



## wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Feb 20, 2018 – 09:44 pm GMT

PDB ID : 6F9B  
EMDB ID: : EMD-4197  
Title : Asymmetric unit of Rift Valley fever virus glycoprotein shell  
Authors : Halldorsson, S.; Bowden, T.A.; Huiskonen, J.T.  
Deposited on : 2017-12-14  
Resolution : 13.30 Å(reported)

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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30686

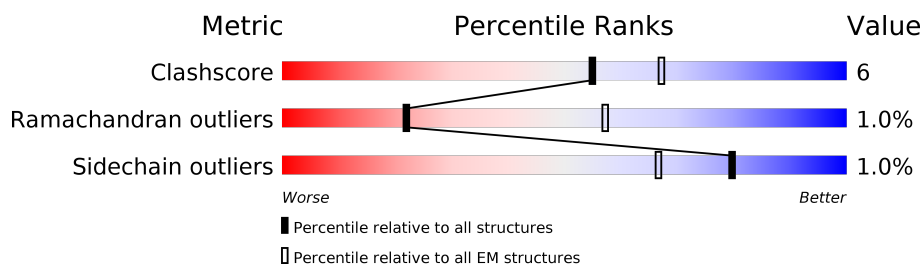
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 13.30 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	136279	1886
Ramachandran outliers	132675	1663
Sidechain outliers	132484	1531

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	316	80% 15% 5%
1	C	316	83% 12% 5%
1	E	316	83% 11% • 5%
1	G	316	82% 14% 5%
1	I	316	79% 15% • 5%
1	K	316	82% 13% • 5%
1	M	316	80% 16% 5%
1	O	316	79% 16% • 5%
1	Q	316	82% 13% 5%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	S	316	 80% 15% 5%
1	U	316	 76% 18% 5%
1	X	316	 76% 19% 5%
2	B	431	 83% 17%
2	D	431	 80% 19%
2	F	431	 82% 18%
2	H	431	 79% 20%
2	J	431	 85% 15%
2	L	431	 83% 17%
2	N	431	 82% 17%
2	P	431	 83% 17%
2	R	431	 82% 18%
2	T	431	 81% 19%
2	V	431	 78% 22%
2	Y	431	 83% 17%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 67152 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 Glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	C	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	E	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	G	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	I	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	K	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	M	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	O	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	Q	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	S	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	U	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		
1	X	301	Total	C	N	O	S	0	0
			2324	1455	400	445	24		

- Molecule 2 is a protein called Glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	D	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	F	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					AltConf	Trace
2	H	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	J	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	L	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	N	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	P	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	R	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	T	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	V	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		
2	Y	431	Total	C	N	O	S	0	0
			3272	2030	561	652	29		

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	688	ASP	-	expression tag	UNP A2T072
B	689	PRO	-	expression tag	UNP A2T072
B	690	GLY	-	expression tag	UNP A2T072
D	688	ASP	-	expression tag	UNP A2T072
D	689	PRO	-	expression tag	UNP A2T072
D	690	GLY	-	expression tag	UNP A2T072
F	688	ASP	-	expression tag	UNP A2T072
F	689	PRO	-	expression tag	UNP A2T072
F	690	GLY	-	expression tag	UNP A2T072
H	688	ASP	-	expression tag	UNP A2T072
H	689	PRO	-	expression tag	UNP A2T072
H	690	GLY	-	expression tag	UNP A2T072
J	688	ASP	-	expression tag	UNP A2T072
J	689	PRO	-	expression tag	UNP A2T072
J	690	GLY	-	expression tag	UNP A2T072
L	688	ASP	-	expression tag	UNP A2T072
L	689	PRO	-	expression tag	UNP A2T072
L	690	GLY	-	expression tag	UNP A2T072
N	688	ASP	-	expression tag	UNP A2T072
N	689	PRO	-	expression tag	UNP A2T072
N	690	GLY	-	expression tag	UNP A2T072

*Continued on next page...*


*Continued from previous page...*

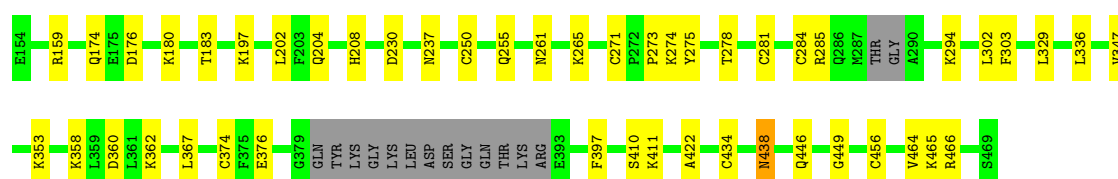
Chain	Residue	Modelled	Actual	Comment	Reference
P	688	ASP	-	expression tag	UNP A2T072
P	689	PRO	-	expression tag	UNP A2T072
P	690	GLY	-	expression tag	UNP A2T072
R	688	ASP	-	expression tag	UNP A2T072
R	689	PRO	-	expression tag	UNP A2T072
R	690	GLY	-	expression tag	UNP A2T072
T	688	ASP	-	expression tag	UNP A2T072
T	689	PRO	-	expression tag	UNP A2T072
T	690	GLY	-	expression tag	UNP A2T072
V	688	ASP	-	expression tag	UNP A2T072
V	689	PRO	-	expression tag	UNP A2T072
V	690	GLY	-	expression tag	UNP A2T072
Y	688	ASP	-	expression tag	UNP A2T072
Y	689	PRO	-	expression tag	UNP A2T072
Y	690	GLY	-	expression tag	UNP A2T072

### 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: Glycoprotein

Chain A: 




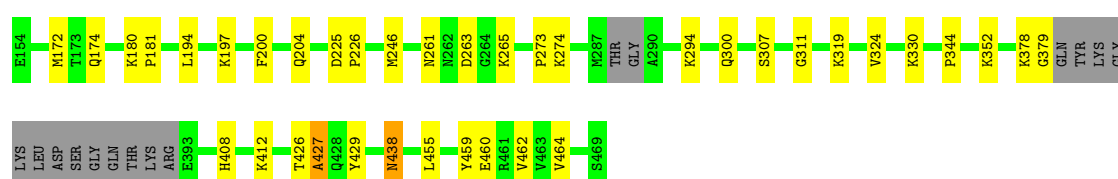
- Molecule 1: Glycoprotein

Chain C: 




- Molecule 1: Glycoprotein

Chain E: 

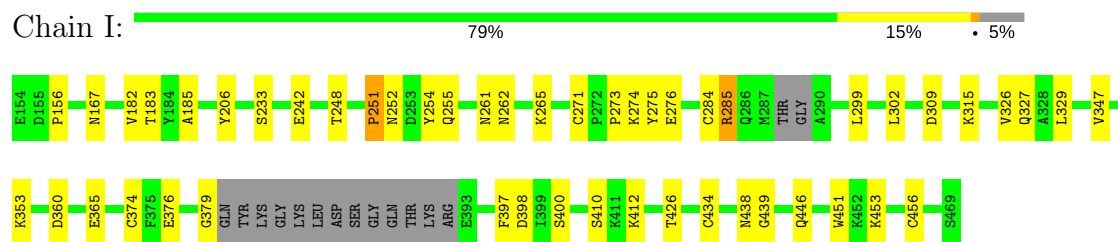


- Molecule 1: Glycoprotein

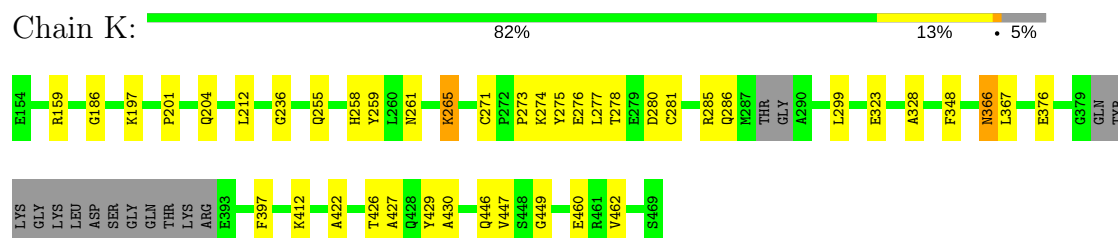
Chain G: 



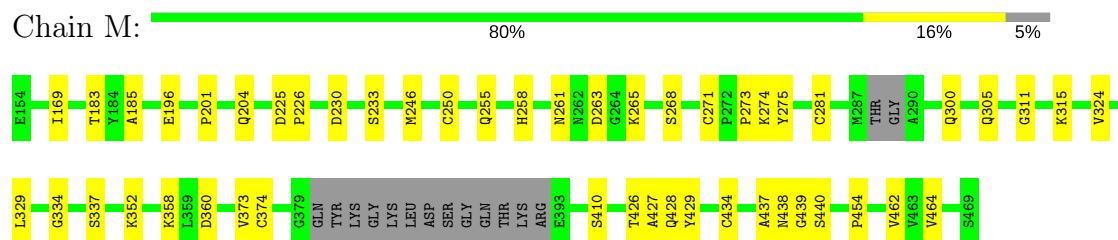
- Molecule 1: Glycoprotein



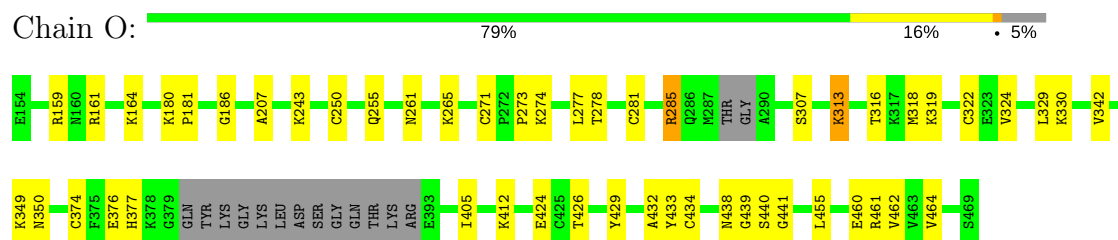
- Molecule 1: Glycoprotein



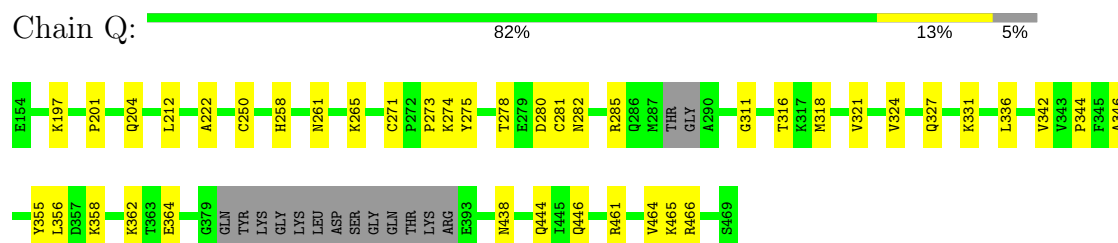
- Molecule 1: Glycoprotein



- Molecule 1: Glycoprotein



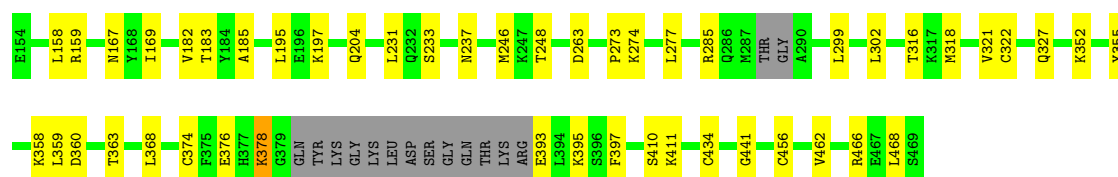
- Molecule 1: Glycoprotein



- Molecule 1: Glycoprotein

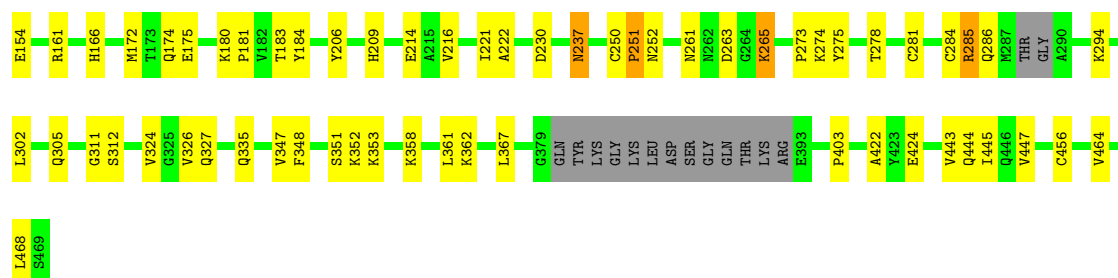






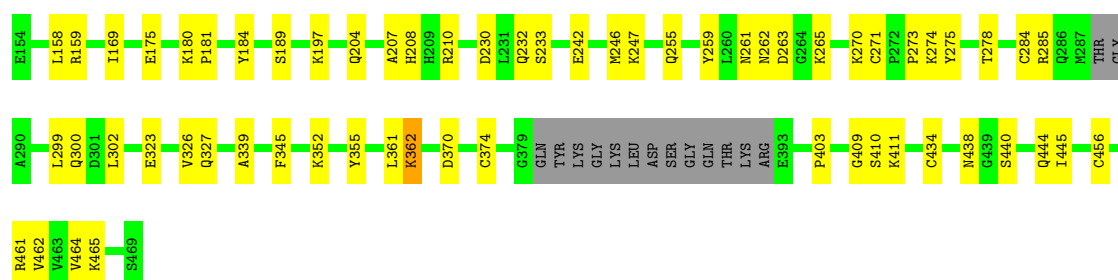
• Molecule 1: Glycoprotein

Chain U: 76% 18% 5%



• Molecule 1: Glycoprotein

Chain X: 76% 19% 5%



• Molecule 2: Glycoprotein

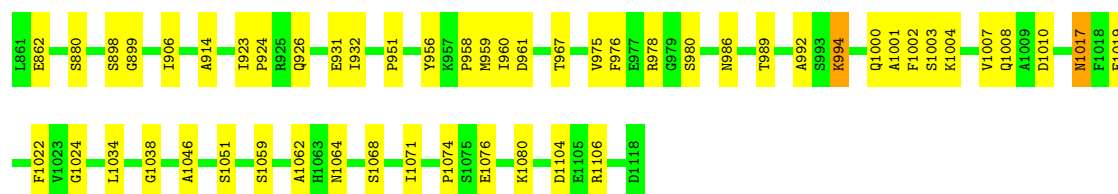
Chain B: 83% 17%



• Molecule 2: Glycoprotein

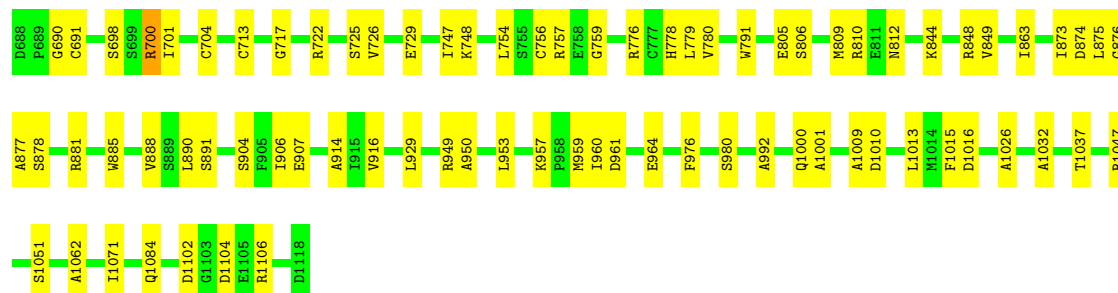
Chain D: 80% 19%





• Molecule 2: Glycoprotein

Chain F: 82% 18%



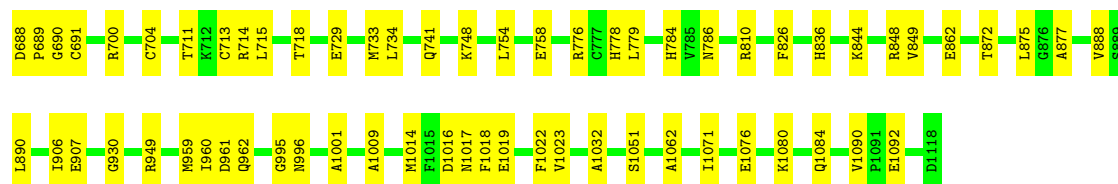
• Molecule 2: Glycoprotein

Chain H: 79% 20%



• Molecule 2: Glycoprotein

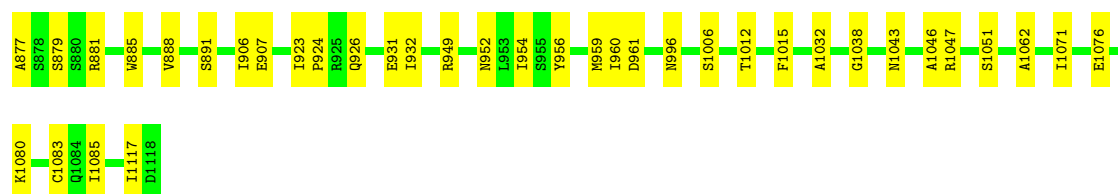
Chain J: 85% 15%



• Molecule 2: Glycoprotein

Chain L: 83% 17%





• Molecule 2: Glycoprotein

Chain N: 82% 17%



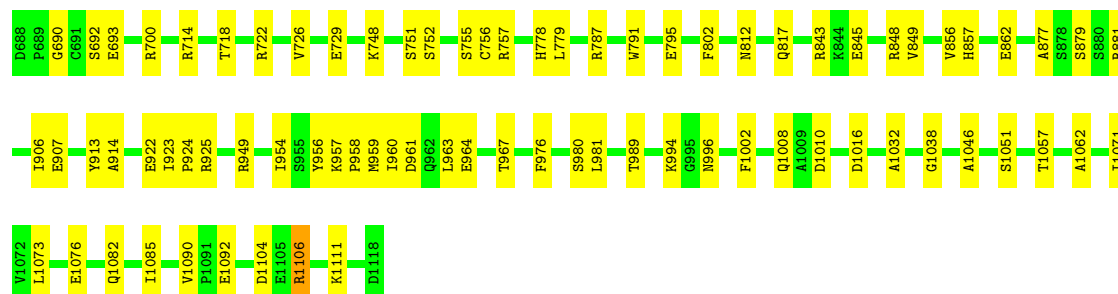
• Molecule 2: Glycoprotein

Chain P: 83% 17%



• Molecule 2: Glycoprotein

Chain R: 82% 18%



• Molecule 2: Glycoprotein

Chain T: 81% 19%





## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of particles used	2995	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TECNAI F30	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	22	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

## 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  >2	RMSZ	# Z  >2
1	A	0.24	0/2373	0.40	0/3193
1	C	0.24	0/2373	0.40	0/3193
1	E	0.24	0/2373	0.40	0/3193
1	G	0.24	0/2373	0.41	0/3193
1	I	0.24	0/2373	0.40	0/3193
1	K	0.24	0/2373	0.41	0/3193
1	M	0.24	0/2373	0.40	0/3193
1	O	0.24	0/2373	0.40	0/3193
1	Q	0.24	0/2373	0.40	0/3193
1	S	0.24	0/2373	0.41	0/3193
1	U	0.24	0/2373	0.41	0/3193
1	X	0.24	0/2373	0.41	0/3193
2	B	0.24	0/3332	0.42	0/4503
2	D	0.24	0/3332	0.42	0/4503
2	F	0.24	0/3332	0.42	0/4503
2	H	0.24	0/3332	0.43	0/4503
2	J	0.24	0/3332	0.43	0/4503
2	L	0.24	0/3332	0.42	0/4503
2	N	0.24	0/3332	0.42	0/4503
2	P	0.24	0/3332	0.42	0/4503
2	R	0.24	0/3332	0.42	0/4503
2	T	0.24	0/3332	0.43	0/4503
2	V	0.24	0/3332	0.42	0/4503
2	Y	0.24	0/3332	0.42	0/4503
All	All	0.24	0/68460	0.42	0/92352

There are no bond length outliers.

There are no bond angle outliers.

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	2324	0	2259	25	0
1	C	2324	0	2259	19	0
1	E	2324	0	2259	22	0
1	G	2324	0	2259	26	0
1	I	2324	0	2259	28	0
1	K	2324	0	2259	23	0
1	M	2324	0	2259	23	0
1	O	2324	0	2259	25	0
1	Q	2324	0	2259	23	0
1	S	2324	0	2259	28	0
1	U	2324	0	2259	31	0
1	X	2324	0	2259	32	0
2	B	3272	0	3143	41	0
2	D	3272	0	3143	44	0
2	F	3272	0	3143	43	0
2	H	3272	0	3143	47	0
2	J	3272	0	3143	33	0
2	L	3272	0	3143	38	0
2	N	3272	0	3143	39	0
2	P	3272	0	3143	37	0
2	R	3272	0	3143	40	0
2	T	3272	0	3143	44	0
2	V	3272	0	3143	50	0
2	Y	3272	0	3143	38	0
All	All	67152	0	64824	766	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 766 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)
2:L:1062:ALA:HB3	2:L:1071:ILE:HB	1.71	0.73
2:J:1062:ALA:HB3	2:J:1071:ILE:HB	1.71	0.71
2:F:691:CYS:HA	2:F:729:GLU:HB2	1.71	0.71

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:N:1062:ALA:HB3	2:N:1071:ILE:HB	1.74	0.69
1:I:262:ASN:HD22	1:I:326:VAL:HG22	1.58	0.69

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	295/316 (93%)	252 (85%)	42 (14%)	1 (0%)	43	81
1	C	295/316 (93%)	262 (89%)	30 (10%)	3 (1%)	17	60
1	E	295/316 (93%)	253 (86%)	40 (14%)	2 (1%)	24	67
1	G	295/316 (93%)	259 (88%)	33 (11%)	3 (1%)	17	60
1	I	295/316 (93%)	258 (88%)	33 (11%)	4 (1%)	12	52
1	K	295/316 (93%)	251 (85%)	41 (14%)	3 (1%)	17	60
1	M	295/316 (93%)	245 (83%)	43 (15%)	7 (2%)	6	40
1	O	295/316 (93%)	262 (89%)	29 (10%)	4 (1%)	12	52
1	Q	295/316 (93%)	249 (84%)	44 (15%)	2 (1%)	24	67
1	S	295/316 (93%)	261 (88%)	32 (11%)	2 (1%)	24	67
1	U	295/316 (93%)	250 (85%)	40 (14%)	5 (2%)	10	49
1	X	295/316 (93%)	256 (87%)	39 (13%)	0	100	100
2	B	429/431 (100%)	377 (88%)	48 (11%)	4 (1%)	19	61
2	D	429/431 (100%)	385 (90%)	40 (9%)	4 (1%)	19	61
2	F	429/431 (100%)	375 (87%)	50 (12%)	4 (1%)	19	61
2	H	429/431 (100%)	378 (88%)	47 (11%)	4 (1%)	19	61
2	J	429/431 (100%)	382 (89%)	43 (10%)	4 (1%)	19	61
2	L	429/431 (100%)	383 (89%)	43 (10%)	3 (1%)	24	67

*Continued on next page...*



Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	N	429/431 (100%)	383 (89%)	39 (9%)	7 (2%)	11	50
2	P	429/431 (100%)	385 (90%)	40 (9%)	4 (1%)	19	61
2	R	429/431 (100%)	378 (88%)	48 (11%)	3 (1%)	24	67
2	T	429/431 (100%)	380 (89%)	44 (10%)	5 (1%)	14	56
2	V	429/431 (100%)	368 (86%)	55 (13%)	6 (1%)	12	52
2	Y	429/431 (100%)	375 (87%)	48 (11%)	6 (1%)	12	52
All	All	8688/8964 (97%)	7607 (88%)	991 (11%)	90 (1%)	22	60

5 of 90 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	H	709	VAL
2	V	709	VAL
2	F	877	ALA
2	H	877	ALA
1	I	251	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 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	259/271 (96%)	258 (100%)	1 (0%)	92	95
1	C	259/271 (96%)	257 (99%)	2 (1%)	83	92
1	E	259/271 (96%)	256 (99%)	3 (1%)	74	87
1	G	259/271 (96%)	258 (100%)	1 (0%)	92	95
1	I	259/271 (96%)	257 (99%)	2 (1%)	83	92
1	K	259/271 (96%)	258 (100%)	1 (0%)	92	95
1	M	259/271 (96%)	257 (99%)	2 (1%)	83	92
1	O	259/271 (96%)	255 (98%)	4 (2%)	67	85
1	Q	259/271 (96%)	257 (99%)	2 (1%)	83	92
1	S	259/271 (96%)	258 (100%)	1 (0%)	92	95

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	U	259/271 (96%)	257 (99%)	2 (1%)	83	92
1	X	259/271 (96%)	257 (99%)	2 (1%)	83	92
2	B	371/371 (100%)	367 (99%)	4 (1%)	76	88
2	D	371/371 (100%)	367 (99%)	4 (1%)	76	88
2	F	371/371 (100%)	368 (99%)	3 (1%)	83	92
2	H	371/371 (100%)	368 (99%)	3 (1%)	83	92
2	J	371/371 (100%)	367 (99%)	4 (1%)	76	88
2	L	371/371 (100%)	367 (99%)	4 (1%)	76	88
2	N	371/371 (100%)	369 (100%)	2 (0%)	90	95
2	P	371/371 (100%)	368 (99%)	3 (1%)	83	92
2	R	371/371 (100%)	363 (98%)	8 (2%)	55	77
2	T	371/371 (100%)	368 (99%)	3 (1%)	83	92
2	V	371/371 (100%)	362 (98%)	9 (2%)	52	75
2	Y	371/371 (100%)	369 (100%)	2 (0%)	90	95
All	All	7560/7704 (98%)	7488 (99%)	72 (1%)	80	89

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	M	315	LYS
2	P	700	ARG
2	V	1047	ARG
1	M	438	ASN
1	O	285	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 78 such sidechains are listed below:

Mol	Chain	Res	Type
1	K	366	ASN
1	M	157	HIS
1	X	232	GLN
1	K	377	HIS
2	L	996	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 ⓘ

There are no carbohydrates in this entry.

### 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

### 5.7 Other polymers ⓘ

There are no such residues in this entry.

### 5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.