wwPDB X-ray Structure Validation Summary Report (i)

Oct 31, 2023 – 02:44 PM JST

PDB ID	:	5B8C
Title	:	High resolution structure of the human PD-1 in complex with pembrolizumab Fv
Authors	:	Horita, S.; Shimamura, T.; Iwata, S.; Nomura, N.
Deposited on	:	2016-06-14
Resolution	:	2.15 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

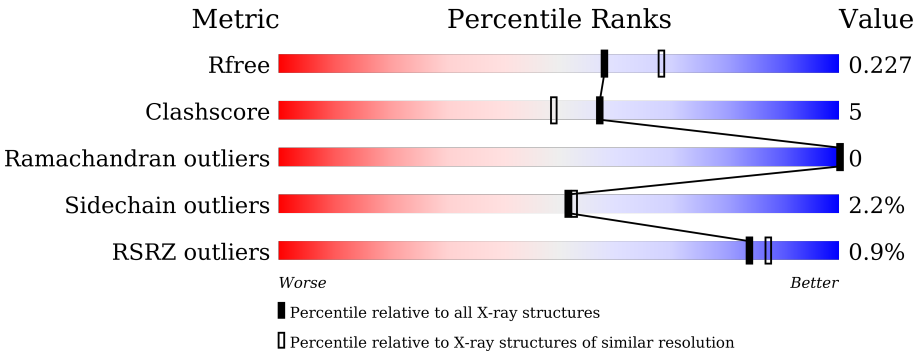
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.15 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.



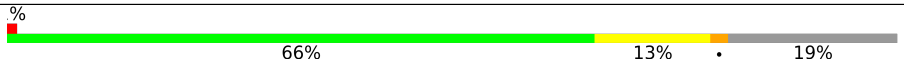
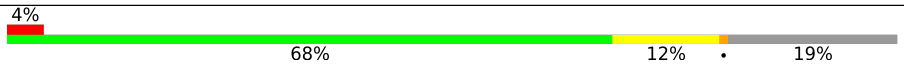
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2523 (2.16-2.12)
Clashscore	141614	2653 (2.16-2.12)
Ramachandran outliers	138981	2618 (2.16-2.12)
Sidechain outliers	138945	2617 (2.16-2.12)
RSRZ outliers	127900	2485 (2.16-2.12)

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	119	82% 10% • 7%
1	D	119	82% 11% • 7%
1	G	119	% 82% 8% • 7%
1	J	119	% 79% 13% • 7%
2	B	120	96% •
2	E	120	2% 82% 17% ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	120	
2	K	120	
3	C	139	
3	F	139	
3	I	139	
3	L	139	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 10858 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 Pembrolizumab light chain variable region (PemVL).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	111	Total	C	N	O	S	0	0	0
			833	532	139	160	2			
1	D	111	Total	C	N	O	S	0	0	0
			830	532	136	160	2			
1	G	111	Total	C	N	O	S	0	0	0
			836	534	138	162	2			
1	J	111	Total	C	N	O	S	0	0	0
			840	536	139	163	2			

- Molecule 2 is a protein called Pembrolizumab heavy chain variable region (PemVH).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	120	Total	C	N	O	S	0	0	0
			938	592	155	185	6			
2	E	119	Total	C	N	O	S	0	0	0
			922	584	153	179	6			
2	H	120	Total	C	N	O	S	0	0	0
			929	587	155	181	6			
2	K	120	Total	C	N	O	S	0	0	0
			940	595	156	183	6			

- Molecule 3 is a protein called Programmed cell death protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	116	Total	C	N	O	S	0	0	0
			903	562	162	175	4			
3	F	116	Total	C	N	O	S	0	0	0
			893	560	159	170	4			
3	I	113	Total	C	N	O	S	0	0	0
			849	537	146	162	4			
3	L	113	Total	C	N	O	S	0	0	0
			874	546	158	166	4			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	30	GLY	-	expression tag	UNP Q15116
C	31	SER	-	expression tag	UNP Q15116
C	93	SER	CYS	engineered mutation	UNP Q15116
C	161	THR	-	expression tag	UNP Q15116
C	162	SER	-	expression tag	UNP Q15116
C	163	GLU	-	expression tag	UNP Q15116
C	164	ASN	-	expression tag	UNP Q15116
C	165	LEU	-	expression tag	UNP Q15116
C	166	TYR	-	expression tag	UNP Q15116
C	167	PHE	-	expression tag	UNP Q15116
C	168	GLN	-	expression tag	UNP Q15116
F	30	GLY	-	expression tag	UNP Q15116
F	31	SER	-	expression tag	UNP Q15116
F	93	SER	CYS	engineered mutation	UNP Q15116
F	161	THR	-	expression tag	UNP Q15116
F	162	SER	-	expression tag	UNP Q15116
F	163	GLU	-	expression tag	UNP Q15116
F	164	ASN	-	expression tag	UNP Q15116
F	165	LEU	-	expression tag	UNP Q15116
F	166	TYR	-	expression tag	UNP Q15116
F	167	PHE	-	expression tag	UNP Q15116
F	168	GLN	-	expression tag	UNP Q15116
I	30	GLY	-	expression tag	UNP Q15116
I	31	SER	-	expression tag	UNP Q15116
I	93	SER	CYS	engineered mutation	UNP Q15116
I	161	THR	-	expression tag	UNP Q15116
I	162	SER	-	expression tag	UNP Q15116
I	163	GLU	-	expression tag	UNP Q15116
I	164	ASN	-	expression tag	UNP Q15116
I	165	LEU	-	expression tag	UNP Q15116
I	166	TYR	-	expression tag	UNP Q15116
I	167	PHE	-	expression tag	UNP Q15116
I	168	GLN	-	expression tag	UNP Q15116
L	30	GLY	-	expression tag	UNP Q15116
L	31	SER	-	expression tag	UNP Q15116
L	93	SER	CYS	engineered mutation	UNP Q15116
L	161	THR	-	expression tag	UNP Q15116
L	162	SER	-	expression tag	UNP Q15116
L	163	GLU	-	expression tag	UNP Q15116
L	164	ASN	-	expression tag	UNP Q15116
L	165	LEU	-	expression tag	UNP Q15116
L	166	TYR	-	expression tag	UNP Q15116

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
L	167	PHE	-	expression tag	UNP Q15116
L	168	GLN	-	expression tag	UNP Q15116

- Molecule 4 is water.

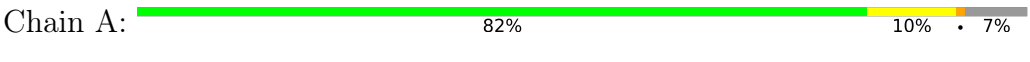
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	29	Total O 29 29	0	0
4	B	37	Total O 37 37	0	0
4	C	24	Total O 24 24	0	0
4	D	19	Total O 19 19	0	0
4	E	22	Total O 22 22	0	0
4	F	17	Total O 17 17	0	0
4	G	32	Total O 32 32	0	0
4	H	26	Total O 26 26	0	0
4	I	10	Total O 10 10	0	0
4	J	28	Total O 28 28	0	0
4	K	20	Total O 20 20	0	0
4	L	7	Total O 7 7	0	0

3 Residue-property plots (i)

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.

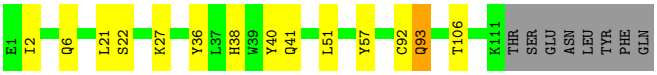
• Molecule 1: Pembrolizumab light chain variable region (PemVL)

Chain A:



• Molecule 1: Pembrolizumab light chain variable region (PemVL)

Chain D:



• Molecule 1: Pembrolizumab light chain variable region (PemVL)

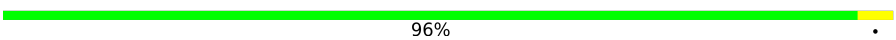
Chain G:

• Molecule 1: Pembrolizumab light chain variable region (PemVL)

Chain J:

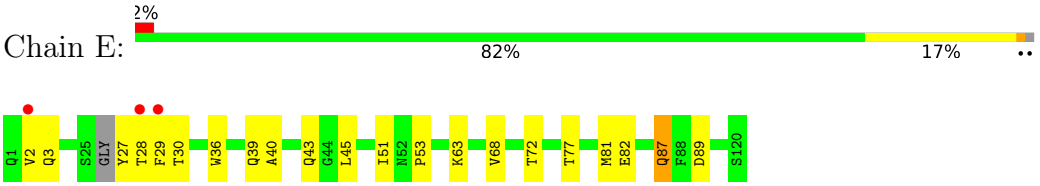
• Molecule 2: Pembrolizumab heavy chain variable region (PemVH)

Chain B:



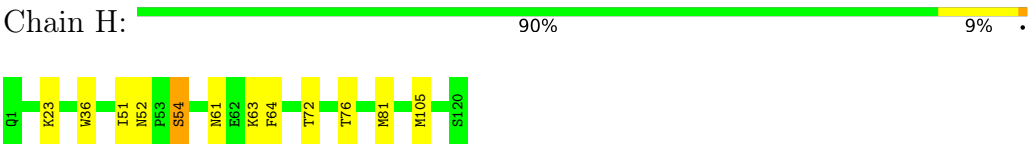
• Molecule 2: Pembrolizumab heavy chain variable region (PemVH)

Chain E:



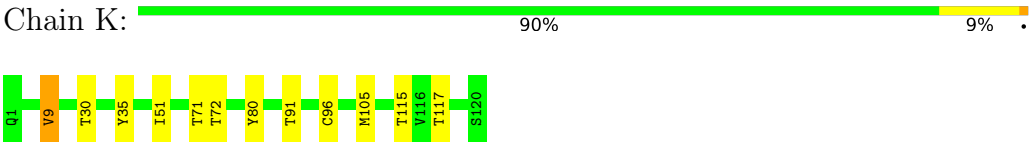
• Molecule 2: Pembrolizumab heavy chain variable region (PemVH)

Chain H:



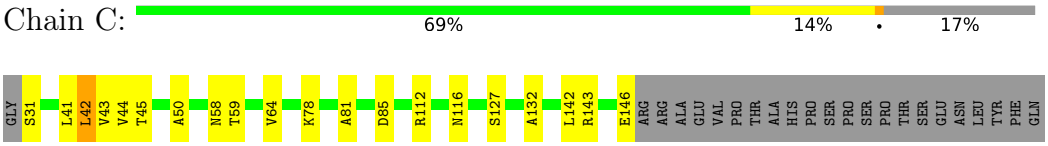
• Molecule 2: Pembrolizumab heavy chain variable region (PemVH)

Chain K:



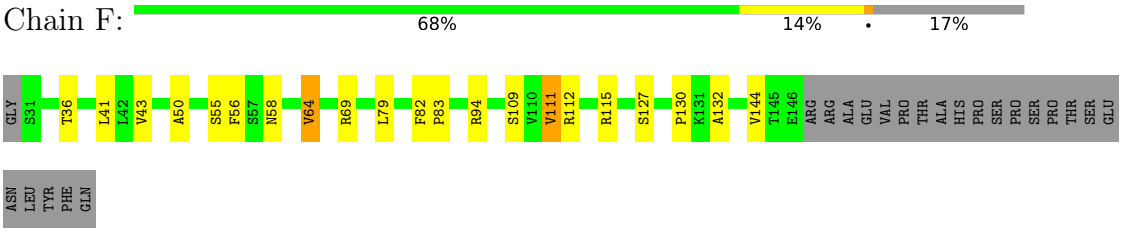
• Molecule 3: Programmed cell death protein 1

Chain C:



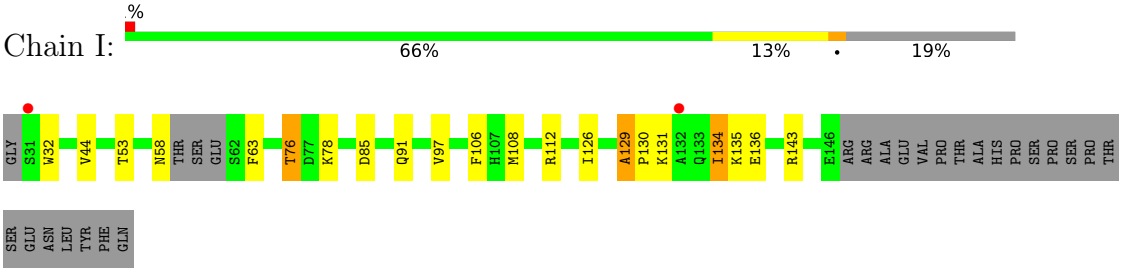
• Molecule 3: Programmed cell death protein 1

Chain F:



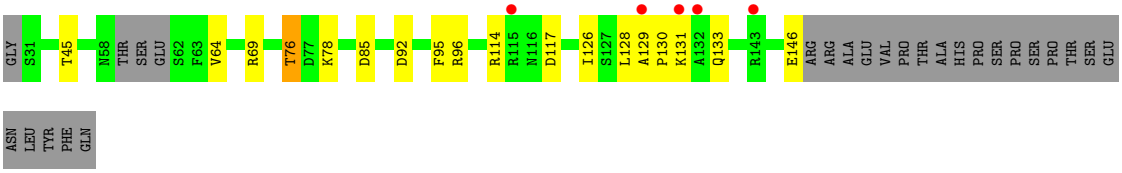
• Molecule 3: Programmed cell death protein 1

Chain I:



• Molecule 3: Programmed cell death protein 1

Chain L:


GLY
S31
T45
N58
THR
SER
GLU
S62
F63
V64
R69
T76
D77
K78
D85
D92
F95
R96
R114
R115
N116
D117
I126
S127
L128
A129
P130
K131
A132
Q133
R143
E146
ARG
ARG
ALA
GLU
VAL
PRO
THR
ALA
HIS
PRO
SER
PRO
SER
PRO
THR
SER
GLU
ASN
LEU
TYR
PHE
GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	143.68Å 143.13Å 76.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.42 – 2.15 45.42 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.7 (45.42-2.15) 98.6 (45.42-2.15)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.15 (at 2.14Å)	Xtriage
Refinement program	PHENIX (dev_2210: ???)	Depositor
R, R_{free}	0.184 , 0.226 0.185 , 0.227	Depositor DCC
R_{free} test set	4438 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	41.0	Xtriage
Anisotropy	0.348	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 28.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.40$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	0.209 for k,h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	10858	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.93% 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	0.27	0/854	0.53	0/1162
1	D	0.27	0/851	0.52	0/1158
1	G	0.27	0/857	0.52	0/1166
1	J	0.26	0/861	0.52	0/1170
2	B	0.29	0/961	0.53	0/1304
2	E	0.28	0/944	0.52	0/1278
2	H	0.28	0/952	0.53	0/1292
2	K	0.27	0/963	0.52	0/1306
3	C	0.27	0/924	0.53	0/1256
3	F	0.29	0/914	0.51	0/1243
3	I	0.53	1/869 (0.1%)	0.51	0/1183
3	L	0.27	0/894	0.52	0/1213
All	All	0.30	1/10844 (0.0%)	0.52	0/14731

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	I	129	ALA	C-N	13.39	1.59	1.34

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	833	0	813	7	0
1	D	830	0	808	8	0
1	G	836	0	812	11	0
1	J	840	0	819	13	0
2	B	938	0	872	3	0
2	E	922	0	850	12	0
2	H	929	0	863	8	0
2	K	940	0	882	7	0
3	C	903	0	864	13	0
3	F	893	0	848	15	0
3	I	849	0	795	12	0
3	L	874	0	834	10	0
4	A	29	0	0	0	0
4	B	37	0	0	0	0
4	C	24	0	0	1	0
4	D	19	0	0	0	0
4	E	22	0	0	0	0
4	F	17	0	0	0	0
4	G	32	0	0	0	0
4	H	26	0	0	2	0
4	I	10	0	0	0	0
4	J	28	0	0	0	0
4	K	20	0	0	0	0
4	L	7	0	0	0	0
All	All	10858	0	10060	102	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 5.

The worst 5 of 102 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)
3:F:58:ASN:HD21	3:I:76:THR:HG22	1.59	0.67
2:K:9:VAL:HG23	2:K:115:THR:HB	1.78	0.65
1:G:24:ARG:NH1	1:G:74:ASP:OD1	2.31	0.63
3:F:43:VAL:HG21	1:J:110:ILE:HD13	1.79	0.63
2:E:2:VAL:HG13	2:E:27:TYR:HD1	1.65	0.62

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	109/119 (92%)	105 (96%)	4 (4%)	0	100	100
1	D	109/119 (92%)	105 (96%)	4 (4%)	0	100	100
1	G	109/119 (92%)	105 (96%)	4 (4%)	0	100	100
1	J	109/119 (92%)	105 (96%)	4 (4%)	0	100	100
2	B	118/120 (98%)	117 (99%)	1 (1%)	0	100	100
2	E	115/120 (96%)	113 (98%)	2 (2%)	0	100	100
2	H	118/120 (98%)	117 (99%)	1 (1%)	0	100	100
2	K	118/120 (98%)	118 (100%)	0	0	100	100
3	C	114/139 (82%)	112 (98%)	2 (2%)	0	100	100
3	F	114/139 (82%)	112 (98%)	2 (2%)	0	100	100
3	I	109/139 (78%)	106 (97%)	3 (3%)	0	100	100
3	L	109/139 (78%)	105 (96%)	4 (4%)	0	100	100
All	All	1351/1512 (89%)	1320 (98%)	31 (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	89/101 (88%)	88 (99%)	1 (1%)	73	76
1	D	88/101 (87%)	87 (99%)	1 (1%)	73	76

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	G	89/101 (88%)	86 (97%)	3 (3%)	37	34
1	J	90/101 (89%)	88 (98%)	2 (2%)	52	53
2	B	99/102 (97%)	99 (100%)	0	100	100
2	E	95/102 (93%)	93 (98%)	2 (2%)	53	54
2	H	97/102 (95%)	96 (99%)	1 (1%)	76	79
2	K	99/102 (97%)	98 (99%)	1 (1%)	76	79
3	C	100/122 (82%)	97 (97%)	3 (3%)	41	39
3	F	95/122 (78%)	93 (98%)	2 (2%)	53	54
3	I	89/122 (73%)	83 (93%)	6 (7%)	16	10
3	L	95/122 (78%)	92 (97%)	3 (3%)	39	37
All	All	1125/1300 (86%)	1100 (98%)	25 (2%)	52	53

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	I	58	ASN
3	I	112	ARG
3	L	114	ARG
3	I	91	GLN
3	I	134	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	J	46	GLN
3	L	91	GLN
2	K	87	GLN
2	E	87	GLN
3	I	91	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 monosaccharides 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 (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, 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	111/119 (93%)	-0.25	0 100 100	30, 40, 54, 59	0
1	D	111/119 (93%)	-0.30	0 100 100	30, 39, 50, 61	0
1	G	111/119 (93%)	-0.32	1 (0%) 84 87	31, 41, 56, 75	0
1	J	111/119 (93%)	-0.08	1 (0%) 84 87	36, 48, 75, 83	0
2	B	120/120 (100%)	-0.31	0 100 100	30, 37, 54, 80	0
2	E	119/120 (99%)	-0.06	3 (2%) 57 64	32, 43, 72, 116	0
2	H	120/120 (100%)	-0.31	0 100 100	33, 41, 56, 85	0
2	K	120/120 (100%)	-0.28	0 100 100	33, 44, 61, 71	0
3	C	116/139 (83%)	-0.23	0 100 100	32, 43, 63, 81	0
3	F	116/139 (83%)	-0.23	0 100 100	32, 42, 58, 94	0
3	I	113/139 (81%)	0.19	2 (1%) 68 74	38, 59, 103, 132	0
3	L	113/139 (81%)	0.30	5 (4%) 34 42	40, 64, 104, 124	0
All	All	1381/1512 (91%)	-0.16	12 (0%) 84 87	30, 43, 74, 132	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	I	132	ALA	6.6
3	L	132	ALA	5.9
2	E	2	VAL	4.0
2	E	28	THR	3.7
1	G	110	ILE	2.8

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 monosaccharides 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.