Full wwPDB X-ray Structure Validation Report (i)

Jul 3, 2024 – 08:37 pm BST

PDB ID	:	4ZBL
Title	:	Phototoxic fluorescent protein mKillerOrange
Authors	:	Pletnev, V.Z.; Pletneva, N.V.; Pletnev, S.V.
Deposited on	:	2015-04-14
Resolution	:	1.57 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 (i) 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 (i)) were used in the production of this report:

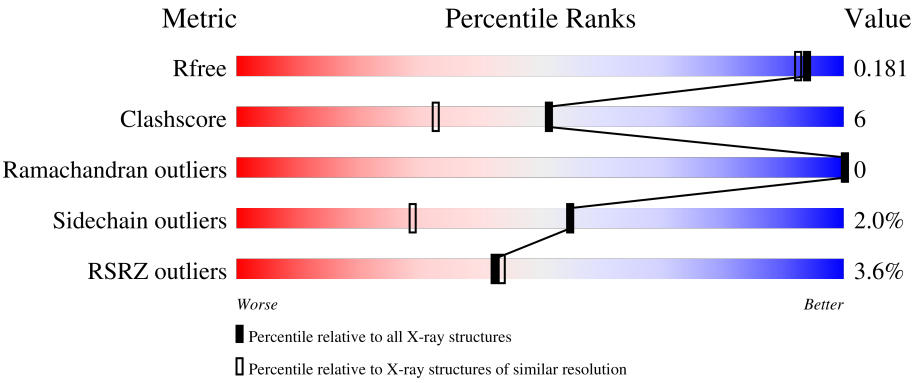
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.57 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	5534 (1.60-1.56)
Clashscore	141614	5861 (1.60-1.56)
Ramachandran outliers	138981	5708 (1.60-1.56)
Sidechain outliers	138945	5703 (1.60-1.56)
RSRZ outliers	127900	5431 (1.60-1.56)

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 <=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	240	3%, 84%, 9%, 6%

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
1	4M9	A	65[A]	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2091 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.

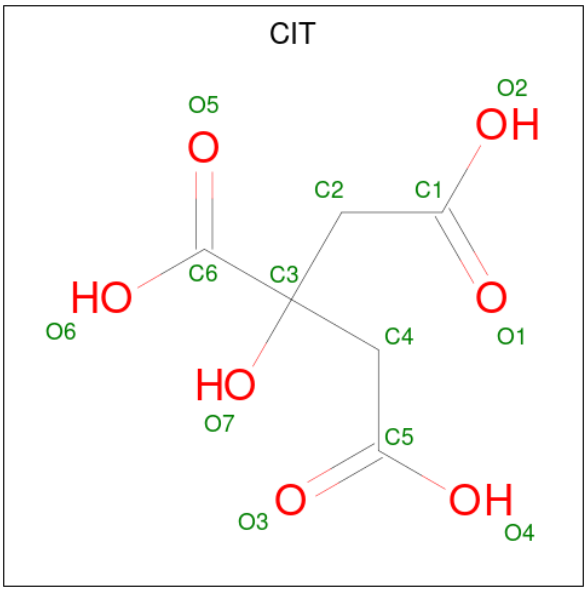
- Molecule 1 is a protein called KillerOrange.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	226	1908	1198	330	363	17	0	14	0

There are 24 discrepancies between the modelled and reference sequences:

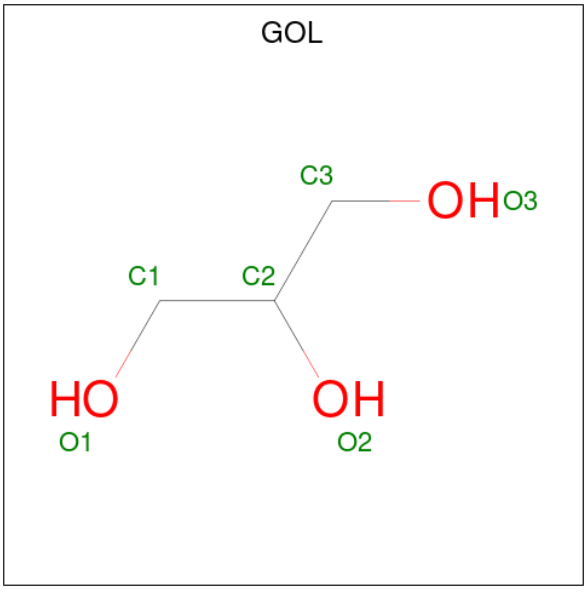
Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	initiating methionine	UNP Q2TCH5
A	-12	ARG	-	expression tag	UNP Q2TCH5
A	-11	GLY	-	expression tag	UNP Q2TCH5
A	-10	SER	-	expression tag	UNP Q2TCH5
A	-9	HIS	-	expression tag	UNP Q2TCH5
A	-8	HIS	-	expression tag	UNP Q2TCH5
A	-7	HIS	-	expression tag	UNP Q2TCH5
A	-6	HIS	-	expression tag	UNP Q2TCH5
A	-5	HIS	-	expression tag	UNP Q2TCH5
A	-4	HIS	-	expression tag	UNP Q2TCH5
A	-3	GLY	-	expression tag	UNP Q2TCH5
A	-2	SER	-	expression tag	UNP Q2TCH5
A	-1	ARG	-	expression tag	UNP Q2TCH5
A	0	GLY	-	expression tag	UNP Q2TCH5
A	1	SER	-	expression tag	UNP Q2TCH5
A	3	VAL	GLY	conflict	UNP Q2TCH5
A	65	4M9	GLN	chromophore	UNP Q2TCH5
A	65	4M9	TYR	chromophore	UNP Q2TCH5
A	65	4M9	GLY	chromophore	UNP Q2TCH5
A	145	SER	ASN	conflict	UNP Q2TCH5
A	160	THR	LEU	conflict	UNP Q2TCH5
A	162	THR	PHE	conflict	UNP Q2TCH5
A	172	LYS	LEU	conflict	UNP Q2TCH5
A	204	THR	MET	conflict	UNP Q2TCH5

- Molecule 2 is CITRIC ACID (three-letter code: CIT) (formula: $C_6H_8O_7$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 13 6 7	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0

- Molecule 4 is water.

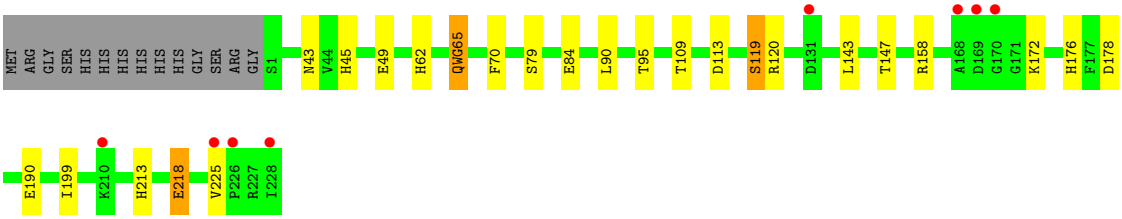
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	158	Total	O	0	0
			158	158		

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: KillerOrange

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	64.16Å 64.16Å 47.35Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	27.80 – 1.57 27.78 – 1.57	Depositor EDS
% Data completeness (in resolution range)	98.2 (27.80-1.57) 98.2 (27.78-1.57)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.11 (at 1.57Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.144 , 0.175 0.156 , 0.181	Depositor DCC
R_{free} test set	1506 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	13.0	Xtriage
Anisotropy	0.023	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 48.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.016 for -h,-k,l 0.031 for h,-h-k,-l 0.017 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2091	wwPDB-VP
Average B, all atoms (Å^2)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.34% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: GOL, 4M9, CIT

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	1.15	3/1907 (0.2%)	1.07	1/2582 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	218	GLU	CD-OE1	-6.63	1.18	1.25
1	A	49	GLU	CD-OE2	-5.74	1.19	1.25
1	A	119	SER	CB-OG	-5.28	1.35	1.42

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	158	ARG	NE-CZ-NH2	-5.85	117.38	120.30

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	1908	0	1768	21	0
2	A	13	0	5	0	0
3	A	12	0	16	3	0
4	A	158	0	0	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	2091	0	1789	21	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 6.

All (21) 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:65[A]:4M9:CD1	4:A:401:HOH:O	1.75	1.27
1:A:65[A]:4M9:NE1	4:A:401:HOH:O	1.77	1.09
1:A:79[B]:SER:OG	1:A:190:GLU:OE1	2.03	0.77
1:A:65[A]:4M9:CZ2	4:A:405:HOH:O	2.38	0.71
1:A:65[A]:4M9:H12	4:A:405:HOH:O	1.90	0.71
1:A:113:ASP:OD2	4:A:402:HOH:O	2.12	0.68
1:A:65[A]:4M9:H10	4:A:401:HOH:O	1.65	0.66
1:A:65[A]:4M9:CE2	4:A:405:HOH:O	2.51	0.59
1:A:43:ASN:HB2	3:A:303:GOL:H2	1.86	0.57
1:A:45:HIS:NE2	3:A:303:GOL:O2	2.42	0.53
1:A:225:VAL:HG22	4:A:533:HOH:O	2.09	0.52
1:A:79[A]:SER:OG	1:A:84[A]:GLU:OE2	2.28	0.52
1:A:176[B]:HIS:CE1	1:A:178:ASP:OD2	2.64	0.51
1:A:95[B]:THR:HG23	4:A:482:HOH:O	2.11	0.50
1:A:143:LEU:HD11	1:A:172:LYS:HD3	1.95	0.48
1:A:62:HIS:HA	1:A:65[A]:4M9:CA2	2.44	0.47
1:A:62:HIS:O	1:A:65[A]:4M9:C2	2.65	0.44
1:A:147[B]:THR:HG21	1:A:199:ILE:HD12	2.01	0.43
1:A:65[A]:4M9:H10	1:A:218:GLU:OE2	2.20	0.42
1:A:45:HIS:CD2	3:A:303:GOL:O2	2.73	0.41
1:A:109:THR:O	1:A:119:SER:HA	2.21	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	234/240 (98%)	226 (97%)	8 (3%)	0	100 100

There are no Ramachandran outliers to report.

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	209/207 (101%)	205 (98%)	4 (2%)	57 31

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

Mol	Chain	Res	Type
1	A	70	PHE
1	A	90	LEU
1	A	120	ARG
1	A	213	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are 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	$\#\|Z\| > 2$
1	4M9	A	65[A]	1	26,28,29	1.11	2 (7%)	29,39,41	1.17	3 (10%)
1	4M9	A	65[B]	1	26,28,29	1.88	6 (23%)	29,39,41	2.70	10 (34%)

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
1	4M9	A	65[A]	1	-	0/8/32/33	0/3/3/3
1	4M9	A	65[B]	1	-	2/8/32/33	0/3/3/3

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	65[B]	4M9	CA1-N1	4.49	1.38	1.27
1	A	65[B]	4M9	C1-N3	4.19	1.45	1.38
1	A	65[B]	4M9	CE3-CD2	-3.23	1.35	1.42
1	A	65[B]	4M9	CZ3-CE3	2.87	1.43	1.36
1	A	65[B]	4M9	CB2-CA2	2.84	1.37	1.35
1	A	65[B]	4M9	CA2-N2	2.73	1.44	1.38
1	A	65[A]	4M9	CB2-CA2	2.22	1.37	1.35
1	A	65[A]	4M9	CA1-N1	2.14	1.32	1.27

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	65[B]	4M9	CB2-CA2-C2	9.13	133.18	122.28
1	A	65[B]	4M9	CB2-CA2-N2	-7.96	117.78	128.83
1	A	65[B]	4M9	CG2-CD1-NE1	3.15	113.90	108.59
1	A	65[A]	4M9	O3-C3-CA3	-3.09	117.06	126.39
1	A	65[B]	4M9	O3-C3-CA3	-2.87	117.74	126.39
1	A	65[B]	4M9	CE3-CD2-CE2	2.77	121.85	118.17
1	A	65[A]	4M9	CE3-CD2-CG2	-2.36	131.44	135.45
1	A	65[B]	4M9	CH2-CZ2-CE2	-2.30	116.78	120.08
1	A	65[B]	4M9	O2-C2-CA2	-2.28	129.68	130.96
1	A	65[B]	4M9	CG2-CB2-CA2	2.22	135.10	130.81
1	A	65[B]	4M9	CE3-CD2-CG2	-2.17	131.77	135.45

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	65[A]	4M9	CE3-CD2-CE2	2.16	121.03	118.17
1	A	65[B]	4M9	CZ3-CE3-CD2	-2.10	117.98	120.89

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	65[B]	4M9	C2-CA2-CB2-CG2
1	A	65[B]	4M9	N2-CA2-CB2-CG2

There are no ring outliers.

1 monomer is involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	65[A]	4M9	9	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

3 ligands are 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	$\#\|Z\| > 2$
3	GOL	A	303	-	5,5,5	0.71	0	5,5,5	1.09	0
2	CIT	A	301	-	12,12,12	1.95	4 (33%)	17,17,17	1.56	4 (23%)
3	GOL	A	302	-	5,5,5	0.48	0	5,5,5	0.70	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	303	-	-	3/4/4/4	-
2	CIT	A	301	-	-	0/16/16/16	-
3	GOL	A	302	-	-	0/4/4/4	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	301	CIT	O2-C1	-3.52	1.18	1.30
2	A	301	CIT	O4-C5	-3.48	1.19	1.30
2	A	301	CIT	C3-C6	-2.78	1.50	1.53
2	A	301	CIT	O6-C6	-2.74	1.20	1.30

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	301	CIT	O6-C6-C3	3.11	118.45	113.05
2	A	301	CIT	C3-C2-C1	-2.35	108.12	113.81
2	A	301	CIT	C2-C3-C6	-2.22	105.33	110.11
2	A	301	CIT	C4-C3-C2	2.08	114.58	109.16

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	303	GOL	O1-C1-C2-C3
3	A	303	GOL	O1-C1-C2-O2
3	A	303	GOL	C1-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	303	GOL	3	0

5.7 Other polymers

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, 95^{th} 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	225/240 (93%)	0.05	8 (3%) 42 44	8, 14, 29, 51	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	228	ILE	5.6
1	A	225	VAL	5.6
1	A	169	ASP	3.8
1	A	210	LYS	3.0
1	A	226	PRO	2.8
1	A	131	ASP	2.7
1	A	170	GLY	2.5
1	A	168	ALA	2.5

6.2 Non-standard residues in protein, DNA, RNA chains

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, 95^{th} 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
1	4M9	A	65[A]	26/27	0.96	0.16	7,7,8,9	26
1	4M9	A	65[B]	26/27	0.96	0.16	8,10,15,15	0

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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, 95^{th} 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
3	GOL	A	303	6/6	0.73	0.22	30,36,38,41	0
3	GOL	A	302	6/6	0.93	0.14	13,23,24,26	0
2	CIT	A	301	13/13	0.96	0.08	10,15,19,20	0

6.5 Other polymers

There are no such residues in this entry.