



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 14, 2026 – 02:01 PM UTC

PDB ID : 9U7Z / pdb_00009u7z
Title : 5hmC specific restriction endonuclease Escherichia coli APEC O1 PD-T4-3 in complex with 5-Hydroxymethyl-dCTP
Authors : Yu, Y.; Liu, R.
Deposited on : 2025-03-25
Resolution : 2.15 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 ⓘ](#)) 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

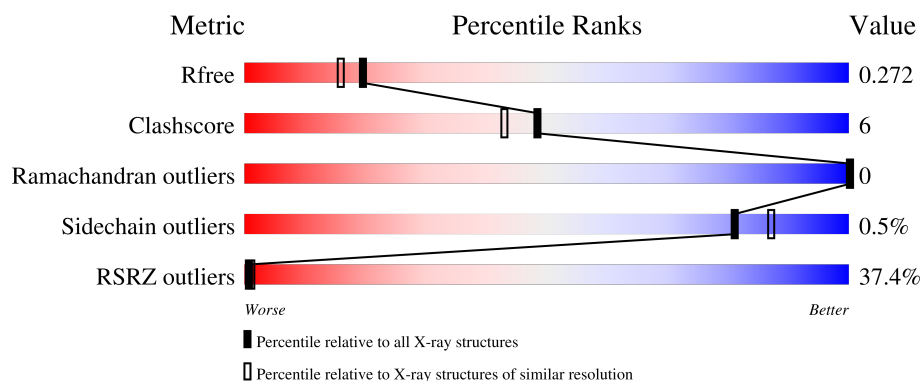
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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	2057 (2.16-2.16)
Clashscore	190562	2159 (2.16-2.16)
Ramachandran outliers	187476	2134 (2.16-2.16)
Sidechain outliers	187428	2133 (2.16-2.16)
RSRZ outliers	180081	2059 (2.16-2.16)

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	262	35% 83% 11% 6%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4126 atoms, of which 1968 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.

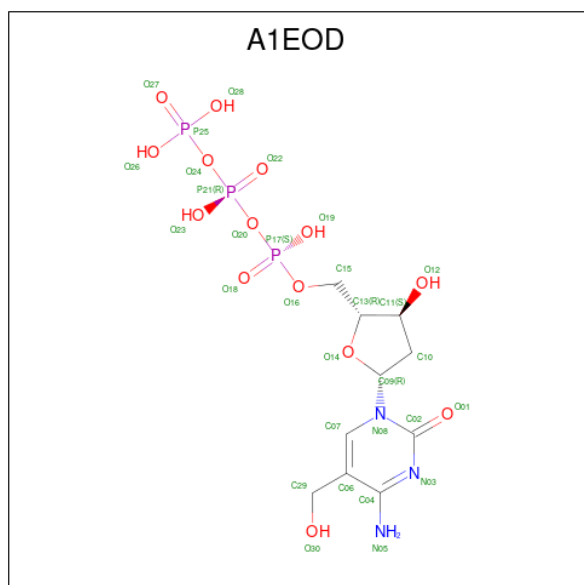
- Molecule 1 is a protein called GIY-YIG domain-containing protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	246	3950	1265	1954	359	365	7	0	2	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	SER	-	expression tag	UNP A0A0H2Z2I3
A	-3	ALA	-	expression tag	UNP A0A0H2Z2I3
A	-2	SER	-	expression tag	UNP A0A0H2Z2I3
A	-1	GLY	-	expression tag	UNP A0A0H2Z2I3
A	0	SER	-	expression tag	UNP A0A0H2Z2I3

- Molecule 2 is [(2R,3S,5R)-5-[4-azanyl-5-(hydroxymethyl)-2-oxidanylidene-pyrimidin-1-yl]-3-oxidanyl-oxolan-2-yl]methoxy-oxidanyl-phosphoryl] phosphono hydrogen phosphate (CCD ID: A1EOD) (formula: C₁₀H₁₈N₃O₁₄P₃).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	P	12	0
			44	10	14	3	14	3		

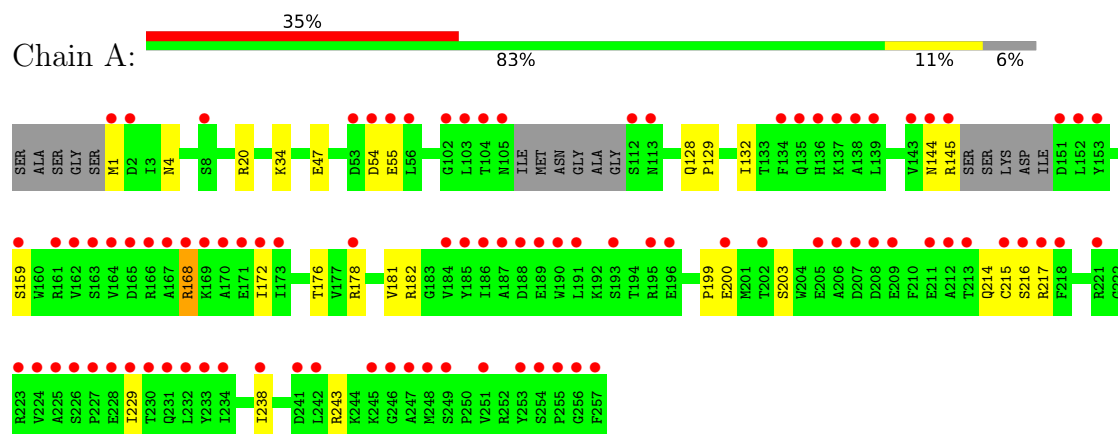
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	132	Total	O	0	0
			132	132		

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: GIY-YIG domain-containing protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 64 2 2	Depositor
Cell constants a, b, c, α , β , γ	124.73Å 124.73Å 80.80Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.37 – 2.15 49.37 – 2.15	Depositor EDS
% Data completeness (in resolution range)	96.5 (49.37-2.15) 96.4 (49.37-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.61 (at 2.16Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.230 , 0.273 0.230 , 0.272	Depositor DCC
R_{free} test set	1037 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	23.8	Xtriage
Anisotropy	0.105	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.41 , 53.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	4126	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 3.95% 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: A1EOD

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	0.36	0/2041	0.53	0/2753

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	1996	1954	1974	21	2
2	A	30	14	0	1	0
3	A	132	0	0	7	0
All	All	2158	1968	1974	22	2

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 6.

All (22) 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:A:301:A1EOD:O14	2:A:301:A1EOD:C13	1.69	1.24

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:182:ARG:NH1	3:A:407:HOH:O	2.24	0.70
1:A:1:MET:HE2	1:A:4:ASN:OD1	1.92	0.69
1:A:144:ASN:O	1:A:145:ARG:HB2	1.93	0.69
1:A:132:ILE:O	3:A:401:HOH:O	2.12	0.68
1:A:20:ARG:NH2	3:A:408:HOH:O	2.26	0.68
1:A:159:SER:OG	1:A:214:GLN:OE1	2.04	0.68
1:A:217:ARG:NE	3:A:411:HOH:O	2.33	0.61
1:A:172:ILE:HG21	1:A:229:ILE:HG21	1.85	0.58
1:A:215:CYS:O	1:A:216:SER:OG	2.16	0.57
1:A:203:SER:OG	3:A:402:HOH:O	2.18	0.56
1:A:55:GLU:HG3	3:A:420:HOH:O	2.09	0.53
1:A:176:THR:HG21	1:A:238:ILE:HD11	1.93	0.50
1:A:172:ILE:HG21	1:A:229:ILE:CG2	2.42	0.49
1:A:144:ASN:O	1:A:145:ARG:CB	2.60	0.48
1:A:178:ARG:O	1:A:243:ARG:NH2	2.47	0.47
1:A:128:GLN:N	1:A:129:PRO:CD	2.79	0.46
1:A:34:LYS:NZ	1:A:54:ASP:O	2.45	0.46
1:A:243:ARG:NH1	3:A:414:HOH:O	2.43	0.43
1:A:132:ILE:HD11	1:A:181:VAL:CG2	2.48	0.43
1:A:199:PRO:HD2	1:A:200:GLU:OE2	2.19	0.42
1:A:215:CYS:C	1:A:216:SER:HG	2.19	0.41

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:47:GLU:OE1	1:A:168[A]:ARG:HH22[12_554]	1.48	0.12
1:A:47:GLU:OE1	1:A:168[A]:ARG:NH2[12_554]	2.08	0.12

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	242/262 (92%)	239 (99%)	3 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	208/217 (96%)	206 (99%)	2 (1%)	68	75

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	168[A]	ARG
1	A	168[B]	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	58	ASN
1	A	81	HIS

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 oligosaccharides in this entry.

5.6 Ligand geometry

1 ligand is modelled in this entry.

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	A1EOD	A	301	-	29,31,31	4.17	15 (51%)	40,48,48	0.81	0

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
2	A1EOD	A	301	-	-	6/24/36/36	0/2/2/2

All (15) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	301	A1EOD	O14-C13	10.95	1.69	1.45
2	A	301	A1EOD	C11-C13	-8.44	1.30	1.53
2	A	301	A1EOD	C04-N03	7.09	1.45	1.34
2	A	301	A1EOD	C02-N03	6.26	1.48	1.36
2	A	301	A1EOD	O14-C09	-5.90	1.29	1.42
2	A	301	A1EOD	C07-C06	5.08	1.49	1.35
2	A	301	A1EOD	P21-O24	4.98	1.64	1.59
2	A	301	A1EOD	P17-O20	4.97	1.64	1.59
2	A	301	A1EOD	P21-O20	4.81	1.64	1.59
2	A	301	A1EOD	C04-N05	4.51	1.45	1.34
2	A	301	A1EOD	C07-N08	4.41	1.45	1.38
2	A	301	A1EOD	C02-N08	3.93	1.48	1.40
2	A	301	A1EOD	C10-C09	2.86	1.60	1.52
2	A	301	A1EOD	O12-C11	2.74	1.49	1.43
2	A	301	A1EOD	O01-C02	-2.64	1.18	1.23

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

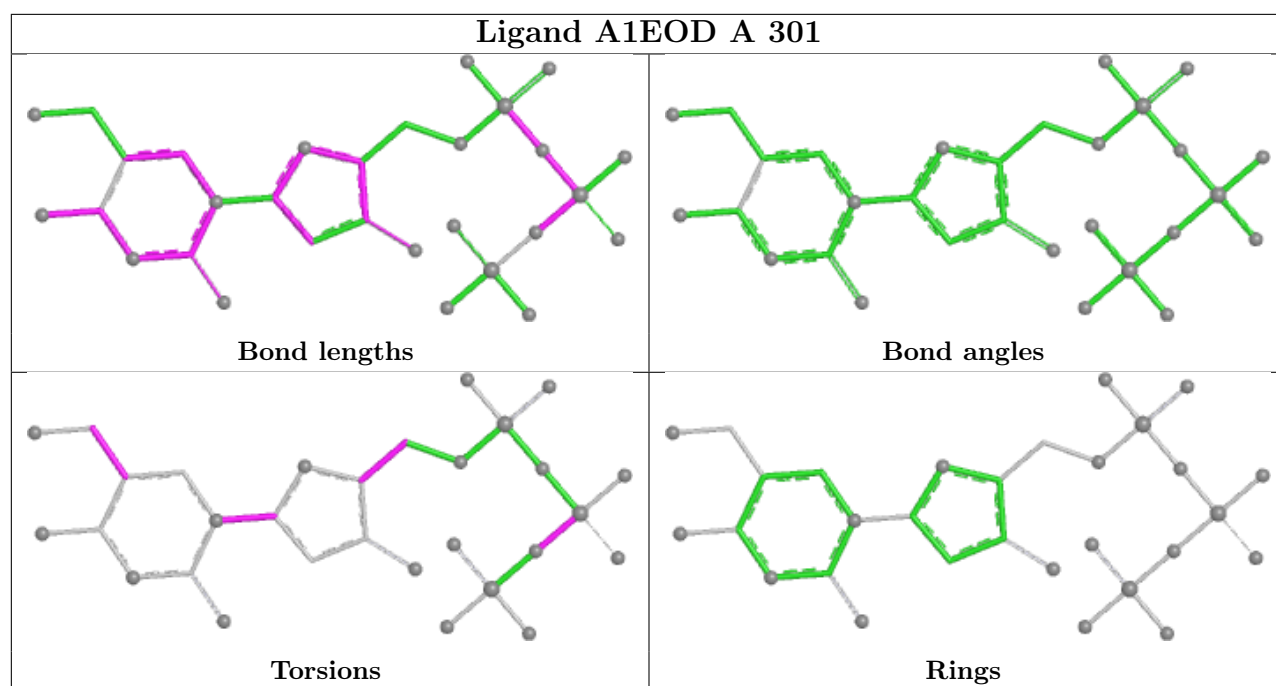
Mol	Chain	Res	Type	Atoms
2	A	301	A1EOD	O14-C09-N08-C07
2	A	301	A1EOD	O14-C09-N08-C02
2	A	301	A1EOD	C10-C09-N08-C02
2	A	301	A1EOD	C04-C06-C29-O30
2	A	301	A1EOD	O14-C13-C15-O16
2	A	301	A1EOD	P25-O24-P21-O23

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	301	A1EOD	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 ⓘ

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	246/262 (93%)	1.47	92 (37%) 1 1	9, 47, 87, 114	2 (0%)

All (92) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	190	TRP	8.1
1	A	168[A]	ARG	7.9
1	A	216	SER	6.1
1	A	104	THR	5.9
1	A	113	ASN	5.9
1	A	206	ALA	5.8
1	A	169	LYS	5.6
1	A	105	ASN	5.1
1	A	255	PRO	5.1
1	A	171	GLU	4.8
1	A	208	ASP	4.8
1	A	245	LYS	4.7
1	A	257	PHE	4.6
1	A	145	ARG	4.6
1	A	225	ALA	4.5
1	A	187	ALA	4.5
1	A	170	ALA	4.5
1	A	186	ILE	4.4
1	A	54	ASP	4.4
1	A	202	THR	4.4
1	A	1	MET	4.3
1	A	112	SER	4.3
1	A	53	ASP	4.2
1	A	207	ASP	4.1
1	A	218	PHE	4.1
1	A	229	ILE	4.0
1	A	191	LEU	3.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	247	ALA	3.9
1	A	138	ALA	3.9
1	A	184	VAL	3.8
1	A	248	MET	3.8
1	A	135	GLN	3.8
1	A	196	GLU	3.7
1	A	224	VAL	3.7
1	A	215	CYS	3.7
1	A	166	ARG	3.6
1	A	230	THR	3.5
1	A	165	ASP	3.5
1	A	212	ALA	3.5
1	A	188	ASP	3.4
1	A	185	TYR	3.4
1	A	159	SER	3.4
1	A	226	SER	3.4
1	A	172	ILE	3.3
1	A	232	LEU	3.3
1	A	189	GLU	3.3
1	A	173	ILE	3.2
1	A	151	ASP	3.2
1	A	8	SER	3.2
1	A	233	TYR	3.2
1	A	152	LEU	3.1
1	A	242	LEU	3.1
1	A	56	LEU	3.0
1	A	238	ILE	3.0
1	A	195	ARG	3.0
1	A	213	THR	3.0
1	A	137	LYS	3.0
1	A	167	ALA	3.0
1	A	228	GLU	2.9
1	A	161	ARG	2.9
1	A	205	GLU	2.8
1	A	209	GLU	2.8
1	A	223	ARG	2.8
1	A	253	TYR	2.8
1	A	227	PRO	2.8
1	A	162	VAL	2.8
1	A	134	PHE	2.7
1	A	211	GLU	2.7
1	A	102	GLY	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	2	ASP	2.6
1	A	163	SER	2.6
1	A	256	GLY	2.6
1	A	221	ARG	2.6
1	A	249	SER	2.6
1	A	234	ILE	2.5
1	A	153	TYR	2.5
1	A	231	GLN	2.4
1	A	254	SER	2.4
1	A	144	ASN	2.4
1	A	217	ARG	2.4
1	A	246	GLY	2.4
1	A	136	HIS	2.3
1	A	251	VAL	2.2
1	A	241	ASP	2.2
1	A	55	GLU	2.2
1	A	139	LEU	2.2
1	A	200	GLU	2.1
1	A	143	VAL	2.1
1	A	103	LEU	2.1
1	A	193	SER	2.1
1	A	178	ARG	2.0
1	A	164	VAL	2.0

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 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.

Continued on next page...

Continued from previous page...

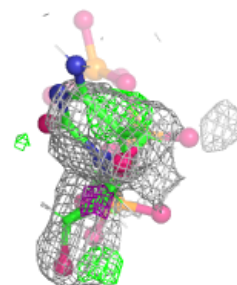
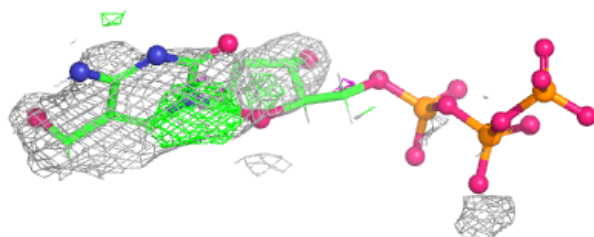
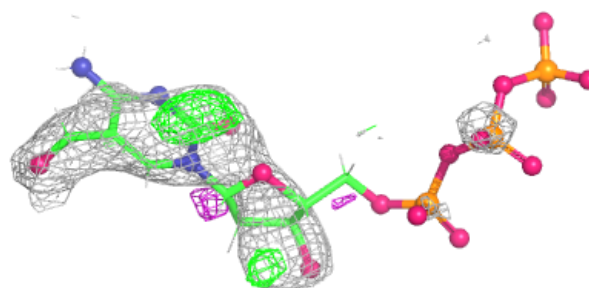
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
-----	------	-------	-----	-------	------	-----	-----------------------------	-------

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
2	A1EOD	A	301	30/30	0.63	0.26	30,71,105,114	12

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.

Electron density around A1EOD A 301:

2mF_o-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



6.5 Other polymers [i](#)

There are no such residues in this entry.