



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 12, 2017 – 05:34 PM EST

PDB ID : 5THO  
Title : Crystal Structure of Mycobacterium Tuberculosis Proteasome in complex with N,C-capped Dipeptide Inhibitor PKS2205  
Authors : Hsu, H.C.; Li, H.  
Deposited on : 2016-09-30  
Resolution : 3.00 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	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
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-20028442

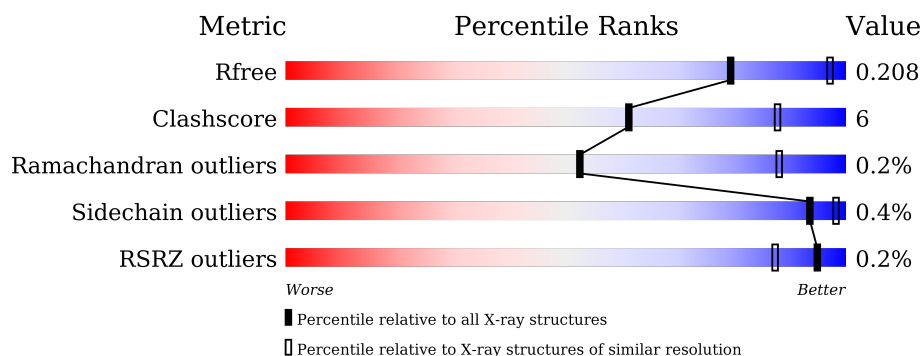
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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.








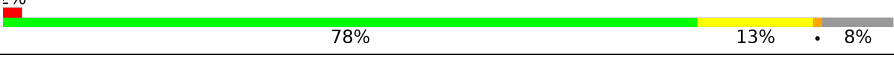


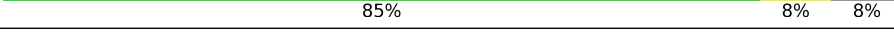


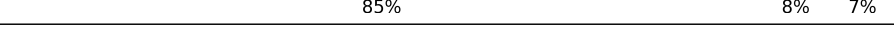

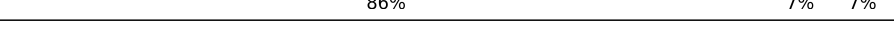


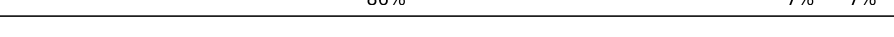

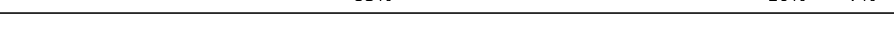



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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	240	<div> <div style="width: 64%; background-color: green;"></div> <div style="width: 26%; background-color: yellow;"></div> <div style="width: 9%; background-color: orange;"></div> <div style="width: 1%; background-color: red;"></div> <div style="width: 0%; background-color: grey;"></div> </div> <div>64% 26% • 9%</div>
1	B	240	<div> <div style="width: 70%; background-color: green;"></div> <div style="width: 20%; background-color: yellow;"></div> <div style="width: 10%; background-color: orange;"></div> <div style="width: 0%; background-color: red;"></div> <div style="width: 0%; background-color: grey;"></div> </div> <div>70% 20% 10%</div>
1	C	240	<div> <div style="width: 71%; background-color: green;"></div> <div style="width: 19%; background-color: yellow;"></div> <div style="width: 9%; background-color: orange;"></div> <div style="width: 1%; background-color: red;"></div> <div style="width: 0%; background-color: grey;"></div> </div> <div>71% 19% • 10%</div>
1	D	240	<div> <div style="width: 73%; background-color: green;"></div> <div style="width: 17%; background-color: yellow;"></div> <div style="width: 10%; background-color: orange;"></div> <div style="width: 0%; background-color: red;"></div> <div style="width: 0%; background-color: grey;"></div> </div> <div>73% 17% 10%</div>
1	E	240	<div> <div style="width: 74%; background-color: green;"></div> <div style="width: 16%; background-color: yellow;"></div> <div style="width: 9%; background-color: orange;"></div> <div style="width: 1%; background-color: red;"></div> <div style="width: 0%; background-color: grey;"></div> </div> <div>74% 16% • 9%</div>
1	F	240	<div> <div style="width: 75%; background-color: green;"></div> <div style="width: 15%; background-color: yellow;"></div> <div style="width: 10%; background-color: orange;"></div> <div style="width: 0%; background-color: red;"></div> <div style="width: 0%; background-color: grey;"></div> </div> <div>% 75% 15% 10%</div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	240	
1	O	240	
1	P	240	
1	Q	240	
1	R	240	
1	S	240	
1	T	240	
1	U	240	
2	H	240	
2	I	240	
2	J	240	
2	K	240	
2	L	240	
2	M	240	
2	N	240	
2	V	240	
2	W	240	
2	X	240	
2	Y	240	
2	Z	240	
2	a	240	
2	b	240	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	7C7	N	301	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 47201 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 subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	218	Total	C	N	O	S	0	0	0
			1677	1050	306	317	4			
1	B	215	Total	C	N	O	S	0	0	0
			1660	1041	303	312	4			
1	C	217	Total	C	N	O	S	0	0	0
			1672	1047	305	316	4			
1	D	216	Total	C	N	O	S	0	0	0
			1663	1041	304	314	4			
1	E	218	Total	C	N	O	S	0	0	0
			1677	1050	306	317	4			
1	F	217	Total	C	N	O	S	0	0	0
			1670	1046	305	315	4			
1	G	218	Total	C	N	O	S	0	0	0
			1677	1050	306	317	4			
1	O	218	Total	C	N	O	S	0	0	0
			1677	1050	306	317	4			
1	P	217	Total	C	N	O	S	0	0	0
			1671	1047	305	315	4			
1	Q	216	Total	C	N	O	S	0	0	0
			1668	1045	304	315	4			
1	R	217	Total	C	N	O	S	0	0	0
			1670	1046	305	315	4			
1	S	221	Total	C	N	O	S	0	0	0
			1701	1064	310	323	4			
1	T	217	Total	C	N	O	S	0	0	0
			1671	1047	305	315	4			
1	U	216	Total	C	N	O	S	0	0	0
			1664	1043	304	313	4			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	MET	-	initiating methionine	UNP A5U4D5

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	9	MET	-	initiating methionine	UNP A5U4D5
C	9	MET	-	initiating methionine	UNP A5U4D5
D	9	MET	-	initiating methionine	UNP A5U4D5
E	9	MET	-	initiating methionine	UNP A5U4D5
F	9	MET	-	initiating methionine	UNP A5U4D5
G	9	MET	-	initiating methionine	UNP A5U4D5
O	9	MET	-	initiating methionine	UNP A5U4D5
P	9	MET	-	initiating methionine	UNP A5U4D5
Q	9	MET	-	initiating methionine	UNP A5U4D5
R	9	MET	-	initiating methionine	UNP A5U4D5
S	9	MET	-	initiating methionine	UNP A5U4D5
T	9	MET	-	initiating methionine	UNP A5U4D5
U	9	MET	-	initiating methionine	UNP A5U4D5

- Molecule 2 is a protein called Proteasome subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	222	Total	C	N	O	S	0	0	0
			1638	1027	282	324	5			
2	I	222	Total	C	N	O	S	0	0	0
			1638	1027	282	324	5			
2	J	222	Total	C	N	O	S	0	0	0
			1638	1027	282	324	5			
2	K	223	Total	C	N	O	S	0	0	0
			1642	1029	283	325	5			
2	L	223	Total	C	N	O	S	0	0	0
			1642	1029	283	325	5			
2	M	223	Total	C	N	O	S	0	0	0
			1642	1029	283	325	5			
2	N	222	Total	C	N	O	S	0	0	0
			1638	1027	282	324	5			
2	V	223	Total	C	N	O	S	0	0	0
			1642	1029	283	325	5			
2	W	223	Total	C	N	O	S	0	0	0
			1642	1029	283	325	5			
2	X	222	Total	C	N	O	S	0	0	0
			1638	1027	282	324	5			
2	Y	223	Total	C	N	O	S	0	0	0
			1642	1029	283	325	5			
2	Z	222	Total	C	N	O	S	0	0	0
			1638	1027	282	324	5			
2	a	223	Total	C	N	O	S	0	0	0
			1642	1029	283	325	5			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	b	223	Total	C	N	O	S	0	0	0
			1642	1029	283	325	5			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	235	HIS	-	expression tag	UNP A5U4D6
H	236	HIS	-	expression tag	UNP A5U4D6
H	237	HIS	-	expression tag	UNP A5U4D6
H	238	HIS	-	expression tag	UNP A5U4D6
H	239	HIS	-	expression tag	UNP A5U4D6
H	240	HIS	-	expression tag	UNP A5U4D6
I	235	HIS	-	expression tag	UNP A5U4D6
I	236	HIS	-	expression tag	UNP A5U4D6
I	237	HIS	-	expression tag	UNP A5U4D6
I	238	HIS	-	expression tag	UNP A5U4D6
I	239	HIS	-	expression tag	UNP A5U4D6
I	240	HIS	-	expression tag	UNP A5U4D6
J	235	HIS	-	expression tag	UNP A5U4D6
J	236	HIS	-	expression tag	UNP A5U4D6
J	237	HIS	-	expression tag	UNP A5U4D6
J	238	HIS	-	expression tag	UNP A5U4D6
J	239	HIS	-	expression tag	UNP A5U4D6
J	240	HIS	-	expression tag	UNP A5U4D6
K	235	HIS	-	expression tag	UNP A5U4D6
K	236	HIS	-	expression tag	UNP A5U4D6
K	237	HIS	-	expression tag	UNP A5U4D6
K	238	HIS	-	expression tag	UNP A5U4D6
K	239	HIS	-	expression tag	UNP A5U4D6
K	240	HIS	-	expression tag	UNP A5U4D6
L	235	HIS	-	expression tag	UNP A5U4D6
L	236	HIS	-	expression tag	UNP A5U4D6
L	237	HIS	-	expression tag	UNP A5U4D6
L	238	HIS	-	expression tag	UNP A5U4D6
L	239	HIS	-	expression tag	UNP A5U4D6
L	240	HIS	-	expression tag	UNP A5U4D6
M	235	HIS	-	expression tag	UNP A5U4D6
M	236	HIS	-	expression tag	UNP A5U4D6
M	237	HIS	-	expression tag	UNP A5U4D6
M	238	HIS	-	expression tag	UNP A5U4D6
M	239	HIS	-	expression tag	UNP A5U4D6
M	240	HIS	-	expression tag	UNP A5U4D6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
N	235	HIS	-	expression tag	UNP A5U4D6
N	236	HIS	-	expression tag	UNP A5U4D6
N	237	HIS	-	expression tag	UNP A5U4D6
N	238	HIS	-	expression tag	UNP A5U4D6
N	239	HIS	-	expression tag	UNP A5U4D6
N	240	HIS	-	expression tag	UNP A5U4D6
V	235	HIS	-	expression tag	UNP A5U4D6
V	236	HIS	-	expression tag	UNP A5U4D6
V	237	HIS	-	expression tag	UNP A5U4D6
V	238	HIS	-	expression tag	UNP A5U4D6
V	239	HIS	-	expression tag	UNP A5U4D6
V	240	HIS	-	expression tag	UNP A5U4D6
W	235	HIS	-	expression tag	UNP A5U4D6
W	236	HIS	-	expression tag	UNP A5U4D6
W	237	HIS	-	expression tag	UNP A5U4D6
W	238	HIS	-	expression tag	UNP A5U4D6
W	239	HIS	-	expression tag	UNP A5U4D6
W	240	HIS	-	expression tag	UNP A5U4D6
X	235	HIS	-	expression tag	UNP A5U4D6
X	236	HIS	-	expression tag	UNP A5U4D6
X	237	HIS	-	expression tag	UNP A5U4D6
X	238	HIS	-	expression tag	UNP A5U4D6
X	239	HIS	-	expression tag	UNP A5U4D6
X	240	HIS	-	expression tag	UNP A5U4D6
Y	235	HIS	-	expression tag	UNP A5U4D6
Y	236	HIS	-	expression tag	UNP A5U4D6
Y	237	HIS	-	expression tag	UNP A5U4D6
Y	238	HIS	-	expression tag	UNP A5U4D6
Y	239	HIS	-	expression tag	UNP A5U4D6
Y	240	HIS	-	expression tag	UNP A5U4D6
Z	235	HIS	-	expression tag	UNP A5U4D6
Z	236	HIS	-	expression tag	UNP A5U4D6
Z	237	HIS	-	expression tag	UNP A5U4D6
Z	238	HIS	-	expression tag	UNP A5U4D6
Z	239	HIS	-	expression tag	UNP A5U4D6
Z	240	HIS	-	expression tag	UNP A5U4D6
a	235	HIS	-	expression tag	UNP A5U4D6
a	236	HIS	-	expression tag	UNP A5U4D6
a	237	HIS	-	expression tag	UNP A5U4D6
a	238	HIS	-	expression tag	UNP A5U4D6
a	239	HIS	-	expression tag	UNP A5U4D6
a	240	HIS	-	expression tag	UNP A5U4D6

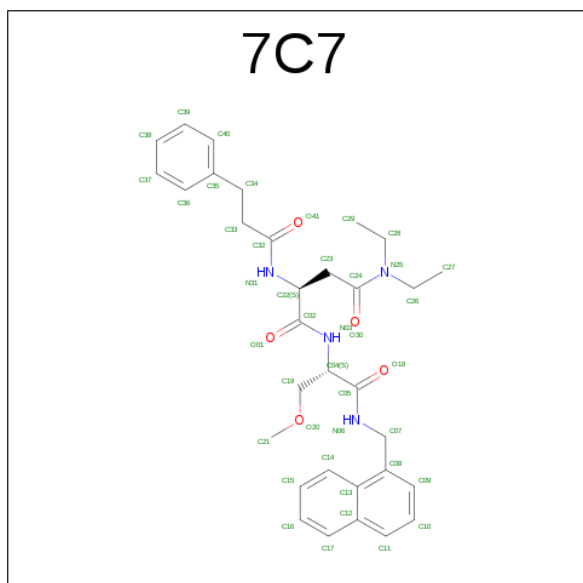
*Continued on next page...*



Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
b	235	HIS	-	expression tag	UNP A5U4D6
b	236	HIS	-	expression tag	UNP A5U4D6
b	237	HIS	-	expression tag	UNP A5U4D6
b	238	HIS	-	expression tag	UNP A5U4D6
b	239	HIS	-	expression tag	UNP A5U4D6
b	240	HIS	-	expression tag	UNP A5U4D6

- Molecule 3 is N,N-diethyl-N 2 -(3-phenylpropanoyl)-L-asparaginyl-O-methyl-N-[(naphthalen-1-yl)methyl]-L-serinamide (three-letter code: 7C7) (formula: C<sub>32</sub>H<sub>40</sub>N<sub>4</sub>O<sub>5</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	H	1	Total	C	N	O	0	0
			41	32	4	5		
3	I	1	Total	C	N	O	0	0
			41	32	4	5		
3	J	1	Total	C	N	O	0	0
			41	32	4	5		
3	K	1	Total	C	N	O	0	0
			41	32	4	5		
3	L	1	Total	C	N	O	0	0
			41	32	4	5		
3	M	1	Total	C	N	O	0	0
			41	32	4	5		
3	N	1	Total	C	N	O	0	0
			41	32	4	5		
3	V	1	Total	C	N	O	0	0
			41	32	4	5		

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	W	1	Total	C	N	O	0	0
			41	32	4	5		
3	X	1	Total	C	N	O	0	0
			41	32	4	5		
3	Y	1	Total	C	N	O	0	0
			41	32	4	5		
3	Z	1	Total	C	N	O	0	0
			41	32	4	5		
3	a	1	Total	C	N	O	0	0
			41	32	4	5		
3	b	1	Total	C	N	O	0	0
			41	32	4	5		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	9	Total	O	0	0
			9	9		
4	B	1	Total	O	0	0
			1	1		
4	C	3	Total	O	0	0
			3	3		
4	D	4	Total	O	0	0
			4	4		
4	E	3	Total	O	0	0
			3	3		
4	F	5	Total	O	0	0
			5	5		
4	G	3	Total	O	0	0
			3	3		
4	H	15	Total	O	0	0
			15	15		
4	I	18	Total	O	0	0
			18	18		
4	J	11	Total	O	0	0
			11	11		
4	K	16	Total	O	0	0
			16	16		
4	L	10	Total	O	0	0
			10	10		
4	M	12	Total	O	0	0
			12	12		

*Continued on next page...*

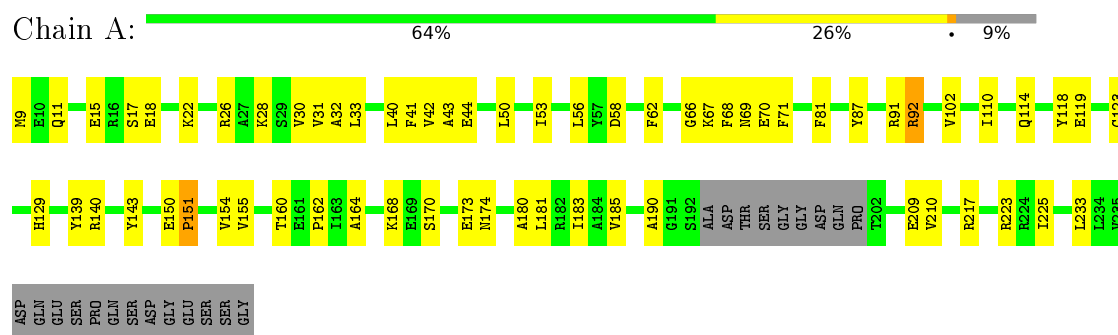
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	N	12	Total 12	O 12	0	0
4	O	6	Total 6	O 6	0	0
4	P	7	Total 7	O 7	0	0
4	Q	3	Total 3	O 3	0	0
4	R	5	Total 5	O 5	0	0
4	S	8	Total 8	O 8	0	0
4	T	6	Total 6	O 6	0	0
4	U	3	Total 3	O 3	0	0
4	V	20	Total 20	O 20	0	0
4	W	13	Total 13	O 13	0	0
4	X	10	Total 10	O 10	0	0
4	Y	10	Total 10	O 10	0	0
4	Z	12	Total 12	O 12	0	0
4	a	10	Total 10	O 10	0	0
4	b	10	Total 10	O 10	0	0

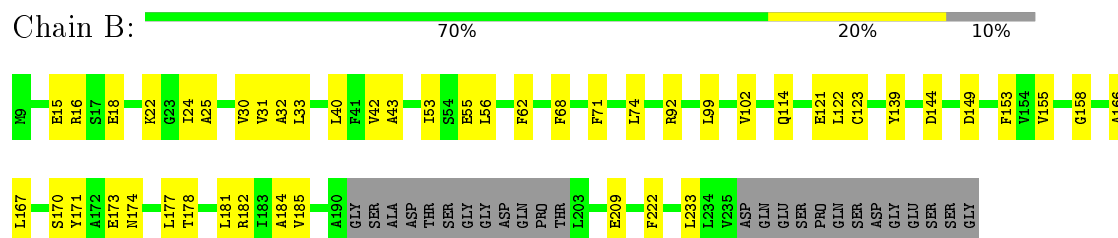
### 3 Residue-property plots

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.

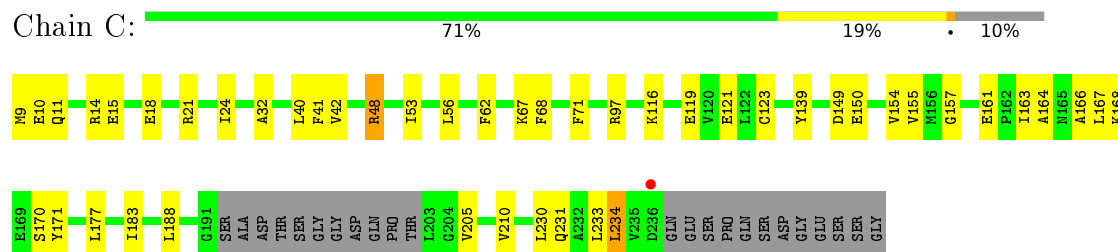
#### • Molecule 1: Proteasome subunit alpha



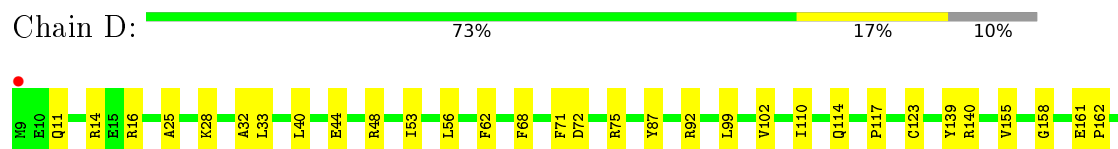
#### • Molecule 1: Proteasome subunit alpha



#### • Molecule 1: Proteasome subunit alpha



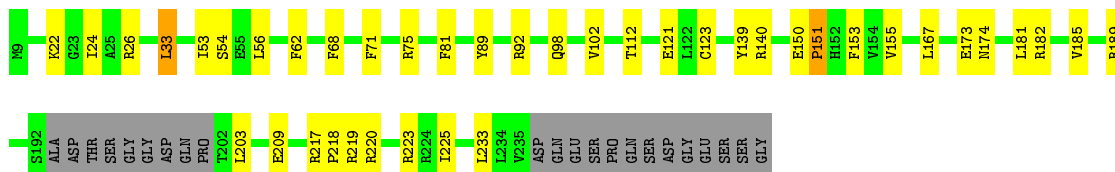
#### • Molecule 1: Proteasome subunit alpha





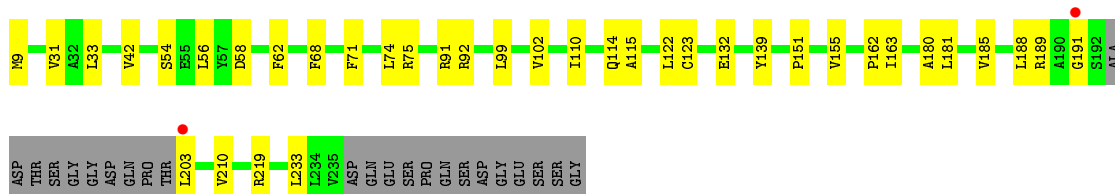
- Molecule 1: Proteasome subunit alpha

Chain E: 74% 16% 9%



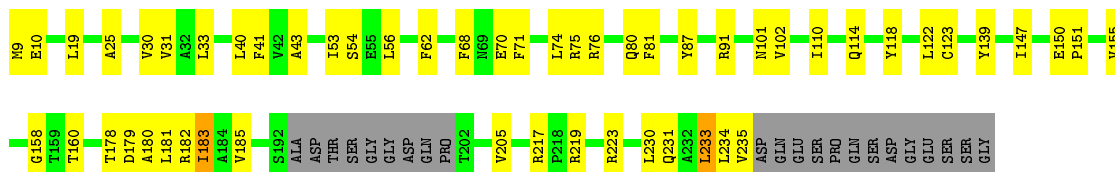
- Molecule 1: Proteasome subunit alpha

Chain F: 75% 15% 10%



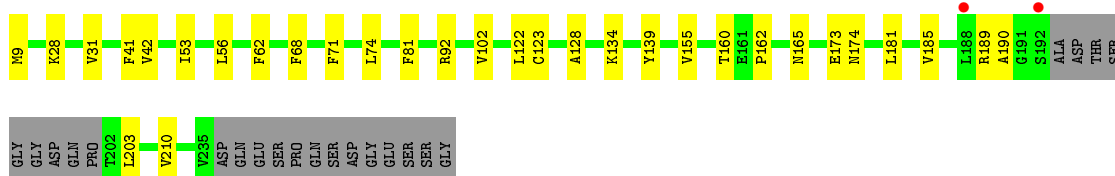
- Molecule 1: Proteasome subunit alpha

Chain G: 68% 22% 9%



- Molecule 1: Proteasome subunit alpha

Chain O: 78% 13% 9%



- Molecule 1: Proteasome subunit alpha

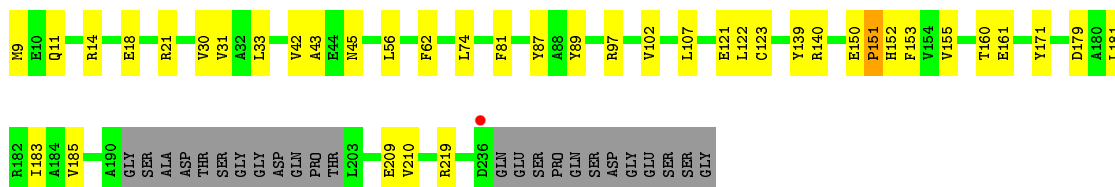
Chain P: 75% 15% 10%





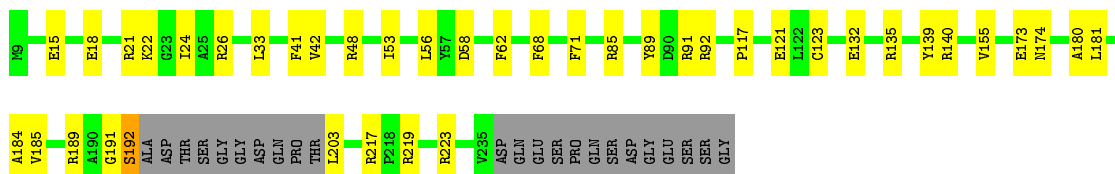
- Molecule 1: Proteasome subunit alpha

Chain Q: 73% 16% 10%



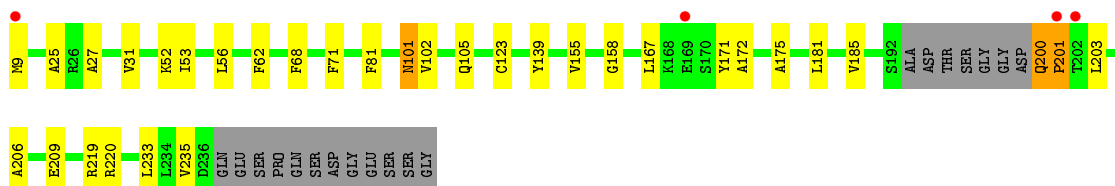
- Molecule 1: Proteasome subunit alpha

Chain R: 73% 17% 10%



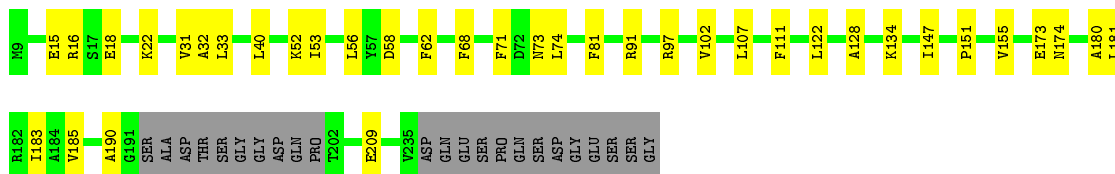
- Molecule 1: Proteasome subunit alpha

Chain S: 2% 78% 13% 8%



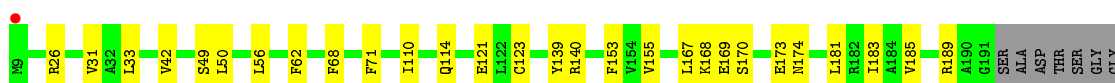
- Molecule 1: Proteasome subunit alpha

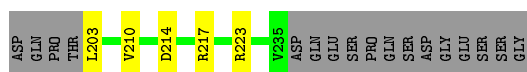
Chain T: 75% 15% 10%



- Molecule 1: Proteasome subunit alpha

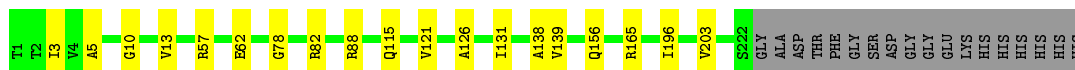
Chain U: 76% 14% 10%





- Molecule 2: Proteasome subunit beta

Chain H: 85% 8% 8%



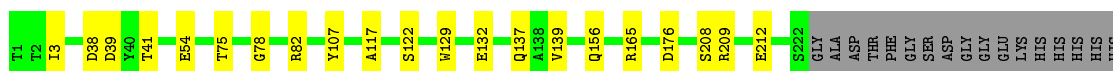
- Molecule 2: Proteasome subunit beta

Chain I: 90% 0% 0%



- Molecule 2: Proteasome subunit beta

Chain J: 84% 9% 8%



HIS

- Molecule 2: Proteasome subunit beta

Chain K: 85% 8% 7%



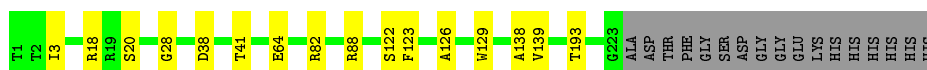
- Molecule 2: Proteasome subunit beta

Chain L: 86% 7% 7%



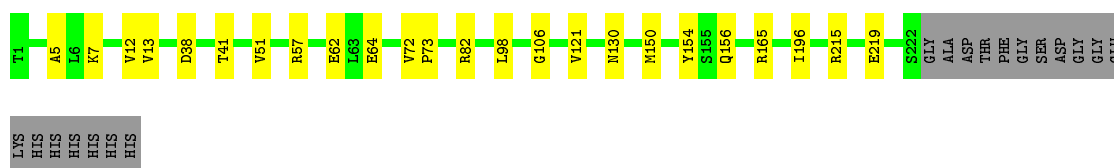
- Molecule 2: Proteasome subunit beta

Chain M: 86% 7% 7%



- Molecule 2: Proteasome subunit beta

Chain N: 83% 10% 8%



- Molecule 2: Proteasome subunit beta

Chain V: 85% 8% 7%



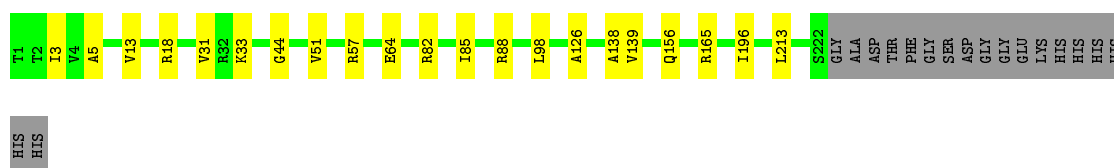
- Molecule 2: Proteasome subunit beta

Chain W: 86% 7% 7%



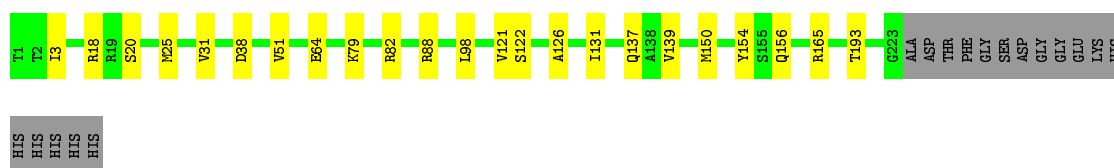
- Molecule 2: Proteasome subunit beta

Chain X: 84% 9% 8%



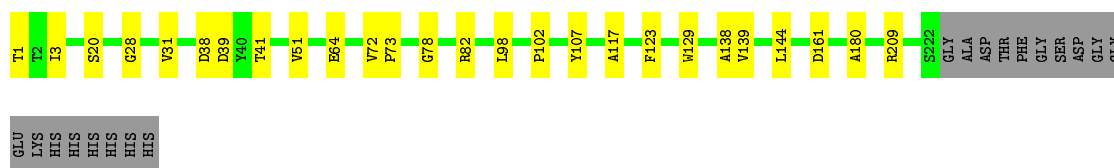
- Molecule 2: Proteasome subunit beta

Chain Y: 83% 10% 7%



- Molecule 2: Proteasome subunit beta

Chain Z: 82% 11% 8%



- Molecule 2: Proteasome subunit beta

Chain a: 93% 7% 0%



TI	6223
	ALA
	ASP
	THR
	PHE
	GLY
	SER
	ASP
	GLY
	GLU
	LYS
	HIS
	HIS
	HIS
	HIS
	HIS

- Molecule 2: Proteasome subunit beta

Chain b: 

93%

7%

TI	6223
	ALA
	ASP
	THR
	PHE
	GLY
	SER
	ASP
	GLY
	GLU
	LYS
	HIS
	HIS
	HIS
	HIS
	HIS

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	120.16Å 198.19Å 166.24Å 90.00° 103.35° 90.00°	Depositor
Resolution (Å)	49.55 – 3.00 49.68 – 3.00	Depositor EDS
% Data completeness (in resolution range)	97.5 (49.55-3.00) 93.3 (49.68-3.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.20 (at 3.01Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.176 , 0.210 0.176 , 0.208	Depositor DCC
$R_{free}$ test set	6990 reflections (4.97%)	DCC
Wilson B-factor (Å <sup>2</sup> )	59.9	Xtriage
Anisotropy	0.004	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 30.0	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	0.94	EDS
Total number of atoms	47201	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.75% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 7C7

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	0.32	0/1701	0.54	0/2297
1	B	0.30	0/1684	0.51	0/2274
1	C	0.30	0/1696	0.52	0/2290
1	D	0.31	0/1687	0.53	0/2277
1	E	0.30	0/1701	0.54	1/2297 (0.0%)
1	F	0.30	0/1694	0.53	0/2287
1	G	0.34	0/1701	0.53	0/2297
1	O	0.29	0/1701	0.50	0/2297
1	P	0.31	0/1695	0.51	0/2289
1	Q	0.31	0/1692	0.52	0/2285
1	R	0.30	0/1694	0.52	0/2287
1	S	0.33	0/1726	0.56	0/2332
1	T	0.30	0/1695	0.52	0/2289
1	U	0.30	0/1688	0.54	0/2279
2	H	0.30	0/1662	0.52	0/2254
2	I	0.30	0/1662	0.52	0/2254
2	J	0.30	0/1662	0.51	0/2254
2	K	0.30	0/1666	0.53	0/2259
2	L	0.30	0/1666	0.54	0/2259
2	M	0.30	0/1666	0.53	0/2259
2	N	0.31	0/1662	0.53	0/2254
2	V	0.30	0/1666	0.52	0/2259
2	W	0.30	0/1666	0.52	0/2259
2	X	0.29	0/1662	0.54	0/2254
2	Y	0.30	0/1666	0.51	0/2259
2	Z	0.30	0/1662	0.53	0/2254
2	a	0.30	0/1666	0.53	0/2259
2	b	0.30	0/1666	0.53	0/2259
All	All	0.31	0/47055	0.53	1/63673 (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	E	33	LEU	CA-CB-CG	5.54	128.03	115.30

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	1677	0	1680	39	0
1	B	1660	0	1665	29	0
1	C	1672	0	1672	32	0
1	D	1663	0	1664	26	0
1	E	1677	0	1680	23	0
1	F	1670	0	1673	26	0
1	G	1677	0	1680	41	0
1	O	1677	0	1680	19	0
1	P	1671	0	1675	26	0
1	Q	1668	0	1669	28	0
1	R	1670	0	1673	28	0
1	S	1701	0	1699	29	0
1	T	1671	0	1675	24	0
1	U	1664	0	1668	23	0
2	H	1638	0	1633	11	0
2	I	1638	0	1633	4	0
2	J	1638	0	1633	14	0
2	K	1642	0	1636	11	0
2	L	1642	0	1636	10	0
2	M	1642	0	1636	11	0
2	N	1638	0	1633	15	0
2	V	1642	0	1636	11	0
2	W	1642	0	1636	10	0
2	X	1638	0	1633	14	0
2	Y	1642	0	1636	13	0
2	Z	1638	0	1633	18	0
2	a	1642	0	1636	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	b	1642	0	1636	0	0
3	H	41	0	0	0	0
3	I	41	0	0	0	0
3	J	41	0	0	0	0
3	K	41	0	0	0	0
3	L	41	0	0	0	0
3	M	41	0	0	0	0
3	N	41	0	0	0	0
3	V	41	0	0	0	0
3	W	41	0	0	0	0
3	X	41	0	0	0	0
3	Y	41	0	0	0	0
3	Z	41	0	0	0	0
3	a	41	0	0	0	0
3	b	41	0	0	0	0
4	A	9	0	0	0	0
4	B	1	0	0	0	0
4	C	3	0	0	0	0
4	D	4	0	0	1	0
4	E	3	0	0	0	0
4	F	5	0	0	0	0
4	G	3	0	0	0	0
4	H	15	0	0	0	0
4	I	18	0	0	1	0
4	J	11	0	0	1	0
4	K	16	0	0	0	0
4	L	10	0	0	0	0
4	M	12	0	0	0	0
4	N	12	0	0	0	0
4	O	6	0	0	0	0
4	P	7	0	0	0	0
4	Q	3	0	0	0	0
4	R	5	0	0	0	0
4	S	8	0	0	2	0
4	T	6	0	0	1	0
4	U	3	0	0	2	0
4	V	20	0	0	1	0
4	W	13	0	0	0	0
4	X	10	0	0	1	0
4	Y	10	0	0	0	0
4	Z	12	0	0	0	0
4	a	10	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	b	10	0	0	0	0
All	All	47201	0	46339	496	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 6.

The worst 5 of 496 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:S:200:GLN:HG3	1:S:201:PRO:CD	1.38	1.49
1:G:179:ASP:O	1:G:183:ILE:HG13	1.38	1.21
1:S:200:GLN:CG	1:S:201:PRO:CD	2.19	1.19
1:S:200:GLN:CG	1:S:201:PRO:HD2	1.74	1.18
1:F:189:ARG:HG3	1:F:203:LEU:CD2	1.85	1.05

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	214/240 (89%)	207 (97%)	6 (3%)	1 (0%)	34	76
1	B	211/240 (88%)	204 (97%)	7 (3%)	0	100	100
1	C	213/240 (89%)	209 (98%)	3 (1%)	1 (0%)	34	76
1	D	212/240 (88%)	204 (96%)	8 (4%)	0	100	100
1	E	214/240 (89%)	206 (96%)	7 (3%)	1 (0%)	34	76
1	F	213/240 (89%)	206 (97%)	6 (3%)	1 (0%)	34	76
1	G	214/240 (89%)	208 (97%)	6 (3%)	0	100	100
1	O	214/240 (89%)	206 (96%)	8 (4%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	P	213/240 (89%)	206 (97%)	7 (3%)	0	100	100
1	Q	212/240 (88%)	207 (98%)	4 (2%)	1 (0%)	34	76
1	R	213/240 (89%)	205 (96%)	8 (4%)	0	100	100
1	S	217/240 (90%)	210 (97%)	4 (2%)	3 (1%)	14	51
1	T	213/240 (89%)	204 (96%)	7 (3%)	2 (1%)	21	64
1	U	212/240 (88%)	203 (96%)	9 (4%)	0	100	100
2	H	220/240 (92%)	217 (99%)	3 (1%)	0	100	100
2	I	220/240 (92%)	217 (99%)	3 (1%)	0	100	100
2	J	220/240 (92%)	217 (99%)	3 (1%)	0	100	100
2	K	221/240 (92%)	217 (98%)	4 (2%)	0	100	100
2	L	221/240 (92%)	218 (99%)	3 (1%)	0	100	100
2	M	221/240 (92%)	216 (98%)	5 (2%)	0	100	100
2	N	220/240 (92%)	216 (98%)	4 (2%)	0	100	100
2	V	221/240 (92%)	218 (99%)	3 (1%)	0	100	100
2	W	221/240 (92%)	218 (99%)	3 (1%)	0	100	100
2	X	220/240 (92%)	216 (98%)	4 (2%)	0	100	100
2	Y	221/240 (92%)	218 (99%)	3 (1%)	0	100	100
2	Z	220/240 (92%)	217 (99%)	3 (1%)	0	100	100
2	a	221/240 (92%)	217 (98%)	4 (2%)	0	100	100
2	b	221/240 (92%)	217 (98%)	4 (2%)	0	100	100
All	All	6073/6720 (90%)	5924 (98%)	139 (2%)	10 (0%)	52	88

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	S	201	PRO
1	T	190	ALA
1	S	206	ALA
1	C	234	LEU
1	F	151	PRO

### 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	167/184 (91%)	166 (99%)	1 (1%)	90	97
1	B	165/184 (90%)	165 (100%)	0	100	100
1	C	166/184 (90%)	164 (99%)	2 (1%)	78	94
1	D	165/184 (90%)	163 (99%)	2 (1%)	78	94
1	E	167/184 (91%)	165 (99%)	2 (1%)	78	94
1	F	166/184 (90%)	166 (100%)	0	100	100
1	G	167/184 (91%)	162 (97%)	5 (3%)	48	83
1	O	167/184 (91%)	167 (100%)	0	100	100
1	P	166/184 (90%)	164 (99%)	2 (1%)	78	94
1	Q	166/184 (90%)	165 (99%)	1 (1%)	90	97
1	R	166/184 (90%)	165 (99%)	1 (1%)	90	97
1	S	170/184 (92%)	167 (98%)	3 (2%)	66	91
1	T	166/184 (90%)	166 (100%)	0	100	100
1	U	165/184 (90%)	165 (100%)	0	100	100
2	H	165/178 (93%)	165 (100%)	0	100	100
2	I	165/178 (93%)	165 (100%)	0	100	100
2	J	165/178 (93%)	165 (100%)	0	100	100
2	K	165/178 (93%)	165 (100%)	0	100	100
2	L	165/178 (93%)	165 (100%)	0	100	100
2	M	165/178 (93%)	165 (100%)	0	100	100
2	N	165/178 (93%)	165 (100%)	0	100	100
2	V	165/178 (93%)	165 (100%)	0	100	100
2	W	165/178 (93%)	165 (100%)	0	100	100
2	X	165/178 (93%)	165 (100%)	0	100	100
2	Y	165/178 (93%)	165 (100%)	0	100	100
2	Z	165/178 (93%)	165 (100%)	0	100	100
2	a	165/178 (93%)	165 (100%)	0	100	100
2	b	165/178 (93%)	165 (100%)	0	100	100
All	All	4639/5068 (92%)	4620 (100%)	19 (0%)	93	98



5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	101	ASN
1	G	233	LEU
1	R	192	SER
1	G	91	ARG
1	S	101	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	129	HIS
1	S	200	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](#)

14 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	7C7	H	301	-	43,43,43	1.21	6 (13%)	55,56,56	1.54	13 (23%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	7C7	I	301	-	43,43,43	1.12	5 (11%)	55,56,56	1.38	8 (14%)
3	7C7	J	301	-	43,43,43	1.16	5 (11%)	55,56,56	1.46	10 (18%)
3	7C7	K	301	-	43,43,43	1.09	5 (11%)	55,56,56	1.39	11 (20%)
3	7C7	L	301	-	43,43,43	1.05	5 (11%)	55,56,56	1.41	12 (21%)
3	7C7	M	301	-	43,43,43	1.07	4 (9%)	55,56,56	1.40	9 (16%)
3	7C7	N	301	-	43,43,43	1.03	4 (9%)	55,56,56	1.43	11 (20%)
3	7C7	V	301	-	43,43,43	1.18	5 (11%)	55,56,56	1.53	12 (21%)
3	7C7	W	301	-	43,43,43	1.04	4 (9%)	55,56,56	1.47	9 (16%)
3	7C7	X	301	-	43,43,43	1.15	5 (11%)	55,56,56	1.53	12 (21%)
3	7C7	Y	301	-	43,43,43	1.09	5 (11%)	55,56,56	1.38	11 (20%)
3	7C7	Z	301	-	43,43,43	0.97	3 (6%)	55,56,56	1.45	10 (18%)
3	7C7	a	301	-	43,43,43	0.95	4 (9%)	55,56,56	1.36	8 (14%)
3	7C7	b	301	-	43,43,43	1.09	5 (11%)	55,56,56	1.43	11 (20%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	7C7	H	301	-	-	0/41/41/41	0/3/3/3
3	7C7	I	301	-	-	0/41/41/41	0/3/3/3
3	7C7	J	301	-	-	0/41/41/41	0/3/3/3
3	7C7	K	301	-	-	0/41/41/41	0/3/3/3
3	7C7	L	301	-	-	0/41/41/41	0/3/3/3
3	7C7	M	301	-	-	0/41/41/41	0/3/3/3
3	7C7	N	301	-	-	0/41/41/41	0/3/3/3
3	7C7	V	301	-	-	0/41/41/41	0/3/3/3
3	7C7	W	301	-	-	0/41/41/41	0/3/3/3
3	7C7	X	301	-	-	0/41/41/41	0/3/3/3
3	7C7	Y	301	-	-	0/41/41/41	0/3/3/3
3	7C7	Z	301	-	-	0/41/41/41	0/3/3/3
3	7C7	a	301	-	-	0/41/41/41	0/3/3/3
3	7C7	b	301	-	-	0/41/41/41	0/3/3/3

The worst 5 of 65 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	W	301	7C7	C02-N03	2.00	1.38	1.34

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	Y	301	7C7	C23-C22	2.02	1.58	1.53
3	a	301	7C7	C23-C22	2.02	1.58	1.53
3	X	301	7C7	C23-C22	2.04	1.58	1.53
3	H	301	7C7	C02-N03	2.04	1.38	1.34

The worst 5 of 147 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	W	301	7C7	O41-C32-N31	-3.48	117.10	122.96
3	X	301	7C7	O41-C32-N31	-3.46	117.15	122.96
3	V	301	7C7	O41-C32-N31	-3.19	117.60	122.96
3	J	301	7C7	O41-C32-N31	-3.18	117.62	122.96
3	L	301	7C7	O41-C32-N31	-3.15	117.66	122.96

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	218/240 (90%)	-0.50	0 100 100	42, 56, 84, 107	0
1	B	215/240 (89%)	-0.37	0 100 100	40, 69, 102, 124	0
1	C	217/240 (90%)	-0.26	1 (0%) 91 76	44, 73, 111, 140	0
1	D	216/240 (90%)	-0.25	1 (0%) 91 76	46, 69, 103, 121	0
1	E	218/240 (90%)	-0.37	0 100 100	43, 64, 96, 129	0
1	F	217/240 (90%)	-0.40	2 (0%) 85 64	44, 67, 101, 132	0
1	G	218/240 (90%)	-0.49	0 100 100	41, 59, 94, 118	0
1	O	218/240 (90%)	-0.31	2 (0%) 85 64	43, 69, 113, 131	0
1	P	217/240 (90%)	-0.40	0 100 100	40, 65, 99, 128	0
1	Q	216/240 (90%)	-0.44	1 (0%) 91 76	41, 62, 88, 129	0
1	R	217/240 (90%)	-0.42	0 100 100	38, 60, 89, 128	0
1	S	221/240 (92%)	-0.39	4 (1%) 71 43	39, 57, 98, 160	0
1	T	217/240 (90%)	-0.34	0 100 100	44, 66, 96, 114	0
1	U	216/240 (90%)	-0.46	1 (0%) 91 76	41, 61, 92, 119	0
2	H	222/240 (92%)	-0.49	0 100 100	39, 49, 71, 96	0
2	I	222/240 (92%)	-0.62	0 100 100	38, 46, 66, 93	0
2	J	222/240 (92%)	-0.64	0 100 100	40, 48, 71, 92	0
2	K	223/240 (92%)	-0.64	0 100 100	39, 47, 66, 87	0
2	L	223/240 (92%)	-0.62	0 100 100	38, 47, 73, 107	0
2	M	223/240 (92%)	-0.58	0 100 100	37, 49, 71, 97	0
2	N	222/240 (92%)	-0.62	0 100 100	41, 51, 75, 96	0
2	V	223/240 (92%)	-0.61	0 100 100	38, 45, 66, 79	0
2	W	223/240 (92%)	-0.65	0 100 100	39, 46, 66, 93	0
2	X	222/240 (92%)	-0.67	0 100 100	39, 48, 68, 85	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
2	Y	223/240 (92%)	-0.63	0 100 100	38, 47, 67, 90	0
2	Z	222/240 (92%)	-0.60	0 100 100	38, 49, 70, 94	0
2	a	223/240 (92%)	-0.61	0 100 100	40, 50, 73, 94	0
2	b	223/240 (92%)	-0.57	0 100 100	37, 49, 72, 84	0
All	All	6157/6720 (91%)	-0.50	12 (0%) 95 87	37, 54, 93, 160	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	S	201	PRO	3.5
1	C	236	ASP	3.1
1	D	9	MET	2.8
1	F	191	GLY	2.8
1	O	192	SER	2.5

## 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	7C7	N	301	41/41	0.94	0.22	2.09	39,46,56,59	0
3	7C7	W	301	41/41	0.96	0.20	1.74	37,42,53,65	0
3	7C7	H	301	41/41	0.95	0.20	1.41	38,43,52,59	0
3	7C7	b	301	41/41	0.95	0.20	1.37	40,44,48,55	0
3	7C7	M	301	41/41	0.95	0.20	1.28	36,45,52,60	0

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	7C7	Z	301	41/41	0.95	0.19	1.13	40,44,55,60	0
3	7C7	J	301	41/41	0.94	0.19	1.10	38,42,49,56	0
3	7C7	X	301	41/41	0.94	0.21	1.09	39,44,58,67	0
3	7C7	Y	301	41/41	0.95	0.19	0.98	33,43,48,52	0
3	7C7	I	301	41/41	0.96	0.18	0.90	36,43,50,57	0
3	7C7	L	301	41/41	0.94	0.19	0.89	38,44,53,57	0
3	7C7	V	301	41/41	0.96	0.19	0.85	38,42,46,59	0
3	7C7	a	301	41/41	0.96	0.18	0.47	42,44,55,57	0
3	7C7	K	301	41/41	0.95	0.18	0.42	42,46,55,61	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.