

Ligand fitting

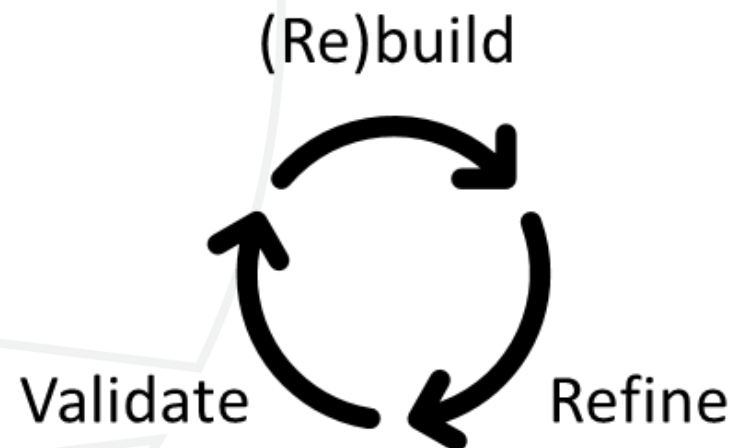
Daren Fearon

Senior Beamline Scientist XChem
Diamond Light Source

Daren.Fearon@diamond.ac.uk

Ligand fitting aims

- To model small molecules 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

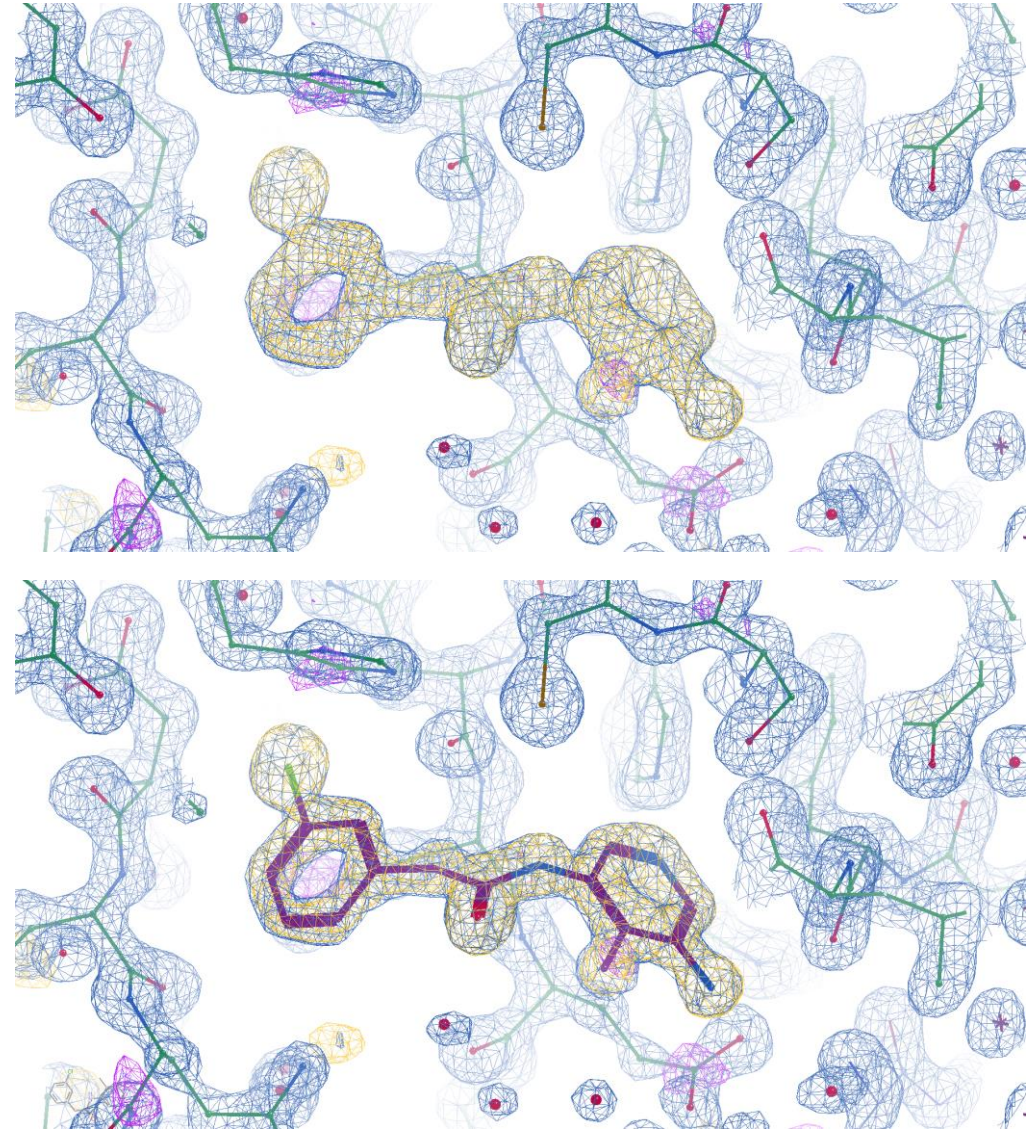


Ligand fitting scenarios

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

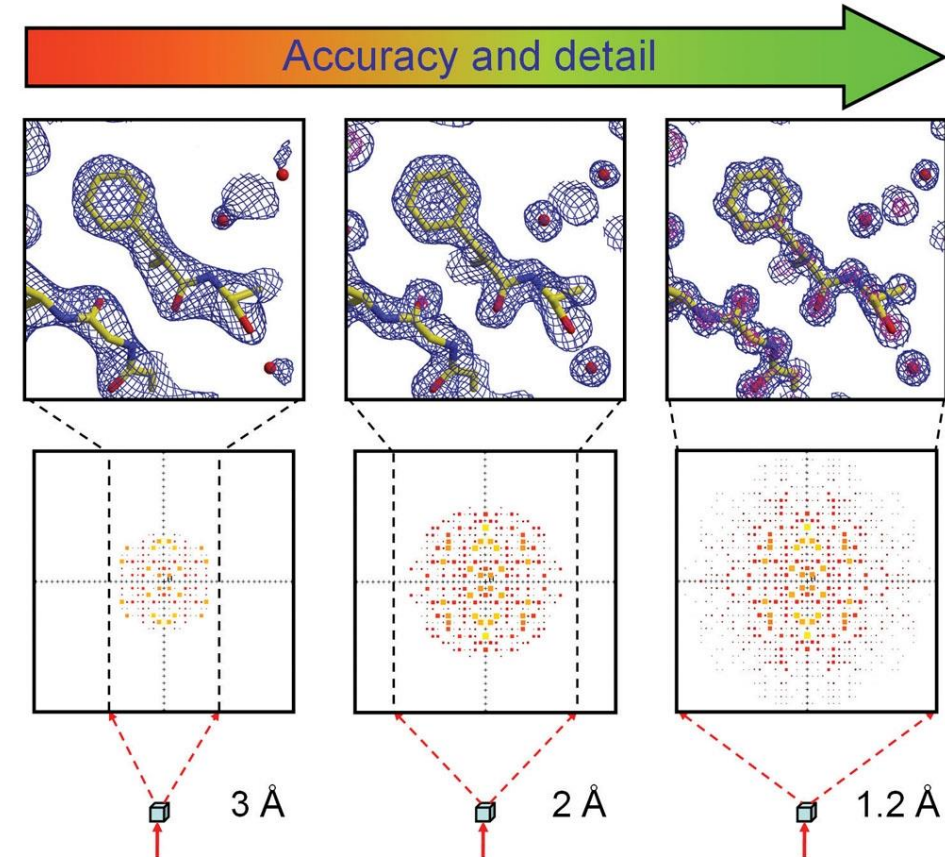
Ligand fitting process

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



Ligand fitting assumptions

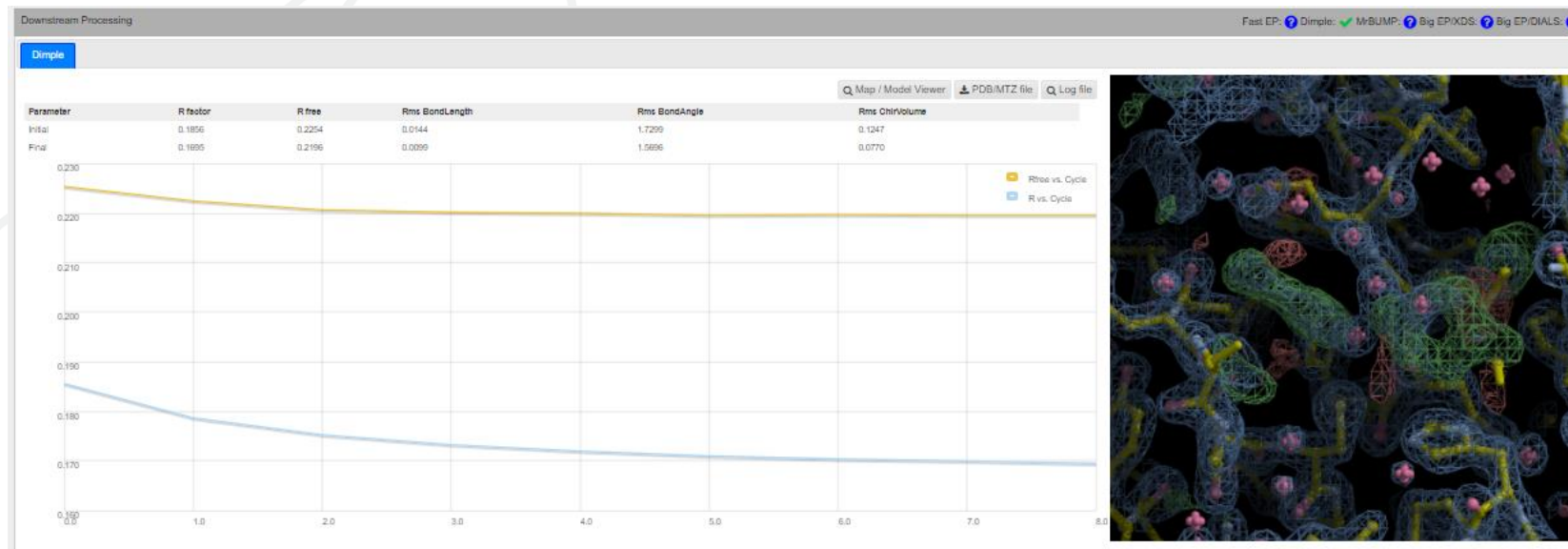
- Data is of sufficient quality to model ligand
 - Ideally 2.5 Å or better
- A good starting model is available
- Ligand is known
 - Description (2D structure/SMILES string) available or ligand is contained in ligand dictionary (e.g. ATP)
- Ligand is actually bound



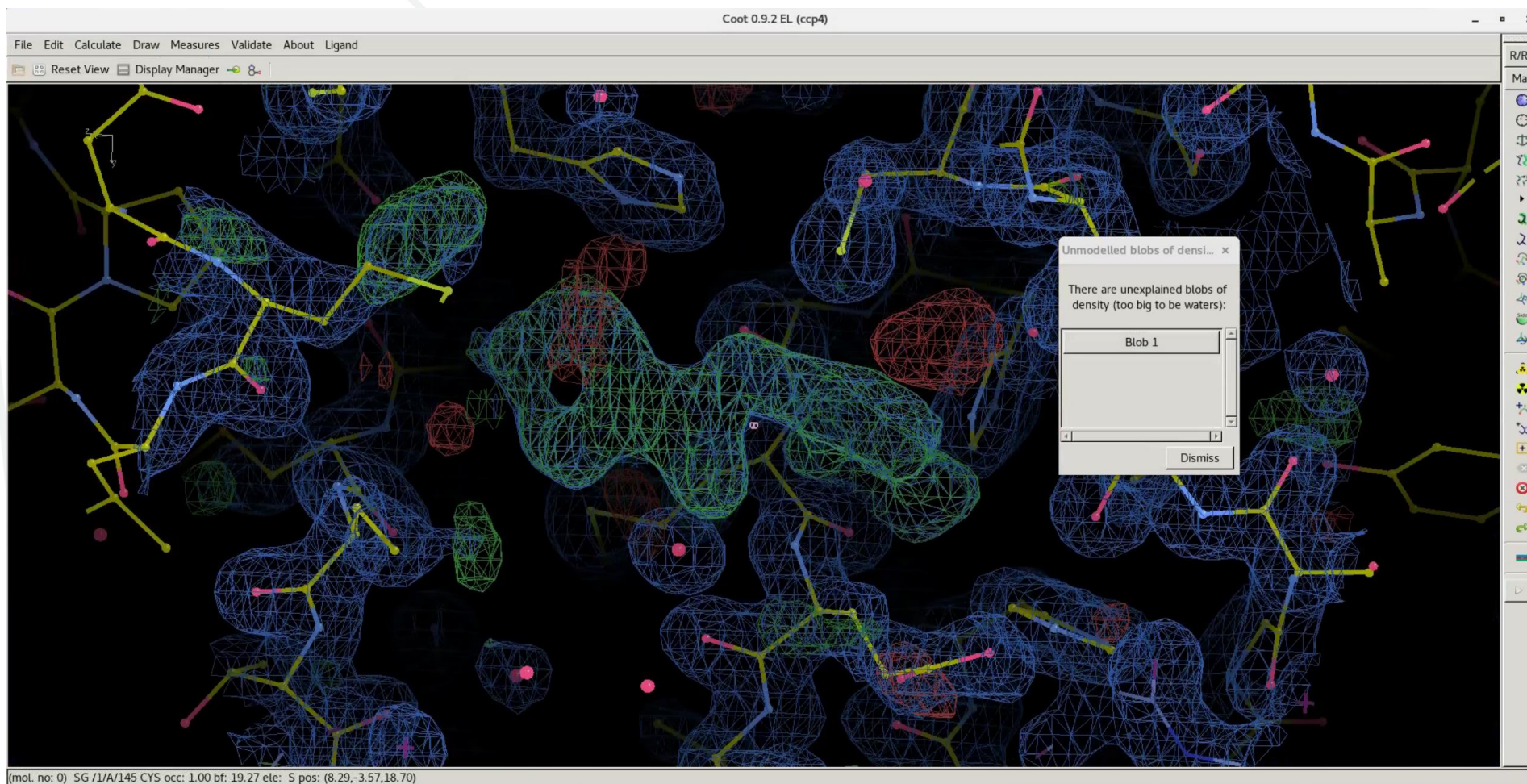
Reproduced from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

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)



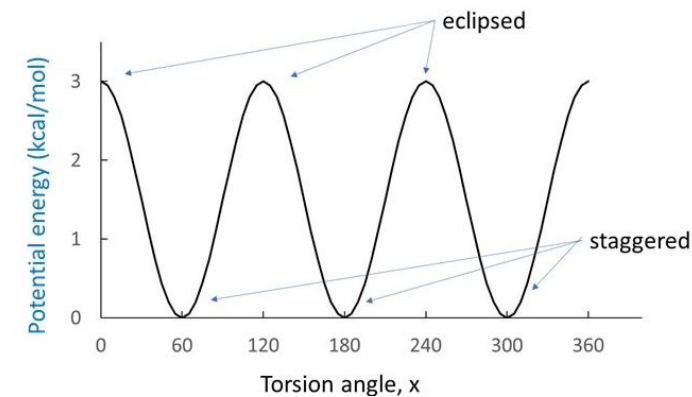
Identifying ligand density – unknown site



Ligand restraints

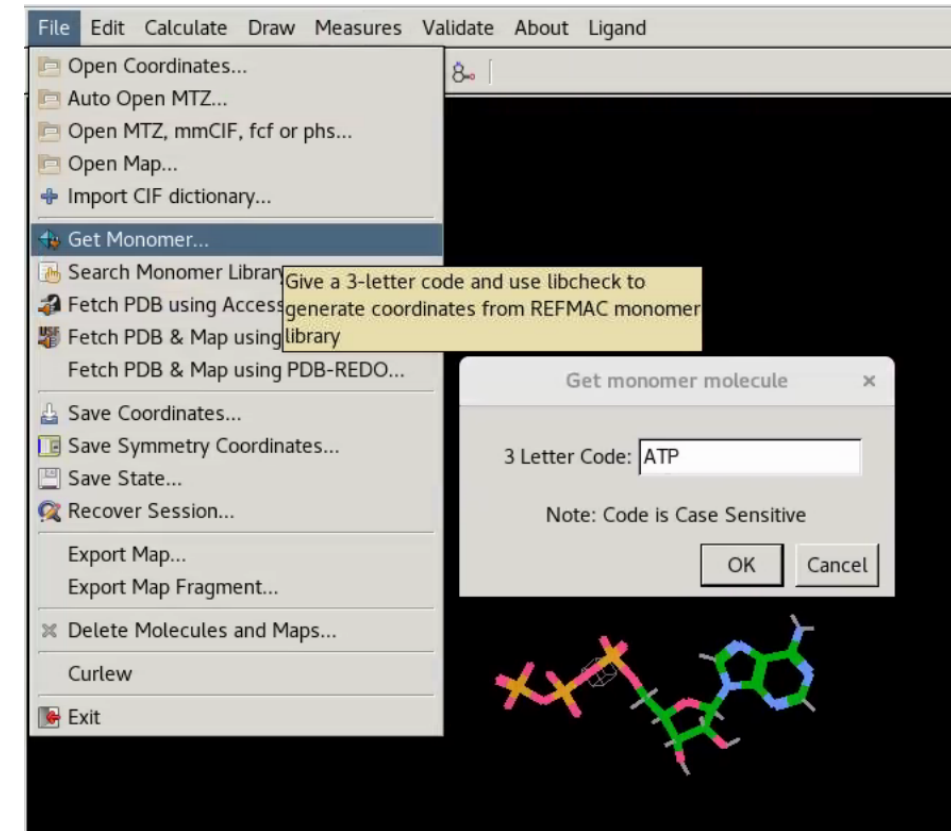
- In macromolecular crystallography experimental data alone insufficient to accurately define 3D structure
- Refinement packages utilize both geometry and experimental data
- For small molecules, a ligand description is used to generate a dictionary used for modelling
- Restraints describe the relative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

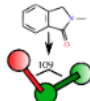
```
#
data_comp_list
loop_
  _chem_comp.id
  _chem_comp.three_letter_code
  _chem_comp.name
  _chem_comp.group
  _chem_comp.number_atoms_all
  _chem_comp.number_atoms_nh
  _chem_comp.desc_level
LIG      LIG      .      non-polymer      34      21      .
#
data_comp_LIG
#
loop_
  _chem_comp_atom.comp_id
  _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
```



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
- Also available on CCP4 cloud



 **Make Ligand with Acedrg**

job description:

output id:

Use

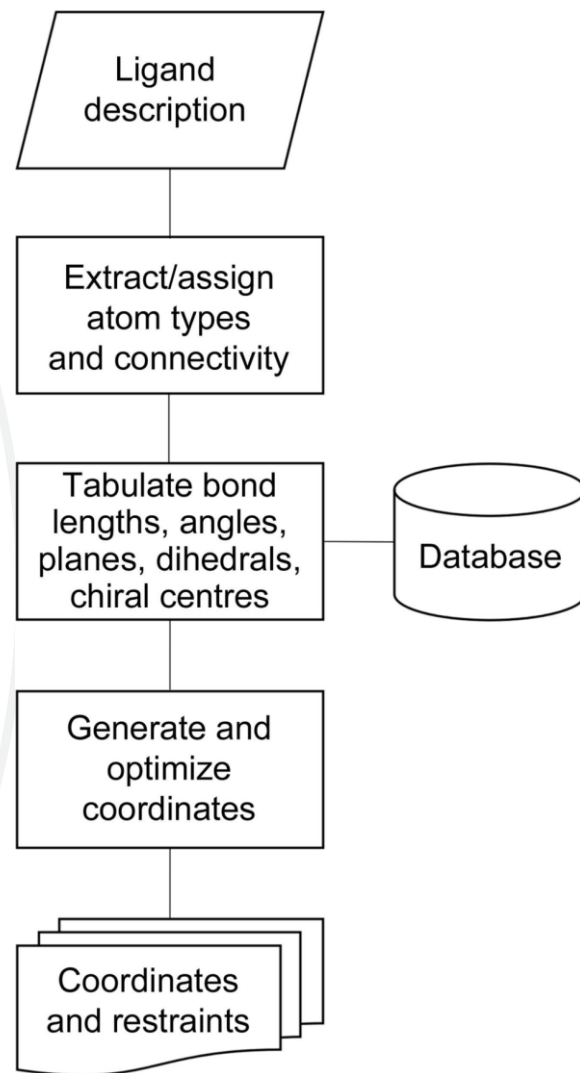
3-letter Code ☐ Recalculate with AceDrg

Codes XXX, LIG, DRG, INH, LG0, LG1, LG2, LG3, LG4, LG5, LG6, LG7, LG8, LG9 are reserved by Coot and cannot be used here.

☐ Suppress protonation/deprotonation by AceDrg

Number of initial conformers to try

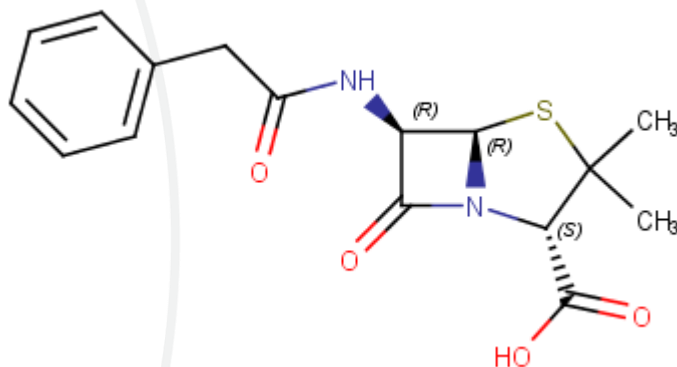
Generating dictionary for novel ligands



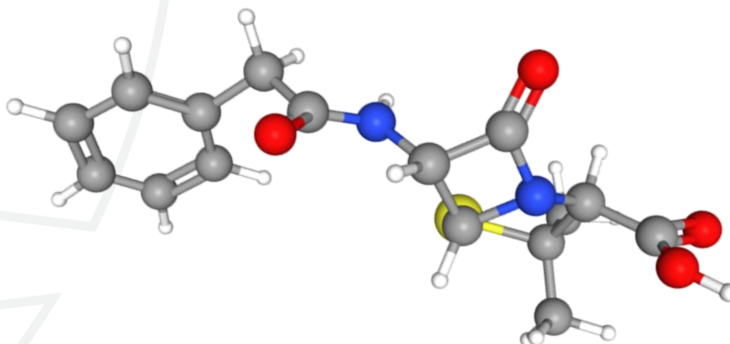
Ligand description – common input formats

- Simplified molecular-input line-entry system (SMILES)
 - 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

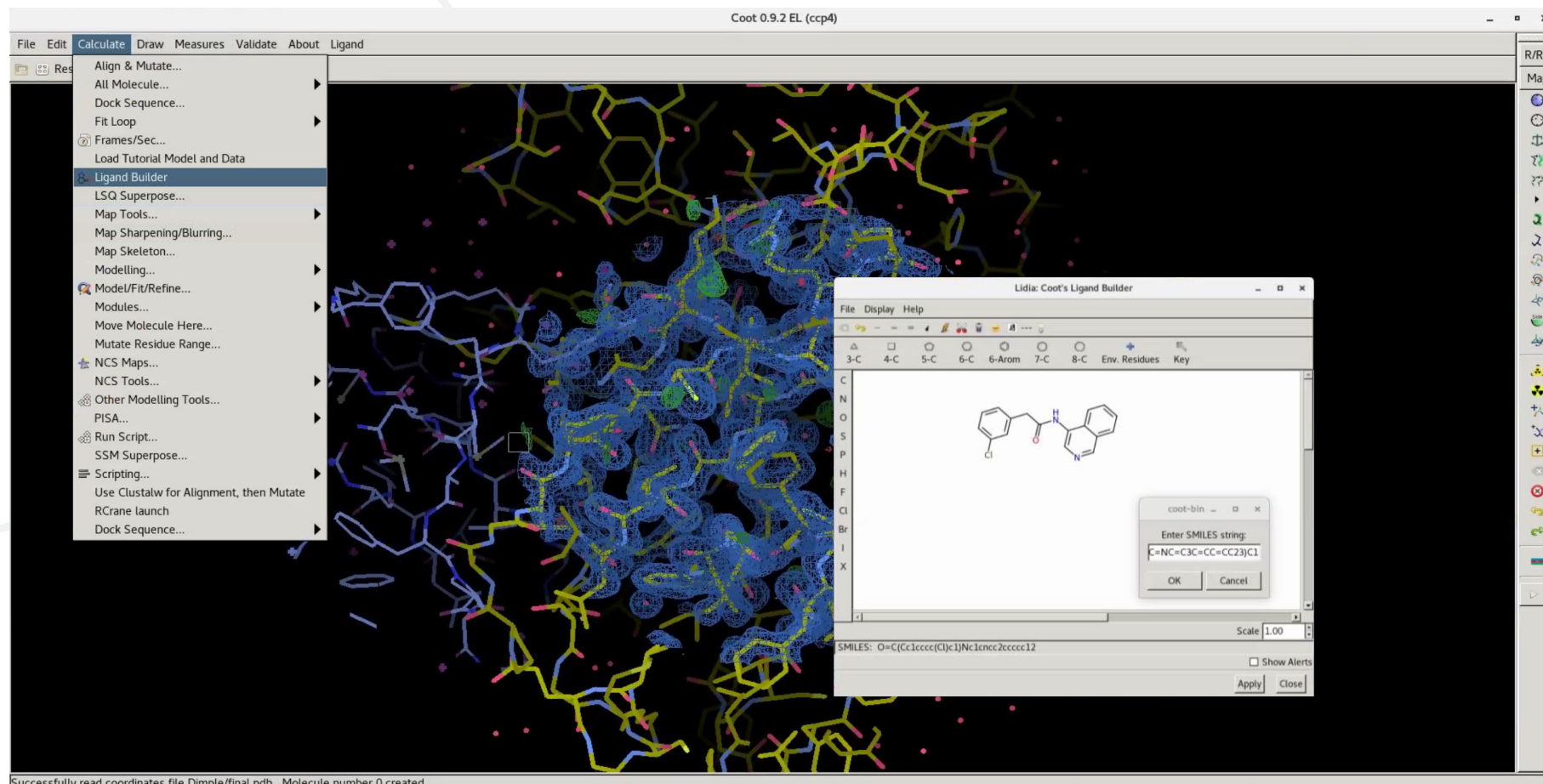
Generating and optimizing coordinates

- Various programs available, mostly free for academic users

Program name	AceDRG	Grade	eLBOW	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

Lidia: Coot's ligand builder



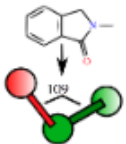
Restraint generation on CCP4 cloud

[0041] make ligand (new)

• Input • Output

Run

Close

 **Make Ligand with Acedrg**

job description:

output id:

Use

SMILES string

Ligand Code *if no code is given (recommended), a suitable new one will be autogenerated*

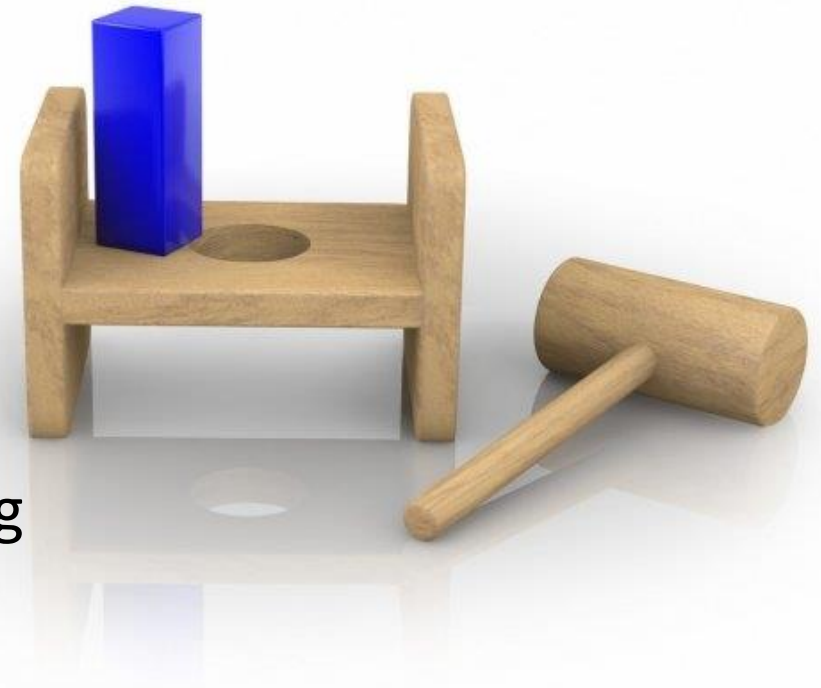
Codes XXX, LIG, DRG, INH, LG0, LG1, LG2, LG3, LG4, LG5, LG6, LG7, LG8, LG9 are reserved by Coot and cannot be used here.

☐ Suppress protonation/deprotonation by AceDrg

Number of initial conformers to try

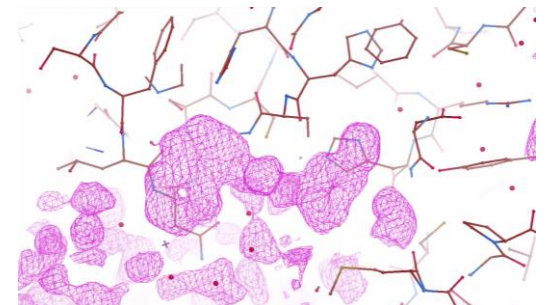
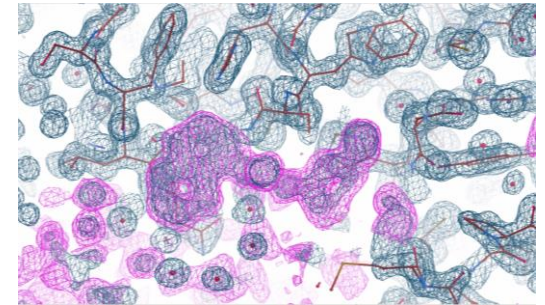
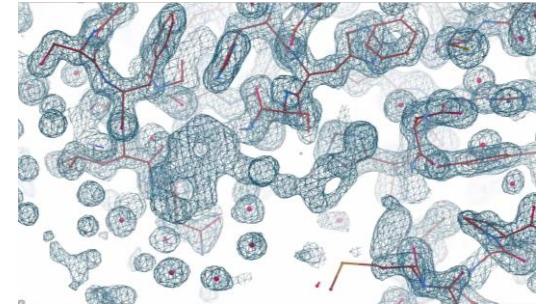
Ligand fitting

- 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**



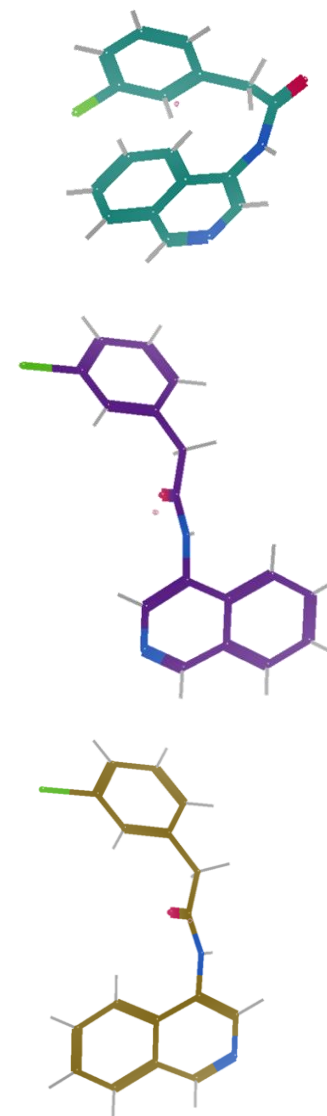
Ligand fitting with Coot – “Find Ligands”

- 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



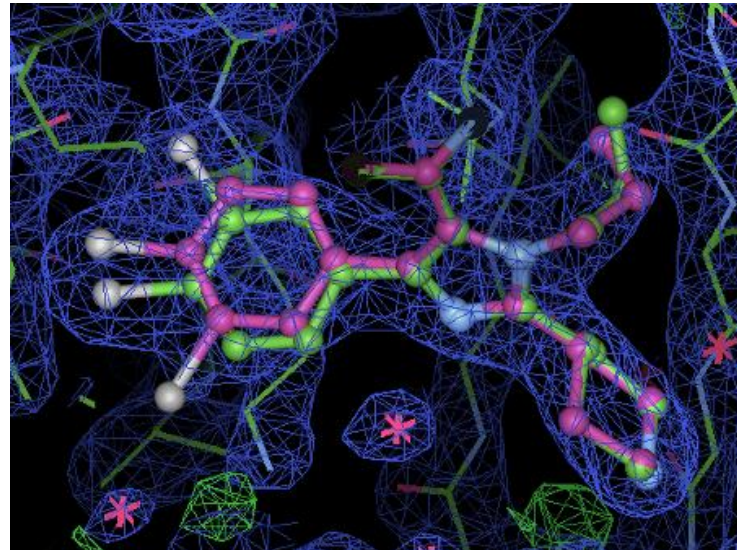
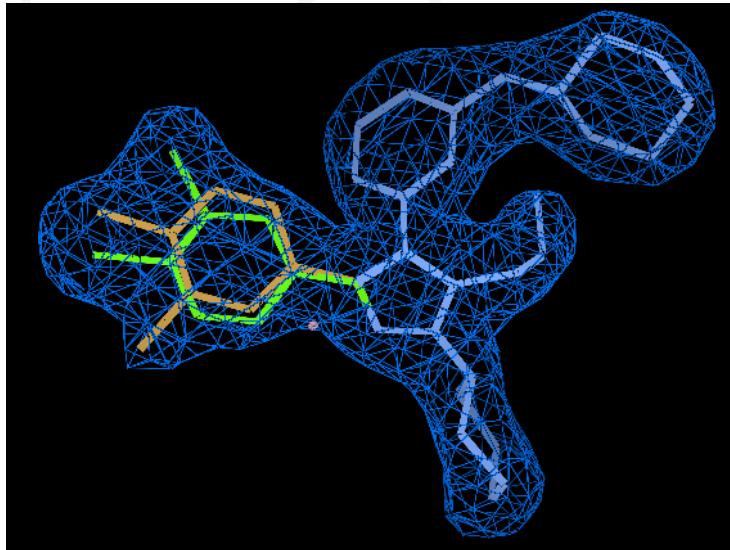
Optimizing ligand fit

- 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



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



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

Heavy atoms can help with ligand placement

- Fraglites - exploit the anomalous scattering of a halogen substituent

Journal of
**Medicinal
Chemistry**

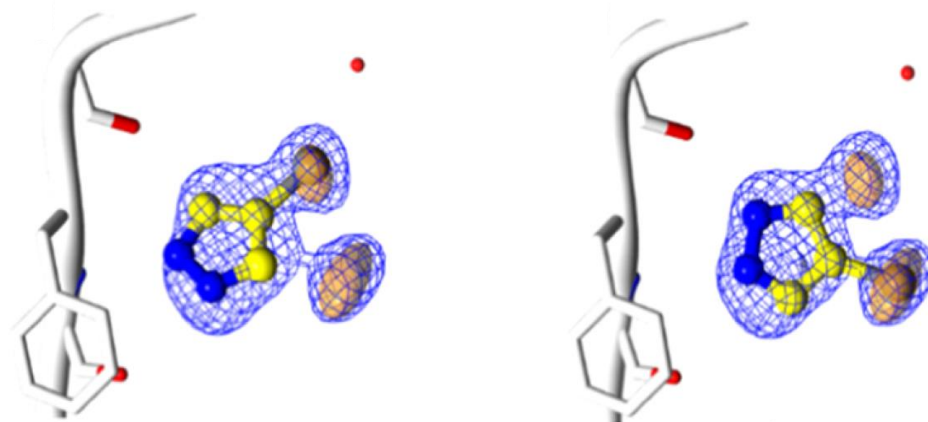
Cite This: *J. Med. Chem.* 2019, 62, 3741–3752

Article

pubs.acs.org/jmc

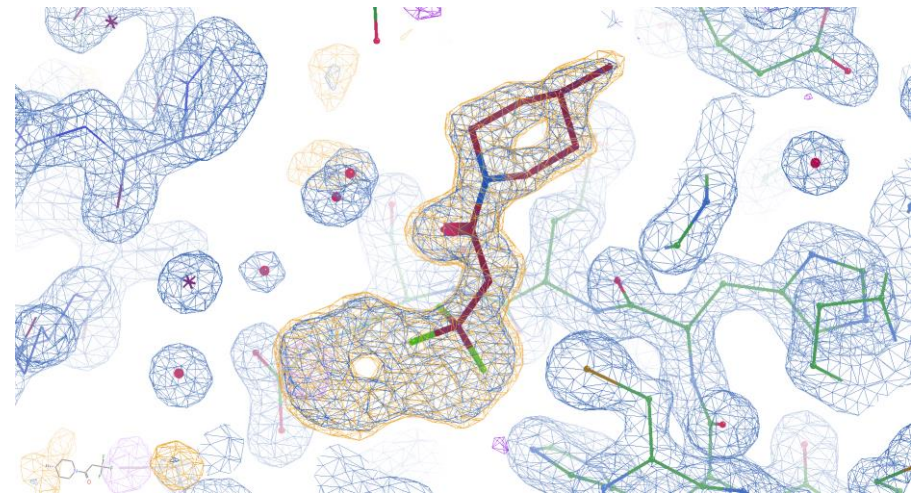
FragLites—Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation

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^{*,‡}



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/>

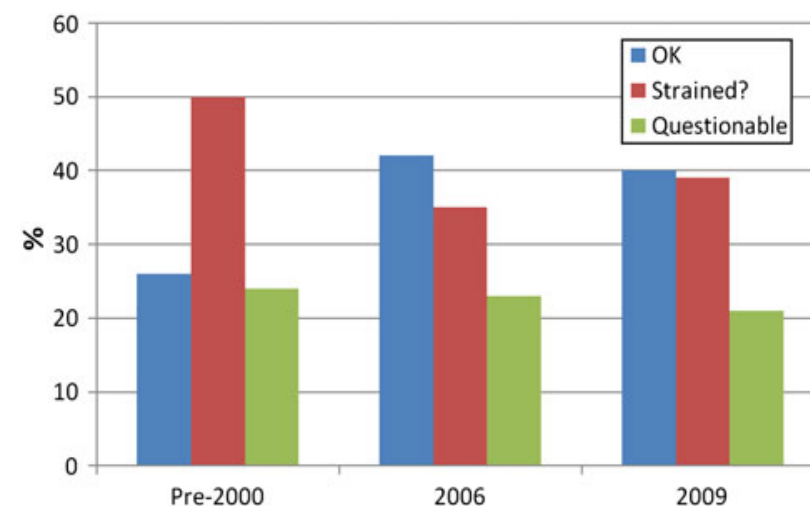
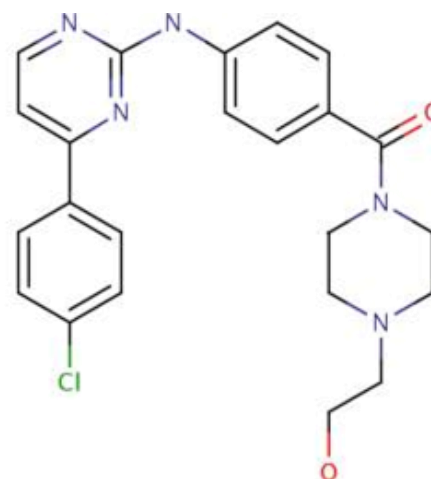
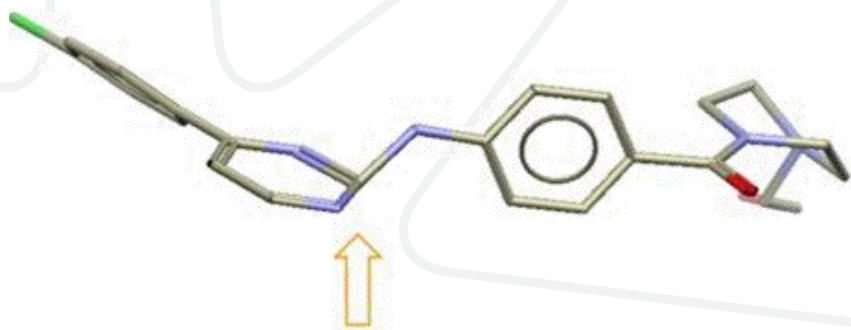
<https://doi.org/10.1093/nar/gkab296>

Non-Crystallographic Symmetry and ligands

- For structures with 2 or more protein chains related by NCS, it can be exploited to help fit ligands in Coot
- 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 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**



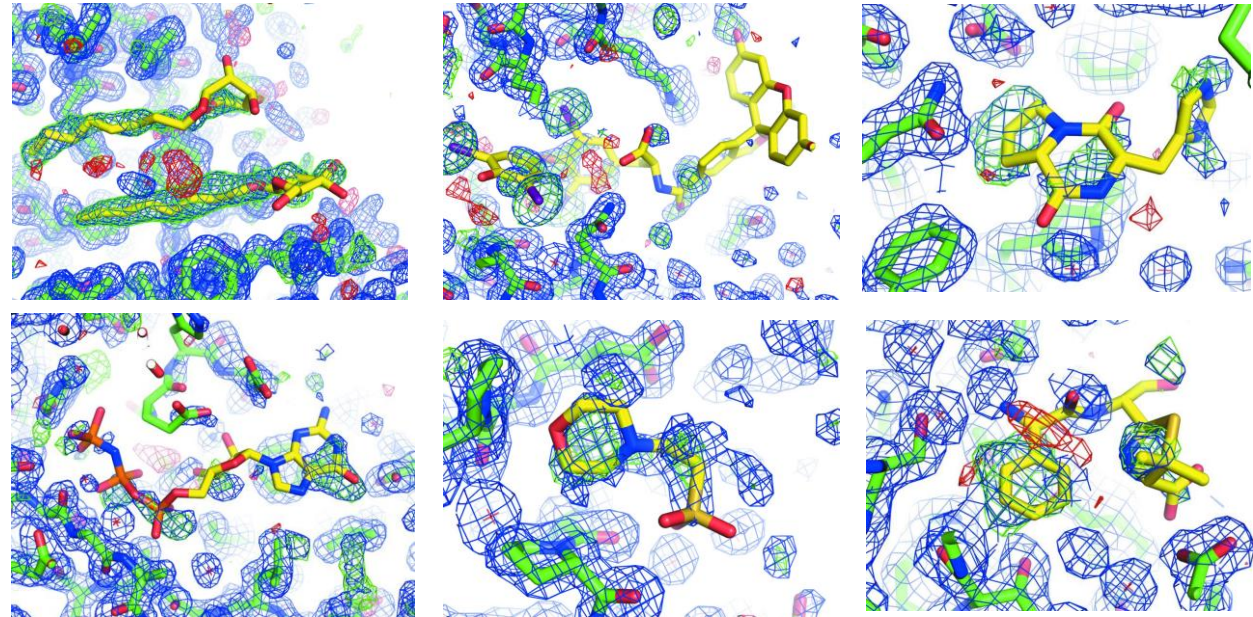
<https://doi.org/10.1107/S0907444912044423>

<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!



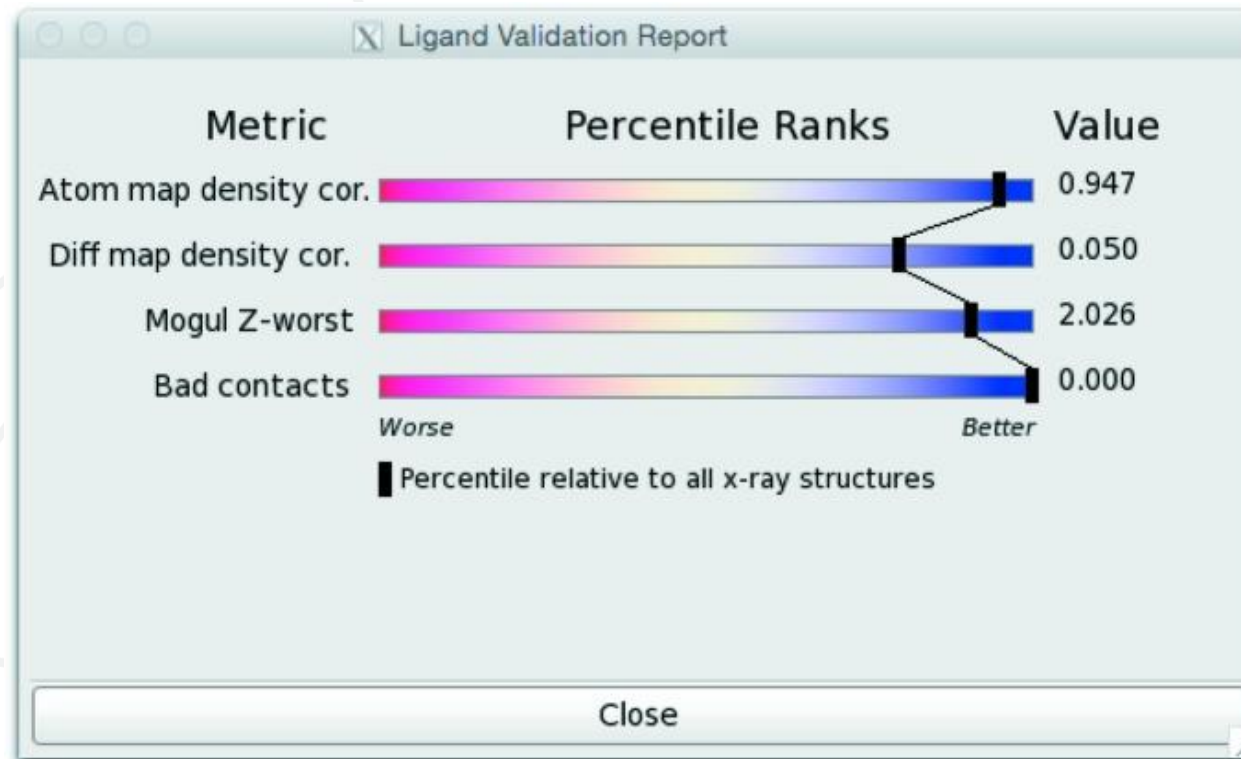
<https://doi.org/10.1107/S0907444912044423>

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

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

Ligand validation in Coot

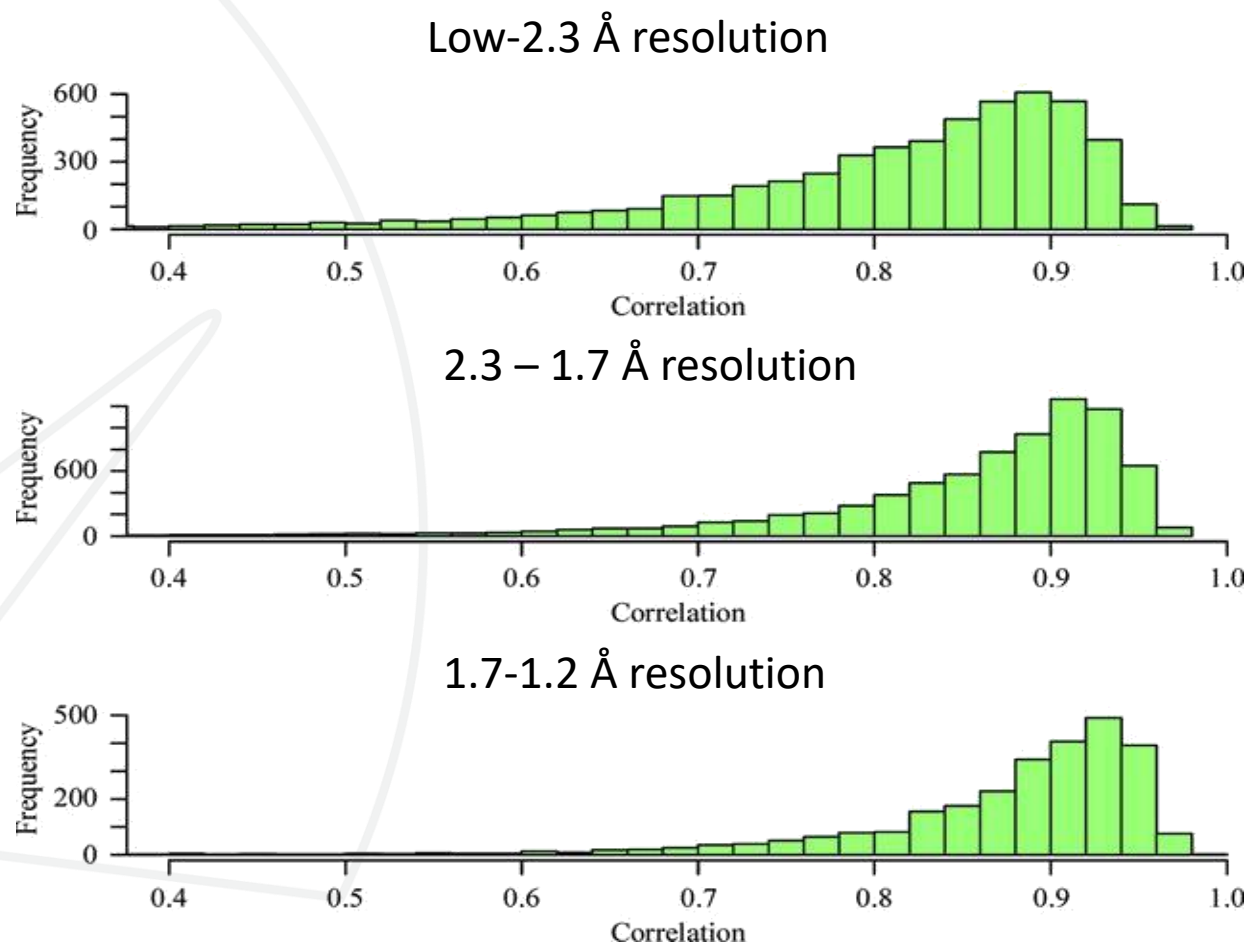
- Ligand Metric Sliders



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

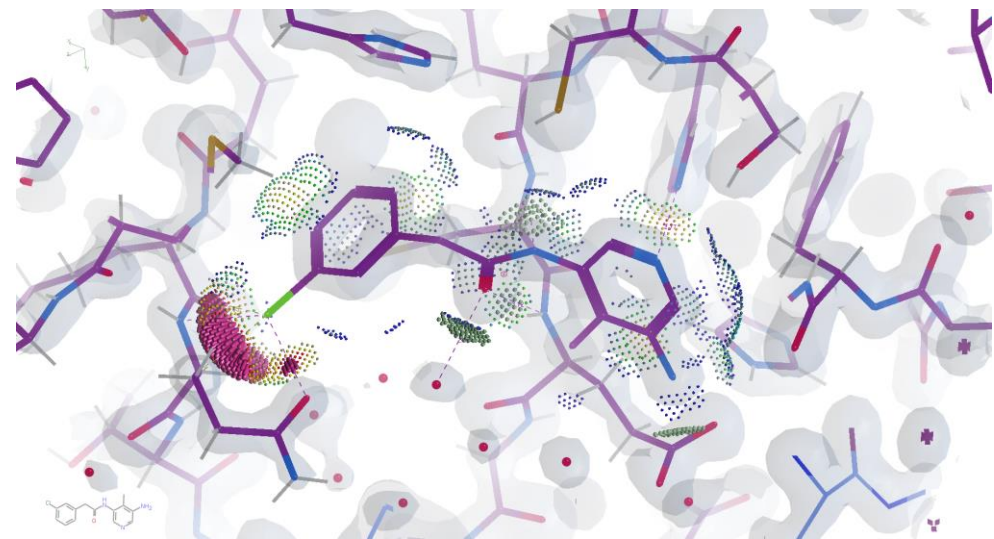
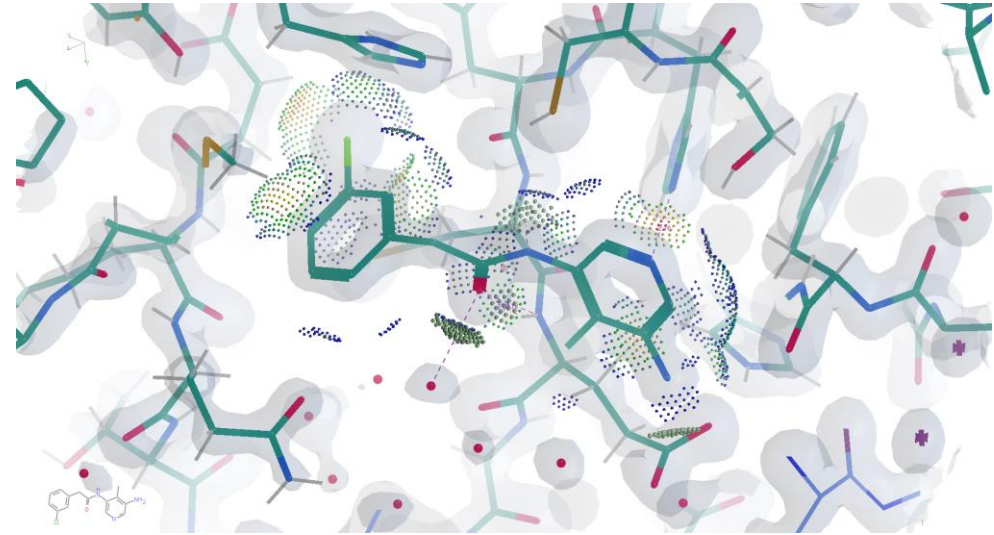
Ligand validation - density correlation



Resolution

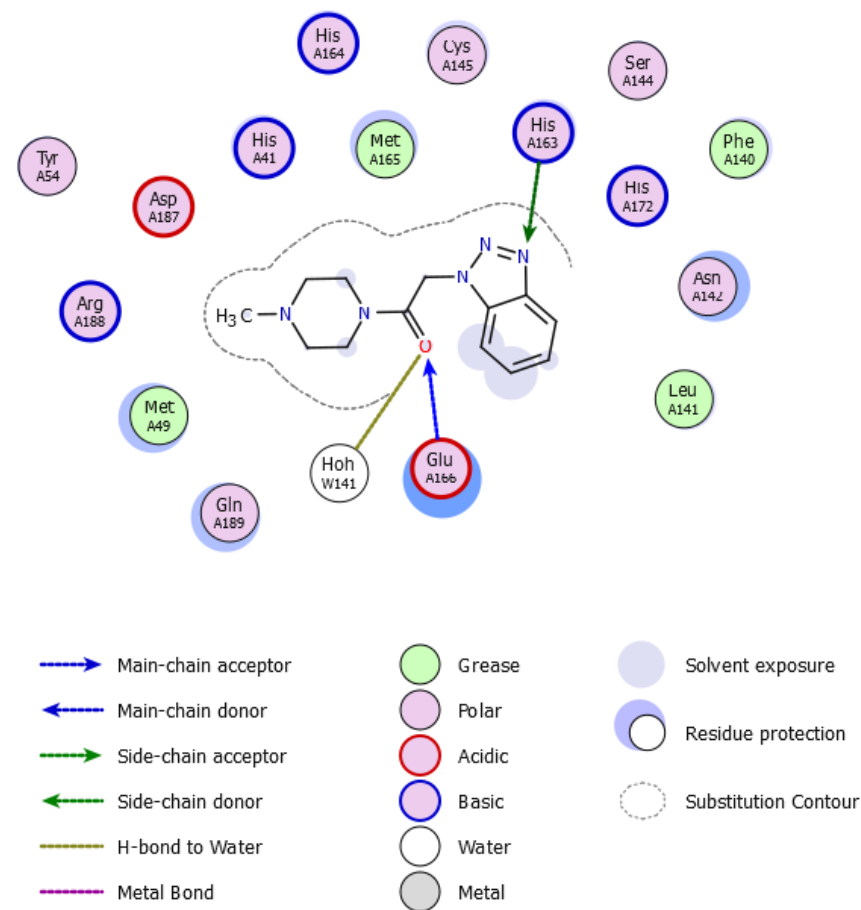
Ligand validation - interactions and clashes

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



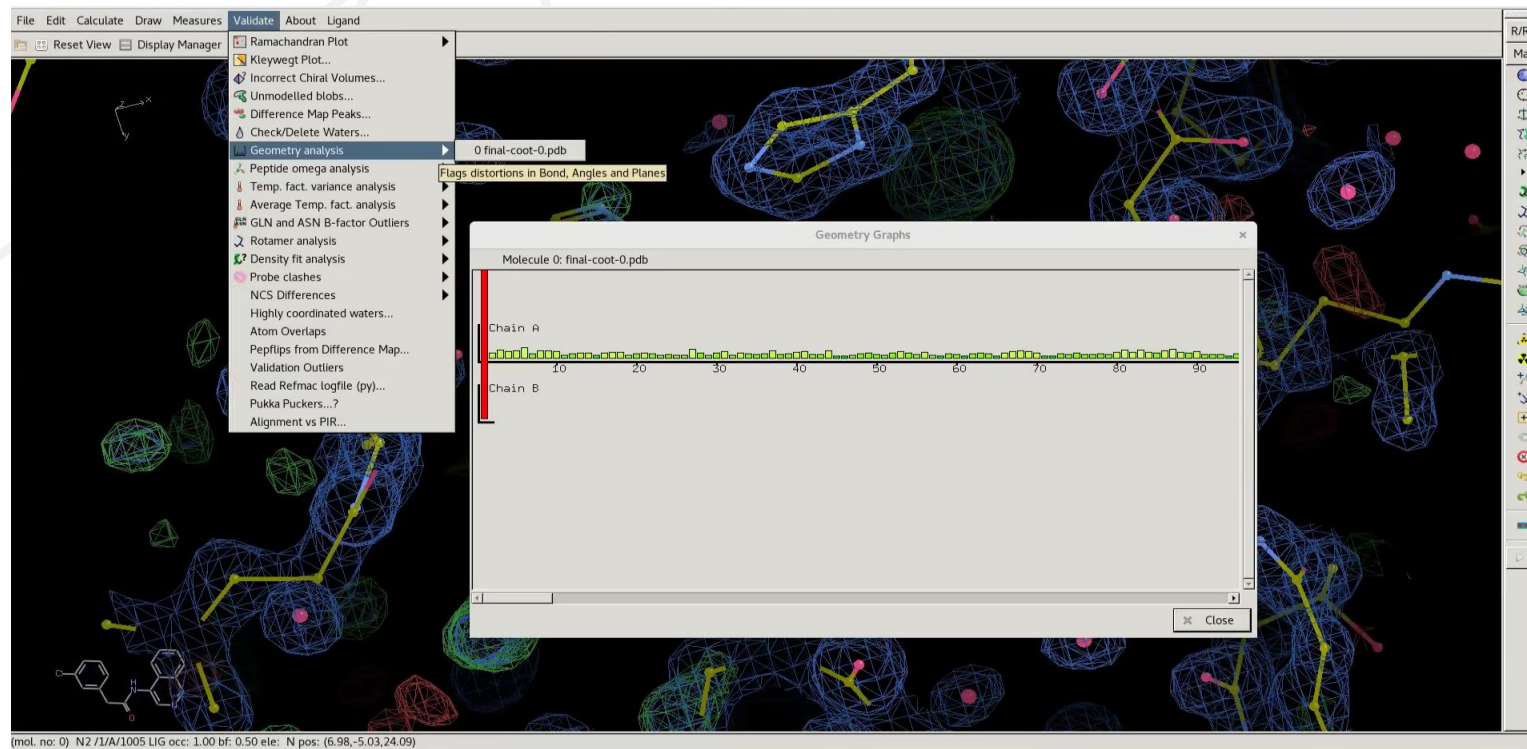
Ligand validation – Ligand environment (FLEV)

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



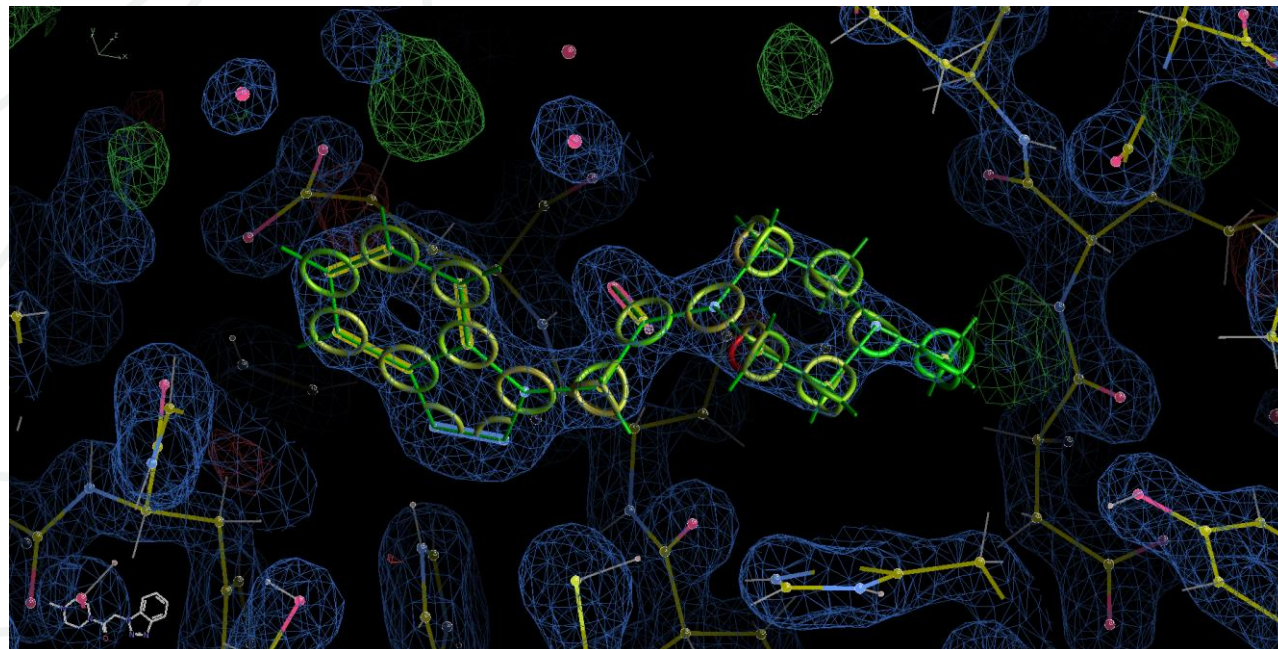
Ligand validation - geometry

- Compare modelled structure to restraints
 - Validate -> Geometry analysis
 - Only shows agreement with restraints file, doesn't flag actual errors



Ligand validation - geometry

- Highlights bonds, angles, or stereochemical features deviating from expected values.
 - Ligand -> Display Ligand Distortions
 - Only shows agreement with restraints file, doesn't flag actual errors



Ligand validation - geometry

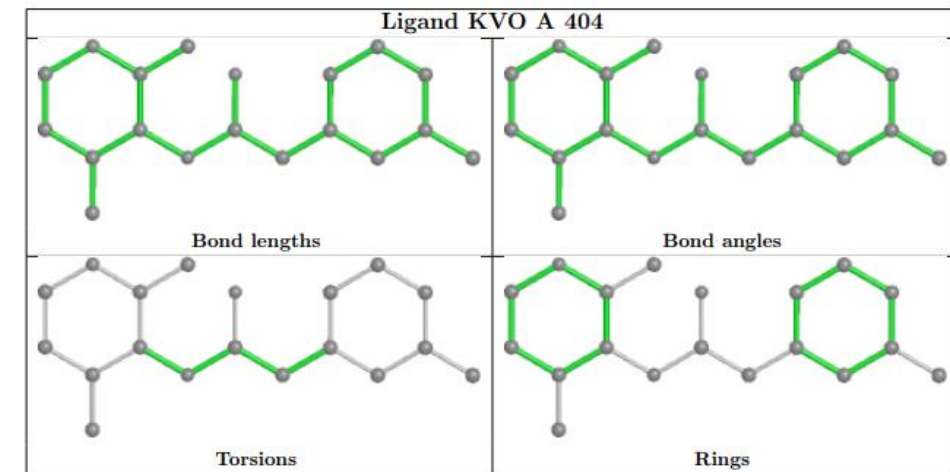
- Comparison with databases:
 - Useful for common ligands (nucleotides, cofactors, etc)
 - Less useful for novel ligands – few examples exist, if any
 - Need to confirm quality of database entry first
- Mogul (CCDC)
 - Compares geometries against CSD - database of small-molecule crystal structures
 - Useful cross-validation if restraints not generated using CSD data (e.g. AceDRG)
- Mogul Z-scores available as part of PDB Validation Report (and in CCP4 Cloud)
 - 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>

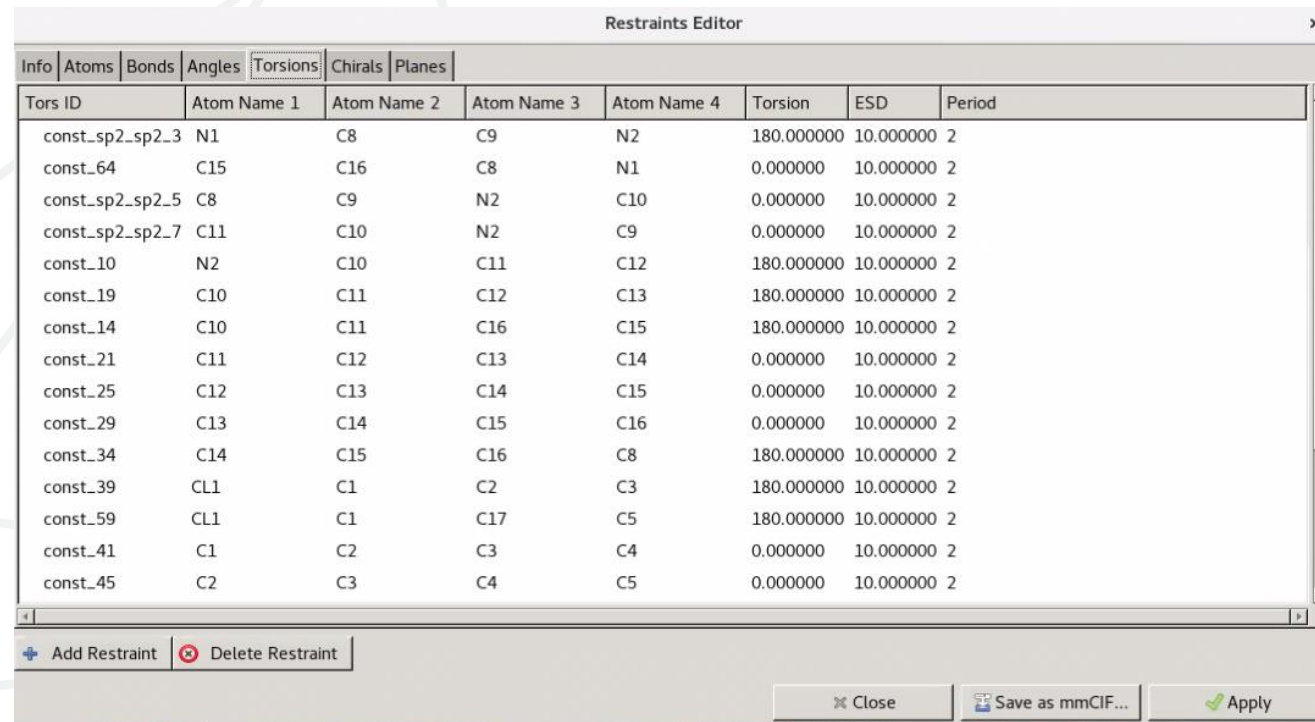
<https://doi.org/10.1021/ci200439d>



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	DMS	A	403	-	3,3,3	0.18	0	3,3,3	0.53	0
3	KVO	A	404	-	20,20,20	0.28	0	25,27,27	0.26	0
2	DMS	A	402	-	3,3,3	0.31	0	3,3,3	0.27	0
2	DMS	A	401	-	3,3,3	0.61	0	3,3,3	0.27	0

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?**