Full wwPDB X-ray Structure Validation Report (i)

May 24, 2020 – 10:38 pm BST

PDB ID	:	5XGH
Title	:	Crystal structure of PI3K complex with an inhibitor
Authors	:	Song, K.; Yang, X.; Zhao, Y.; Jian, Z.
Deposited on	:	2017-04-13
Resolution	:	2.97 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

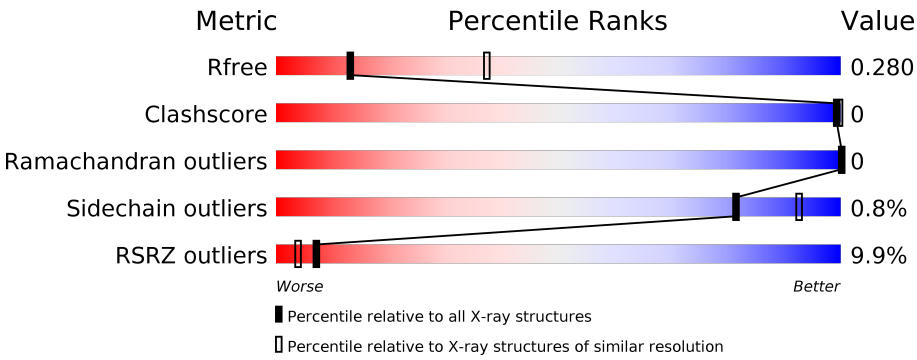
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.97 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2754 (3.00-2.96)
Clashscore	141614	3103 (3.00-2.96)
Ramachandran outliers	138981	2993 (3.00-2.96)
Sidechain outliers	138945	2996 (3.00-2.96)
RSRZ outliers	127900	2644 (3.00-2.96)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1048	10% 94% • 5%
2	B	277	6% 99% •

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 10540 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

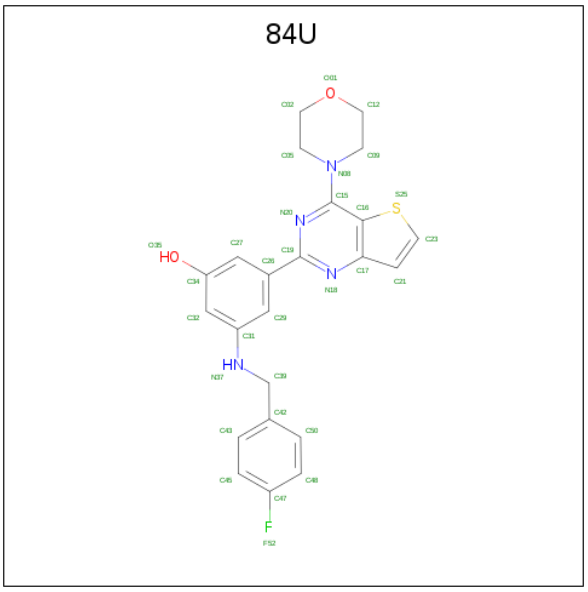
- Molecule 1 is a protein called Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	993	8126	5196	1390	1473	67	0	0	0

- Molecule 2 is a protein called Phosphatidylinositol 3-kinase regulatory subunit alpha.

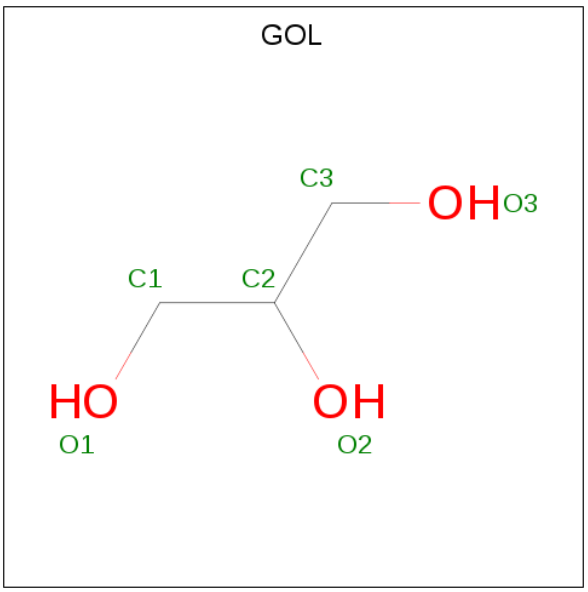
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	277	2354	1474	420	452	8	0	0	0

- Molecule 3 is 3-[(4-fluorophenyl)methylamino]-5-(4-morpholin-4-ylthieno[3,2-d]pyrimidin-2-yl)phenol (three-letter code: 84U) (formula: $C_{23}H_{21}FN_4O_2S$).



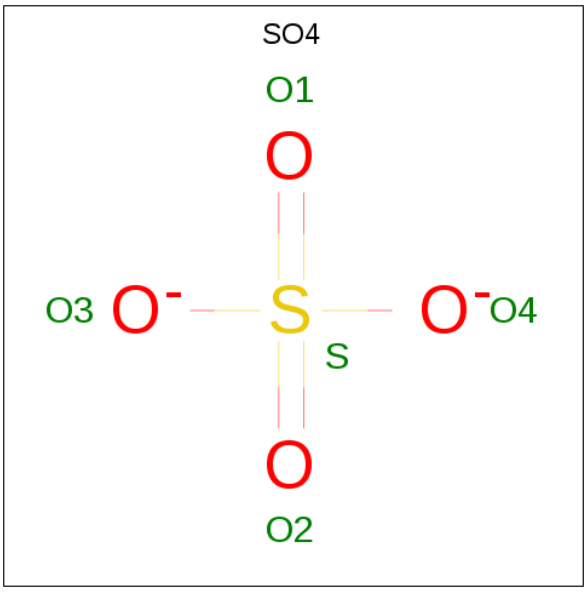
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	F	N	O	S		
3	A	1	31	23	1	4	2	1	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	B	1	6	3	3	0	0

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
5	B	1	5	4	1	0	0

- Molecule 6 is water.

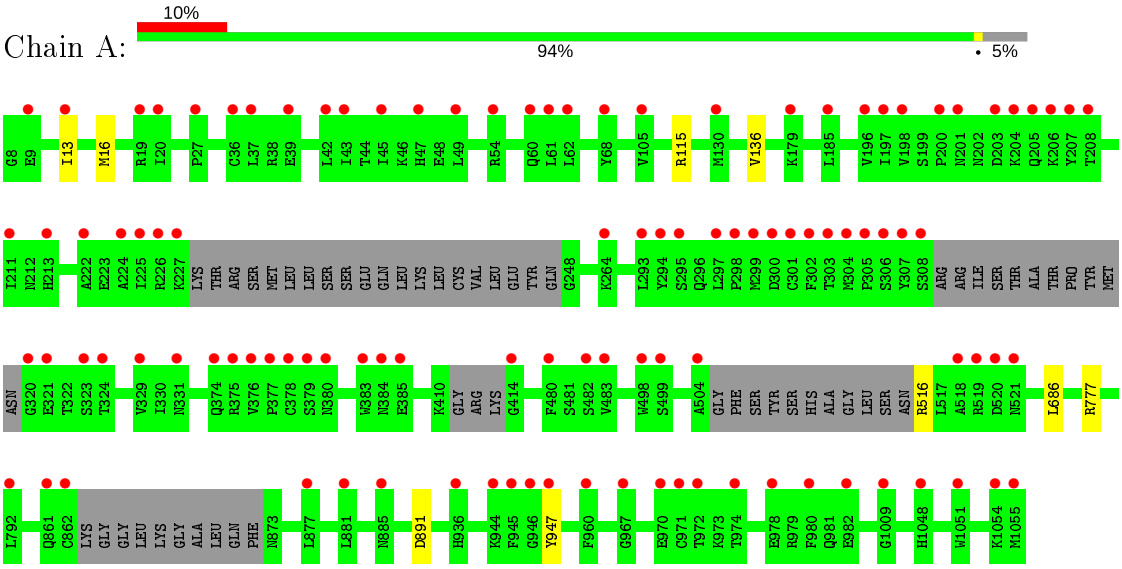
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	14	14	14	0	0
			Total	O		
6	B	4	4	4	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

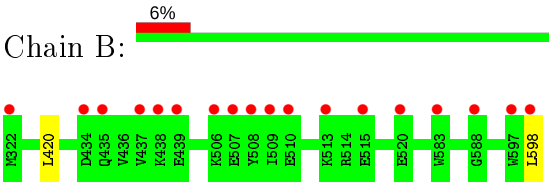
- Molecule 1: Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform

Chain A:



- Molecule 2: Phosphatidylinositol 3-kinase regulatory subunit alpha

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	70.36Å 136.31Å 149.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.97 48.95 – 2.97	Depositor EDS
% Data completeness (in resolution range)	94.2 (50.00-2.97) 94.3 (48.95-2.97)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.62 (at 2.96Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.231 , 0.280 0.231 , 0.280	Depositor DCC
R_{free} test set	1362 reflections (4.75%)	wwPDB-VP
Wilson B-factor (Å^2)	77.3	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 43.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	10540	wwPDB-VP
Average B, all atoms (Å^2)	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.48% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, 84U, SO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| >5$	RMSZ	$\#\|Z\| >5$
1	A	0.36	0/8307	0.52	0/11224
2	B	0.37	0/2394	0.53	0/3207
All	All	0.36	0/10701	0.53	0/14431

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	8126	0	8127	1	0
2	B	2354	0	2330	0	0
3	A	31	0	0	0	0
4	B	6	0	8	0	0
5	B	5	0	0	0	0
6	A	14	0	0	0	0
6	B	4	0	0	0	0
All	All	10540	0	10465	1	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 0.

All (1) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:136:VAL:HG13	1:A:686:LEU:HD11	2.04	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	981/1048 (94%)	944 (96%)	37 (4%)	0	100	100
2	B	275/277 (99%)	267 (97%)	8 (3%)	0	100	100
All	All	1256/1325 (95%)	1211 (96%)	45 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	909/955 (95%)	902 (99%)	7 (1%)	81	92
2	B	258/258 (100%)	256 (99%)	2 (1%)	81	92
All	All	1167/1213 (96%)	1158 (99%)	9 (1%)	81	92

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	13	ILE
1	A	16	MET
1	A	115	ARG
1	A	516	ARG
1	A	777	ARG
1	A	891	ASP
1	A	947	TYR
2	B	420	LEU
2	B	598	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	384	ASN
1	A	958	GLN
2	B	453	ASN
2	B	457	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the

expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	GOL	B	601	-	5,5,5	0.34	0	5,5,5	0.22	0
5	SO4	B	602	-	4,4,4	0.32	0	6,6,6	0.09	0
3	84U	A	1101	-	33,35,35	1.17	1 (3%)	40,49,49	1.78	8 (20%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	GOL	B	601	-	-	4/4/4/4	-
3	84U	A	1101	-	-	4/13/21/21	0/5/5/5

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1101	84U	C15-N20	4.37	1.38	1.32

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	1101	84U	C21-C17-C16	4.98	114.25	110.62
3	A	1101	84U	C16-C15-N08	3.85	129.24	121.57
3	A	1101	84U	N18-C19-N20	-3.66	123.24	126.11
3	A	1101	84U	C16-C15-N20	-3.46	113.90	122.60
3	A	1101	84U	C15-N20-C19	3.32	123.96	116.17
3	A	1101	84U	C27-C26-C19	3.06	124.55	120.05
3	A	1101	84U	C26-C29-C31	2.62	124.97	121.07
3	A	1101	84U	C05-N08-C15	2.27	125.20	118.73

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1101	84U	C16-C15-N08-C05
3	A	1101	84U	N20-C15-N08-C05
3	A	1101	84U	C29-C31-N37-C39
3	A	1101	84U	C32-C31-N37-C39
4	B	601	GOL	O2-C2-C3-O3
4	B	601	GOL	O1-C1-C2-C3

Continued on next page...

Continued from previous page...

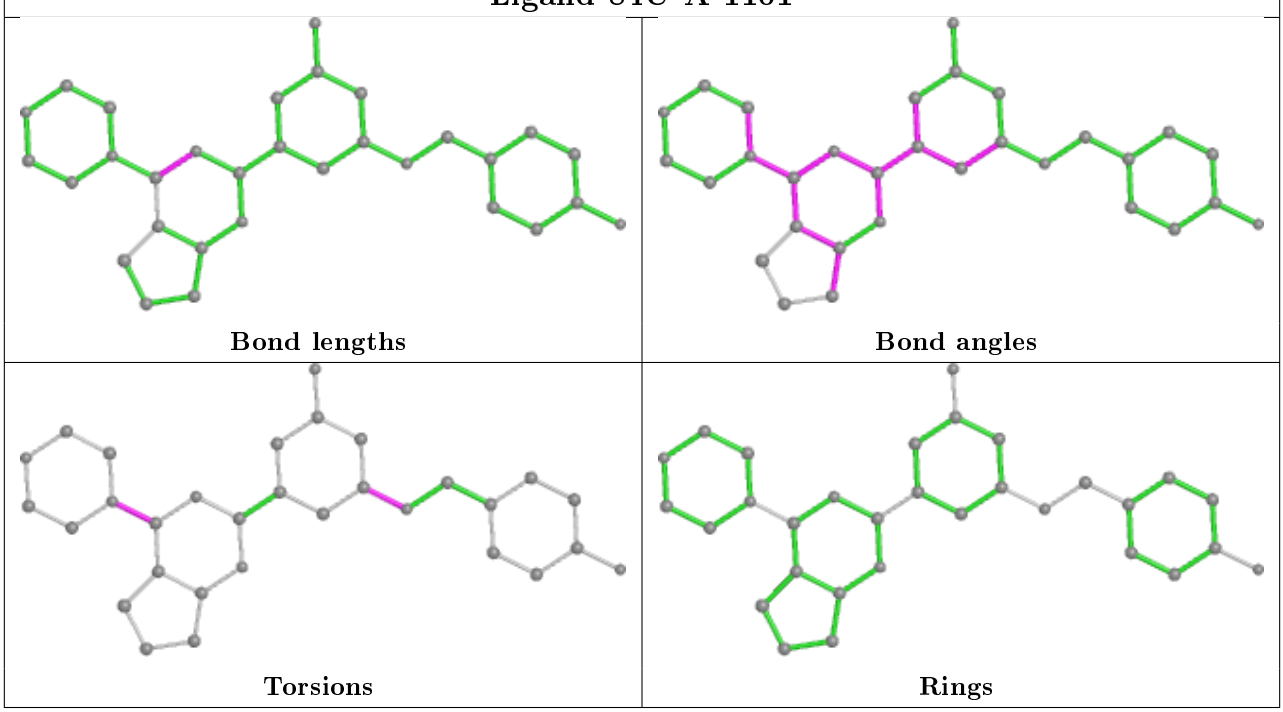
Mol	Chain	Res	Type	Atoms
4	B	601	GOL	C1-C2-C3-O3
4	B	601	GOL	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand 84U A 1101



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	993/1048 (94%)	0.71	108 (10%) 5 3	49, 81, 135, 163	0
2	B	277/277 (100%)	0.54	18 (6%) 18 10	60, 86, 145, 156	0
All	All	1270/1325 (95%)	0.67	126 (9%) 7 4	49, 83, 138, 163	0

All (126) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	298	PRO	8.2
1	A	205	GLN	7.7
1	A	482	SER	7.2
1	A	308	SER	7.0
1	A	946	GLY	6.9
1	A	383	TRP	5.9
1	A	320	GLY	5.8
1	A	301	CYS	5.6
2	B	438	LYS	5.6
1	A	227	LYS	5.4
1	A	519	ARG	4.6
2	B	322	MET	4.6
1	A	1055	MET	4.6
1	A	518	ALA	4.5
1	A	297	LEU	4.5
1	A	201	ASN	4.5
1	A	61	LEU	4.5
1	A	302	PHE	4.2
1	A	378	CYS	4.1
2	B	439	GLU	4.1
1	A	37	LEU	4.0
1	A	60	GLN	4.0
1	A	862	CYS	3.9
1	A	375	ARG	3.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	435	GLN	3.8
1	A	307	TYR	3.8
1	A	944	LYS	3.8
1	A	380	ASN	3.7
2	B	583	TRP	3.7
1	A	945	PHE	3.7
1	A	200	PRO	3.6
1	A	504	ALA	3.6
2	B	506	LYS	3.6
1	A	385	GLU	3.6
1	A	304	MET	3.6
2	B	437	VAL	3.5
1	A	377	PRO	3.5
1	A	323	SER	3.4
1	A	207	TYR	3.4
1	A	972	THR	3.4
1	A	303	THR	3.3
1	A	43	ILE	3.3
1	A	376	VAL	3.3
1	A	295	SER	3.3
1	A	224	ALA	3.3
1	A	42	LEU	3.3
1	A	1048	HIS	3.2
1	A	19	ARG	3.2
2	B	508	TYR	3.2
2	B	513	LYS	3.2
1	A	305	PRO	3.2
1	A	414	GLY	3.1
1	A	9	GLU	3.1
1	A	198	VAL	3.1
1	A	204	LYS	3.1
2	B	509	ILE	3.0
1	A	222	ALA	3.0
1	A	971	CYS	2.9
1	A	300	ASP	2.9
1	A	179	LYS	2.9
1	A	384	ASN	2.9
1	A	20	ILE	2.9
1	A	947	TYR	2.9
1	A	62	LEU	2.9
1	A	208	THR	2.9
1	A	1051	TRP	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	374	GLN	2.8
1	A	861	GLN	2.8
1	A	47	HIS	2.8
1	A	54	ARG	2.8
2	B	598	LEU	2.8
1	A	498	TRP	2.8
1	A	982	GLU	2.8
1	A	520	ASP	2.8
2	B	434	ASP	2.7
1	A	13	ILE	2.7
1	A	264	LYS	2.7
1	A	196	VAL	2.7
1	A	321	GLU	2.6
1	A	521	ASN	2.6
1	A	203	ASP	2.6
1	A	331	ASN	2.6
1	A	974	THR	2.6
2	B	507	GLU	2.5
1	A	294	TYR	2.5
1	A	68	TYR	2.5
1	A	967	GLY	2.4
1	A	306	SER	2.4
1	A	39	GLU	2.4
2	B	597	TRP	2.4
1	A	885	ASN	2.4
1	A	211	ILE	2.4
1	A	185	LEU	2.4
2	B	510	GLU	2.4
1	A	299	MET	2.3
1	A	226	ARG	2.3
1	A	225	ILE	2.3
2	B	588	GLY	2.3
1	A	480	PHE	2.3
1	A	293	LEU	2.3
1	A	877	LEU	2.3
1	A	36	CYS	2.3
1	A	970	GLU	2.3
1	A	45	ILE	2.2
1	A	197	ILE	2.2
1	A	1054	LYS	2.2
1	A	960	PHE	2.2
2	B	520	GLU	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	379	SER	2.2
1	A	499	SER	2.2
1	A	329	VAL	2.2
1	A	881	LEU	2.1
1	A	936	HIS	2.1
1	A	324	THR	2.1
1	A	206	LYS	2.1
1	A	130	MET	2.1
1	A	483	VAL	2.1
1	A	980	PHE	2.1
1	A	978	GLU	2.0
1	A	27	PRO	2.0
1	A	213	HIS	2.0
1	A	792	LEU	2.0
1	A	1009	GLY	2.0
1	A	105	VAL	2.0
1	A	49	LEU	2.0
2	B	515	GLU	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

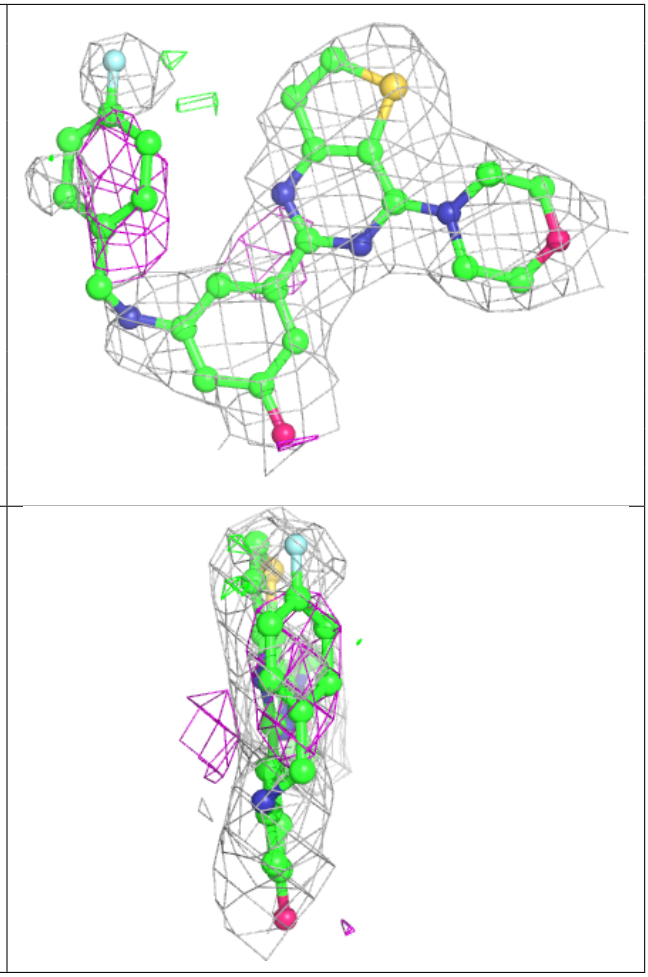
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	GOL	B	601	6/6	0.84	0.16	91,92,92,93	0
3	84U	A	1101	31/31	0.89	0.26	56,61,84,85	0
5	SO4	B	602	5/5	0.97	0.13	79,80,81,81	0

The following is a graphical depiction of the model fit to experimental electron density of all

instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around 84U A 1101:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.