



Full wwPDB X-ray Structure Validation Report ⓘ

Aug 29, 2026 – 11:40 pm BST

PDB ID : 9SIM / pdb_00009sim
Title : Phage epsilon15 tailspike gp20 containing domains beta-helix, beta-sandwich and petal domains
Authors : Seoane-Blanco, M.; van Raaij, M.J.
Deposited on : 2025-08-28
Resolution : 1.94 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	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.50

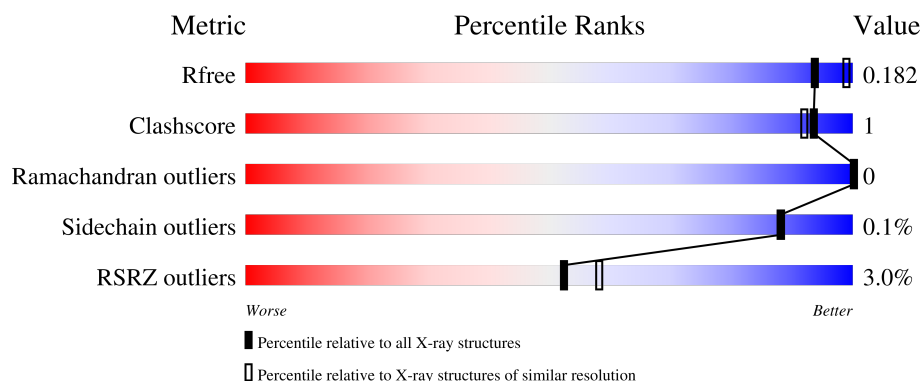
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

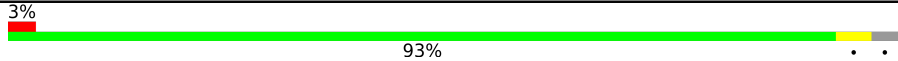
The reported resolution of this entry is 1.94 Å.

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	1452 (1.94-1.94)
Clashscore	190562	1494 (1.94-1.94)
Ramachandran outliers	187476	1479 (1.94-1.94)
Sidechain outliers	187428	1479 (1.94-1.94)
RSRZ outliers	180081	1453 (1.94-1.94)

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	857	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	FMT	A	1113	-	-	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 13479 atoms, of which 6263 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 Tail spike protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	832	Total	C	H	N	O	S	218	28	0
			12621	3997	6201	1116	1282	25			

There are 35 discrepancies between the modelled and reference sequences:

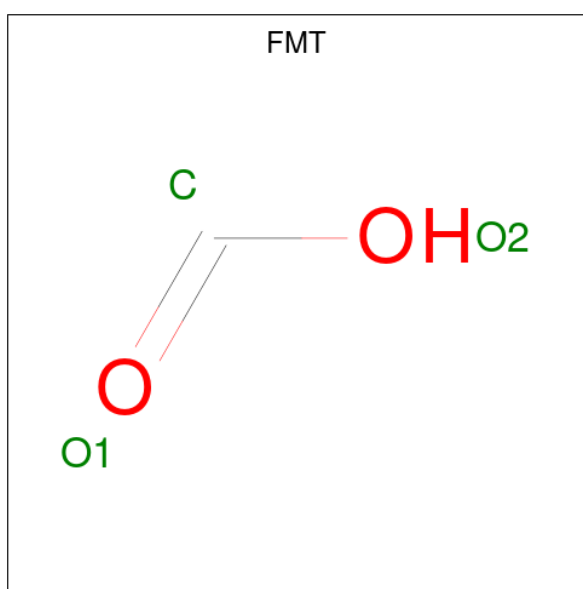
Chain	Residue	Modelled	Actual	Comment	Reference
A	214	MET	-	initiating methionine	UNP Q858F5
A	215	GLY	-	expression tag	UNP Q858F5
A	216	SER	-	expression tag	UNP Q858F5
A	217	SER	-	expression tag	UNP Q858F5
A	218	HIS	-	expression tag	UNP Q858F5
A	219	HIS	-	expression tag	UNP Q858F5
A	220	HIS	-	expression tag	UNP Q858F5
A	221	HIS	-	expression tag	UNP Q858F5
A	222	HIS	-	expression tag	UNP Q858F5
A	223	HIS	-	expression tag	UNP Q858F5
A	224	SER	-	expression tag	UNP Q858F5
A	225	SER	-	expression tag	UNP Q858F5
A	226	GLY	-	expression tag	UNP Q858F5
A	227	LEU	-	expression tag	UNP Q858F5
A	228	VAL	-	expression tag	UNP Q858F5
A	229	PRO	-	expression tag	UNP Q858F5
A	230	ARG	-	expression tag	UNP Q858F5
A	231	GLY	-	expression tag	UNP Q858F5
A	232	SER	-	expression tag	UNP Q858F5
A	233	HIS	-	expression tag	UNP Q858F5
A	234	MET	-	expression tag	UNP Q858F5
A	235	ALA	-	expression tag	UNP Q858F5
A	236	SER	-	expression tag	UNP Q858F5
A	237	MET	-	expression tag	UNP Q858F5
A	238	THR	-	expression tag	UNP Q858F5
A	239	GLY	-	expression tag	UNP Q858F5
A	240	GLY	-	expression tag	UNP Q858F5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	241	GLN	-	expression tag	UNP Q858F5
A	242	GLN	-	expression tag	UNP Q858F5
A	243	MET	-	expression tag	UNP Q858F5
A	244	GLY	-	expression tag	UNP Q858F5
A	245	ARG	-	expression tag	UNP Q858F5
A	246	ILE	-	expression tag	UNP Q858F5
A	247	LEU	-	expression tag	UNP Q858F5
A	308	ILE	CYS	conflict	UNP Q858F5

- Molecule 2 is FORMIC ACID (CCD ID: FMT) (formula: CH_2O_2).



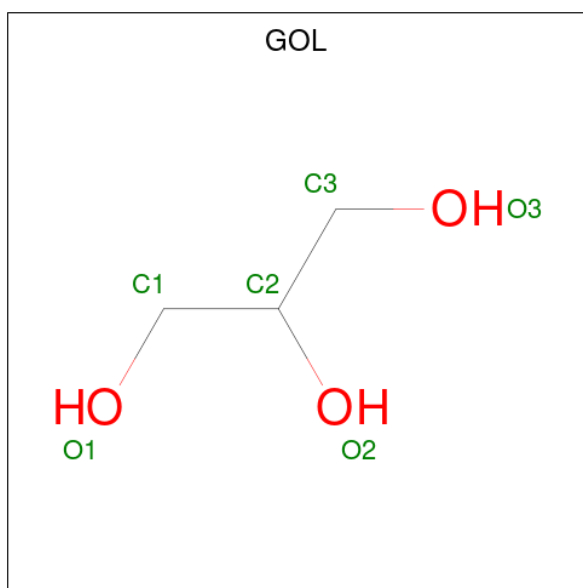
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		
2	A	1	Total	C	H	O	1	0
			5	1	2	2		

- Molecule 3 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	3	0
			14	3	8	3		
3	A	1	Total	C	H	O	3	0
			14	3	8	3		
3	A	1	Total	C	H	O	3	0
			14	3	8	3		

- Molecule 4 is SODIUM ION (CCD ID: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Na	0	0
			2	2		

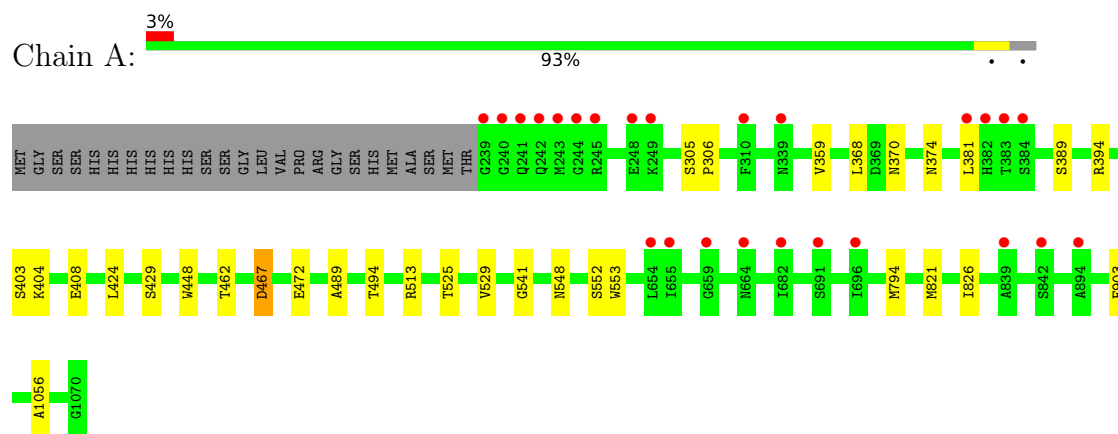
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	714	Total	O	0	5
			719	719		

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: Tail spike protein



4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α , β , γ	117.33Å 117.33Å 251.29Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	83.90 – 1.94 83.90 – 1.94	Depositor EDS
% Data completeness (in resolution range)	96.3 (83.90-1.94) 96.3 (83.90-1.94)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.14 (at 1.94Å)	Xtriage
Refinement program	REFMAC 5.8.0430 (refmacat 0.4.105)	Depositor
R, R_{free}	0.159 , 0.178 0.163 , 0.182	Depositor DCC
R_{free} test set	4737 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å ²)	29.8	Xtriage
Anisotropy	0.217	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.43 , 44.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.024 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	13479	wwPDB-VP
Average B, all atoms (Å ²)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.30% 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: GOL, NA, FMT

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	0.47	0/6630	0.77	1/9018 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	467	ASP	CA-CB-CG	5.36	117.95	112.60

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	6420	6201	6145	18	0
2	A	57	38	19	0	0
3	A	18	24	24	0	0
4	A	2	0	0	0	0
5	A	719	0	0	0	0
All	All	7216	6263	6188	18	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.

All (18) 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:A:462:THR:HA	1:A:489:ALA:O	2.11	0.50
1:A:794:MET:HG3	1:A:1056:ALA:HB2	1.95	0.49
1:A:467:ASP:HA	1:A:494:THR:O	2.13	0.48
1:A:525:THR:HA	1:A:548:ASN:O	2.15	0.46
1:A:389:SER:OG	1:A:394:ARG:NH2	2.49	0.46
1:A:403[B]:SER:HA	1:A:424:LEU:O	2.17	0.44
1:A:529:VAL:O	1:A:552:SER:HA	2.18	0.44
1:A:359:VAL:HG11	1:A:368[B]:LEU:HD11	1.99	0.43
1:A:370:ASN:HA	1:A:404[A]:LYS:O	2.18	0.43
1:A:408:GLU:HA	1:A:429:SER:O	2.18	0.43
1:A:305:SER:HB2	1:A:306:PRO:CD	2.49	0.43
1:A:821:MET:HG2	1:A:826:ILE:HG13	2.01	0.42
1:A:513:ARG:NH2	1:A:541:GLY:O	2.43	0.41
1:A:381:LEU:HD21	1:A:448:TRP:CH2	2.55	0.41
1:A:821:MET:HE1	1:A:903:PHE:HD1	1.85	0.41
1:A:374:ASN:HA	1:A:408:GLU:O	2.21	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	858/857 (100%)	824 (96%)	34 (4%)	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	707/703 (101%)	706 (100%)	1 (0%)	88	88

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	472	GLU

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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 24 ligands modelled in this entry, 2 are monoatomic - leaving 22 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	FMT	A	1115	-	2,2,2	1.04	0	1,1,1	0.22	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	FMT	A	1101	-	2,2,2	1.03	0	1,1,1	0.23	0
2	FMT	A	1108	-	2,2,2	1.09	0	1,1,1	0.22	0
2	FMT	A	1111	-	2,2,2	1.04	0	1,1,1	0.22	0
2	FMT	A	1104	-	2,2,2	0.91	0	1,1,1	0.29	0
3	GOL	A	1120	-	5,5,5	0.10	0	5,5,5	0.29	0
2	FMT	A	1102	-	2,2,2	1.01	0	1,1,1	0.22	0
2	FMT	A	1117	-	2,2,2	1.07	0	1,1,1	0.22	0
2	FMT	A	1107	-	2,2,2	0.95	0	1,1,1	0.25	0
2	FMT	A	1112	-	2,2,2	1.04	0	1,1,1	0.24	0
3	GOL	A	1122	-	5,5,5	0.10	0	5,5,5	0.29	0
2	FMT	A	1109	-	2,2,2	1.00	0	1,1,1	0.25	0
2	FMT	A	1110	-	2,2,2	1.01	0	1,1,1	0.24	0
2	FMT	A	1113	-	2,2,2	1.04	0	1,1,1	0.24	0
2	FMT	A	1105	-	2,2,2	1.00	0	1,1,1	0.23	0
2	FMT	A	1116	-	2,2,2	1.00	0	1,1,1	0.24	0
2	FMT	A	1118	-	2,2,2	1.05	0	1,1,1	0.23	0
2	FMT	A	1106	-	2,2,2	1.02	0	1,1,1	0.24	0
2	FMT	A	1114	-	2,2,2	1.03	0	1,1,1	0.22	0
3	GOL	A	1121	-	5,5,5	0.11	0	5,5,5	0.33	0
2	FMT	A	1103	-	2,2,2	1.06	0	1,1,1	0.23	0
2	FMT	A	1119	-	2,2,2	1.03	0	1,1,1	0.24	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
3	GOL	A	1121	-	-	0/4/4/4	-
3	GOL	A	1120	-	-	0/4/4/4	-
3	GOL	A	1122	-	-	1/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1122	GOL	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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	832/857 (97%)	-0.05	25 (3%)	52 58	11, 28, 52, 84	21 (2%)

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	242[A]	GLN	5.0
1	A	239	GLY	3.6
1	A	241	GLN	3.5
1	A	243	MET	3.3
1	A	381	LEU	3.3
1	A	310	PHE	3.2
1	A	682[A]	ILE	3.1
1	A	240	GLY	3.1
1	A	382	HIS	3.0
1	A	249	LYS	2.7
1	A	894	ALA	2.6
1	A	384	SER	2.5
1	A	664	ASN	2.5
1	A	839	ALA	2.5
1	A	383	THR	2.4
1	A	248	GLU	2.4
1	A	696	ILE	2.4
1	A	659	GLY	2.3
1	A	654	LEU	2.2
1	A	655	ILE	2.1
1	A	691	SER	2.1
1	A	339	ASN	2.1
1	A	842	SER	2.0
1	A	244	GLY	2.0
1	A	245	ARG	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.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
2	FMT	A	1113	3/3	0.69	0.43	71,72,72,73	1
2	FMT	A	1115	3/3	0.75	0.28	67,67,69,69	1
3	GOL	A	1121	6/6	0.76	0.21	52,58,61,61	3
3	GOL	A	1122	6/6	0.81	0.26	61,62,64,64	3
2	FMT	A	1108	3/3	0.82	0.25	54,55,57,57	1
2	FMT	A	1119	3/3	0.82	0.26	61,62,63,63	1
2	FMT	A	1109	3/3	0.83	0.19	68,68,68,68	1
2	FMT	A	1110	3/3	0.84	0.22	68,68,68,69	1
2	FMT	A	1116	3/3	0.87	0.23	67,67,68,69	1
2	FMT	A	1118	3/3	0.88	0.27	57,59,60,60	1
2	FMT	A	1114	3/3	0.88	0.18	59,60,61,61	1
2	FMT	A	1106	3/3	0.89	0.15	54,56,56,56	1
2	FMT	A	1112	3/3	0.89	0.19	54,56,57,58	1
3	GOL	A	1120	6/6	0.91	0.11	33,36,39,39	3
2	FMT	A	1105	3/3	0.91	0.18	51,53,53,53	1
2	FMT	A	1107	3/3	0.91	0.13	39,39,41,41	1
2	FMT	A	1117	3/3	0.92	0.25	53,54,55,55	1
2	FMT	A	1103	3/3	0.92	0.16	48,49,49,50	1
2	FMT	A	1111	3/3	0.93	0.17	44,45,48,49	1
2	FMT	A	1104	3/3	0.93	0.12	35,36,40,40	1
2	FMT	A	1102	3/3	0.94	0.13	41,41,44,44	1
2	FMT	A	1101	3/3	0.97	0.08	35,36,37,37	1
4	NA	A	1123	1/1	0.98	0.03	25,25,25,25	0
4	NA	A	1124	1/1	0.98	0.06	40,40,40,40	0

6.5 Other polymers [i](#)

There are no such residues in this entry.