WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2023 – 03:42 PM EST

PDB ID	:	4PJF
Title	:	Structure of human MR1-Ac-6-FP in complex with human MAIT B-C10 TCR
Authors	:	Birkinshaw, R.W.; Rossjohn, J.
Deposited on	:	2014-05-12
Resolution	:	2.45 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	:	2.36
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)	:	2.36

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 2.45 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		0.214
Clashscore		1
Ramachandran outliers		0.1%
Sidechain outliers		2.2%
RSRZ outliers		2.1%

Worse ... *Better*

▮ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
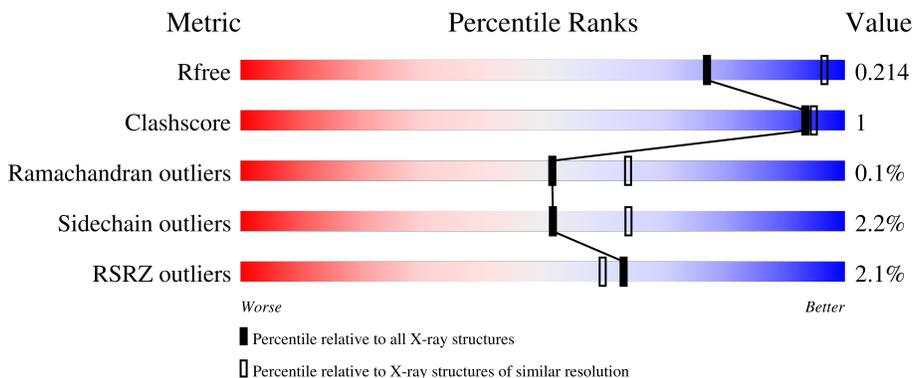
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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	271	% — 91% — 6% ·
1	C	271	3% — 87% — 7% 5%
2	B	100	% — 96% — · ·
2	D	100	7% — 93% — · ·
3	E	205	91% — 5% ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	205	3% 85% 7% 7%
4	F	246	92% 5%
4	H	246	4% 95%
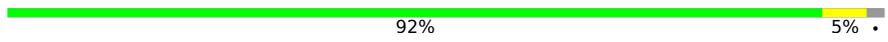
2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 13209 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 Major histocompatibility complex class I-related gene protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	263	Total 2141	C 1371	N 372	O 388	S 10	0	0	0
1	C	257	Total 2065	C 1320	N 360	O 375	S 10	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP Q95460
A	261	SER	CYS	engineered mutation	UNP Q95460
C	0	MET	-	initiating methionine	UNP Q95460
C	261	SER	CYS	engineered mutation	UNP Q95460

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	99	Total 792	C 506	N 136	O 147	S 3	0	0	0
2	D	97	Total 762	C 491	N 125	O 144	S 2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P61769
D	0	MET	-	initiating methionine	UNP P61769

- Molecule 3 is a protein called TCR-alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	197	Total 1491	C 947	N 235	O 300	S 9	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	190	Total 1407	C 902	N 221	O 275	S 9	0	0	0

- Molecule 4 is a protein called TCR-beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	240	Total 1842	C 1163	N 316	O 354	S 9	0	0	0
4	H	240	Total 1831	C 1153	N 320	O 349	S 9	0	0	0

- Molecule 5 is N-(6-formyl-4-oxo-3,4-dihydropteridin-2-yl)acetamide (three-letter code: 30W) (formula: $C_9H_7N_5O_3$).
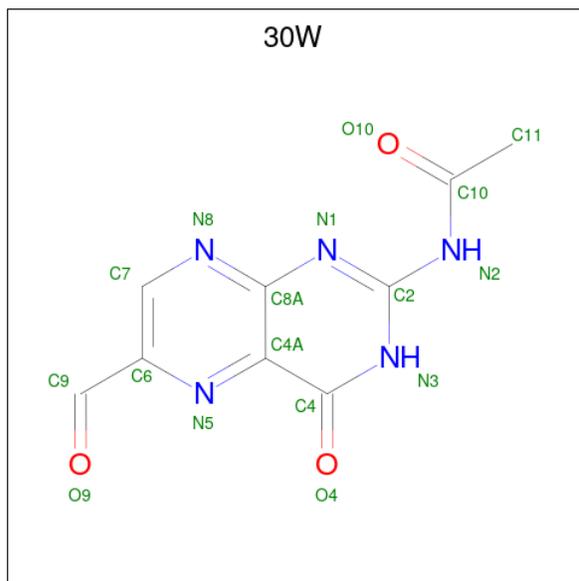
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total 16	C 9	N 5	O 2	0	0
5	C	1	Total 16	C 9	N 5	O 2	0	0

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

GOL

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

- Molecule 7 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	160	Total 160	O 160	0	0
7	B	62	Total 62	O 62	0	0
7	C	135	Total 135	O 135	0	0
7	D	29	Total 29	O 29	0	0
7	E	91	Total 91	O 91	0	0
7	F	147	Total 147	O 147	0	0
7	G	78	Total 78	O 78	0	0
7	H	132	Total 132	O 132	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: Major histocompatibility complex class I-related gene protein

Chain A:

• Molecule 1: Major histocompatibility complex class I-related gene protein

Chain C:

• Molecule 2: Beta-2-microglobulin

Chain B:

• Molecule 2: Beta-2-microglobulin

Chain D:

• Molecule 3: TCR-alpha

Chain E:

HIS MET G1 T26 L32 F45 L51 S93 L121 D138 C157 V158 L159 M164 S178 ASP F180 P198 SER PRO GLU SER SER

• Molecule 3: TCR-alpha

Chain G:

3%

85%   7%   7%

HIS MET G1 T26 S27 G28 L32 L51 L64 S65 R66 E77 S93 D123 SER LYS SER SER ASP K129 S130 V131 D138 S146 K147 D148 S149 D150 V151 C157 S163 W174 S175 ASN LYS SER D179 F180 A181 A185 P198 SER PRO GLU SER SER

• Molecule 4: TCR-beta

Chain F:

92%   5% •

HIS MET ASN ALA G3 L13 M19 D26 E79 A100 N101 V102 L111 V127 D153 C171 P176 R193 R205 A237 R242 ALA ASP

• Molecule 4: TCR-beta

Chain H:

4%

95%   • •

HIS MET ASN ALA G3 K14 V127 D153 H154 V161 R193 W201 Q202 N203 P204 R205 N206 Y215 N220 D226 A237 E238 A239 W240 G241 R242 ALA ASP
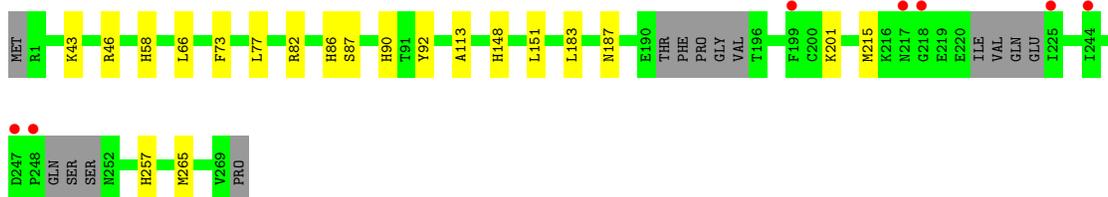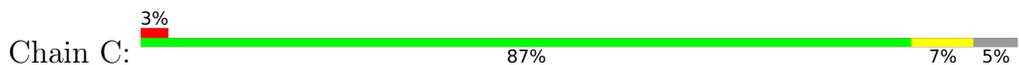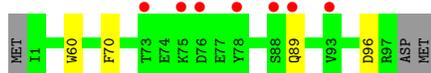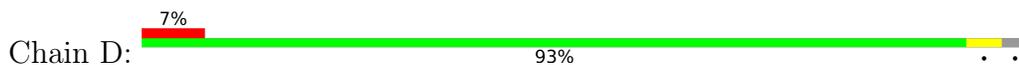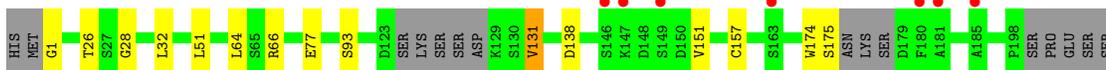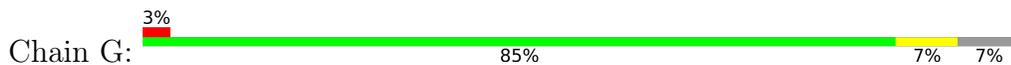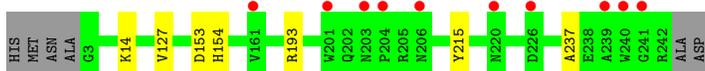
4   Data and refinement statistics   ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	218.99Å 71.70Å 144.03Å 90.00° 104.85° 90.00°	Depositor
Resolution (Å)	35.87 – 2.45 35.69 – 2.45	Depositor EDS
% Data completeness (in resolution range)	99.0 (35.87-2.45) 99.0 (35.69-2.45)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.16 (at 2.45Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.178 , 0.212 0.180 , 0.214	Depositor DCC
R_{free} test set	3969 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	42.7	Xtriage
Anisotropy	0.460	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 52.5	EDS
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	13209	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.58% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 30W, 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	0.47	0/2206	0.67	0/3001
1	C	0.45	0/2126	0.66	0/2892
2	B	0.44	0/815	0.64	0/1111
2	D	0.40	0/785	0.64	0/1074
3	E	0.46	0/1524	0.69	0/2073
3	G	0.46	0/1438	0.67	0/1958
4	F	0.44	0/1891	0.70	0/2582
4	H	0.46	0/1879	0.69	0/2568
All	All	0.45	0/12664	0.67	0/17259

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	2141	0	2016	7	0
1	C	2065	0	1901	11	0
2	B	792	0	730	2	0
2	D	762	0	680	1	0
3	E	1491	0	1358	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	1407	0	1257	6	0
4	F	1842	0	1705	8	0
4	H	1831	0	1691	2	0
5	A	16	0	6	0	0
5	C	16	0	6	0	0
6	B	12	0	16	1	0
7	A	160	0	0	0	0
7	B	62	0	0	0	0
7	C	135	0	0	0	0
7	D	29	0	0	0	0
7	E	91	0	0	1	0
7	F	147	0	0	2	0
7	G	78	0	0	0	0
7	H	132	0	0	0	0
All	All	13209	0	11366	34	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.

The worst 5 of 34 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:F:13:LEU:HD21	4:F:19:MET:HG3	1.59	0.84
3:G:131:VAL:HG23	3:G:174:TRP:HB3	1.75	0.68
1:A:77:LEU:HD13	1:A:92:TYR:HB2	1.86	0.57
3:G:1:GLY:HA3	3:G:26:THR:HA	1.88	0.56
2:B:96:ASP:HB3	6:B:102:GOL:H12	1.87	0.55

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	259/271 (96%)	252 (97%)	7 (3%)	0	100	100
1	C	249/271 (92%)	244 (98%)	5 (2%)	0	100	100
2	B	97/100 (97%)	96 (99%)	1 (1%)	0	100	100
2	D	95/100 (95%)	93 (98%)	1 (1%)	1 (1%)	14	14
3	E	193/205 (94%)	192 (100%)	1 (0%)	0	100	100
3	G	184/205 (90%)	182 (99%)	2 (1%)	0	100	100
4	F	238/246 (97%)	233 (98%)	5 (2%)	0	100	100
4	H	238/246 (97%)	235 (99%)	2 (1%)	1 (0%)	34	41
All	All	1553/1644 (94%)	1527 (98%)	24 (2%)	2 (0%)	51	64

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	96	ASP
4	H	153	ASP

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	224/241 (93%)	219 (98%)	5 (2%)	52	64
1	C	209/241 (87%)	205 (98%)	4 (2%)	57	69
2	B	85/95 (90%)	84 (99%)	1 (1%)	71	81
2	D	78/95 (82%)	76 (97%)	2 (3%)	46	58
3	E	156/181 (86%)	151 (97%)	5 (3%)	39	50
3	G	137/181 (76%)	132 (96%)	5 (4%)	35	46
4	F	193/213 (91%)	189 (98%)	4 (2%)	53	66
4	H	190/213 (89%)	188 (99%)	2 (1%)	73	82
All	All	1272/1460 (87%)	1244 (98%)	28 (2%)	52	64

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	E	121	LEU
4	H	193	ARG
4	F	19	MET
3	G	138	ASP
3	E	157	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	187	ASN
3	G	169	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 monosaccharides 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	#$|Z| > 2$
5	30W	A	301	1	16,17,18	0.71	1 (6%)	18,24,25	2.20	3 (16%)
5	30W	C	301	1	16,17,18	0.78	1 (6%)	18,24,25	2.34	3 (16%)

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

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#|Z| > 2	Counts	RMSZ	#|Z| > 2
6	GOL	B	102	-	5,5,5	0.15	0	5,5,5	0.51	0
6	GOL	B	101	-	5,5,5	0.10	0	5,5,5	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
5	30W	A	301	1	-	0/4/4/6	0/2/2/2
5	30W	C	301	1	-	0/4/4/6	0/2/2/2
6	GOL	B	102	-	-	0/4/4/4	-
6	GOL	B	101	-	-	0/4/4/4	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	301	30W	C2-N2	-2.20	1.33	1.37
5	C	301	30W	C4A-C8A	2.01	1.44	1.40

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	C	301	30W	N2-C2-N3	8.34	130.71	117.98
5	A	301	30W	N2-C2-N3	7.48	129.40	117.98
5	A	301	30W	N2-C2-N1	-4.10	112.15	117.72
5	C	301	30W	N2-C2-N1	-3.93	112.38	117.72
5	C	301	30W	C4A-C8A-N1	-2.45	119.67	123.18

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	B	102	GOL	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(Å²)	Q<0.9
1	A	263/271 (97%)	-0.19	2 (0%)	86	86	28, 40, 66, 88	0
1	C	257/271 (94%)	-0.19	7 (2%)	54	50	29, 44, 76, 103	0
2	B	99/100 (99%)	-0.13	1 (1%)	82	83	31, 47, 72, 82	0
2	D	97/100 (97%)	0.26	7 (7%)	15	11	36, 65, 96, 109	0
3	E	197/205 (96%)	-0.14	0	100	100	30, 45, 73, 83	0
3	G	190/205 (92%)	-0.03	7 (3%)	41	38	25, 49, 97, 130	0
4	F	240/246 (97%)	-0.33	0	100	100	32, 43, 66, 94	0
4	H	240/246 (97%)	-0.08	10 (4%)	36	33	26, 45, 87, 128	0
All	All	1583/1644 (96%)	-0.14	34 (2%)	63	60	25, 45, 82, 130	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	0	MET	4.7
1	A	222	VAL	4.6
4	H	204	PRO	3.8
4	H	239	ALA	3.5
1	C	218	GLY	3.3

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
6	GOL	B	102	6/6	0.90	0.20	52,60,62,62	0
6	GOL	B	101	6/6	0.97	0.19	35,38,43,45	0
5	30W	C	301	16/17	0.98	0.14	29,35,37,37	0
5	30W	A	301	16/17	0.99	0.20	29,31,39,40	0

6.5 Other polymers ⓘ

There are no such residues in this entry.