Full wwPDB X-ray Structure Validation Report (i)

Sep 7, 2023 – 01:37 AM EDT

PDB ID	:	4GF2
Title	:	Crystal structure of Plasmodium falciparum Erythrocyte Binding Antigen 140 (PfEBA-140/BAEBL)
Authors	:	Lin, D.H.; Malpede, B.M.; Batchelor, J.D.; Tolia, N.H.
Deposited on	:	2012-08-02
Resolution	:	2.40 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
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)	:	2.35

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 2.40 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	Percentile Ranks	Value
Rfree		0.236
Clashscore		1
Ramachandran outliers		0.5%
Sidechain outliers		2.0%
RSRZ outliers		11.3%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
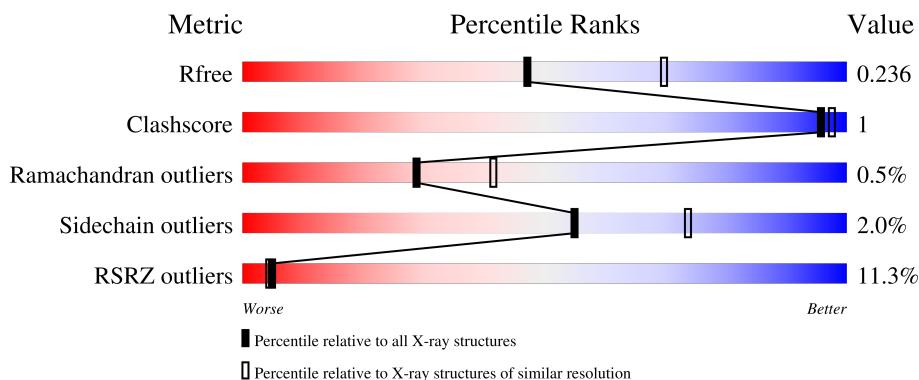
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	615	11% ... 92% ...
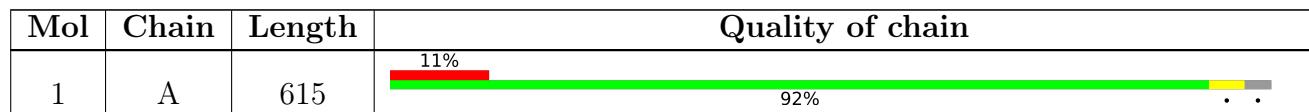
The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GOL	A	801	-	-	-	X

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 10216 atoms, of which 5006 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 Erythrocyte binding antigen 140.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	595	Total 9863	C 3156	H 4886	N 850	O 935	S 36	0	1	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	303	ALA	SER	engineered mutation	UNP Q8WS31
A	469	ALA	THR	engineered mutation	UNP Q8WS31
A	727	ALA	SER	engineered mutation	UNP Q8WS31

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
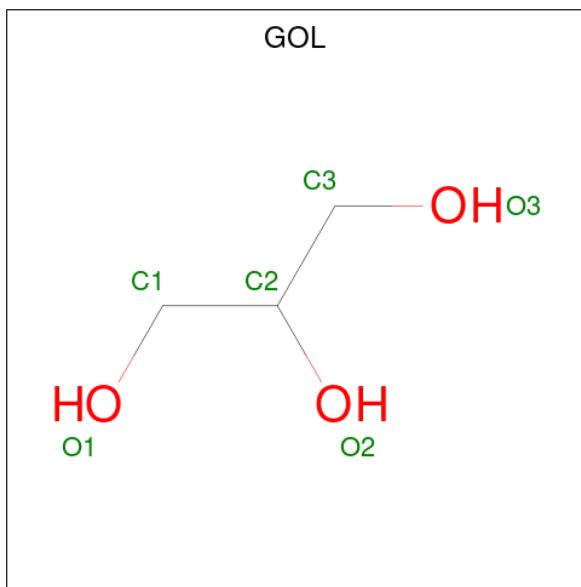
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 14	C 3	H 8	O 3	0	0
2	A	1	Total 14	C 3	H 8	O 3	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	143	Total 143	O 143	0	0

3 Residue-property plots ⓘ

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: Erythrocyte binding antigen 140
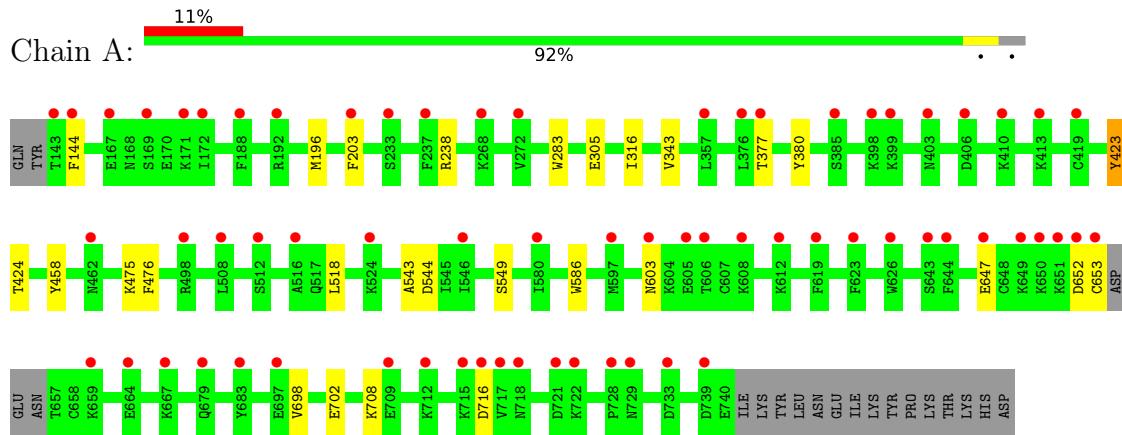
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	65.42Å 76.48Å 82.34Å 90.00° 96.80° 90.00°	Depositor
Resolution (Å)	19.75 – 2.40 19.75 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.6 (19.75-2.40) 99.6 (19.75-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.47 (at 2.41Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.200 , 0.237 0.205 , 0.236	Depositor DCC
R_{free} test set	1591 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	48.2	Xtriage
Anisotropy	0.088	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 52.5	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.51, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10216	wwPDB-VP
Average B, all atoms (Å^2)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.66% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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

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	0.24	0/5080	0.36	0/6793

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 ⓘ

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	4977	4886	4905	9	0
2	A	90	120	120	3	0
3	A	143	0	0	2	0
All	All	5210	5006	5025	11	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 (11) 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:423:TYR:HH	2:A:807:GOL:HO1	1.23	0.81

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:652:ASP:O	1:A:653:CYS:HB3	1.96	0.65
1:A:652:ASP:O	1:A:653:CYS:CB	2.50	0.58
2:A:809:GOL:O3	3:A:957:HOH:O	2.18	0.55
1:A:458:TYR:HA	1:A:518:LEU:HD22	1.90	0.54
1:A:543:ALA:HB2	1:A:586:TRP:CZ2	2.46	0.51
1:A:475:LYS:HD2	1:A:476:PHE:CE2	2.48	0.49
1:A:544:ASP:OD1	1:A:549:SER:OG	2.29	0.48
2:A:812:GOL:O3	3:A:1041:HOH:O	2.20	0.46
1:A:316:ILE:HD13	1:A:380:TYR:HA	2.00	0.44
1:A:238:ARG:HB2	1:A:283:TRP:CZ2	2.55	0.42

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	592/615 (96%)	576 (97%)	13 (2%)	3 (0%)	29	41

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	343	VAL
1	A	603	ASN
1	A	698	VAL

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	555/574 (97%)	544 (98%)	11 (2%)	55	74

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	144	PHE
1	A	196	MET
1	A	203	PHE
1	A	305	GLU
1	A	377	THR
1	A	423	TYR
1	A	424	THR
1	A	647	GLU
1	A	702	GLU
1	A	708	LYS
1	A	716	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 ⓘ

15 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	#$|Z| > 2$
2	GOL	A	804	-	5,5,5	0.38	0	5,5,5	0.34	0
2	GOL	A	806	-	5,5,5	0.35	0	5,5,5	0.23	0
2	GOL	A	810	-	5,5,5	0.38	0	5,5,5	0.28	0
2	GOL	A	808	-	5,5,5	0.36	0	5,5,5	0.27	0
2	GOL	A	813	-	5,5,5	0.38	0	5,5,5	0.28	0
2	GOL	A	809	-	5,5,5	0.37	0	5,5,5	0.26	0
2	GOL	A	811	-	5,5,5	0.36	0	5,5,5	0.33	0
2	GOL	A	801	-	5,5,5	0.35	0	5,5,5	0.13	0
2	GOL	A	805	-	5,5,5	0.39	0	5,5,5	0.19	0
2	GOL	A	812	-	5,5,5	0.36	0	5,5,5	0.23	0
2	GOL	A	807	-	5,5,5	0.36	0	5,5,5	0.35	0
2	GOL	A	802	-	5,5,5	0.38	0	5,5,5	0.18	0
2	GOL	A	803	-	5,5,5	0.39	0	5,5,5	0.23	0
2	GOL	A	815	-	5,5,5	0.37	0	5,5,5	0.27	0
2	GOL	A	814	-	5,5,5	0.39	0	5,5,5	0.19	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	GOL	A	804	-	-	2/4/4/4	-
2	GOL	A	806	-	-	0/4/4/4	-
2	GOL	A	810	-	-	2/4/4/4	-
2	GOL	A	808	-	-	0/4/4/4	-
2	GOL	A	813	-	-	2/4/4/4	-
2	GOL	A	809	-	-	2/4/4/4	-
2	GOL	A	811	-	-	2/4/4/4	-
2	GOL	A	801	-	-	2/4/4/4	-
2	GOL	A	805	-	-	1/4/4/4	-
2	GOL	A	812	-	-	2/4/4/4	-
2	GOL	A	807	-	-	0/4/4/4	-
2	GOL	A	802	-	-	2/4/4/4	-
2	GOL	A	803	-	-	4/4/4/4	-
2	GOL	A	815	-	-	2/4/4/4	-

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	814	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (25) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	801	GOL	O1-C1-C2-C3
2	A	802	GOL	O1-C1-C2-C3
2	A	804	GOL	O1-C1-C2-C3
2	A	812	GOL	O1-C1-C2-C3
2	A	813	GOL	O1-C1-C2-C3
2	A	804	GOL	O1-C1-C2-O2
2	A	813	GOL	O1-C1-C2-O2
2	A	803	GOL	O1-C1-C2-C3
2	A	803	GOL	C1-C2-C3-O3
2	A	805	GOL	O1-C1-C2-C3
2	A	809	GOL	O1-C1-C2-C3
2	A	810	GOL	O1-C1-C2-C3
2	A	811	GOL	O1-C1-C2-C3
2	A	814	GOL	O1-C1-C2-C3
2	A	815	GOL	O1-C1-C2-C3
2	A	801	GOL	O1-C1-C2-O2
2	A	812	GOL	O1-C1-C2-O2
2	A	802	GOL	O1-C1-C2-O2
2	A	810	GOL	O1-C1-C2-O2
2	A	811	GOL	O1-C1-C2-O2
2	A	814	GOL	O1-C1-C2-O2
2	A	815	GOL	O1-C1-C2-O2
2	A	809	GOL	O1-C1-C2-O2
2	A	803	GOL	O2-C2-C3-O3
2	A	803	GOL	O1-C1-C2-O2

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	809	GOL	1	0
2	A	812	GOL	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	807	GOL	1	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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	595/615 (96%)	0.63	67 (11%)	5	4	34, 60, 118, 171	0

All (67) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	729	ASN	7.4
1	A	143	THR	5.9
1	A	650	LYS	5.5
1	A	652	ASP	5.3
1	A	653	CYS	5.0
1	A	651	LYS	5.0
1	A	715	LYS	4.9
1	A	733	ASP	4.7
1	A	649	LYS	4.7
1	A	659	LYS	4.5
1	A	144	PHE	4.4
1	A	643	SER	4.4
1	A	410	LYS	4.3
1	A	406	ASP	4.2
1	A	644	PHE	4.2
1	A	623	PHE	4.1
1	A	721	ASP	3.7
1	A	619	PHE	3.7
1	A	546	ILE	3.5
1	A	171	LYS	3.5
1	A	647	GLU	3.4
1	A	399	LYS	3.3
1	A	718	ASN	3.3
1	A	605	GLU	3.2
1	A	683	TYR	3.2
1	A	739	ASP	3.2
1	A	603	ASN	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	722	LYS	3.1
1	A	728	PRO	3.1
1	A	357	LEU	3.1
1	A	608	LYS	3.0
1	A	606	THR	3.0
1	A	679	GLN	2.9
1	A	385	SER	2.8
1	A	398	LYS	2.8
1	A	272	VAL	2.8
1	A	192	ARG	2.6
1	A	712	LYS	2.6
1	A	233	SER	2.6
1	A	580	ILE	2.5
1	A	237	PHE	2.5
1	A	413	LYS	2.4
1	A	203	PHE	2.4
1	A	268	LYS	2.4
1	A	376	LEU	2.4
1	A	664	GLU	2.4
1	A	169	SER	2.3
1	A	667	LYS	2.3
1	A	716	ASP	2.3
1	A	172	ILE	2.2
1	A	403	ASN	2.2
1	A	167	GLU	2.2
1	A	597	MET	2.2
1	A	626	TRP	2.2
1	A	709	GLU	2.1
1	A	516	ALA	2.1
1	A	419	CYS	2.1
1	A	512	SER	2.1
1	A	462	ASN	2.1
1	A	498	ARG	2.1
1	A	508	LEU	2.1
1	A	717	VAL	2.1
1	A	524	LYS	2.0
1	A	612	LYS	2.0
1	A	377	THR	2.0
1	A	188	PHE	2.0
1	A	697	GLU	2.0

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

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

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, 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(Å^2)	Q<0.9
2	GOL	A	813	6/6	0.61	0.25	74,89,97,100	0
2	GOL	A	814	6/6	0.61	0.32	97,116,124,127	0
2	GOL	A	805	6/6	0.65	0.36	70,84,94,96	0
2	GOL	A	801	6/6	0.71	0.44	68,82,89,92	0
2	GOL	A	803	6/6	0.74	0.27	66,79,79,79	0
2	GOL	A	802	6/6	0.79	0.26	79,95,110,113	0
2	GOL	A	815	6/6	0.81	0.39	94,113,119,121	0
2	GOL	A	806	6/6	0.84	0.26	60,72,81,82	0
2	GOL	A	812	6/6	0.85	0.33	73,87,98,102	0
2	GOL	A	809	6/6	0.88	0.25	56,67,77,80	0
2	GOL	A	810	6/6	0.88	0.24	69,83,87,87	0
2	GOL	A	811	6/6	0.89	0.22	87,104,105,105	0
2	GOL	A	808	6/6	0.91	0.25	86,103,104,104	0
2	GOL	A	807	6/6	0.91	0.19	72,87,91,92	0
2	GOL	A	804	6/6	0.92	0.21	61,73,81,85	0

6.5　Other polymers ⓘ

There are no such residues in this entry.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK