wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 03:18 pm BST

PDB ID	:	5WKH
Title	:	D30 TCR in complex with HLA-A*11:01-GTS3
Authors	:	Gras, S.; Rossjohn, J.
Deposited on	:	2017-07-25
Resolution	:	3.20 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:

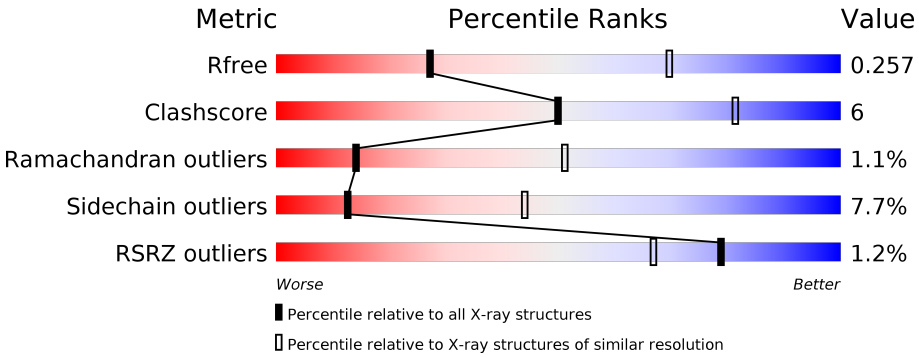
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.20 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.



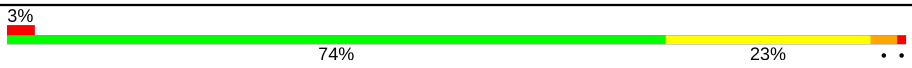
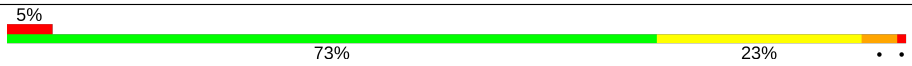
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	274	86% 13%
1	F	274	% 83% 16%
2	B	100	79% 19%
2	G	100	% 87% 13%
3	C	10	80% 20%
3	H	10	70% 30%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	198	
4	I	198	
5	E	244	
5	J	244	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 13218 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 HLA class I histocompatibility antigen, A-11 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	274	Total	C	N	O	S	0	0	0
			2235	1389	407	430	9			
1	F	274	Total	C	N	O	S	0	0	0
			2235	1389	407	430	9			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	99	Total	C	N	O	S	0	0	0
			829	528	140	158	3			
2	G	100	Total	C	N	O	S	0	0	0
			837	533	141	159	4			

There are 2 discrepancies between the modelled and reference sequences:

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

- Molecule 3 is a protein called GTS3 peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	10	Total	C	N	O	0	0	0
			70	41	14	15			
3	H	10	Total	C	N	O	0	0	0
			70	41	14	15			

- Molecule 4 is a protein called T-cell receptor alpha variable 30,T-cell receptor, sp3.4 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	198	Total	C	N	O	S	0	0	0
			1535	961	259	305	10			
4	I	198	Total	C	N	O	S	0	0	0
			1535	961	259	305	10			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	51	ILE	VAL	ENGINEERED MUTATION	UNP A0A087WSZ9
D	102	LEU	-	linker	UNP A0A087WSZ9
D	103	GLY	-	linker	UNP A0A087WSZ9
D	105	ASP	-	linker	UNP A0A087WSZ9
D	106	ALA	-	linker	UNP A0A087WSZ9
D	107	GLY	-	linker	UNP A0A087WSZ9
D	108	ASN	-	linker	UNP A0A087WSZ9
D	109	MET	-	linker	UNP A0A087WSZ9
D	110	LEU	-	linker	UNP A0A087WSZ9
D	111	THR	-	linker	UNP A0A087WSZ9
D	112	PHE	-	linker	UNP A0A087WSZ9
D	113	GLY	-	linker	UNP A0A087WSZ9
D	114	GLY	-	linker	UNP A0A087WSZ9
D	115	GLY	-	linker	UNP A0A087WSZ9
D	116	THR	-	linker	UNP A0A087WSZ9
D	117	ARG	-	linker	UNP A0A087WSZ9
D	118	LEU	-	linker	UNP A0A087WSZ9
D	119	MET	-	linker	UNP A0A087WSZ9
D	120	VAL	-	linker	UNP A0A087WSZ9
D	121	LYS	-	linker	UNP A0A087WSZ9
D	122	PRO	-	linker	UNP A0A087WSZ9
D	123	HIS	-	linker	UNP A0A087WSZ9
I	51	ILE	VAL	ENGINEERED MUTATION	UNP A0A087WSZ9
I	102	LEU	-	linker	UNP A0A087WSZ9
I	103	GLY	-	linker	UNP A0A087WSZ9
I	105	ASP	-	linker	UNP A0A087WSZ9
I	106	ALA	-	linker	UNP A0A087WSZ9
I	107	GLY	-	linker	UNP A0A087WSZ9
I	108	ASN	-	linker	UNP A0A087WSZ9
I	109	MET	-	linker	UNP A0A087WSZ9
I	110	LEU	-	linker	UNP A0A087WSZ9
I	111	THR	-	linker	UNP A0A087WSZ9
I	112	PHE	-	linker	UNP A0A087WSZ9
I	113	GLY	-	linker	UNP A0A087WSZ9
I	114	GLY	-	linker	UNP A0A087WSZ9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
I	115	GLY	-	linker	UNP A0A087WSZ9
I	116	THR	-	linker	UNP A0A087WSZ9
I	117	ARG	-	linker	UNP A0A087WSZ9
I	118	LEU	-	linker	UNP A0A087WSZ9
I	119	MET	-	linker	UNP A0A087WSZ9
I	120	VAL	-	linker	UNP A0A087WSZ9
I	121	LYS	-	linker	UNP A0A087WSZ9
I	122	PRO	-	linker	UNP A0A087WSZ9
I	123	HIS	-	linker	UNP A0A087WSZ9

- Molecule 5 is a protein called D30 TCR beta chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	244	Total	C	N	O	S	0	0	0
			1935	1229	336	365	5			
5	J	244	Total	C	N	O	S	0	0	0
			1935	1229	336	365	5			

- Molecule 6 is water.

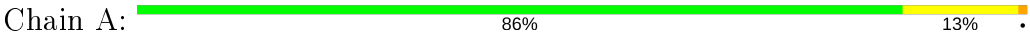
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	E	1	Total	O	0	0
			1	1		
6	F	1	Total	O	0	0
			1	1		

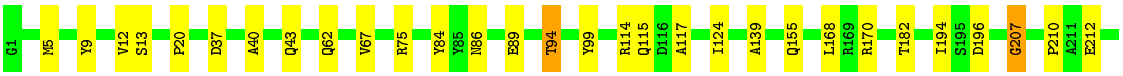
3 Residue-property plots (i)

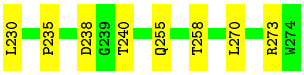
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: HLA class I histocompatibility antigen, A-11 alpha chain

Chain A:

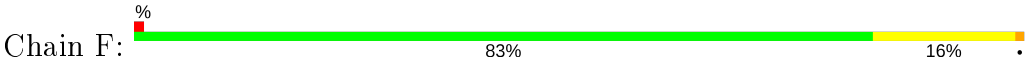


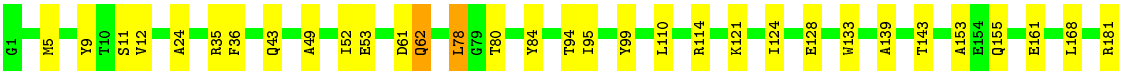


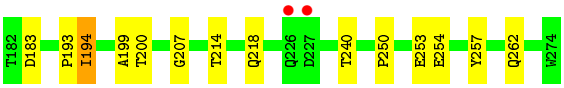


- Molecule 1: HLA class I histocompatibility antigen, A-11 alpha chain

Chain F:

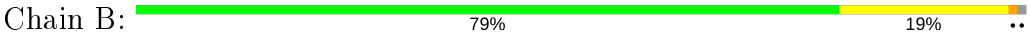






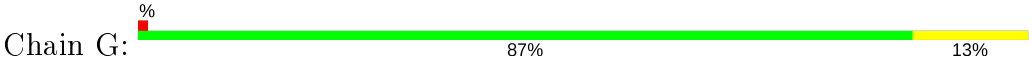
- Molecule 2: Beta-2-microglobulin

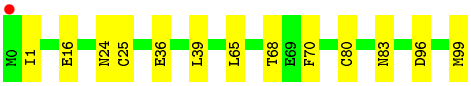
Chain B:



- Molecule 2: Beta-2-microglobulin

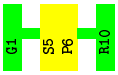
Chain G:





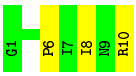
- Molecule 3: GTS3 peptide

Chain C:



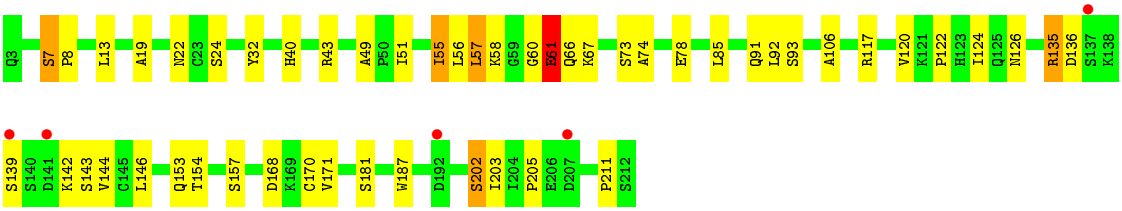
• Molecule 3: GTS3 peptide

Chain H:



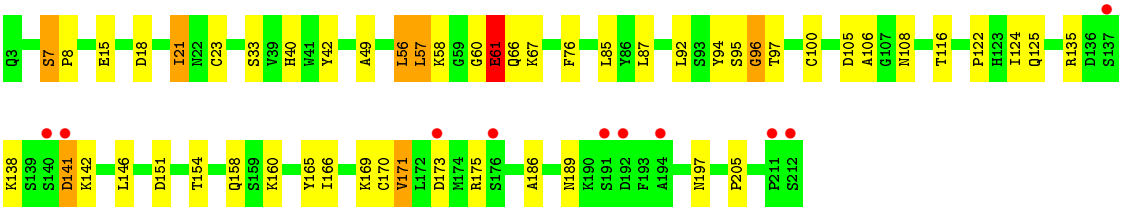
• Molecule 4: T-cell receptor alpha variable 30,T-cell receptor, sp3.4 alpha chain

Chain D:



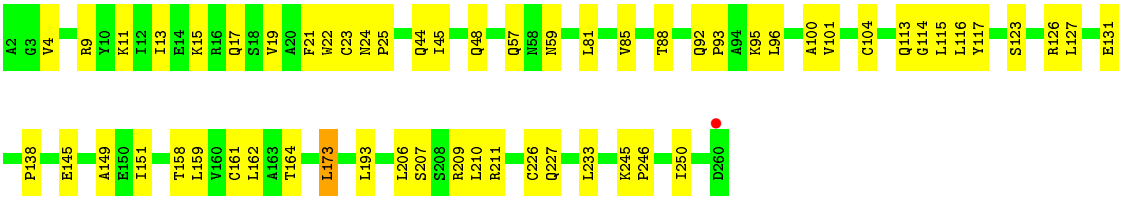
• Molecule 4: T-cell receptor alpha variable 30,T-cell receptor, sp3.4 alpha chain

Chain I:



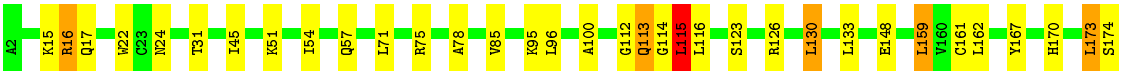
• Molecule 5: D30 TCR beta chain

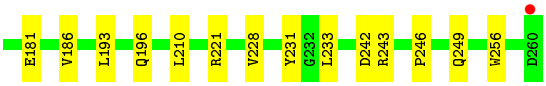
Chain E:



• Molecule 5: D30 TCR beta chain

Chain J:




E181
V186
L193
Q196
L210
R221
V228
Y231
G232
L233
D242
R243
P246
Q249
W256
D260

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	77.39Å 170.54Å 85.69Å 90.00° 113.89° 90.00°	Depositor
Resolution (Å)	46.01 – 3.20 44.34 – 3.20	Depositor EDS
% Data completeness (in resolution range)	99.8 (46.01-3.20) 99.8 (44.34-3.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.59 (at 3.19Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.177 , 0.246 0.190 , 0.257	Depositor DCC
R_{free} test set	1663 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	38.6	Xtriage
Anisotropy	0.683	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 39.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.030 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	13218	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.11% 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

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	0.42	0/2296	0.65	0/3117
1	F	0.43	0/2296	0.64	0/3117
2	B	0.40	0/852	0.64	0/1152
2	G	0.41	0/860	0.63	0/1162
3	C	0.47	0/70	0.84	0/92
3	H	0.51	0/70	0.76	0/92
4	D	0.50	0/1568	0.74	0/2116
4	I	0.44	0/1568	0.74	0/2116
5	E	0.43	0/1986	0.67	0/2702
5	J	0.45	0/1986	0.68	0/2702
All	All	0.44	0/13552	0.68	0/18368

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	2235	0	2085	15	0
1	F	2235	0	2089	21	0
2	B	829	0	794	10	0
2	G	837	0	805	5	0
3	C	70	0	73	1	0
3	H	70	0	73	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	1535	0	1477	34	0
4	I	1535	0	1476	35	0
5	E	1935	0	1878	28	0
5	J	1935	0	1879	22	0
6	E	1	0	0	0	0
6	F	1	0	0	0	0
All	All	13218	0	12629	148	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.

The worst 5 of 148 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:I:60:GLY:HA2	4:I:61:GLU:HB2	1.13	1.12
4:D:60:GLY:HA2	4:D:61:GLU:HB2	1.09	1.07
4:D:60:GLY:CA	4:D:61:GLU:HB2	1.99	0.93
1:F:155:GLN:NE2	5:J:114:GLY:O	2.02	0.93
4:I:60:GLY:HA2	4:I:61:GLU:CB	2.01	0.89

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	272/274 (99%)	260 (96%)	11 (4%)	1 (0%)	34 69
1	F	272/274 (99%)	254 (93%)	17 (6%)	1 (0%)	34 69
2	B	97/100 (97%)	95 (98%)	2 (2%)	0	100 100
2	G	98/100 (98%)	91 (93%)	6 (6%)	1 (1%)	15 54
3	C	8/10 (80%)	7 (88%)	1 (12%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
3	H	8/10 (80%)	7 (88%)	1 (12%)	0	100 100
4	D	196/198 (99%)	173 (88%)	18 (9%)	5 (3%)	5 31
4	I	196/198 (99%)	174 (89%)	17 (9%)	5 (3%)	5 31
5	E	242/244 (99%)	227 (94%)	13 (5%)	2 (1%)	19 58
5	J	242/244 (99%)	219 (90%)	20 (8%)	3 (1%)	13 49
All	All	1631/1652 (99%)	1507 (92%)	106 (6%)	18 (1%)	14 51

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	61	GLU
4	D	135	ARG
5	J	15	LYS
4	D	139	SER
5	E	15	LYS

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	231/231 (100%)	216 (94%)	15 (6%)	17 51
1	F	231/231 (100%)	217 (94%)	14 (6%)	18 54
2	B	94/95 (99%)	89 (95%)	5 (5%)	22 58
2	G	95/95 (100%)	93 (98%)	2 (2%)	53 79
3	C	8/8 (100%)	7 (88%)	1 (12%)	4 21
3	H	8/8 (100%)	8 (100%)	0	100 100
4	D	173/173 (100%)	159 (92%)	14 (8%)	11 42
4	I	173/173 (100%)	153 (88%)	20 (12%)	5 24
5	E	209/209 (100%)	196 (94%)	13 (6%)	18 53
5	J	209/209 (100%)	183 (88%)	26 (12%)	4 21
All	All	1431/1432 (100%)	1321 (92%)	110 (8%)	13 44

5 of 110 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	35	ARG
1	F	194	ILE
5	J	174	SER
1	F	53	GLU
1	F	114	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
5	E	229	GLN
1	F	62	GLN
5	J	57	GLN
5	E	223	HIS
5	J	113	GLN

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)

There are no ligands in this entry.

5.7 Other polymers (i)

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, 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	274/274 (100%)	-0.30	0 100 100	13, 33, 71, 94	0
1	F	274/274 (100%)	-0.29	2 (0%) 87 81	9, 32, 73, 87	0
2	B	99/100 (99%)	-0.45	0 100 100	16, 27, 52, 70	0
2	G	100/100 (100%)	-0.12	1 (1%) 82 72	19, 42, 70, 85	0
3	C	10/10 (100%)	-0.51	0 100 100	11, 21, 29, 54	0
3	H	10/10 (100%)	-0.72	0 100 100	17, 22, 27, 34	0
4	D	198/198 (100%)	-0.04	5 (2%) 57 43	21, 47, 81, 89	1 (0%)
4	I	198/198 (100%)	-0.00	10 (5%) 28 16	15, 46, 83, 95	1 (0%)
5	E	244/244 (100%)	-0.37	1 (0%) 92 89	11, 31, 58, 92	6 (2%)
5	J	244/244 (100%)	-0.32	1 (0%) 92 89	13, 34, 60, 91	6 (2%)
All	All	1651/1652 (99%)	-0.25	20 (1%) 79 67	9, 36, 75, 95	14 (0%)

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	G	0	MET	3.8
4	I	140	SER	3.4
4	D	141	ASP	3.4
4	I	141	ASP	3.1
4	I	211	PRO	2.9

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 (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.