wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 15, 2020 – 12:17 am BST

PDB ID	:	3HG1
Title	:	Germline-governed recognition of a cancer epitope by an immunodominant human T cell receptor
Authors	:	Cole, D.K.; Yuan, F.; Rizkallah, P.J.; Miles, J.J.; Gostick, E.; Price, D.A.; Gao, G.F.; Jakobsen, B.K.; Sewell, A.K.
Deposited on	:	2009-05-13
Resolution	:	3.00 Å(reported)

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

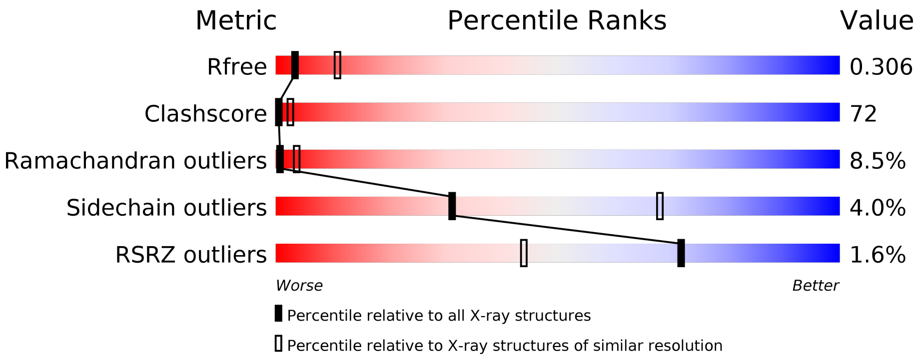
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

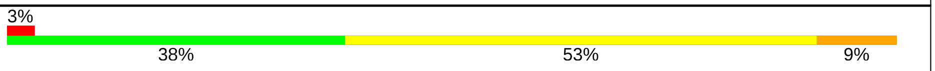
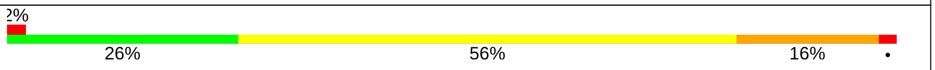
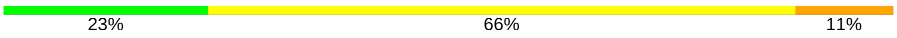
The reported resolution of this entry is 3.00 Å.

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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	
2	B	100	
3	C	10	
4	D	194	
5	E	244	

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 6649 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 MHC class I antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	276	2253	1408	410	426	9	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	100	837	533	141	159	4	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	INITIATING METHIONINE	UNP P61769

- Molecule 3 is a protein called CANCER/MART-1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	10	69	45	10	14	0	0	0

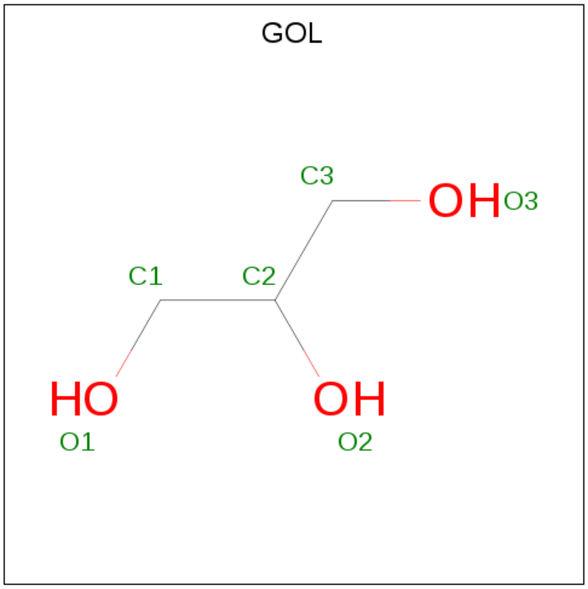
- Molecule 4 is a protein called T-CELL RECEPTOR, ALPHA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	194	1498	927	249	314	8	0	0	0

- Molecule 5 is a protein called T-cell Receptor, Beta Chain.

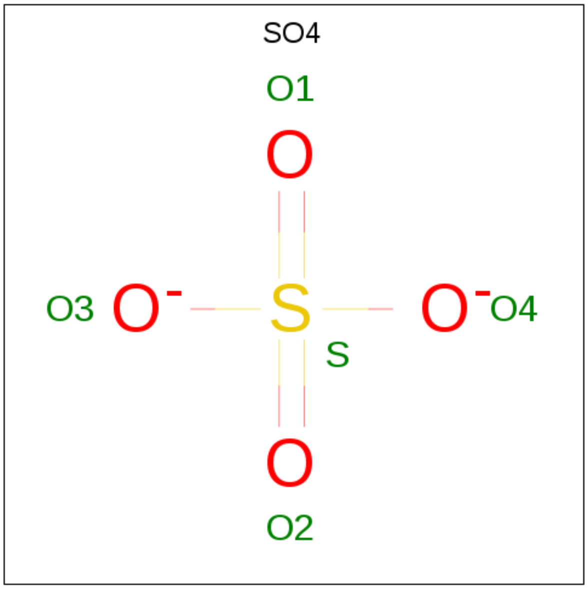
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	244	1926	1221	330	370	5	0	0	0

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



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

- Molecule 7 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	1	Total O S 5 4 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	1	Total O S 5 4 1	0	0

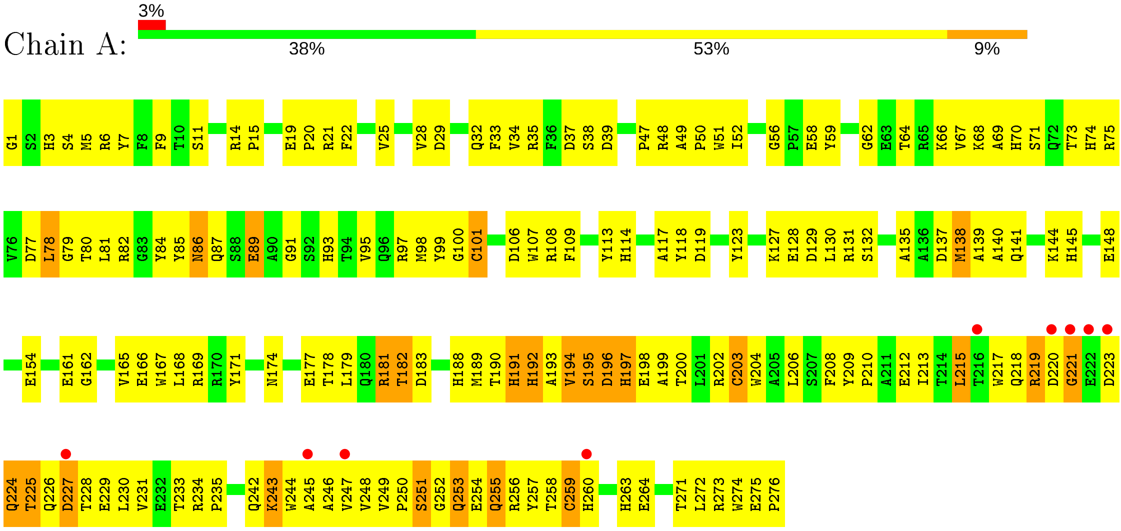
- Molecule 8 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	17	Total O 17 17	0	0
8	B	6	Total O 6 6	0	0
8	C	1	Total O 1 1	0	0
8	D	8	Total O 8 8	0	0
8	E	12	Total O 12 12	0	0

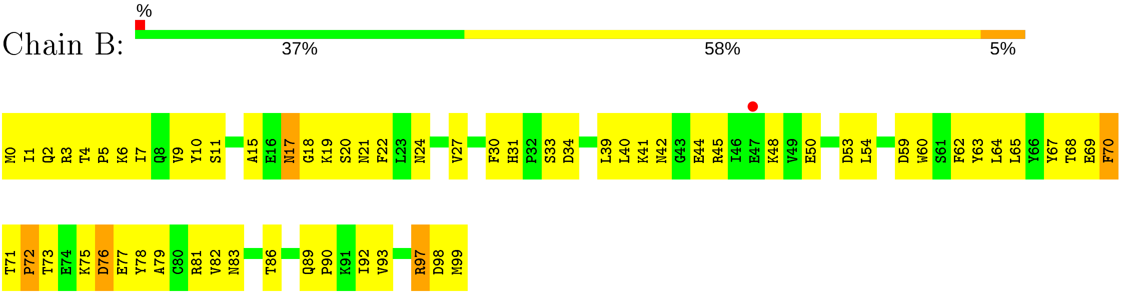
3 Residue-property plots

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.

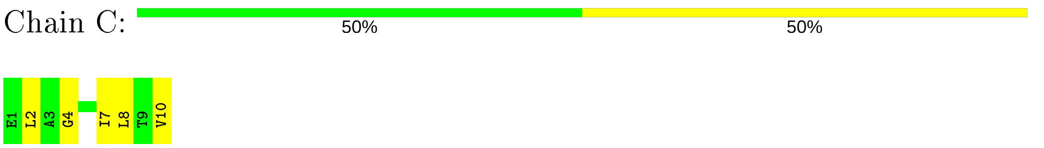
- Molecule 1: MHC class I antigen



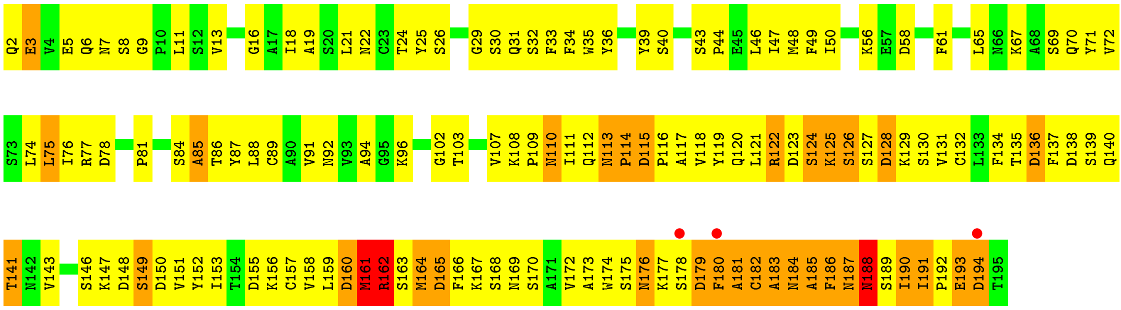
- Molecule 2: Beta-2-microglobulin



- Molecule 3: CANCER/MART-1

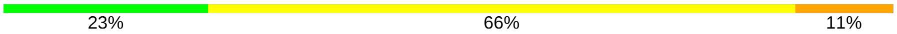


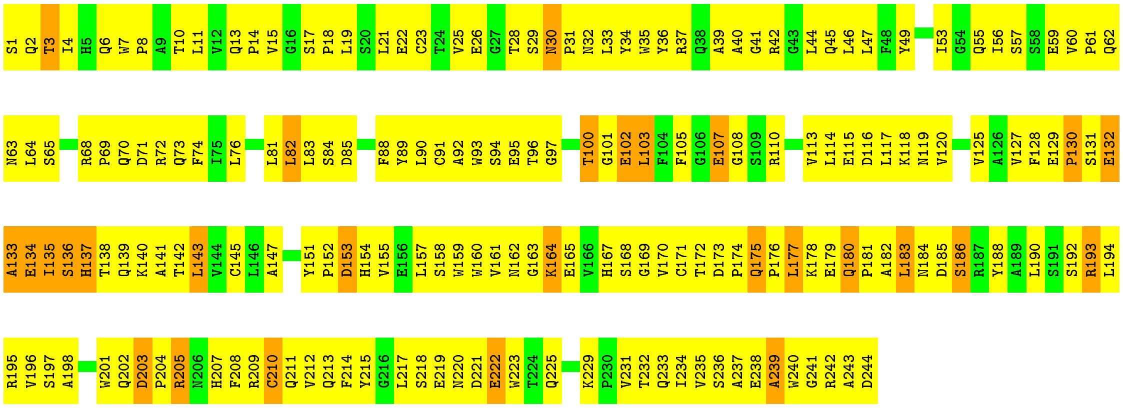
- Molecule 4: T-CELL RECEPTOR, ALPHA CHAIN



- Molecule 5: T-cell Receptor, Beta Chain

Chain E:





4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	120.90Å 120.90Å 81.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.65 – 3.00 48.65 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.65-3.00) 99.5 (48.65-3.00)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	0.16	Depositor
$< I/\sigma(I) >$ [1]	1.91 (at 3.01Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.225 , 0.304 0.233 , 0.306	Depositor DCC
R_{free} test set	1211 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	54.9	Xtriage
Anisotropy	0.349	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 58.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.055 for h,-k,-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	6649	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.17% 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, SO4

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	0.69	2/2319 (0.1%)	0.49	0/3149
2	B	0.67	0/860	0.51	0/1162
3	C	0.70	0/68	0.50	0/90
4	D	0.75	1/1527 (0.1%)	0.54	0/2070
5	E	0.73	2/1979 (0.1%)	0.55	0/2698
All	All	0.71	5/6753 (0.1%)	0.52	0/9169

All (5) bond length outliers are listed below:

Mol	**Chain**	**Res**	**Type**	**Atoms**	**Z**	**Observed(Å)**	**Ideal(Å)**
1	A	101	CYS	CB-SG	-10.60	1.64	1.82
5	E	210	CYS	CB-SG	-5.69	1.72	1.81
1	A	89	GLU	CG-CD	5.57	1.60	1.51
4	D	89	CYS	CB-SG	-5.33	1.73	1.81
5	E	145	CYS	CB-SG	-5.33	1.73	1.81

There are no bond angle outliers.

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	2253	0	2101	274	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	837	0	801	81	0
3	C	69	0	79	10	0
4	D	1498	0	1408	275	1
5	E	1926	0	1845	323	0
6	A	6	0	8	0	0
6	E	6	0	8	1	0
7	A	10	0	0	1	0
8	A	17	0	0	0	0
8	B	6	0	0	1	0
8	C	1	0	0	0	0
8	D	8	0	0	1	0
8	E	12	0	0	6	0
All	All	6649	0	6250	922	1

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 72.

The worst 5 of 922 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)
4:D:161:MET:O	4:D:162:ARG:CD	1.63	1.45
4:D:184:ASN:ND2	4:D:185:ALA:H	0.95	1.44
4:D:184:ASN:HD22	4:D:185:ALA:N	1.12	1.41
4:D:147:LYS:HD3	4:D:148:ASP:N	1.40	1.34
1:A:215:LEU:HD12	1:A:215:LEU:O	1.31	1.23

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:D:140:GLN:OE1	4:D:176:ASN:O[4_555]	2.04	0.16

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	274/276 (99%)	225 (82%)	29 (11%)	20 (7%)	1	5
2	B	98/100 (98%)	80 (82%)	14 (14%)	4 (4%)	3	16
3	C	8/10 (80%)	7 (88%)	1 (12%)	0	100	100
4	D	192/194 (99%)	147 (77%)	19 (10%)	26 (14%)	0	1
5	E	242/244 (99%)	189 (78%)	34 (14%)	19 (8%)	1	4
All	All	814/824 (99%)	648 (80%)	97 (12%)	69 (8%)	1	4

5 of 69 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	182	THR
1	A	192	HIS
1	A	196	ASP
1	A	221	GLY
1	A	224	GLN

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	232/232 (100%)	227 (98%)	5 (2%)	52	81
2	B	95/95 (100%)	92 (97%)	3 (3%)	39	74
3	C	7/7 (100%)	7 (100%)	0	100	100
4	D	172/172 (100%)	160 (93%)	12 (7%)	15	47
5	E	211/211 (100%)	202 (96%)	9 (4%)	29	66
All	All	717/717 (100%)	688 (96%)	29 (4%)	31	68

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	161	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	D	184	ASN
5	E	175	GLN
4	D	162	ARG
4	D	188	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	2	GLN
4	D	110	ASN
5	E	180	GLN
4	D	6	GLN
4	D	22	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

4 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	E	802	-	5,5,5	0.55	0	5,5,5	0.46	0
6	GOL	A	801	-	5,5,5	0.54	0	5,5,5	0.64	0
7	SO4	A	803	-	4,4,4	0.13	0	6,6,6	0.08	0
7	SO4	A	804	-	4,4,4	0.25	0	6,6,6	0.36	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	E	802	-	-	2/4/4/4	-
6	GOL	A	801	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	E	802	GOL	O1-C1-C2-C3
6	E	802	GOL	O1-C1-C2-O2
6	A	801	GOL	O1-C1-C2-O2
6	A	801	GOL	O1-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	E	802	GOL	1	0
7	A	804	SO4	1	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, 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	276/276 (100%)	-0.02	9 (3%) 46 20	22, 38, 70, 107	0
2	B	100/100 (100%)	-0.16	1 (1%) 82 59	19, 40, 57, 70	0
3	C	10/10 (100%)	-0.23	0 100 100	28, 31, 38, 45	0
4	D	194/194 (100%)	-0.07	3 (1%) 73 46	21, 38, 70, 92	0
5	E	244/244 (100%)	-0.22	0 100 100	24, 40, 62, 101	0
All	All	824/824 (100%)	-0.11	13 (1%) 72 44	19, 39, 68, 107	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	221	GLY	3.4
1	A	260	HIS	3.3
1	A	227	ASP	3.0
1	A	247	VAL	2.9
1	A	222	GLU	2.7

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 carbohydrates 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
6	GOL	A	801	6/6	0.79	0.24	71,79,81,82	0
7	SO4	A	804	5/5	0.79	0.38	60,60,60,60	5
6	GOL	E	802	6/6	0.88	0.20	47,58,59,59	0
7	SO4	A	803	5/5	0.94	0.36	61,62,65,67	5

6.5 Other polymers ⓘ

There are no such residues in this entry.