



wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 20, 2022 – 07:09 pm BST

PDB ID : 7AY2
Title : Crystal structure of truncated USP1-UAF1 reacted with ubiquitin-prg
Authors : Arkinson, C.; Rennie, M.L.; Walden, H.
Deposited on : 2020-11-10
Resolution : 3.20 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:

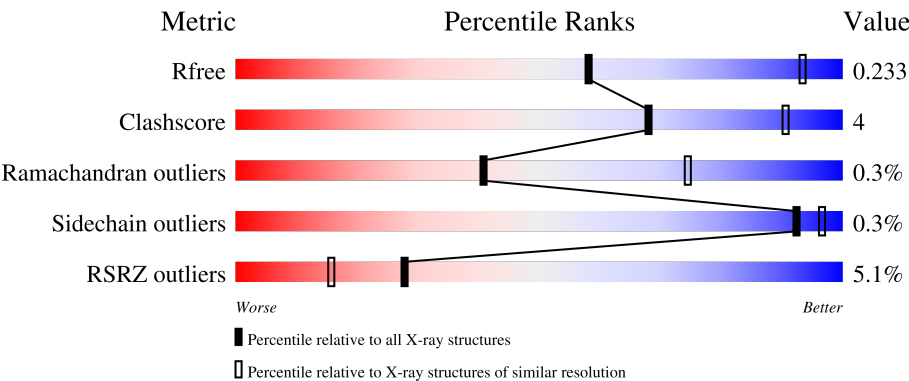
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.29
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	563	<div><div></div><div></div><div></div><div></div><div></div></div> <div>%82%9%9%</div>
1	D	563	<div><div></div><div></div><div></div><div></div><div></div></div> <div>%81%9%10%</div>
2	B	397	<div><div></div><div></div><div></div><div></div><div></div></div> <div>%76%8%15%</div>
2	E	397	<div><div></div><div></div><div></div><div></div><div></div></div> <div>16%44%7%49%</div>
3	C	76	<div><div></div><div></div><div></div><div></div><div></div></div> <div>92%7%</div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	76	<div><div></div><div>8%</div><div>66%</div><div>11%</div><div>24%</div></div>

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 13451 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 WD repeat-containing protein 48.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	514	Total	C	N	O	S	0	0	0
			4038	2546	710	759	23			
1	D	508	Total	C	N	O	S	0	0	0
			3990	2516	700	751	23			

- Molecule 2 is a protein called Ubiquitin carboxyl-terminal hydrolase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	337	Total	C	N	O	S	0	0	0
			2711	1737	441	519	14			
2	E	203	Total	C	N	O	S	0	0	0
			1649	1065	267	306	11			

- Molecule 3 is a protein called Polyubiquitin-B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	75	Total	C	N	O	S	0	0	0
			597	376	104	116	1			
3	F	58	Total	C	N	O	S	0	0	0
			460	293	74	92	1			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	76	AYE	GLY	engineered mutation	UNP P0CG47
F	76	AYE	GLY	engineered mutation	UNP P0CG47

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

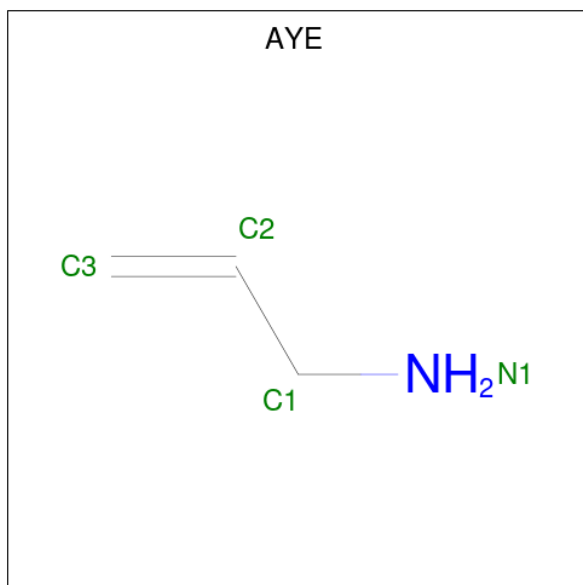
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Zn	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total	Zn	0	0
			1	1		

- Molecule 5 is prop-2-en-1-amine (three-letter code: AYE) (formula: C₃H₇N) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	N	0	0
			4	3	1		

- Molecule 1: WD repeat-containing protein 48



4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α , β , γ	134.24Å 134.24Å 274.67Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	67.12 – 3.20 88.73 – 3.20	Depositor EDS
% Data completeness (in resolution range)	100.0 (67.12-3.20) 100.0 (88.73-3.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.76 (at 3.19Å)	Xtriage
Refinement program	PHENIX 1.18	Depositor
R, R_{free}	0.204 , 0.234 0.204 , 0.233	Depositor DCC
R_{free} test set	2353 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å ²)	86.7	Xtriage
Anisotropy	0.248	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.055 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	13451	wwPDB-VP
Average B, all atoms (Å ²)	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.36% 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

5.1 Standard geometry

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

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	0.29	0/4118	0.60	1/5589 (0.0%)
1	D	0.29	0/4069	0.62	0/5523
2	B	0.28	0/2759	0.50	0/3721
2	E	0.27	0/1677	0.47	0/2258
3	C	0.25	0/603	0.45	0/811
3	F	0.24	0/463	0.42	0/621
All	All	0.28	0/13689	0.56	1/18523 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	528	LEU	CB-CG-CD1	5.01	119.52	111.00

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	4038	0	4038	33	0
1	D	3990	0	3987	33	0
2	B	2711	0	2709	22	0
2	E	1649	0	1656	18	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	597	0	626	5	0
3	F	460	0	475	5	0
4	B	1	0	0	0	0
4	E	1	0	0	0	0
5	B	4	0	4	0	0
All	All	13451	0	13495	101	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 4.

The worst 5 of 101 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:D:191:VAL:HG11	1:D:205:LYS:HG2	1.66	0.77
2:B:152:LEU:HD13	2:B:178:PRO:HB3	1.66	0.76
1:A:250:VAL:O	1:A:287:ARG:NH1	2.23	0.70
2:E:109:VAL:O	2:E:150:GLN:NE2	2.25	0.70
1:A:114:ARG:NH1	1:D:155:SER:O	2.26	0.68

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	506/563 (90%)	487 (96%)	18 (4%)	1 (0%)	47	79
1	D	500/563 (89%)	485 (97%)	15 (3%)	0	100	100
2	B	323/397 (81%)	301 (93%)	19 (6%)	3 (1%)	17	56
2	E	187/397 (47%)	176 (94%)	10 (5%)	1 (0%)	29	67
3	C	73/76 (96%)	73 (100%)	0	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	F	50/76 (66%)	49 (98%)	1 (2%)	0	100	100
All	All	1639/2072 (79%)	1571 (96%)	63 (4%)	5 (0%)	41	74

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	739	SER
1	A	157	ASN
2	E	155	SER
2	B	740	LEU
2	B	178	PRO

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	450/492 (92%)	449 (100%)	1 (0%)	93	98
1	D	444/492 (90%)	443 (100%)	1 (0%)	93	98
2	B	310/364 (85%)	309 (100%)	1 (0%)	92	96
2	E	190/364 (52%)	189 (100%)	1 (0%)	88	95
3	C	68/68 (100%)	68 (100%)	0	100	100
3	F	54/68 (79%)	54 (100%)	0	100	100
All	All	1516/1848 (82%)	1512 (100%)	4 (0%)	92	96

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	221	ASP
2	B	201	GLN
1	D	73	HIS
2	E	574	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are

no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 2 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	$\# Z > 2$
5	AYE	B	802	3,2	3,3,3	0.71	0	1,2,2	2.07	1 (100%)

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
5	AYE	B	802	3,2	-	0/1/1/1	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	802	AYE	C1-C2-C3	-2.07	116.44	125.74

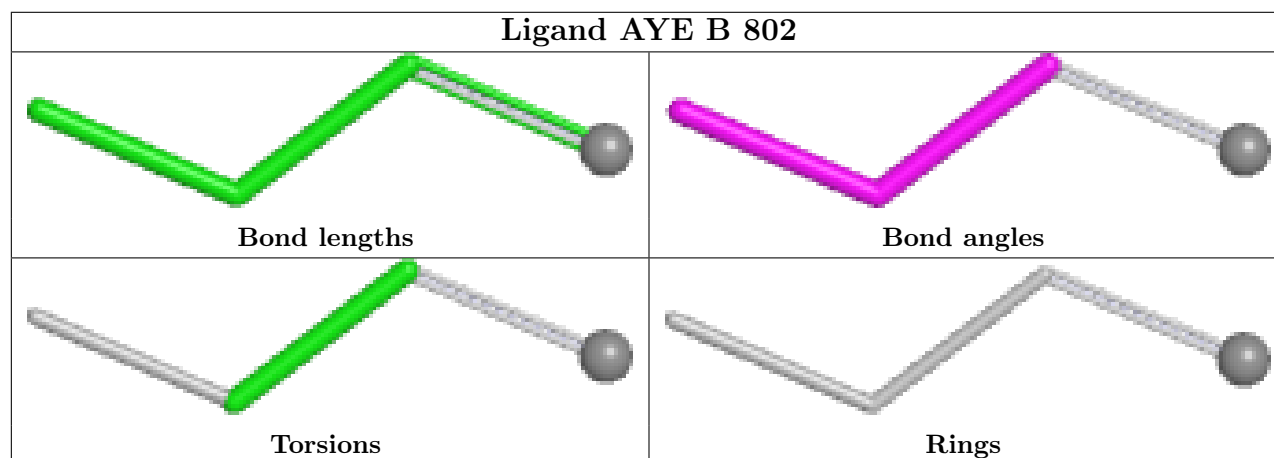
There are no chirality outliers.

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 [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	514/563 (91%)	0.20	6 (1%) 79 67	58, 84, 133, 175	0
1	D	508/563 (90%)	0.30	6 (1%) 79 67	54, 88, 127, 165	0
2	B	337/397 (84%)	0.15	4 (1%) 79 67	56, 90, 145, 175	0
2	E	203/397 (51%)	1.41	64 (31%) 0 0	68, 143, 189, 211	0
3	C	75/76 (98%)	0.09	0 100 100	53, 79, 104, 126	0
3	F	58/76 (76%)	0.76	6 (10%) 6 4	109, 138, 160, 177	0
All	All	1695/2072 (81%)	0.38	86 (5%) 28 16	53, 91, 157, 211	0

The worst 5 of 86 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	110	LYS	6.9
2	E	153	ILE	5.8
2	E	780	LEU	5.5
2	E	109	VAL	5.5
2	E	112	LEU	5.0

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](#)

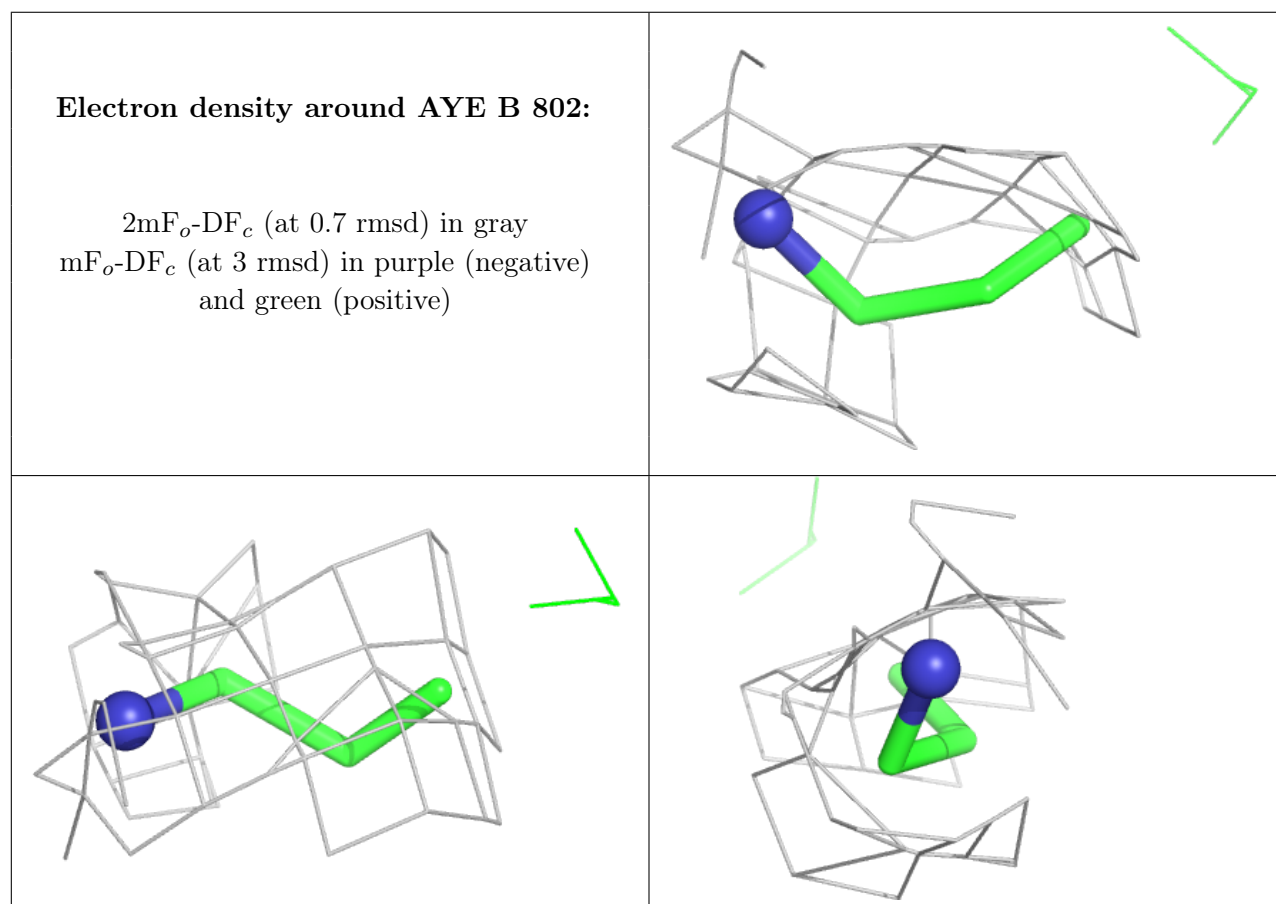
There are no monosaccharides in this entry.

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
5	AYE	B	802	4/4	0.98	0.21	92,93,96,98	0
4	ZN	E	1000	1/1	1.00	0.21	82,82,82,82	0
4	ZN	B	801	1/1	1.00	0.24	70,70,70,70	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.