W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 12:01 am BST

PDB ID	:	4MJI
Title	:	T cell response to a HIV reverse transcriptase epitope presented by the protective allele HLA-B*51:01
Authors	:	Rizkallah, P.J.; Cole, D.K.; Sewell, A.K.; Motozono, C.; Takiguchi, M.
Deposited on	:	2013-09-03
Resolution	:	2.99 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:

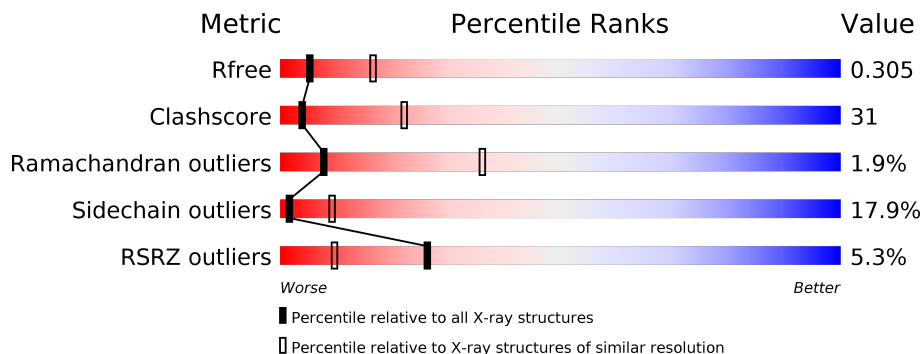|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance ⓘ

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

The reported resolution of this entry is 2.99 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.
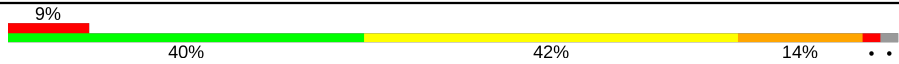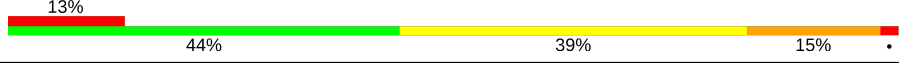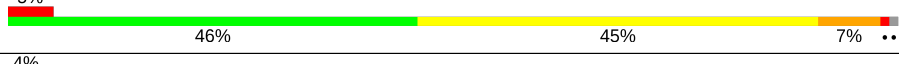
Metric	Percentile Ranks	Value
Rfree		0.305
Clashscore		31
Ramachandran outliers		1.9%
Sidechain outliers		17.9%
RSRZ outliers		5.3%

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	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	5% — 52% / 39% / 9%
1	F	276	2% — 46% / 44% / 9% •
2	B	99	63% / 32% / 5%
2	G	99	3% — 59% / 30% / 11%
3	C	8	38% / 63%
3	H	8	38% / 63%

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	195	9% / 40% / 42% / 14%
4	I	195	13% / 44% / 39% / 15%
5	E	242	5% / 46% / 45% / 7%
5	J	242	4% / 52% / 40% / 8%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13067 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, B-51 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	276	Total 2263	C 1412	N 412	O 432	S 7	0	0	0
1	F	275	Total 2255	C 1407	N 411	O 430	S 7	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	99	Total 829	C 528	N 140	O 158	S 3	0	0	0
2	G	99	Total 829	C 528	N 140	O 158	S 3	0	0	0

- Molecule 3 is a protein called HIV Reverse Transcriptase peptide Marker.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	8	Total 60	C 40	N 8	O 12	0	0	0
3	H	8	Total 60	C 40	N 8	O 12	0	0	0

- Molecule 4 is a protein called T-Cell Receptor Chain alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	192	Total 1492	C 912	N 261	O 311	S 8	0	0	0
4	I	194	Total 1502	C 915	N 263	O 316	S 8	0	0	0

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

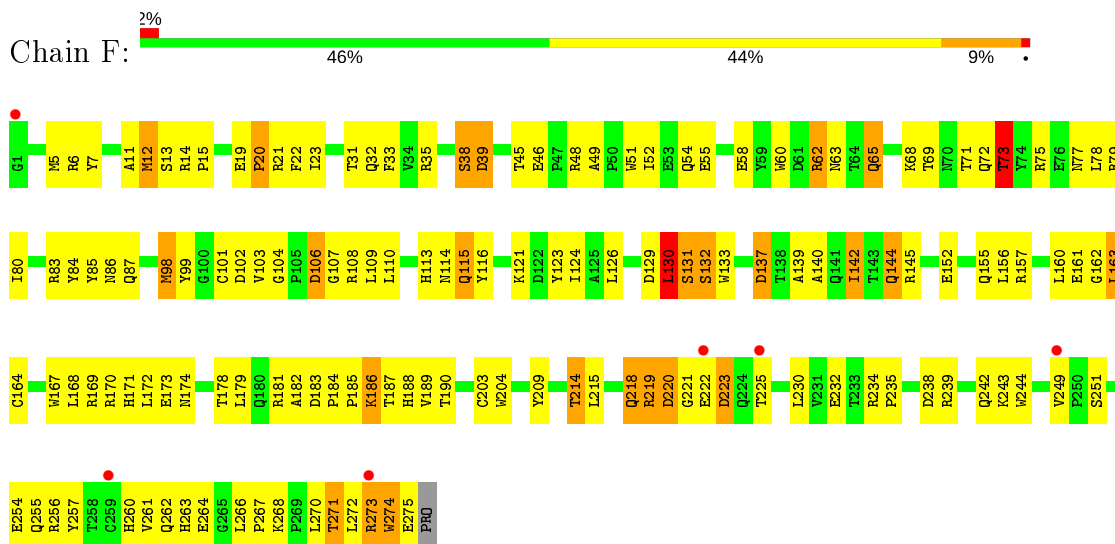| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	240	Total 1884	C 1189	N 327	O 363	S 5	0	0	0
5	J	241	Total 1893	C 1193	N 328	O 367	S 5	0	0	0

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.

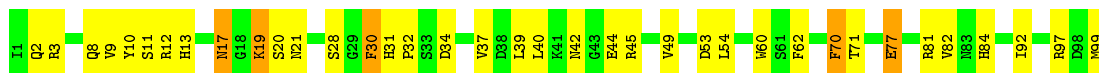- Molecule 1: HLA class I histocompatibility antigen, B-51 alpha chain
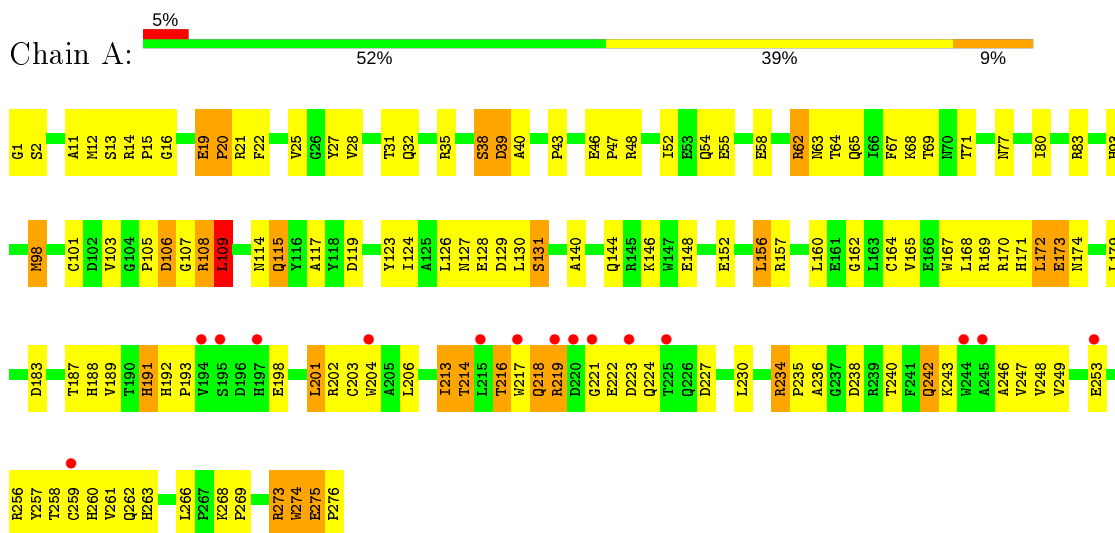
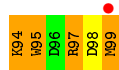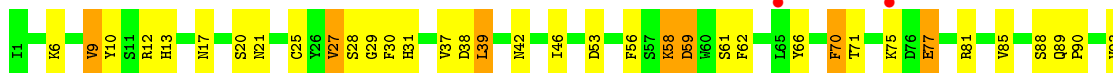- Molecule 1: HLA class I histocompatibility antigen, B-51 alpha chain

- Molecule 2: Beta-2-microglobulin

I1 Q2 R3 Q8 V9 Y10 S11 R12 H13 N17 G18 K19 S20 N21 S28 G29 F30 H31 P32 S33 D34 V37 D38 L39 L40 K41 N42 G43 E44 R45 V49 D53 L54 W60 S61 F62 F70 T71 E77 R81 V82 N83 H84 I92 R97 D98 M99

- Molecule 2: Beta-2-microglobulin

Chain G:

I1 K6 V9 Y10 S11 R12 H13 N17 S20 N21 C25 Y26 V27 S28 G29 F30 H31 V37 D38 L39 N42 I46 D53 F56 S57 K58 D59 W60 S61 F62 L65 Y66 F70 T71 K75 D76 E77 R81 V85 S88 Q89 P90 V93

K94 W95 D96 R97 D98 M99

- Molecule 3: HIV Reverse Transcriptase peptide Marker

Chain C:

T1 I5 P6 S7 I8

- Molecule 3: HIV Reverse Transcriptase peptide Marker

Chain H:

T1 T4 I5 P6 S7 I8

- Molecule 4: T-Cell Receptor Chain alpha

Chain D:

GLY GLU GLU D7 P8 Q9 S12 I13 E17 M21 S24 Y25 K26 T27 S28 I29 N30 N31 L32 Q33 W34 Y35 R36 Q37 N38 S39 G40 R41 G42 L43 V44 H45 L46 I47 L48 I49 R50 S51 N52 E53 R54 H57 R60 L61 R62 D66 K69 S72 L75 I76

T77 R80 D83 T84 F88 C89 A90 T91 D92 D93 R97 Q98 L99 T100 F101 G102 S103 G104 T105 T108 V109 L110 P111 D112 I113 Q114 N115 P118 A119 V120 Y121 Q122 L123 R124 D125 S126 K127 S128 S129 D130 K131 S132 V133 C134 L135 F136 T137 D138 Q142 V145 S146 Q147

S148 K149 D152 V153 Y154 I155 T156 D157 K158 C159 V160 R164 S165 M166 D167 N171 S172 A173 V174 A175 W176 S177 N178 K179 S180 D181 F182 A183 C184 A185 N186 A187 F188 N189 M190 S191 I192 I193 P194 E195 D196 T197 F198

- Molecule 4: T-Cell Receptor Chain alpha

Chain I:

G4 E5 E6 D7 P8 Q9 A10 L11 S12 I13 Q14 E17 N18 M21 M22 C23 S24 Y25 K26 T27 S28 I29 N30 N31 L32 Q33 W34 Y35 R36 Q37 N38 R41 G42 L43 V44 H45 L46 I47 L48 I49 R50 S51 N52 E53 R54 E55 S58 G59 R60 L61 R62 V63 T64 L65 D66 T67

• Molecule 5: T-cell Receptor Beta chain

Chain E:

• Molecule 5: T-cell Receptor Beta chain

Chain J:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	45.51Å 88.87Å 129.27Å 102.91° 95.92° 90.09°	Depositor
Resolution (Å)	64.75 – 2.99 64.75 – 2.99	Depositor EDS
% Data completeness (in resolution range)	98.7 (64.75-2.99) 98.7 (64.75-2.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$<I/\sigma(I)>$ [1]	2.62 (at 3.01Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.242 , 0.304 0.242 , 0.305	Depositor DCC
R_{free} test set	1965 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	47.3	Xtriage
Anisotropy	0.885	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 46.6	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	0.88	EDS
Total number of atoms	13067	wwPDB-VP
Average B, all atoms (Å^2)	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.01% 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)

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	0.57	0/2329	0.83	1/3170 (0.0%)
1	F	0.58	0/2320	0.84	3/3158 (0.1%)
2	B	0.62	0/852	0.76	0/1152
2	G	0.56	0/852	0.81	1/1152 (0.1%)
3	C	0.60	0/61	0.87	0/82
3	H	0.73	0/61	0.96	0/82
4	D	0.66	1/1515 (0.1%)	0.93	2/2049 (0.1%)
4	I	0.64	0/1524	0.93	4/2062 (0.2%)
5	E	0.55	0/1934	0.78	2/2631 (0.1%)
5	J	0.55	0/1943	0.77	2/2643 (0.1%)
All	All	0.59	1/13391 (0.0%)	0.84	15/18181 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	F	0	1
4	D	0	2
4	I	0	8
5	E	0	1
All	All	0	12

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	47	ILE	N-CA	6.01	1.58	1.46

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	I	43	LEU	CA-CB-CG	7.77	133.17	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	42	GLY	N-CA-C	6.58	129.55	113.10
4	I	123	LEU	CA-CB-CG	6.52	130.31	115.30
5	J	63	LEU	CA-CB-CG	6.37	129.94	115.30
4	D	46	LEU	N-CA-C	6.30	128.02	111.00

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	D	45	HIS	Peptide
4	D	46	LEU	Peptide
5	E	208	ASN	Peptide
1	F	221	GLY	Peptide
4	I	7	ASP	Peptide

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	2263	0	2113	137	0
1	F	2255	0	2106	135	0
2	B	829	0	794	30	0
2	G	829	0	794	38	0
3	C	60	0	64	17	0
3	H	60	0	64	11	0
4	D	1492	0	1429	121	0
4	I	1502	0	1433	151	0
5	E	1884	0	1794	120	0
5	J	1893	0	1798	109	0
All	All	13067	0	12389	777	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 31.

The worst 5 of 777 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:36:ARG:HA	4:D:45:HIS:O	1.31	1.27
1:F:249:VAL:HG11	1:F:257:TYR:CE1	1.76	1.19
1:F:189:VAL:O	1:F:274:TRP:HE3	1.24	1.17
4:I:29:ILE:N	4:I:29:ILE:HD12	1.57	1.16
4:I:36:ARG:HA	4:I:45:HIS:O	1.44	1.14

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	274/276 (99%)	245 (89%)	25 (9%)	4 (2%)	10	42
1	F	273/276 (99%)	239 (88%)	26 (10%)	8 (3%)	4	24
2	B	97/99 (98%)	91 (94%)	6 (6%)	0	100	100
2	G	97/99 (98%)	88 (91%)	8 (8%)	1 (1%)	15	53
3	C	6/8 (75%)	6 (100%)	0	0	100	100
3	H	6/8 (75%)	6 (100%)	0	0	100	100
4	D	190/195 (97%)	166 (87%)	17 (9%)	7 (4%)	3	19
4	I	192/195 (98%)	163 (85%)	23 (12%)	6 (3%)	4	23
5	E	238/242 (98%)	213 (90%)	22 (9%)	3 (1%)	12	45
5	J	239/242 (99%)	219 (92%)	19 (8%)	1 (0%)	34	72
All	All	1612/1640 (98%)	1436 (89%)	146 (9%)	30 (2%)	8	36

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	9	GLN
4	D	28	SER
4	D	42	GLY

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	D	197	THR
1	F	130	LEU

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	234/234 (100%)	194 (83%)	40 (17%)	2	10
1	F	233/234 (100%)	187 (80%)	46 (20%)	1	7
2	B	94/94 (100%)	82 (87%)	12 (13%)	4	19
2	G	94/94 (100%)	81 (86%)	13 (14%)	3	17
3	C	7/7 (100%)	7 (100%)	0	100	100
3	H	7/7 (100%)	6 (86%)	1 (14%)	3	15
4	D	172/174 (99%)	131 (76%)	41 (24%)	0	3
4	I	173/174 (99%)	138 (80%)	35 (20%)	1	6
5	E	203/204 (100%)	168 (83%)	35 (17%)	2	10
5	J	204/204 (100%)	173 (85%)	31 (15%)	3	14
All	All	1421/1426 (100%)	1167 (82%)	254 (18%)	2	9

5 of 254 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	E	177	GLN
1	F	106	ASP
5	J	126	GLU
5	E	197	ARG
1	F	12	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 51 such sidechains are listed below:

Mol	Chain	Res	Type
5	E	235	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	114	ASN
5	J	164	ASN
1	F	65	GLN
1	F	115	GLN

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 ⓘ

There are no ligands in this entry.

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 (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	276/276 (100%)	0.41	15 (5%)	25	9	35, 59, 113, 130	0
1	F	275/276 (99%)	0.34	6 (2%)	62	33	37, 59, 103, 124	0
2	B	99/99 (100%)	0.03	0	100	100	45, 57, 76, 82	0
2	G	99/99 (100%)	0.21	3 (3%)	50	22	49, 59, 79, 82	0
3	C	8/8 (100%)	0.50	0	100	100	48, 50, 58, 58	0
3	H	8/8 (100%)	0.65	0	100	100	45, 46, 53, 53	0
4	D	192/195 (98%)	0.57	18 (9%)	8	3	39, 63, 112, 124	0
4	I	194/195 (99%)	0.68	25 (12%)	3	1	39, 65, 113, 122	0
5	E	240/242 (99%)	0.39	11 (4%)	32	12	37, 69, 100, 129	0
5	J	241/242 (99%)	0.34	9 (3%)	41	17	37, 65, 95, 122	0
All	All	1632/1640 (99%)	0.40	87 (5%)	26	10	35, 61, 107, 130	0

The worst 5 of 87 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	1	GLY	4.8
1	A	225	THR	4.5
4	D	154	TYR	4.5
4	I	130	ASP	4.4
4	I	191	SER	4.2

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.