wwPDB X-ray Structure Validation Summary Report (i)

May 21, 2020 – 09:17 am BST

PDB ID	:	5M02
Title	:	Crystal structure of murine P14 TCR / H-2Db with PF, modified gp33 peptide from LCMV
Authors	:	Achour, A.; Sandalova, T.; Sun, R.
Deposited on	:	2016-10-03
Resolution	:	1.75 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 following versions of software and data (see references (1)) were used in the production of this report:

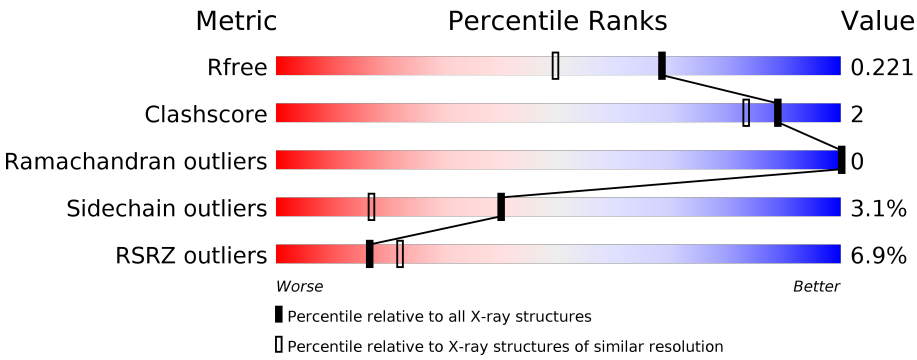
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.75 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	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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 on 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	276	3% 92% 7%
2	B	119	5% 77% 8% 15%
3	G	205	17% 84% 7% 9%
4	H	238	3% 94% 5%
5	P	9	78% 22%

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
6	GOL	H	301	-	-	-	X

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 6998 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 H-2 class I histocompatibility antigen, D-B alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	275	2315	1463	411	432	9	0	8	0

- Molecule 2 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	101	853	545	145	155	8	0	3	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-19	MET	-	initiating methionine	UNP P01887
B	-18	ALA	-	expression tag	UNP P01887
B	-17	ARG	-	expression tag	UNP P01887
B	-16	SER	-	expression tag	UNP P01887
B	-15	VAL	-	expression tag	UNP P01887
B	-14	THR	-	expression tag	UNP P01887
B	-13	LEU	-	expression tag	UNP P01887
B	-12	VAL	-	expression tag	UNP P01887
B	-11	PHE	-	expression tag	UNP P01887
B	-10	LEU	-	expression tag	UNP P01887
B	-9	VAL	-	expression tag	UNP P01887
B	-8	LEU	-	expression tag	UNP P01887
B	-7	VAL	-	expression tag	UNP P01887
B	-6	SER	-	expression tag	UNP P01887
B	-5	LEU	-	expression tag	UNP P01887
B	-4	THR	-	expression tag	UNP P01887
B	-3	GLY	-	expression tag	UNP P01887
B	-2	LEU	-	expression tag	UNP P01887
B	-1	MET	-	expression tag	UNP P01887
B	0	GLY	-	expression tag	UNP P01887

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	85	ASP	ALA	variant	UNP P01887

- Molecule 3 is a protein called Protein Trav14-1,T-cell receptor alpha chain C region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	G	187	1467	933	235	291	8	0	0	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	100	LEU	-	linker	UNP A0A0G2JF94
G	101	TYR	-	linker	UNP A0A0G2JF94
G	102	GLY	-	linker	UNP A0A0G2JF94
G	103	ASN	-	linker	UNP A0A0G2JF94
G	104	GLU	-	linker	UNP A0A0G2JF94
G	105	LYS	-	linker	UNP A0A0G2JF94
G	106	ILE	-	linker	UNP A0A0G2JF94
G	107	THR	-	linker	UNP A0A0G2JF94
G	108	PHE	-	linker	UNP A0A0G2JF94
G	109	GLY	-	linker	UNP A0A0G2JF94
G	110	ALA	-	linker	UNP A0A0G2JF94
G	111	GLY	-	linker	UNP A0A0G2JF94
G	112	THR	-	linker	UNP A0A0G2JF94
G	113	LYS	-	linker	UNP A0A0G2JF94
G	114	LEU	-	linker	UNP A0A0G2JF94
G	115	THR	-	linker	UNP A0A0G2JF94
G	116	ILE	-	linker	UNP A0A0G2JF94
G	117	LYS	-	linker	UNP A0A0G2JF94
G	118	PRO	-	linker	UNP A0A0G2JF94
G	119	ASN	-	linker	UNP A0A0G2JF94
G	166	CYS	THR	conflict	UNP P01849

- Molecule 4 is a protein called T-cell receptor beta chain V region C5,T-cell receptor beta-2 chain C region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	H	237	1881	1184	333	358	6	0	1	0

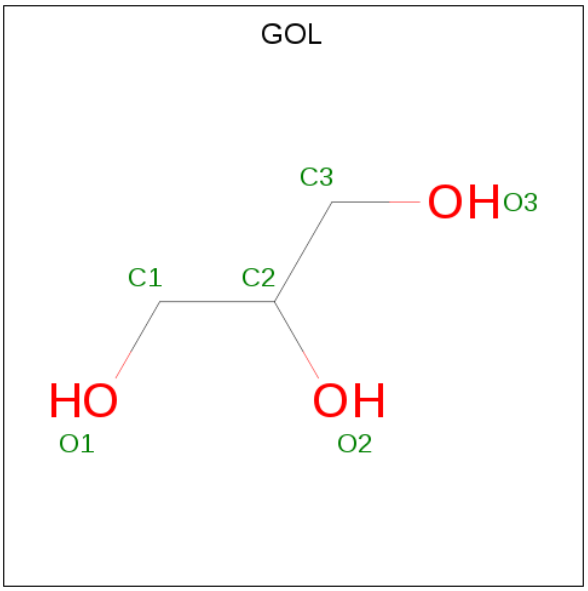
There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	93	ASP	-	linker	UNP P04213
H	94	ALA	-	linker	UNP P04213
H	95	GLY	-	linker	UNP P04213
H	96	GLY	-	linker	UNP P04213
H	97	ARG	-	linker	UNP P04213
H	98	ASN	-	linker	UNP P04213
H	99	THR	-	linker	UNP P04213
H	100	LEU	-	linker	UNP P04213
H	101	TYR	-	linker	UNP P04213
H	102	PHE	-	linker	UNP P04213
H	103	GLY	-	linker	UNP P04213
H	104	ALA	-	linker	UNP P04213
H	105	GLY	-	linker	UNP P04213
H	106	THR	-	linker	UNP P04213
H	107	ARG	-	linker	UNP P04213
H	108	LEU	-	linker	UNP P04213
H	109	SER	-	linker	UNP P04213
H	110	VAL	-	linker	UNP P04213
H	111	LEU	-	linker	UNP P04213
H	168	CYS	SER	conflict	UNP P01851
H	182	SER	CYS	conflict	UNP P01851

- Molecule 5 is a protein called LCMV-DERIVED GP33 ALTERED PEPTIDE LIGAND PF.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	P	9	72	48	11	12	1	0	0	0

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 7 is water.

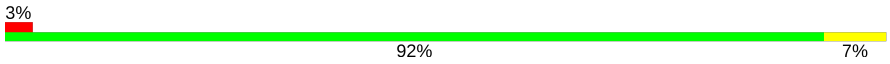
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	138	138	138	0	0
			Total	O		
7	B	56	56	56	0	0
			Total	O		
7	G	63	63	63	0	0
			Total	O		
7	H	131	131	131	0	0
			Total	O		
7	P	4	4	4	0	0

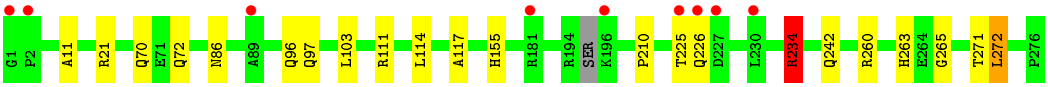
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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: H-2 class I histocompatibility antigen, D-B alpha chain

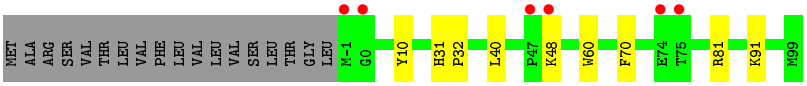
Chain A:





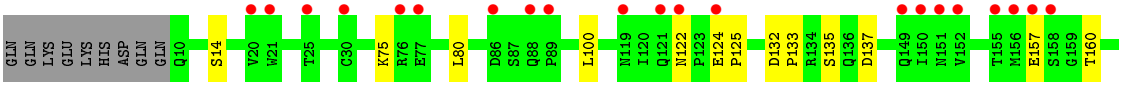
- Molecule 2: Beta-2-microglobulin

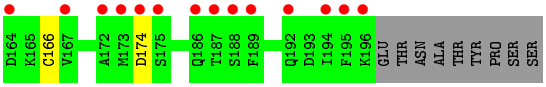
Chain B:



- Molecule 3: Protein Trav14-1,T-cell receptor alpha chain C region

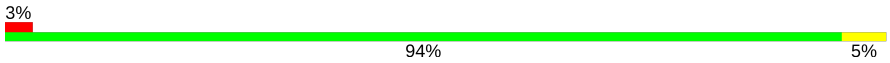
Chain G:

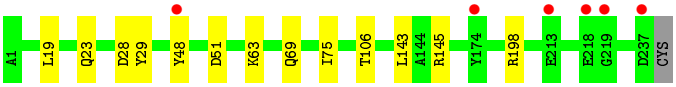




- Molecule 4: T-cell receptor beta chain V region C5,T-cell receptor beta-2 chain C region

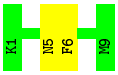
Chain H:





- Molecule 5: LCMV-DERIVED GP33 ALTERED PEPTIDE LIGAND PF

Chain P:


K1
N5
F6
M9

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	255.11Å 46.50Å 89.02Å 90.00° 94.60° 90.00°	Depositor
Resolution (Å)	127.14 – 1.75 53.76 – 1.75	Depositor EDS
% Data completeness (in resolution range)	98.1 (127.14-1.75) 98.1 (53.76-1.75)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.50 (at 1.75Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.188 , 0.214 0.197 , 0.221	Depositor DCC
R_{free} test set	5198 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	27.4	Xtriage
Anisotropy	0.326	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 41.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	6998	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.58% 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 (i)

5.1 Standard geometry (i)

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	0.61	0/2405	0.98	7/3259 (0.2%)
2	B	0.54	0/888	0.74	0/1200
3	G	0.67	2/1502 (0.1%)	0.78	4/2037 (0.2%)
4	H	0.57	0/1937	0.76	2/2633 (0.1%)
5	P	0.56	0/74	0.74	0/97
All	All	0.60	2/6806 (0.0%)	0.85	13/9226 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	133	PRO	N-CD	5.24	1.55	1.47
3	G	125	PRO	N-CD	5.11	1.54	1.47

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	234	ARG	NE-CZ-NH1	23.91	132.26	120.30
1	A	234	ARG	NE-CZ-NH2	-20.27	110.17	120.30
1	A	272	LEU	CA-CB-CG	7.82	133.29	115.30
1	A	234	ARG	CD-NE-CZ	7.51	134.11	123.60
1	A	272	LEU	CB-CG-CD1	6.33	121.75	111.00

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	2315	0	2206	19	0
2	B	853	0	842	6	0
3	G	1467	0	1412	0	0
4	H	1881	0	1792	5	0
5	P	72	0	72	4	0
6	A	12	0	16	0	0
6	H	6	0	8	0	0
7	A	138	0	0	0	0
7	B	56	0	0	0	0
7	G	63	0	0	0	0
7	H	131	0	0	0	0
7	P	4	0	0	0	0
All	All	6998	0	6348	25	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 2.

The worst 5 of 25 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:234:ARG:HD3	2:B:10:TYR:CE2	2.21	0.75
1:A:111[B]:ARG:HG2	1:A:111[B]:ARG:HH11	1.55	0.71
1:A:97:GLN:HE22	5:P:5:ASN:HD21	1.40	0.67
1:A:96:GLN:OE1	2:B:31:HIS:HE1	1.84	0.61
1:A:234:ARG:HD3	2:B:10:TYR:CZ	2.36	0.60

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	279/276 (101%)	274 (98%)	5 (2%)	0	100	100
2	B	102/119 (86%)	101 (99%)	1 (1%)	0	100	100
3	G	185/205 (90%)	178 (96%)	7 (4%)	0	100	100
4	H	236/238 (99%)	233 (99%)	3 (1%)	0	100	100
5	P	7/9 (78%)	6 (86%)	1 (14%)	0	100	100
All	All	809/847 (96%)	792 (98%)	17 (2%)	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	241/234 (103%)	235 (98%)	6 (2%)	47	25
2	B	98/111 (88%)	95 (97%)	3 (3%)	40	17
3	G	166/184 (90%)	157 (95%)	9 (5%)	22	5
4	H	204/204 (100%)	200 (98%)	4 (2%)	55	34
5	P	7/7 (100%)	7 (100%)	0	100	100
All	All	716/740 (97%)	694 (97%)	22 (3%)	40	17

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	14	SER
3	G	135	SER
4	H	51	ASP
3	G	75	LYS
3	G	100	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	31	HIS
3	G	13	GLN
4	H	39	HIS
2	B	17	ASN
4	H	22	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 carbohydrates 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	$\#\|Z\| > 2$
6	GOL	A	301	-	5,5,5	0.31	0	5,5,5	0.11	0
6	GOL	H	301	-	5,5,5	0.30	0	5,5,5	0.39	0
6	GOL	A	302	-	5,5,5	0.40	0	5,5,5	0.22	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
6	GOL	A	301	-	-	2/4/4/4	-
6	GOL	H	301	-	-	3/4/4/4	-
6	GOL	A	302	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	301	GOL	O1-C1-C2-C3
6	H	301	GOL	O2-C2-C3-O3
6	H	301	GOL	C1-C2-C3-O3
6	A	301	GOL	O1-C1-C2-O2
6	H	301	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 (i)

6.1 Protein, DNA and RNA chains (i)

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	275/276 (99%)	0.12	9 (3%) 46 53	18, 28, 47, 70	0
2	B	101/119 (84%)	0.08	6 (5%) 22 27	19, 33, 49, 60	0
3	G	187/205 (91%)	1.01	35 (18%) 1 1	24, 44, 84, 113	0
4	H	237/238 (99%)	0.09	6 (2%) 57 63	23, 31, 50, 68	0
5	P	9/9 (100%)	-0.19	0 100 100	20, 21, 25, 25	0
All	All	809/847 (95%)	0.31	56 (6%) 16 22	18, 32, 62, 113	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	G	187	THR	14.8
3	G	173	MET	12.7
3	G	172	ALA	11.7
3	G	189	PHE	10.8
2	B	-1	MET	6.3

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 carbohydrates 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(Å^2)	Q<0.9
6	GOL	H	301	6/6	0.54	0.41	64,67,68,72	0
6	GOL	A	301	6/6	0.72	0.26	61,63,65,66	0
6	GOL	A	302	6/6	0.81	0.24	31,36,43,49	0

6.5 Other polymers (i)

There are no such residues in this entry.