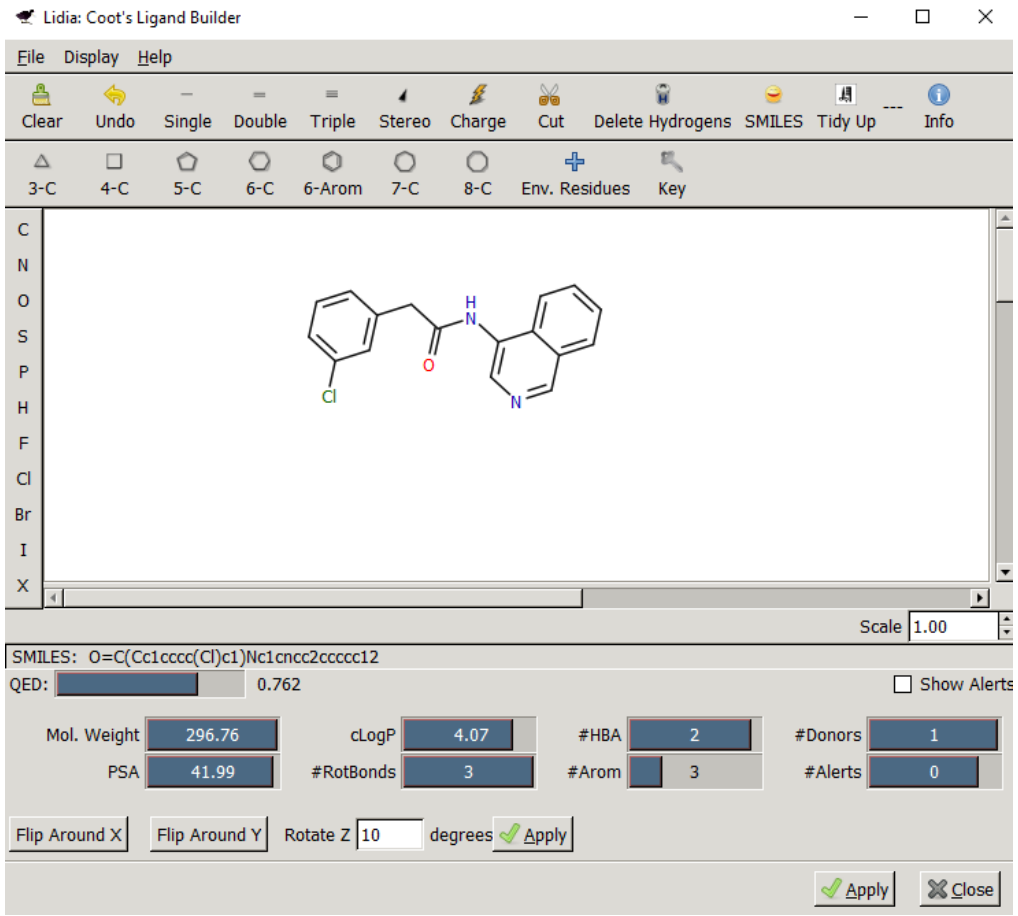
Lidia: Coot’s ligand builder



https://doi.org/10.1107/S0907444910007493

Quantitative Evaluation of Drug-Likeness

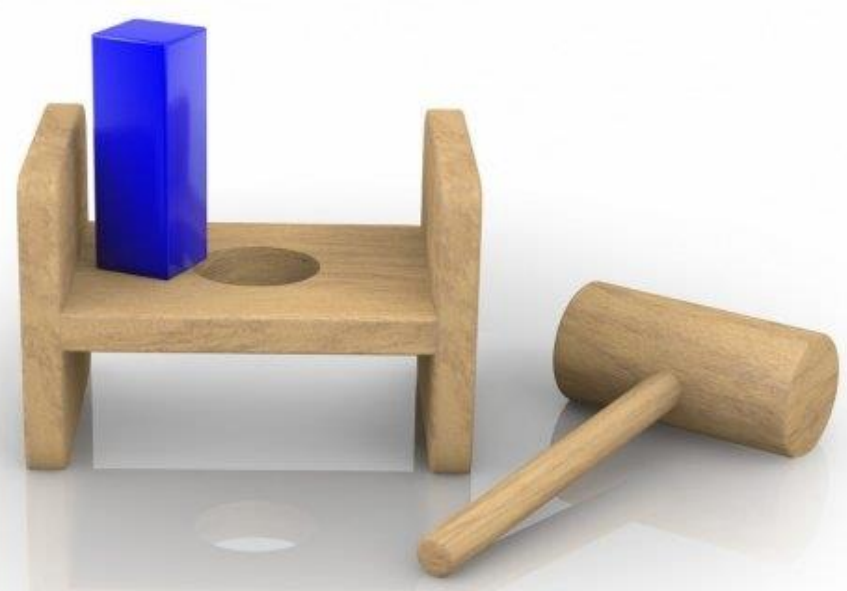
- QED score - measure of drug likeness
- Lidia displays various physical chemical properties
  - Molecular weight
  - cLogP
  - Polar Surface Area
  - Hydrogen bond donors/acceptors
  - Rotatable bonds



DOI: 10.1038/nchem.1243

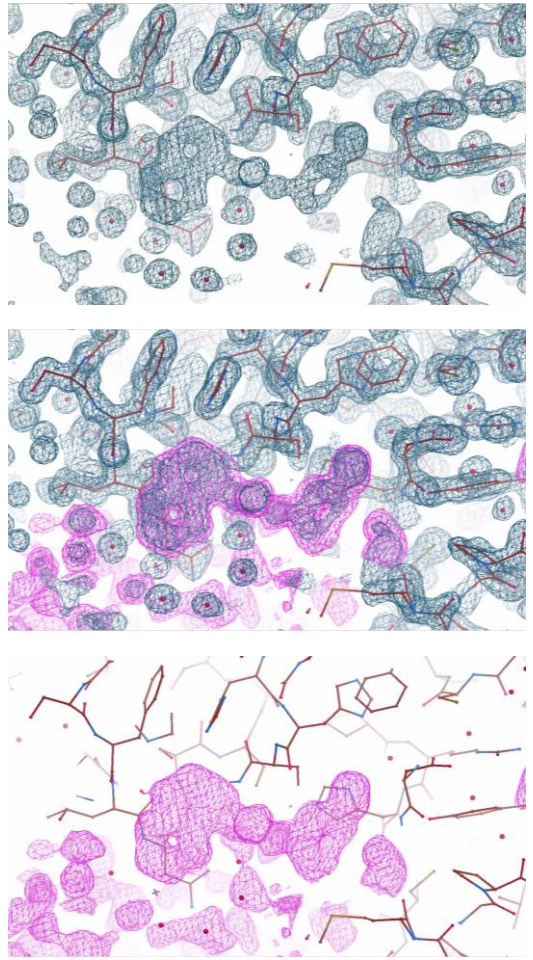
Ligand fitting with Coot

- 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 real space using Coot with real-space/rigid body refinement
  - **Remember to merge ligand and protein model**

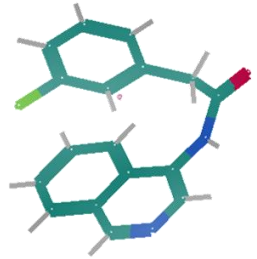


Ligand fitting with Coot – “Find Ligands”

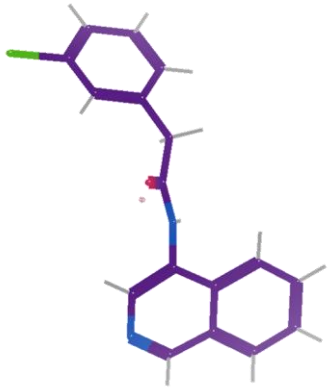
- Automated procedure:
  - Mask the map – protein coordinates used to identify unmodelled density
  - Find blobs – masked map searched for regions with substantial density
  - Cut out blobs and rank them according to density volume
  - Fit ligand into each blob
    - Match centres
    - Orientate ligand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

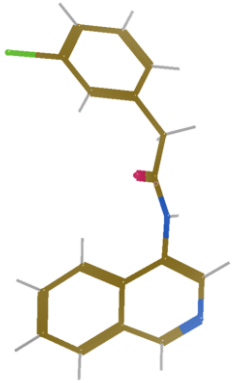


Conformation generation



- 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



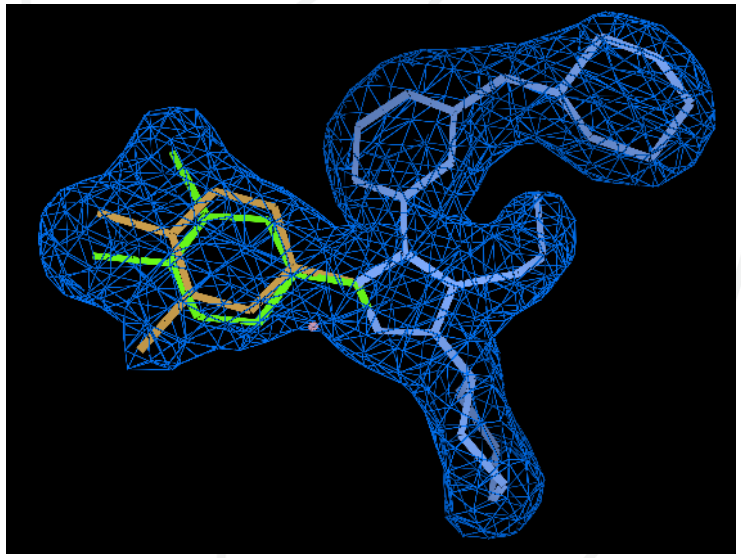


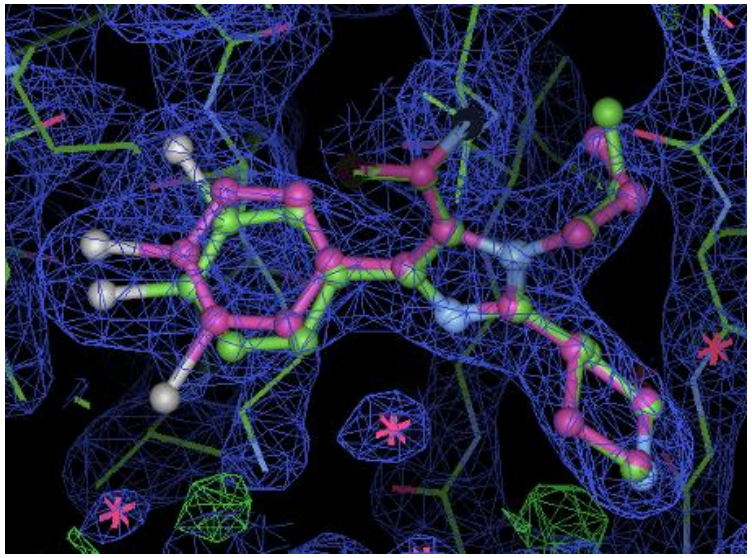
Optimizing ligand fit

- 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”

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





diamond

Disorder and mobility

- Atomic displacement parameters (B-factor) are a statistical measure of uncertainty in position of atoms
  - Provided in PDB file and Validation Report (average for each molecule)
- Ligand B-factor >>> surrounding protein: ligand unlikely to be present where modelled and/or occupancy too high
- Atomic B-factor >>> rest of molecule: part of molecule disordered
  - Proceed with caution - several approaches taken for partially ordered ligands

https://doi.org/10.1007/s10822-015-9833-8

Heavy atoms can help with ligand placement

- Fraglites - exploit the anomalous scattering of a halogen substituent

Journal of Medicinal Chemistry

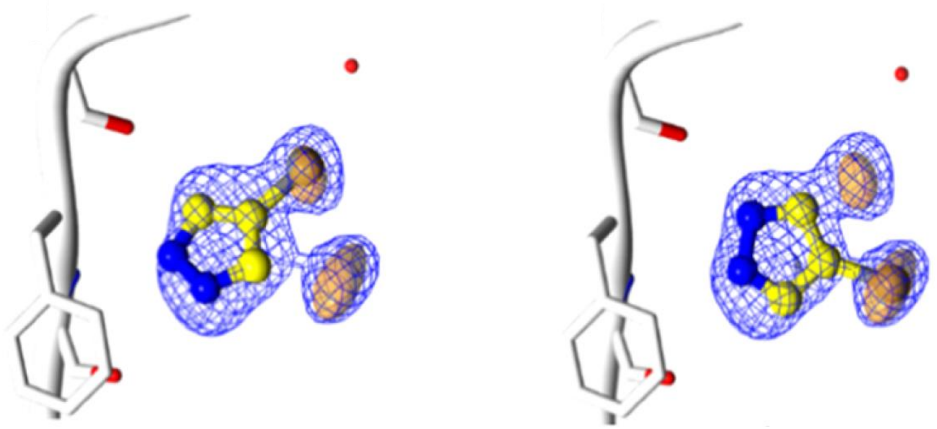
Article

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

pubs.acs.org/jmc

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

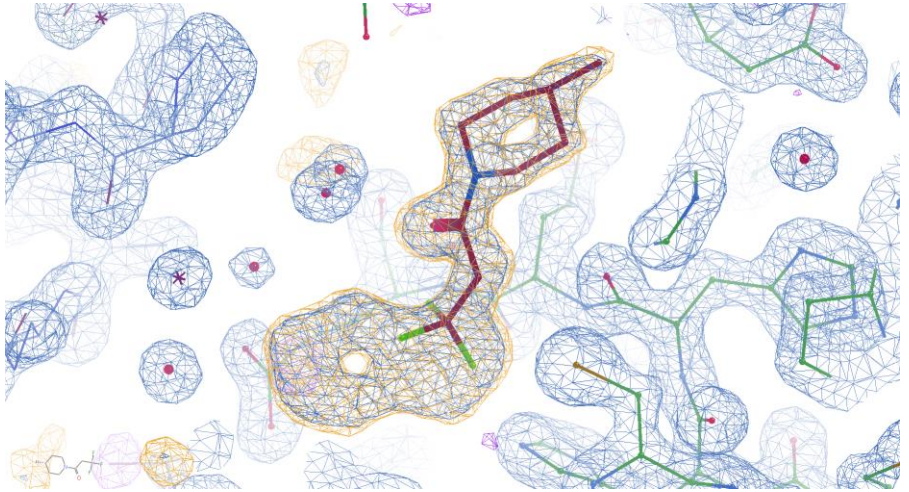
Daniel J. Wood,[†] J. Daniel Lopez-Fernandez,[‡] Leanne E. Knight,[‡] Islam Al-Khawaldeh,[‡] Conghao Gai,[‡] Shengying Lin,[‡] Mathew P. Martin,[†] Duncan C. Miller,[‡] Céline Cano,[‡] Jane A. Endicott,[†] Ian R. Hardcastle,[‡] Martin E. M. Noble,*[,†] and Michael J. Waring*[,‡]



https://doi.org/10.1021/acs.jmedchem.9b00304

diamond

Fitting unknown ligand



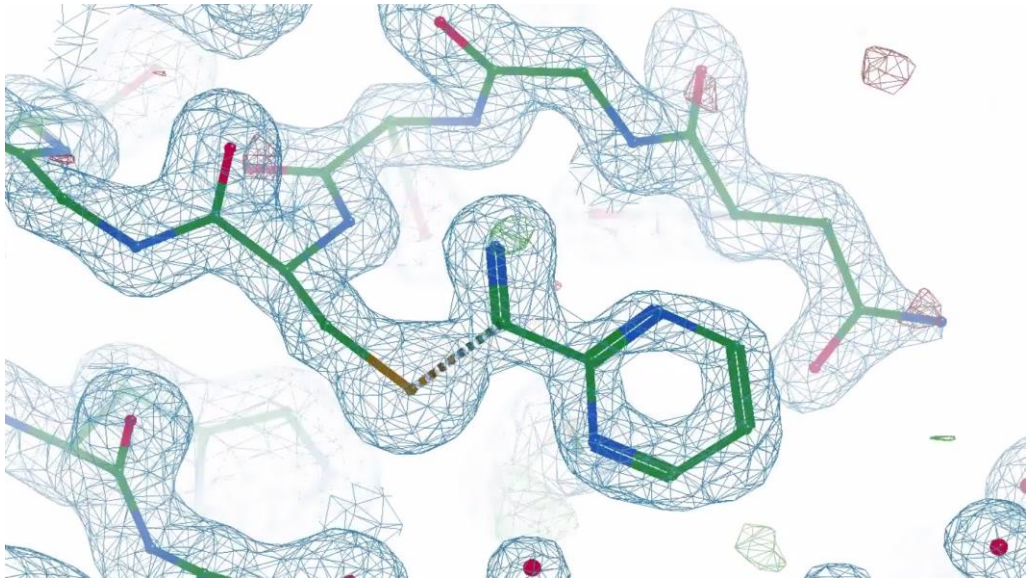
- Sometimes additional density observed for unexpected ligand
  - Crystallisation buffer components (HEPES, MES, salts)
  - Cryoprotectants (PEG, glycerol)
  - Compounds from purification (co-factors, substrates)
  - Solvents (DMSO, ethylene glycol)
  - Wrong ligand
- Possible to search for some common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

https://doi.org/10.1107/S2059798316020143
https://checkmyblob.bioreproducibility.org/server/
https://doi.org/10.1093/nar/gkab296

diamond

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



http://www.ysbl.york.ac.uk/mxstat/JLigand/tutorial_link.html

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

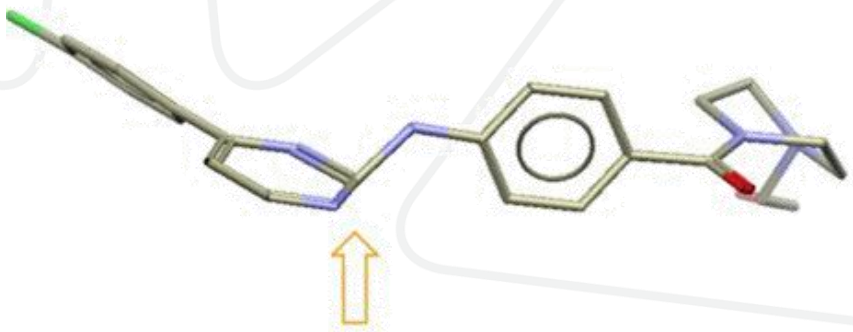
diamond

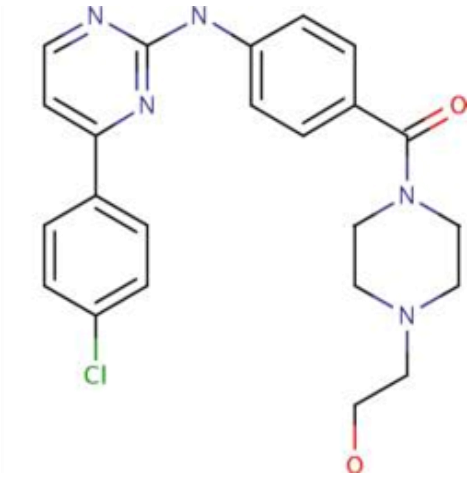
Refinement

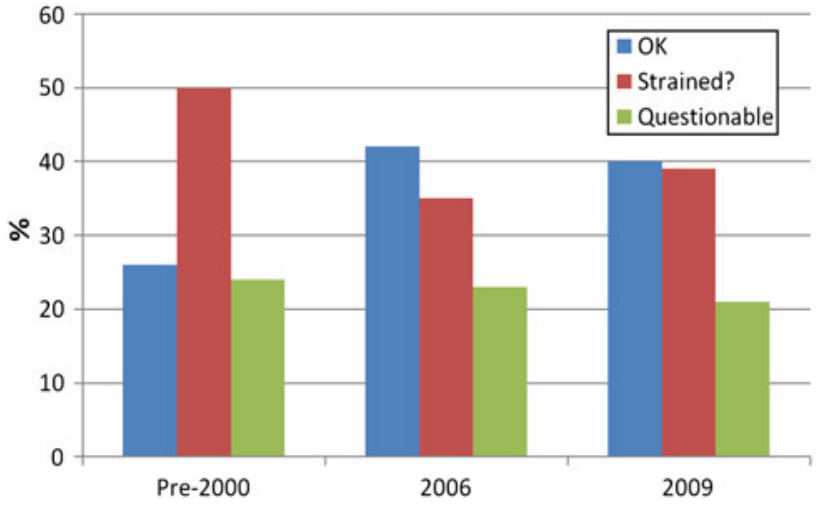
- Following ligand fitting, full refinement should be carried out with ligand restraints you generated
  - REFMAC/Buster/Phenix
- Allows protein and ligand to co-refine, optimizing agreement between your model and the experimental data
- As ligand now contributes to model phases, further density interpretation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

Ligand fitting – sources of errors

- Over 70% of crystal structures in PDB contain a ligand (>100k structures)
- Quality of small molecules in protein-ligand complexes is varied
- Ligand quality not necessarily reflected by overall quality of structure
- **Crystal structures are models and can (will) contain errors**



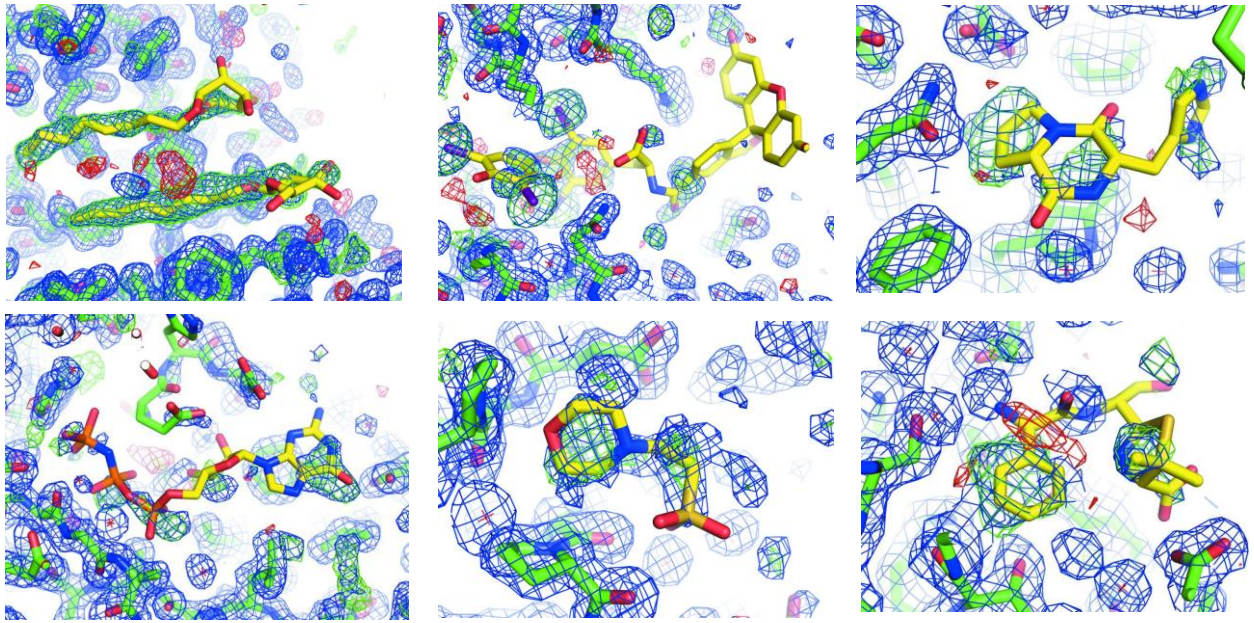




https://doi.org/10.1107/S0907444912044423
https://doi.org/10.1021/ml500220a
https://doi.org/10.1517/17460441.2011.585154

Ligand fitting – 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/disordered ligands
  - Wrong/unjustified ligands
  - Wrong ligand conformation
  - Multiple ligand conformations
  - Partial occupancy
- Important to validate structure!

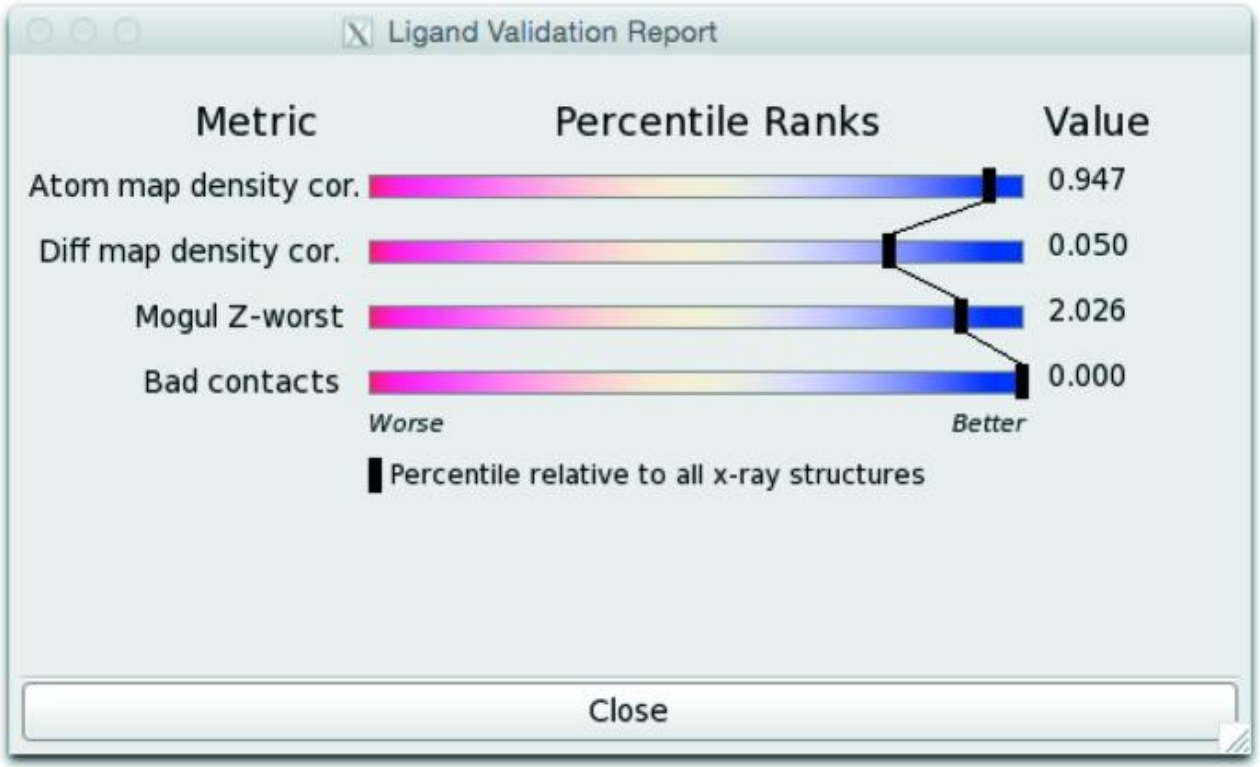


https://doi.org/10.1107/S0907444912044423
https://doi.org/10.1021/ml500220a
https://doi.org/10.1517/17460441.2011.585154

diamond

Ligand validation in Coot

- Ligand Metric Sliders



https://doi.org/10.1107/S2059798317003382

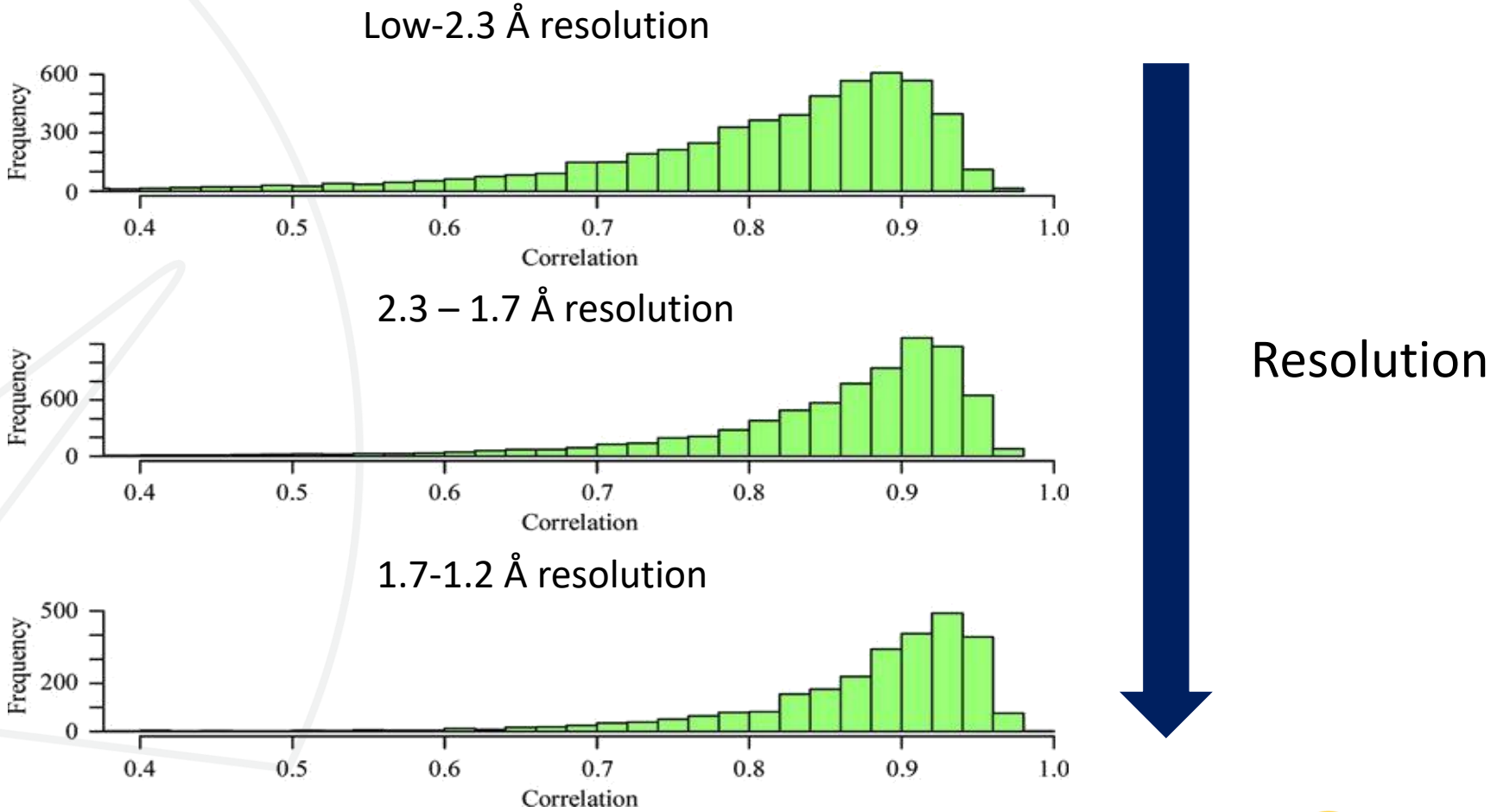
Quality of fit to experimental data

- Real-Space Correlation Coefficient (RSCC) is a measure of the fit of residues/ligands to electron density
- Varies between 1 (perfect correlation) and -1 (perfect anticorrelation)
  - >0.95 considered very good fit (43.85% of PDB ligands)
  - <0.8 considered poor fit (11.3% PDB ligands)
- Approx. 11-12% of ligands modelled in PDB considered dubious or poor

https://doi.org/10.1107/S2059798318002541
https://doi.org/10.1007/s10822-015-9833-8

diamond

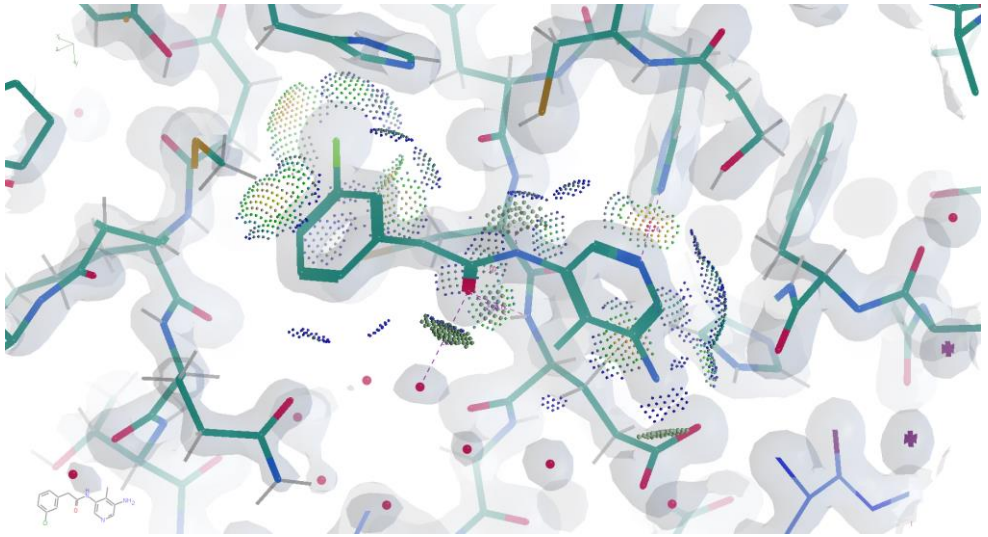
Ligand validation - density correlation

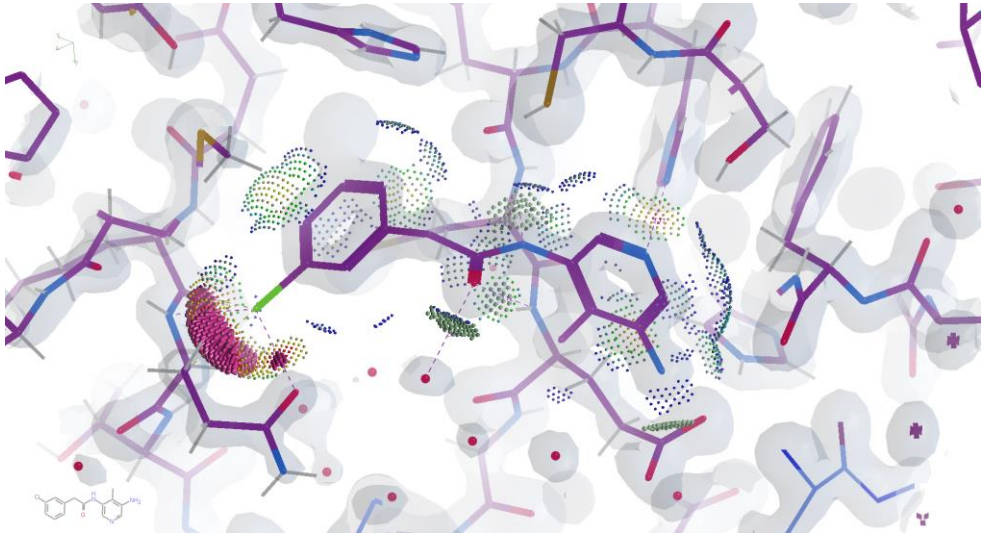


https://doi.org/10.1107/S2059798317003382

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

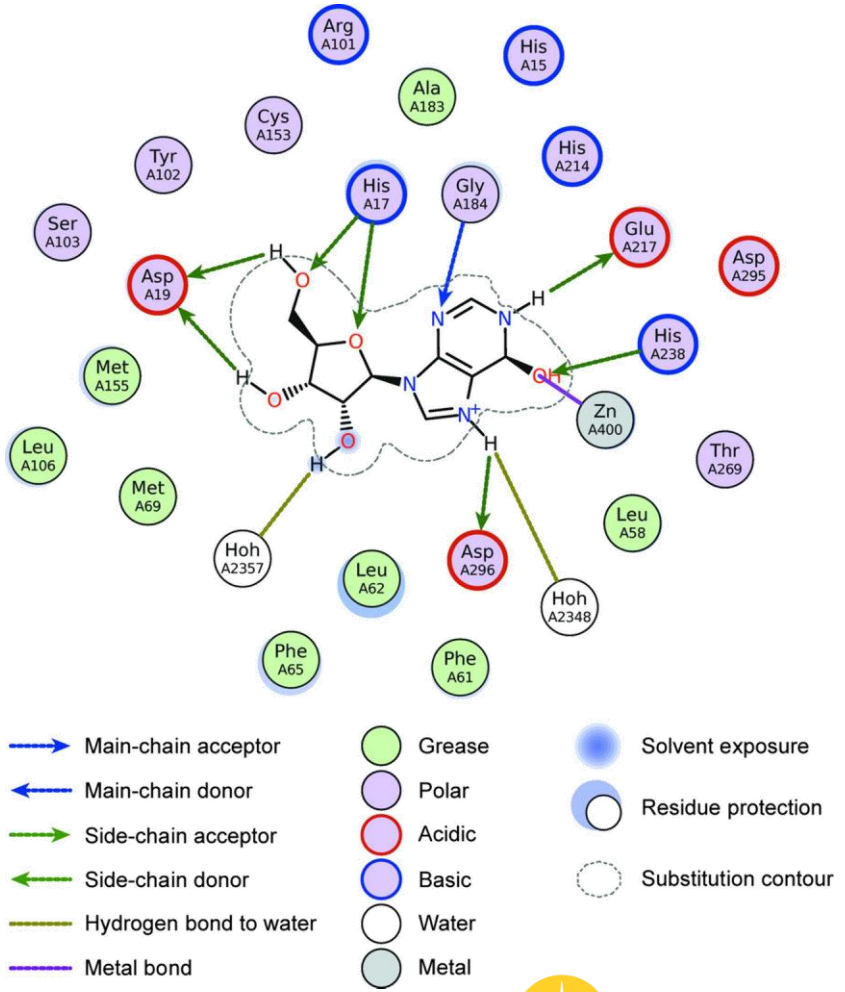




https://doi.org/10.1107/S2059798317003382

Ligand validation – Ligand environment (FLEV)

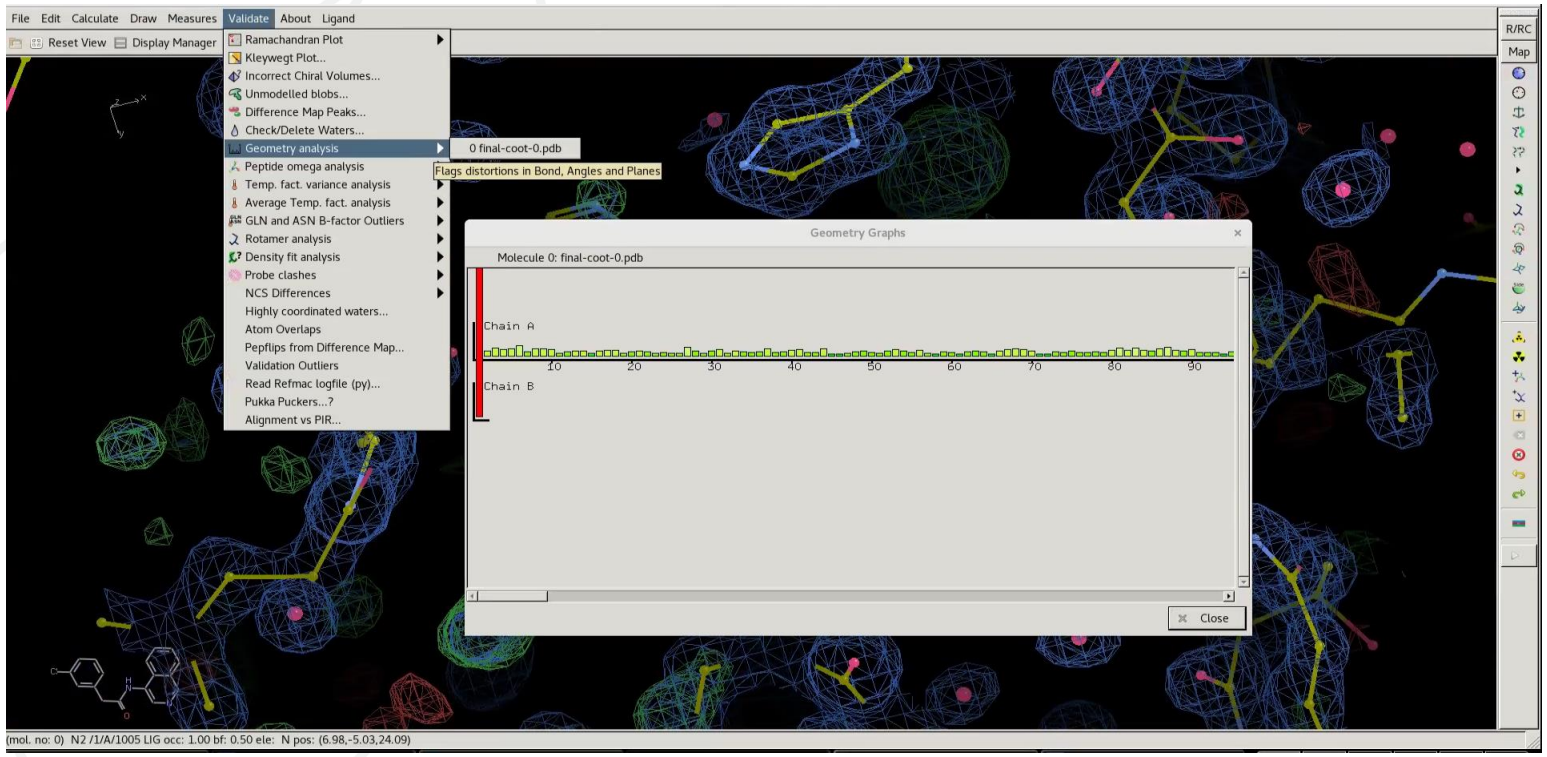
- Flatland Ligand Environment View
  - Ligand -> Toggle FLEV Ligand Interactions
- 2D representation of ligand environment
  - Binding pocket residues
  - Intermolecular interactions
  - Solvent accessible regions
  - Substitution contour



https://doi.org/10.1107/S2059798317003382

Ligand validation - geometry

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

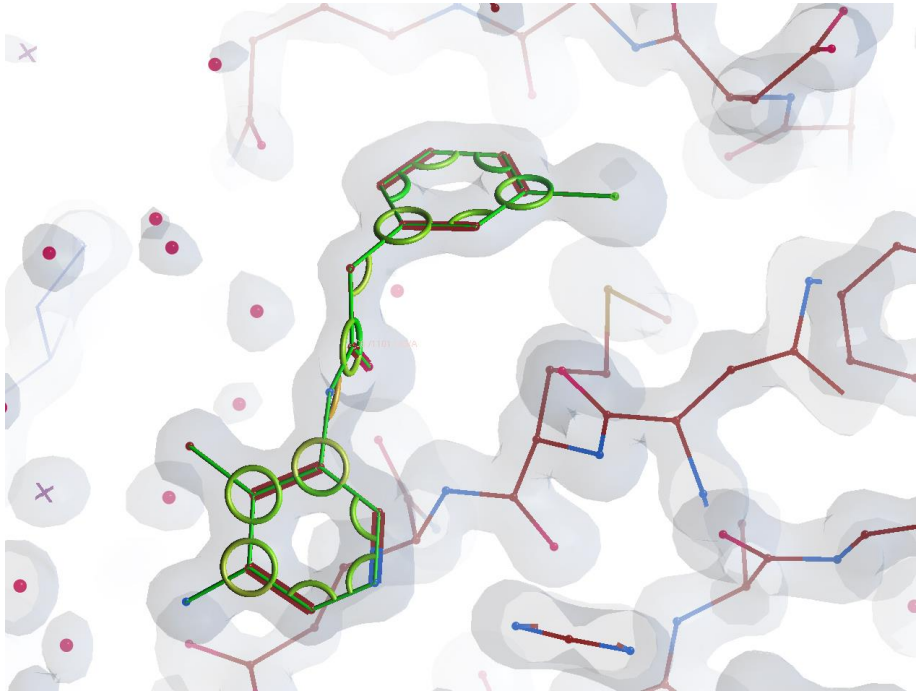


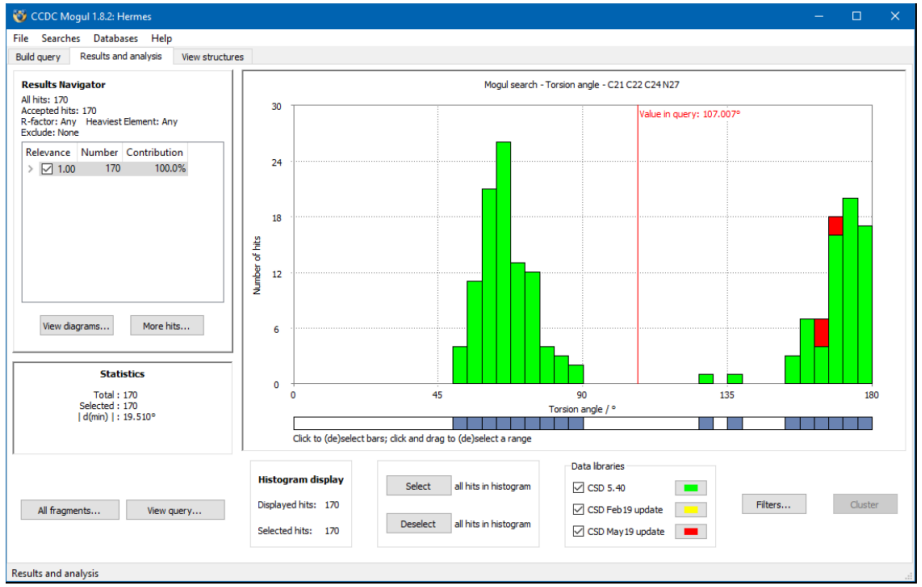
Ligand validation - geometry

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

Ligand validation - geometry

- Mogul (CCDC)
  - Compares geometries against CSD - database of small-molecule crystal structures
  - Useful cross-validation if restraints not generated using CSD data (e.g. AceDRG)
- Mogul Z-scores available as part of PDB Validation Report
  - Z-score > 2.0 flagged as ‘outliers’ but values <5 can be acceptable

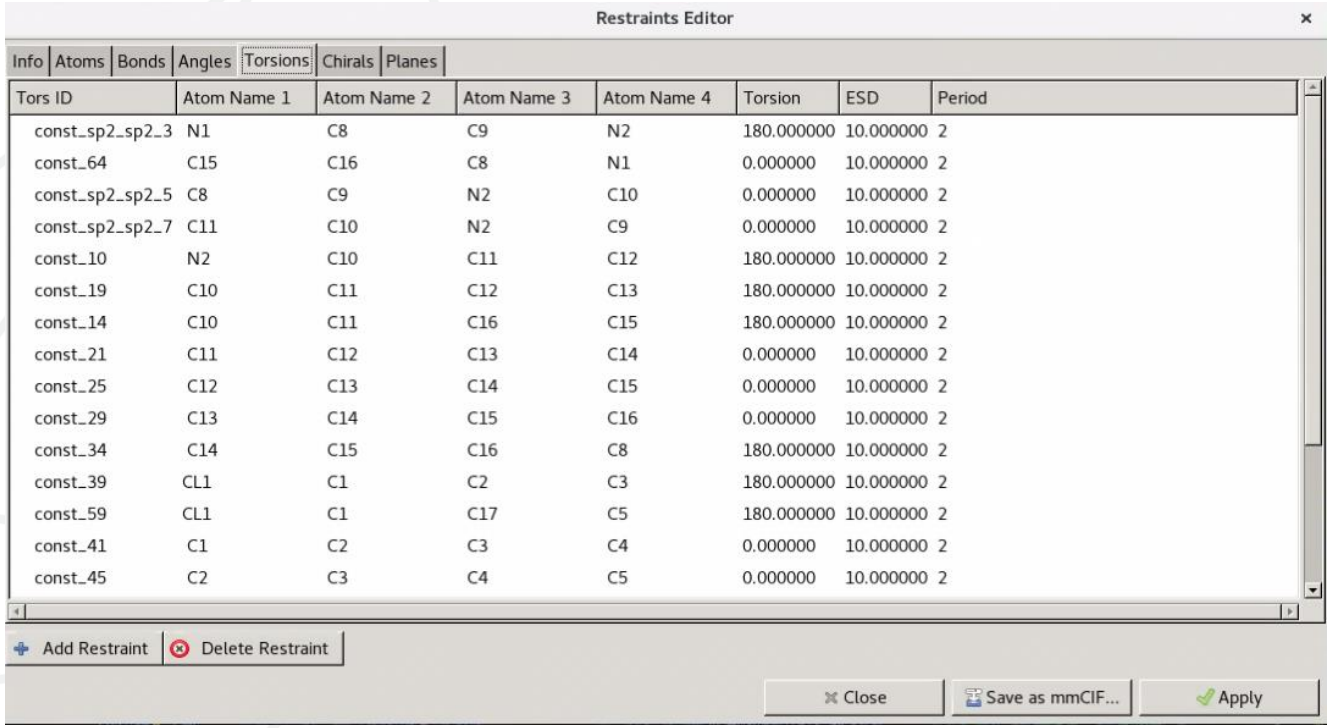




https://doi.org/10.1107/S2059798318002541
https://doi.org/10.1107/S2059798317003382
https://doi.org/10.1021/ci049780b
https://doi.org/10.1021/ci200439d

Ligand validation - geometry

- 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

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…

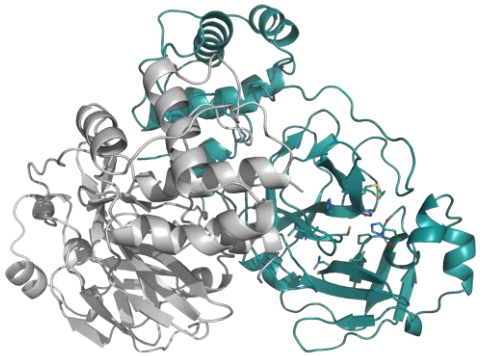
Summary

- Crystallography is subjective and burdened with confirmation bias
  - Data quality, electron density map interpretation, knowledge of chemistry
- Ligand binding is often incomplete (particularly with fragments)
- Crystallographers produce models
  - “All models are wrong, but some are useful”
  - **How can we improve our models?**
- Shouldn’t make assumptions when working with models
  - Speak to crystallographer and/or carry out own validation
  - **How can we better communicate information about these models?**

diamond

Ligand Fitting with Coot Tutorial

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
  - Ligand coordinates (pdb) and restraints (cif)

https://doi.org/10.1101/2020.10.29.339317
https://doi.org/10.1126/science.abo7201

diamond

Tutorial – Part 1: Identify binding site

- Copy data: /dls/i04-1/data/2023/mx37045-5/processing/DLS-CCP4_Covid_Moonshot
- 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
- Validate -> Unmodelled blobs…
- Select difference map and protein model
- Click “Find blobs” and search through identified blobs
- Did you find a blob near Cys145?

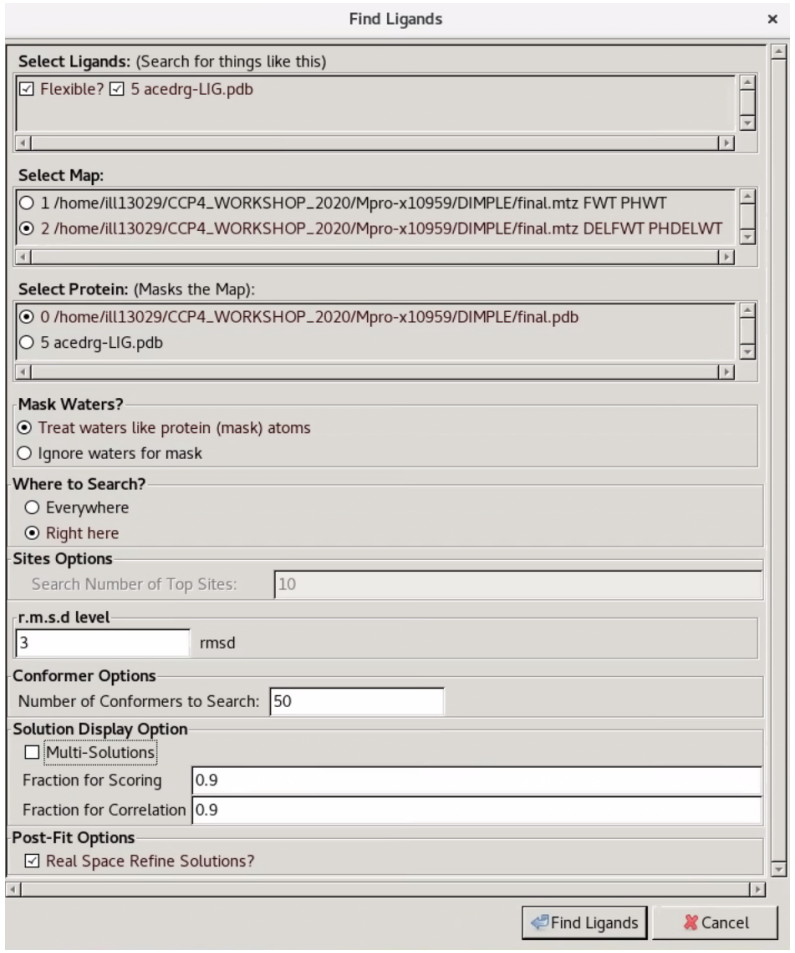
Tutorial – Part 2: Generating restraints

You can either…

- Generate restraints in Coot:
  - 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
- Generate restraints using CCP4 Cloud (“Make Ligand with Acedrg”)
- Use pre-prepared compound pdb and cif files (from Grade webserver)

Tutorial – Part 3: Fitting ligand

- We now know where to place the ligand and have the ligand restraints
- *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**
  - Edit -> merge molecules
- Tidy up any surrounding residues/solvent molecules

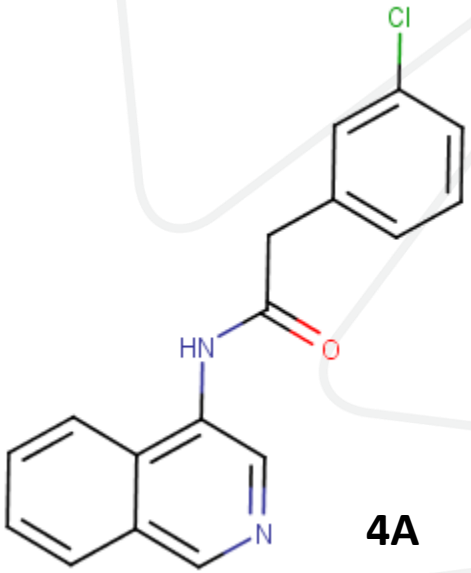


Tutorial – Part 4: Validation

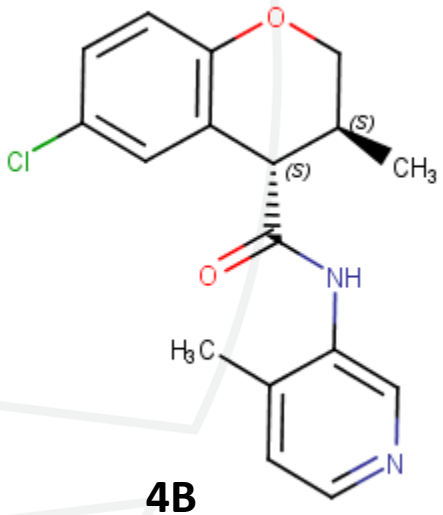
- Measures -> Environment Distances -> Show Residue Environment?
  - Are molecular interactions sensible?
- Ligand -> FLEV this residue -> Show Env. Residues
- Ligand -> Isolated Molprobity dots for this ligand
  - Do you see any significant clashes?
- Ligand -> Display Ligand Distortions
  - How does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

Tutorial – Part 5

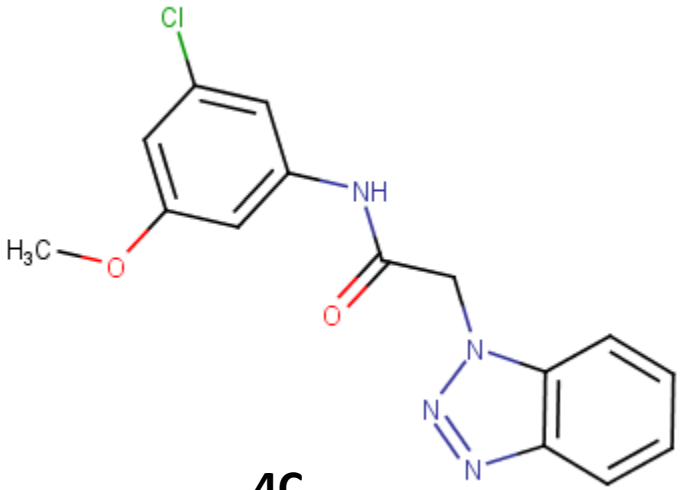
- Dataset Mpro_4 contains DIMPLE files and SMILES strings for 3 ligands
- Using the methods described previously, identify which ligand is bound to the protein



4A



4B



4C

diamond

Tutorial – Part 6: 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 as necessary and repeat refinement

diamond