Full wwPDB X-ray Structure Validation Report (i)

May 25, 2020 – 06:33 pm BST

PDB ID	:	6CBY
Title	:	Crystal structure of human SET and MYND Domain Containing protein 2 with MTF9975
Authors	:	ZENG, H.; DONG, A.; Hutchinson, A.; Seitova, A.; TATLOCK, J.; KUMPF, R.; OWEN, A.; TAYLOR, A.; Casimiro-Garcia, A.; Bountra, C.; Arrowsmith, C.H.; Edwards, A.M.; BROWN, P.J.; WU, H.; Structural Genomics Consortium (SGC)
Deposited on	:	2018-02-05
Resolution	:	2.55 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:

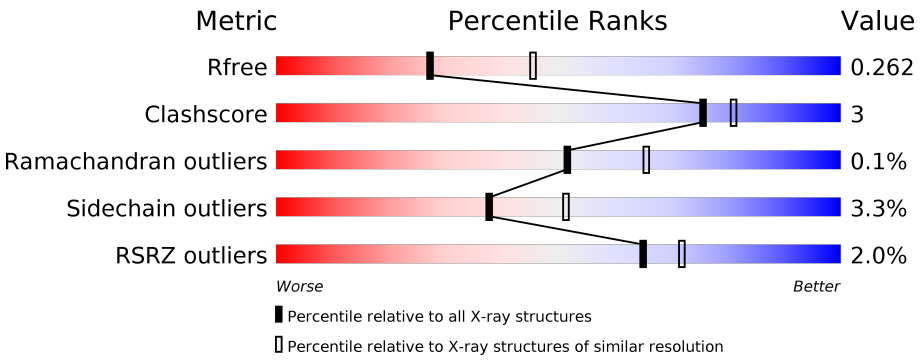
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (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 2.55 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	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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	434	88% 10% •
1	B	434	4% 88% 10% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6540 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 N-lysine methyltransferase SMYD2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	428	3266	2088	543	605	30	0	0	0
1	B	422	3142	2000	527	585	30	0	0	0

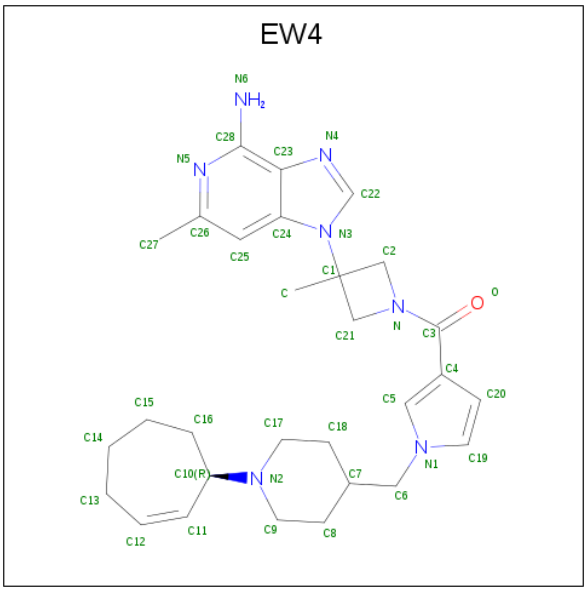
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q9NRG4
A	165	GLU	GLY	conflict	UNP Q9NRG4
B	0	GLY	-	expression tag	UNP Q9NRG4
B	165	GLU	GLY	conflict	UNP Q9NRG4

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
2	B	3	3	3	0	0
2	A	3	3	3	0	0

- Molecule 3 is [3-(4-amino-6-methyl-1H-imidazo[4,5-c]pyridin-1-yl)-3-methylazetidin-1-yl][1-({1-[(1R)-cyclohept-2-en-1-yl]piperidin-4-yl}methyl)-1H-pyrrol-3-yl]methanone (three-letter code: EW4) (formula: $C_{29}H_{39}N_7O$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	37	29	7	1	0	0
			Total	C	N	O		
3	B	1	37	29	7	1	0	0

- Molecule 4 is water.

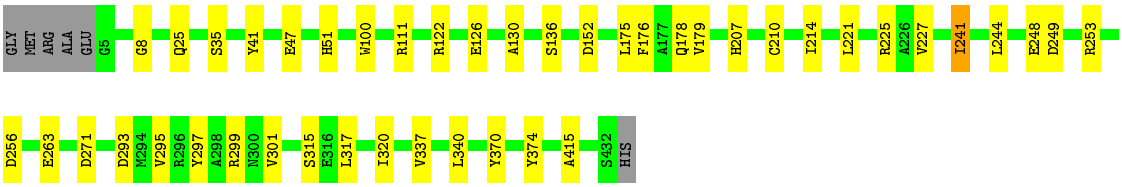
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	31	31	31	0	0
			Total	O		
4	B	21	21	21	0	0

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.

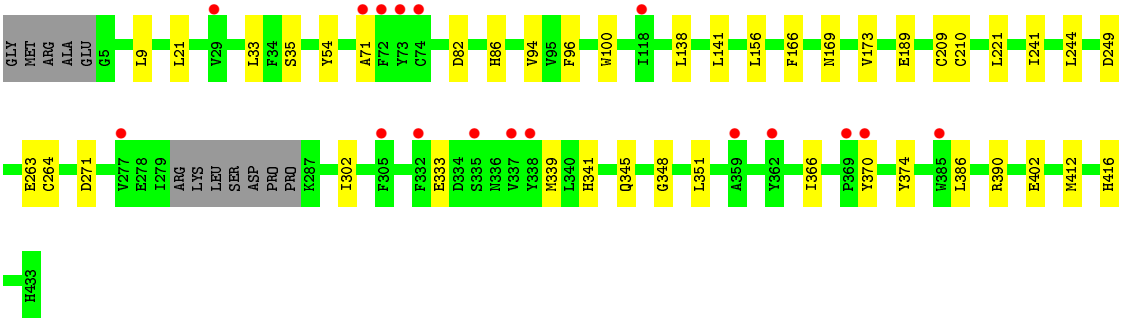
- Molecule 1: N-lysine methyltransferase SMYD2

Chain A:



- Molecule 1: N-lysine methyltransferase SMYD2

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	57.82Å 117.58Å 65.01Å 90.00° 92.42° 90.00°	Depositor
Resolution (Å)	43.59 – 2.55 43.59 – 2.52	Depositor EDS
% Data completeness (in resolution range)	98.9 (43.59-2.55) 97.3 (43.59-2.52)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.66 (at 2.51Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.199 , 0.258 0.209 , 0.262	Depositor DCC
R_{free} test set	893 reflections (3.11%)	wwPDB-VP
Wilson B-factor (Å^2)	43.7	Xtriage
Anisotropy	0.454	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 44.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.047 for h,-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6540	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.85% 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: ZN, EW4

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	0.53	0/3343	0.68	0/4541
1	B	0.51	0/3215	0.65	0/4374
All	All	0.52	0/6558	0.67	0/8915

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	3266	0	3018	19	0
1	B	3142	0	2792	19	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
3	A	37	0	0	0	0
3	B	37	0	0	0	0
4	A	31	0	0	0	0
4	B	21	0	0	0	0
All	All	6540	0	5810	38	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 3.

All (38) 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:214:ILE:CD1	1:A:214:ILE:CG1	1.77	1.55
1:B:366:ILE:O	1:B:370:TYR:HB2	2.06	0.56
1:A:152:ASP:HB3	1:A:176:PHE:CE1	2.44	0.53
1:B:244:LEU:HB2	1:B:374:TYR:CZ	2.43	0.53
1:A:244:LEU:HB2	1:A:374:TYR:CZ	2.44	0.52
1:A:317:LEU:HA	1:A:320:ILE:HD12	1.92	0.50
1:B:156:LEU:HG	1:B:166:PHE:CE1	2.49	0.47
1:A:35:SER:HB2	1:A:221:LEU:HD11	1.96	0.47
1:A:47:GLU:HB3	1:A:51:HIS:CD2	2.50	0.47
1:B:169:ASN:O	1:B:173:VAL:HG23	2.15	0.47
1:A:225:ARG:HD2	1:A:415:ALA:O	2.15	0.46
1:B:71:ALA:HB2	1:B:86:HIS:CE1	2.50	0.46
1:A:214:ILE:CD1	1:A:214:ILE:CB	2.82	0.46
1:B:9:LEU:HD21	1:B:33:LEU:HD23	1.97	0.46
1:B:9:LEU:HB3	1:B:21:LEU:HD11	1.98	0.46
1:B:209:CYS:HB2	1:B:264:CYS:SG	2.57	0.45
1:A:136:SER:HB3	1:A:178:GLN:HG3	1.98	0.45
1:B:348:GLY:HA2	1:B:351:LEU:HD12	2.00	0.43
1:A:175:LEU:O	1:A:179:VAL:HG23	2.18	0.43
1:B:96:PHE:HB2	1:B:100:TRP:HB2	2.01	0.43
1:A:41:TYR:O	1:A:111:ARG:HD3	2.18	0.43
1:B:412:MET:O	1:B:416:HIS:HB2	2.18	0.43
1:A:214:ILE:HG13	1:A:227:VAL:HG11	2.00	0.43
1:A:47:GLU:HB3	1:A:51:HIS:HD2	1.83	0.43
1:A:8:GLY:HA3	1:A:25:GLN:CD	2.38	0.43
1:A:295:VAL:O	1:A:299:ARG:HG3	2.18	0.42
1:A:126:GLU:HG2	1:A:130:ALA:HA	2.00	0.42
1:B:189:GLU:HG2	1:B:386:LEU:HD21	2.01	0.42
1:A:297:TYR:CZ	1:A:301:VAL:HG21	2.55	0.42
1:B:333:GLU:O	1:B:339:MET:HG3	2.19	0.42
1:B:138:LEU:HD12	1:B:141:LEU:HD12	2.02	0.41
1:B:54:TYR:HB2	1:B:82:ASP:HB2	2.02	0.41
1:A:337:VAL:HA	1:A:340:LEU:HD12	2.02	0.41
1:B:302:ILE:HG12	1:B:341:HIS:CD2	2.55	0.41
1:A:207:HIS:CE1	1:A:241:ILE:HD13	2.56	0.40
1:B:35:SER:HB3	1:B:221:LEU:HD11	2.03	0.40
1:B:156:LEU:HG	1:B:166:PHE:HE1	1.86	0.40
1:B:341:HIS:O	1:B:345:GLN:HG2	2.22	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	426/434 (98%)	408 (96%)	17 (4%)	1 (0%)	47	60
1	B	418/434 (96%)	408 (98%)	10 (2%)	0	100	100
All	All	844/868 (97%)	816 (97%)	27 (3%)	1 (0%)	51	65

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	100	TRP

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	322/382 (84%)	310 (96%)	12 (4%)	34	46
1	B	293/382 (77%)	285 (97%)	8 (3%)	44	59
All	All	615/764 (80%)	595 (97%)	20 (3%)	38	51

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	122	ARG
1	A	210	CYS
1	A	241	ILE
1	A	248	GLU
1	A	249	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	253	ARG
1	A	256	ASP
1	A	263	GLU
1	A	271	ASP
1	A	293	ASP
1	A	315	SER
1	A	370	TYR
1	B	94	VAL
1	B	210	CYS
1	B	241	ILE
1	B	249	ASP
1	B	263	GLU
1	B	271	ASP
1	B	390	ARG
1	B	402	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	182	ASN

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)

Of 8 ligands modelled in this entry, 6 are monoatomic - leaving 2 for Mogul analysis.

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	$\#\|Z\| > 2$
3	EW4	A	1004	-	36,42,42	0.59	0	31,62,62	1.47	2 (6%)
3	EW4	B	1004	-	36,42,42	0.54	0	31,62,62	1.34	2 (6%)

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
3	EW4	A	1004	-	-	4/14/51/51	0/6/6/6
3	EW4	B	1004	-	-	4/14/51/51	0/6/6/6

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	1004	EW4	C1-N3-C24	6.06	130.99	124.46
3	B	1004	EW4	C1-N3-C24	5.20	130.07	124.46
3	A	1004	EW4	C23-C28-N6	2.01	123.41	120.35
3	B	1004	EW4	C23-C28-N5	-2.00	119.70	121.01

There are no chirality outliers.

All (8) torsion outliers are listed below:

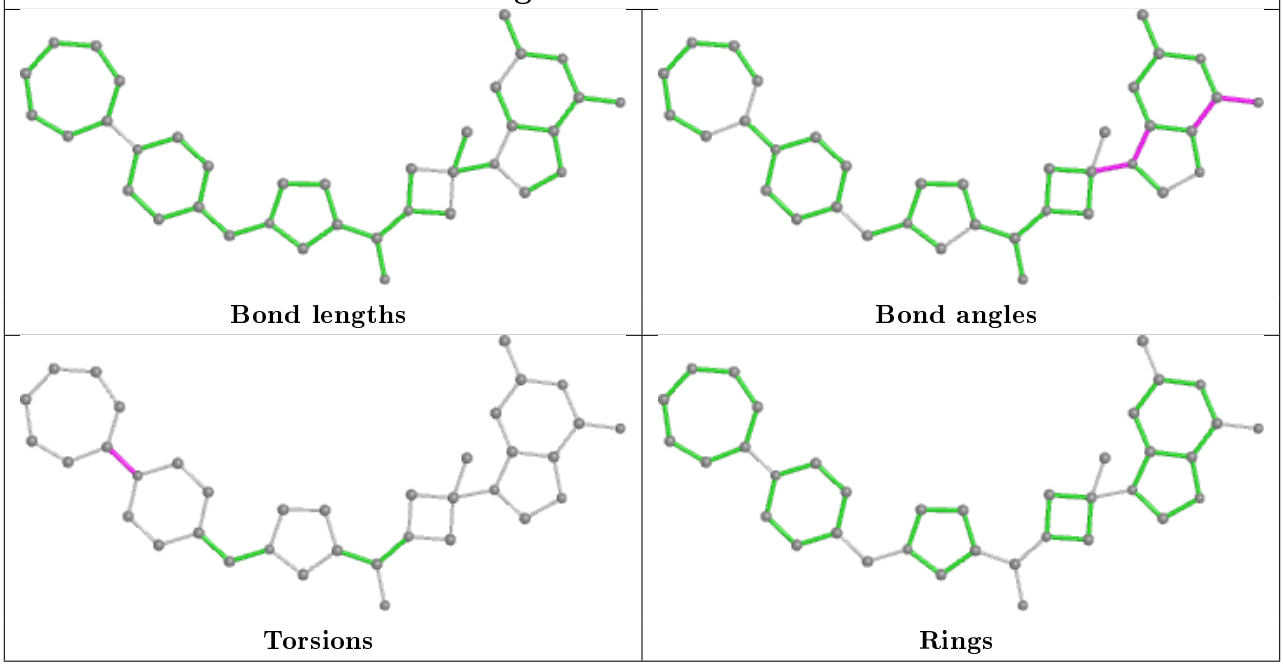
Mol	Chain	Res	Type	Atoms
3	B	1004	EW4	C11-C10-N2-C9
3	B	1004	EW4	C16-C10-N2-C9
3	B	1004	EW4	C11-C10-N2-C17
3	B	1004	EW4	C16-C10-N2-C17
3	A	1004	EW4	C11-C10-N2-C9
3	A	1004	EW4	C16-C10-N2-C9
3	A	1004	EW4	C16-C10-N2-C17
3	A	1004	EW4	C11-C10-N2-C17

There are no ring outliers.

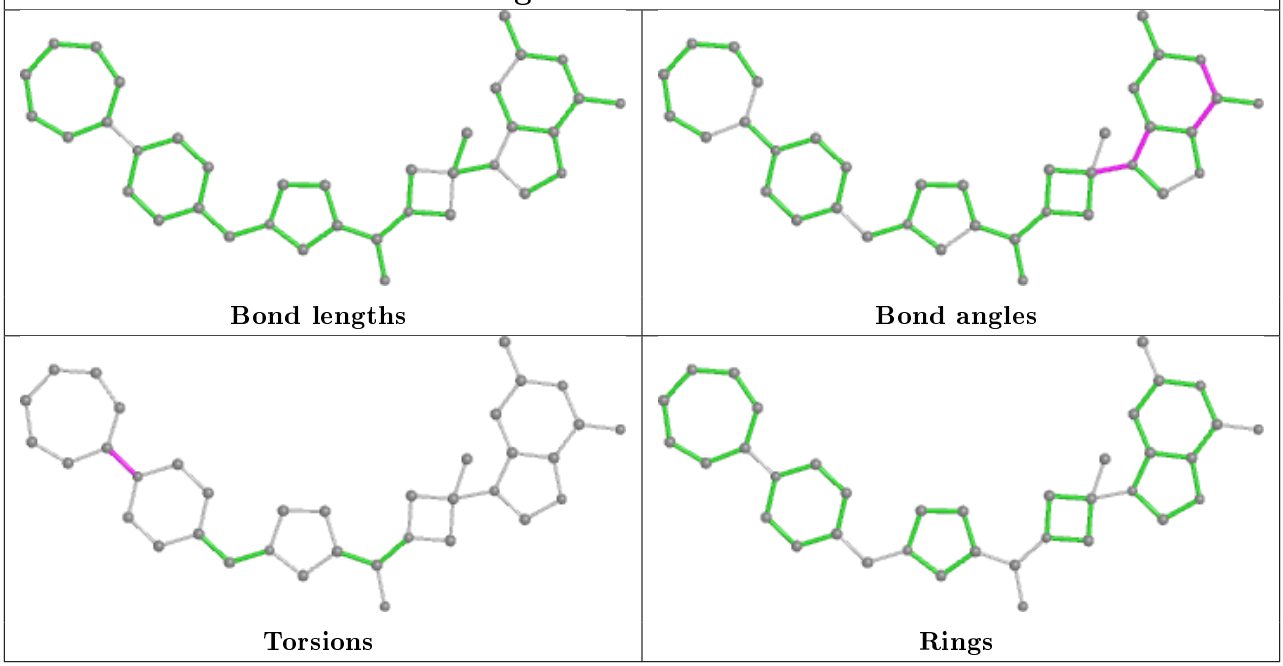
No monomer is involved in short contacts.

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 EW4 A 1004



Ligand EW4 B 1004



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	428/434 (98%)	-0.28	0 100 100	28, 46, 68, 106	0
1	B	422/434 (97%)	-0.00	17 (4%) 38 45	30, 57, 89, 122	0
All	All	850/868 (97%)	-0.14	17 (2%) 65 72	28, 50, 83, 122	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	73	TYR	3.7
1	B	332	PHE	3.0
1	B	337	VAL	3.0
1	B	359	ALA	3.0
1	B	71	ALA	2.9
1	B	385	TRP	2.8
1	B	118	ILE	2.5
1	B	338	TYR	2.5
1	B	72	PHE	2.4
1	B	277	VAL	2.4
1	B	29	VAL	2.4
1	B	370	TYR	2.4
1	B	74	CYS	2.3
1	B	335	SER	2.2
1	B	369	PRO	2.2
1	B	305	PHE	2.0
1	B	362	TYR	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)

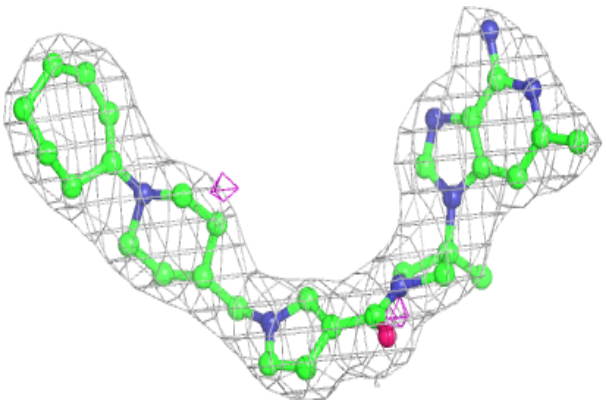
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

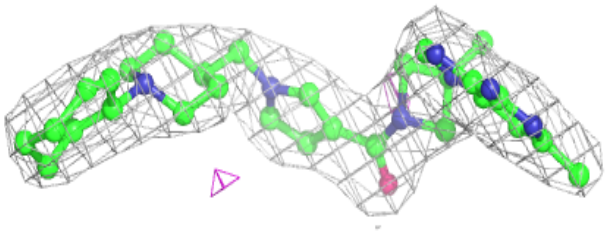
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	EW4	B	1004	37/37	0.95	0.15	33,38,41,43	0
3	EW4	A	1004	37/37	0.95	0.14	27,33,41,41	0
2	ZN	B	1001	1/1	0.98	0.07	80,80,80,80	0
2	ZN	A	1002	1/1	0.99	0.08	53,53,53,53	0
2	ZN	B	1002	1/1	0.99	0.09	64,64,64,64	0
2	ZN	A	1001	1/1	0.99	0.12	50,50,50,50	0
2	ZN	A	1003	1/1	0.99	0.13	39,39,39,39	0
2	ZN	B	1003	1/1	1.00	0.08	45,45,45,45	0

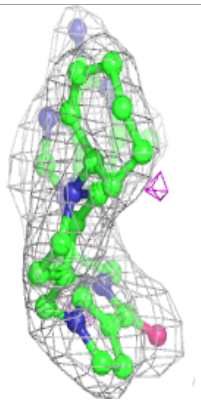
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around EW4 B 1004:

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

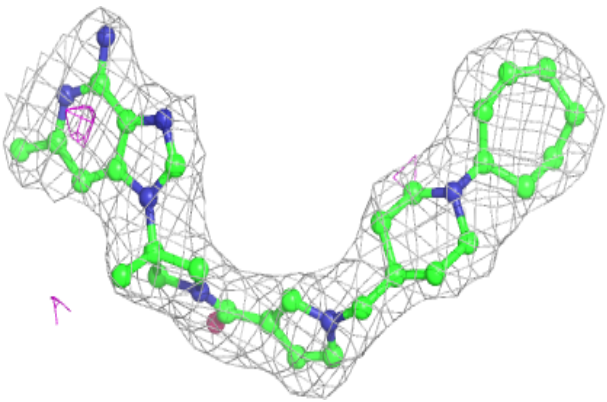


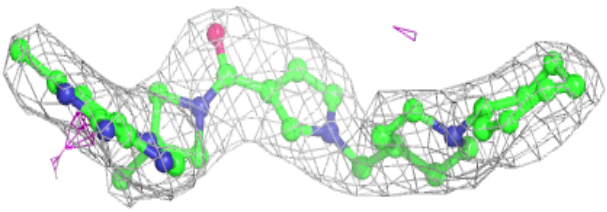


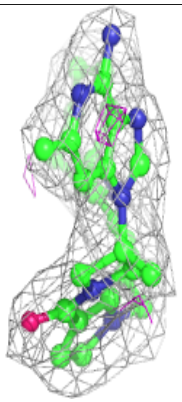


Electron density around EW4 A 1004:

$2mF_o$-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.