



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 25, 2023 – 11:24 AM EDT

PDB ID : 5WG6
Title : Human Polycomb Repressive Complex 2 in complex with GSK126 inhibitor
Authors : Bratkowski, M.A.; Liu, X.
Deposited on : 2017-07-13
Resolution : 3.90 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 ⓘ 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 ⓘ](#)) 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

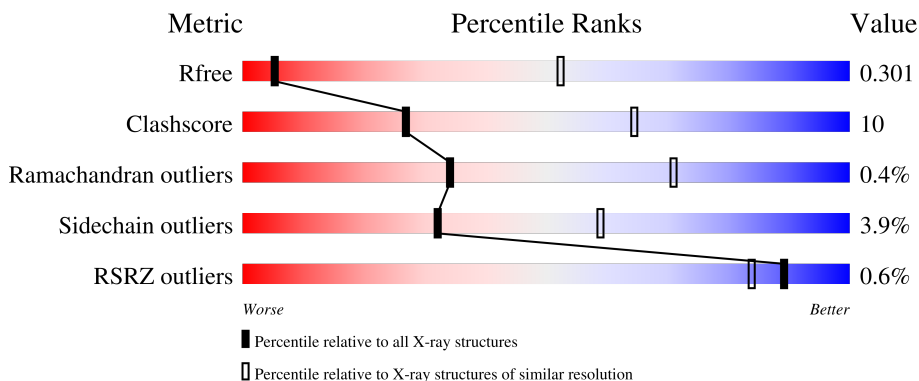
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.90 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	1002 (4.14-3.66)
Clashscore	141614	1004 (4.12-3.68)
Ramachandran outliers	138981	1021 (4.14-3.66)
Sidechain outliers	138945	1014 (4.14-3.66)
RSRZ outliers	127900	1275 (4.20-3.60)

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 $\leq 5\%$. 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	859	
1	C	859	
2	B	442	
2	D	442	

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 13474 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 Histone-lysine N-methyltransferase EZH2,Polycomb protein SUZ12 (E.C.2.1.1.43) chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	517	4210	2642	752	770	46	0	0	0
1	C	504	4092	2565	729	751	47	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	SER	-	expression tag	UNP Q15910
A	-6	ASN	-	expression tag	UNP Q15910
A	-5	ALA	-	expression tag	UNP Q15910
A	-4	HIS	-	expression tag	UNP Q15910
A	-3	HIS	-	expression tag	UNP Q15910
A	-2	HIS	-	expression tag	UNP Q15910
A	-1	HIS	-	expression tag	UNP Q15910
A	0	HIS	-	expression tag	UNP Q15910
A	1	HIS	-	expression tag	UNP Q15910
A	594	ARG	TRP	see remark 999	UNP Q15910
C	-7	SER	-	expression tag	UNP Q15910
C	-6	ASN	-	expression tag	UNP Q15910
C	-5	ALA	-	expression tag	UNP Q15910
C	-4	HIS	-	expression tag	UNP Q15910
C	-3	HIS	-	expression tag	UNP Q15910
C	-2	HIS	-	expression tag	UNP Q15910
C	-1	HIS	-	expression tag	UNP Q15910
C	0	HIS	-	expression tag	UNP Q15910
C	1	HIS	-	expression tag	UNP Q15910
C	594	ARG	TRP	see remark 999	UNP Q15910

- Molecule 2 is a protein called Polycomb protein EED.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	355	Total 2873	C 1819	N 503	O 530	S 21	0	0	0
2	D	274	Total 2244	C 1425	N 394	O 408	S 17	0	0	0

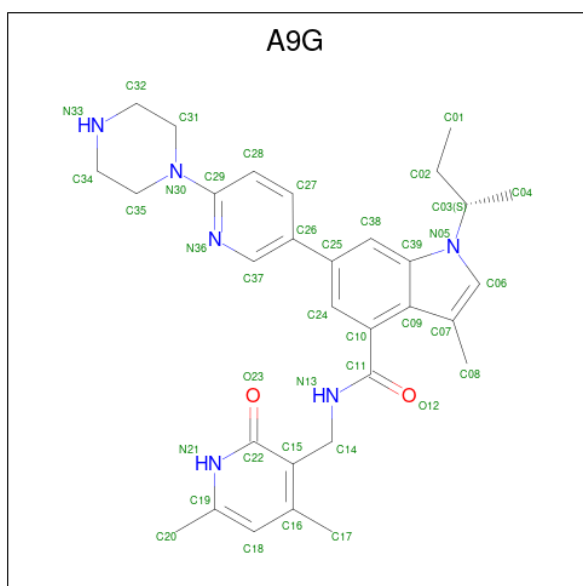
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP O75530
B	1	SER	-	expression tag	UNP O75530
D	0	MET	-	initiating methionine	UNP O75530
D	1	SER	-	expression tag	UNP O75530

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

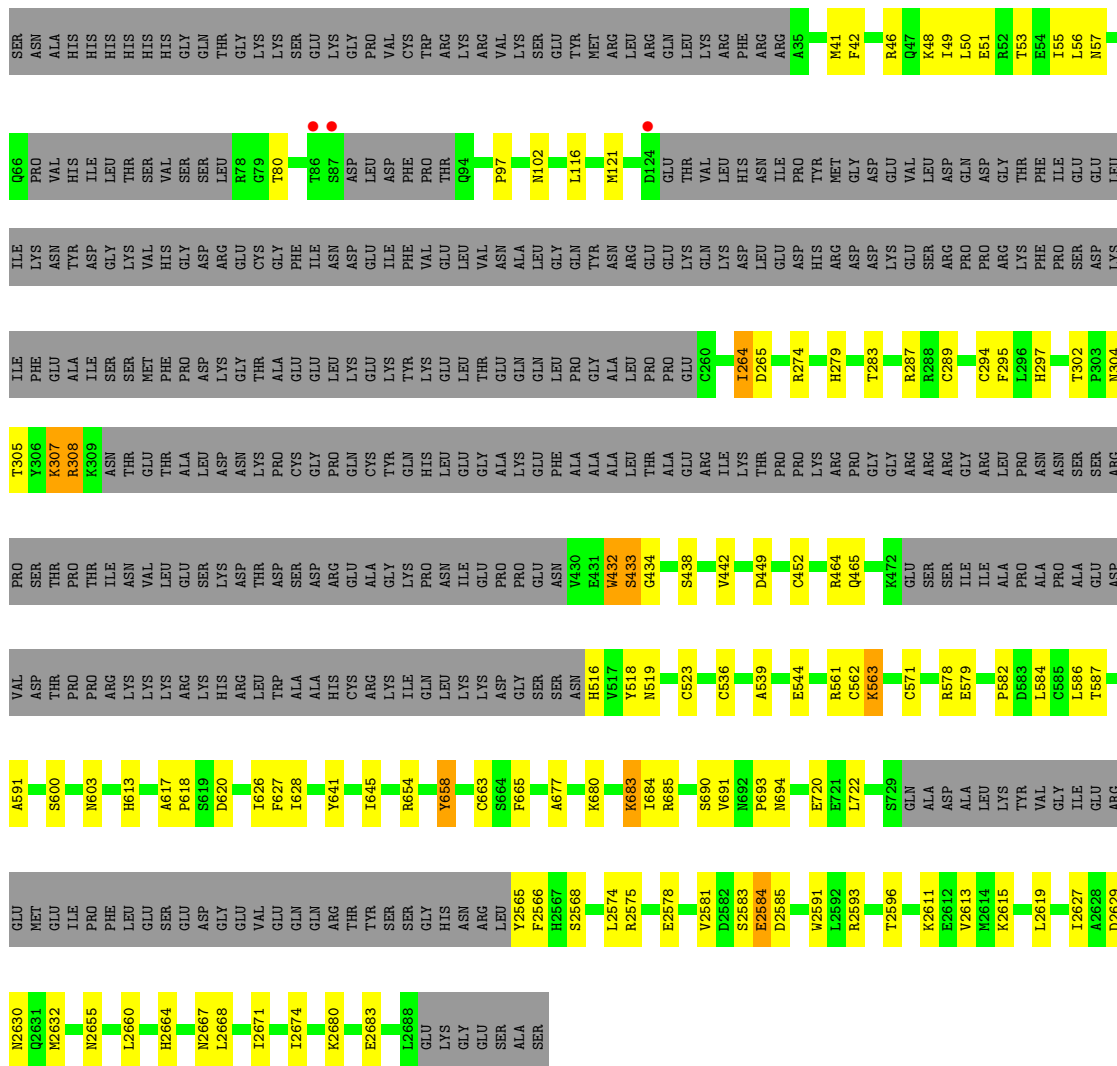
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	8	Total 8	Zn 8	0	0
3	C	8	Total 8	Zn 8	0	0

- 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: C₃₁H₃₈N₆O₂).



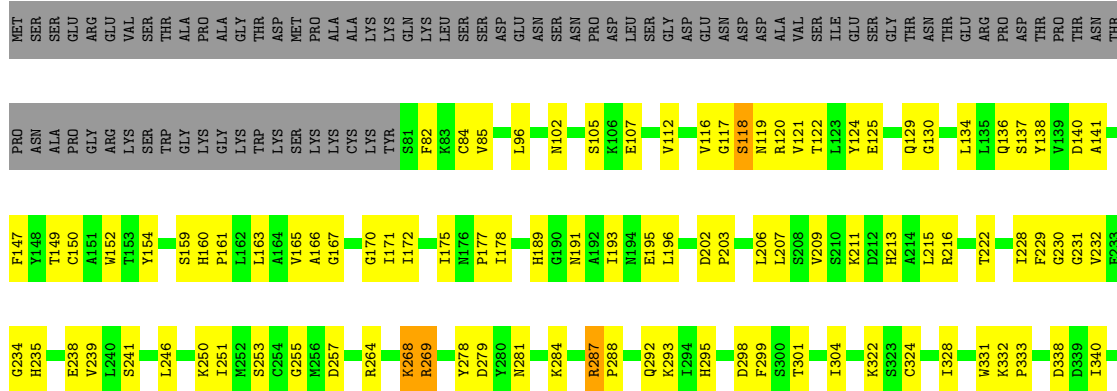
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
4	A	1	39	31	6	2	0	0

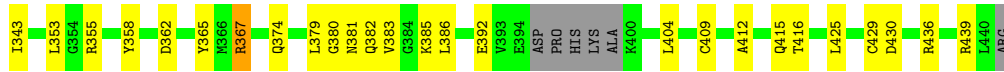
Chain C: 45% 12% 41%



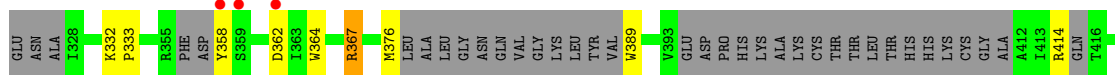
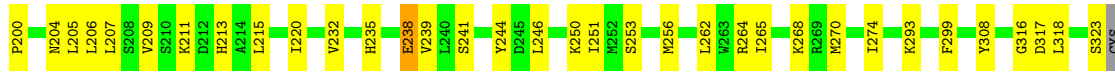
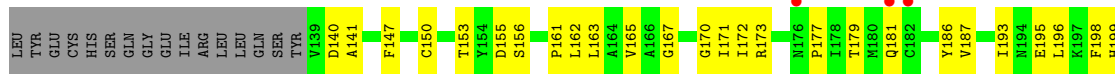
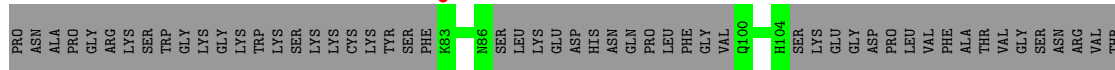
● Molecule 2: Polycomb protein EED

Chain B: 52% 27% 20%





● Molecule 2: Polycomb protein EED



4 Data and refinement statistics i

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, α , β , γ	243.88Å 243.88Å 243.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.65 – 3.90 40.65 – 3.90	Depositor EDS
% Data completeness (in resolution range)	79.7 (40.65-3.90) 79.7 (40.65-3.90)	Depositor EDS
R_{merge}	0.38	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.84 (at 3.87Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575)	Depositor
R, R_{free}	0.242 , 0.300 0.242 , 0.301	Depositor DCC
R_{free} test set	1367 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å ²)	28.0	Xtriage
Anisotropy	0.798	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.23 , 15.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.015 for l,-k,h 0.017 for -h,-l,-k	Xtriage
F_o, F_c correlation	0.80	EDS
Total number of atoms	13474	wwPDB-VP
Average B, all atoms (Å ²)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.11% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: A9G, ZN

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	0.28	0/4301	0.50	0/5790
1	C	0.29	0/4178	0.50	0/5619
2	B	0.28	0/2944	0.50	0/3986
2	D	0.26	0/2296	0.47	0/3100
All	All	0.28	0/13719	0.50	0/18495

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	4210	0	4084	90	0
1	C	4092	0	3964	72	0
2	B	2873	0	2791	80	0
2	D	2244	0	2176	49	0
3	A	8	0	0	0	0
3	C	8	0	0	0	0
4	A	39	0	0	1	0
All	All	13474	0	13015	269	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 269 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)
2:D:358:TYR:CE1	2:D:389:TRP:HH2	1.46	1.33
2:D:358:TYR:CE1	2:D:389:TRP:CH2	2.20	1.29
2:D:358:TYR:HE1	2:D:389:TRP:CH2	1.51	1.24
1:A:2568:SER:OG	1:A:2584:GLU:OE2	1.65	1.10
1:A:271:SER:HA	1:A:438:SER:HB2	1.63	0.81

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	505/859 (59%)	457 (90%)	46 (9%)	2 (0%)	34	71
1	C	490/859 (57%)	439 (90%)	47 (10%)	4 (1%)	19	57
2	B	351/442 (79%)	303 (86%)	48 (14%)	0	100	100
2	D	258/442 (58%)	229 (89%)	28 (11%)	1 (0%)	34	71
All	All	1604/2602 (62%)	1428 (89%)	169 (10%)	7 (0%)	34	71

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	463	CYS
1	C	683	LYS
1	C	2584	GLU
1	C	2581	VAL
2	D	238	GLU

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	474/773 (61%)	452 (95%)	22 (5%)	27	55
1	C	460/773 (60%)	444 (96%)	16 (4%)	36	62
2	B	319/393 (81%)	302 (95%)	17 (5%)	22	52
2	D	249/393 (63%)	245 (98%)	4 (2%)	62	79
All	All	1502/2332 (64%)	1443 (96%)	59 (4%)	32	59

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	268	LYS
2	D	241	SER
2	B	362	ASP
1	C	2575	ARG
1	C	563	LYS

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 17 ligands modelled in this entry, 16 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	A	9009	-	37,43,43	1.94	8 (21%)	46,62,62	3.25	19 (41%)

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	A	9009	-	-	4/19/31/31	0/5/5/5

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	9009	A9G	O23-C22	5.19	1.33	1.23
4	A	9009	A9G	C11-N13	4.85	1.44	1.33
4	A	9009	A9G	C29-N30	4.68	1.47	1.37
4	A	9009	A9G	C10-C11	2.88	1.55	1.50
4	A	9009	A9G	C35-N30	2.72	1.51	1.46

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	9009	A9G	C15-C22-N21	14.05	126.19	115.69
4	A	9009	A9G	C22-N21-C19	-8.26	117.19	125.20
4	A	9009	A9G	O23-C22-C15	-6.65	116.42	124.54
4	A	9009	A9G	N36-C29-N30	4.71	123.84	116.79
4	A	9009	A9G	C28-C29-N36	-4.06	115.97	123.41

There are no chirality outliers.

All (4) torsion outliers are listed below:

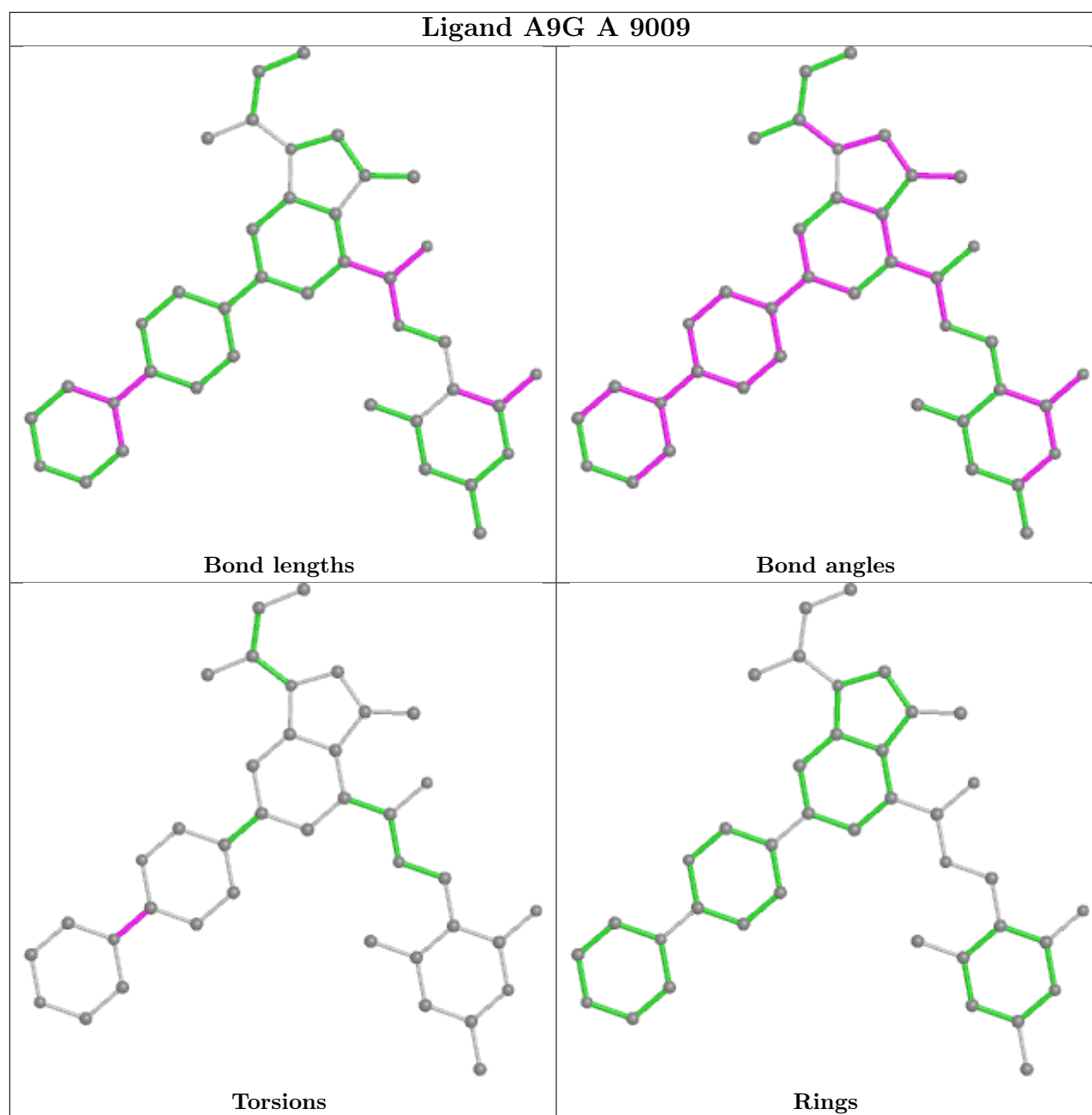
Mol	Chain	Res	Type	Atoms
4	A	9009	A9G	C28-C29-N30-C31
4	A	9009	A9G	N36-C29-N30-C31
4	A	9009	A9G	C28-C29-N30-C35
4	A	9009	A9G	N36-C29-N30-C35

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	9009	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 than 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.



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	517/859 (60%)	-0.39	0 100 100	11, 41, 85, 125	0
1	C	504/859 (58%)	-0.28	3 (0%) 89 84	16, 47, 88, 120	0
2	B	355/442 (80%)	-0.55	0 100 100	13, 31, 54, 91	0
2	D	274/442 (61%)	-0.02	7 (2%) 56 45	35, 59, 91, 127	0
All	All	1650/2602 (63%)	-0.33	10 (0%) 89 84	11, 44, 85, 127	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	124	ASP	2.7
2	D	83	LYS	2.6
1	C	86	THR	2.4
1	C	87	SER	2.3
2	D	358	TYR	2.2

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](#)

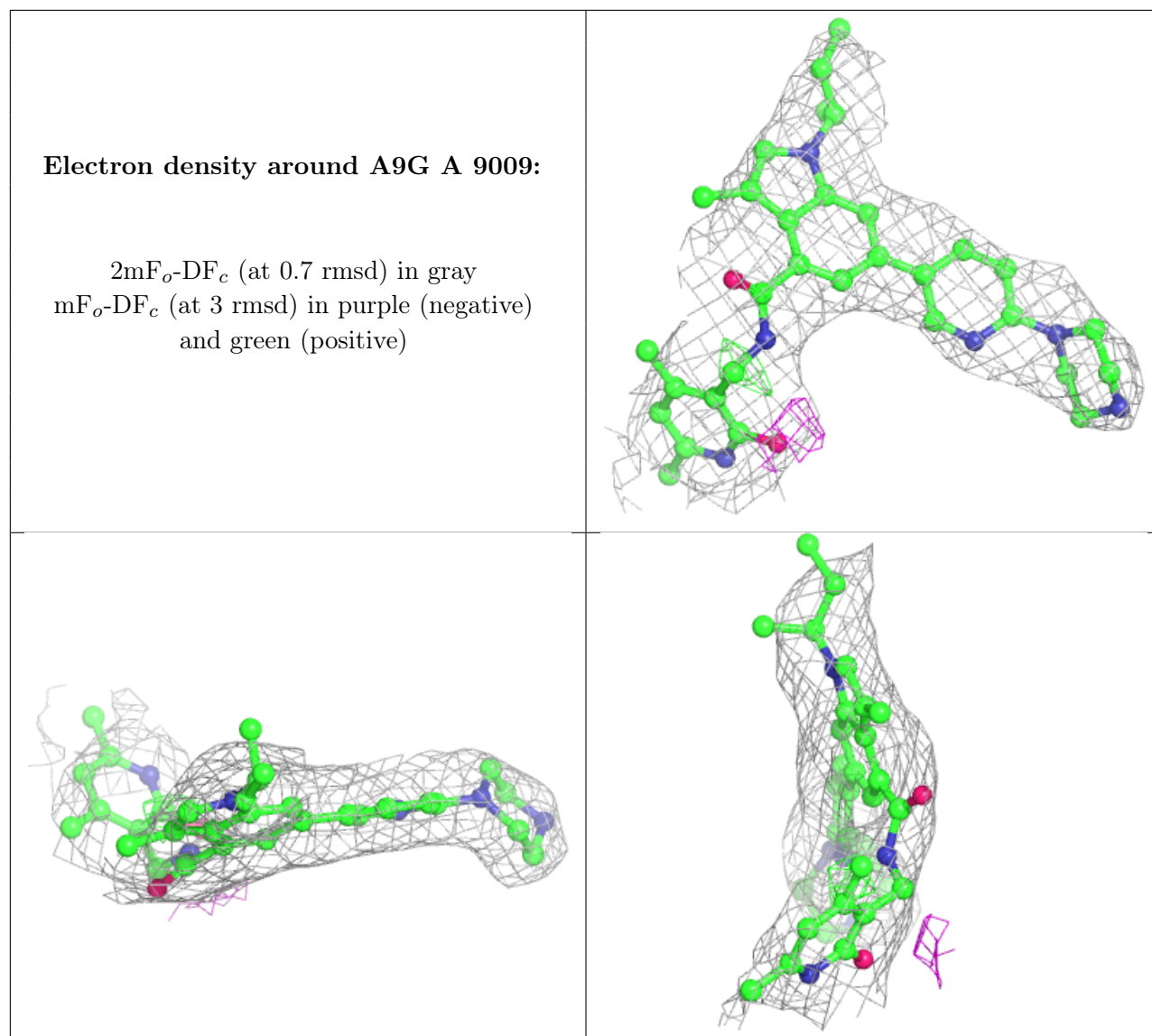
There are no monosaccharides in this entry.

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(\AA^2)	Q<0.9
3	ZN	A	9001	1/1	0.83	0.10	99,99,99,99	0
4	A9G	A	9009	39/39	0.93	0.22	17,21,24,24	0
3	ZN	C	9004	1/1	0.96	0.06	48,48,48,48	0
3	ZN	C	9002	1/1	0.97	0.06	43,43,43,43	0
3	ZN	A	9002	1/1	0.97	0.06	48,48,48,48	0
3	ZN	C	9001	1/1	0.97	0.05	71,71,71,71	0
3	ZN	C	9005	1/1	0.98	0.14	36,36,36,36	0
3	ZN	C	9007	1/1	0.98	0.11	37,37,37,37	0
3	ZN	A	9008	1/1	0.98	0.13	38,38,38,38	0
3	ZN	A	9005	1/1	0.99	0.09	30,30,30,30	0
3	ZN	C	9003	1/1	0.99	0.04	37,37,37,37	0
3	ZN	A	9006	1/1	0.99	0.10	29,29,29,29	0
3	ZN	A	9007	1/1	0.99	0.10	28,28,28,28	0
3	ZN	C	9006	1/1	0.99	0.11	36,36,36,36	0
3	ZN	A	9003	1/1	0.99	0.08	37,37,37,37	0
3	ZN	C	9008	1/1	0.99	0.11	35,35,35,35	0
3	ZN	A	9004	1/1	0.99	0.06	41,41,41,41	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.



6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.