



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 01:26 am GMT

PDB ID : 5HYJ  
Title : 1E6 TCR in Complex with HLA-A02 carrying AQWGPDPAAA  
Authors : Rizkallah, P.J.; Bulek, A.M.; Cole, D.K.; Sewell, A.K.  
Deposited on : 2016-02-01  
Resolution : 3.06 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](mailto: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

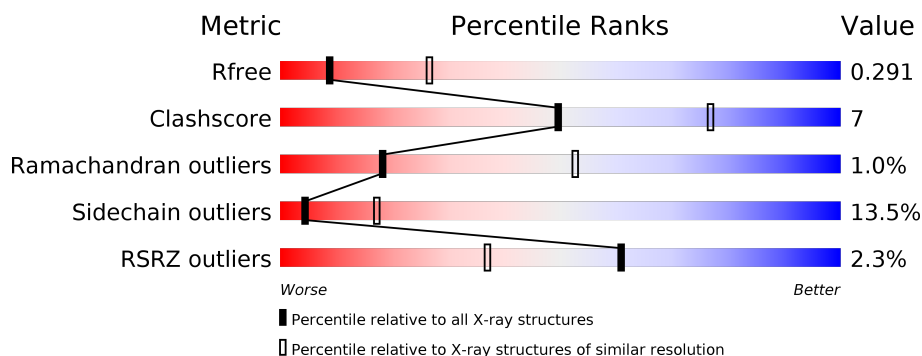
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.06 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	1348 (3.10-3.02)
Clashscore	112137	1462 (3.10-3.02)
Ramachandran outliers	110173	1410 (3.10-3.02)
Sidechain outliers	110143	1410 (3.10-3.02)
RSRZ outliers	101464	1355 (3.10-3.02)

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  $\geq 3$ , 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></div> <div>75%</div> <div>21%</div> <div>.</div> </div>
1	F	276	<div> <div>77%</div> <div>18%</div> <div>.</div> </div>
2	B	100	<div> <div>76%</div> <div>21%</div> <div>.</div> </div>
2	G	100	<div> <div>76%</div> <div>17%</div> <div>7%</div> </div>
3	C	10	<div> <div>60%</div> <div>30%</div> <div>10%</div> </div>
3	H	10	<div> <div>70%</div> <div>30%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	193	
4	I	193	
5	E	246	
5	J	246	

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 13392 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-2 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	276	Total	C	N	O	S	0	0	0
			2254	1408	410	427	9			
1	F	276	Total	C	N	O	S	0	0	0
			2254	1408	410	427	9			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	100	Total	C	N	O	S	0	0	0
			837	533	141	159	4			
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 ALA-GLN-TRP-GLY-PRO-ASP-PRO-ALA-ALA-ALA.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	10	Total	C	N	O	0	0	0
			70	44	12	14			
3	H	10	Total	C	N	O	0	0	0
			70	44	12	14			

- Molecule 4 is a protein called Human T-cell Receptor, Class I, Light alpha Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	193	Total	C	N	O	S	0	0	0
			1520	950	253	307	10			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	I	193	Total	C	N	O	S	0	0	0
			1520	950	253	307	10			

- Molecule 5 is a protein called Human T-cell Receptor, Class I, Heavy beta Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	246	Total	C	N	O	S	0	0	0
			1974	1249	341	374	10			
5	J	246	Total	C	N	O	S	0	0	0
			1974	1249	341	374	10			

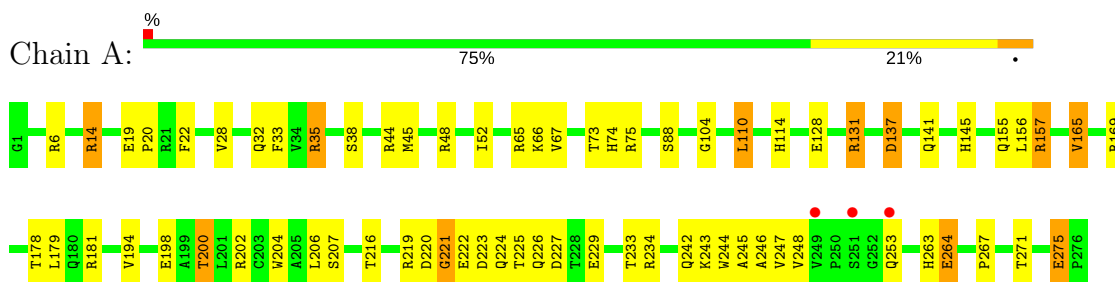
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	12	Total	O	0	0
			12	12		
6	B	2	Total	O	0	0
			2	2		
6	D	14	Total	O	0	0
			14	14		
6	E	13	Total	O	0	0
			13	13		
6	F	20	Total	O	0	0
			20	20		
6	G	2	Total	O	0	0
			2	2		
6	I	11	Total	O	0	0
			11	11		
6	J	8	Total	O	0	0
			8	8		

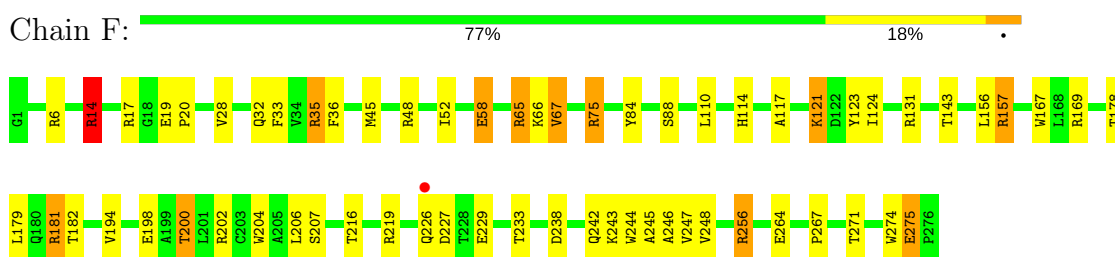
### 3 Residue-property plots [i](#)

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.

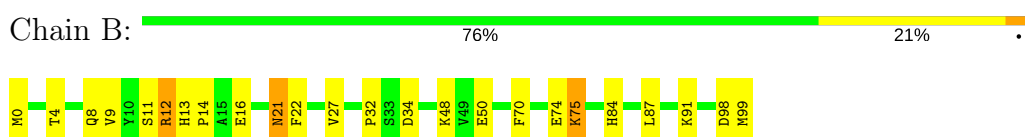
- Molecule 1: HLA class I histocompatibility antigen, A-2 alpha chain



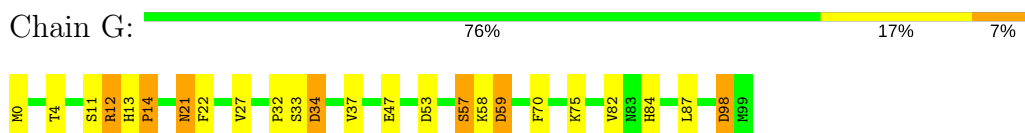
- Molecule 1: HLA class I histocompatibility antigen, A-2 alpha chain



- Molecule 2: Beta-2-microglobulin

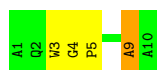


- Molecule 2: Beta-2-microglobulin

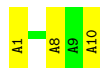


- Molecule 3: ALA-GLN-TRP-GLY-PRO-ASP-PRO-ALA-ALA-ALA

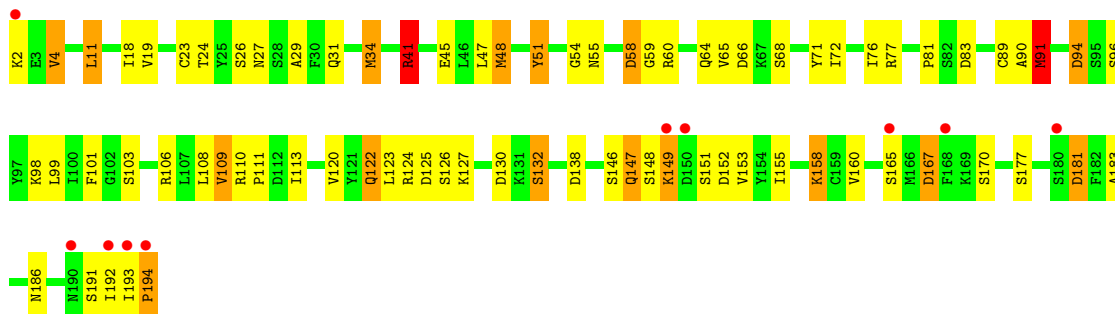




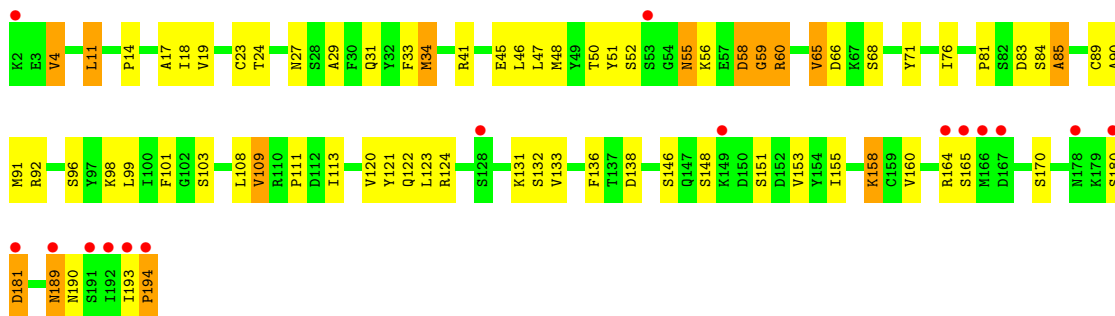
- Molecule 3: ALA-GLN-TRP-GLY-PRO-ASP-PRO-ALA-ALA-ALA



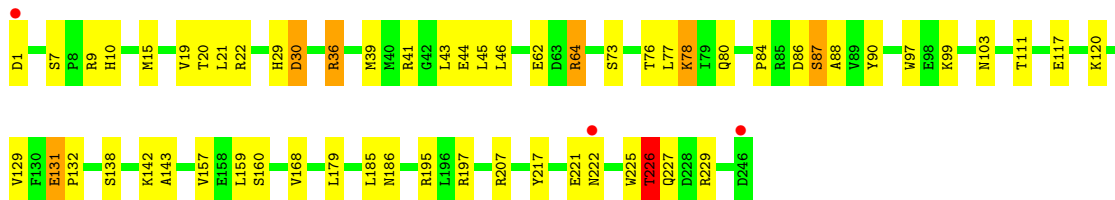
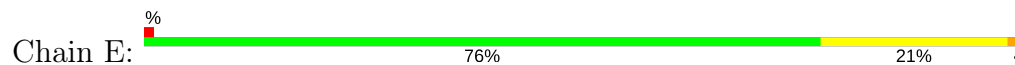
- Molecule 4: Human T-cell Receptor, Class I, Light alpha Chain



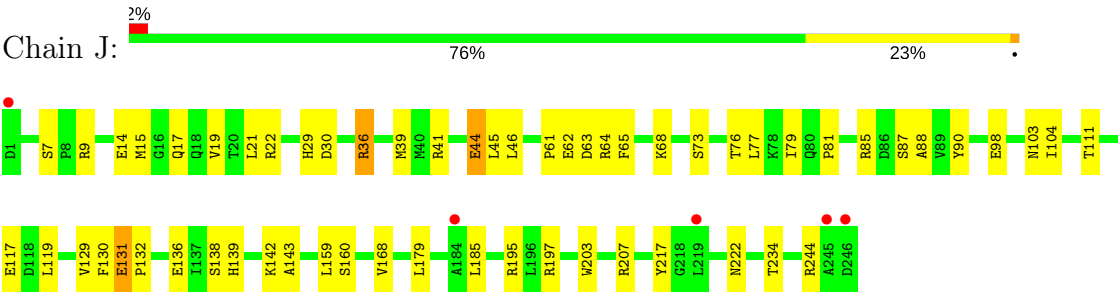
- Molecule 4: Human T-cell Receptor, Class I, Light alpha Chain



- Molecule 5: Human T-cell Receptor, Class I, Heavy beta Chain



- Molecule 5: Human T-cell Receptor, Class I, Heavy beta Chain





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	41.44Å 97.31Å 121.37Å 97.29° 97.66° 92.73°	Depositor
Resolution (Å)	38.63 – 3.06 38.63 – 3.06	Depositor EDS
% Data completeness (in resolution range)	97.2 (38.63-3.06) 93.0 (38.63-3.06)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.36 (at 3.06Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.212 , 0.293 0.213 , 0.291	Depositor DCC
$R_{free}$ test set	1726 reflections (5.34%)	DCC
Wilson B-factor (Å <sup>2</sup> )	56.0	Xtriage
Anisotropy	0.285	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 55.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	13392	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.76% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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	0.89	0/2320	1.05	10/3149 (0.3%)
1	F	0.95	2/2320 (0.1%)	1.09	15/3149 (0.5%)
2	B	0.90	0/860	1.02	2/1162 (0.2%)
2	G	0.97	1/860 (0.1%)	1.12	5/1162 (0.4%)
3	C	1.14	0/73	0.96	0/100
3	H	1.15	0/73	1.13	0/100
4	D	0.91	1/1553 (0.1%)	1.08	10/2100 (0.5%)
4	I	0.85	1/1553 (0.1%)	1.01	4/2100 (0.2%)
5	E	0.84	1/2029 (0.0%)	1.00	4/2759 (0.1%)
5	J	0.75	1/2029 (0.0%)	0.97	6/2759 (0.2%)
All	All	0.88	7/13670 (0.1%)	1.04	56/18540 (0.3%)

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	A	0	2
2	G	0	1
3	C	0	1
3	H	0	1
4	D	0	2
4	I	0	2
5	E	0	2
All	All	0	11

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	194	PRO	C-O	10.49	1.44	1.23
4	I	194	PRO	C-O	10.06	1.43	1.23
1	F	58	GLU	CB-CG	7.30	1.66	1.52

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	58	GLU	CG-CD	7.08	1.62	1.51
5	J	44	GLU	CD-OE1	5.14	1.31	1.25

The worst 5 of 56 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	14	ARG	NE-CZ-NH2	-10.03	115.28	120.30
1	A	44	ARG	NE-CZ-NH1	8.98	124.79	120.30
2	G	12	ARG	NE-CZ-NH2	-8.25	116.17	120.30
1	F	14	ARG	NE-CZ-NH1	8.18	124.39	120.30
1	A	48	ARG	NE-CZ-NH2	8.00	124.30	120.30

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	137	ASP	Peptide
1	A	226	GLN	Peptide
3	C	9	ALA	Peptide
4	D	149	LYS	Peptide
4	D	158	LYS	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	2254	0	2103	26	0
1	F	2254	0	2103	21	0
2	B	837	0	803	12	0
2	G	837	0	803	10	0
3	C	70	0	61	4	0
3	H	70	0	61	2	0
4	D	1520	0	1447	36	0
4	I	1520	0	1447	30	0
5	E	1974	0	1887	28	0
5	J	1974	0	1887	25	0
6	A	12	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	2	0	0	0	0
6	D	14	0	0	1	0
6	E	13	0	0	1	0
6	F	20	0	0	1	0
6	G	2	0	0	1	0
6	I	11	0	0	0	0
6	J	8	0	0	0	0
All	All	13392	0	12602	174	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 7.

The worst 5 of 174 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:41:ARG:NH2	5:E:157:VAL:O	2.06	0.87
4:I:34:MET:HE2	5:J:103:ASN:HB2	1.58	0.85
2:B:21:ASN:HD22	2:B:22:PHE:H	1.27	0.82
4:D:183:ALA:HB3	4:D:186:ASN:HD21	1.46	0.80
5:E:15:MET:CE	5:E:117:GLU:HA	2.15	0.77

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 42
1	F	274/276 (99%)	253 (92%)	19 (7%)	2 (1%)	25 61
2	B	98/100 (98%)	91 (93%)	6 (6%)	1 (1%)	18 52
2	G	98/100 (98%)	96 (98%)	1 (1%)	1 (1%)	18 52

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	C	8/10 (80%)	8 (100%)	0	0	100	100
3	H	8/10 (80%)	7 (88%)	1 (12%)	0	100	100
4	D	191/193 (99%)	166 (87%)	22 (12%)	3 (2%)	11	40
4	I	191/193 (99%)	167 (87%)	19 (10%)	5 (3%)	6	27
5	E	244/246 (99%)	226 (93%)	17 (7%)	1 (0%)	38	72
5	J	244/246 (99%)	225 (92%)	19 (8%)	0	100	100
All	All	1630/1650 (99%)	1484 (91%)	129 (8%)	17 (1%)	18	52

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	221	GLY
1	A	137	ASP
4	D	181	ASP
4	I	158	LYS
4	D	158	LYS

### 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	232/232 (100%)	206 (89%)	26 (11%)	7	26
1	F	232/232 (100%)	207 (89%)	25 (11%)	7	27
2	B	95/95 (100%)	84 (88%)	11 (12%)	6	24
2	G	95/95 (100%)	84 (88%)	11 (12%)	6	24
3	C	5/5 (100%)	5 (100%)	0	100	100
3	H	5/5 (100%)	5 (100%)	0	100	100
4	D	174/174 (100%)	137 (79%)	37 (21%)	1	5
4	I	174/174 (100%)	139 (80%)	35 (20%)	1	5
5	E	216/216 (100%)	191 (88%)	25 (12%)	6	24
5	J	216/216 (100%)	191 (88%)	25 (12%)	6	24

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1444/1444 (100%)	1249 (86%)	195 (14%)	<b>4</b> <b>17</b>

5 of 195 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	E	131	GLU
1	F	121	LYS
5	J	68	LYS
5	E	168	VAL
1	F	14	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
5	E	29	HIS
1	F	174	ASN
4	I	189	ASN
5	E	141	GLN
1	F	54	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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	276/276 (100%)	-0.18	3 (1%) 80 61	45, 65, 108, 138	0
1	F	276/276 (100%)	-0.18	1 (0%) 92 81	30, 57, 123, 169	0
2	B	100/100 (100%)	-0.33	0 100 100	47, 60, 88, 109	0
2	G	100/100 (100%)	-0.44	0 100 100	38, 55, 76, 91	0
3	C	10/10 (100%)	-0.40	0 100 100	47, 54, 56, 61	0
3	H	10/10 (100%)	-0.22	0 100 100	42, 50, 59, 60	0
4	D	193/193 (100%)	0.17	10 (5%) 28 12	46, 77, 130, 155	0
4	I	193/193 (100%)	0.30	16 (8%) 12 4	51, 86, 144, 173	0
5	E	246/246 (100%)	-0.18	3 (1%) 79 59	34, 63, 120, 143	0
5	J	246/246 (100%)	0.03	5 (2%) 65 42	49, 80, 117, 151	0
All	All	1650/1650 (100%)	-0.08	38 (2%) 61 37	30, 67, 126, 173	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	150	ASP	7.5
4	D	194	PRO	6.7
4	I	180	SER	4.9
4	I	194	PRO	4.8
4	I	53	SER	4.7

### 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.