



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 9, 2026 – 06:16 AM UTC

PDB ID : 4ZM5 / pdb_00004zm5
Title : Shigella flexneri lipopolysaccharide O-antigen chain-length regulator WzzBSF
- A107P mutant
Authors : Ericsson, D.J.; Chang, C.-W.; Lonhienne, T.; Casey, L.; Benning, F.; Kobe,
B.; Tran, E.N.H.; Morona, R.
Deposited on : 2015-05-02
Resolution : 2.47 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
Xtriage (Phenix)	:	2.0
EDS	:	3.0
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

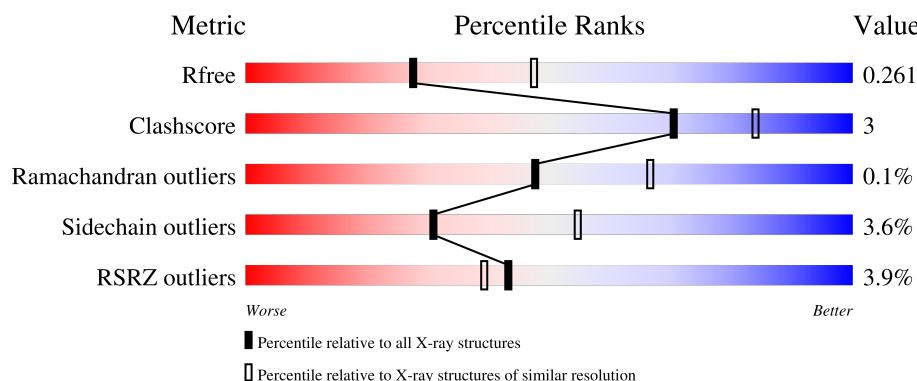
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

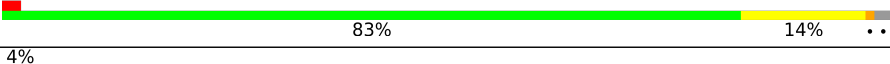
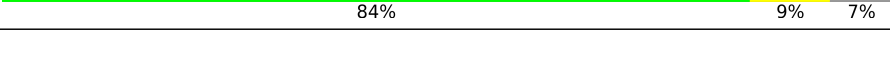
The reported resolution of this entry is 2.47 Å.

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	7589 (2.50-2.46)
Clashscore	190562	8295 (2.50-2.46)
Ramachandran outliers	187476	8164 (2.50-2.46)
Sidechain outliers	187428	8166 (2.50-2.46)
RSRZ outliers	180081	7593 (2.50-2.46)

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	243	 5% 85% 8% 5%
1	B	243	 2% 83% 14% ..
1	C	243	 4% 84% 9% 7%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5699 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 Chain length determinant protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	230	Total	C	N	O	S	0	0	0
			1841	1153	315	369	4			
1	B	237	Total	C	N	O	S	0	0	0
			1895	1186	326	379	4			
1	C	227	Total	C	N	O	S	0	1	0
			1820	1143	315	358	4			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP P37792
A	2	ASN	-	expression tag	UNP P37792
A	3	ALA	-	expression tag	UNP P37792
A	57	PRO	ALA	engineered mutation	UNP P37792
B	1	SER	-	expression tag	UNP P37792
B	2	ASN	-	expression tag	UNP P37792
B	3	ALA	-	expression tag	UNP P37792
B	57	PRO	ALA	engineered mutation	UNP P37792
C	1	SER	-	expression tag	UNP P37792
C	2	ASN	-	expression tag	UNP P37792
C	3	ALA	-	expression tag	UNP P37792
C	57	PRO	ALA	engineered mutation	UNP P37792

- Molecule 2 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Cl	0	0
			1	1		

- Molecule 3 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

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

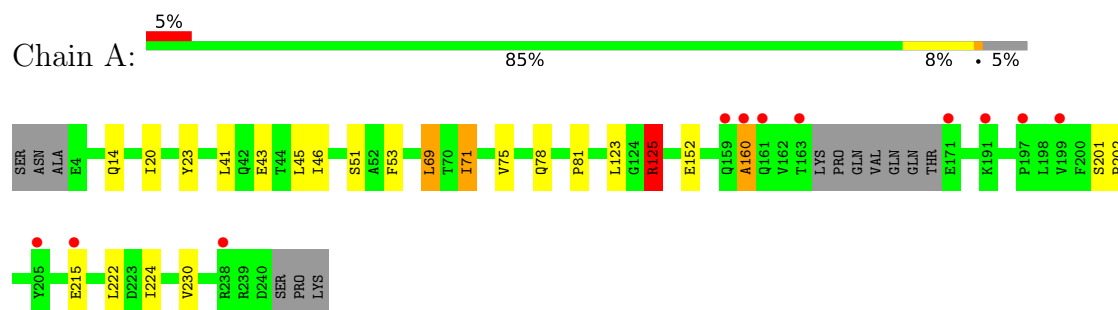
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	51	Total 51	O 51	0	0
4	B	53	Total 53	O 53	0	0
4	C	37	Total 37	O 37	0	0

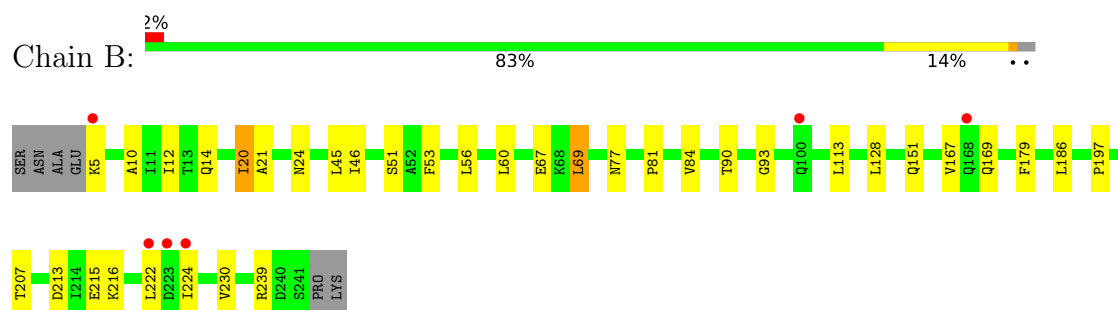
3 Residue-property plots [i](#)

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.

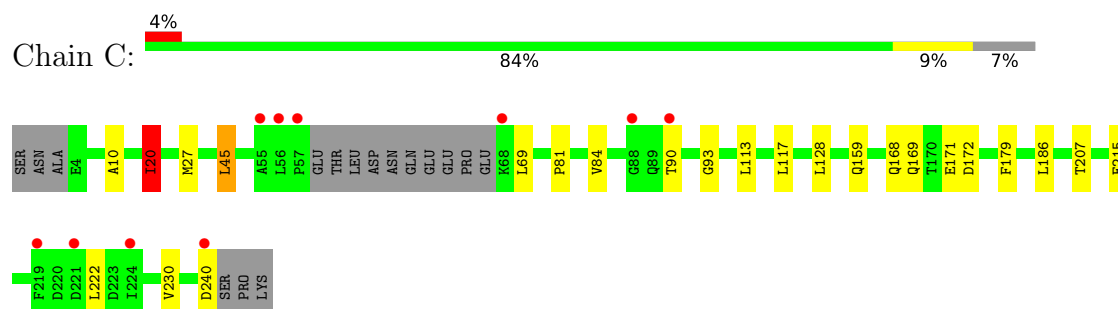
- Molecule 1: Chain length determinant protein



- Molecule 1: Chain length determinant protein



- Molecule 1: Chain length determinant protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	80.57Å 62.46Å 90.24Å 90.00° 94.06° 90.00°	Depositor
Resolution (Å)	32.14 – 2.47 32.14 – 2.48	Depositor EDS
% Data completeness (in resolution range)	97.7 (32.14-2.47) 97.4 (32.14-2.48)	Depositor EDS
R_{merge}	0.22	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.82 (at 2.48Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.195 , 0.242 0.209 , 0.261	Depositor DCC
R_{free} test set	1572 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å ²)	34.0	Xtriage
Anisotropy	0.437	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 57.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	5699	wwPDB-VP
Average B, all atoms (Å ²)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.83% 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: CL, MG

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	0.85	0/1867	1.37	8/2529 (0.3%)
1	B	0.82	0/1923	1.36	4/2607 (0.2%)
1	C	0.81	0/1849	1.34	3/2504 (0.1%)
All	All	0.83	0/5639	1.36	15/7640 (0.2%)

There are no bond length outliers.

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	230	VAL	N-CA-C	-8.57	104.44	111.81
1	B	230	VAL	N-CA-C	-7.69	103.43	111.58
1	B	77	ASN	CA-CB-CG	7.52	120.12	112.60
1	C	230	VAL	N-CA-C	-7.42	103.71	111.58
1	A	71	ILE	N-CA-CB	5.88	119.28	111.64
1	C	172	ASP	CA-CB-CG	5.49	118.09	112.60
1	A	160	ALA	CA-C-N	5.42	129.45	120.94
1	A	160	ALA	C-N-CA	5.42	129.45	120.94
1	C	20	ILE	N-CA-CB	-5.33	102.76	111.98
1	A	125	ARG	CA-C-N	5.19	127.54	120.54
1	A	125	ARG	C-N-CA	5.19	127.54	120.54
1	B	21	ALA	CA-C-N	5.11	125.61	119.94
1	B	21	ALA	C-N-CA	5.11	125.61	119.94
1	A	202	PRO	CA-C-N	5.07	127.58	120.28
1	A	202	PRO	C-N-CA	5.07	127.58	120.28

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	1841	0	1833	13	0
1	B	1895	0	1893	20	0
1	C	1820	0	1832	11	0
2	B	1	0	0	0	0
3	B	1	0	0	0	0
4	A	51	0	0	0	0
4	B	53	0	0	1	0
4	C	37	0	0	0	0
All	All	5699	0	5558	38	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 3.

All (38) 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:125:ARG:HH12	1:A:215:GLU:HG3	1.25	1.00
1:C:20:ILE:HD11	1:C:45:LEU:HD13	1.70	0.73
1:B:53:PHE:CE1	1:B:69:LEU:HD21	2.26	0.70
1:C:45:LEU:HD21	1:C:113:LEU:HD11	1.74	0.69
1:A:123:LEU:HD11	1:B:24:ASN:ND2	2.09	0.68
1:C:90:THR:HG23	1:C:93:GLY:H	1.60	0.66
1:C:45:LEU:CD2	1:C:113:LEU:HD11	2.27	0.64
1:A:125:ARG:NH1	1:A:215:GLU:HG3	2.06	0.62
1:B:113:LEU:HD13	1:B:224:ILE:HD12	1.82	0.62
1:A:53:PHE:HE1	1:A:69:LEU:HD21	1.64	0.61
1:A:23:TYR:HB2	1:A:224:ILE:HD13	1.84	0.58
1:B:20:ILE:HD11	1:B:45:LEU:HD13	1.86	0.57
1:A:75:VAL:HB	1:A:78:GLN:HB2	1.87	0.57
1:B:10:ALA:HB3	1:B:84:VAL:HG23	1.87	0.56
1:C:10:ALA:HB3	1:C:84:VAL:HG23	1.90	0.54
1:B:5:LYS:HG2	1:B:239:ARG:HD3	1.89	0.54
1:A:123:LEU:HD11	1:B:24:ASN:HD22	1.73	0.53
1:B:151:GLN:HG2	4:B:551:HOH:O	2.10	0.51
1:A:152:GLU:HG3	1:B:197:PRO:HD3	1.95	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:41:LEU:O	1:A:45:LEU:HG	2.13	0.48
1:C:128:LEU:HD22	1:C:207:THR:HG23	1.97	0.47
1:B:179:PHE:HA	1:B:186:LEU:HD11	1.98	0.46
1:C:179:PHE:HA	1:C:186:LEU:HD11	1.98	0.46
1:B:5:LYS:HG2	1:B:239:ARG:HH11	1.81	0.46
1:C:27:MET:HE3	1:C:117:LEU:HD12	1.97	0.46
1:B:169:GLN:HB3	1:C:169:GLN:HG2	1.98	0.45
1:B:14:GLN:HA	1:B:46:ILE:HB	1.99	0.44
1:A:75:VAL:HG23	1:A:81:PRO:HD2	1.99	0.44
1:A:14:GLN:HA	1:A:46:ILE:HB	2.00	0.43
1:B:53:PHE:HE1	1:B:69:LEU:HD21	1.78	0.43
1:B:90:THR:HG23	1:B:93:GLY:H	1.84	0.42
1:C:45:LEU:HD21	1:C:113:LEU:CD1	2.45	0.42
1:B:60:LEU:HB3	1:B:67:GLU:HB2	2.01	0.42
1:B:51:SER:HA	1:C:81:PRO:HG3	2.03	0.41
1:A:51:SER:HA	1:B:81:PRO:HG3	2.03	0.41
1:A:43:GLU:HA	1:A:46:ILE:HG22	2.01	0.41
1:B:128:LEU:HD22	1:B:207:THR:HG23	2.02	0.40
1:B:213:ASP:HA	1:B:216:LYS:HD3	2.02	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	226/243 (93%)	221 (98%)	4 (2%)	1 (0%)	30	46
1	B	235/243 (97%)	233 (99%)	2 (1%)	0	100	100
1	C	224/243 (92%)	222 (99%)	2 (1%)	0	100	100
All	All	685/729 (94%)	676 (99%)	8 (1%)	1 (0%)	48	66

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	160	ALA

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	204/216 (94%)	198 (97%)	6 (3%)	37	62
1	B	211/216 (98%)	204 (97%)	7 (3%)	33	58
1	C	202/216 (94%)	193 (96%)	9 (4%)	24	46
All	All	617/648 (95%)	595 (96%)	22 (4%)	31	55

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	20	ILE
1	A	69	LEU
1	A	71	ILE
1	A	125	ARG
1	A	201	SER
1	A	222	LEU
1	B	12	ILE
1	B	20	ILE
1	B	56	LEU
1	B	69	LEU
1	B	167	VAL
1	B	215	GLU
1	B	222	LEU
1	C	20	ILE
1	C	45	LEU
1	C	69	LEU
1	C	159	GLN
1	C	168	GLN
1	C	171	GLU
1	C	215	GLU
1	C	222	LEU
1	C	240	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (14) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	14	GLN
1	A	28	ASN
1	A	63	GLN
1	A	100	GLN
1	A	110	ASN
1	A	135	GLN
1	A	161	GLN
1	B	14	GLN
1	B	100	GLN
1	B	155	GLN
1	B	159	GLN
1	C	77	ASN
1	C	175	GLN
1	C	192	HIS

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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	230/243 (94%)	0.27	11 (4%) 35 32	22, 49, 90, 127	0
1	B	237/243 (97%)	0.13	6 (2%) 58 55	23, 47, 85, 98	0
1	C	227/243 (93%)	0.34	10 (4%) 39 35	24, 52, 93, 123	1 (0%)
All	All	694/729 (95%)	0.25	27 (3%) 43 39	22, 49, 90, 127	1 (0%)

All (27) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	57	PRO	7.7
1	C	219	PHE	4.3
1	A	197	PRO	3.7
1	B	222	LEU	3.7
1	C	56	LEU	3.5
1	C	55	ALA	3.3
1	C	68	LYS	3.2
1	C	224	ILE	3.0
1	B	168	GLN	3.0
1	A	171	GLU	2.8
1	A	163	THR	2.7
1	B	100	GLN	2.7
1	A	161	GLN	2.6
1	A	199	VAL	2.5
1	B	5	LYS	2.5
1	A	191	LYS	2.4
1	A	160	ALA	2.3
1	B	224	ILE	2.2
1	A	215	GLU	2.2
1	A	159	GLN	2.2
1	C	88	GLY	2.1
1	C	240	ASP	2.1
1	A	238	ARG	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	223	ASP	2.1
1	C	221	ASP	2.1
1	A	205	TYR	2.1
1	C	90	THR	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
3	MG	B	401	1/1	0.94	0.10	36,36,36,36	0
2	CL	B	400	1/1	0.95	0.09	57,57,57,57	0

6.5 Other polymers [i](#)

There are no such residues in this entry.