



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2026 – 11:35 AM UTC

PDB ID : 5WF7 / pdb_00005wf7
Title : Chaetomium thermophilum Polycomb Repressive Complex 2 bound to GSK126
Authors : Bratkowski, M.A.; Liu, X.
Deposited on : 2017-07-11
Resolution : 2.50 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Buster-report	:	wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

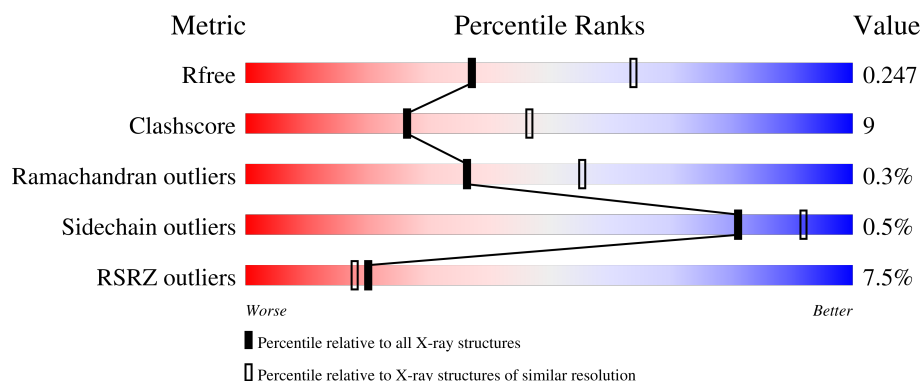
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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}	180053	5829 (2.50-2.50)
Clashscore	190562	6492 (2.50-2.50)
Ramachandran outliers	187476	6378 (2.50-2.50)
Sidechain outliers	187428	6380 (2.50-2.50)
RSRZ outliers	180081	5833 (2.50-2.50)

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	605	
2	B	937	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10065 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	453	3579	2289	614	657	19	0	0	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	765	Total	C	N	O	S	0	0	0
			6151	3885	1112	1114	40			

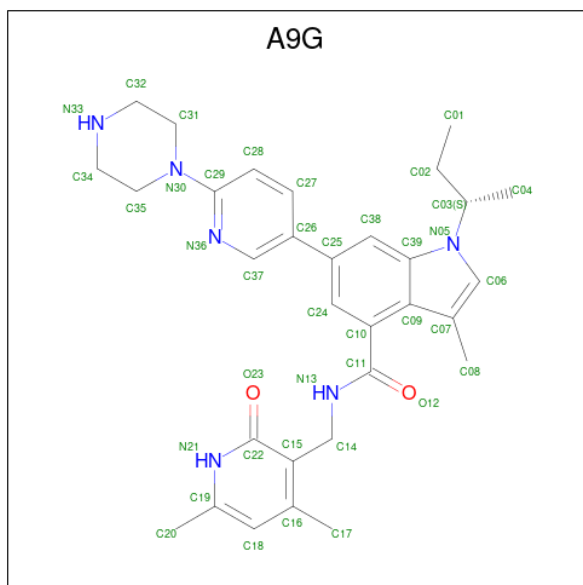
There are 15 discrepancies between the modelled and reference sequences:

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	2524	LEU	-	linker	UNP G0SDW4
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 (CCD ID: ZN) (formula: Zn).

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

- 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 (CCD ID: A9G) (formula: C₃₁H₃₈N₆O₂).



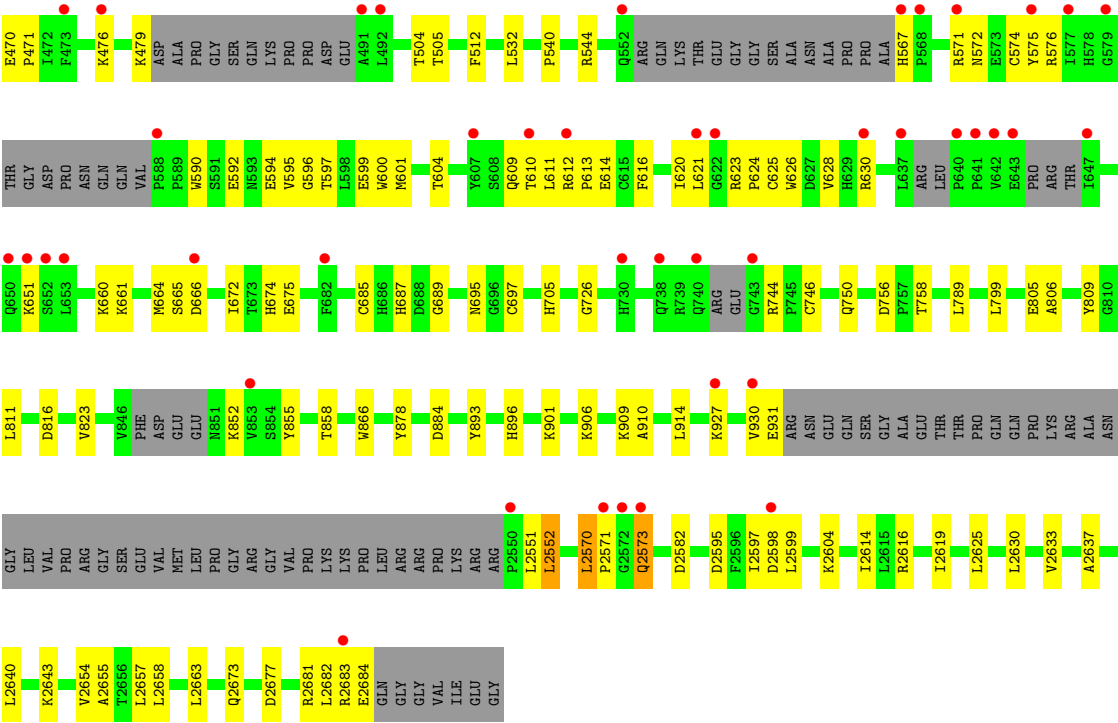
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total	C	N	O	0	0
			39	31	6	2		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	134	Total	O	0	0
			134	134		
5	B	154	Total	O	0	0
			154	154		

- Molecule 1: Polycomb Protein EED





4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	115.88Å 136.30Å 223.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.14 – 2.50 44.14 – 2.50	Depositor EDS
% Data completeness (in resolution range)	93.5 (44.14-2.50) 93.9 (44.14-2.50)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.35 (at 2.48Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575)	Depositor
R, R_{free}	0.191 , 0.247 0.191 , 0.247	Depositor DCC
R_{free} test set	2875 reflections (4.66%)	wwPDB-VP
Wilson B-factor (Å ²)	37.0	Xtriage
Anisotropy	0.125	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 57.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	10065	wwPDB-VP
Average B, all atoms (Å ²)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.65% 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: 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	0.40	0/3683	0.62	1/5013 (0.0%)
2	B	0.40	2/6290 (0.0%)	0.61	3/8491 (0.0%)
All	All	0.40	2/9973 (0.0%)	0.62	4/13504 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	823	VAL	C-N	-6.74	1.29	1.33
2	B	435	ARG	CA-C	-5.44	1.45	1.52

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	2570	LEU	CA-C-N	-6.54	112.77	119.83
2	B	2570	LEU	C-N-CA	-6.54	112.77	119.83
2	B	2552	LEU	CD1-CG-CD2	-6.48	96.55	110.80
1	A	392	LEU	N-CA-C	-5.68	98.63	108.52

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	3579	0	3437	54	0
2	B	6151	0	6065	134	0
3	B	8	0	0	0	0
4	B	39	0	0	0	0
5	A	134	0	0	1	0
5	B	154	0	0	6	0
All	All	10065	0	9502	180	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 9.

The worst 5 of 180 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:B:612:ARG:HG2	2:B:614:GLU:OE2	1.44	1.17
2:B:626:TRP:HE3	2:B:630:ARG:NH2	1.43	1.16
2:B:394:THR:HG22	2:B:397:ASN:HB2	1.41	1.02
2:B:612:ARG:HG3	2:B:613:PRO:HD2	1.46	0.96
2:B:626:TRP:CE3	2:B:630:ARG:NH2	2.31	0.96

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	443/605 (73%)	425 (96%)	17 (4%)	1 (0%)	43 63
2	B	739/937 (79%)	706 (96%)	30 (4%)	3 (0%)	30 49
All	All	1182/1542 (77%)	1131 (96%)	47 (4%)	4 (0%)	36 55

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	443	ALA
2	B	806	ALA
1	A	393	ALA
2	B	260	GLU

5.3.2 Protein sidechains ⓘ

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	382/495 (77%)	381 (100%)	1 (0%)	86	94
2	B	667/816 (82%)	663 (99%)	4 (1%)	78	91
All	All	1049/1311 (80%)	1044 (100%)	5 (0%)	81	92

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	300	THR
2	B	264	THR
2	B	388	LEU
2	B	628	VAL
2	B	2573	GLN

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

Mol	Chain	Res	Type
1	A	72	GLN
1	A	335	HIS
2	B	281	ASN
2	B	549	GLN

5.3.3 RNA ⓘ

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 oligosaccharides 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	-	40,43,43	2.82	20 (50%)	52,62,62	2.50	19 (36%)

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	-	-	4/22/31/31	0/5/5/5

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	8009	A9G	O23-C22	6.44	1.35	1.23
4	B	8009	A9G	C29-N30	6.34	1.51	1.37
4	B	8009	A9G	C10-C11	5.66	1.61	1.50
4	B	8009	A9G	C35-N30	4.86	1.55	1.46
4	B	8009	A9G	C11-N13	4.80	1.44	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	8009	A9G	C15-C22-N21	9.53	122.94	115.70
4	B	8009	A9G	O23-C22-C15	-6.27	116.84	124.56
4	B	8009	A9G	N36-C29-N30	4.72	123.51	116.76
4	B	8009	A9G	C28-C29-N36	-3.87	115.82	123.46
4	B	8009	A9G	C38-C25-C26	3.75	127.12	120.84

There are no chirality outliers.

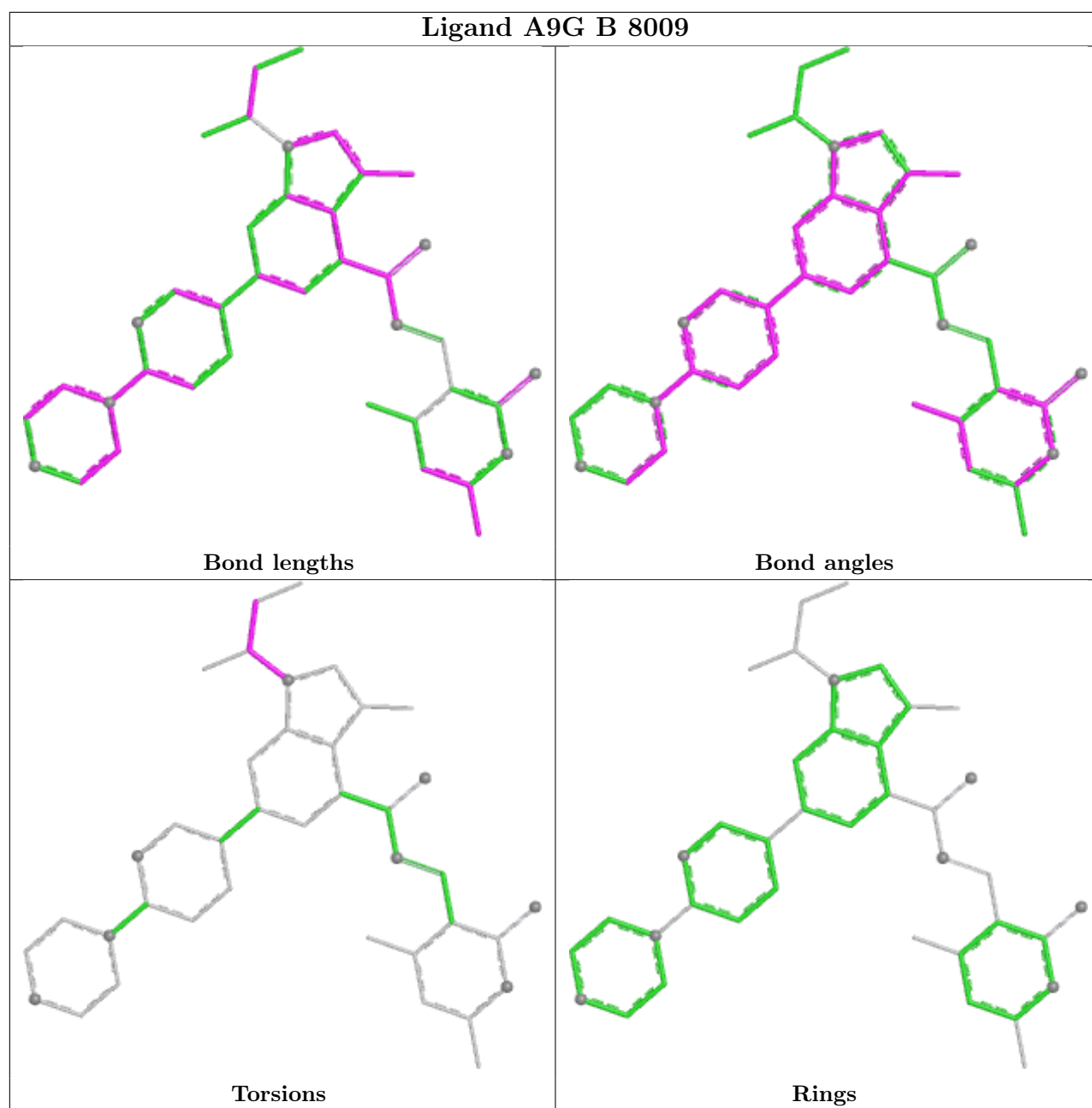
All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	8009	A9G	C01-C02-C03-C04
4	B	8009	A9G	C01-C02-C03-N05
4	B	8009	A9G	C02-C03-N05-C06
4	B	8009	A9G	C02-C03-N05-C39

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 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	453/605 (74%)	-0.19	21 (4%) 37 33	18, 35, 95, 126	0
2	B	765/937 (81%)	0.40	70 (9%) 14 13	18, 53, 93, 137	0
All	All	1218/1542 (78%)	0.18	91 (7%) 20 18	18, 47, 94, 137	0

The worst 5 of 91 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	579	GLY	6.4
2	B	254	ALA	6.2
2	B	2572	GLY	5.8
2	B	2550	PRO	5.6
2	B	435	ARG	5.1

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](#)

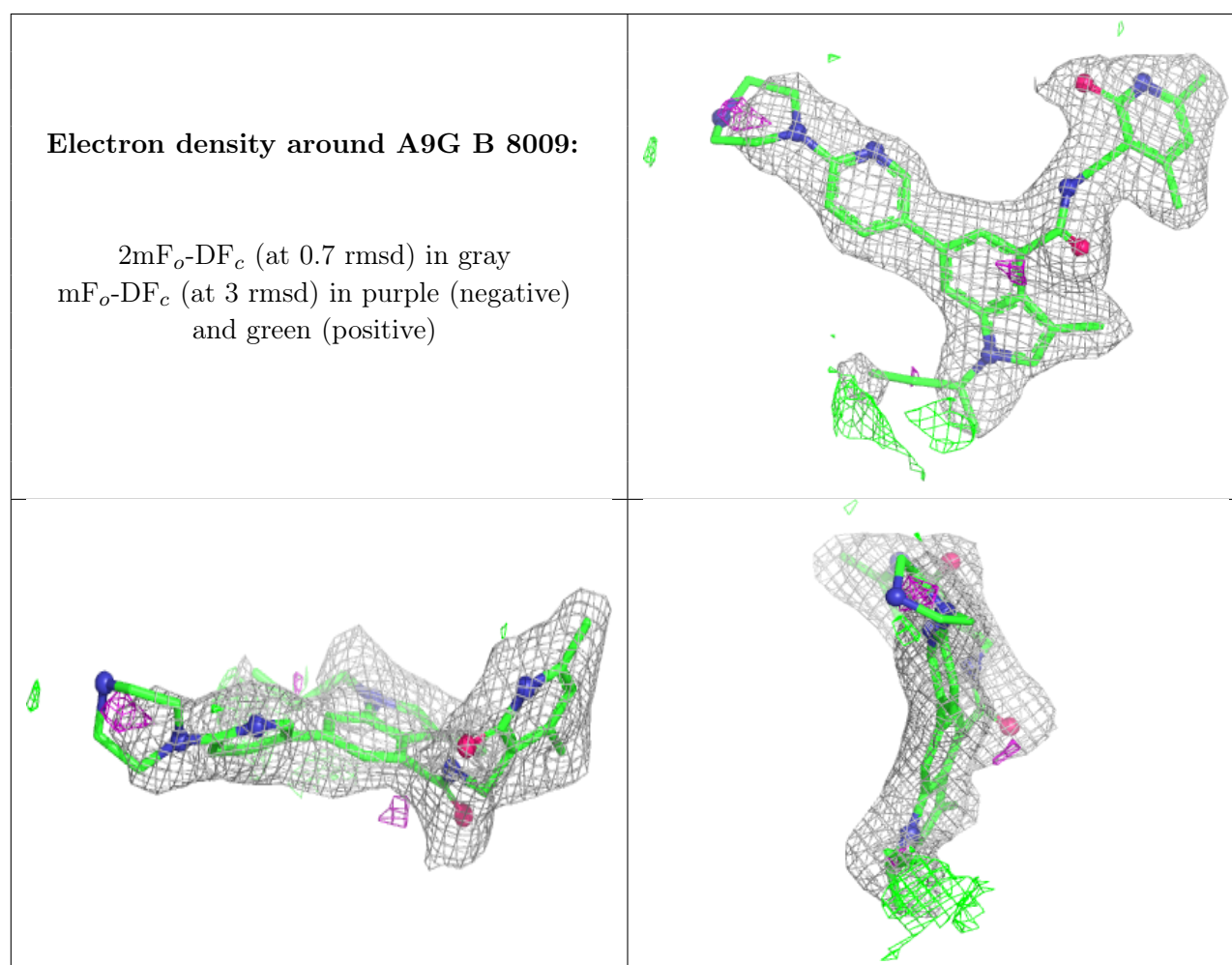
There are no oligosaccharides in this entry.

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($\text{\AA}^2$)	Q<0.9
4	A9G	B	8009	39/39	0.87	0.14	27,46,69,71	0
3	ZN	B	8006	1/1	0.99	0.03	48,48,48,48	0
3	ZN	B	8008	1/1	0.99	0.04	64,64,64,64	0
3	ZN	B	8004	1/1	0.99	0.02	52,52,52,52	0
3	ZN	B	8005	1/1	1.00	0.01	53,53,53,53	0
3	ZN	B	8002	1/1	1.00	0.04	44,44,44,44	0
3	ZN	B	8007	1/1	1.00	0.02	30,30,30,30	0
3	ZN	B	8003	1/1	1.00	0.04	47,47,47,47	0
3	ZN	B	8001	1/1	1.00	0.04	44,44,44,44	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 ⓘ

There are no such residues in this entry.