



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 12:46 am BST

PDB ID : 5KZF  
Title : Crystal structure of near full-length hexameric Mycobacterium tuberculosis proteasomal ATPase Mpa in apo form  
Authors : Li, H.; Hu, K.; Yang, S.; Bai, L.  
Deposited on : 2016-07-25  
Resolution : 3.49 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

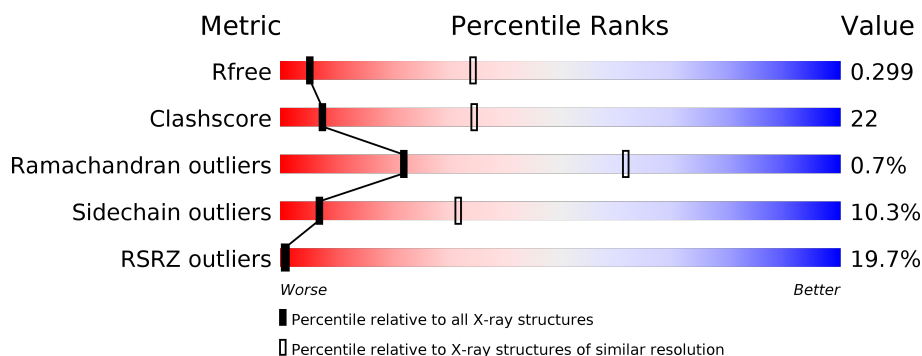
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.49 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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	513	<div> <div>21%</div> <div>57%</div> <div>31%</div> <div>6%</div> </div>
1	B	513	<div> <div>19%</div> <div>60%</div> <div>29%</div> <div>7%</div> </div>
1	C	513	<div> <div>19%</div> <div>60%</div> <div>27%</div> <div>8%</div> </div>
1	D	513	<div> <div>22%</div> <div>56%</div> <div>29%</div> <div>5%</div> <div>10%</div> </div>
1	E	513	<div> <div>27%</div> <div>59%</div> <div>26%</div> <div>5%</div> <div>10%</div> </div>
1	F	513	<div> <div>18%</div> <div>55%</div> <div>29%</div> <div>5%</div> <div>10%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	513	<div><div></div><div>17%</div><div>56%</div><div>28%</div><div>5%</div><div>10%</div></div>
1	H	513	<div><div></div><div>16%</div><div>61%</div><div>28%</div><div></div><div>7%</div></div>
1	I	513	<div><div></div><div>13%</div><div>56%</div><div>31%</div><div>6%</div><div>7%</div></div>
1	J	513	<div><div></div><div>12%</div><div>59%</div><div>30%</div><div></div><div>6%</div></div>
1	K	513	<div><div></div><div>10%</div><div>58%</div><div>32%</div><div>6%</div><div></div></div>
1	L	513	<div><div></div><div>23%</div><div>56%</div><div>31%</div><div>5%</div><div>8%</div></div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 44298 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 Proteasome-associated ATPase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	480	Total	C	N	O	S	0	0	0
			3747	2360	647	729	11			
1	B	475	Total	C	N	O	S	0	0	0
			3707	2337	637	722	11			
1	C	472	Total	C	N	O	S	0	0	0
			3649	2300	632	706	11			
1	D	460	Total	C	N	O	S	0	0	0
			3593	2266	621	695	11			
1	E	461	Total	C	N	O	S	0	0	0
			3597	2268	622	696	11			
1	F	460	Total	C	N	O	S	0	0	0
			3589	2267	615	696	11			
1	G	461	Total	C	N	O	S	0	0	0
			3595	2270	615	698	12			
1	H	478	Total	C	N	O	S	0	0	0
			3729	2351	639	727	12			
1	I	478	Total	C	N	O	S	0	0	0
			3735	2352	645	726	12			
1	J	484	Total	C	N	O	S	0	0	0
			3775	2377	646	740	12			
1	K	498	Total	C	N	O	S	0	0	0
			3887	2442	672	761	12			
1	L	473	Total	C	N	O	S	0	0	0
			3695	2329	635	720	11			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP A5U4E1
B	1	MET	-	initiating methionine	UNP A5U4E1
C	1	MET	-	initiating methionine	UNP A5U4E1
D	1	MET	-	initiating methionine	UNP A5U4E1
E	1	MET	-	initiating methionine	UNP A5U4E1

*Continued on next page...*

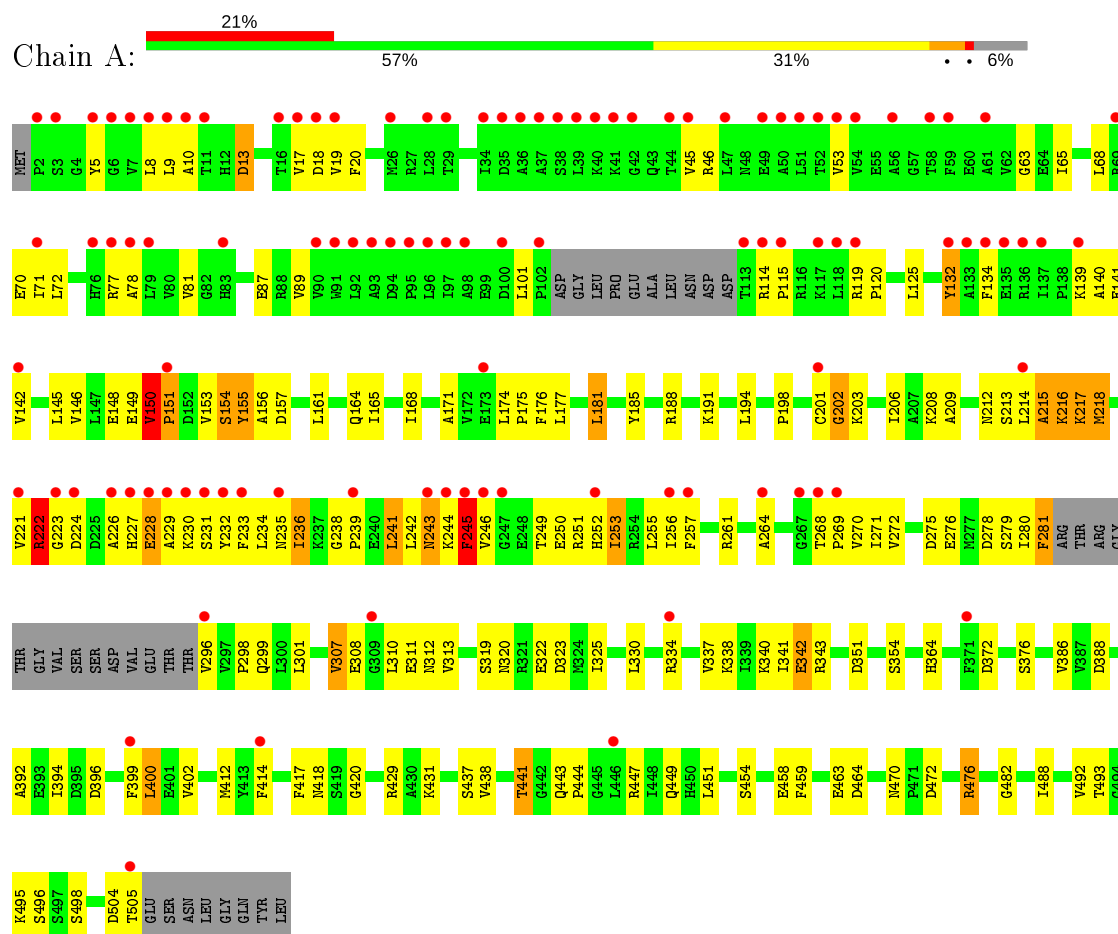
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	1	MET	-	initiating methionine	UNP A5U4E1
G	1	MET	-	initiating methionine	UNP A5U4E1
H	1	MET	-	initiating methionine	UNP A5U4E1
I	1	MET	-	initiating methionine	UNP A5U4E1
J	1	MET	-	initiating methionine	UNP A5U4E1
K	1	MET	-	initiating methionine	UNP A5U4E1
L	1	MET	-	initiating methionine	UNP A5U4E1

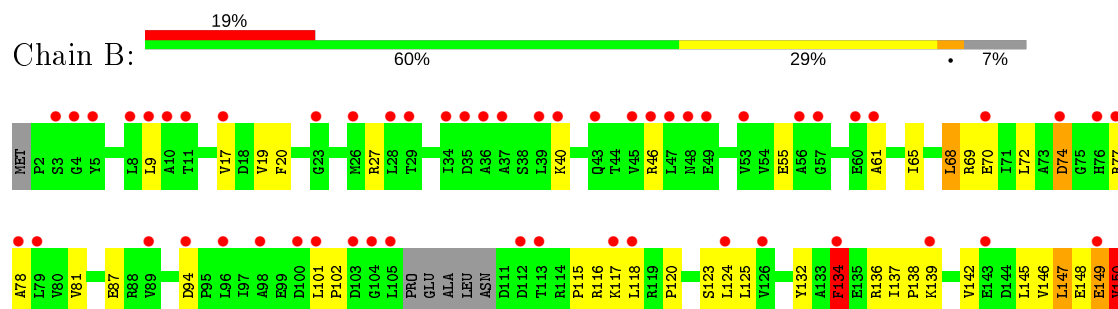
### 3 Residue-property plots [i](#)

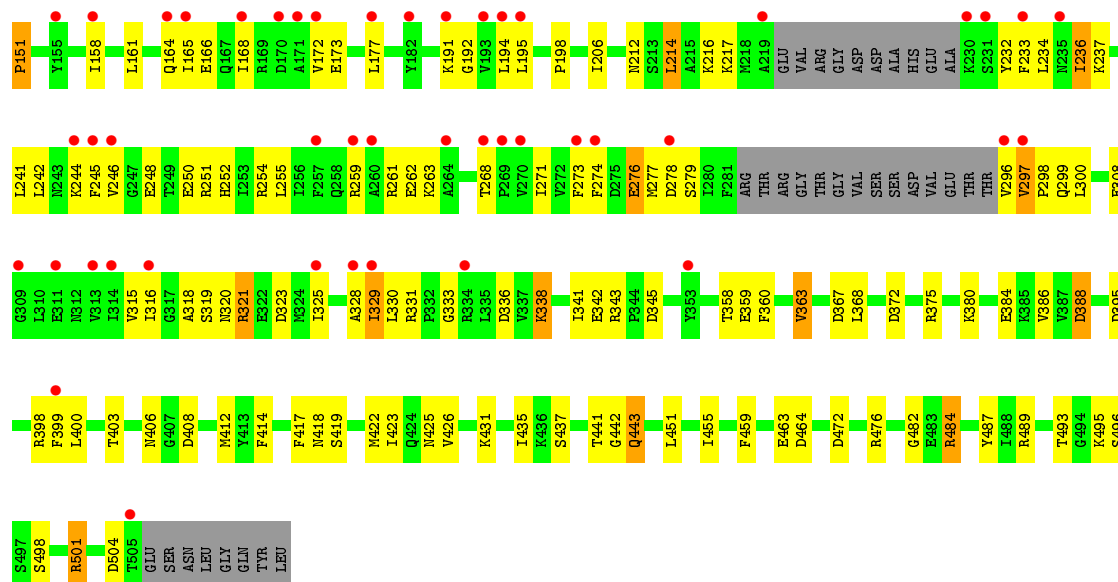
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Proteasome-associated ATPase

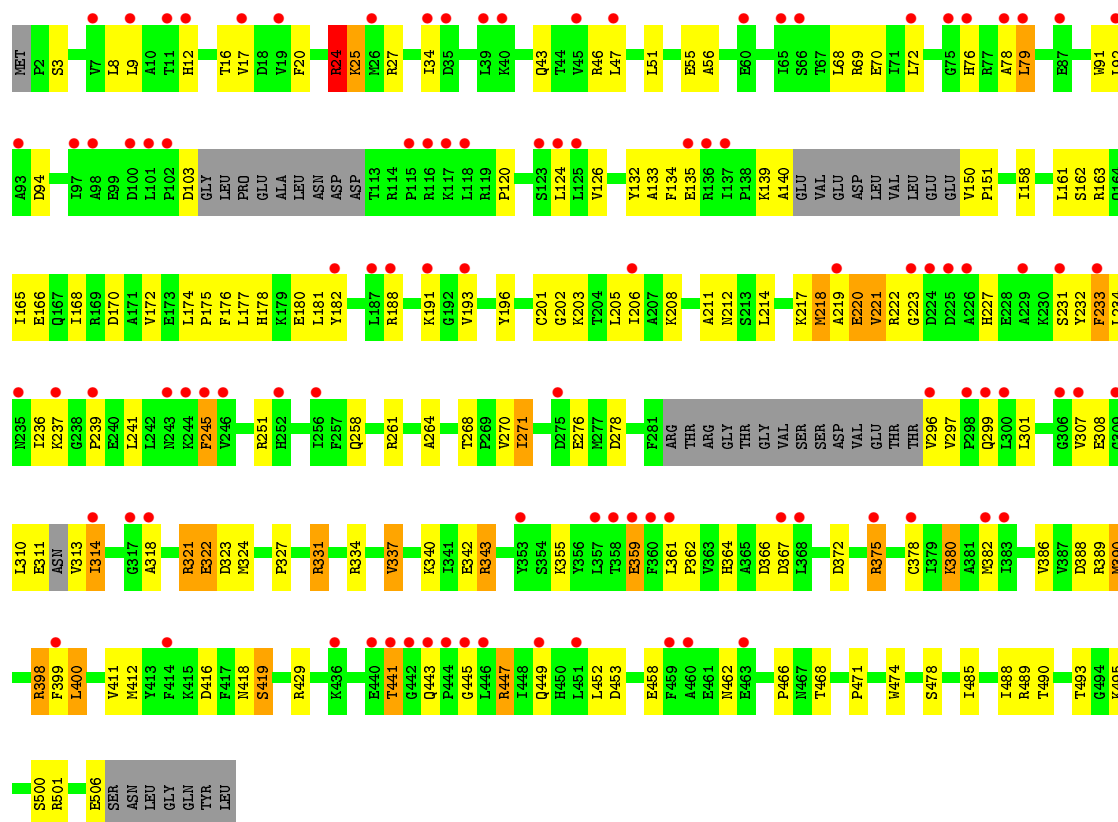


#### • Molecule 1: Proteasome-associated ATPase



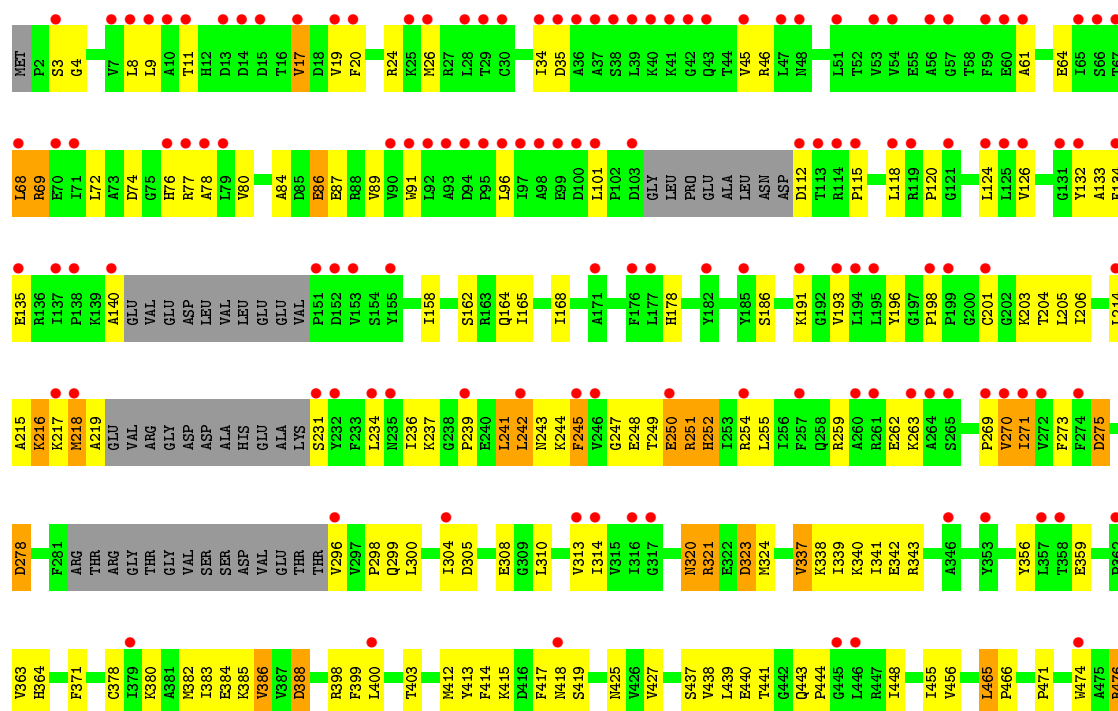


• Molecule 1: Proteasome-associated ATPase

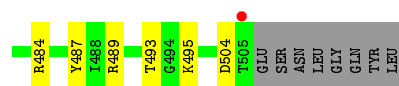


• Molecule 1: Proteasome-associated ATPase

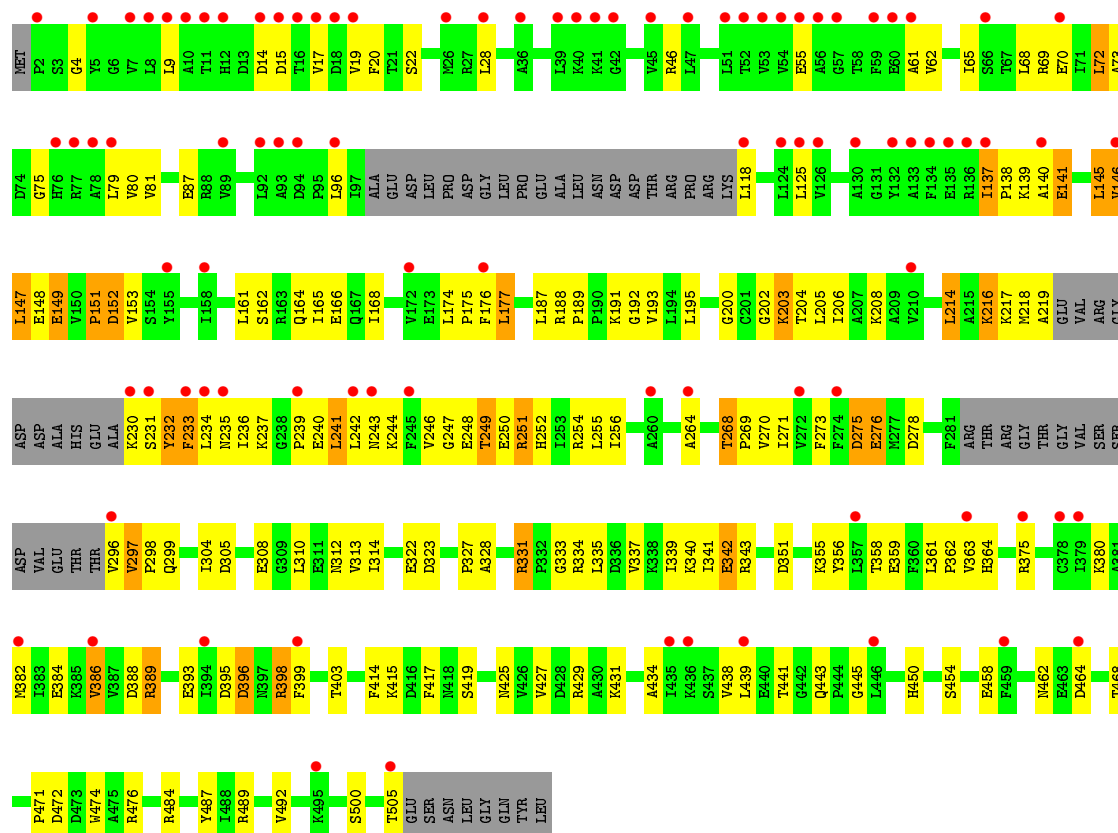




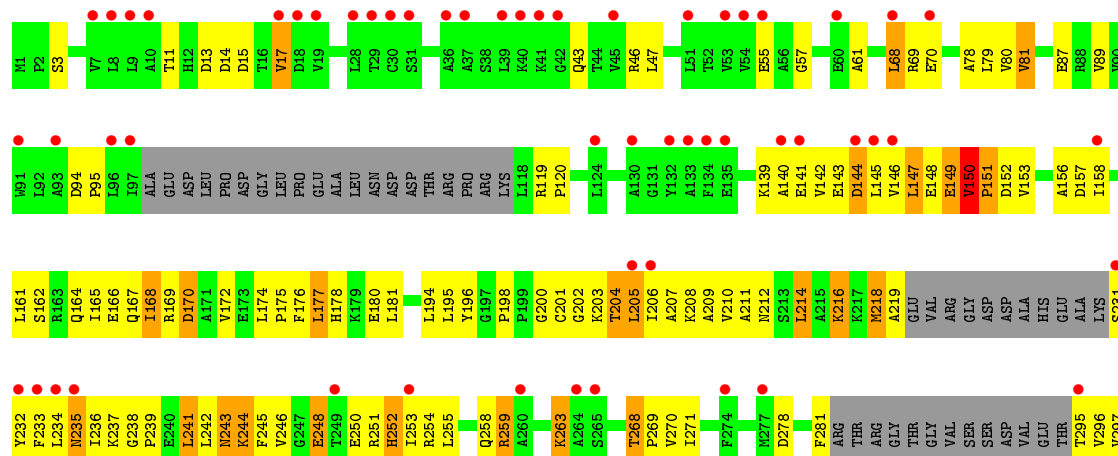


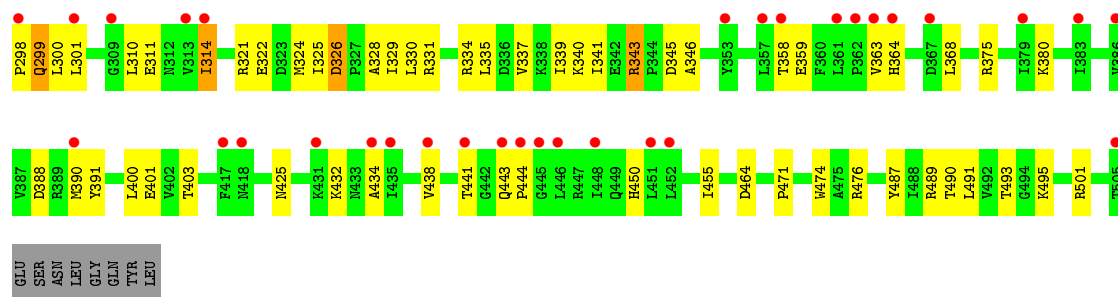


● Molecule 1: Proteasome-associated ATPase

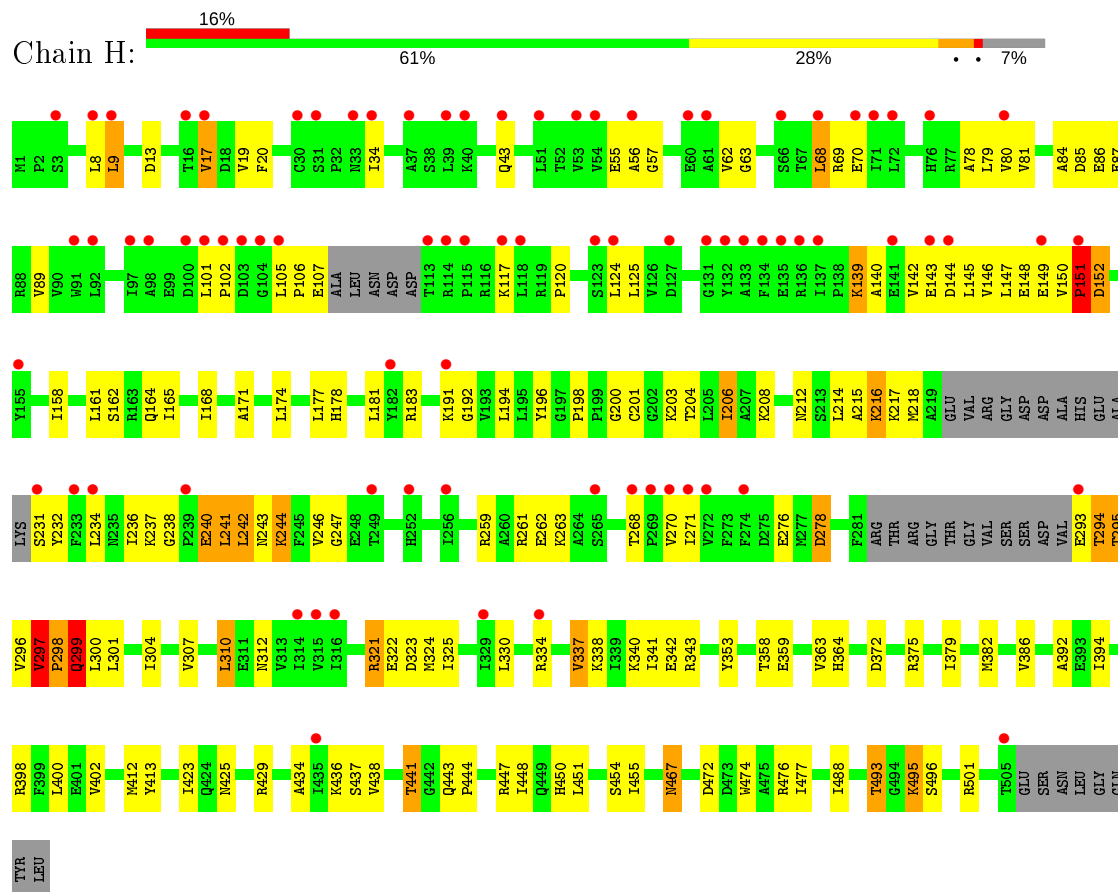


● Molecule 1: Proteasome-associated ATPase

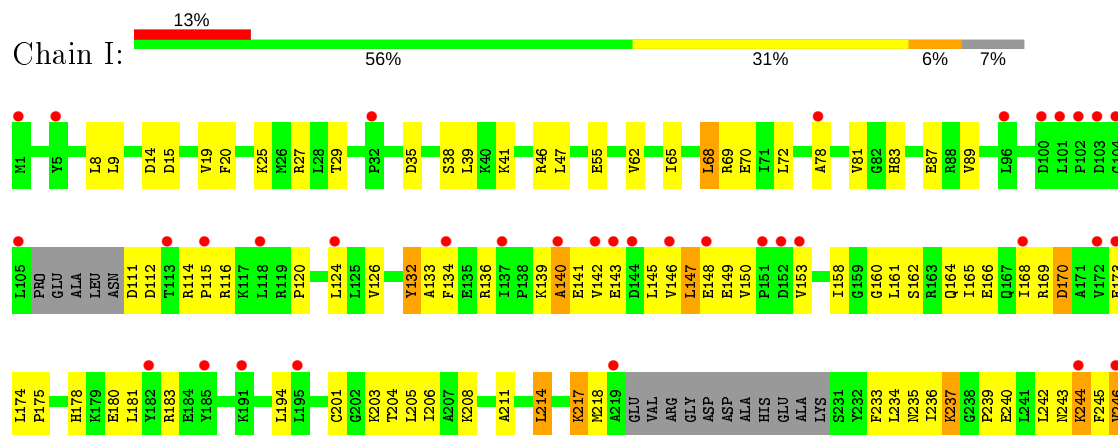


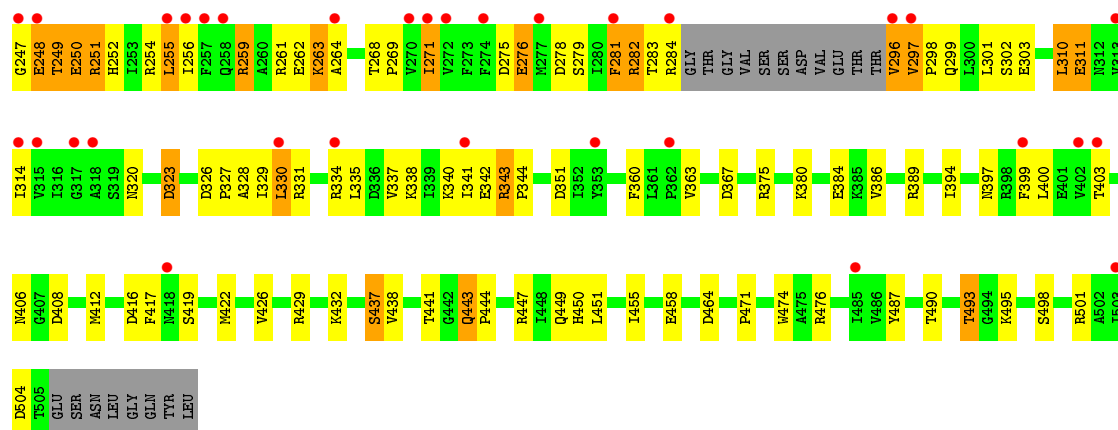


• Molecule 1: Proteasome-associated ATPase

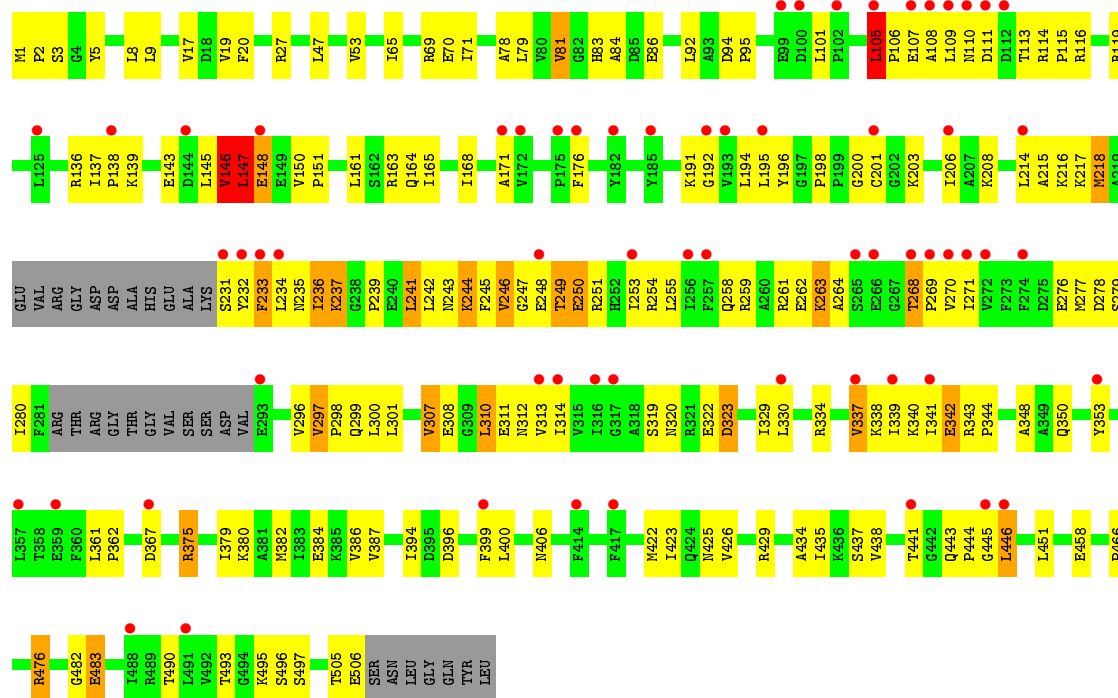


• Molecule 1: Proteasome-associated ATPase

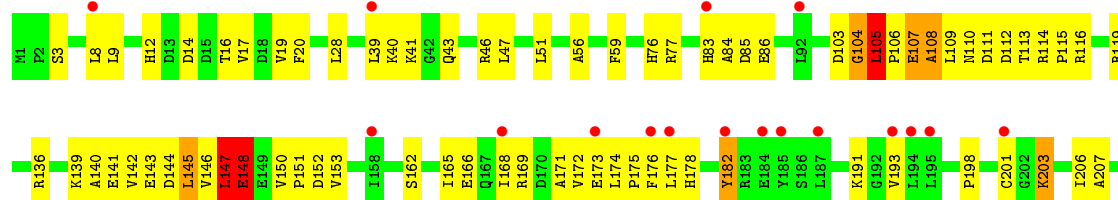


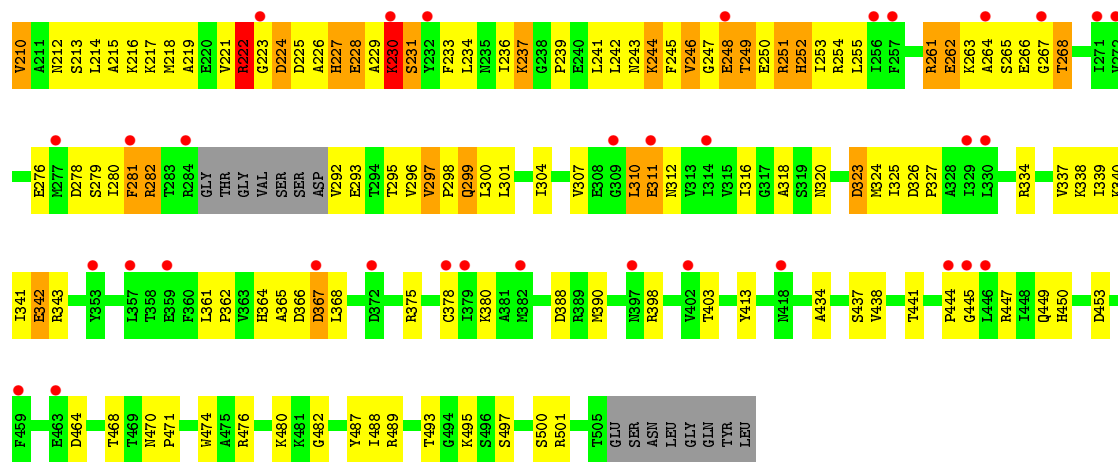


• Molecule 1: Proteasome-associated ATPase

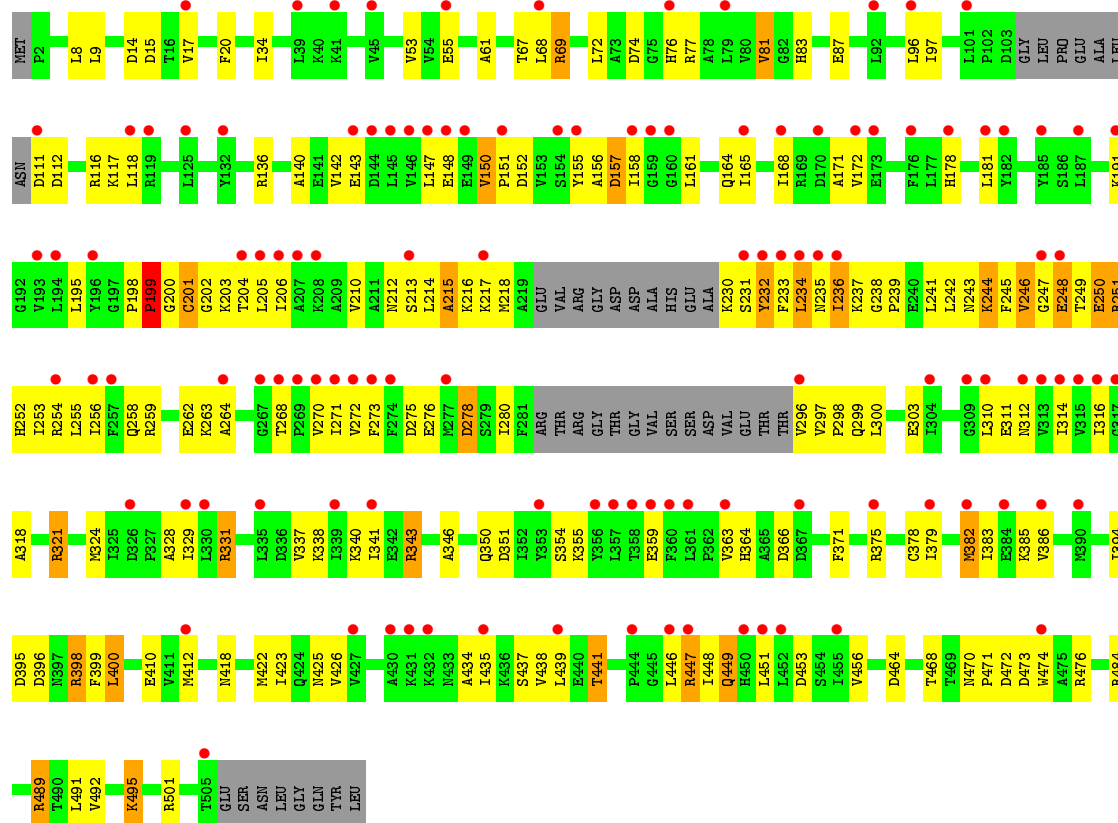


• Molecule 1: Proteasome-associated ATPase





• Molecule 1: Proteasome-associated ATPase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	115.27Å 202.59Å 303.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	71.13 – 3.49 71.13 – 3.49	Depositor EDS
% Data completeness (in resolution range)	99.4 (71.13-3.49) 99.4 (71.13-3.49)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.61 (at 3.49Å)	Xtriage
Refinement program	REFMAC 5.8.0123	Depositor
R, $R_{free}$	0.268 , 0.306 0.262 , 0.299	Depositor DCC
$R_{free}$ test set	1996 reflections (2.21%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	119.4	Xtriage
Anisotropy	0.038	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 92.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	44298	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	130.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.94% 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	0.35	1/3808 (0.0%)	0.60	1/5150 (0.0%)
1	B	0.32	0/3766	0.56	3/5092 (0.1%)
1	C	0.31	0/3706	0.56	1/5011 (0.0%)
1	D	0.30	0/3651	0.53	1/4933 (0.0%)
1	E	0.31	1/3655 (0.0%)	0.55	3/4939 (0.1%)
1	F	0.32	0/3646	0.57	2/4929 (0.0%)
1	G	0.35	1/3652 (0.0%)	0.56	1/4939 (0.0%)
1	H	0.35	1/3789 (0.0%)	0.58	2/5126 (0.0%)
1	I	0.34	0/3794	0.60	1/5130 (0.0%)
1	J	0.32	0/3836	0.58	2/5192 (0.0%)
1	K	0.37	0/3950	0.62	3/5346 (0.1%)
1	L	0.31	0/3754	0.56	0/5076
All	All	0.33	4/45007 (0.0%)	0.57	20/60863 (0.0%)

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	3
1	B	0	1
1	C	0	1
1	D	0	1
1	F	0	1
1	G	0	2
1	H	0	3
1	I	0	2
1	J	0	1
1	K	0	4
1	L	0	2
All	All	0	21

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	466	PRO	N-CD	5.37	1.55	1.47
1	A	151	PRO	N-CD	5.20	1.55	1.47
1	H	298	PRO	N-CD	5.18	1.55	1.47
1	G	151	PRO	N-CD	5.08	1.54	1.47

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	K	148	GLU	N-CA-C	7.45	131.12	111.00
1	E	251	ARG	N-CA-C	7.28	130.66	111.00
1	H	150	VAL	C-N-CD	-6.30	106.75	120.60
1	K	251	ARG	N-CA-C	-5.73	95.53	111.00
1	G	150	VAL	C-N-CD	5.69	140.36	128.40

There are no chirality outliers.

5 of 21 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	202	GLY	Peptide
1	A	215	ALA	Peptide
1	A	222	ARG	Peptide
1	B	297	VAL	Peptide
1	C	24	ARG	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	3747	0	3755	190	0
1	B	3707	0	3718	145	0
1	C	3649	0	3644	136	3
1	D	3593	0	3607	120	0
1	E	3597	0	3610	131	0
1	F	3589	0	3606	137	0
1	G	3595	0	3611	219	0
1	H	3729	0	3741	180	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	I	3735	0	3749	205	0
1	J	3775	0	3778	206	0
1	K	3887	0	3888	270	3
1	L	3695	0	3704	166	0
All	All	44298	0	44411	1957	3

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 22.

The worst 5 of 1957 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:K:229:ALA:CA	1:K:230:LYS:HB2	1.46	1.45
1:A:150:VAL:CG2	1:A:231:SER:HA	1.52	1.38
1:H:296:VAL:CG2	1:I:242:LEU:HD21	1.51	1.38
1:K:153:VAL:O	1:K:212:ASN:ND2	1.57	1.38
1:G:148:GLU:CG	1:G:150:VAL:HG13	1.54	1.34

All (3) 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:C:180:GLU:CG	1:K:230:LYS:CD[3_444]	1.97	0.23
1:C:180:GLU:CG	1:K:230:LYS:CE[3_444]	2.05	0.15
1:C:180:GLU:CG	1:K:230:LYS:NZ[3_444]	2.11	0.09

## 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	474/513 (92%)	440 (93%)	29 (6%)	5 (1%)	14 52

Continued on next page...



Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	467/513 (91%)	441 (94%)	23 (5%)	3 (1%)	25	64
1	C	462/513 (90%)	429 (93%)	30 (6%)	3 (1%)	25	64
1	D	450/513 (88%)	433 (96%)	17 (4%)	0	100	100
1	E	451/513 (88%)	431 (96%)	19 (4%)	1 (0%)	47	81
1	F	452/513 (88%)	424 (94%)	24 (5%)	4 (1%)	17	56
1	G	453/513 (88%)	428 (94%)	21 (5%)	4 (1%)	17	56
1	H	470/513 (92%)	442 (94%)	24 (5%)	4 (1%)	17	56
1	I	470/513 (92%)	446 (95%)	23 (5%)	1 (0%)	47	81
1	J	478/513 (93%)	442 (92%)	33 (7%)	3 (1%)	25	64
1	K	494/513 (96%)	457 (92%)	29 (6%)	8 (2%)	9	43
1	L	465/513 (91%)	434 (93%)	27 (6%)	4 (1%)	17	56
All	All	5586/6156 (91%)	5247 (94%)	299 (5%)	40 (1%)	22	61

5 of 40 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	151	PRO
1	E	250	GLU
1	H	151	PRO
1	H	152	ASP
1	K	222	ARG

### 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	405/433 (94%)	371 (92%)	34 (8%)	11	40
1	B	402/433 (93%)	362 (90%)	40 (10%)	7	32
1	C	389/433 (90%)	351 (90%)	38 (10%)	8	33
1	D	389/433 (90%)	346 (89%)	43 (11%)	6	28
1	E	389/433 (90%)	351 (90%)	38 (10%)	8	33

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	F	389/433 (90%)	345 (89%)	44 (11%)	6	27
1	G	390/433 (90%)	347 (89%)	43 (11%)	6	29
1	H	405/433 (94%)	362 (89%)	43 (11%)	6	30
1	I	405/433 (94%)	362 (89%)	43 (11%)	6	30
1	J	410/433 (95%)	376 (92%)	34 (8%)	11	40
1	K	421/433 (97%)	371 (88%)	50 (12%)	5	25
1	L	401/433 (93%)	356 (89%)	45 (11%)	6	27
All	All	4795/5196 (92%)	4300 (90%)	495 (10%)	7	32

5 of 495 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	363	VAL
1	G	441	THR
1	L	230	LYS
1	F	425	ASN
1	G	218	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	178	HIS
1	I	320	ASN
1	L	252	HIS
1	H	467	ASN
1	J	110	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 [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	480/513 (93%)	1.17	107 (22%) 0 0	56, 117, 190, 215	0
1	B	475/513 (92%)	1.19	100 (21%) 1 1	57, 133, 189, 213	0
1	C	472/513 (92%)	1.13	100 (21%) 0 1	55, 134, 193, 231	0
1	D	460/513 (89%)	1.33	111 (24%) 0 0	66, 127, 215, 230	0
1	E	461/513 (89%)	1.61	136 (29%) 0 0	71, 149, 212, 231	0
1	F	460/513 (89%)	1.14	93 (20%) 1 1	66, 136, 191, 226	0
1	G	461/513 (89%)	1.11	88 (19%) 1 1	69, 128, 195, 223	0
1	H	478/513 (93%)	1.10	81 (16%) 1 2	61, 107, 191, 209	0
1	I	478/513 (93%)	0.93	68 (14%) 2 3	62, 107, 176, 217	0
1	J	484/513 (94%)	0.88	63 (13%) 3 4	70, 120, 191, 215	0
1	K	498/513 (97%)	0.78	51 (10%) 6 7	67, 114, 164, 191	0
1	L	473/513 (92%)	1.36	119 (25%) 0 0	64, 146, 205, 219	0
All	All	5680/6156 (92%)	1.14	1117 (19%) 1 1	55, 125, 200, 231	0

The worst 5 of 1117 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	231	SER	11.2
1	D	78	ALA	11.2
1	C	231	SER	10.9
1	L	146	VAL	10.4
1	L	232	TYR	10.1

### 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.