



Full wwPDB X-ray Structure Validation Report ⓘ

Dec 10, 2023 – 12:52 am GMT

PDB ID : 8QF8
Title : GH146 beta-L-arabinofuranosidase from *Bacteroides thetaioatomicron* in complex with beta-l-arabinofurano cyclophellitol aziridine
Authors : Borlandelli, V.; Offen, W.; Moroz, O.V.; Nin-Hill, A.; McGregor, N.; Binkhorst, L.; Armstrong, Z.; Ishiwata, A.; Artola, M.; Rovira, C.; Davies, G.J.; Overkleeft, H.
Deposited on : 2023-09-04
Resolution : 2.40 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 ⓘ 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](#) ①) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtrriage (Phenix) : 1.13
EDS : 2.36
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

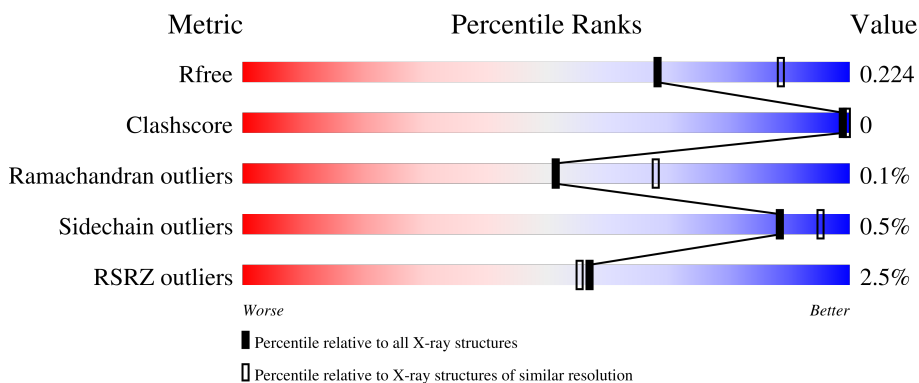
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	802	 3% (Poor fit), 96% (Quality), 2% (Not modelled)
1	B	802	 2% (Poor fit), 94% (Quality), 4% (Not modelled)

2 Entry composition [i](#)

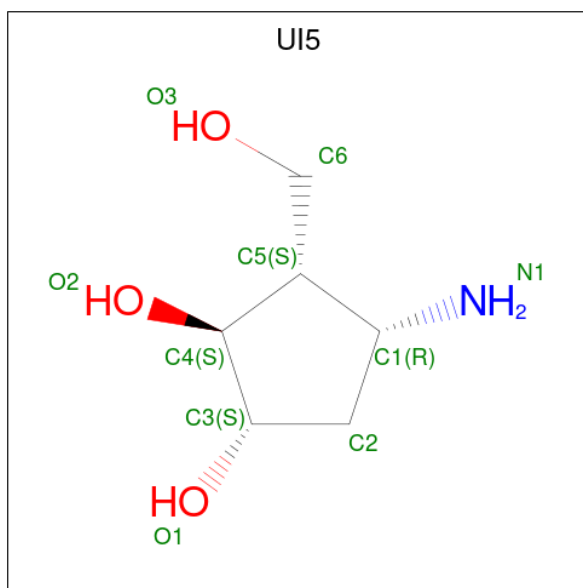
There are 5 unique types of molecules in this entry. The entry contains 24508 atoms, of which 11753 are hydrogens and 0 are deuteriums.

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 Glycosyl hydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	776	Total	C	H	N	O	S	0	0	0
			12163	3960	5918	1091	1169	25			
1	B	771	Total	C	H	N	O	S	0	0	0
			11989	3915	5813	1079	1157	25			

- Molecule 2 is (1 {S},2 {S},3 {S},4 {R})-4-azanyl-3-(hydroxymethyl)cyclopentane-1,2-diol (three-letter code: UI5) (formula: C₆H₁₃NO₃) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	N	O		
2	A	1	Total	C	H	N	O	0	0
			22	6	12	1	3		

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

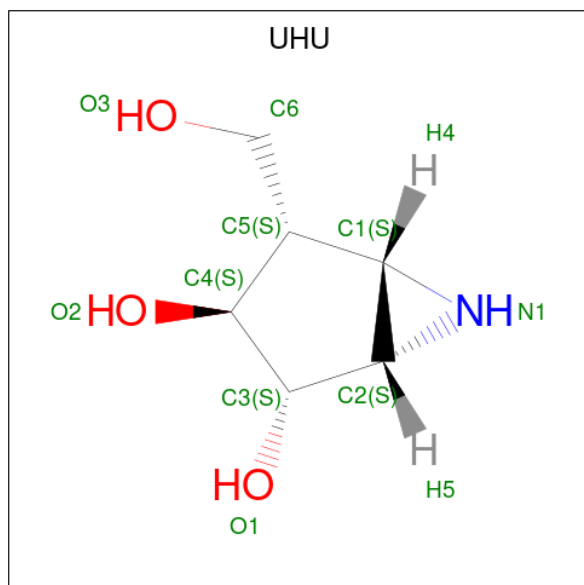
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Zn	0	0
			1	1		

- Molecule 4 is (1 {S},2 {S},3 {S},4 {S},5 {S})-4-(hydroxymethyl)-6-azabicyclo[3.1.0]hexane-2,3-diol (three-letter code: UHU) (formula: C₆H₁₁NO₃).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	B	1	Total	C	H	N	O	0	0
			20	6	10	1	3		

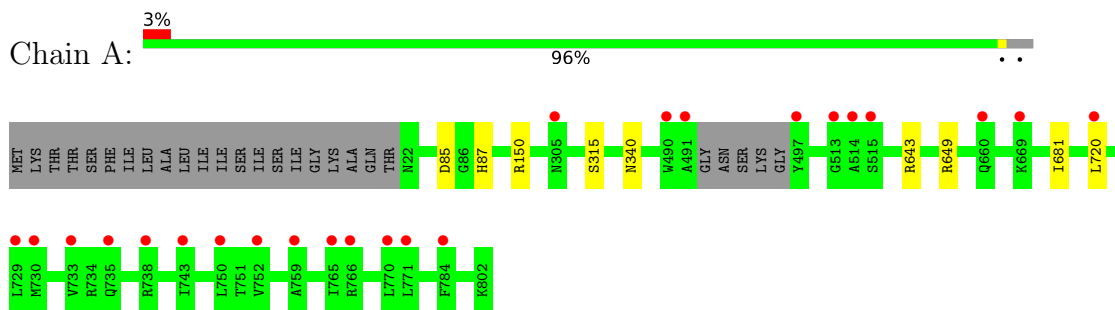
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	164	Total	O	0	0
			164	164		
5	B	148	Total	O	0	0
			148	148		

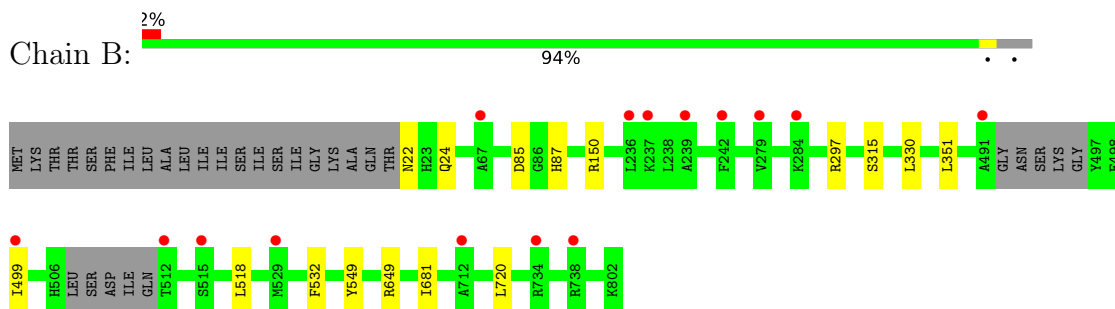
3 Residue-property plots [i](#)

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: Glycosyl hydrolase



- Molecule 1: Glycosyl hydrolase



4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	93.84Å 98.55Å 196.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	55.90 – 2.40 55.90 – 2.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (55.90-2.40) 100.0 (55.90-2.40)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.29 (at 2.40Å)	Xtrriage
Refinement program	REFMAC 5.8.0419, REFMAC 5.8.0419	Depositor
R, R_{free}	0.190 , 0.221 0.195 , 0.224	Depositor DCC
R_{free} test set	3622 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	48.4	Xtrriage
Anisotropy	0.542	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 31.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.012 for k,h,-l	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	24508	wwPDB-VP
Average B, all atoms (Å ²)	64.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.20% 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: UI5, ZN, UHU

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	0.34	0/6401	0.64	0/8671
1	B	0.33	0/6331	0.64	0/8583
All	All	0.34	0/12732	0.64	0/17254

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2
1	B	0	1
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	150	ARG	Sidechain
1	A	643	ARG	Sidechain
1	B	150	ARG	Sidechain

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	6245	5918	6087	1	0
1	B	6176	5813	5972	8	0
2	A	10	12	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	B	10	10	0	0	0
5	A	164	0	0	0	0
5	B	148	0	0	0	0
All	All	12755	11753	12059	9	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 0.

All (9) 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:B:22:ASN:OD1	1:B:24:GLN:HG2	1.39	1.19
1:B:22:ASN:OD1	1:B:24:GLN:CG	2.02	1.07
1:B:22:ASN:OD1	1:B:24:GLN:HG3	1.92	0.67
1:B:330:LEU:HD13	1:B:549:TYR:CD1	2.43	0.54
1:B:499:ILE:HD13	1:B:518:LEU:HD21	1.93	0.50
1:B:499:ILE:HG22	1:B:532:PHE:HB3	1.94	0.49
1:A:681:ILE:HD11	1:A:720:LEU:HD23	2.01	0.43
1:B:297:ARG:NH1	1:B:351:LEU:O	2.52	0.41
1:B:681:ILE:HD11	1:B:720:LEU:HD23	2.02	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	772/802 (96%)	754 (98%)	17 (2%)	1 (0%)	51 68
1	B	765/802 (95%)	744 (97%)	20 (3%)	1 (0%)	51 68
All	All	1537/1604 (96%)	1498 (98%)	37 (2%)	2 (0%)	51 68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	315	SER
1	B	315	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.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	673/697 (97%)	669 (99%)	4 (1%)	86 94
1	B	660/697 (95%)	657 (100%)	3 (0%)	88 95
All	All	1333/1394 (96%)	1326 (100%)	7 (0%)	88 95

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	85	ASP
1	A	87	HIS
1	A	340	ASN
1	A	649	ARG
1	B	85	ASP
1	B	87	HIS
1	B	649	ARG

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	360	ASN
1	A	728	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	407	GLN
1	B	728	GLN

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 4 ligands modelled in this entry, 2 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
4	UHU	B	901	-	10,11,11	1.72	2 (20%)	10,17,17	1.57	2 (20%)
2	UI5	A	901	1	10,10,10	0.65	0	8,14,14	1.82	2 (25%)

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	UHU	B	901	-	-	0/2/23/23	0/2/2/2
2	UI5	A	901	1	-	0/2/18/18	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	901	UHU	C1-N1	4.32	1.51	1.48
4	B	901	UHU	C2-N1	2.73	1.49	1.48

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	901	UI5	C2-C1-C5	3.91	107.02	103.05
4	B	901	UHU	O1-C3-C2	-3.23	103.25	111.16
4	B	901	UHU	O1-C3-C4	2.53	119.99	111.82
2	A	901	UI5	O1-C3-C4	-2.40	106.72	111.27

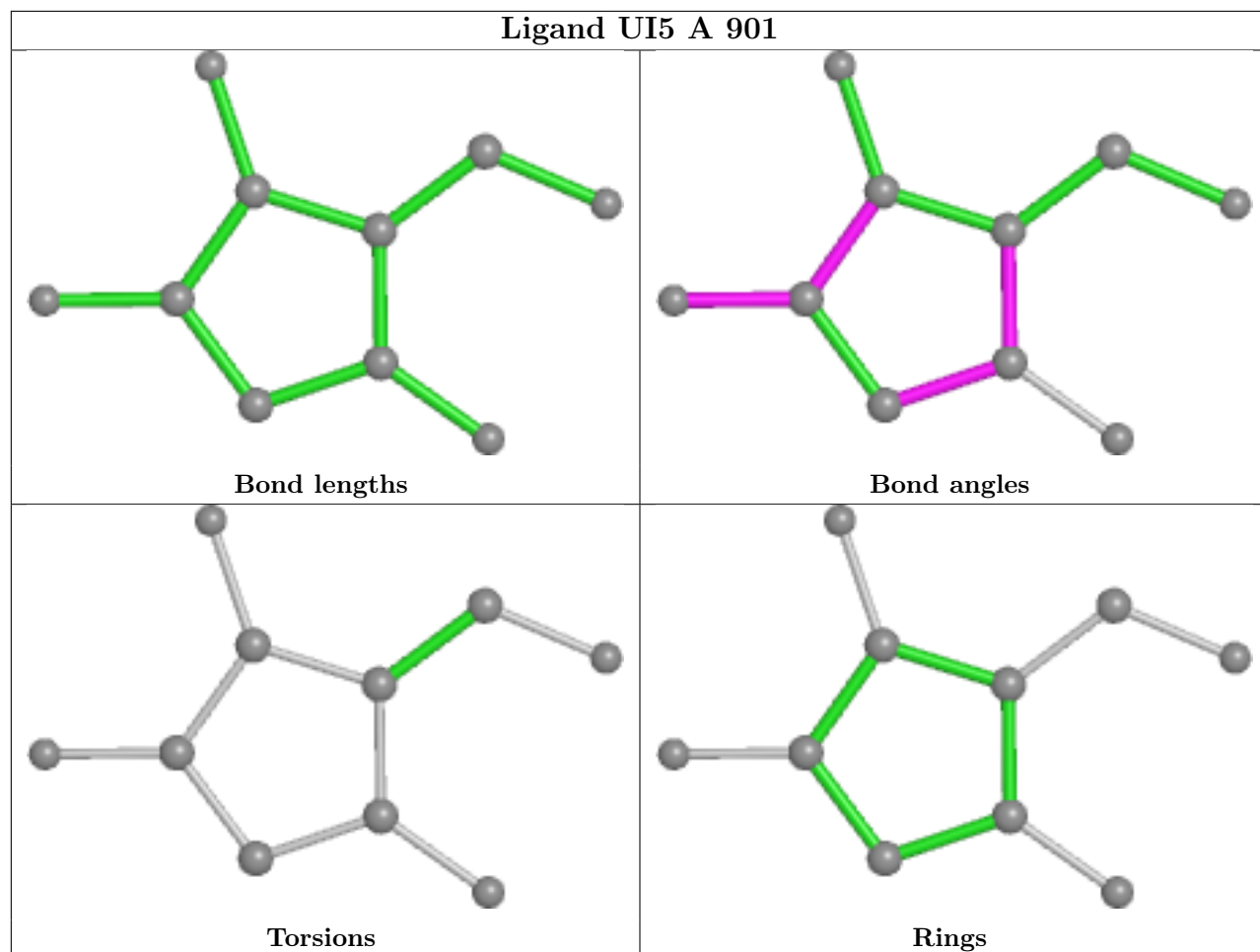
There are no chirality outliers.

There are no torsion outliers.

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

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, 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	776/802 (96%)	0.22	24 (3%)	49 47	34, 55, 109, 159	0
1	B	771/802 (96%)	0.14	15 (1%)	66 64	40, 63, 103, 141	0
All	All	1547/1604 (96%)	0.18	39 (2%)	57 55	34, 59, 106, 159	0

All (39) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	512	THR	5.2
1	A	752	VAL	5.0
1	A	735	GLN	4.3
1	B	67	ALA	3.9
1	A	491	ALA	3.7
1	A	729	LEU	3.7
1	A	784	PHE	3.5
1	B	242	PHE	3.2
1	A	733	VAL	3.1
1	B	515	SER	3.1
1	B	284	LYS	3.0
1	A	730	MET	2.9
1	A	514	ALA	2.8
1	B	491	ALA	2.7
1	A	513	GLY	2.7
1	A	738	ARG	2.7
1	B	279	VAL	2.6
1	A	669	LYS	2.6
1	A	497	TYR	2.6
1	A	766	ARG	2.6
1	A	750	LEU	2.5
1	B	529	MET	2.5
1	B	738	ARG	2.5
1	A	765	ILE	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	515	SER	2.3
1	B	734	ARG	2.2
1	A	743	ILE	2.2
1	B	499	ILE	2.2
1	A	720	LEU	2.2
1	A	490	TRP	2.2
1	A	770	LEU	2.2
1	B	236	LEU	2.2
1	B	237	LYS	2.1
1	A	660	GLN	2.1
1	B	712	ALA	2.1
1	A	771	LEU	2.1
1	A	759	ALA	2.1
1	B	239	ALA	2.1
1	A	305	ASN	2.1

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](#)

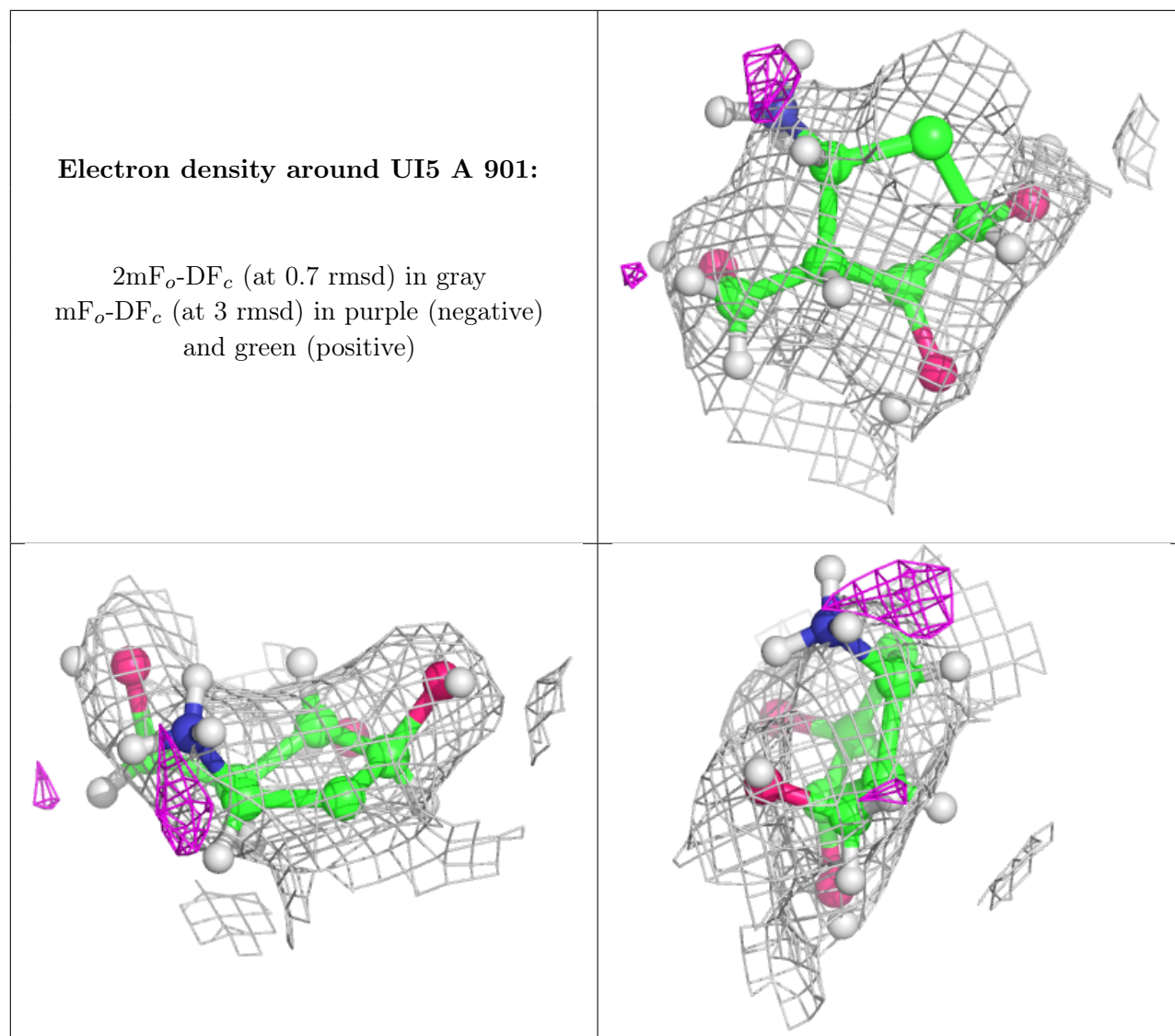
There are no monosaccharides in this entry.

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
4	UHU	B	901	10/10	0.95	0.27	47,49,52,54	0
2	UI5	A	901	10/10	0.96	0.21	47,48,54,56	0
3	ZN	B	902	1/1	0.99	0.12	47,47,47,47	0
3	ZN	A	902	1/1	0.99	0.13	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.



6.5 Other polymers [i](#)

There are no such residues in this entry.