



wwPDB X-ray Structure Validation Summary Report ⓘ

May 28, 2024 – 04:25 PM EDT

PDB ID : 8SW8
Title : Crystal Structure of HaloTag7 bound to JF669-HaloTag ligand
Authors : Farrants, H.; Schreiter, E.R.
Deposited on : 2023-05-17
Resolution : 1.90 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 : 2.36.2
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) : 2.36.2

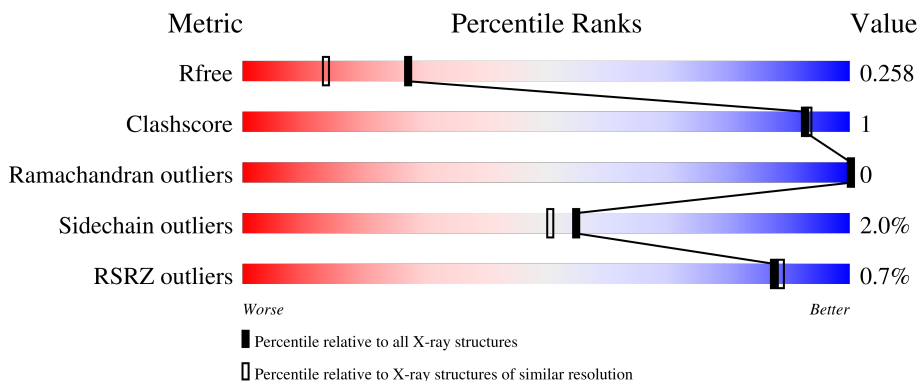
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	AAA	304	

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4952 atoms, of which 2363 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 Haloalkane dehalogenase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	AAA	293	4682	1537	2318	397	420	10	49	2	0

There are 34 discrepancies between the modelled and reference sequences:

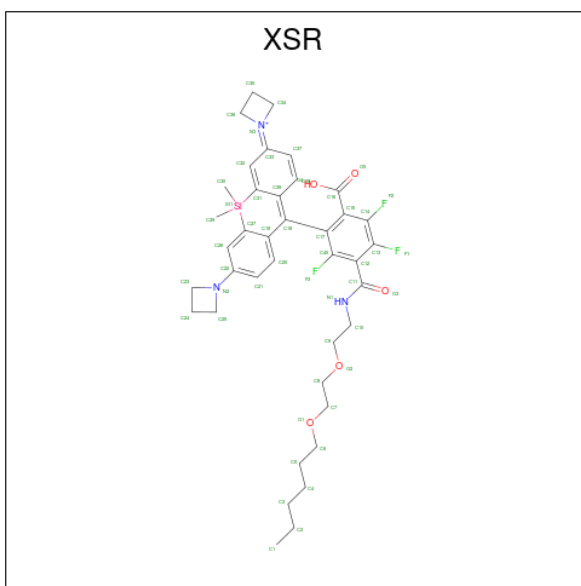
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	-6	MET	-	initiating methionine	UNP P0A3G3
AAA	-5	HIS	-	expression tag	UNP P0A3G3
AAA	-4	HIS	-	expression tag	UNP P0A3G3
AAA	-3	HIS	-	expression tag	UNP P0A3G3
AAA	-2	HIS	-	expression tag	UNP P0A3G3
AAA	-1	HIS	-	expression tag	UNP P0A3G3
AAA	0	HIS	-	expression tag	UNP P0A3G3
AAA	1	GLY	-	expression tag	UNP P0A3G3
AAA	2	ALA	-	expression tag	UNP P0A3G3
AAA	47	VAL	LEU	engineered mutation	UNP P0A3G3
AAA	58	THR	SER	engineered mutation	UNP P0A3G3
AAA	78	GLY	ASP	engineered mutation	UNP P0A3G3
AAA	87	PHE	TYR	engineered mutation	UNP P0A3G3
AAA	88	MET	LEU	engineered mutation	UNP P0A3G3
AAA	128	PHE	CYS	engineered mutation	UNP P0A3G3
AAA	155	THR	ALA	engineered mutation	UNP P0A3G3
AAA	160	LYS	GLU	engineered mutation	UNP P0A3G3
AAA	167	VAL	ALA	engineered mutation	UNP P0A3G3
AAA	172	THR	ALA	engineered mutation	UNP P0A3G3
AAA	175	MET	LYS	engineered mutation	UNP P0A3G3
AAA	176	GLY	CYS	engineered mutation	UNP P0A3G3
AAA	195	ASN	LYS	engineered mutation	UNP P0A3G3
AAA	224	GLU	ALA	engineered mutation	UNP P0A3G3
AAA	227	ASP	ASN	engineered mutation	UNP P0A3G3
AAA	257	LYS	GLU	engineered mutation	UNP P0A3G3
AAA	264	ALA	THR	engineered mutation	UNP P0A3G3
AAA	272	ASN	HIS	engineered mutation	UNP P0A3G3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	273	LEU	TYR	engineered mutation	UNP P0A3G3
AAA	291	SER	PRO	engineered mutation	UNP P0A3G3
AAA	292	THR	ALA	engineered mutation	UNP P0A3G3
AAA	294	GLU	-	insertion	UNP P0A3G3
AAA	295	ILE	-	insertion	UNP P0A3G3
AAA	296	SER	-	insertion	UNP P0A3G3
AAA	297	GLY	-	insertion	UNP P0A3G3

- Molecule 2 is 1-[(4aM,10P)-7-(azetidin-1-yl)-10-[2-carboxy-3,4,6-trifluoro-5-({2-[2-(hexyloxy)ethoxy]ethyl}carbamoyl)phenyl]-5,5-dimethyldibenzo[b,e]silin-3(5H)-ylidene]azetidin-1-ium (three-letter code: XSR) (formula: C₃₉H₄₇F₃N₃O₅Si) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	
			Total	C	F	H	N	O			Si
2	AAA	1	96	39	3	45	3	5	1	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
3	AAA	1	1	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	AAA	173	Total 173	O 173	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: Haloalkane dehalogenase

Chain AAA:  %



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	62.74Å 62.74Å 163.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	58.65 – 1.90 58.58 – 1.90	Depositor EDS
% Data completeness (in resolution range)	98.9 (58.65-1.90) 98.9 (58.58-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.17 (at 1.90Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.194 , 0.246 0.206 , 0.258	Depositor DCC
R_{free} test set	1325 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	22.6	Xtrriage
Anisotropy	0.019	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.43 , 38.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.52$, $\langle L^2 \rangle = 0.37$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4952	wwPDB-VP
Average B, all atoms (Å ²)	24.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.75% 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: CL, XSR

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	AAA	0.71	0/2446	0.83	0/3344

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	AAA	2364	2318	2307	7	0
2	AAA	51	45	0	0	0
3	AAA	1	0	0	0	0
4	AAA	173	0	0	0	0
All	All	2589	2363	2307	7	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 1.

The worst 5 of 7 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:AAA:131:PHE:HZ	1:AAA:255:LEU:HD13	1.79	0.47
1:AAA:152:PHE:CG	1:AAA:162:ILE:HD11	2.50	0.46
1:AAA:30:ARG:NE	1:AAA:30:ARG:HA	2.33	0.44
1:AAA:133:ARG:HA	1:AAA:226[B]:MET:HE2	2.01	0.42
1:AAA:29:PRO:O	1:AAA:60:ARG:HB2	2.19	0.42

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	AAA	293/304 (96%)	283 (97%)	10 (3%)	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	AAA	255/261 (98%)	250 (98%)	5 (2%)	55 51

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

Mol	Chain	Res	Type
1	AAA	42	PRO
1	AAA	86	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	AAA	168	PHE
1	AAA	225	TYR
1	AAA	267	ILE

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 2 ligands modelled in this entry, 1 is 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
2	XSR	AAA	301	1	51,56,56	1.17	3 (5%)	65,81,81	1.43	10 (15%)

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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	XSR	AAA	301	1	-	5/29/78/78	0/5/6/6

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	AAA	301	XSR	C33-N3	3.43	1.42	1.35
2	AAA	301	XSR	C32-C31	2.81	1.40	1.35
2	AAA	301	XSR	SI1-C27	-2.48	1.82	1.87

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	AAA	301	XSR	C34-N3-C36	4.41	100.50	94.20
2	AAA	301	XSR	C19-C18-C39	3.88	127.76	119.99
2	AAA	301	XSR	C38-C39-C31	3.49	119.00	115.66
2	AAA	301	XSR	C35-C36-N3	-3.49	83.83	88.05
2	AAA	301	XSR	C40-C17-C18	-3.41	118.07	121.90

There are no chirality outliers.

All (5) torsion outliers are listed below:

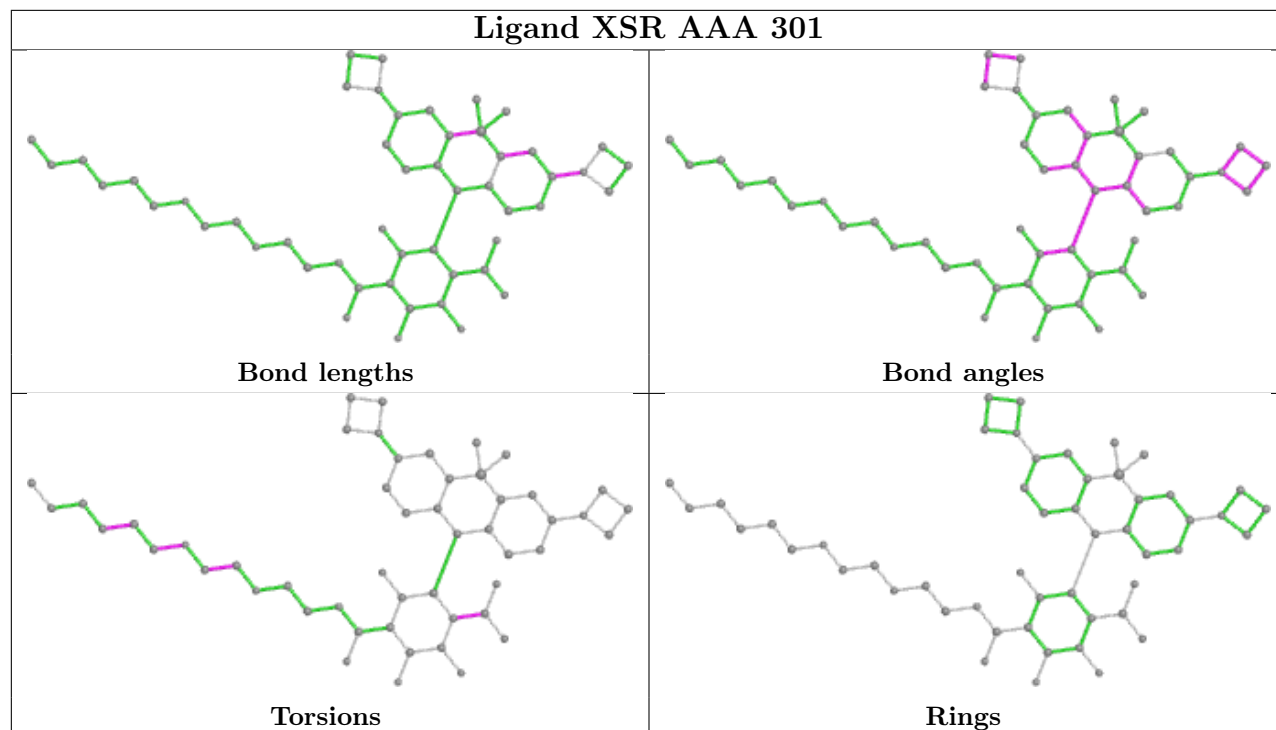
Mol	Chain	Res	Type	Atoms
2	AAA	301	XSR	O1-C7-C8-O2
2	AAA	301	XSR	C14-C15-C16-O4
2	AAA	301	XSR	C14-C15-C16-O5
2	AAA	301	XSR	C3-C4-C5-C6
2	AAA	301	XSR	C5-C6-O1-C7

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	AAA	293/304 (96%)	-0.06	2 (0%) 87 88	14, 22, 36, 51	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	155	THR	5.2
1	AAA	154	THR	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.

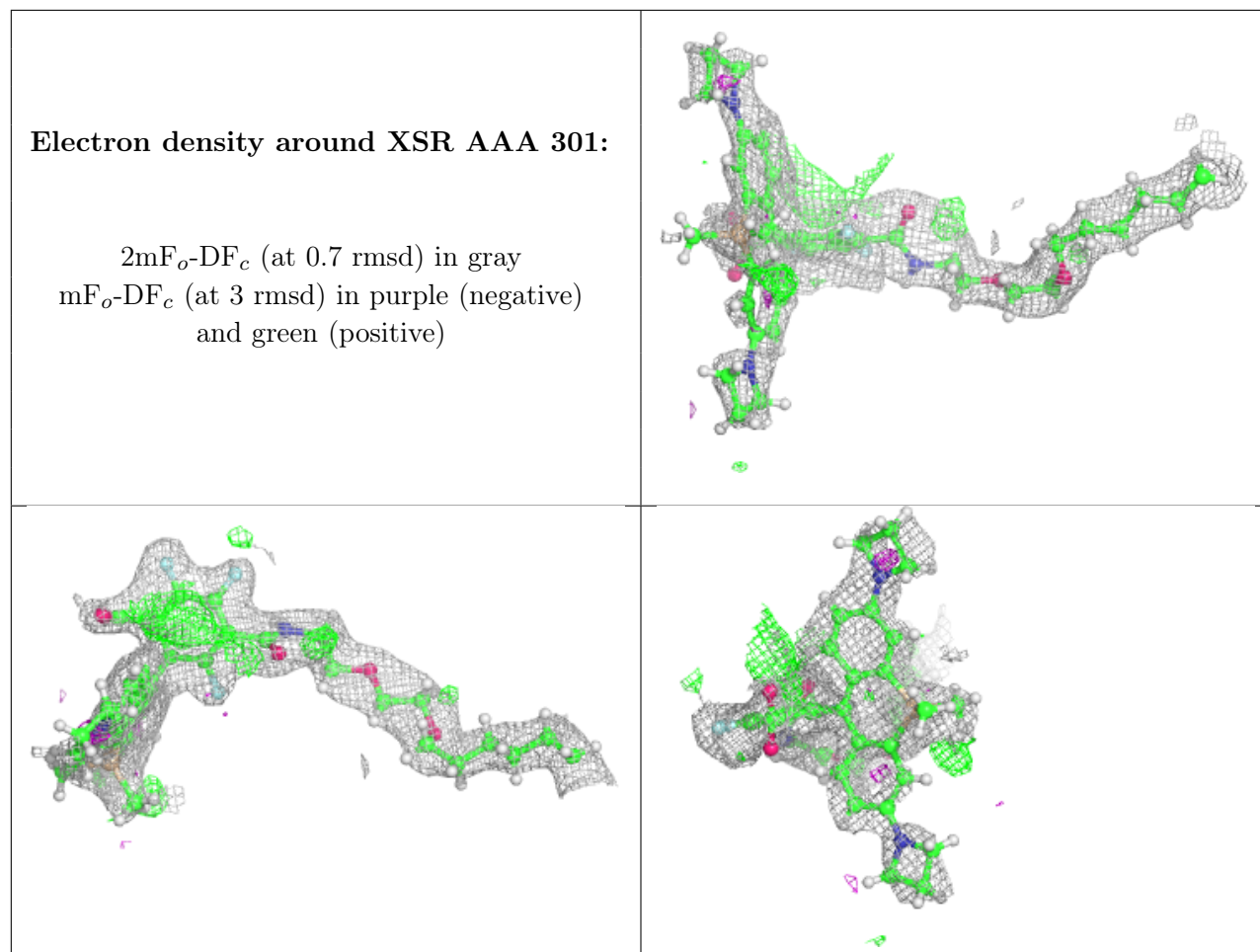
6.4 Ligands [i](#)

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(Å ²)	Q<0.9
2	XSR	AAA	301	51/51	0.83	0.19	24,45,61,64	0
3	CL	AAA	302	1/1	0.99	0.10	19,19,19,19	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.