



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 9, 2016 – 05:29 AM EDT

PDB ID : 5BYV  
Title : Crystal structure of MSM-13, a putative T1-like thiolase from Mycobacterium smegmatis  
Authors : Janardan, N.; Harijan, R.K.; Keima, T.R.; Wierenga, R.; Murthy, M.R.N.  
Deposited on : 2015-06-11  
Resolution : 2.16 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  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027457  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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-20027457

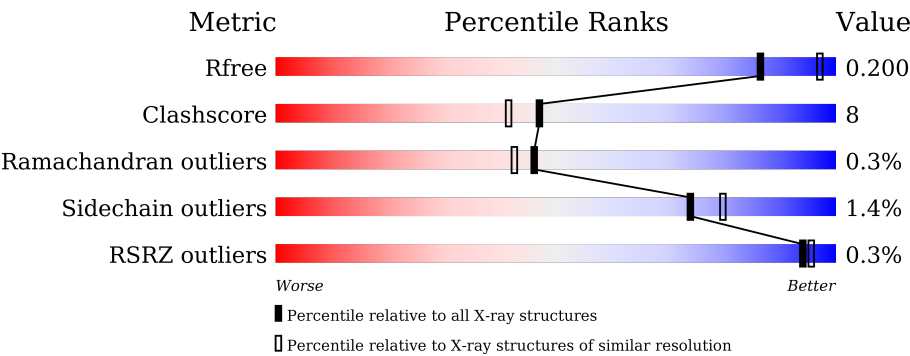
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.16 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))
$R_{free}$	91344	1045 (2.16-2.16)
Clashscore	102246	1152 (2.16-2.16)
Ramachandran outliers	100387	1131 (2.16-2.16)
Sidechain outliers	100360	1131 (2.16-2.16)
RSRZ outliers	91569	1050 (2.16-2.16)

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	407	<div><div></div><div>81%16%.</div></div>
1	B	407	<div><div></div><div>82%15%.</div></div>
1	C	407	<div><div>%</div><div>81%16%..</div></div>
1	D	407	<div><div></div><div>85%11%.</div></div>
1	E	407	<div><div></div><div>83%14%..</div></div>
1	F	407	<div><div></div><div>82%15%.</div></div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	G	407	 82% 15% ..
1	H	407	 81% 16% ..
1	J	407	 83% 14% .
1	K	407	 86% 12% .
1	L	407	 % 76% 21% ..
1	M	407	 80% 16% ..

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 36700 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 Beta-ketothiolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	396	Total	C	N	O	S	0	0	0
			2895	1795	529	555	16			
1	B	398	Total	C	N	O	S	0	0	0
			2938	1819	540	563	16			
1	C	398	Total	C	N	O	S	0	0	0
			2933	1814	539	564	16			
1	D	396	Total	C	N	O	S	0	0	0
			2924	1811	538	559	16			
1	E	396	Total	C	N	O	S	0	0	0
			2925	1810	538	561	16			
1	F	399	Total	C	N	O	S	0	0	0
			2947	1824	542	565	16			
1	G	398	Total	C	N	O	S	0	0	0
			2940	1820	540	564	16			
1	H	397	Total	C	N	O	S	0	0	0
			2933	1816	540	561	16			
1	J	396	Total	C	N	O	S	0	0	0
			2925	1810	538	561	16			
1	K	399	Total	C	N	O	S	0	0	0
			2949	1825	542	566	16			
1	L	398	Total	C	N	O	S	0	0	0
			2932	1816	540	560	16			
1	M	397	Total	C	N	O	S	0	0	0
			2933	1816	540	561	16			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	125	Total	O	0	0
			125	125		
2	B	145	Total	O	0	0
			145	145		

*Continued on next page...*

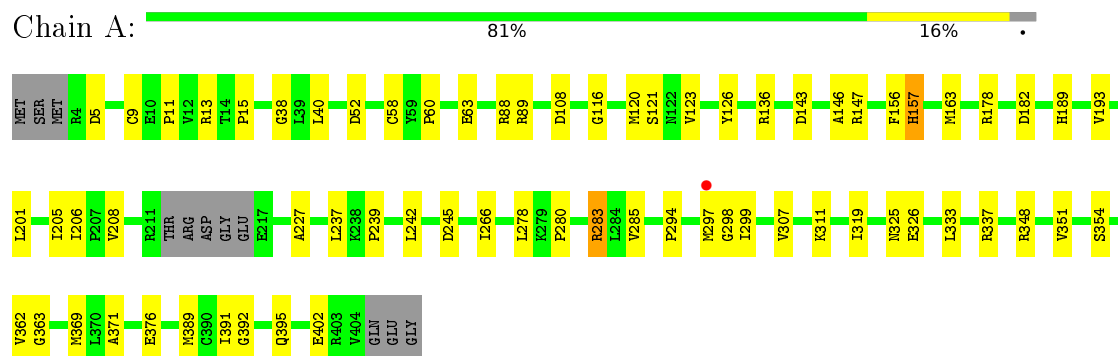
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	131	Total 131	O 131	0	0
2	D	156	Total 156	O 156	0	0
2	E	105	Total 105	O 105	0	0
2	F	129	Total 129	O 129	0	0
2	G	114	Total 114	O 114	0	0
2	H	154	Total 154	O 154	0	0
2	J	120	Total 120	O 120	0	0
2	K	142	Total 142	O 142	0	0
2	L	99	Total 99	O 99	0	0
2	M	106	Total 106	O 106	0	0

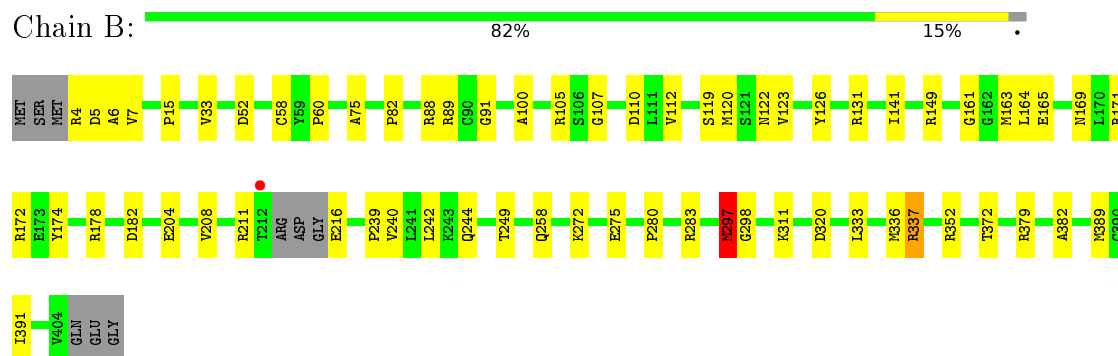
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

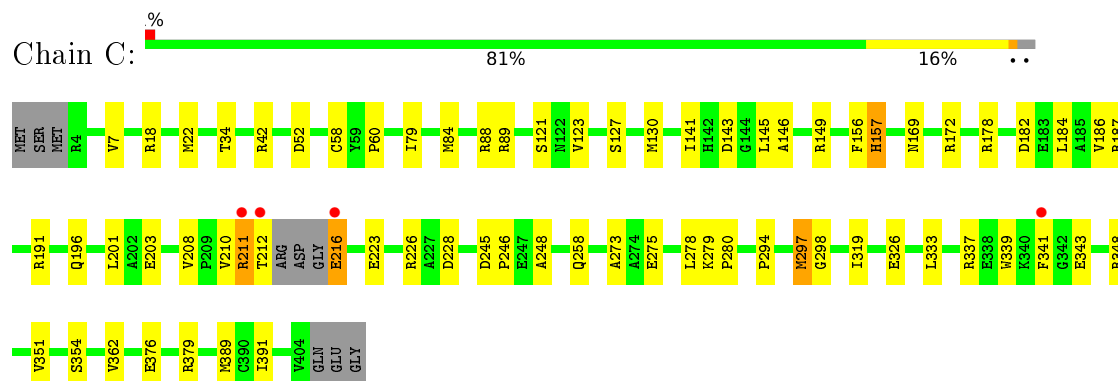
#### • Molecule 1: Beta-ketothiolase




#### • Molecule 1: Beta-ketothiolase

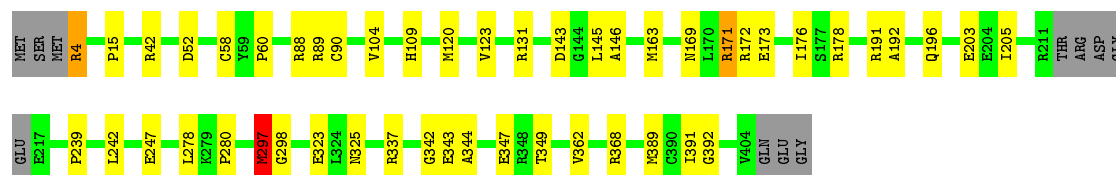


#### • Molecule 1: Beta-ketothiolase




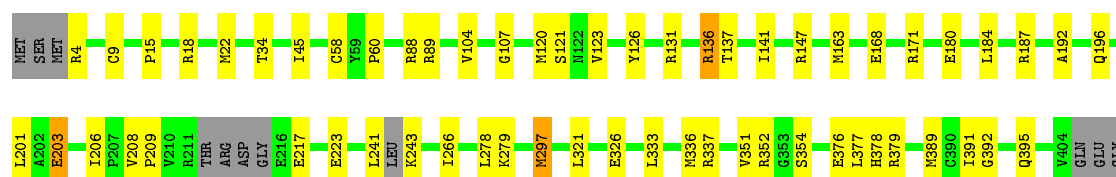
- Molecule 1: Beta-ketothiolase

Chain D:  85% 11%




- Molecule 1: Beta-ketothiolase

Chain E:  83% 14%




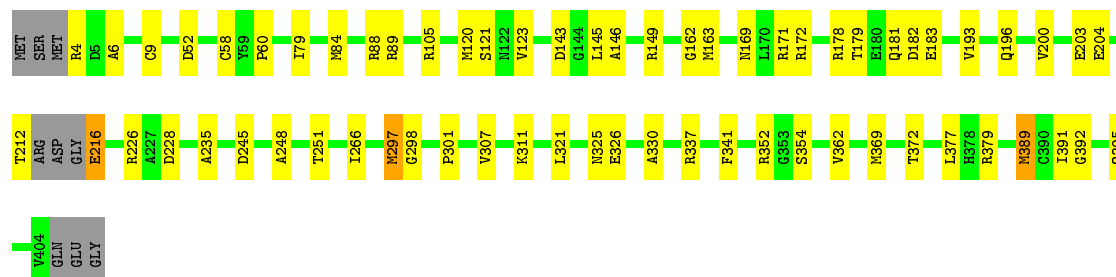
- Molecule 1: Beta-ketothiolase

Chain F:  82% 15%




- Molecule 1: Beta-ketothiolase

Chain G:  82% 15%



- Molecule 1: Beta-ketothiolase

Chain H:  81% 16%

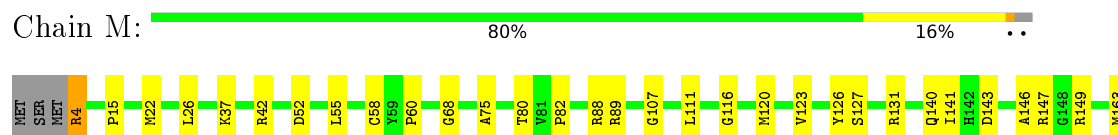


- Molecule 1: Beta-ketothiolase

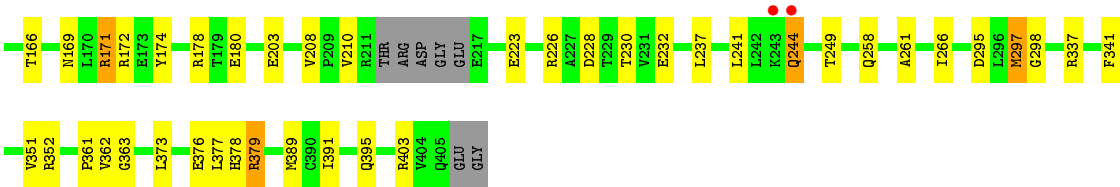
- Molecule 1: Beta-ketothiolase

- Molecule 1: Beta-ketothiolase

- Molecule 1: Beta-ketothiolase







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	190.96Å 190.96Å 264.68Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.19 – 2.16 45.19 – 2.16	Depositor EDS
% Data completeness (in resolution range)	99.9 (45.19-2.16) 99.9 (45.19-2.16)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.32 (at 2.16Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.169 , 0.200 0.169 , 0.200	Depositor DCC
$R_{free}$ test set	14686 reflections (5.06%)	DCC
Wilson B-factor (Å <sup>2</sup> )	28.6	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 36.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.226 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	36700	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 25.40 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.1919e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.41	0/2939	0.60	0/3990
1	B	0.42	0/2982	0.60	0/4044
1	C	0.63	4/2977 (0.1%)	0.75	7/4039 (0.2%)
1	D	0.46	0/2968	0.63	2/4025 (0.0%)
1	E	0.41	0/2968	0.59	0/4023
1	F	0.42	0/2991	0.61	1/4056 (0.0%)
1	G	0.42	0/2984	0.60	0/4047
1	H	0.42	1/2977 (0.0%)	0.61	2/4037 (0.0%)
1	J	0.40	0/2968	0.59	0/4023
1	K	0.42	0/2993	0.60	0/4059
1	L	0.39	0/2976	0.60	0/4037
1	M	0.40	0/2977	0.59	0/4037
All	All	0.44	5/35700 (0.0%)	0.61	12/48417 (0.0%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	216	GLU	CB-CG	17.57	1.85	1.52
1	C	216	GLU	CG-CD	13.99	1.73	1.51
1	C	216	GLU	CD-OE1	-6.48	1.18	1.25
1	C	216	GLU	CA-CB	6.09	1.67	1.53
1	H	389	MET	CG-SD	-5.19	1.67	1.81

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	216	GLU	CA-CB-CG	13.89	143.95	113.40
1	C	216	GLU	OE1-CD-OE2	-12.54	108.26	123.30
1	D	297	MET	CG-SD-CE	-10.06	84.11	100.20
1	C	216	GLU	CG-CD-OE1	7.88	134.07	118.30
1	F	389	MET	CA-CB-CG	-7.06	101.31	113.30

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	2895	0	2889	50	0
1	B	2938	0	2947	46	1
1	C	2933	0	2932	60	0
1	D	2924	0	2939	35	1
1	E	2925	0	2933	44	0
1	F	2947	0	2955	41	0
1	G	2940	0	2952	54	0
1	H	2933	0	2947	51	0
1	J	2925	0	2933	42	0
1	K	2949	0	2960	39	0
1	L	2932	0	2944	66	0
1	M	2933	0	2947	48	0
2	A	125	0	0	2	0
2	B	145	0	0	0	0
2	C	131	0	0	2	0
2	D	156	0	0	4	0
2	E	105	0	0	1	0
2	F	129	0	0	2	0
2	G	114	0	0	7	0
2	H	154	0	0	1	0
2	J	120	0	0	1	0
2	K	142	0	0	2	0
2	L	99	0	0	2	0
2	M	106	0	0	1	0
All	All	36700	0	35278	529	1

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 8.

The worst 5 of 529 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:C:216:GLU:CG	1:C:216:GLU:CB	1.85	1.49

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:297:MET:HG3	1:D:298:GLY:N	1.77	0.96
1:C:211:ARG:HD2	1:C:212:THR:H	1.33	0.93
1:F:163:MET:SD	1:F:297:MET:HE1	2.12	0.90
1:B:149:ARG:NH2	1:B:258:GLN:OE1	2.04	0.90

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:337:ARG:NH2	1:D:342:GLY:CA[5_554]	2.08	0.12

## 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	392/407 (96%)	381 (97%)	9 (2%)	2 (0%)	34	26
1	B	394/407 (97%)	382 (97%)	11 (3%)	1 (0%)	46	42
1	C	394/407 (97%)	384 (98%)	9 (2%)	1 (0%)	46	42
1	D	392/407 (96%)	380 (97%)	9 (2%)	3 (1%)	24	15
1	E	390/407 (96%)	382 (98%)	8 (2%)	0	100	100
1	F	395/407 (97%)	386 (98%)	8 (2%)	1 (0%)	46	42
1	G	394/407 (97%)	382 (97%)	10 (2%)	2 (0%)	34	26
1	H	393/407 (97%)	385 (98%)	8 (2%)	0	100	100
1	J	390/407 (96%)	378 (97%)	11 (3%)	1 (0%)	46	42
1	K	395/407 (97%)	384 (97%)	10 (2%)	1 (0%)	46	42
1	L	394/407 (97%)	381 (97%)	13 (3%)	0	100	100
1	M	393/407 (97%)	383 (98%)	8 (2%)	2 (0%)	34	26
All	All	4716/4884 (97%)	4588 (97%)	114 (2%)	14 (0%)	46	42

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	343	GLU
1	A	136	ARG
1	D	297	MET
1	G	297	MET
1	K	297	MET

### 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	294/309 (95%)	290 (99%)	4 (1%)	74	80
1	B	301/309 (97%)	299 (99%)	2 (1%)	88	93
1	C	300/309 (97%)	296 (99%)	4 (1%)	76	82
1	D	300/309 (97%)	294 (98%)	6 (2%)	63	67
1	E	300/309 (97%)	296 (99%)	4 (1%)	76	82
1	F	302/309 (98%)	298 (99%)	4 (1%)	76	82
1	G	302/309 (98%)	299 (99%)	3 (1%)	82	87
1	H	301/309 (97%)	299 (99%)	2 (1%)	88	93
1	J	300/309 (97%)	299 (100%)	1 (0%)	94	97
1	K	303/309 (98%)	299 (99%)	4 (1%)	76	82
1	L	300/309 (97%)	292 (97%)	8 (3%)	52	53
1	M	301/309 (97%)	294 (98%)	7 (2%)	58	62
All	All	3604/3708 (97%)	3555 (99%)	49 (1%)	74	80

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	243	LYS
1	H	337	ARG
1	M	244	GLN
1	G	216	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	H	347	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	157	HIS
1	M	181	GLN
1	L	157	HIS
1	D	181	GLN
1	L	181	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 ⓘ

### 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	396/407 (97%)	-0.26	1 (0%) 94 95	16, 31, 50, 66	0
1	B	398/407 (97%)	-0.29	1 (0%) 94 95	14, 27, 49, 73	0
1	C	398/407 (97%)	-0.22	4 (1%) 84 88	15, 30, 53, 73	0
1	D	396/407 (97%)	-0.34	0 100 100	15, 24, 45, 62	0
1	E	396/407 (97%)	-0.22	0 100 100	17, 32, 53, 70	0
1	F	399/407 (98%)	-0.28	0 100 100	16, 28, 50, 75	0
1	G	398/407 (97%)	-0.25	0 100 100	16, 30, 54, 72	0
1	H	397/407 (97%)	-0.33	0 100 100	16, 25, 45, 67	0
1	J	396/407 (97%)	-0.22	0 100 100	15, 31, 54, 68	0
1	K	399/407 (98%)	-0.30	1 (0%) 94 95	15, 26, 50, 76	0
1	L	398/407 (97%)	-0.12	3 (0%) 87 90	18, 36, 56, 76	0
1	M	397/407 (97%)	-0.25	2 (0%) 91 93	20, 31, 53, 73	0
All	All	4768/4884 (97%)	-0.26	12 (0%) 94 95	14, 29, 52, 76	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	216	GLU	4.8
1	C	211	ARG	4.1
1	L	174	TYR	3.0
1	B	212	THR	2.9
1	C	212	THR	2.5

### 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.