



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 08:32 pm BST

PDB ID : 5WKF  
Title : D30 TCR in complex with HLA-A\*11:01-GTS1  
Authors : Gras, S.; Rossjohn, J.  
Deposited on : 2017-07-25  
Resolution : 2.95 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

<https://www.wwpdb.org/validation/2017/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) : 1.13  
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

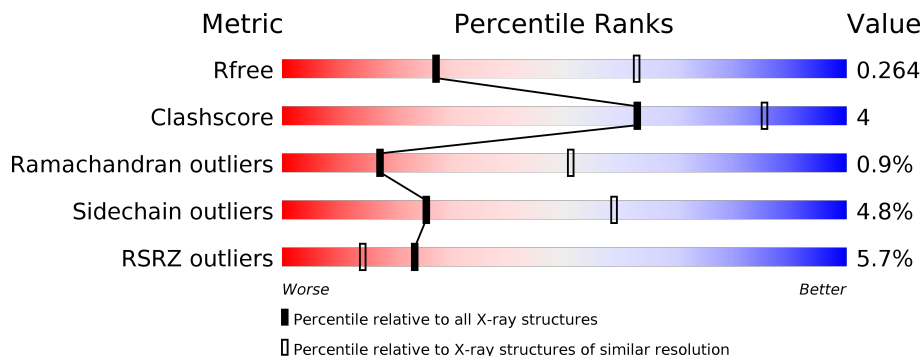
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*


The reported resolution of this entry is 2.95 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.



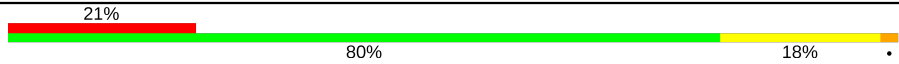

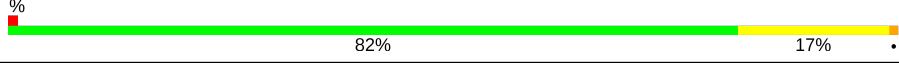

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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 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  $\leq 5\%$ . 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	274	
1	F	274	
2	B	100	
2	G	100	
3	C	10	
3	H	10	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	198	 <p>21% 80% 18%</p>
4	I	198	 <p>18% 78% 20%</p>
5	E	244	 <p>1% 82% 17%</p>
5	J	244	 <p>2% 83% 15%</p>

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 13580 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
			Total	C	N	O	S			
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
			Total	C	N	O	S			
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 GTS1 peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	10	Total	C	N	O	0	0	0
			69	40	14	15			
3	H	10	Total	C	N	O	0	0	0
			69	40	14	15			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	198	Total	C	N	O	S	0	0	0
			1536	962	259	305	10			
4	I	198	Total	C	N	O	S	0	0	0
			1536	962	259	305	10			

There are 46 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	20	ILE	VAL	conflict	UNP A0A087WSZ9
D	44	ILE	VAL	conflict	UNP A0A087WSZ9
D	91	LEU	-	linker	UNP A0A087WSZ9
D	92	GLY	-	linker	UNP A0A087WSZ9
D	93	ASP	-	linker	UNP A0A087WSZ9
D	94	ALA	-	linker	UNP A0A087WSZ9
D	95	GLY	-	linker	UNP A0A087WSZ9
D	96	ASN	-	linker	UNP A0A087WSZ9
D	97	MET	-	linker	UNP A0A087WSZ9
D	98	LEU	-	linker	UNP A0A087WSZ9
D	99	THR	-	linker	UNP A0A087WSZ9
D	100	PHE	-	linker	UNP A0A087WSZ9
D	101	GLY	-	linker	UNP A0A087WSZ9
D	102	GLY	-	linker	UNP A0A087WSZ9
D	103	GLY	-	linker	UNP A0A087WSZ9
D	104	THR	-	linker	UNP A0A087WSZ9
D	105	ARG	-	linker	UNP A0A087WSZ9
D	106	LEU	-	linker	UNP A0A087WSZ9
D	107	MET	-	linker	UNP A0A087WSZ9
D	108	VAL	-	linker	UNP A0A087WSZ9
D	109	LYS	-	linker	UNP A0A087WSZ9
D	110	PRO	-	linker	UNP A0A087WSZ9
D	111	HIS	-	linker	UNP A0A087WSZ9
I	20	ILE	VAL	conflict	UNP A0A087WSZ9
I	51	ILE	VAL	conflict	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

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
I	113	GLY	-	linker	UNP A0A087WSZ9
I	114	GLY	-	linker	UNP A0A087WSZ9
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
			Total	C	N	O	S			
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			

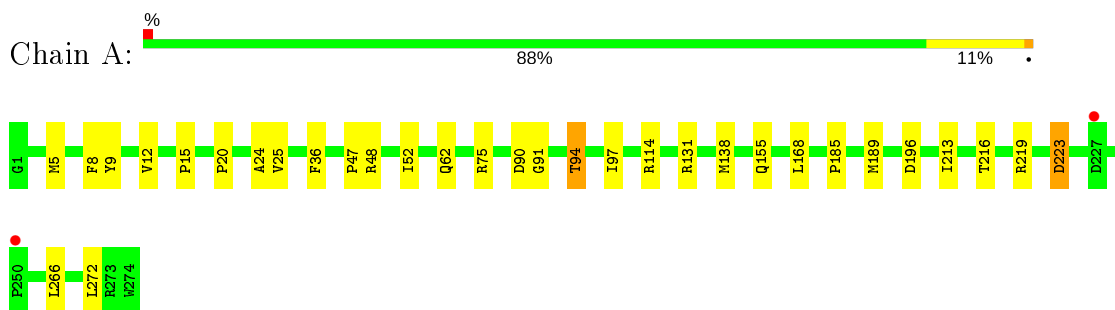
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	60	Total	O	0	0
			60	60		
6	B	28	Total	O	0	0
			28	28		
6	C	4	Total	O	0	0
			4	4		
6	D	35	Total	O	0	0
			35	35		
6	E	36	Total	O	0	0
			36	36		
6	F	69	Total	O	0	0
			69	69		
6	G	34	Total	O	0	0
			34	34		
6	H	2	Total	O	0	0
			2	2		
6	I	31	Total	O	0	0
			31	31		
6	J	65	Total	O	0	0
			65	65		

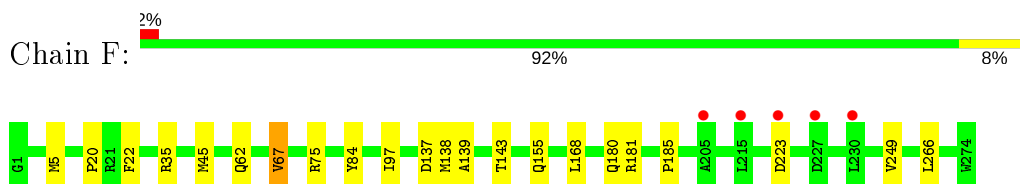
### 3 Residue-property plots [i](#)

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.

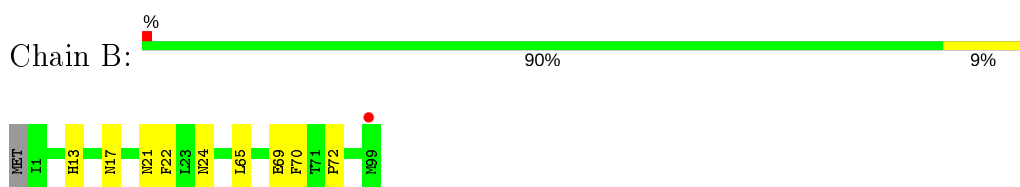
- Molecule 1: HLA class I histocompatibility antigen, A-11 alpha chain



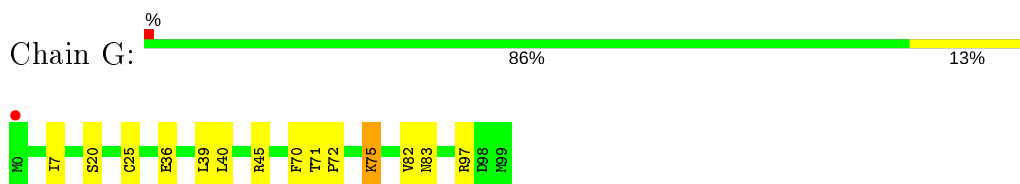
- Molecule 1: HLA class I histocompatibility antigen, A-11 alpha chain



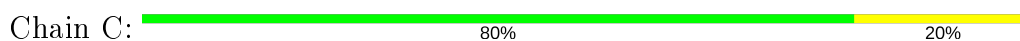
- Molecule 2: Beta-2-microglobulin



- Molecule 2: Beta-2-microglobulin



- Molecule 3: GTS1 peptide

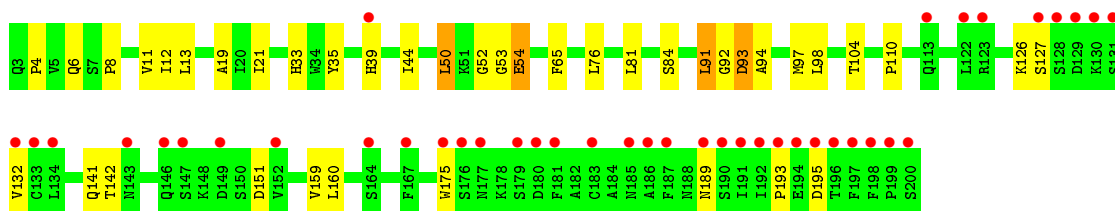
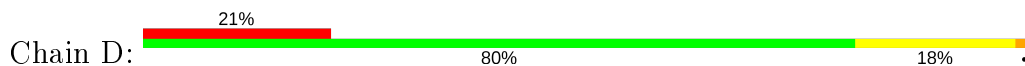




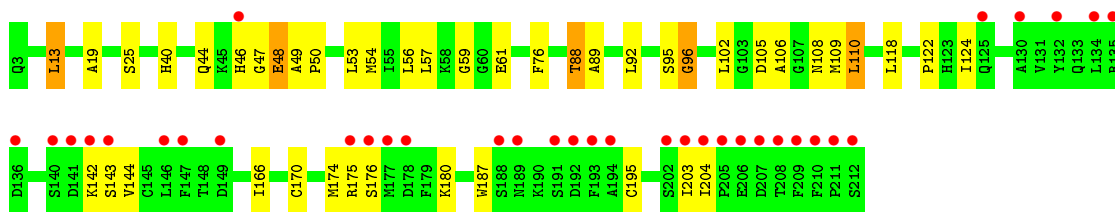
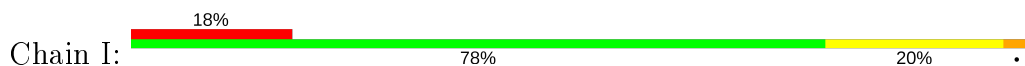
- Molecule 3: GTS1 peptide



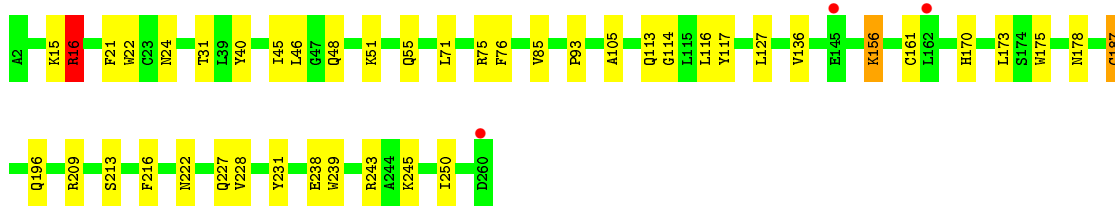
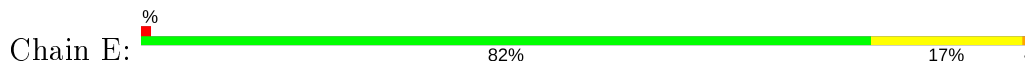
- Molecule 4: T-cell receptor alpha variable 30,T-cell receptor, sp3.4 alpha chain Chimera



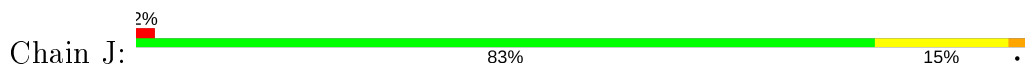
- Molecule 4: T-cell receptor alpha variable 30,T-cell receptor, sp3.4 alpha chain Chimera



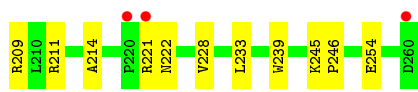
- Molecule 5: D30 TCR beta chain



- Molecule 5: D30 TCR beta chain







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	88.06Å 146.38Å 167.56Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.17 – 2.95 47.17 – 2.95	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.17-2.95) 100.0 (47.17-2.95)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.04 (at 2.96Å)	Xtrriage
Refinement program	BUSTER 2.10.3	Depositor
R, $R_{free}$	0.190 , 0.253 0.201 , 0.264	Depositor DCC
$R_{free}$ test set	2276 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.0	Xtrriage
Anisotropy	0.587	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 48.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	13580	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.51% 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 [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	0.36	0/2296	0.61	0/3117
1	F	0.37	0/2296	0.61	0/3117
2	B	0.36	0/852	0.59	0/1152
2	G	0.37	0/860	0.63	0/1162
3	C	0.37	0/69	0.74	0/91
3	H	0.40	0/69	0.80	0/91
4	D	0.47	0/1569	0.70	0/2117
4	I	0.41	0/1569	0.69	0/2117
5	E	0.37	0/1986	0.63	0/2702
5	J	0.39	0/1986	0.65	0/2702
All	All	0.39	0/13552	0.64	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 [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	2235	0	2087	16	0
1	F	2235	0	2087	10	0
2	B	829	0	794	4	0
2	G	837	0	803	7	0
3	C	69	0	71	2	0
3	H	69	0	71	3	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	1536	0	1479	26	0
4	I	1536	0	1477	21	0
5	E	1935	0	1874	20	0
5	J	1935	0	1874	19	0
6	A	60	0	0	0	0
6	B	28	0	0	0	0
6	C	4	0	0	0	0
6	D	35	0	0	0	0
6	E	36	0	0	0	0
6	F	69	0	0	1	0
6	G	34	0	0	0	0
6	H	2	0	0	0	0
6	I	31	0	0	0	0
6	J	65	0	0	0	0
All	All	13580	0	12617	111	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 4.

The worst 5 of 111 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:53:GLY:HA2	4:D:54:GLU:HB2	1.26	1.11
1:A:62:GLN:HE22	4:D:94:ALA:H	1.06	0.97
4:D:53:GLY:HA2	4:D:54:GLU:CB	2.09	0.82
4:D:53:GLY:CA	4:D:54:GLU:HB2	2.10	0.78
1:F:155:GLN:NE2	5:J:114:GLY:O	2.21	0.74

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%)	265 (97%)	7 (3%)	0	100	100
1	F	272/274 (99%)	260 (96%)	12 (4%)	0	100	100
2	B	97/100 (97%)	93 (96%)	4 (4%)	0	100	100
2	G	98/100 (98%)	93 (95%)	5 (5%)	0	100	100
3	C	8/10 (80%)	7 (88%)	1 (12%)	0	100	100
3	H	8/10 (80%)	7 (88%)	1 (12%)	0	100	100
4	D	196/198 (99%)	174 (89%)	17 (9%)	5 (3%)	5	24
4	I	196/198 (99%)	179 (91%)	11 (6%)	6 (3%)	4	19
5	E	242/244 (99%)	228 (94%)	12 (5%)	2 (1%)	19	53
5	J	242/244 (99%)	226 (93%)	14 (6%)	2 (1%)	19	53
All	All	1631/1652 (99%)	1532 (94%)	84 (5%)	15 (1%)	17	51

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	54	GLU
4	I	176	SER
5	J	81	LEU
4	D	92	GLY
4	I	48	GLU

### 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%)	220 (95%)	11 (5%)	25	59
1	F	231/231 (100%)	222 (96%)	9 (4%)	32	65
2	B	94/95 (99%)	93 (99%)	1 (1%)	73	89
2	G	95/95 (100%)	93 (98%)	2 (2%)	53	80
3	C	8/8 (100%)	8 (100%)	0	100	100
3	H	8/8 (100%)	7 (88%)	1 (12%)	4	17

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	D	173/173 (100%)	162 (94%)	11 (6%)	17	47
4	I	173/173 (100%)	163 (94%)	10 (6%)	20	51
5	E	209/209 (100%)	199 (95%)	10 (5%)	25	59
5	J	209/209 (100%)	195 (93%)	14 (7%)	16	45
All	All	1431/1432 (100%)	1362 (95%)	69 (5%)	25	59

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	E	216	PHE
1	F	180	GLN
5	J	162	LEU
5	E	238	GLU
1	F	67	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
5	E	55	GLN
5	E	219	ASN
5	J	113	GLN
5	E	113	GLN
5	E	196	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, 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	274/274 (100%)	-0.19	2 (0%) 87 76	18, 37, 71, 96	0
1	F	274/274 (100%)	-0.14	5 (1%) 68 51	16, 36, 85, 100	0
2	B	99/100 (99%)	-0.05	1 (1%) 82 68	23, 48, 72, 81	0
2	G	100/100 (100%)	-0.35	1 (1%) 82 68	22, 39, 63, 71	0
3	C	10/10 (100%)	0.12	0 100 100	23, 24, 30, 47	0
3	H	10/10 (100%)	-0.33	0 100 100	17, 19, 26, 37	0
4	D	198/198 (100%)	0.91	41 (20%) 1 0	25, 59, 130, 163	1 (0%)
4	I	198/198 (100%)	1.04	35 (17%) 1 1	18, 54, 137, 166	1 (0%)
5	E	244/244 (100%)	0.01	3 (1%) 79 63	18, 41, 86, 102	6 (2%)
5	J	244/244 (100%)	-0.07	6 (2%) 57 40	15, 39, 76, 106	6 (2%)
All	All	1651/1652 (99%)	0.15	94 (5%) 23 14	15, 41, 96, 166	14 (0%)

The worst 5 of 94 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	I	205	PRO	17.3
4	I	207	ASP	14.5
4	I	208	THR	13.6
4	I	212	SER	13.2
4	I	206	GLU	9.9

### 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.