



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 13, 2017 – 07:46 PM EDT

PDB ID : 5FT1  
Title : Crystal structure of gp37(Dip) from bacteriophage phiKZ bound to RNase E of *Pseudomonas aeruginosa*  
Authors : Van den Bossche, A.; Hardwick, S.W.; Ceyssens, P.J.; Hendrix, H.; Voet, M.; Dendooven, T.; Bandyra, K.J.; De Maeyer, M.; Aertsen, A.; Noben, J.P.; Luisi, B.F.; Lavigne, R.  
Deposited on : 2016-01-08  
Resolution : 2.75 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	:	rb-20029077
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)	:	rb-20029077

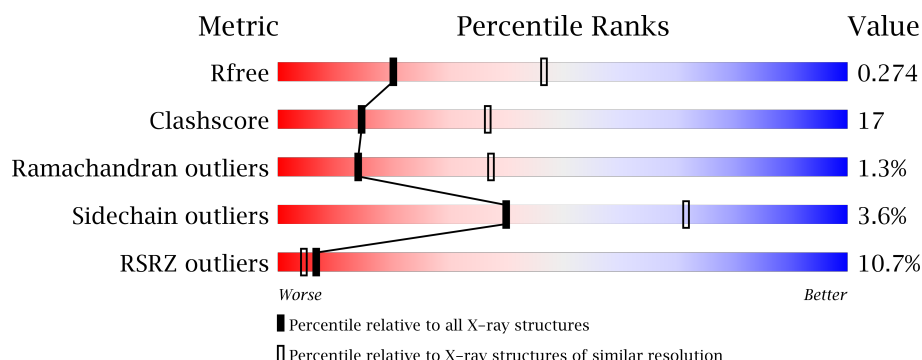
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.75 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	3666 (2.80-2.72)
Clashscore	112137	4174 (2.80-2.72)
Ramachandran outliers	110173	4103 (2.80-2.72)
Sidechain outliers	110143	4106 (2.80-2.72)
RSRZ outliers	101464	3697 (2.80-2.72)

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	281	<div> <div>11%</div> <div> <div></div> <div>56%</div> <div>30%</div> <div>•</div> <div>11%</div> </div> </div>
1	B	281	<div> <div>7%</div> <div> <div></div> <div>57%</div> <div>27%</div> <div>•</div> <div>14%</div> </div> </div>
1	E	281	<div> <div>9%</div> <div> <div></div> <div>59%</div> <div>27%</div> <div>•</div> <div>13%</div> </div> </div>
1	G	281	<div> <div>12%</div> <div> <div></div> <div>65%</div> <div>23%</div> <div>•</div> <div>11%</div> </div> </div>
1	I	281	<div> <div>10%</div> <div> <div></div> <div>58%</div> <div>30%</div> <div>•</div> <div>12%</div> </div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	K	281	
2	C	20	
2	D	20	
2	F	20	
2	H	20	
2	J	20	
2	L	20	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11561 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 GP37.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	250	Total	C	N	O	S	0	0	0
			1891	1217	306	363	5			
1	B	243	Total	C	N	O	S	0	0	0
			1867	1205	301	354	7			
1	E	245	Total	C	N	O	S	0	0	0
			1843	1191	299	346	7			
1	G	249	Total	C	N	O	S	0	0	0
			1892	1218	306	361	7			
1	I	248	Total	C	N	O	S	0	0	0
			1869	1207	298	358	6			
1	K	240	Total	C	N	O	S	0	0	0
			1851	1195	298	352	6			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	274	LYS	-	EXPRESSION TAG	UNP Q8SDC5
A	275	GLY	-	EXPRESSION TAG	UNP Q8SDC5
A	276	HIS	-	EXPRESSION TAG	UNP Q8SDC5
A	277	HIS	-	EXPRESSION TAG	UNP Q8SDC5
A	278	HIS	-	EXPRESSION TAG	UNP Q8SDC5
A	279	HIS	-	EXPRESSION TAG	UNP Q8SDC5
A	280	HIS	-	EXPRESSION TAG	UNP Q8SDC5
A	281	HIS	-	EXPRESSION TAG	UNP Q8SDC5
B	274	LYS	-	EXPRESSION TAG	UNP Q8SDC5
B	275	GLY	-	EXPRESSION TAG	UNP Q8SDC5
B	276	HIS	-	EXPRESSION TAG	UNP Q8SDC5
B	277	HIS	-	EXPRESSION TAG	UNP Q8SDC5
B	278	HIS	-	EXPRESSION TAG	UNP Q8SDC5
B	279	HIS	-	EXPRESSION TAG	UNP Q8SDC5
B	280	HIS	-	EXPRESSION TAG	UNP Q8SDC5
B	281	HIS	-	EXPRESSION TAG	UNP Q8SDC5
E	274	LYS	-	EXPRESSION TAG	UNP Q8SDC5

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	275	GLY	-	EXPRESSION TAG	UNP Q8SDC5
E	276	HIS	-	EXPRESSION TAG	UNP Q8SDC5
E	277	HIS	-	EXPRESSION TAG	UNP Q8SDC5
E	278	HIS	-	EXPRESSION TAG	UNP Q8SDC5
E	279	HIS	-	EXPRESSION TAG	UNP Q8SDC5
E	280	HIS	-	EXPRESSION TAG	UNP Q8SDC5
E	281	HIS	-	EXPRESSION TAG	UNP Q8SDC5
G	274	LYS	-	EXPRESSION TAG	UNP Q8SDC5
G	275	GLY	-	EXPRESSION TAG	UNP Q8SDC5
G	276	HIS	-	EXPRESSION TAG	UNP Q8SDC5
G	277	HIS	-	EXPRESSION TAG	UNP Q8SDC5
G	278	HIS	-	EXPRESSION TAG	UNP Q8SDC5
G	279	HIS	-	EXPRESSION TAG	UNP Q8SDC5
G	280	HIS	-	EXPRESSION TAG	UNP Q8SDC5
G	281	HIS	-	EXPRESSION TAG	UNP Q8SDC5
I	274	LYS	-	EXPRESSION TAG	UNP Q8SDC5
I	275	GLY	-	EXPRESSION TAG	UNP Q8SDC5
I	276	HIS	-	EXPRESSION TAG	UNP Q8SDC5
I	277	HIS	-	EXPRESSION TAG	UNP Q8SDC5
I	278	HIS	-	EXPRESSION TAG	UNP Q8SDC5
I	279	HIS	-	EXPRESSION TAG	UNP Q8SDC5
I	280	HIS	-	EXPRESSION TAG	UNP Q8SDC5
I	281	HIS	-	EXPRESSION TAG	UNP Q8SDC5
K	274	LYS	-	EXPRESSION TAG	UNP Q8SDC5
K	275	GLY	-	EXPRESSION TAG	UNP Q8SDC5
K	276	HIS	-	EXPRESSION TAG	UNP Q8SDC5
K	277	HIS	-	EXPRESSION TAG	UNP Q8SDC5
K	278	HIS	-	EXPRESSION TAG	UNP Q8SDC5
K	279	HIS	-	EXPRESSION TAG	UNP Q8SDC5
K	280	HIS	-	EXPRESSION TAG	UNP Q8SDC5
K	281	HIS	-	EXPRESSION TAG	UNP Q8SDC5

- Molecule 2 is a protein called RIBONUCLEASE E.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	9	Total	C	N	O	0	0	0
			61	34	16	11			
2	D	9	Total	C	N	O	0	0	0
			62	35	18	9			
2	F	6	Total	C	N	O	0	0	0
			43	24	12	7			
2	H	8	Total	C	N	O	0	0	0
			63	35	18	10			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	J	10	Total	C	N	O	0	0	0
			72	40	20	12			
2	L	6	Total	C	N	O	0	0	0
			46	25	13	8			

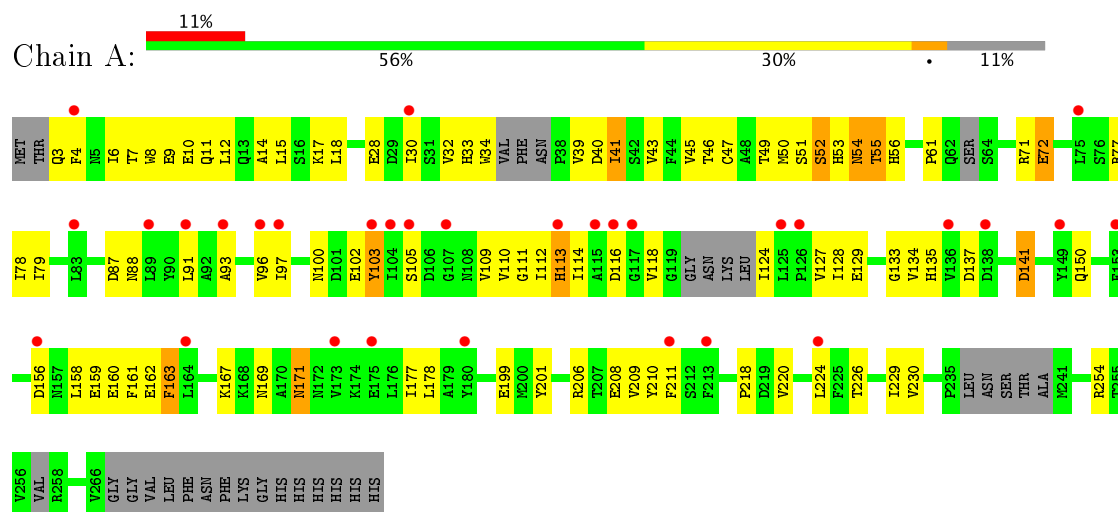
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	O	0	0
			1	1		

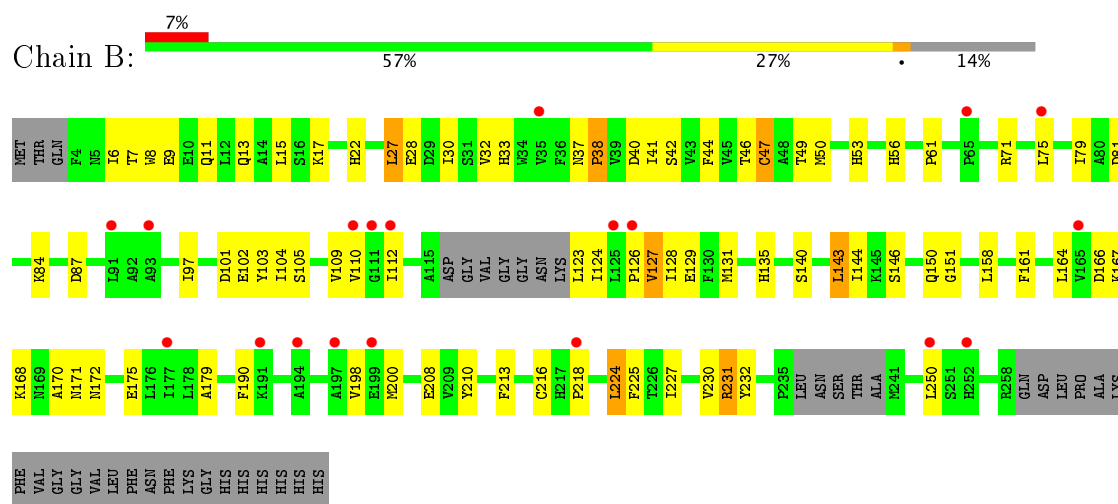
### 3 Residue-property plots

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: GP37

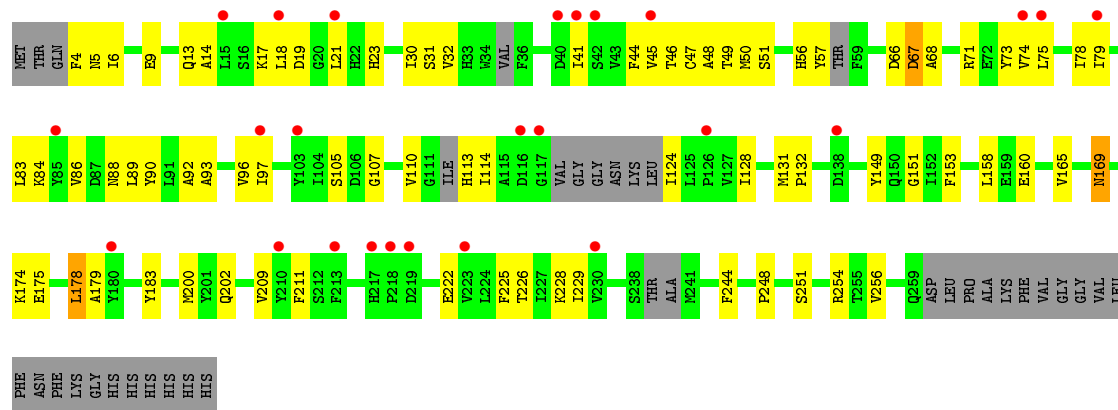


#### • Molecule 1: GP37



#### • Molecule 1: GP37





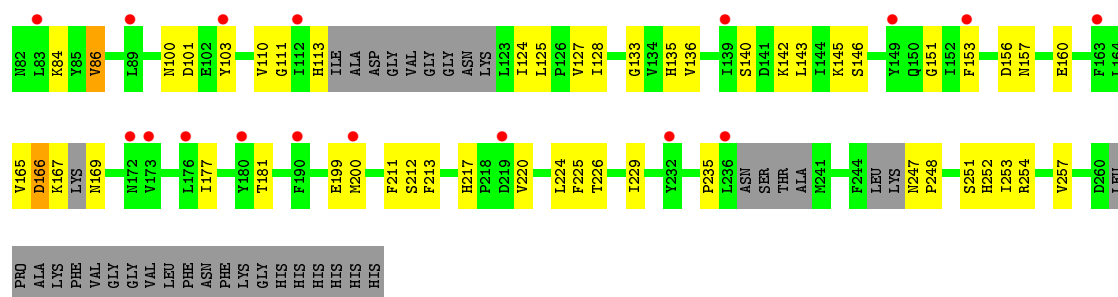
• Molecule 1: GP37



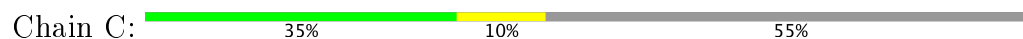
• Molecule 1: GP37



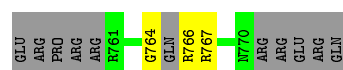
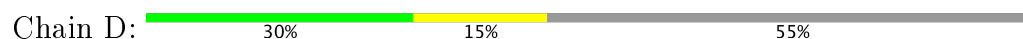




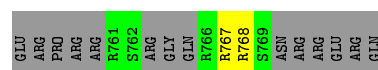
• Molecule 2: RIBONUCLEASE E



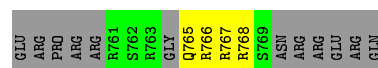
• Molecule 2: RIBONUCLEASE E



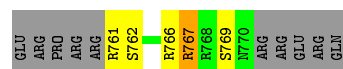
• Molecule 2: RIBONUCLEASE E



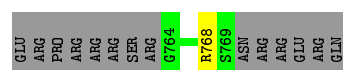
• Molecule 2: RIBONUCLEASE E



• Molecule 2: RIBONUCLEASE E



• Molecule 2: RIBONUCLEASE E



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	84.61Å 84.65Å 84.60Å 107.86° 107.88° 107.90°	Depositor
Resolution (Å)	72.19 – 2.75 72.19 – 2.75	Depositor EDS
% Data completeness (in resolution range)	79.8 (72.19-2.75) 77.5 (72.19-2.75)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.07 (at 2.73Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.263 , 0.293 0.238 , 0.274	Depositor DCC
$R_{free}$ test set	1989 reflections (5.46%)	DCC
Wilson B-factor (Å <sup>2</sup> )	83.6	Xtriage
Anisotropy	0.034	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 69.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.467 for k,l,h 0.467 for l,h,k 0.476 for -l,-k,-h 0.466 for -h,-l,-k 0.467 for -k,-h,-l	Xtriage
Reported twinning fraction	0.500 for -L,-K,-H	Depositor
Outliers	0 of 39553 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	11561	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.24% 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	0.28	0/1935	0.48	0/2643
1	B	0.31	0/1913	0.46	0/2614
1	E	0.28	0/1887	0.49	0/2576
1	G	0.30	0/1936	0.49	1/2641 (0.0%)
1	I	0.29	0/1911	0.48	0/2608
1	K	0.28	0/1894	0.47	0/2579
2	C	0.20	0/59	0.43	0/75
2	D	0.19	0/60	0.36	0/76
2	F	0.19	0/41	0.27	0/51
2	H	0.24	0/61	0.49	0/77
2	J	0.24	0/71	0.43	0/92
2	L	0.19	0/45	0.36	0/57
All	All	0.29	0/11813	0.48	1/16089 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	91	LEU	CA-CB-CG	5.79	128.63	115.30

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	1891	0	1705	86	0
1	B	1867	0	1722	56	0
1	E	1843	0	1650	62	0
1	G	1892	0	1711	56	0
1	I	1869	0	1691	70	0
1	K	1851	0	1694	61	0
2	C	61	0	48	3	0
2	D	62	0	50	1	0
2	F	43	0	35	1	0
2	H	63	0	56	6	0
2	J	72	0	62	3	0
2	L	46	0	43	1	0
3	A	1	0	0	0	0
All	All	11561	0	10467	376	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 17.

The worst 5 of 376 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)
1:A:55:THR:O	1:A:56:HIS:ND1	1.80	1.14
1:A:51:SER:CB	1:A:52:SER:HA	2.05	0.87
1:A:102:GLU:OE1	1:A:113:HIS:NE2	2.09	0.85
1:B:168:LYS:O	1:B:171:ASN:ND2	2.09	0.85
1:A:30:ILE:HG23	1:A:47:CYS:HB3	1.58	0.85

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	238/281 (85%)	218 (92%)	17 (7%)	3 (1%)	14	37
1	B	237/281 (84%)	222 (94%)	13 (6%)	2 (1%)	22	52
1	E	233/281 (83%)	220 (94%)	10 (4%)	3 (1%)	14	37
1	G	239/281 (85%)	224 (94%)	13 (5%)	2 (1%)	22	52
1	I	237/281 (84%)	221 (93%)	12 (5%)	4 (2%)	11	30
1	K	226/281 (80%)	211 (93%)	13 (6%)	2 (1%)	20	49
2	C	5/20 (25%)	2 (40%)	3 (60%)	0	100	100
2	D	5/20 (25%)	2 (40%)	2 (40%)	1 (20%)	0	0
2	F	2/20 (10%)	0	1 (50%)	1 (50%)	0	0
2	H	4/20 (20%)	3 (75%)	1 (25%)	0	100	100
2	J	8/20 (40%)	5 (62%)	3 (38%)	0	100	100
2	L	4/20 (20%)	3 (75%)	1 (25%)	0	100	100
All	All	1438/1806 (80%)	1331 (93%)	89 (6%)	18 (1%)	14	37

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	41	ILE
1	A	114	ILE
1	I	55	THR
1	I	54	ASN
1	A	113	HIS

### 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	190/251 (76%)	181 (95%)	9 (5%)	30	61
1	B	194/251 (77%)	185 (95%)	9 (5%)	31	62
1	E	182/251 (72%)	176 (97%)	6 (3%)	43	74
1	G	191/251 (76%)	187 (98%)	4 (2%)	59	85

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	I	187/251 (74%)	182 (97%)	5 (3%)	50	80
1	K	192/251 (76%)	185 (96%)	7 (4%)	40	71
2	C	4/19 (21%)	4 (100%)	0	100	100
2	D	3/19 (16%)	3 (100%)	0	100	100
2	F	3/19 (16%)	3 (100%)	0	100	100
2	H	5/19 (26%)	5 (100%)	0	100	100
2	J	5/19 (26%)	3 (60%)	2 (40%)	0	0
2	L	4/19 (21%)	4 (100%)	0	100	100
All	All	1160/1620 (72%)	1118 (96%)	42 (4%)	40	71

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	44	PHE
1	E	244	PHE
1	K	44	PHE
1	E	67	ASP
1	E	169	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	62	GLN
1	G	186	ASN
1	I	186	ASN
1	A	172	ASN
1	I	169	ASN

### 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 [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 ⓘ

### 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<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	250/281 (88%)	0.54	31 (12%) 4 3	61, 100, 134, 162	0
1	B	243/281 (86%)	0.43	19 (7%) 14 10	59, 97, 127, 144	0
1	E	245/281 (87%)	0.48	25 (10%) 7 5	61, 93, 132, 161	0
1	G	249/281 (88%)	0.67	34 (13%) 3 2	67, 96, 136, 152	0
1	I	248/281 (88%)	0.48	27 (10%) 6 4	64, 98, 131, 140	0
1	K	240/281 (85%)	0.59	27 (11%) 6 4	62, 94, 127, 150	0
2	C	9/20 (45%)	-0.22	0 100 100	83, 106, 120, 121	0
2	D	9/20 (45%)	-0.18	0 100 100	107, 112, 135, 137	0
2	F	6/20 (30%)	0.25	0 100 100	106, 121, 128, 128	0
2	H	8/20 (40%)	0.09	0 100 100	93, 108, 121, 145	0
2	J	10/20 (50%)	0.00	0 100 100	91, 104, 142, 145	0
2	L	6/20 (30%)	-0.44	0 100 100	107, 117, 131, 133	0
All	All	1523/1806 (84%)	0.51	163 (10%) 7 4	59, 97, 132, 162	0

The worst 5 of 163 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	83	LEU	8.1
1	A	116	ASP	7.6
1	A	115	ALA	7.4
1	G	89	LEU	6.4
1	A	117	GLY	6.0

### 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 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.