



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 01:50 am GMT

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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

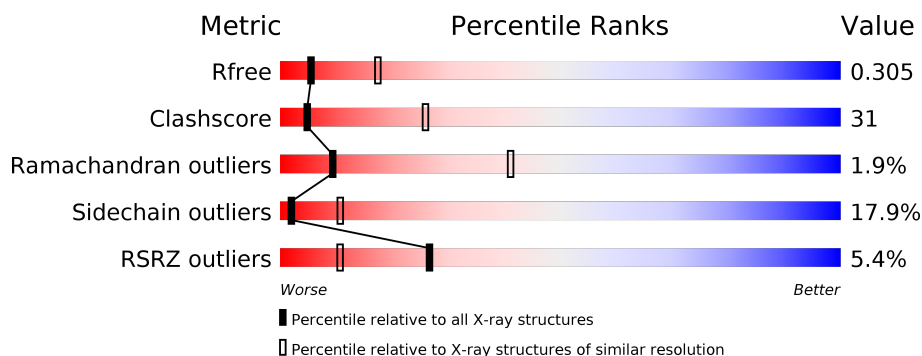
1 Overall quality at a glance ⓘ

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (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. 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 $\leq 5\%$. 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	<div> <div>5%</div> <div> <div>52%</div> <div>39%</div> <div>9%</div> </div> </div>
1	F	276	<div> <div>3%</div> <div> <div>46%</div> <div>44%</div> <div>9%</div> </div> </div>
2	B	99	<div> <div>63%</div> <div>32%</div> <div>5%</div> </div>
2	G	99	<div> <div>3%</div> <div> <div>59%</div> <div>30%</div> <div>11%</div> </div> </div>
3	C	8	<div> <div>38%</div> <div>63%</div> </div>
3	H	8	<div> <div>38%</div> <div>63%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	195	<div><div></div><div>10%</div><div>40%</div><div>42%</div><div>14%</div><div></div><div></div></div>
4	I	195	<div><div></div><div>14%</div><div>44%</div><div>39%</div><div>15%</div><div></div><div></div></div>
5	E	242	<div><div></div><div>4%</div><div>46%</div><div>45%</div><div>7%</div><div></div><div></div></div>
5	J	242	<div><div></div><div>4%</div><div>52%</div><div>40%</div><div>8%</div><div></div><div></div></div>

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	C	N	O	S	0	0	0
			2263	1412	412	432	7			
1	F	275	Total	C	N	O	S	0	0	0
			2255	1407	411	430	7			

- 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	99	Total	C	N	O	S	0	0	0
			829	528	140	158	3			

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

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

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

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

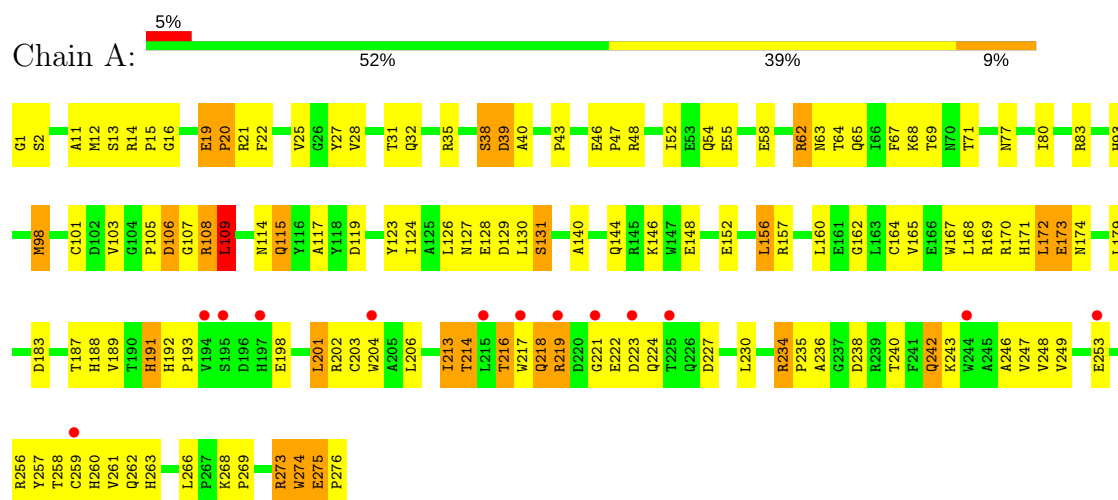
- Molecule 5 is a protein called T-cell Receptor Beta chain.

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

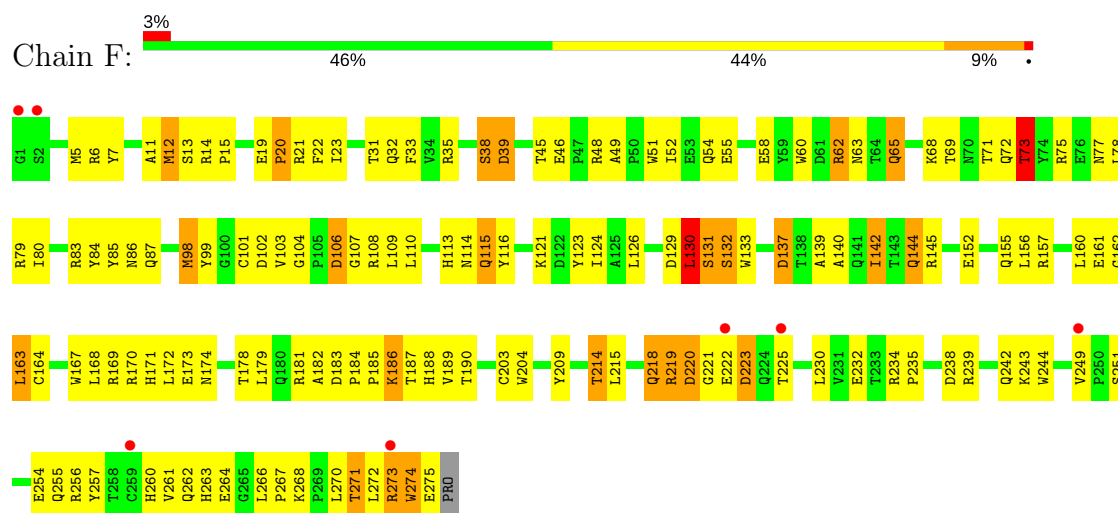
3 Residue-property plots

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

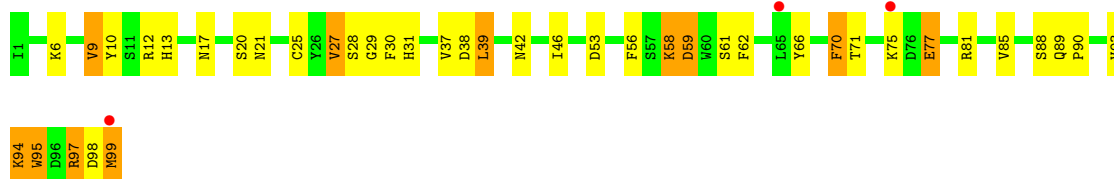


- Molecule 2: Beta-2-microglobulin





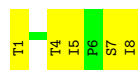
- Molecule 2: Beta-2-microglobulin



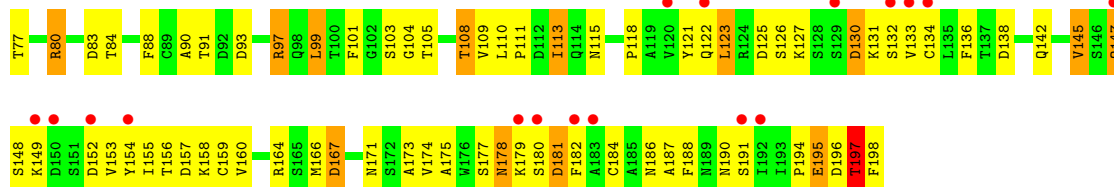
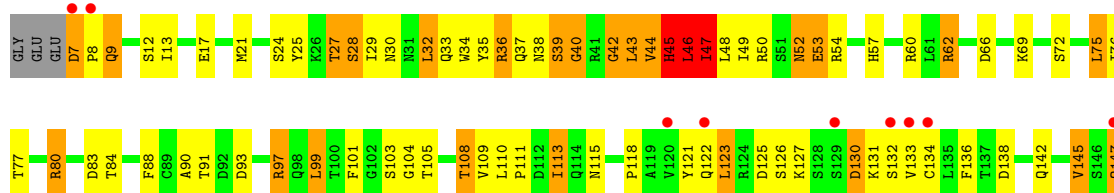
- Molecule 3: HIV Reverse Transcriptase peptide Marker



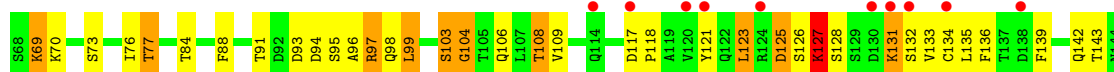
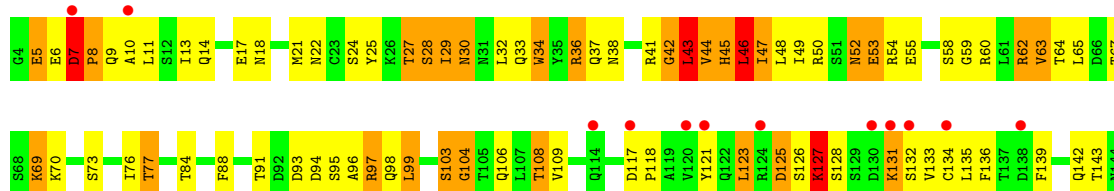
- Molecule 3: HIV Reverse Transcriptase peptide Marker

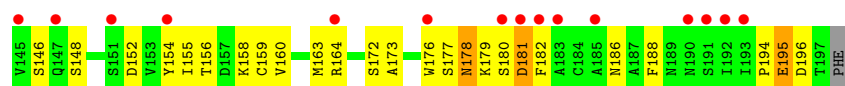


- Molecule 4: T-Cell Receptor Chain alpha

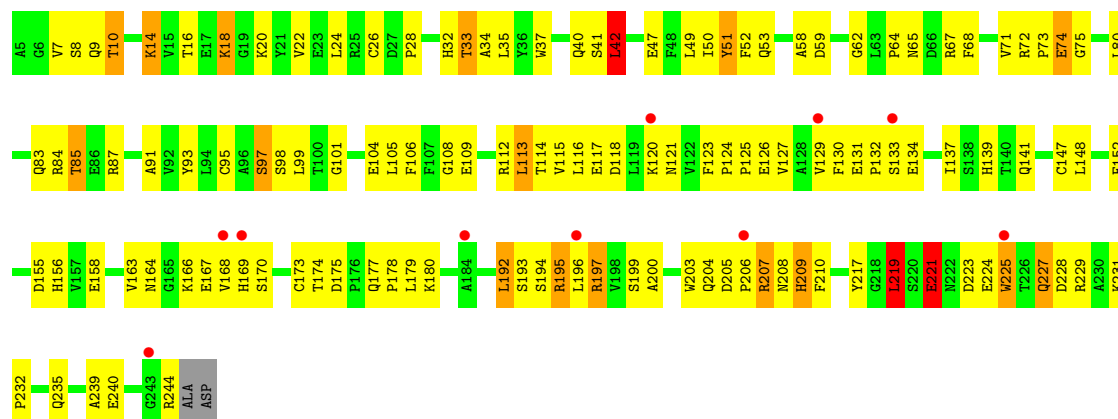


- Molecule 4: T-Cell Receptor Chain alpha

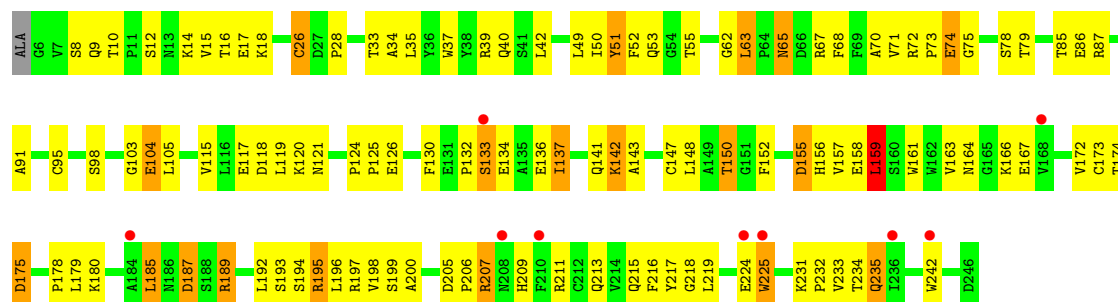




• Molecule 5: T-cell Receptor Beta chain



• Molecule 5: T-cell Receptor Beta chain



4 Data and refinement statistics

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) 91.3 (64.75-2.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.62 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.242 , 0.304 0.242 , 0.305	Depositor DCC
R_{free} test set	1965 reflections (5.28%)	DCC
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 , 43.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	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.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.57	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 ⓘ

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%)	12	48
1	F	273/276 (99%)	239 (88%)	26 (10%)	8 (3%)	5	28
2	B	97/99 (98%)	91 (94%)	6 (6%)	0	100	100
2	G	97/99 (98%)	88 (91%)	8 (8%)	1 (1%)	18	59
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%)	4	22
4	I	192/195 (98%)	163 (85%)	23 (12%)	6 (3%)	5	26
5	E	238/242 (98%)	213 (90%)	22 (9%)	3 (1%)	14	51
5	J	239/242 (99%)	219 (92%)	19 (8%)	1 (0%)	38	78
All	All	1612/1640 (98%)	1436 (89%)	146 (9%)	30 (2%)	9	41

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 ⓘ

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	12
1	F	233/234 (100%)	187 (80%)	46 (20%)	1	8
2	B	94/94 (100%)	82 (87%)	12 (13%)	5	22
2	G	94/94 (100%)	81 (86%)	13 (14%)	4	19
3	C	7/7 (100%)	7 (100%)	0	100	100
3	H	7/7 (100%)	6 (86%)	1 (14%)	4	18
4	D	172/174 (99%)	131 (76%)	41 (24%)	1	4
4	I	173/174 (99%)	138 (80%)	35 (20%)	1	7
5	E	203/204 (100%)	168 (83%)	35 (17%)	2	12
5	J	204/204 (100%)	173 (85%)	31 (15%)	3	16
All	All	1421/1426 (100%)	1167 (82%)	254 (18%)	2	11

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 [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 [i](#)

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%)	0.40	13 (4%) 32 13	35, 59, 113, 130	0
1	F	275/276 (99%)	0.33	7 (2%) 58 29	37, 59, 103, 124	0
2	B	99/99 (100%)	0.02	0 100 100	45, 57, 76, 82	0
2	G	99/99 (100%)	0.20	3 (3%) 51 23	49, 59, 79, 82	0
3	C	8/8 (100%)	0.51	0 100 100	48, 50, 58, 58	0
3	H	8/8 (100%)	0.67	0 100 100	45, 46, 53, 53	0
4	D	192/195 (98%)	0.56	19 (9%) 8 3	39, 63, 112, 124	0
4	I	194/195 (99%)	0.67	27 (13%) 3 1	39, 65, 113, 122	0
5	E	240/242 (99%)	0.39	10 (4%) 37 15	37, 69, 100, 129	0
5	J	241/242 (99%)	0.33	9 (3%) 42 18	37, 65, 95, 122	0
All	All	1632/1640 (99%)	0.39	88 (5%) 26 11	35, 61, 107, 130	0

The worst 5 of 88 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	1	GLY	4.9
1	A	225	THR	4.5
4	D	154	TYR	4.4
4	I	117	ASP	4.3
4	I	130	ASP	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.