wwPDB X-ray Structure Validation Summary Report (i)

Sep 24, 2023 – 11:49 PM EDT

PDB ID	:	5WFD
Title	:	Humanized mutant of the Chaetomium thermophilum Polycomb Repressive Complex 2 bound to the inhibitor GSK126
Authors	:	Bratkowski, M.A.; Liu, X.
Deposited on	:	2017-07-11
Resolution	:	2.65 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

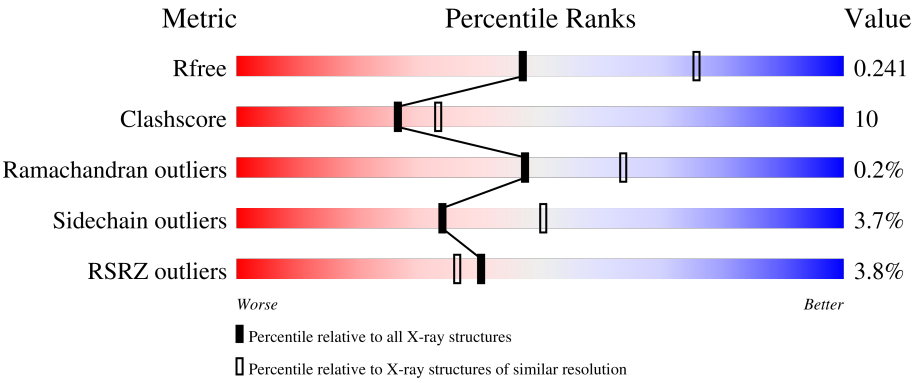
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.65 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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 of 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	605	2% 62% 14% 23%
2	B	936	4% 63% 19% 17%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 9956 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 Polycomb Protein EED.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	466	3682	2354	635	674	19	0	2	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-39	MET	-	initiating methionine	UNP G0S8H7
A	-38	ALA	-	expression tag	UNP G0S8H7
A	-37	SER	-	expression tag	UNP G0S8H7
A	-36	ALA	-	expression tag	UNP G0S8H7
A	-35	TRP	-	expression tag	UNP G0S8H7
A	-34	SER	-	expression tag	UNP G0S8H7
A	-33	HIS	-	expression tag	UNP G0S8H7
A	-32	PRO	-	expression tag	UNP G0S8H7
A	-31	GLN	-	expression tag	UNP G0S8H7
A	-30	PHE	-	expression tag	UNP G0S8H7
A	-29	GLU	-	expression tag	UNP G0S8H7
A	-28	LYS	-	expression tag	UNP G0S8H7
A	-27	GLY	-	expression tag	UNP G0S8H7
A	-26	GLY	-	expression tag	UNP G0S8H7
A	-25	GLY	-	expression tag	UNP G0S8H7
A	-24	SER	-	expression tag	UNP G0S8H7
A	-23	GLY	-	expression tag	UNP G0S8H7
A	-22	GLY	-	expression tag	UNP G0S8H7
A	-21	GLY	-	expression tag	UNP G0S8H7
A	-20	SER	-	expression tag	UNP G0S8H7
A	-19	GLY	-	expression tag	UNP G0S8H7
A	-18	GLY	-	expression tag	UNP G0S8H7
A	-17	SER	-	expression tag	UNP G0S8H7
A	-16	ALA	-	expression tag	UNP G0S8H7
A	-15	TRP	-	expression tag	UNP G0S8H7
A	-14	SER	-	expression tag	UNP G0S8H7
A	-13	HIS	-	expression tag	UNP G0S8H7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	PRO	-	expression tag	UNP G0S8H7
A	-11	GLN	-	expression tag	UNP G0S8H7
A	-10	PHE	-	expression tag	UNP G0S8H7
A	-9	GLU	-	expression tag	UNP G0S8H7
A	-8	LYS	-	expression tag	UNP G0S8H7
A	-7	LEU	-	expression tag	UNP G0S8H7
A	-6	GLU	-	expression tag	UNP G0S8H7
A	-5	VAL	-	expression tag	UNP G0S8H7
A	-4	LEU	-	expression tag	UNP G0S8H7
A	-3	PHE	-	expression tag	UNP G0S8H7
A	-2	GLN	-	expression tag	UNP G0S8H7
A	-1	GLY	-	expression tag	UNP G0S8H7
A	0	PRO	-	expression tag	UNP G0S8H7

- Molecule 2 is a protein called Histone-lysine-N-methyltransferase EZH2, Polycomb protein SUZ12 chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	774	Total	C	N	O	S	0	0	0
			6227	3925	1125	1137	40			

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	182	SER	-	expression tag	UNP G0SDW4
B	183	ASN	-	expression tag	UNP G0SDW4
B	184	HIS	-	expression tag	UNP G0SDW4
B	185	HIS	-	expression tag	UNP G0SDW4
B	186	HIS	-	expression tag	UNP G0SDW4
B	187	HIS	-	expression tag	UNP G0SDW4
B	188	HIS	-	expression tag	UNP G0SDW4
B	189	HIS	-	expression tag	UNP G0SDW4
B	190	ALA	-	expression tag	UNP G0SDW4
B	302	ILE	PRO	engineered mutation	UNP G0SDW4
B	304	TYR	ARG	engineered mutation	UNP G0SDW4
B	?	-	GLU	deletion	UNP G0SDW4
B	850	LYS	GLU	engineered mutation	UNP G0SDW4
B	851	TYR	ASN	engineered mutation	UNP G0SDW4
B	852	MET	LYS	engineered mutation	UNP G0SDW4
B	853	CYS	VAL	engineered mutation	UNP G0SDW4
B	855	PHE	TYR	engineered mutation	UNP G0SDW4
B	2524	LEU	-	linker	UNP G0SDW4

Continued on next page...

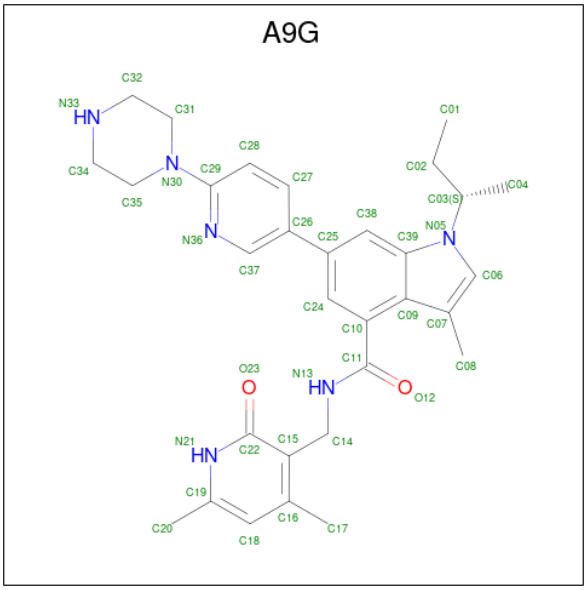
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	2525	VAL	-	linker	UNP G0SDW4
B	2526	PRO	-	linker	UNP G0SDW4
B	2527	ARG	-	linker	UNP G0SDW4
B	2528	GLY	-	linker	UNP G0SDW4
B	2529	SER	-	linker	UNP G0SDW4

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	8	Total Zn 8 8	0	0

- Molecule 4 is 1-[(2S)-butan-2-yl]-N-[(4,6-dimethyl-2-oxo-1,2-dihydropyridin-3-yl)methyl]-3-methyl-6-[6-(piperazin-1-yl)pyridin-3-yl]-1H-indole-4-carboxamide (three-letter code: A9G) (formula: $C_{31}H_{38}N_6O_2$).



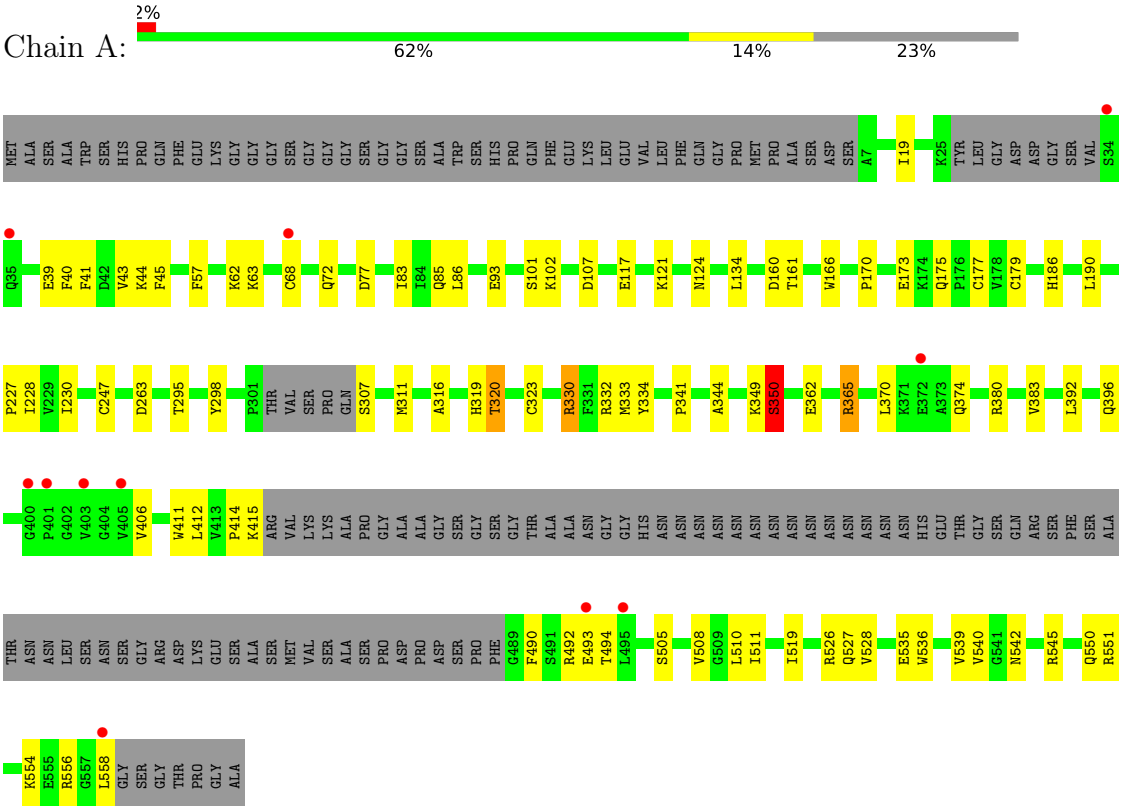
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C N O 39 31 6 2	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

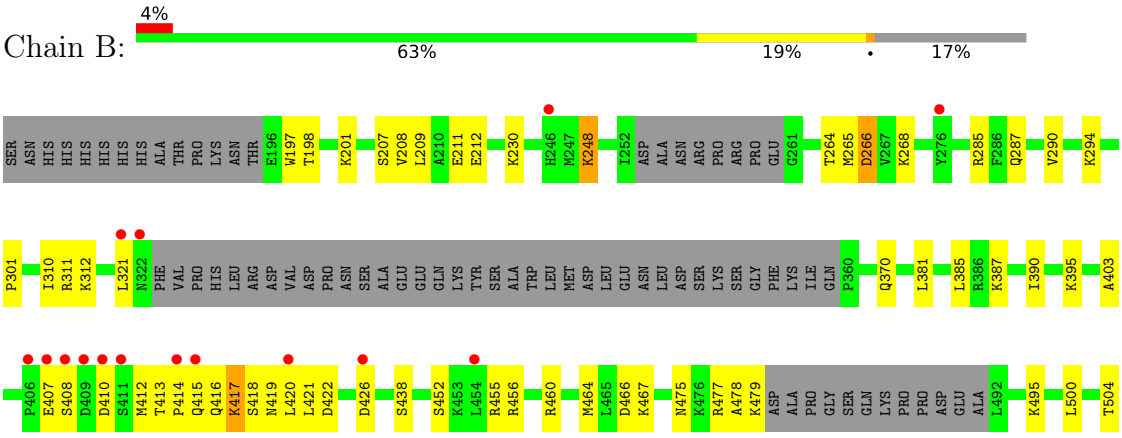
• Molecule 1: Polycomb Protein EED

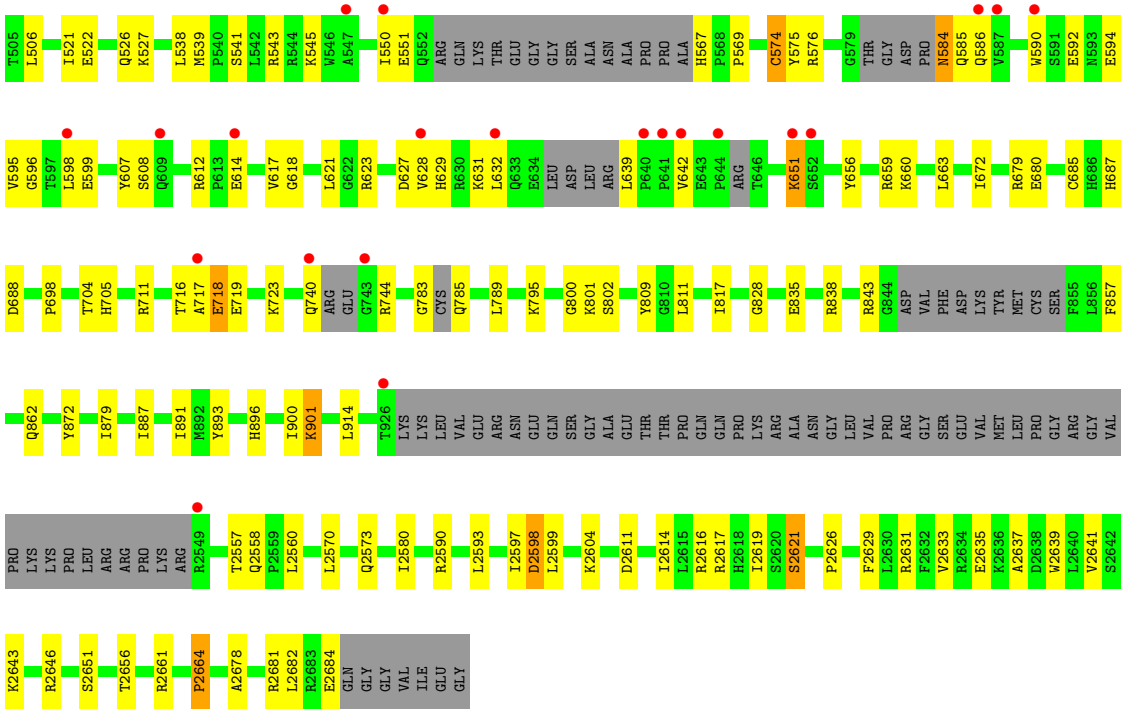
Chain A:



• Molecule 2: Histone-lysine-N-methyltransferase EZH2, Polycomb protein SUZ12 chimera

Chain B:




T505 L506 I521 E522 Q526 K527 L538 M539 P540 S541 L542 R543 R544 K545 W546 A547 I550 E551 Q552 ARG GLN LYS THR GLU GLY GLY SER ALA ASN ALA PRO PRO ALA H567 P568 P569 C574 Y575 R576 G579 THR GLY ASP PRO N584 Q585 Q586 V587 W590 S591 E592 N593 E594
V595 G596 T597 L598 E599 Y607 S608 Q609 R612 P613 E614 V617 G618 L621 G622 R623 D627 V628 H629 R630 K631 L632 Q633 E634 LEU ASP LEU ARG L639 P640 P641 V642 E643 P644 ARG T646 K651 S652 Y656 R659 K660 L663 I672 R679 E680 C685 H686 H687
D688 P698 T704 H705 R711 T716 A717 E718 E719 K723 Q740 ARG GLU G743 R744 G783 CYS Q785 L789 K795 G800 K801 S802 Y809 G810 L811 I817 G828 E835 R838 R843 G844 ASP VAL PHE ASP LYS TYR MET CYS SER F855 L856 F857
Q862 Y872 I879 I887 I891 M892 Y893 H896 I900 K901 L914 T926 LYS LYS LEU VAL GLU ARG ASN GLU GLN SER GLY ALA GLU THR THR PRO GLN GLN PRO LYS ARG ALA ASN GLY LEU VAL PRO ARG GLY SER GLU VAL MET LEU PRO GLY ARG GLY VAL
PRO LYS LYS PRO LEU ARG ARG PRO LYS ARG R2549 T2557 Q2558 P2559 L2560 L2570 Q2573 I2580 R2590 L2593 I2597 D2598 L2599 K2604 D2611 I2614 L2615 R2616 R2617 H2618 I2619 S2620 S2621 P2626 F2629 L2630 R2631 F2632 V2633 R2634 E2635 K2636 A2637 D2638 W2639 L2640 V2641 S2642
K2643 R2646 S2651 T2656 R2661 P2664 A2678 R2681 L2682 R2683 E2684 GLN GLY GLY VAL ILE GLU GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	117.53Å 136.97Å 223.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.15 – 2.65 46.15 – 2.65	Depositor EDS
% Data completeness (in resolution range)	95.8 (46.15-2.65) 95.8 (46.15-2.65)	Depositor EDS
R_{merge}	0.21	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.59 (at 2.65Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575)	Depositor
R, R_{free}	0.181 , 0.240 0.182 , 0.241	Depositor DCC
R_{free} test set	2476 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å^2)	42.1	Xtriage
Anisotropy	0.057	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 42.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9956	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.63% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, A9G

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	0.51	0/3794	0.66	0/5163
2	B	0.48	0/6368	0.63	0/8602
All	All	0.49	0/10162	0.64	0/13765

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

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	3682	0	3557	73	0
2	B	6227	0	6113	134	0
3	B	8	0	0	0	0
4	B	39	0	0	1	0
All	All	9956	0	9670	202	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 10.

The worst 5 of 202 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:68:CYS:SG	1:A:83:ILE:HG12	1.90	1.10
1:A:68:CYS:SG	1:A:83:ILE:CG1	2.56	0.92
1:A:72:GLN:OE1	1:A:554:LYS:NZ	2.03	0.91
2:B:415:GLN:HA	2:B:418:SER:HB3	1.53	0.91
2:B:717:ALA:HB2	2:B:723:LYS:HG2	1.57	0.85

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	460/605 (76%)	443 (96%)	16 (4%)	1 (0%)	47	64
2	B	750/936 (80%)	710 (95%)	39 (5%)	1 (0%)	51	69
All	All	1210/1541 (78%)	1153 (95%)	55 (4%)	2 (0%)	47	64

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	718	GLU
1	A	350	SER

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	393/495 (79%)	380 (97%)	13 (3%)	38	54
2	B	677/815 (83%)	650 (96%)	27 (4%)	31	47
All	All	1070/1310 (82%)	1030 (96%)	40 (4%)	34	50

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	651	LYS
2	B	2590	ARG
2	B	659	ARG
2	B	740	GLN
2	B	2621	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 8 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
4	A9G	B	8009	-	37,43,43	2.66	17 (45%)	46,62,62	2.68	14 (30%)

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	A9G	B	8009	-	-	2/19/31/31	0/5/5/5

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	8009	A9G	C29-N30	6.35	1.50	1.37
4	B	8009	A9G	O23-C22	6.19	1.35	1.23
4	B	8009	A9G	C10-C11	5.44	1.59	1.50
4	B	8009	A9G	C09-C39	4.20	1.48	1.42
4	B	8009	A9G	C24-C25	4.14	1.46	1.39

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	8009	A9G	C15-C22-N21	10.08	123.22	115.69
4	B	8009	A9G	O23-C22-C15	-8.05	114.71	124.54
4	B	8009	A9G	C22-N21-C19	-5.12	120.24	125.20
4	B	8009	A9G	N36-C29-N30	4.63	123.73	116.79
4	B	8009	A9G	C28-C29-N36	-4.09	115.90	123.41

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	8009	A9G	C28-C29-N30-C35
4	B	8009	A9G	N36-C29-N30-C35

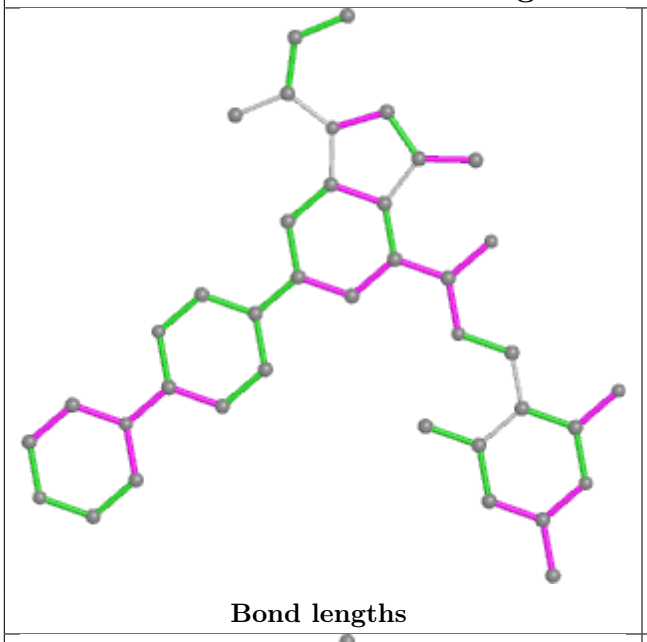
There are no ring outliers.

1 monomer is involved in 1 short contact:

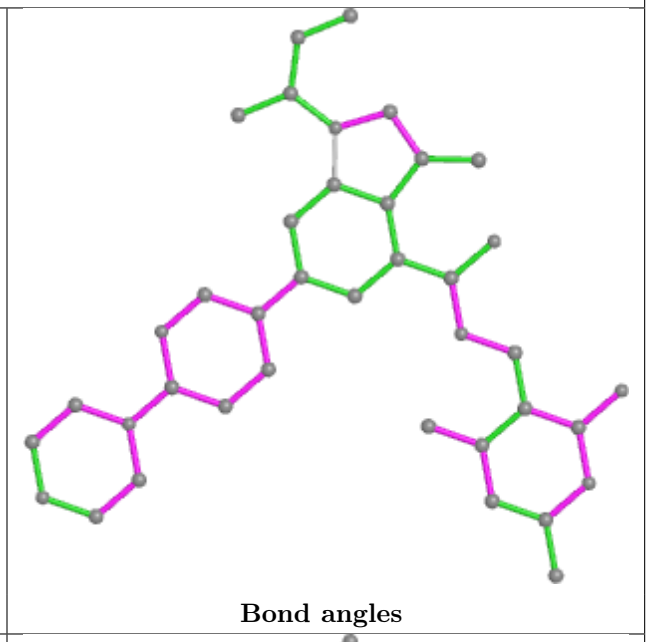
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	8009	A9G	1	0

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.

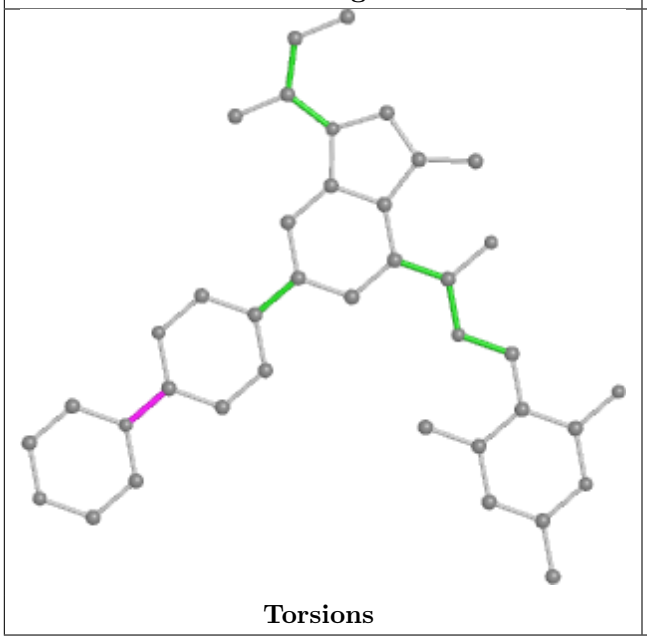
Ligand A9G B 8009



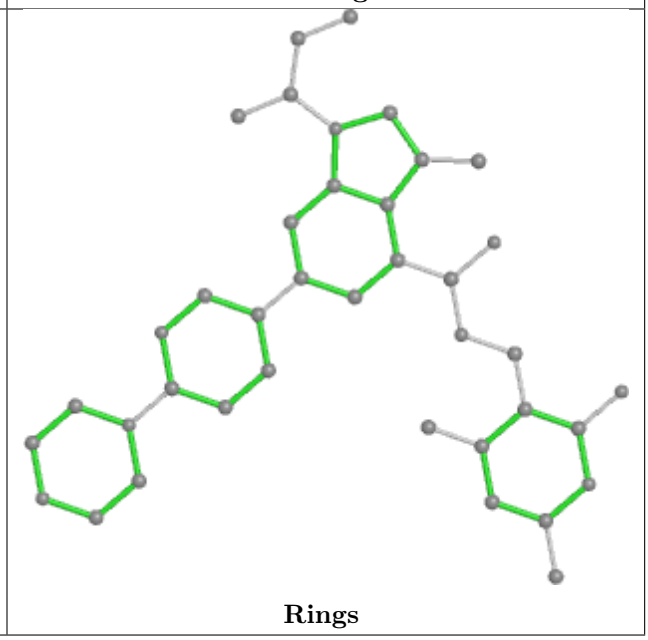
Bond lengths



Bond angles



Torsions



Rings

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	466/605 (77%)	-0.12	11 (2%)	59	54	17, 35, 87, 123	0
2	B	774/936 (82%)	-0.01	36 (4%)	31	28	19, 48, 97, 137	0
All	All	1240/1541 (80%)	-0.05	47 (3%)	40	36	17, 43, 95, 137	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	403	VAL	5.5
2	B	408	SER	5.0
2	B	628	VAL	4.6
2	B	926	THR	4.5
2	B	586	GLN	4.2

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

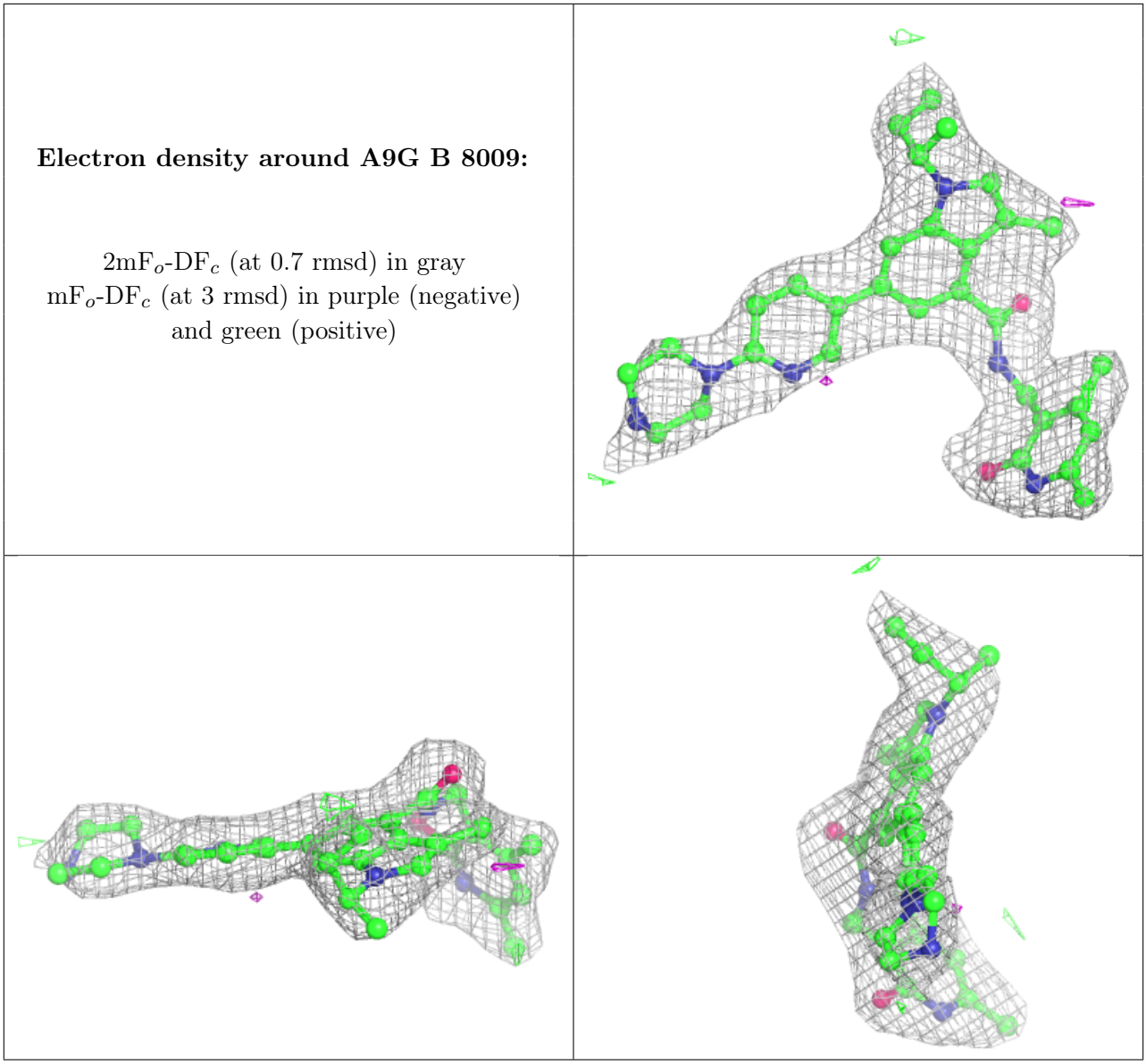
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	ZN	B	8008	1/1	0.93	0.06	88,88,88,88	0
4	A9G	B	8009	39/39	0.95	0.16	16,37,67,73	0
3	ZN	B	8002	1/1	0.97	0.18	34,34,34,34	0
3	ZN	B	8005	1/1	0.99	0.11	44,44,44,44	0
3	ZN	B	8006	1/1	0.99	0.14	44,44,44,44	0
3	ZN	B	8001	1/1	0.99	0.12	34,34,34,34	0
3	ZN	B	8004	1/1	0.99	0.12	43,43,43,43	0
3	ZN	B	8003	1/1	1.00	0.14	42,42,42,42	0
3	ZN	B	8007	1/1	1.00	0.13	28,28,28,28	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 A9G B 8009:

$2mF_o$-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.