



wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Aug 22, 2017 – 08:32 PM EDT

PDB ID : 5O4U
EMDB ID: : EMD-3746
Title : The flagellin of *Pyrococcus furiosus*
Authors : Daum, B.; Vonck, J.
Deposited on : unknown
Resolution : 4.20 Å(reported)
Based on PDB ID : 5TFY

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report
for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

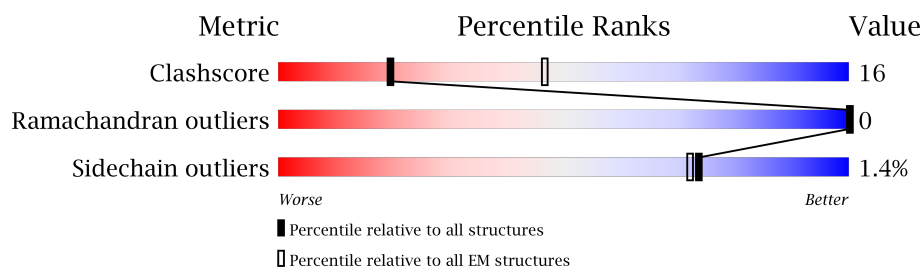
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 4.20 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)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for ≥ 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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	212	
1	B	212	
1	C	212	
1	D	212	
1	E	212	
1	F	212	
1	G	212	
1	H	212	
1	I	212	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	J	212	<div><div></div><div>69%</div><div>28%</div><div>..</div></div>
1	K	212	<div><div></div><div>61%</div><div>35%</div><div>..</div></div>
1	L	212	<div><div></div><div>64%</div><div>33%</div><div>..</div></div>
1	M	212	<div><div></div><div>64%</div><div>33%</div><div>.</div></div>
1	N	212	<div><div></div><div>64%</div><div>34%</div><div>.</div></div>
1	O	212	<div><div></div><div>68%</div><div>28%</div><div>..</div></div>
1	P	212	<div><div></div><div>69%</div><div>29%</div><div>.</div></div>

2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 24528 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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.

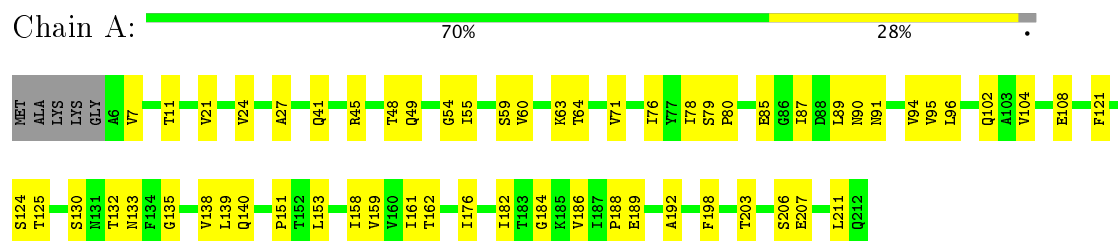
- Molecule 1 is a protein called Flagellin.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	B	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	C	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	D	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	E	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	F	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	G	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	H	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	I	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	J	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	K	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	L	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	M	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	N	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	O	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		
1	P	207	Total	C	N	O	S	0	0
			1533	975	254	303	1		

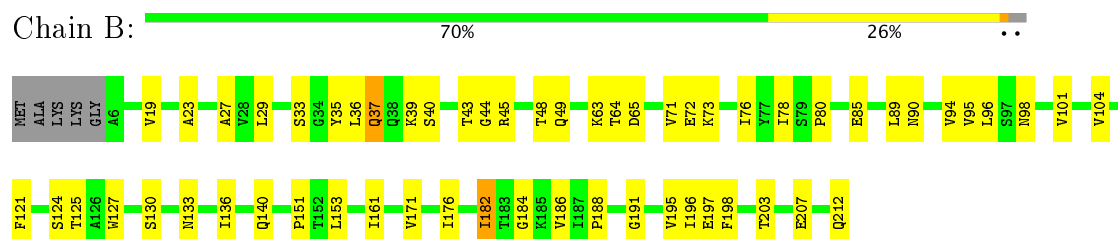
3 Residue-property plots [i](#)

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. 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. 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.

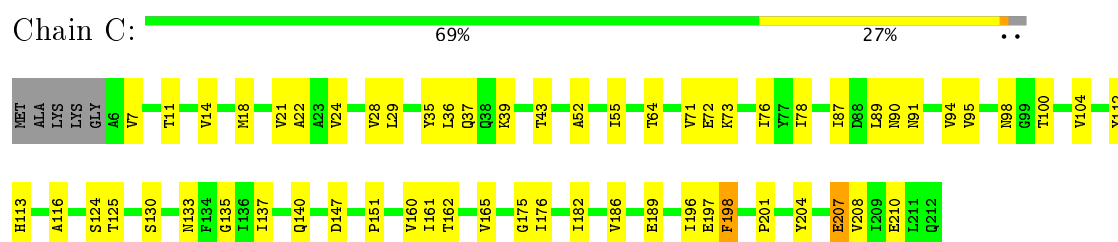
• Molecule 1: Flagellin



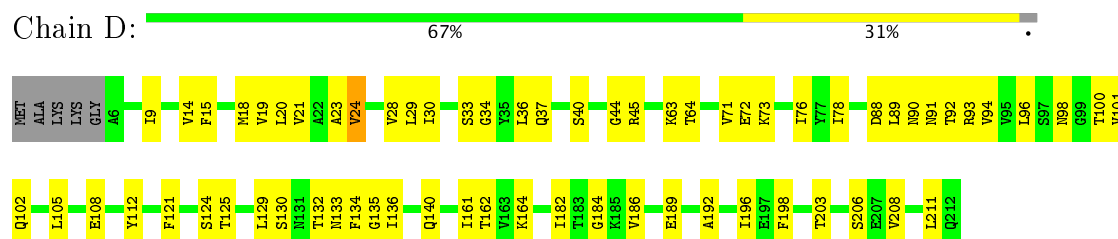
• Molecule 1: Flagellin




• Molecule 1: Flagellin

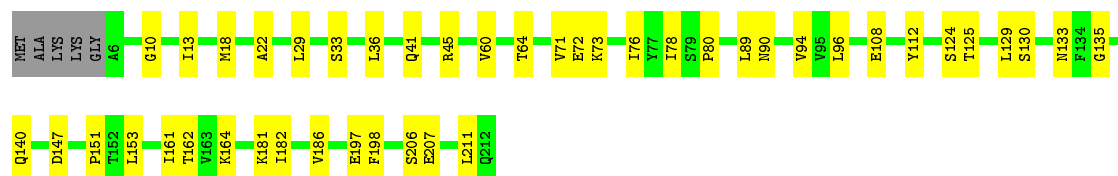


• Molecule 1: Flagellin



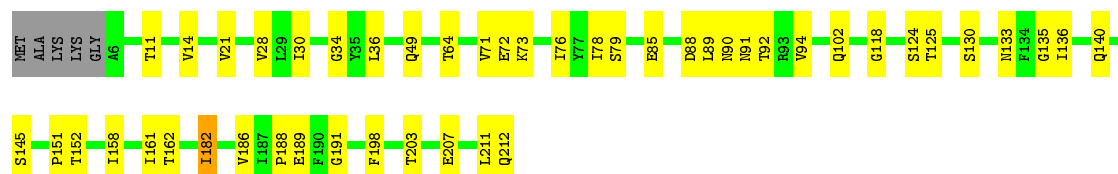
• Molecule 1: Flagellin

Chain E:  77% 21% .



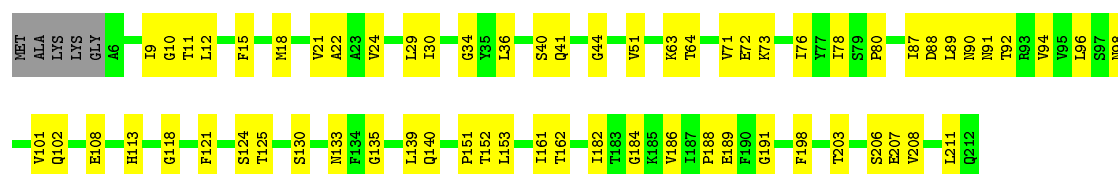
- Molecule 1: Flagellin

Chain F:  75% 22% .



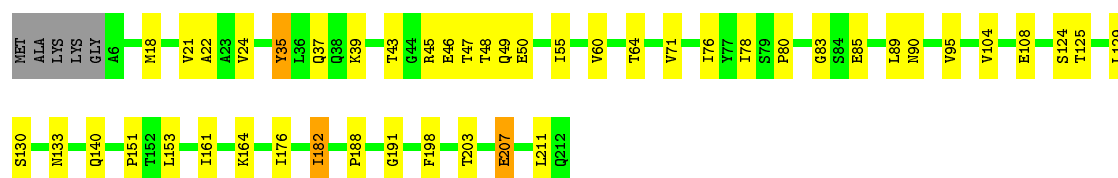
- Molecule 1: Flagellin

Chain G:  67% 30% .



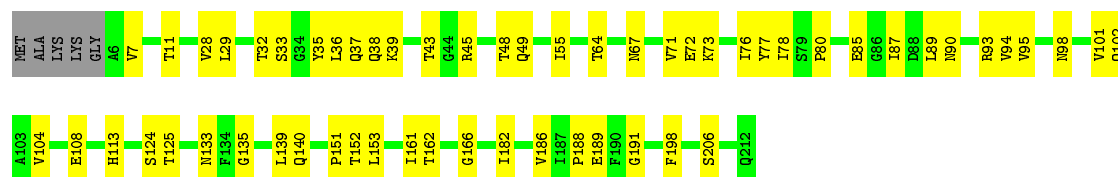
- Molecule 1: Flagellin

Chain H:  76% 20% ..



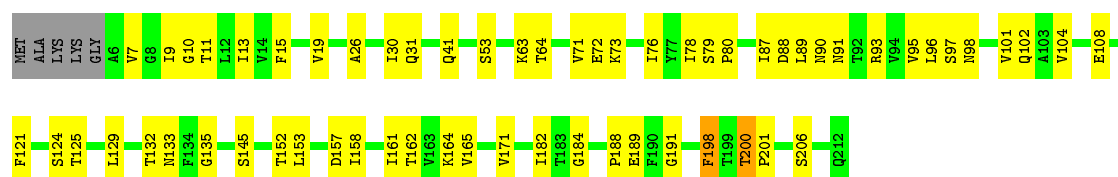
- Molecule 1: Flagellin

Chain I:  71% 27% .



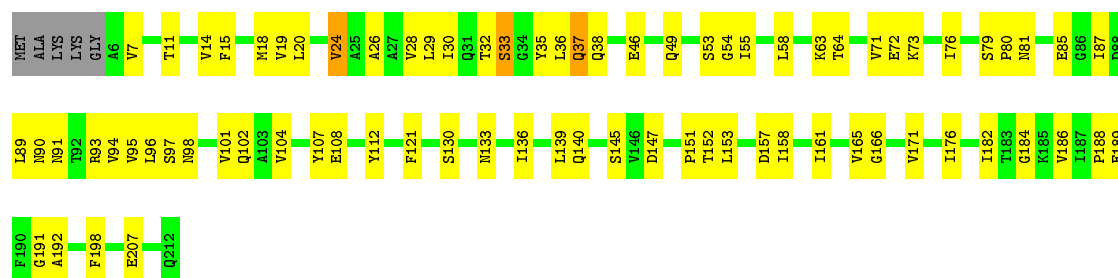
- Molecule 1: Flagellin

Chain J:  69% 28% ..



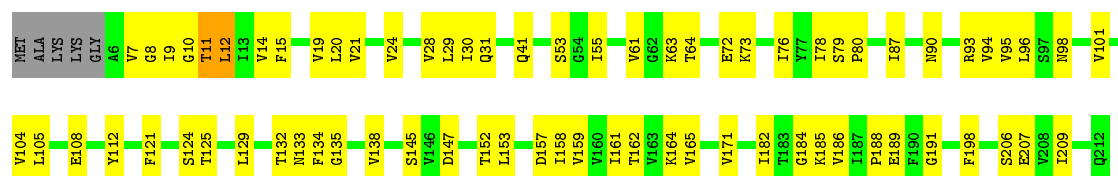
• Molecule 1: Flagellin

Chain K: 61% 35% ..



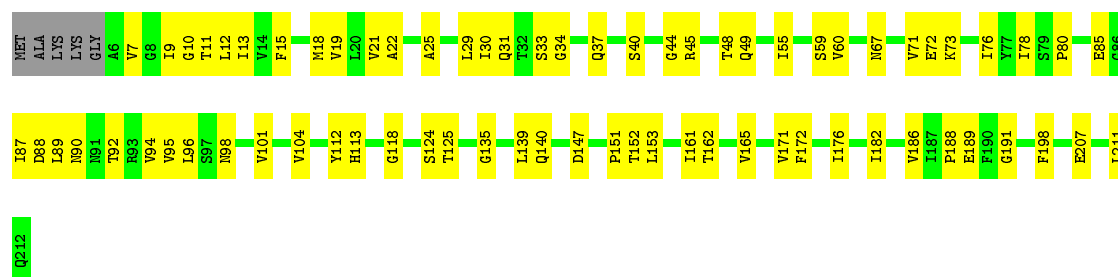
• Molecule 1: Flagellin

Chain L: 64% 33% ..



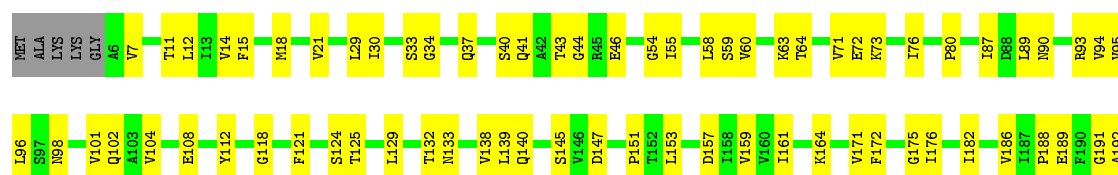
• Molecule 1: Flagellin

Chain M: 64% 33% .



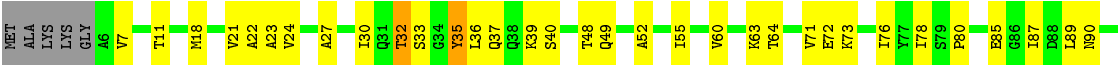
• Molecule 1: Flagellin

Chain N: 64% 34% .





• Molecule 1: Flagellin



• Molecule 1: Flagellin



4 Experimental information

Property	Value	Source
Reconstruction method	HELICAL	Depositor
Imposed symmetry	HELICAL, twist=108°, rise=5.5 Å, axial sym=C1	Depositor
Number of segments used	13965	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	JEOL 3200FSC	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	60	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	44642	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

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 >2	RMSZ	# Z >2
1	A	0.36	0/1555	0.67	0/2119
1	B	0.35	0/1555	0.67	0/2119
1	C	0.36	0/1555	0.69	0/2119
1	D	0.36	0/1555	0.66	0/2119
1	E	0.35	0/1555	0.67	0/2119
1	F	0.36	0/1555	0.66	0/2119
1	G	0.36	0/1555	0.67	0/2119
1	H	0.34	0/1555	0.67	0/2119
1	I	0.37	0/1555	0.67	1/2119 (0.0%)
1	J	0.35	0/1555	0.67	1/2119 (0.0%)
1	K	0.35	0/1555	0.67	0/2119
1	L	0.35	0/1555	0.66	0/2119
1	M	0.37	0/1555	0.67	1/2119 (0.0%)
1	N	0.36	0/1555	0.66	0/2119
1	O	0.37	0/1555	0.68	0/2119
1	P	0.34	0/1555	0.64	0/2119
All	All	0.36	0/24880	0.67	3/33904 (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	B	0	1
1	C	0	1
1	D	0	1
1	G	0	1
1	H	0	1
1	I	0	1
1	L	0	1
1	M	0	1
1	P	0	1

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
All	All	0	9

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	J	7	VAL	N-CA-C	5.54	125.95	111.00
1	M	29	LEU	CA-CB-CG	5.02	126.84	115.30
1	I	166	GLY	N-CA-C	5.02	125.64	113.10

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	207	GLU	Peptide
1	C	207	GLU	Peptide
1	D	208	VAL	Peptide
1	G	208	VAL	Peptide
1	H	207	GLU	Peptide

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	1533	0	1568	54	0
1	B	1533	0	1566	67	0
1	C	1533	0	1568	88	0
1	D	1533	0	1568	65	0
1	E	1533	0	1566	32	0
1	F	1533	0	1568	39	0
1	G	1533	0	1568	62	0
1	H	1533	0	1568	52	0
1	I	1533	0	1568	93	0
1	J	1533	0	1568	73	0
1	K	1533	0	1567	89	0
1	L	1533	0	1568	69	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	M	1533	0	1568	78	0
1	N	1533	0	1568	59	0
1	O	1533	0	1568	70	0
1	P	1533	0	1568	51	0
All	All	24528	0	25083	777	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 16.

The worst 5 of 777 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:I:37:GLN:HG3	1:K:28:VAL:CG2	1.20	1.58
1:A:48:THR:CG2	1:I:36:LEU:HD21	1.29	1.53
1:A:48:THR:HG21	1:I:36:LEU:CD2	1.46	1.46
1:C:37:GLN:HE22	1:G:24:VAL:CB	1.30	1.45
1:I:37:GLN:CG	1:K:28:VAL:CG2	2.02	1.35

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 PDB entries followed by that with respect to all EM entries.

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	205/212 (97%)	174 (85%)	31 (15%)	0	100	100
1	B	205/212 (97%)	164 (80%)	41 (20%)	0	100	100
1	C	205/212 (97%)	165 (80%)	40 (20%)	0	100	100
1	D	205/212 (97%)	169 (82%)	36 (18%)	0	100	100
1	E	205/212 (97%)	167 (82%)	38 (18%)	0	100	100
1	F	205/212 (97%)	169 (82%)	36 (18%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	G	205/212 (97%)	171 (83%)	34 (17%)	0	100	100
1	H	205/212 (97%)	168 (82%)	37 (18%)	0	100	100
1	I	205/212 (97%)	172 (84%)	33 (16%)	0	100	100
1	J	205/212 (97%)	174 (85%)	31 (15%)	0	100	100
1	K	205/212 (97%)	170 (83%)	35 (17%)	0	100	100
1	L	205/212 (97%)	170 (83%)	35 (17%)	0	100	100
1	M	205/212 (97%)	170 (83%)	35 (17%)	0	100	100
1	N	205/212 (97%)	171 (83%)	34 (17%)	0	100	100
1	O	205/212 (97%)	169 (82%)	36 (18%)	0	100	100
1	P	205/212 (97%)	171 (83%)	34 (17%)	0	100	100
All	All	3280/3392 (97%)	2714 (83%)	566 (17%)	0	100	100

There are no Ramachandran outliers to report.

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 PDB entries followed by that with respect to all EM entries.

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	168/171 (98%)	167 (99%)	1 (1%)	89	94
1	B	168/171 (98%)	165 (98%)	3 (2%)	64	84
1	C	168/171 (98%)	166 (99%)	2 (1%)	75	88
1	D	168/171 (98%)	165 (98%)	3 (2%)	64	84
1	E	168/171 (98%)	167 (99%)	1 (1%)	89	94
1	F	168/171 (98%)	166 (99%)	2 (1%)	75	88
1	G	168/171 (98%)	167 (99%)	1 (1%)	89	94
1	H	168/171 (98%)	165 (98%)	3 (2%)	64	84
1	I	168/171 (98%)	166 (99%)	2 (1%)	75	88
1	J	168/171 (98%)	165 (98%)	3 (2%)	64	84
1	K	168/171 (98%)	163 (97%)	5 (3%)	46	73

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	L	168/171 (98%)	164 (98%)	4 (2%)	54	78
1	M	168/171 (98%)	167 (99%)	1 (1%)	89	94
1	N	168/171 (98%)	167 (99%)	1 (1%)	89	94
1	O	168/171 (98%)	164 (98%)	4 (2%)	54	78
1	P	168/171 (98%)	167 (99%)	1 (1%)	89	94
All	All	2688/2736 (98%)	2651 (99%)	37 (1%)	74	87

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	32	THR
1	J	200	THR
1	O	90	ASN
1	I	90	ASN
1	J	90	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 46 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	140	GLN
1	I	102	GLN
1	O	140	GLN
1	H	37	GLN
1	H	140	GLN

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 ⓘ

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.