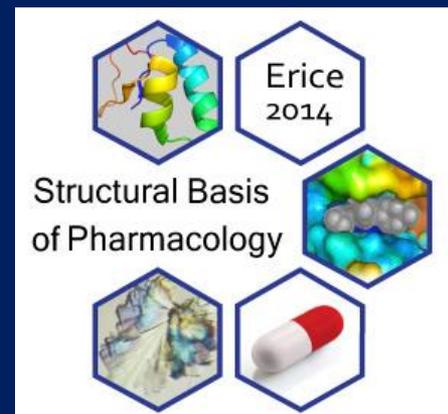


# Achieving high quality protein-ligand X-ray structures for drug design

Oliver Smart  
Global Phasing Ltd

GΦL

*Structural Basis of Pharmacology:  
Deeper Understanding of Drug  
Discovery through Crystallography  
Erica 4<sup>th</sup> June 2014*



# Structure-guided drug discovery

- Relies on X-ray structure of protein ligand complex (XSPLC)
- Obtaining such a structure involves many stages.
- Mistakes made in any of the stages will adversely affect results.
- PDB is a data bank of “complete” structures
- Despite this, the PDB has entries showing mistakes for each stage.
- Users of XSPLCs should have critical awareness

# Steps involved in solving a XSPLC

See page 135 of the lecture notes for list of steps:

- a) Experimental X-ray data collection from a protein crystal with the ligand soaked or co-crystallized using synchrotron or an in-house diffractometer.
- b) Image integration and data processing to give space group, unit cell and structure factor amplitudes (SF).
- c) Molecular replacement to reposition the protein model for the cell and SF from **(b)**.
- d) Initial refinement of model from **(c)** without a ligand.
- e) Assessment of whether the difference electron density (ED) for the model from **(d)** supports placing a ligand.
- f) Production of a molecular model and a restraint dictionary for the ligand.
- g) Fitting the model of ligand **(f)** into difference density and protein model from **(d)**.
- h) Refinement of combined protein and ligand model.
- i) Assessment of refined protein-ligand complex **(h)**.
- j) If assessment shows issues then rebuild/refit protein, ligand and/or solvent and back to step **(h)**.
- k) Deposition of the structure model, SF, maps and validation data to an in-house database (or the PDB).

We will look at each one in turn

Slide are marked **step (a)**, **step (b)** ...

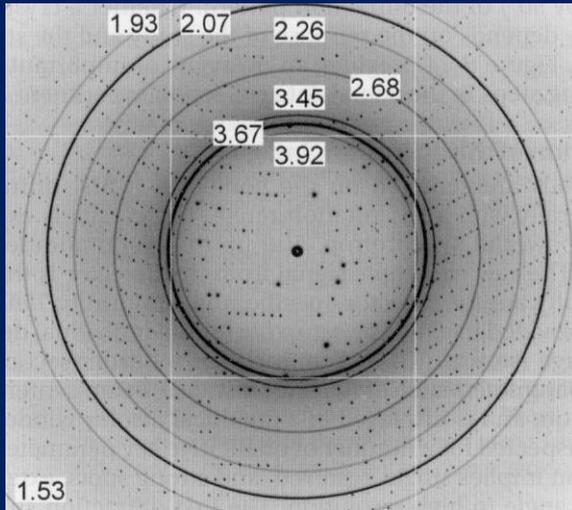
# Crystallization/soaking

Often ligands are dissolved in nasty solvents like DMSO. Using this to soak protein crystals can cause:

- Crystal cracking
- Unit-cell changes
- Reduction in the resolution limits of measurable diffraction data ...

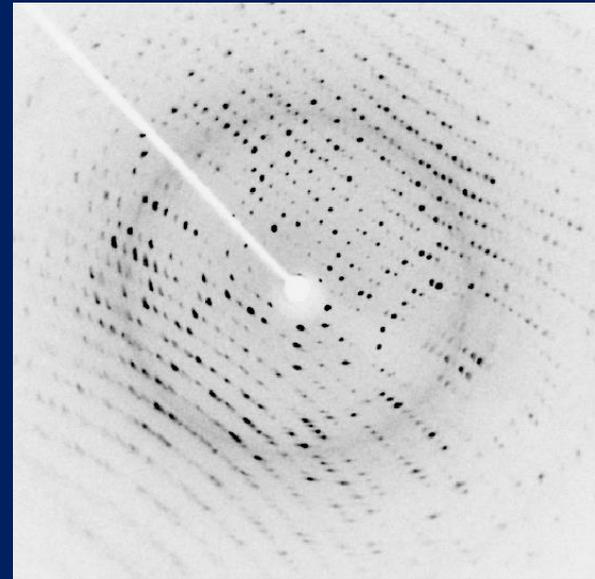
# X-ray data resolution

The further out spots go, the more information. Measure this in Å data resolution:



High resolution data but with “ice rings” (Rupp figure 8-25)

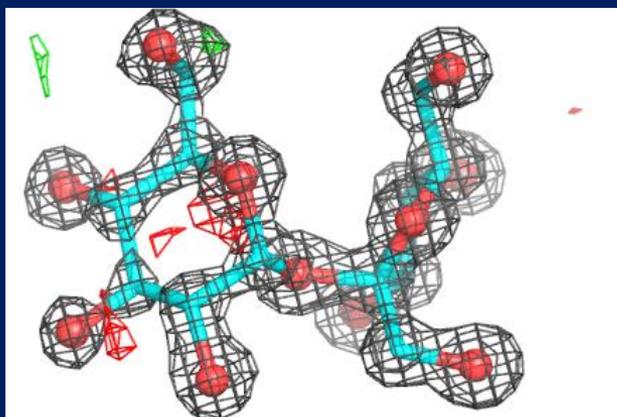
In practice diffraction is often more limited:



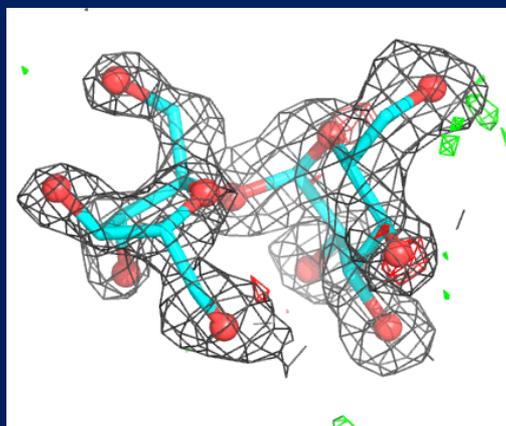
[http://commons.wikimedia.org/wiki/File:Lysozym\\_diffraction.png](http://commons.wikimedia.org/wiki/File:Lysozym_diffraction.png)

# Data resolution limit affects ligand electron density detail

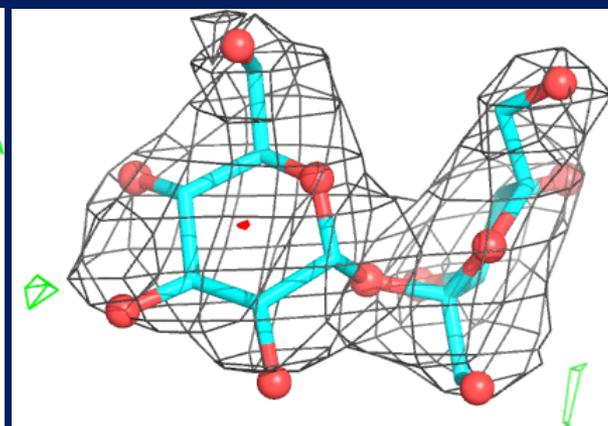
Well placed/refined sucrose ligand at different data resolutions:



**1ylt 1.2Å resolution**  
*“atomic resolution”*  
*2Fo-Fc ED shows atoms as individual blobs. Need higher resolution for hydrogen atoms*



**2pwe 2.0Å resolution**  
Typical medium resolution for ligand studies. Can see ring pucker

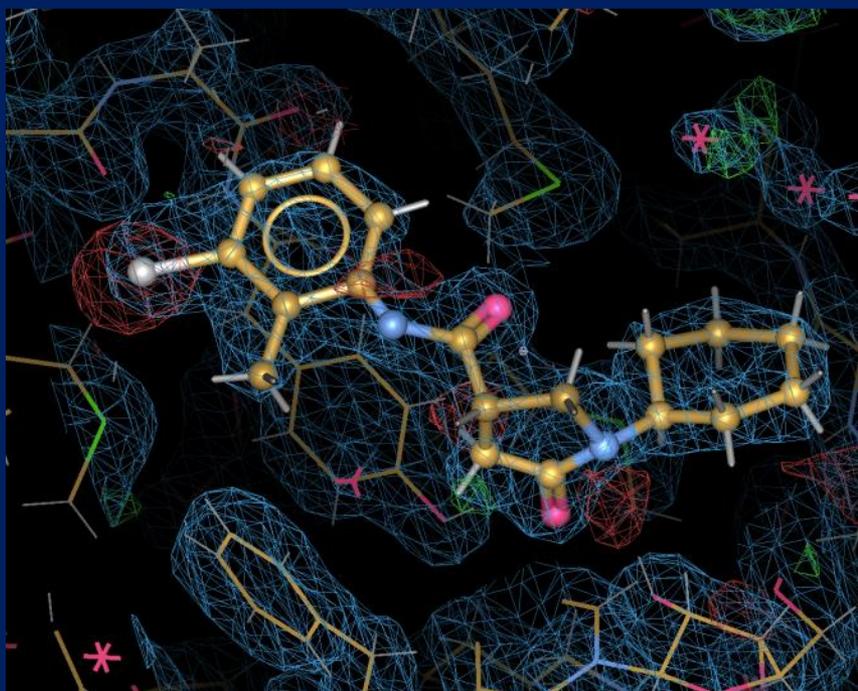


**2qqv 3.0Å resolution**  
Low resolution. Ligand placement unambiguous but fine detail cannot be seen

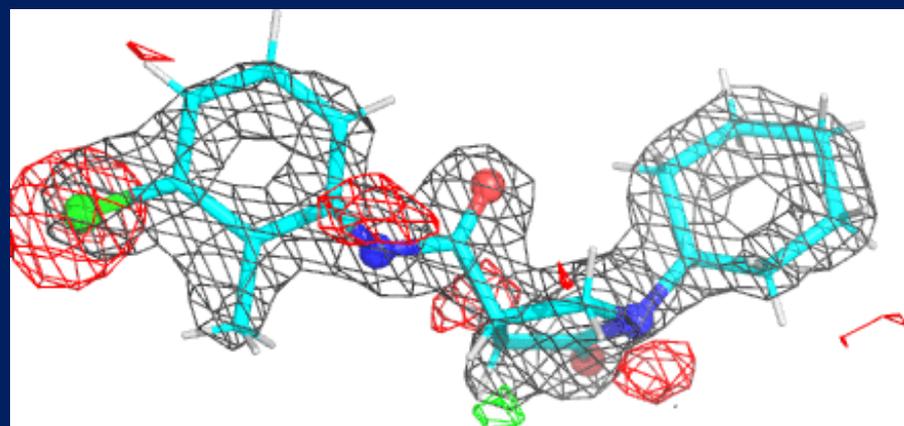
# Electron density (ED)

- ED maps are as important as the model:
- 2Fo-Fc map indicates where electrons are (according to SF and model). Normally color blue or grey.
- Fo-Fc difference map:
  - green for positive difference: where the current model fails to place sufficient electrons
  - Red for negative difference: where the current model places too many electrons

# Ligand electron density



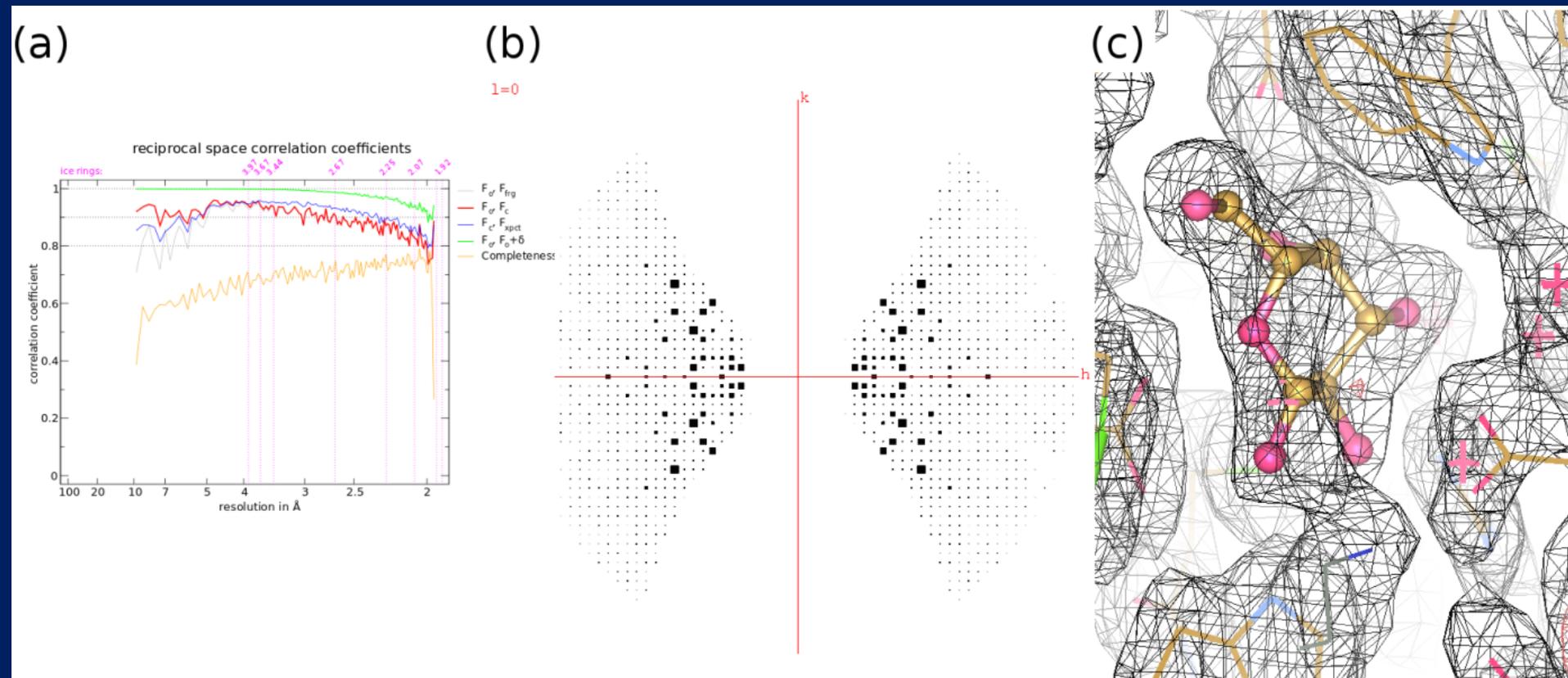
2h7p.pdb: ED visualized in coot  
Notice difference density around ligand



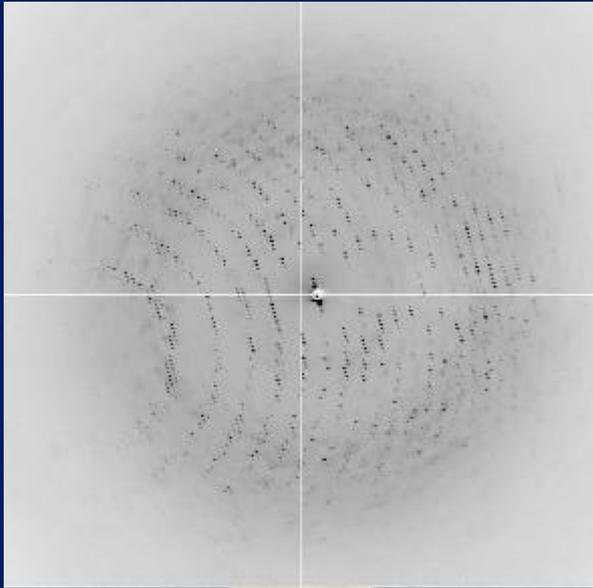
2h7p.pdb: ED around ligand, as visualized in  
buster-report

# Data collection

- Mistakes here are serious (uncorrectable)
- For example 1t0o



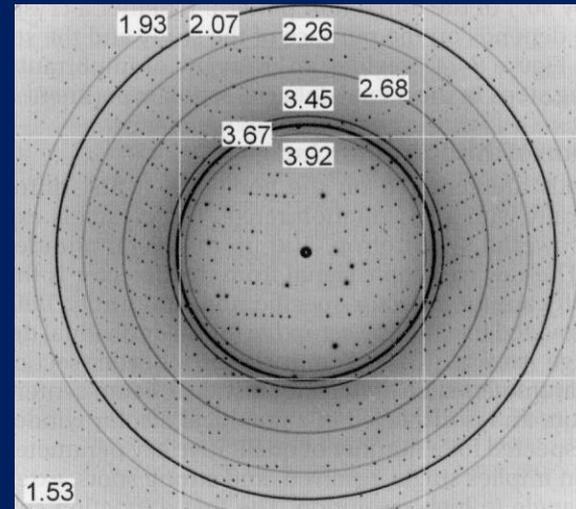
# Data processing and integration



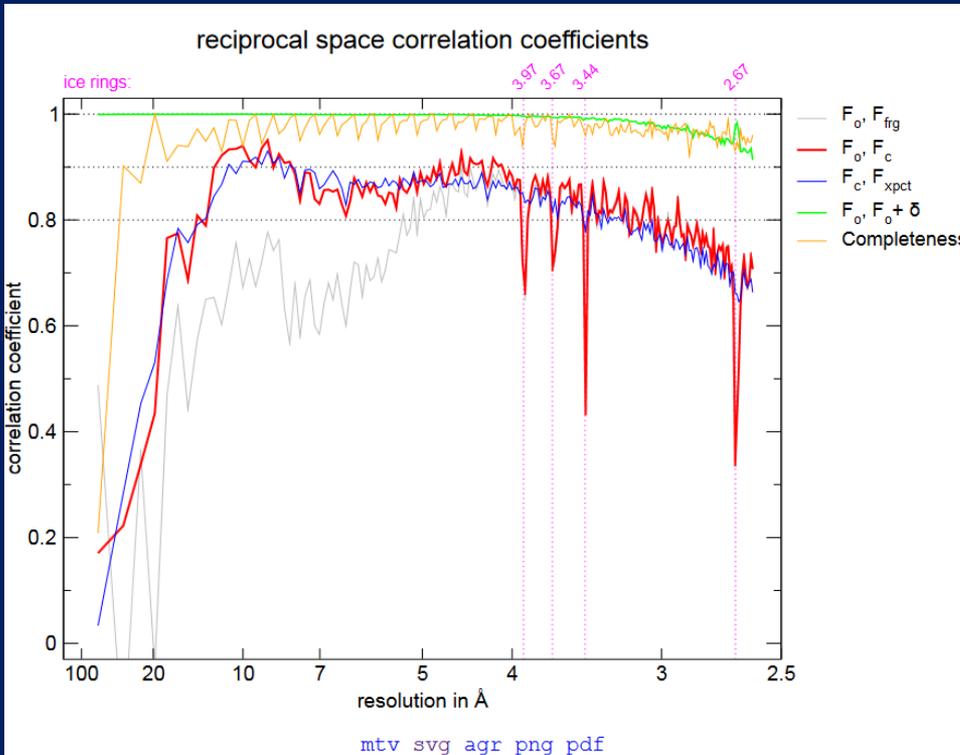
- GΦL tool [autoPROC](#) provides a robust, intelligent tool to go from images to intensities
- Uses XDS, ccp4 tools, GΦL tools ...
- Important to exclude ice rings

LIST OF REFLECTIONS  
=====

0	0	3	11433.95	470.53	11512.56	844.61	11398.58	566.59
			1067.97	23.85	0.00	0.00	1069.75	39.60
			1066.19	26.61	0.00			
0	0	6	3224.50	133.67	3097.33	182.53	3371.56	196.28
			567.80	11.81	0.00	0.00	555.75	16.45
			579.84	16.96	0.00			
0	0	9	41700.96	1612.14	39626.43	2283.65	43761.96	2276.19
			2036.07	39.70	0.00	0.00	1985.15	57.63
			2086.98	54.62	0.00			
0	0	12	6319.44	363.28	?	?	6319.44	363.28
			793.41	22.94	0.00	0.00	?	?
			793.41	22.94	2.00			
0	0	15	33102.28	2280.71	33102.28	2280.71	?	?
			1811.78	63.11	0.00	0.00	1811.78	63.11
			?	?	1.00			



# Data processing problems revealed in BUSTER RecipSCC plot



help? see BUSTER wiki [interpreting reciprocal-space CC plots](#)

- Shows up ice rings
- Shows up problems with beam stop
- Shows up anisotropic diffraction
- We've tried to persuade people to look at it for years
- Now buster-report makes it hard to ignore!

# Phasing or Molecular Replacement

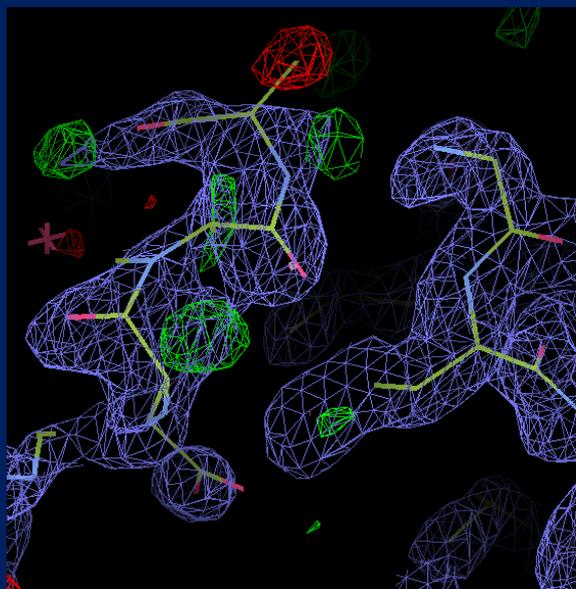
- For simple ligand soaks this is usually not a problem
- But sometimes unit-cell changes can be large
- The GΦL Pipedream tool automates limited molecular replacement using ccp4 Phaser

# Initial Refinement of protein before ligand placement

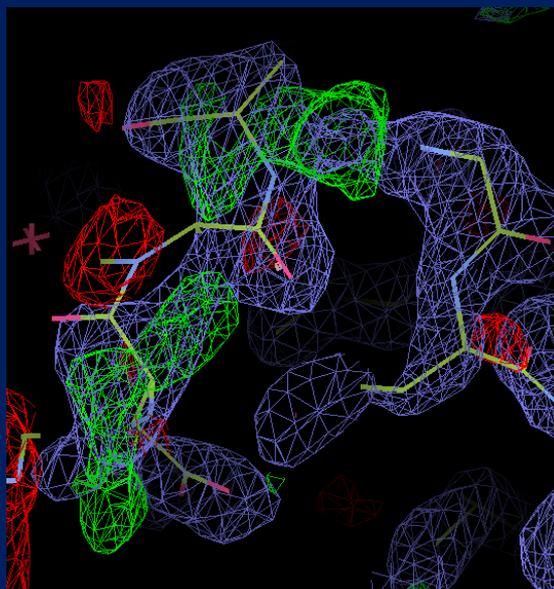
- Model is adjusted to better fit density
- This is crucial because ED maps improve as model is improved, so poor initial ligand ED becomes interpretable
- Maximum-likelihood refinement is now universally used.
- BUSTER maps are particularly good.

# BUSTER maps are bias-resistant

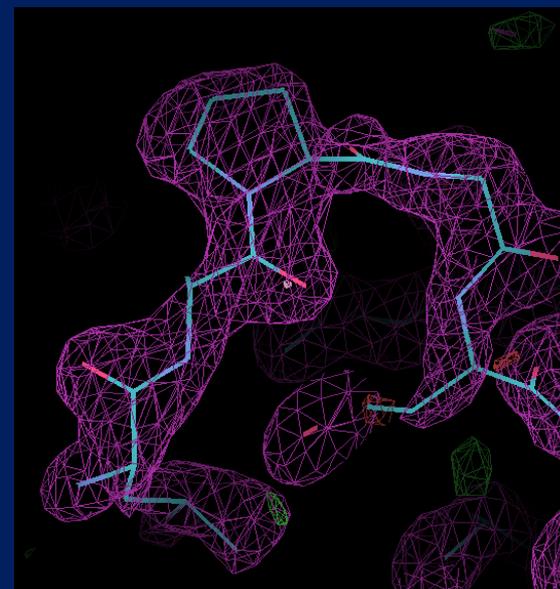
Electron  
Density  
Server



BUSTER  
maps  
2wfw.pdb



BUSTER  
rebuilt &  
re-refined model



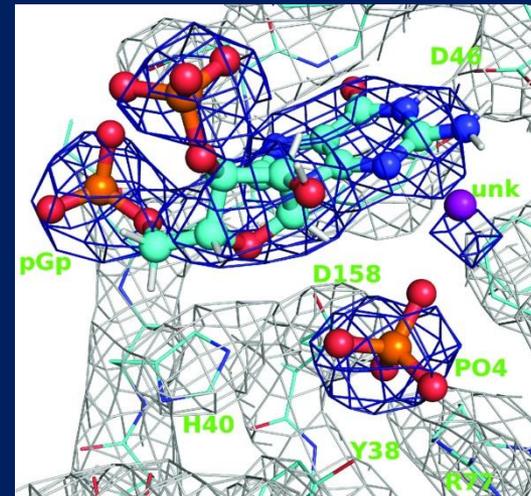
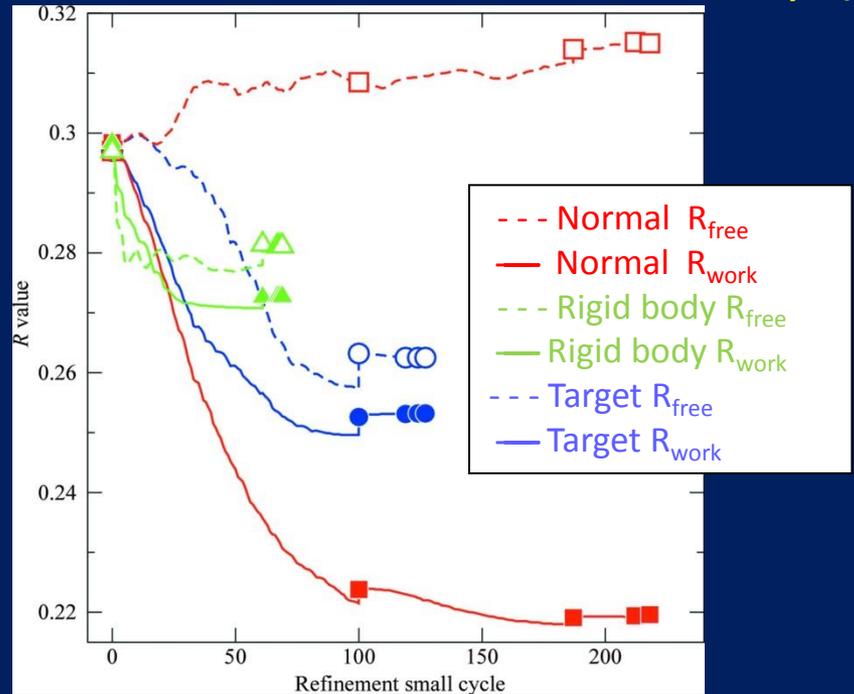
Example 2wfw 1.6Å resolution

# 2wfw example 1.6Å resolution

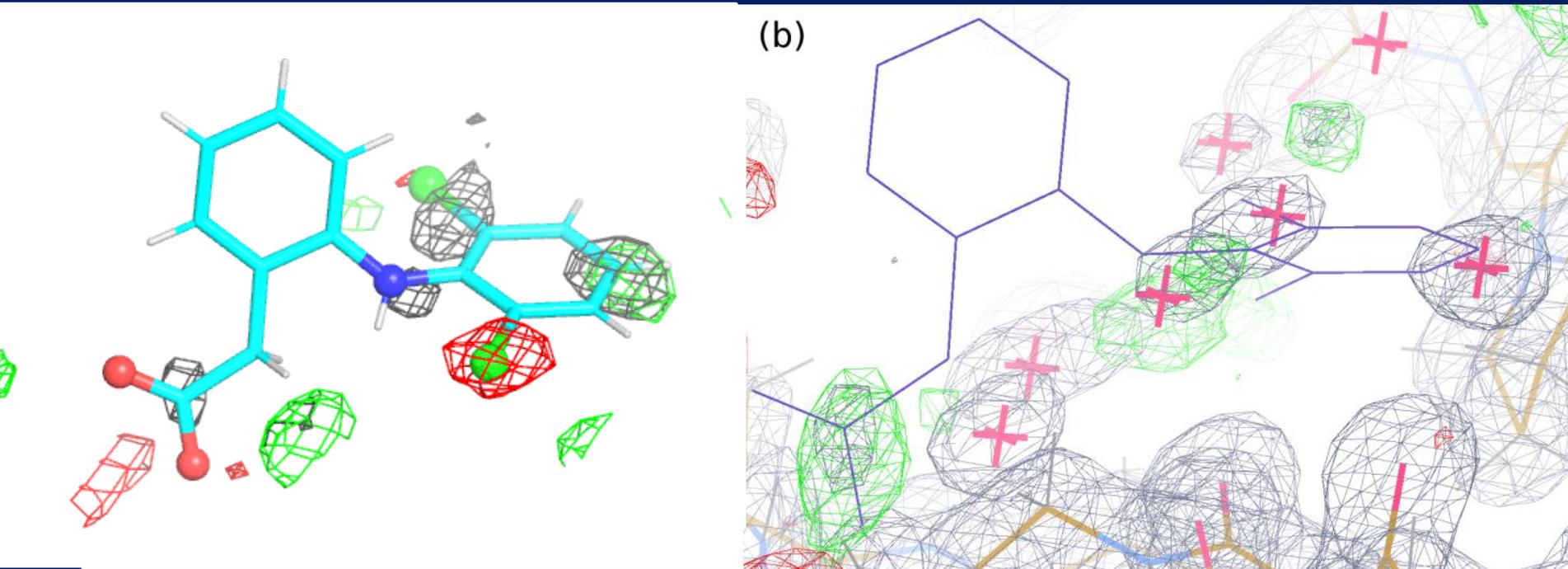
	As deposited	After default refinement	After remediation
Rwork	0.210	0.228	0.193
Rfree	0.246	0.263	0.215
Rotamer outliers %	4.70	3.69	1.78
Ramach. outliers %	3.32	3.88	0.29
Ramach. favoured %	94.5	93.35	98.3
Molprobit score	2.55	2.27	1.28
Molprobit percentile	12	26	97

# Target restraints

- Use similarity restraints “LSSR” to an external fixed PDB structure.
- Useful for low resolution ligand complex with higher resolution apo/other ligand
- [3urp](#), 3.2Å resolution poor data, [Acta Cryst D \(2012\) 68:368-380](#)



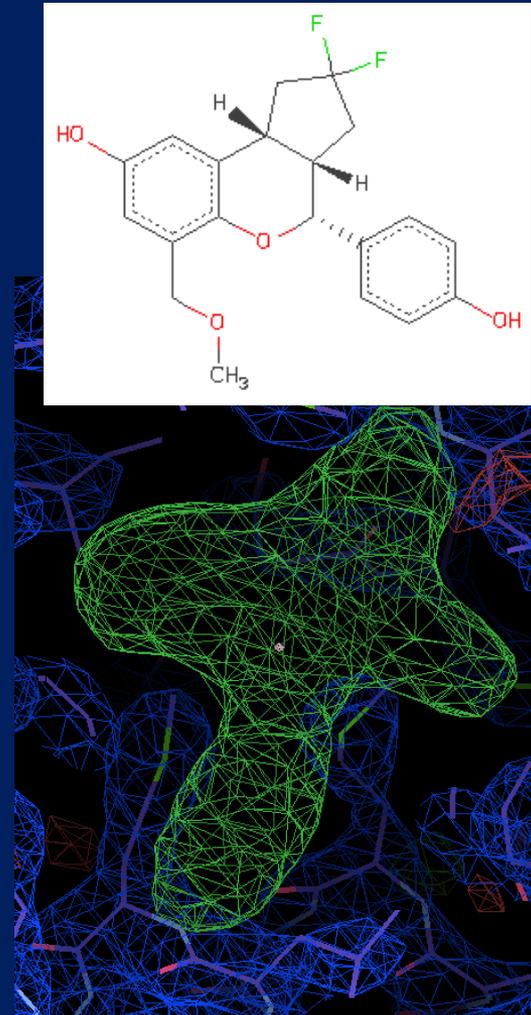
After initial refinement, need to assess ED difference density: is there any ligand bound?



**Figure 4.** [Pozharsk et al. \[2013\]](#) classify the diclofenac ligand in PDB entry [3IB0](#) (1.4 Å resolution) as “absent”. BUSTER-REPORT supports this classification: the ligand has high B-factors and a CC (2Fo-Fc) of 0.57 and as shown in panel (a) there is only a small amount of disconnected ED around it. Panel (b) shows COOT image of the result of BUSTER re-refinement of the protein after the diclofenac (purple “ghost”) has been removed. The re-refinement included automated water placement and this shows the ED can be well modeled by three water molecules (red crosses) that form good hydrogen bonds.

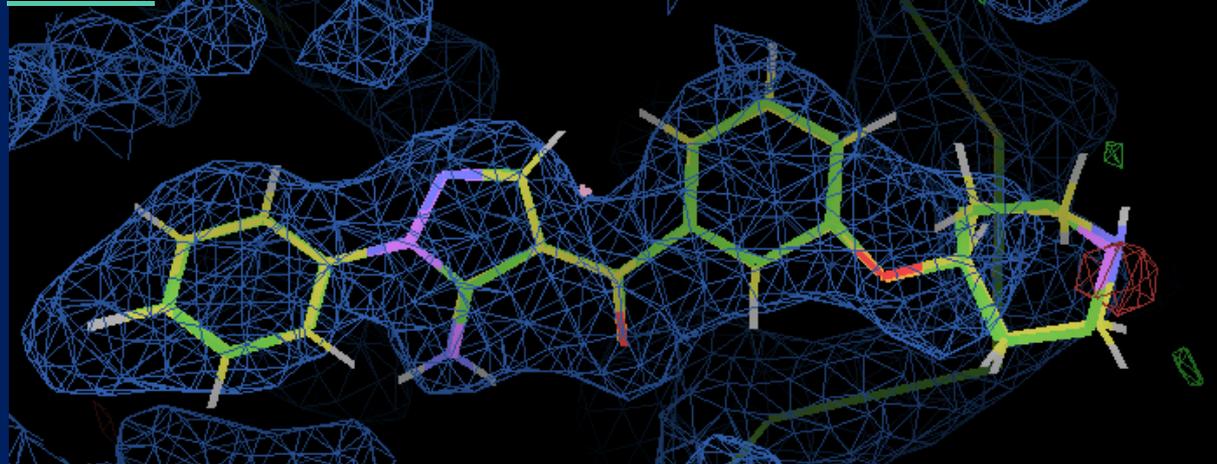
# Fitting a protein-bound ligand

- The task is to fit ligand into  $F_o - F_c$  and then refine complex
- Prior knowledge of ligand chemistry is needed to interpret density
- Use this prior knowledge
  - in ligand fitting step (g) to assess accessible low-strain conformations that the ligand can adopt
  - in refinement step (h) to keep ligand conformation realistic

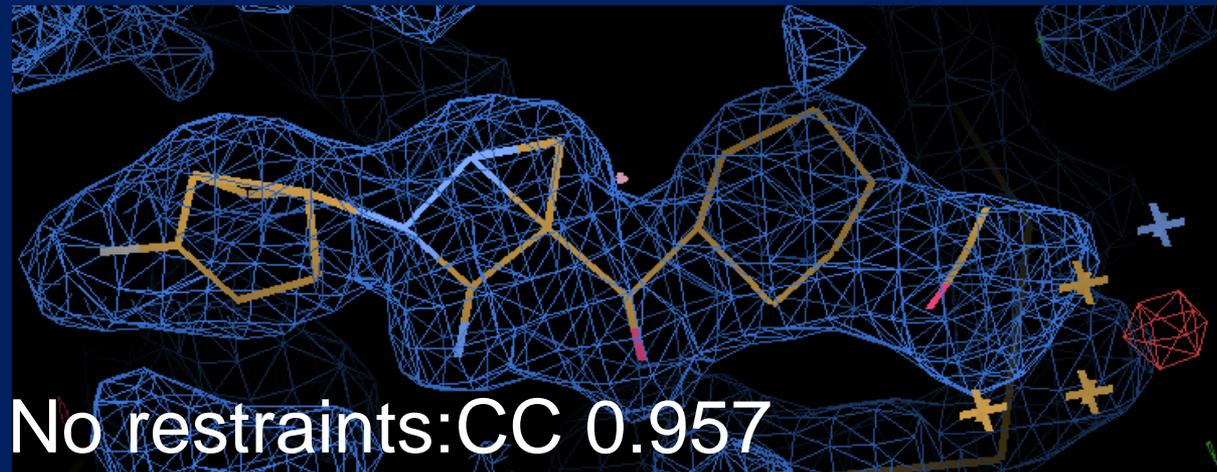


# Prior knowledge is essential!

2bal kinase 2.1Å resolution.

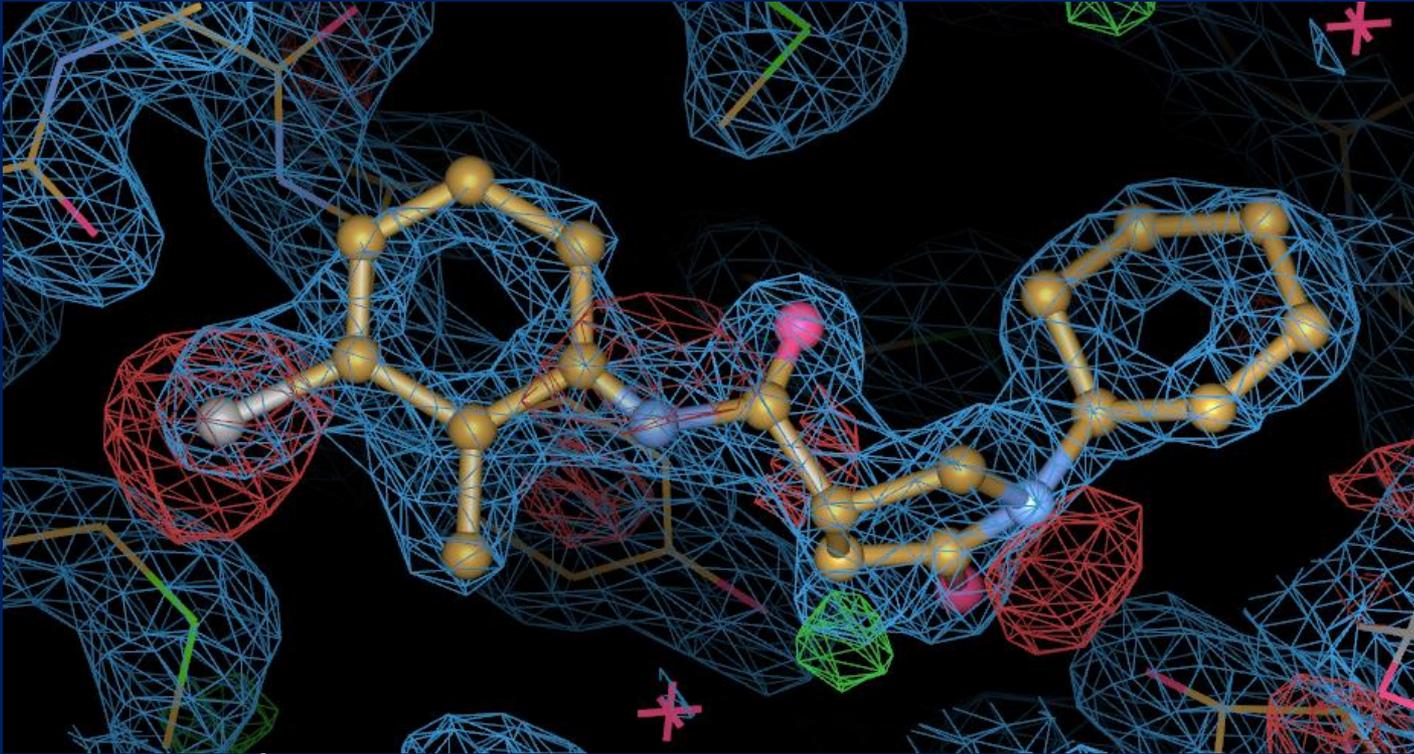


Restraints: 2Fo-Fc real space CC 0.943



No restraints: CC 0.957

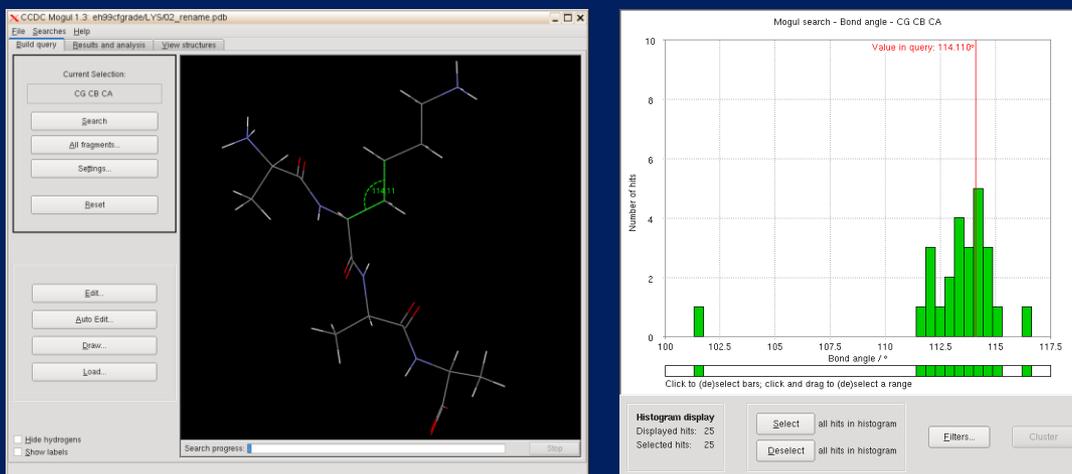
# 2h7p.pdb shows how the use of a poor dictionary leads to poor fit



- [2h7p.pdb](#) 1.86Å resolution shows nice density
- But ligand has poor geometry
  - 5-membered saturated lactam ring is flat while it should be puckered
  - Bent amide group
  - “Unusual” cyclohexyl group conformation
  - Difference density shows X-ray data are unhappy with the modeled ligand<sup>20</sup>

# grade: ligand dictionaries based on CSD information where possible

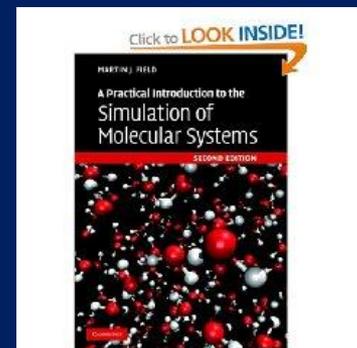
- Use CCDC [Mogul](#) program to survey the [CSD](#) small-molecule structure database



- [GFL grade](#) uses mogul in batch mode!
- Mogul is used as a *source* of information for restraints – *not* for validation
- Uses CSD dihedral information for restraints on clear sp<sup>2</sup>-sp<sup>2</sup> or sp<sup>3</sup>-sp<sup>3</sup> torsions

# What if CSD has nothing to say?

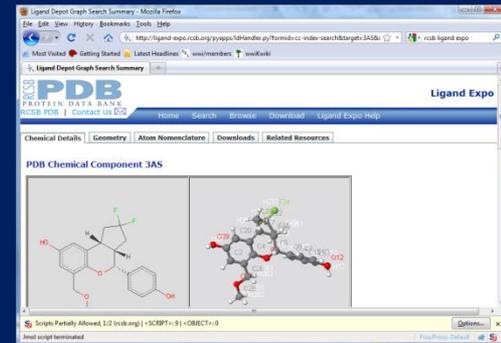
- Use quantum chemistry
  - as pioneered by eLBOW
  - normally use RM1 semi-empirical method as implemented in [dynamo](#) (Martin Field, IBS Grenoble)
- Only used for particular restraints where Mogul does not provide CSD data.
  - Angles to H atoms (bond lengths taken from neutron data)



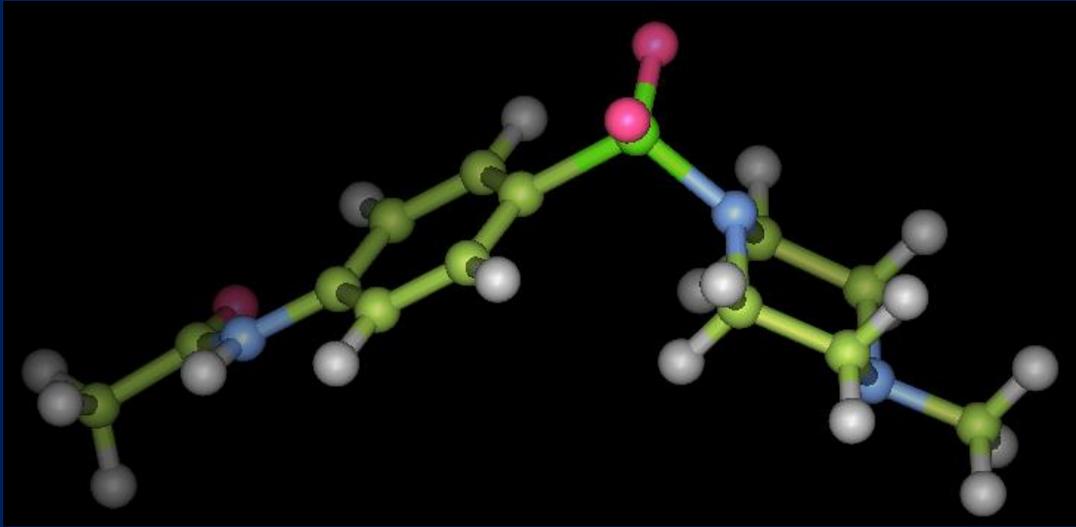
*pDynamo*

# grade tool

- Needs [CSDS](#) installation
- Input:
  - mol2 coordinates
  - cif dictionary from other dictionary generators
  - smi smiles string *uses libcheck*
- **grade\_PDB\_ligand**
  - tool for existing PDB ligands
  - fetches info from [pdbechem](#) or [ligand\\_expo](#)
  - produces grade dictionary
- Produces cif restraint dictionary for use in
  - BUSTER and rhotit
  - coot
  - refmac
  - ....
- In current BUSTER academic & consortium releases



# Let's look at one ligand-like example from CSD: EVIDUI



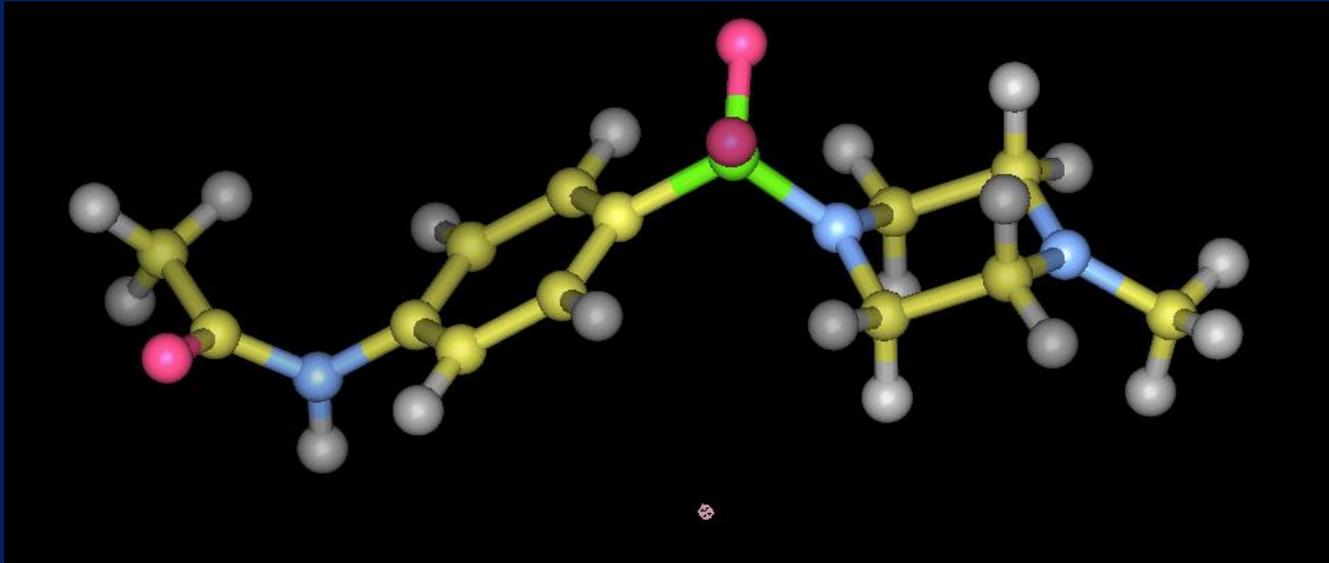
S(=O)(=O)(N1CCN(CC1)C)c1ccc(NC(=O)C)cc1



grade-XXX.cif

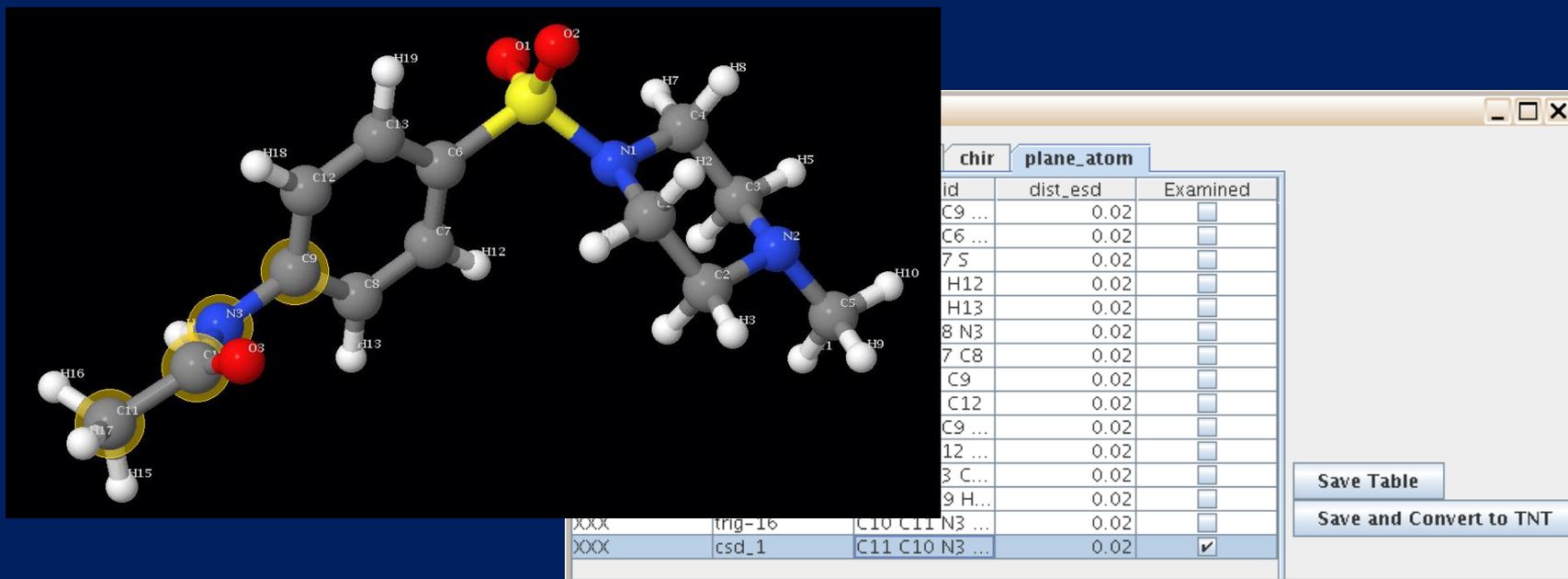
- Like ½ viagra with a amide attached
- CSD structure:
  - amide planar but not coplanar with phenyl ring
  - Both nitrogen atoms pyramidal
- Create dictionaries from smiles
- Score dictionary against CSD structure

# EVIDUI – 3<sup>rd</sup> party dictionary



- dictionary from X, ideal coords
- Test the dictionary by “scoring” the CSD structure
- piperazine nitrogens exactly planar but should be ~tetrahedral
- amide plane missed.
- rms bond deviation 0.048Å rms angle 5.0°

# EVIDUI grade with mogul



```
grade `S(=O)(=O)(N1CCN(CC1)C)c1ccc(NC(=O)C)cc1`
```

- Piperazine nitrogen atoms correct pyramidal
- Amide group: torsions restraints hold trans.
- rms bond 0.006Å rms angle 0.6°

# Grade Web Server



<http://grade.globalphasing.org>

- For good results grade requires mogul and so CSDS licence
- Grade Web Server
  - By kind permission of CCDC
  - Launched in May 2012
  - Provides easy access to Grade for non-confidential compounds
  - Provides GΦL with useful test examples for development
  - >3500 cif dictionaries produced to date (May 2014)
  - Dictionaries can be used in BUSTER, reftmac, phenix.refine

**GΦL** grade web server <http://grade.globalphasing.org>

---

Tom Womack ([twomack@globalphasing.com](mailto:twomack@globalphasing.com)) thank you for agreeing to the [conditions of use](#)

---

Please select an input type:

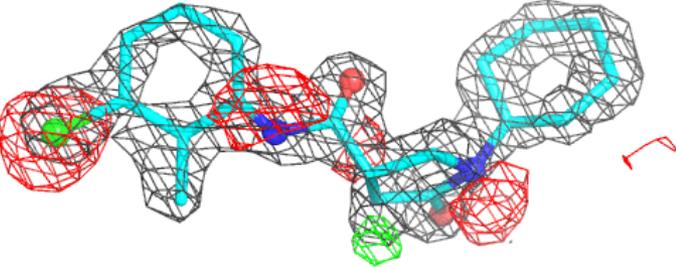
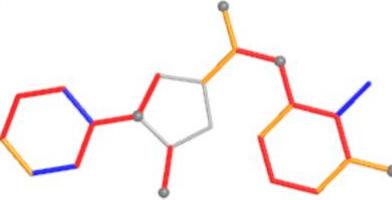
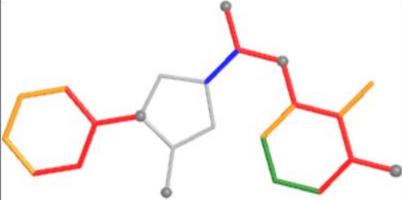
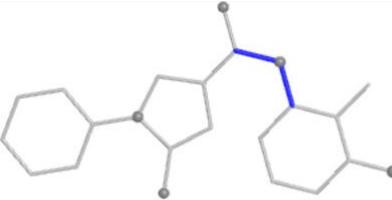
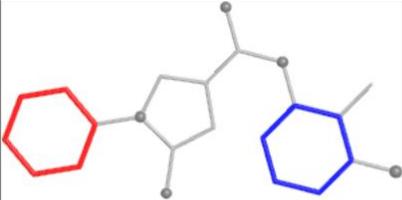
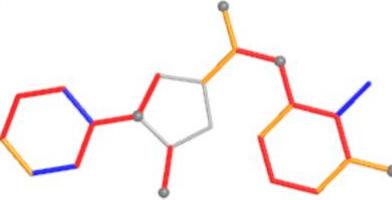
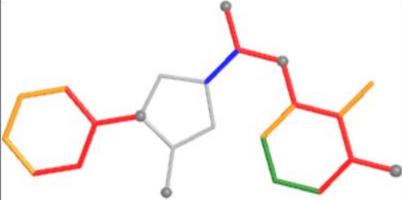
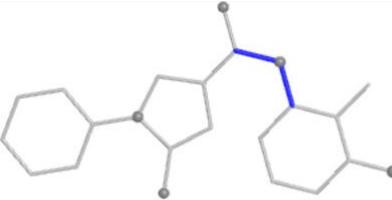
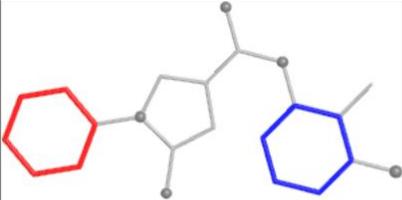
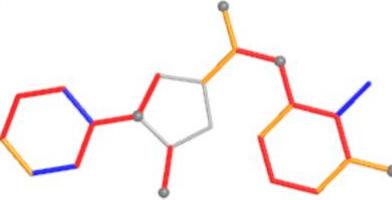
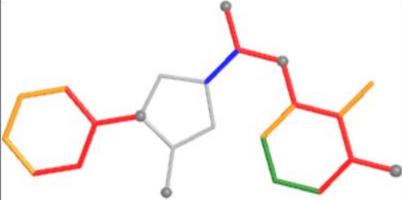
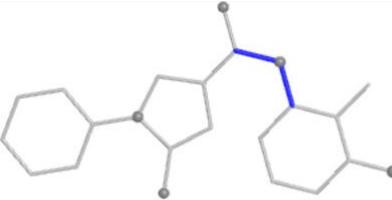
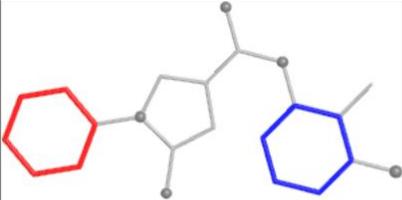
- SMILES string (for example from the Structure Editor at the [CACTVS Online SMILES Translator](#))
- Run grade on mol2 file (must have [hydrogen atoms](#). This option is useful for getting correct atom names [further details](#))
- Produce dictionary for an existing PDB chemical component (for example "ATP") [further details](#)

---

For help and further details see [Grade Web Server Help Page](#) on the BUSTER wiki.

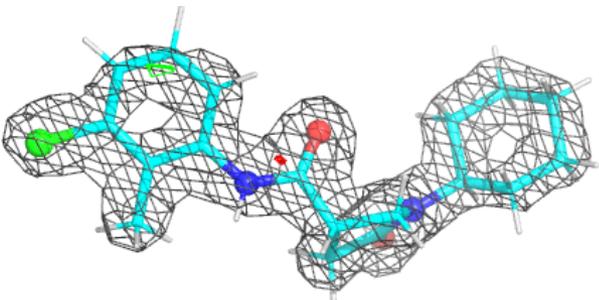
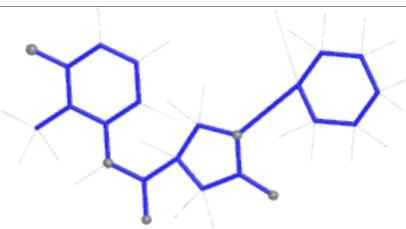
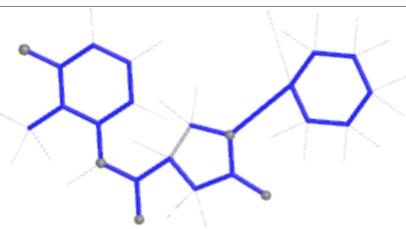
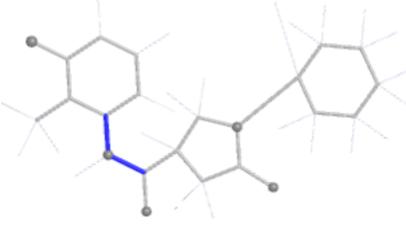
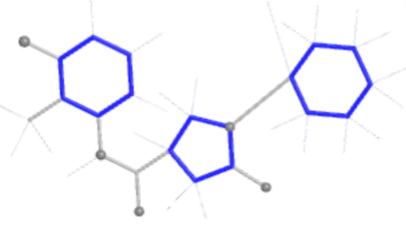
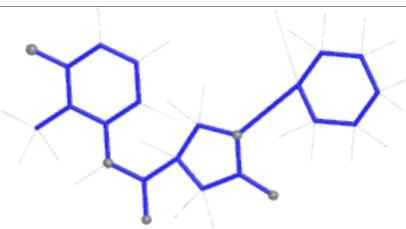
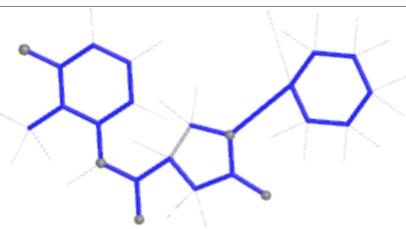
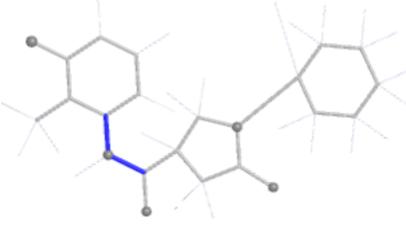
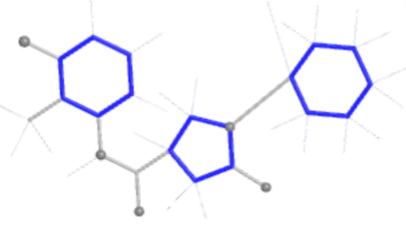
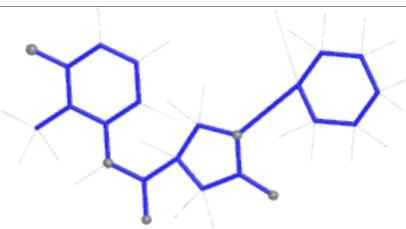
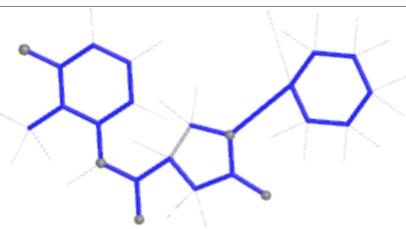
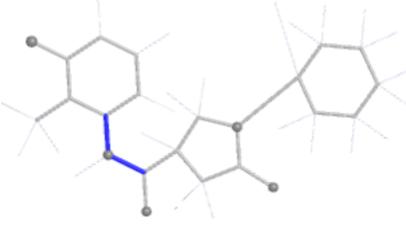
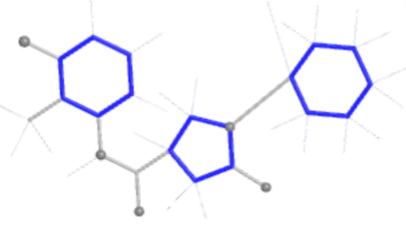
Please email [buster-develop@globalphasing.com](mailto:buster-develop@globalphasing.com) with any problems, queries or suggestions.

## 2h7p.pdb buster-report

Residue	Picture	Statistics	Geometry Analysis																																
468 A 501	 <p>Density around 468 A 501 : 2mF<sub>o</sub>-DF<sub>c</sub> (at 1.3 rnsd) in grey. mF<sub>o</sub>-DF<sub>c</sub> (at 3 rnsd) in red (negative) and green (positive). (picture produced with <a href="#">PyMOL</a>)</p>	<table border="1"> <tr> <td>CC(2F<sub>o</sub>-F<sub>c</sub>)</td> <td>0.887</td> </tr> <tr> <td>min(B-factor)<sup>‡</sup></td> <td>25.1</td> </tr> <tr> <td>avg(B-factor)<sup>‡</sup></td> <td>27.3</td> </tr> <tr> <td>max(B-factor)<sup>‡</sup></td> <td>31.1</td> </tr> <tr> <td>min(occ)<sup>‡</sup></td> <td>1.00</td> </tr> <tr> <td>max(occ)<sup>‡</sup></td> <td>1.00</td> </tr> <tr> <td colspan="2"><sup>‡</sup> hydrogen atoms excluded</td> </tr> <tr> <td colspan="2"><b>Mogul Analysis:</b></td> </tr> <tr> <td>"bad" bonds</td> <td>12/21</td> </tr> <tr> <td>"bad" angles</td> <td>7/22</td> </tr> <tr> <td>"unusual" dihedrals</td> <td>0/2</td> </tr> <tr> <td>"bad" rings</td> <td>1/2</td> </tr> <tr> <td>bonds rms Z</td> <td>9.6</td> </tr> <tr> <td>angles rms Z</td> <td>5.3</td> </tr> </table> <p><a href="#">Detailed report</a></p>	CC(2F <sub>o</sub> -F <sub>c</sub> )	0.887	min(B-factor) <sup>‡</sup>	25.1	avg(B-factor) <sup>‡</sup>	27.3	max(B-factor) <sup>‡</sup>	31.1	min(occ) <sup>‡</sup>	1.00	max(occ) <sup>‡</sup>	1.00	<sup>‡</sup> hydrogen atoms excluded		<b>Mogul Analysis:</b>		"bad" bonds	12/21	"bad" angles	7/22	"unusual" dihedrals	0/2	"bad" rings	1/2	bonds rms Z	9.6	angles rms Z	5.3	<table border="1"> <tr> <td>  <p><b>Mogul bonds</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p> </td> <td>  <p><b>Mogul angles</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p> </td> </tr> <tr> <td>  <p><b>Mogul dihedrals</b> "common" "unknown" <a href="#">Details</a></p> </td> <td>  <p><b>Mogul rings</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p> </td> </tr> </table>	 <p><b>Mogul bonds</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p>	 <p><b>Mogul angles</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p>	 <p><b>Mogul dihedrals</b> "common" "unknown" <a href="#">Details</a></p>	 <p><b>Mogul rings</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p>
CC(2F <sub>o</sub> -F <sub>c</sub> )	0.887																																		
min(B-factor) <sup>‡</sup>	25.1																																		
avg(B-factor) <sup>‡</sup>	27.3																																		
max(B-factor) <sup>‡</sup>	31.1																																		
min(occ) <sup>‡</sup>	1.00																																		
max(occ) <sup>‡</sup>	1.00																																		
<sup>‡</sup> hydrogen atoms excluded																																			
<b>Mogul Analysis:</b>																																			
"bad" bonds	12/21																																		
"bad" angles	7/22																																		
"unusual" dihedrals	0/2																																		
"bad" rings	1/2																																		
bonds rms Z	9.6																																		
angles rms Z	5.3																																		
 <p><b>Mogul bonds</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p>	 <p><b>Mogul angles</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p>																																		
 <p><b>Mogul dihedrals</b> "common" "unknown" <a href="#">Details</a></p>	 <p><b>Mogul rings</b> "bad" "poor" "ok" "good" "unknown" <a href="#">Details</a></p>																																		

- 2h7p.pdb 1.86Å resolution shows nice density
- But ligand has poor geometry
  - 5-membered saturated lactam ring is flat while it should be puckered
  - Bent amide group
  - “Unusual” cyclohexyl group conformation
  - Difference density shows X-ray knows better!

# 2h7p BUSTER re-refinement with grade dictionary fixes all problems

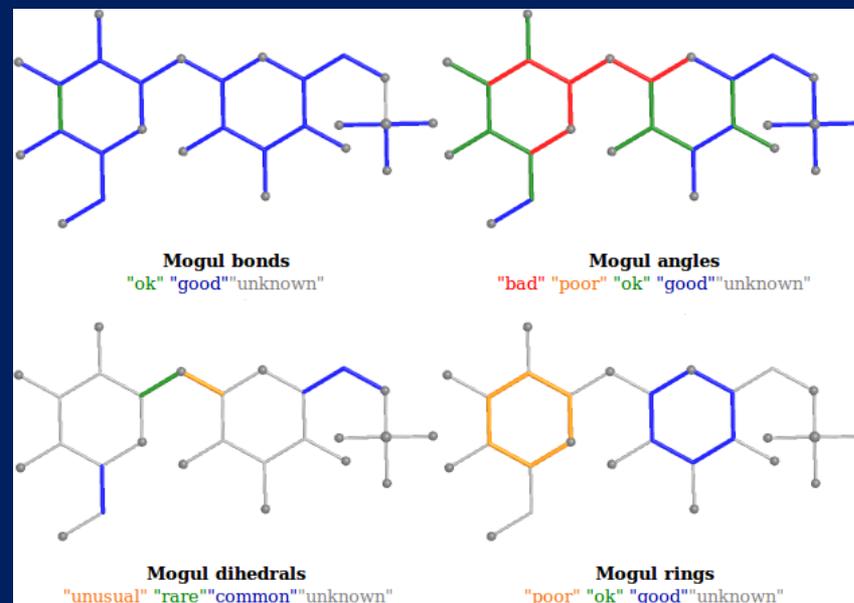
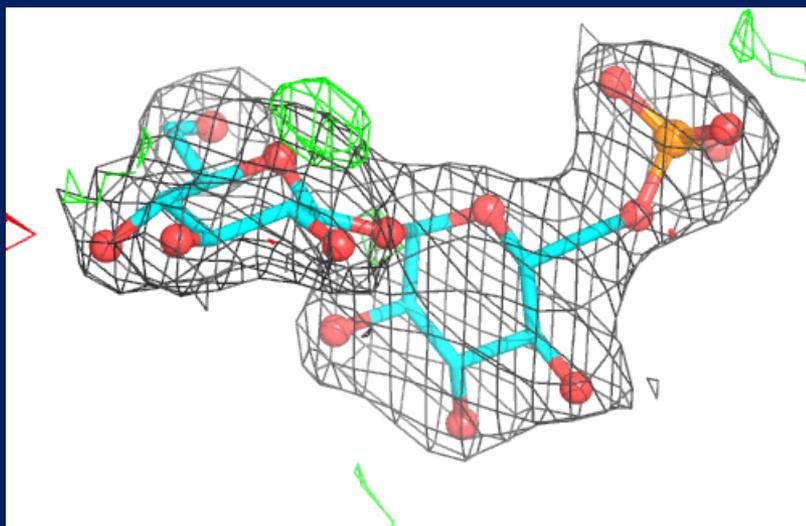
Residue	Picture	Statistics	Geometry Analysis																																
468 A 501	 <p><b>Density around 468 A 501 :</b>  <math>2mF_o - DF_c</math> (at 1.3 rmsd) in grey.  <math>mF_o - DF_c</math> (at 3 rmsd) in red (negative) and green (positive),            (picture produced with <a href="#">PyMOL</a>)</p>	<table border="1"> <tbody> <tr> <td><math>CC(2F_o - F_c)</math></td> <td>0.966</td> </tr> <tr> <td>min(B-factor)<sup>‡</sup></td> <td>16.9</td> </tr> <tr> <td>avg(B-factor)<sup>‡</sup></td> <td>24.7</td> </tr> <tr> <td>max(B-factor)<sup>‡</sup></td> <td>45.9</td> </tr> <tr> <td>min(occ)<sup>‡</sup></td> <td>0.87</td> </tr> <tr> <td>max(occ)<sup>‡</sup></td> <td>0.87</td> </tr> <tr> <td colspan="2"><sup>‡</sup> hydrogen atoms excluded</td> </tr> <tr> <td colspan="2"><b>Mogul Analysis:</b></td> </tr> <tr> <td>"bad" bonds</td> <td>0/25</td> </tr> <tr> <td>"bad" angles</td> <td>0/27</td> </tr> <tr> <td>"unusual" dihedrals</td> <td>0/2</td> </tr> <tr> <td>"bad" rings</td> <td>0/3</td> </tr> <tr> <td>bonds rms Z</td> <td>0.3</td> </tr> <tr> <td>angles rms Z</td> <td>0.5</td> </tr> </tbody> </table> <p><a href="#">Detailed report</a></p>	$CC(2F_o - F_c)$	0.966	min(B-factor) <sup>‡</sup>	16.9	avg(B-factor) <sup>‡</sup>	24.7	max(B-factor) <sup>‡</sup>	45.9	min(occ) <sup>‡</sup>	0.87	max(occ) <sup>‡</sup>	0.87	<sup>‡</sup> hydrogen atoms excluded		<b>Mogul Analysis:</b>		"bad" bonds	0/25	"bad" angles	0/27	"unusual" dihedrals	0/2	"bad" rings	0/3	bonds rms Z	0.3	angles rms Z	0.5	<table border="1"> <tbody> <tr> <td>  <p><b>Mogul bonds</b> "good""unknown" <a href="#">Details</a></p> </td> <td>  <p><b>Mogul angles</b> "good""unknown" <a href="#">Details</a></p> </td> </tr> <tr> <td>  <p><b>Mogul dihedrals</b> "common""unknown" <a href="#">Details</a></p> </td> <td>  <p><b>Mogul rings</b> "good""unknown" <a href="#">Details</a></p> </td> </tr> </tbody> </table>	 <p><b>Mogul bonds</b> "good""unknown" <a href="#">Details</a></p>	 <p><b>Mogul angles</b> "good""unknown" <a href="#">Details</a></p>	 <p><b>Mogul dihedrals</b> "common""unknown" <a href="#">Details</a></p>	 <p><b>Mogul rings</b> "good""unknown" <a href="#">Details</a></p>
$CC(2F_o - F_c)$	0.966																																		
min(B-factor) <sup>‡</sup>	16.9																																		
avg(B-factor) <sup>‡</sup>	24.7																																		
max(B-factor) <sup>‡</sup>	45.9																																		
min(occ) <sup>‡</sup>	0.87																																		
max(occ) <sup>‡</sup>	0.87																																		
<sup>‡</sup> hydrogen atoms excluded																																			
<b>Mogul Analysis:</b>																																			
"bad" bonds	0/25																																		
"bad" angles	0/27																																		
"unusual" dihedrals	0/2																																		
"bad" rings	0/3																																		
bonds rms Z	0.3																																		
angles rms Z	0.5																																		
 <p><b>Mogul bonds</b> "good""unknown" <a href="#">Details</a></p>	 <p><b>Mogul angles</b> "good""unknown" <a href="#">Details</a></p>																																		
 <p><b>Mogul dihedrals</b> "common""unknown" <a href="#">Details</a></p>	 <p><b>Mogul rings</b> "good""unknown" <a href="#">Details</a></p>																																		

- Grouped occupancy refinement sorts negative density for chlorine atom
- Very nice fit to density

# Importance of providing correct chemistry to the dictionary generator.

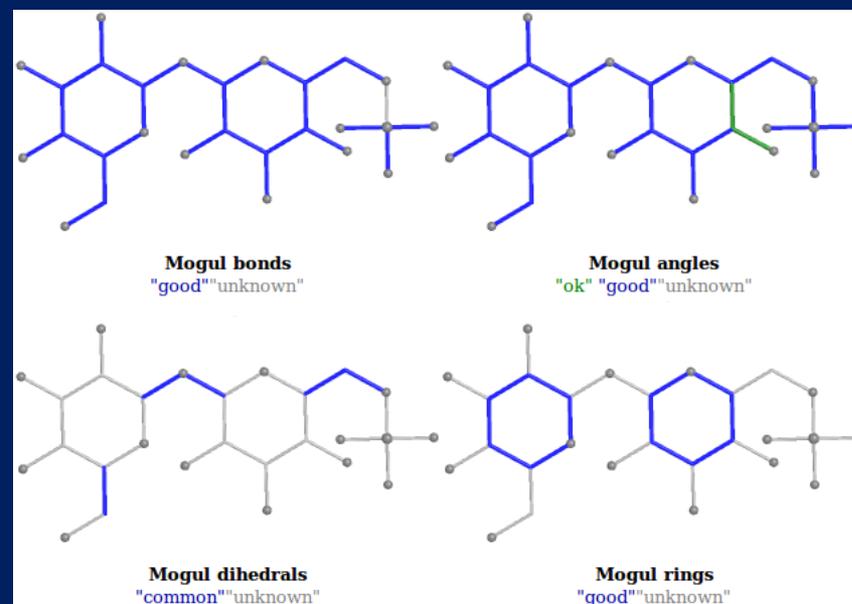
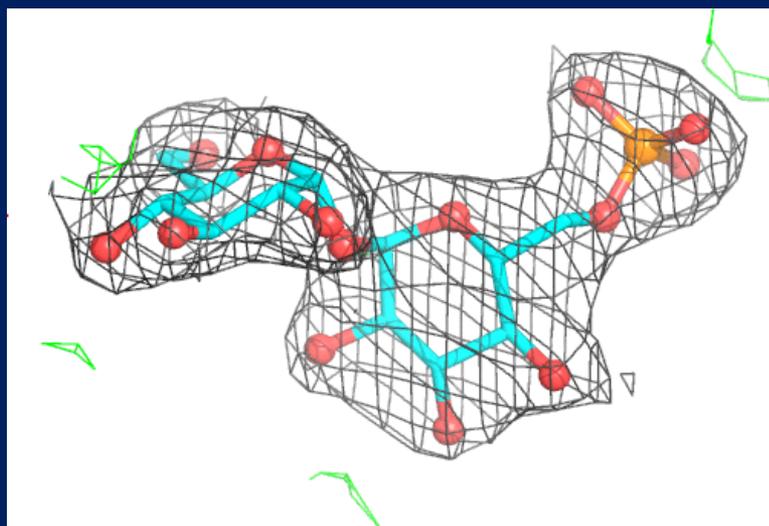
- [1byk.pdb](#) *E. coli* trehalose-repressor in complex with trehalose-6-phosphate.
- From [Kay Diederichs](#) University of Konstanz
- 2.5Å resolution structure solved in 1998. They used  $\beta$ -glucopyranose as part of trehalose-6-phosphate rather than the correct  $\alpha$  anomer
- PDB chemical components [T6P](#) propagates mistake.

# 1byk BUSTER re-refinement with incorrect T6P grade dictionary



ED and mogul show problem

# 1byk BUSTER re-refinement with correct-chemistry T6P grade dictionary



Now working with Kay to deposit correction to PDB  
step (k)

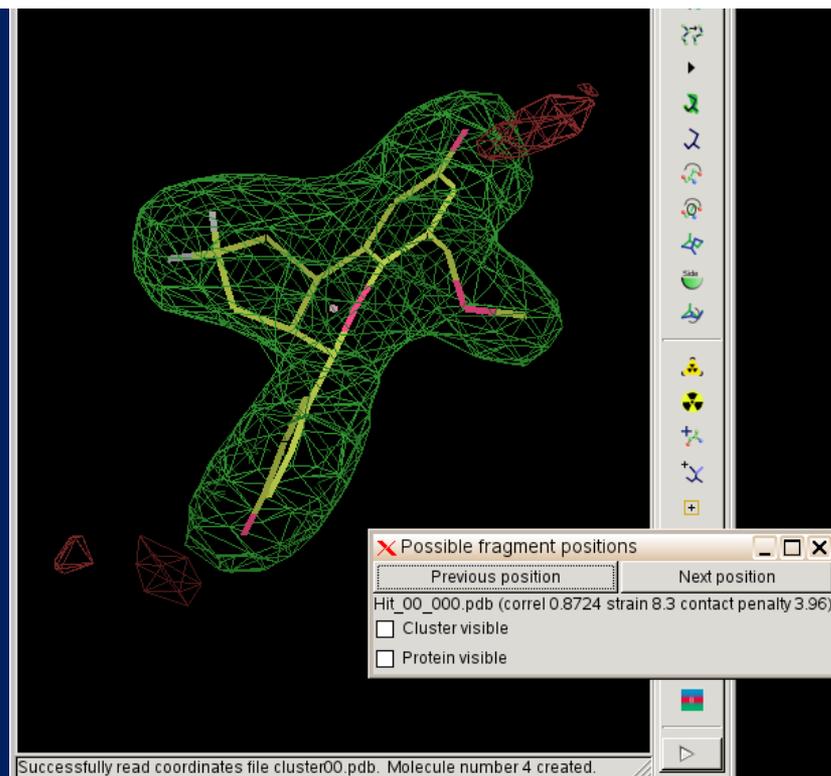
# Ligand fitting by hand

- Erice Workshop
- Will use as examples:
  - 2h7p (Bob Stroud)
  - 1pmq (Giovanna Scapin)

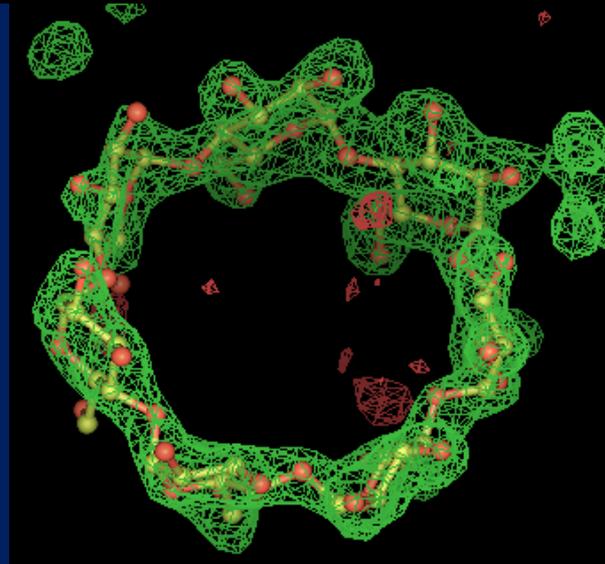
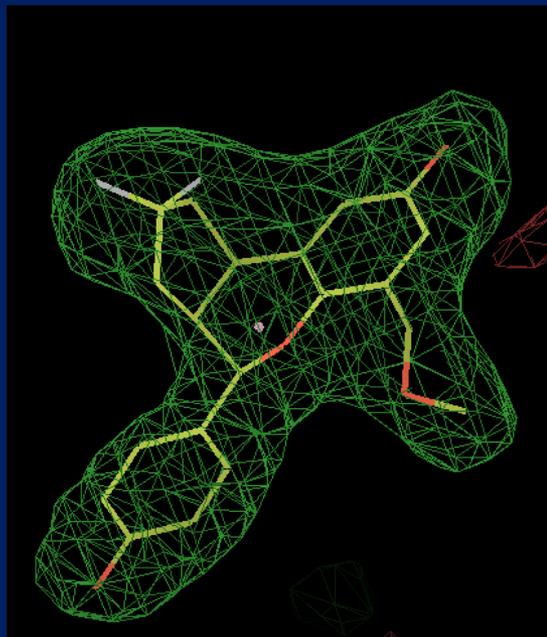
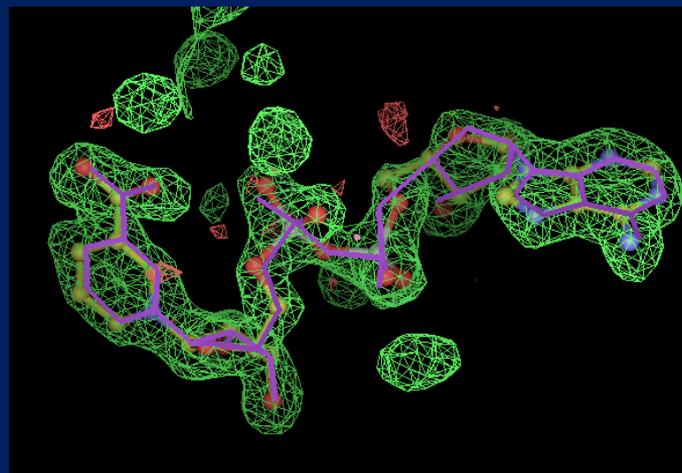
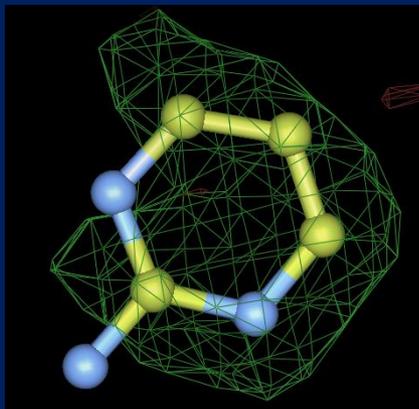
# Rhofit ligand fitter

- GΦL Rhofit is our flexible ligand Fo-Fc fitter
- Uses cif restraint dictionary to describe ligand conformational properties.
- Works well with macrocycles
- Can automatically sample chiralities if these are unknown.

File	Chain	rhofit total score	Correl coeff	ligand strain score	LigProt contact score	Poorly fitting atoms	Chirals (s:swap) (k:keep)
Hit_00_000.pdb	A	-1679.2	0.8724	8.3	4.0	0/26	sks
Hit_01_000.pdb	A	-1645.1	0.8477	11.7	4.9	0/26	skk
Hit_02_000.pdb	A	-1610.8	0.8227	20.2	8.9	0/26	ssk
Hit_02_001.pdb	A	-1608.9	0.8148	13.3	10.4	0/26	ssk
Hit_03_000.pdb	A	-1602.8	0.8336	20.8	13.8	0/26	sss
Hit_03_001.pdb	A	-1600.2	0.8427	40.5	11.3	0/26	sss



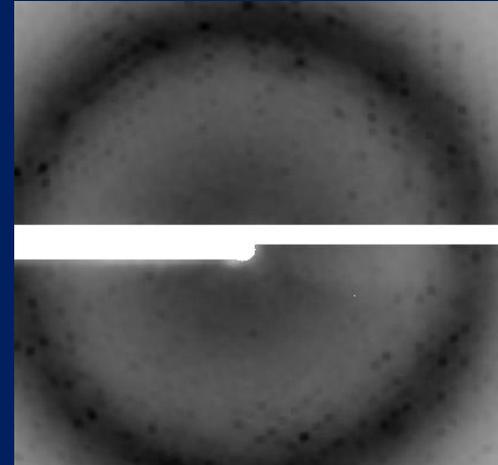
# Rhofit can handle a wide diversity of ligands



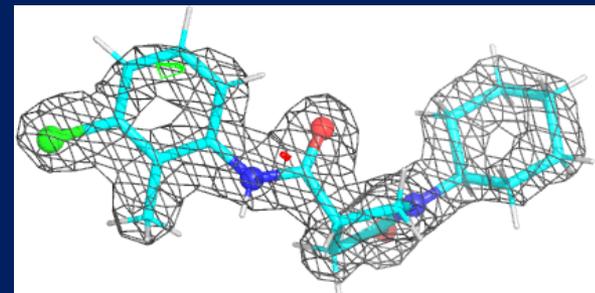
steps (b), (c), (d), (e), (g), (h) and (i)

# Pipedream: an intelligent ligand screening pipeline

- Industrial crystallographers 50+ structures per month.
- GΦL [Pipedream](#): automated data processing, structure refinement, ligand fitting and post-refinement report
- Seamlessly links autoPROC, BUSTER, rhotit & buster-report
- Grade not included as it should be run before you get images
- Automation is useful but human checking and intervention are essential!



Cc1c(C1)cccc1NC(=O)[C@H]1CN(C2CCCC2)C(=O)C1



# buster-report

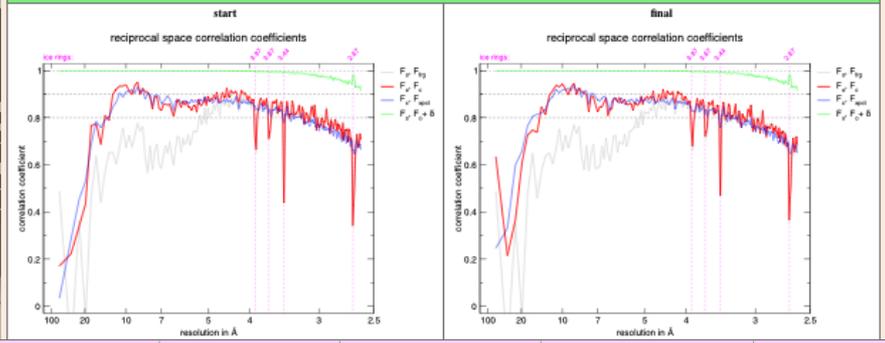
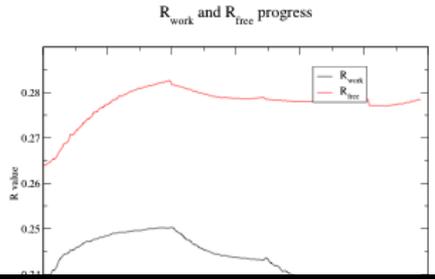
- Produce rich but concise report of a BUSTER run
- html, pdf, xml output
- Including useful analysis like

## MolProbity and Mogul

MolProbity is a great tool for assessing protein geometry issues.

run summary		
	start	final
N <sub>cycles</sub> big	0	5
N <sub>cycles</sub> small	0	294
X-ray weight	4.00	7.12
R <sub>work</sub>	0.2356	0.2341
R <sub>free</sub>	0.2639	0.2784
100*(R <sub>free</sub> -R <sub>work</sub> )	2.88	4.28
LLG <sub>work</sub> (cumulative Log-Likelihood Gain, working set)	0	-5.0036
LLG <sub>free</sub> (cumulative Log-Likelihood Gain, free set)	0	-0.0761
rms bond in Å	0.0102	0.0080
rms angle in degrees	1.15	1.
high resolution in Å	2.60	2.7
low resolution in Å	43.56	43
help? see BUSTER wiki <a href="#">run summary info</a>		

pdb entry 3qgd has bad ice ring data contamination



<a href="#">Main</a>	<a href="#">X-ray</a>	<a href="#">Ligand analysis</a>	<a href="#">Geometry details</a>	<a href="#">Refinement pro</a>
----------------------	-----------------------	---------------------------------	----------------------------------	--------------------------------

### MolProbity analysis

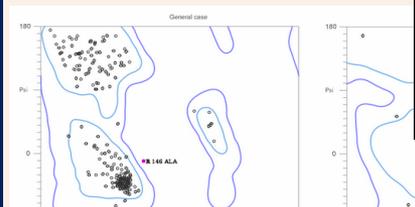
#### Summary statistics

All-Atom Contacts	Clashscore, all atoms:	3.63	99 <sup>th</sup> percentile* (N=
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	28	6.64% Goal: <1%
	Ramachandran outliers	4	0.79% Goal: <0.05%
	Ramachandran favored	480	94.86% Goal: >98%
	MolProbity score <sup>†</sup>	2.13	92 <sup>nd</sup> percentile* (N=
	Cβ deviations >0.25Å	0	0.00% Goal: 0
	Bad backbone bonds:	2 / 2038	0.10% Goal: 0%
Bad backbone angles:	8 / 2544	0.31% Goal: <0.1%	

† In the two column results, the left column gives the raw count, right column gives the percentage.  
\* 100<sup>th</sup> percentile is the best among structures of comparable resolution; 0<sup>th</sup> percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score the comparative set was selected in 2007.  
<sup>†</sup> MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

For full MolProbity multichart with per-residue violation information

#### Ramachandran plot



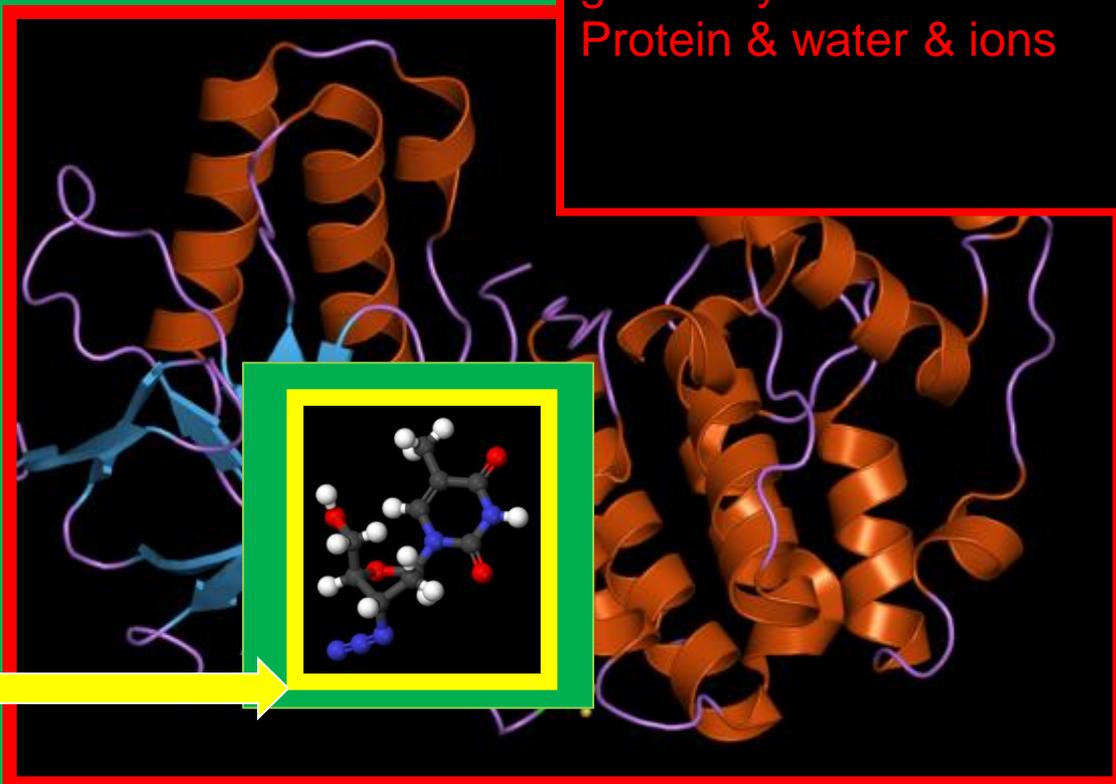
Current work: use QM or force field in BUSTER refinement to assess ligand strain energy

- Use a weighted QM or FF energy as part of the geometry function

X-ray BUSTER ML for everything!

conventional TNT+ geometry function:  
Protein & water & ions

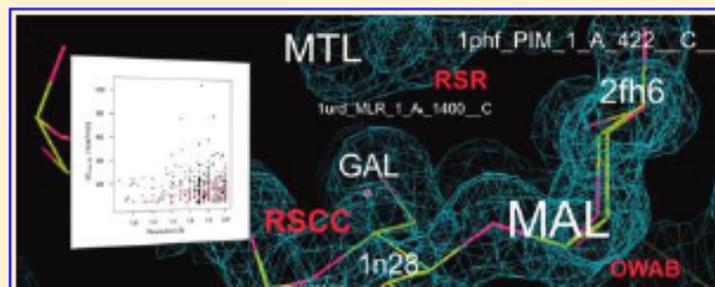
Weighted QM or force field energy for ligand



# PDB Ligand Conformational Energies Calculated Quantum-Mechanically

Markus Sitzmann,<sup>†,#</sup> Iwona E. Weidlich,<sup>†,#,∇</sup> Igor V. Filippov,<sup>‡</sup> Chenzhong Liao,<sup>†,○</sup> Megan L. Peach,<sup>‡</sup> Wolf-Dietrich Ihlenfeldt,<sup>§</sup> Rajeshri G. Karki,<sup>†,◆</sup> Yulia V. Borodina,<sup>||,¶</sup> Raul E. Cachau,<sup>⊥</sup> and Marc C. Nicklaus<sup>\*,†</sup>

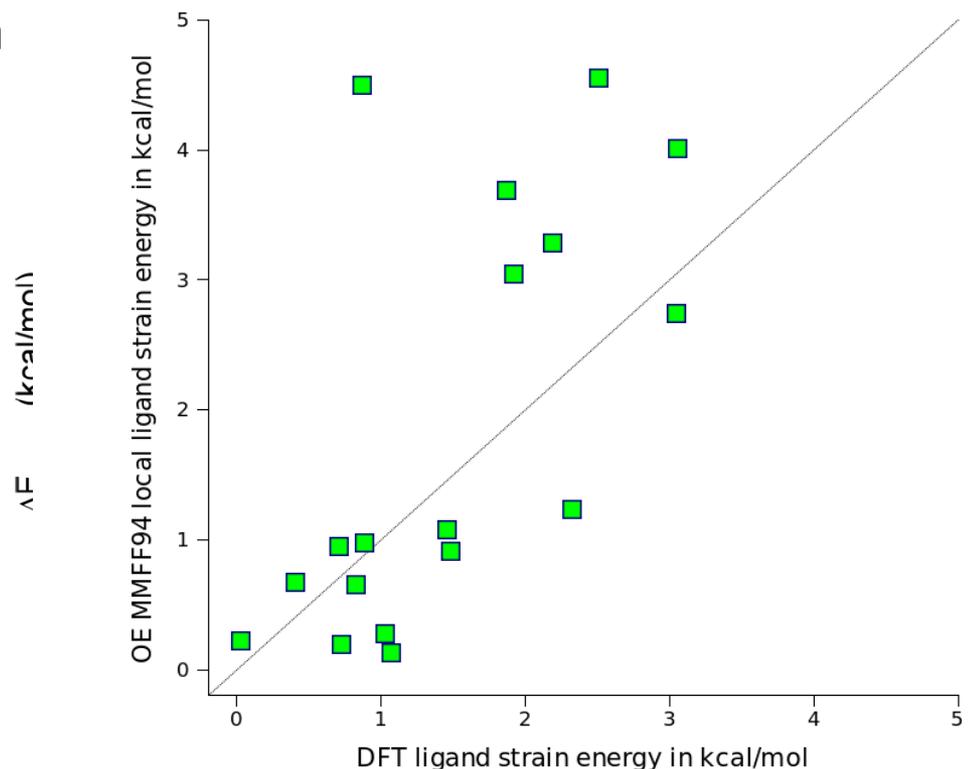
**ABSTRACT:** We present here a greatly updated version of an earlier study on the conformational energies of protein–ligand complexes in the Protein Data Bank (PDB) [Nicklaus et al. *Bioorg. Med. Chem.* 1995, 3, 411–428], with the goal of improving on all possible aspects such as number and selection of ligand instances, energy calculations performed, and additional analyses conducted. Starting from about 357,000 ligand instances deposited in the 2008 version of the Ligand Expo database of the experimental 3D coordinates of all small-molecule instances in the PDB, we created a “high-quality” subset of ligand instances by various filtering steps including application of crystallographic quality criteria and structural unambiguousness. Submission of 640 Gaussian 03 jobs yielded a set of about 415 successfully concluded runs. We used a stepwise optimization of internal degrees of freedom at the DFT level of theory with the B3LYP/6-31G(d) basis set and a single-point energy calculation at B3LYP/6-311++G(3df,2p) after each round of (partial) optimization to separate energy changes due to bond length stretches vs bond angle changes vs torsion changes. Even for the most “conservative” choice of all the possible conformational energies—the energy difference between the conformation in which all internal degrees of freedom except torsions have been optimized and the fully optimized conformer—significant energy values were found. The range of 0 to ~25 kcal/mol was populated quite evenly and independently of the crystallographic resolution. A smaller number of “outliers” of yet higher energies were seen only at resolutions above 1.3 Å. The energies showed some correlation with molecular size and flexibility but not with crystallographic



J. Chem. Inf. Model. 2012, 52, 739–756

# BUSTER Refinement with QM or Force Field for ligands

- PDB ligand structures reported by Sitzmann *et al.* to have high local strain energy using DFT
- Select 19 examples
  - Refine in BUSTER with DFT or MMFF94
  - Determine local strain energy
- BUSTER + DFT results
  - Low strain energy (<5 kcal/mol)
- BUSTER + MMFF94 results
  - strain compares to DFT
  - CPU requirements: DFT days, MMFF94 negligible



GΦL

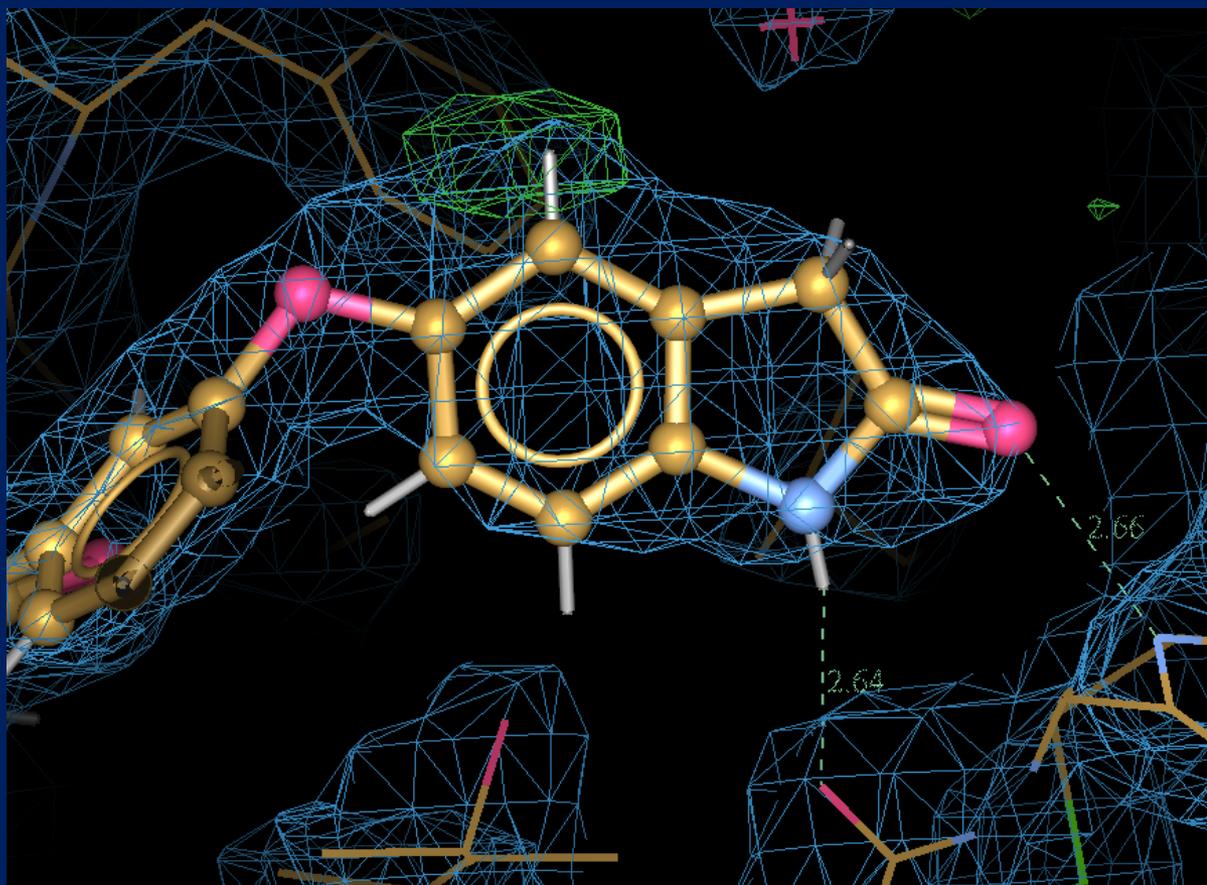
Global Phasing Limited

M. Sitzmann *et al.*, *JCIM*, **52**, 739 (2012).

## Use QM or force field in BUSTER refinement to assess ligand strain energy

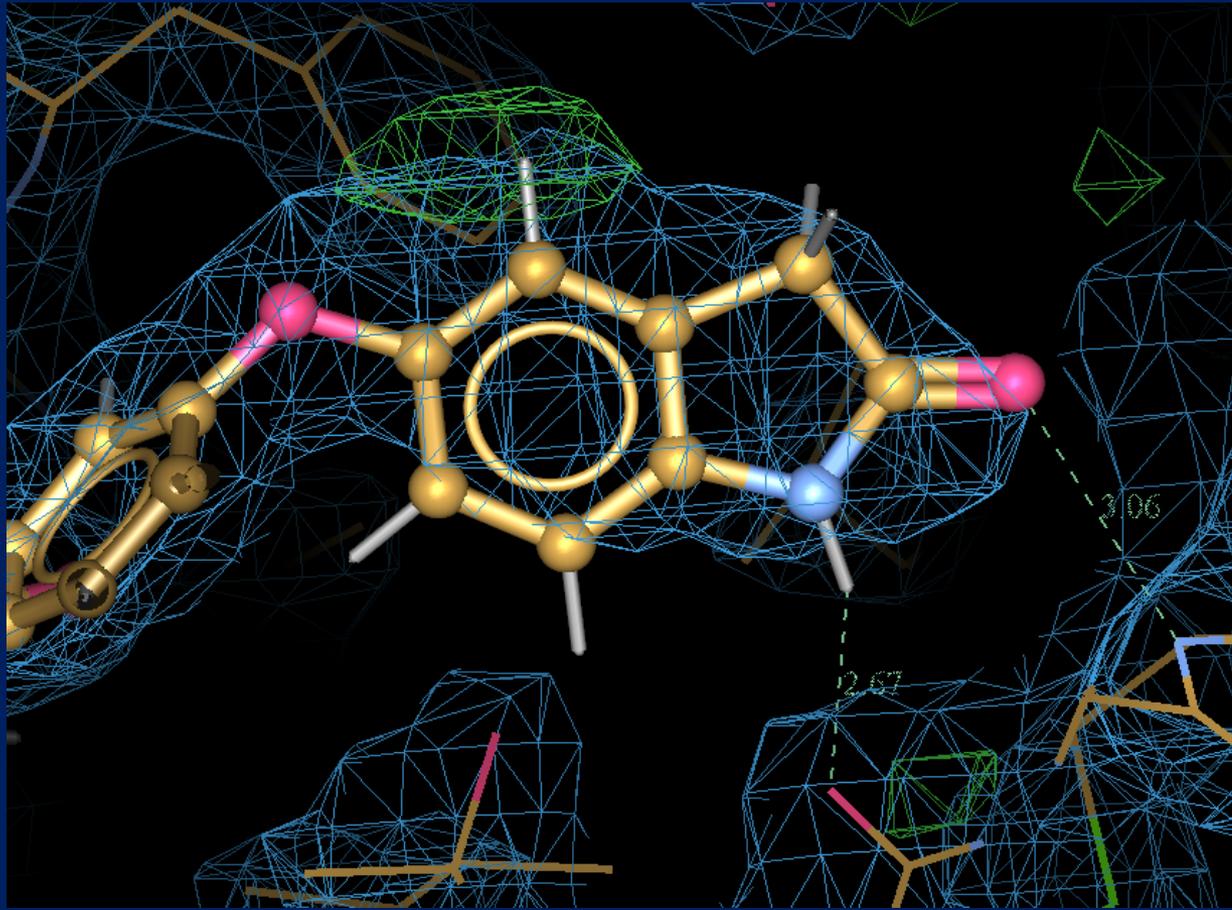
- Conformational strain energy provides an additional method of ligand validation
- Well fitted/refined structures have strain energies < 5kcal/mol
- FF suitable for general use.
- QM suitable for difficult cases.
- Links refinement to computational modelling
- Indications of strain agree with Mogul
- Joint work with John Liebeschuetz (CCDC) and Greg Warren OpenEye

# Identification of mis-fitted ligand using ligand strain & difference density



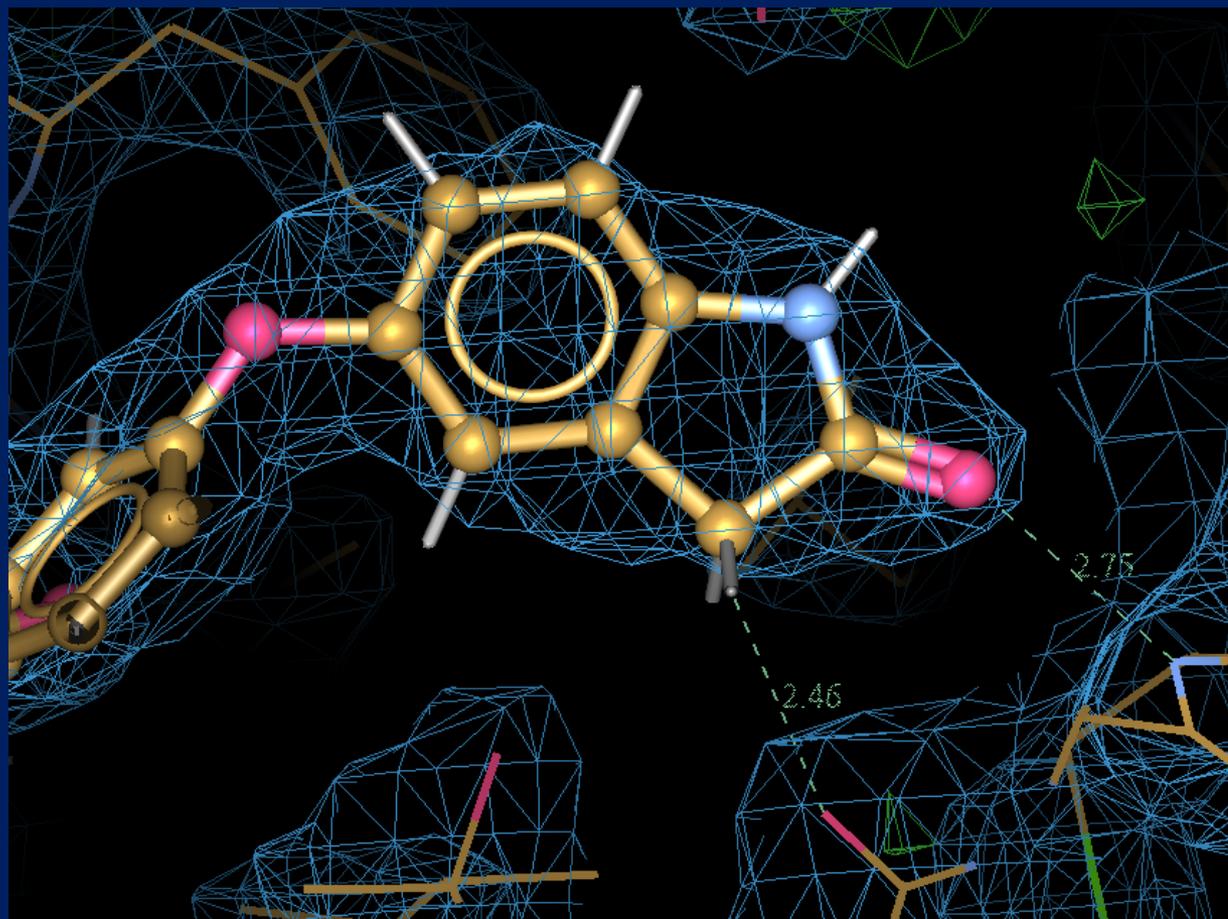
4bki DDR1 Receptor Tyrosine Kinase Inhibitor PDB entry fitted  
as per design, but the ligand geometry is strained

# 4bki re-refinement with BUSTER reduces strain and reveals mis-fit



indolin-2-one ring flip required

# BUSTER re-refinement after ring flip



indolin-2-one ring happy improve C=O H-bond

# Corrected PDB entry 4CKR

ACS Chem Biol. Mar 21, 2014; 9(3): 840.

PMCID: PMC3971954

Published online Feb 14, 2014. doi: [10.1021/cb5000949](https://doi.org/10.1021/cb5000949)

## Correction to Discovery of a Potent and Selective DDR1 Receptor Tyrosine Kinase Inhibitor

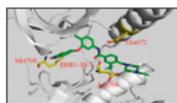
[Hyung-Gu Kim](#), [Li Tan](#), [Ellen L. Weisberg](#), [Feiyang Liu](#), [Peter Canning](#), [Hwan-Geun Choi](#), [Scott Ezell](#), [Zheng Zhao](#), [Hong Wu](#), [Jinhua Wang](#), [Anna Mandinova](#), [Alex N. Bullock,\\*](#) [Qingsong Liu,\\*](#) [Sam W. Lee,\\*](#) and [Nathanael S. Gray\\*](#)

[Copyright and License information](#) ►

This corrects the article "[Discovery of a Potent and Selective DDR1 Receptor Tyrosine Kinase Inhibitor](#)" in *ACS Chem Biol*, volume 8 on page 2145.

In the reported X-ray cocrystal structure of the DDR1 kinase domain in complex with the inhibitor DDR1-IN-1 (PDB code 4BKI), the indolin-2-one moiety was modeled with two hydrogen bonds to the kinase hinge residues Met704 and Asp702.

Subsequent analysis of the electron density has revealed that the indolin-2-one group is flipped allowing only a single hydrogen bond to Met704. The amended coordinates have been released with the new PDB code 4CKR. The corrected Figure 2A is shown, together with the updated refinement statistics (Table 5) reported in Supporting Information Table 5 of the [original paper](#). This change does not otherwise affect the scientific integrity of the article. We thank Oliver Smart, Global Phasing Ltd., for drawing our attention to the error.



**Figure 2**

Binding information of DDR1-IN-1/2 against DDR1.

(A) X-ray cocrystal structure of DDR1-IN-1 with DDR1 kinase.

# Deposition to in-house database or PDB

- Deposition minimum:
  - Model (~~PDB~~, mmCIF)
  - Structure factors (mtz)
- Ideally should include:
  - Diffraction images! For 1byk Kay Diederichs kept these. Completeness 1998: 67.5% → 2014: 99.4%
  - Ligand chemistry and restraint dictionaries
  - Final map that the crystallographer saw
  - Validation information (buster-report pdf output)

# Acknowledgments

## Ideas, Feedback,... 🤖

### Global Phasing Colleagues

- Gérard Bricogne
- Claus Flensburg
- Andrew Sharff
- Tom Womack
- Clemens Vornrhein

### Global Phasing Consortium Users

#### BUSTER users

#### CCDC

- John Liebeschuetz
- Colin Groom

#### OpenEye

- Greg Warren
- Brian Kelley

## Support £\$€

### Global Phasing Consortium

#### EU project:

SILVER FP7-HEALTH-F3-2010-260644

## PDB & depositors

Thanks to all depositors of structures used.

*“We do not learn from experience... we learn from reflecting on experience.” John Dewey*