



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Dec 28, 2017 – 08:30 AM EST

PDB ID : 5VU2
EMDB ID: : EMD-8734
Title : Electron cryo-microscopy of "immature" Chikungunya VLP
Authors : Rossmann, M.G.; Yap, M.L.
Deposited on : 2017-05-18
Resolution : 6.80 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

A user guide is available at

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

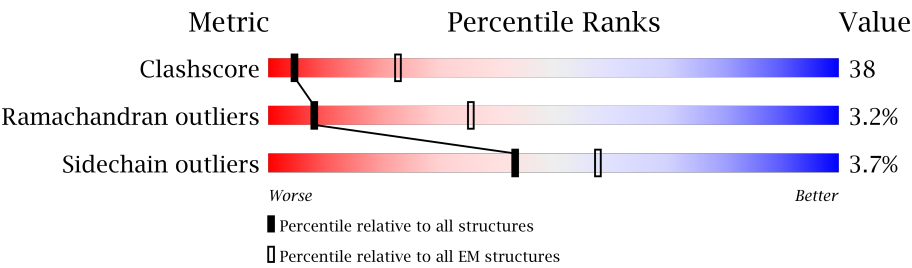
MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030736

1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 6.80 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	866	<div><div>37%</div><div>•</div><div>61%</div></div>
1	B	866	<div><div>36%</div><div>•</div><div>61%</div></div>
1	C	866	<div><div>36%</div><div>•</div><div>61%</div></div>
1	D	866	<div><div>36%</div><div>•</div><div>61%</div></div>
1	M	866	<div><div>41%</div><div>••</div><div>56%</div></div>
1	N	866	<div><div>41%</div><div>•</div><div>56%</div></div>
1	O	866	<div><div>39%</div><div>5%</div><div>56%</div></div>
1	P	866	<div><div>40%</div><div>•</div><div>56%</div></div>
2	U	60	<div><div>97%</div><div>•</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	V	60	 97% .
2	W	60	 97% .
2	X	60	 97% .
3	I	149	 64% 30% . .
3	J	149	 64% 32% . .
3	K	149	 62% 32% . .
3	L	149	 62% 32% . .
4	E	46	 33% 50% 13% .
4	F	46	 33% 50% 13% .
4	G	46	 30% 50% 15% .
4	H	46	 30% 50% 15% .
5	Q	81	 17% 48% 31% .
5	R	81	 15% 53% 26% 6% .
5	S	81	 21% 49% 26% .
5	T	81	 17% 54% 25% .

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 32676 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 E1 envelope glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	M	382	Total	C	N	O	S	1	0
			2911	1839	487	561	24		
1	A	341	Total	C	N	O	S	0	0
			2691	1677	487	507	20		
1	B	341	Total	C	N	O	S	0	0
			2691	1677	487	507	20		
1	N	382	Total	C	N	O	S	1	0
			2911	1839	487	561	24		
1	C	341	Total	C	N	O	S	0	0
			2691	1677	487	507	20		
1	O	382	Total	C	N	O	S	1	0
			2911	1839	487	561	24		
1	D	341	Total	C	N	O	S	0	0
			2691	1677	487	507	20		
1	P	382	Total	C	N	O	S	1	0
			2911	1839	487	561	24		

- Molecule 2 is a protein called E3 envelope glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	U	60	Total	C	N	O	S	0	0
			477	296	81	91	9		
2	V	60	Total	C	N	O	S	0	0
			477	296	81	91	9		
2	W	60	Total	C	N	O	S	0	0
			477	296	81	91	9		
2	X	60	Total	C	N	O	S	0	0
			477	296	81	91	9		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
U	64	GLU	-	expression tag	UNP C7S7A1
V	64	GLU	-	expression tag	UNP C7S7A1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
W	64	GLU	-	expression tag	UNP C7S7A1
X	64	GLU	-	expression tag	UNP C7S7A1

- Molecule 3 is a protein called capsid protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	I	149	Total	C	N	O	S	0	0
			1141	723	201	211	6		
3	J	149	Total	C	N	O	S	0	0
			1141	723	201	211	6		
3	K	149	Total	C	N	O	S	0	0
			1141	723	201	211	6		
3	L	149	Total	C	N	O	S	0	0
			1141	723	201	211	6		

- Molecule 4 is a protein called E1 envelope glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	46	Total	C	N	O	S	0	0
			336	218	57	59	2		
4	F	46	Total	C	N	O	S	0	0
			336	218	57	59	2		
4	G	46	Total	C	N	O	S	0	0
			336	218	57	59	2		
4	H	46	Total	C	N	O	S	0	0
			336	218	57	59	2		

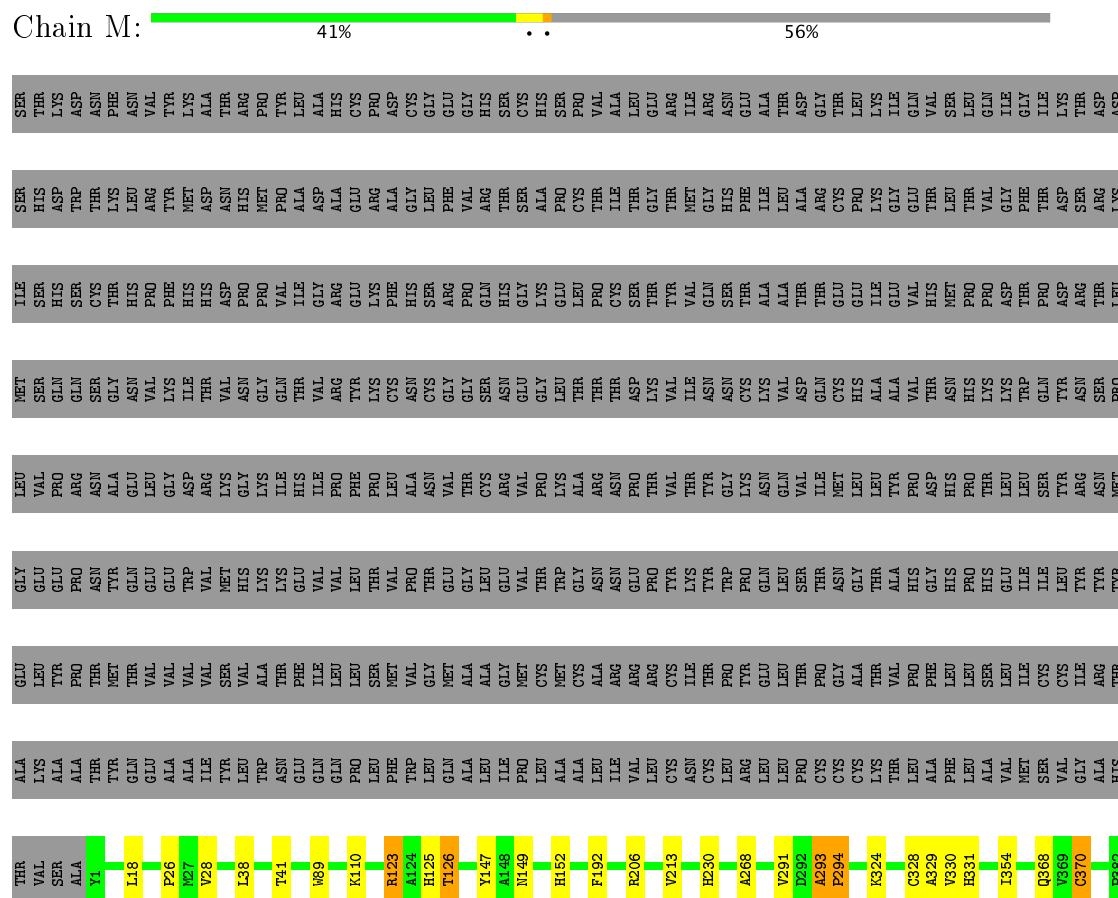
- Molecule 5 is a protein called E2 envelope glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	Q	81	Total	C	N	O	S	0	0
			613	396	101	108	8		
5	R	81	Total	C	N	O	S	0	0
			613	396	101	108	8		
5	S	81	Total	C	N	O	S	0	0
			613	396	101	108	8		
5	T	81	Total	C	N	O	S	0	0
			613	396	101	108	8		

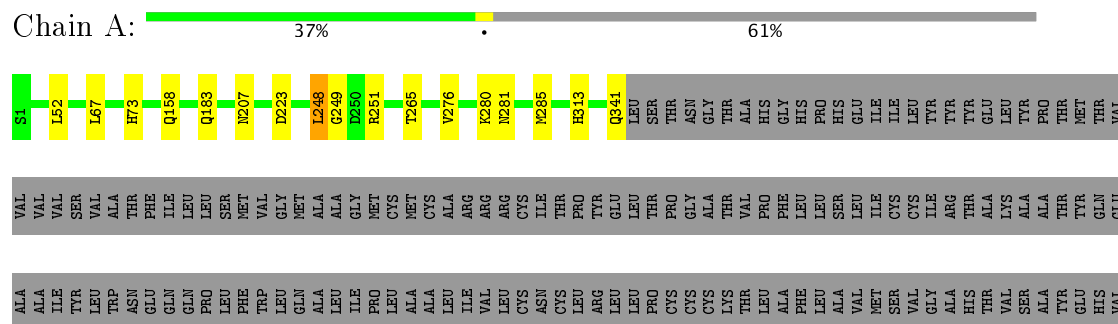
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: E1 envelope glycoprotein



- Molecule 1: E1 envelope glycoprotein





- Chain N:  41% . 56%

- Molecule 1: E1 envelope glycoprotein

Chain C: 36% . 61%

[illegible]

ASN	GLY	THR	ALA	HIS	GLY	HIS	HIS	PRO	PRO	HIS	GLU	HIS	ILE	ILE	LEU	TYR	TYR	TYR	GLU	LEU	TYR	PRO	PRO	THR	THR	VAL	VAL	VAL	VAL	SER	VAL	ALA	ALA	ALA	ALA	GLY	CYS	CYS	ALA	ARG	ARG	ARG	CYS	ILE	THR	PRO	TYR	GLU	LEU	THR	PRO
S1	H48	G19	P20	E24	G25	H26	S27	T42	L52	L67	H73	E109	T110	I136	H142	S143	G144	P145	Q146	Q158	Q183	N207	D223	L248	G249	D250	R251	T265	R272	T275	V276	K280	M285	H313	Q341	SER	THR														

[illegible]

- Molecule 1: E1 envelope glycoprotein

Chain P:  40% . 56%

[illegible]

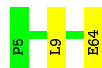
- Molecule 2: E3 envelope glycoprotein

Chain U:  97%



- Molecule 2: E3 envelope glycoprotein

Chain V:  97%



- Molecule 2: E3 envelope glycoprotein

Chain W:  97%



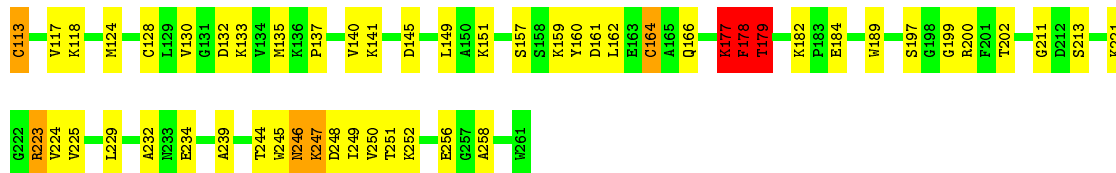
- Molecule 2: E3 envelope glycoprotein

Chain X:  97%



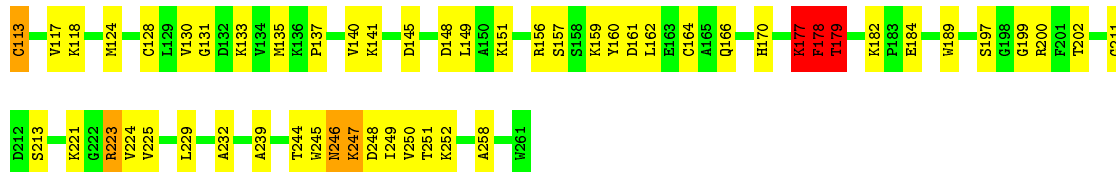
- Molecule 3: capsid protein

Chain I:  64% 30%



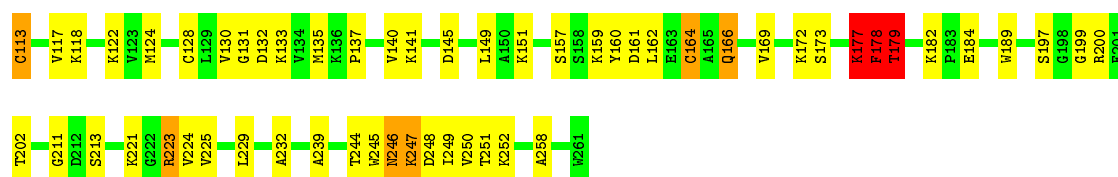
- Molecule 3: capsid protein

Chain J:  64% 32%



- Molecule 3: capsid protein

Chain K:  62% 32%



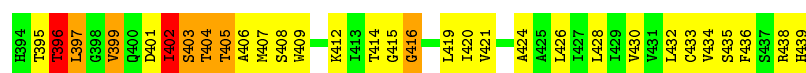
• Molecule 3: capsid protein

Chain L: 62% 32% . .



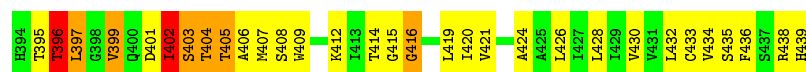
• Molecule 4: E1 envelope glycoprotein

Chain E: 33% 50% 13% .



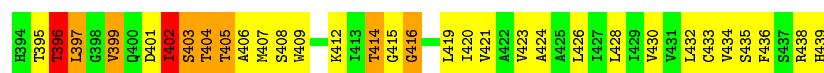
• Molecule 4: E1 envelope glycoprotein

Chain F: 33% 50% 13% .



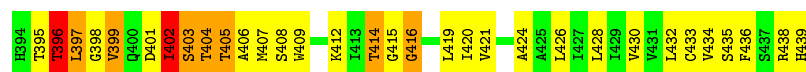
• Molecule 4: E1 envelope glycoprotein

Chain G: 30% 50% 15% .



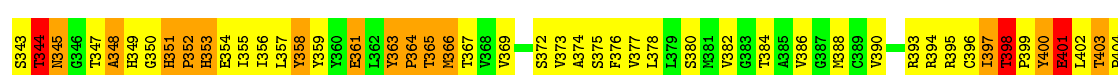
• Molecule 4: E1 envelope glycoprotein

Chain H: 30% 50% 15% .



• Molecule 5: E2 envelope glycoprotein

Chain Q: 17% 48% 31% .



G405
A406
T407
V408
P409
F410
L411
L412
S413
L414
L415
C416
C417
V418
R419
T420
T421
K422
A423

• Molecule 5: E2 envelope glycoprotein

Chain R: 15% 53% 26% 6%

S343
T344
R345
A348
R349
G350
R351
P352
R353
E354
I355
I356
L357
Y358
Y359
Y360
E361
L362
Y363
P364
T365
M366
T367
Y368
Y369
I370
V371
S372
V373
A374
F375
S376
S377
F378
L379
S380
M381
V382
G383
T384
A385
V386
G387
M388
G389
V390
R393
R394
R395
C396
I397
T398
P399
Y400
E401
L402
T403
P404
G405

G405
A406
T407
V408
P409
F410
L411
L412
S413
L414
L415
C416
C417
V418
R419
T420
T421
K422
A423

• Molecule 5: E2 envelope glycoprotein

Chain S: 21% 49% 26% .

S343
T344
R345
G346
T347
A348
R349
G350
R351
P352
R353
E354
I355
I356
L357
Y358
Y359
L362
Y363
P364
T365
M366
V369
S372
V373
A374
S375
F376
V377
L378
L379
S380
M381
V382
G383
T384
A385
V386
G387
M388
G389
V390
R393
R394
R395
C396
I397
T398
P399
Y400
E401
L402
T403
P404
G405

T407
V408
P409
F410
L411
L412
S413
L414
L415
C416
C417
V418
R419
T420
T421
K422
A423

• Molecule 5: E2 envelope glycoprotein

Chain T: 17% 54% 25% .

S343
T344
R345
A348
R349
G350
R351
P352
R353
E354
I355
I356
L357
Y358
Y359
Y360
E361
L362
Y363
P364
T365
M366
V369
I370
V371
S372
V373
A374
S375
F376
V377
L378
L379
S380
M381
V382
G383
T384
A385
V386
G387
M388
G389
V390
R393
R394
R395
C396
I397
T398
P399
Y400
E401
L402
T403
P404
G405

A406
T407
V408
P409
F410
L411
L412
S413
L414
L415
C416
C417
V418
R419
T420
T421
K422
A423

4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	72944	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	36	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	0.35	0/2763	0.58	0/3760
1	B	0.35	0/2763	0.58	0/3760
1	C	0.35	0/2763	0.58	0/3760
1	D	0.35	0/2763	0.59	0/3760
1	M	0.33	0/2984	0.57	0/4069
1	N	0.33	0/2984	0.57	0/4069
1	O	0.33	0/2984	0.57	0/4069
1	P	0.33	0/2984	0.57	0/4069
2	U	0.38	0/489	0.59	0/665
2	V	0.38	0/489	0.59	0/665
2	W	0.38	0/489	0.59	0/665
2	X	0.38	0/489	0.60	0/665
3	I	0.81	3/1169 (0.3%)	1.75	8/1577 (0.5%)
3	J	0.81	3/1169 (0.3%)	1.74	8/1577 (0.5%)
3	K	0.81	3/1169 (0.3%)	1.75	8/1577 (0.5%)
3	L	0.81	3/1169 (0.3%)	1.74	8/1577 (0.5%)
4	E	0.76	0/340	1.10	5/464 (1.1%)
4	F	0.76	0/340	1.10	5/464 (1.1%)
4	G	0.76	0/340	1.10	5/464 (1.1%)
4	H	0.76	0/340	1.10	5/464 (1.1%)
5	Q	0.34	0/627	0.49	0/858
5	R	0.33	0/627	0.48	0/858
5	S	0.33	0/627	0.48	0/858
5	T	0.32	0/627	0.46	0/858
All	All	0.46	12/33488 (0.0%)	0.86	52/45572 (0.1%)

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
3	I	0	2

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
3	J	0	2
3	K	0	2
3	L	0	2
4	E	0	1
4	F	0	1
4	G	0	1
4	H	0	1
All	All	0	12

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	J	113	CYS	CB-SG	-14.10	1.58	1.82
3	K	113	CYS	CB-SG	-14.09	1.58	1.82
3	L	113	CYS	CB-SG	-14.03	1.58	1.82
3	I	113	CYS	CB-SG	-13.99	1.58	1.82
3	J	128	CYS	CB-SG	-12.10	1.61	1.82

The worst 5 of 52 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	I	177	LYS	O-C-N	-38.39	61.27	122.70
3	K	177	LYS	O-C-N	-38.37	61.31	122.70
3	L	177	LYS	O-C-N	-38.34	61.35	122.70
3	J	177	LYS	O-C-N	-38.34	61.35	122.70
3	J	178	PHE	O-C-N	-26.35	80.54	122.70

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	I	177	LYS	Mainchain
3	I	178	PHE	Mainchain
3	J	177	LYS	Mainchain
3	J	178	PHE	Mainchain
3	K	177	LYS	Mainchain

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	2691	0	2608	17	0
1	B	2691	0	2603	59	0
1	C	2691	0	2605	49	0
1	D	2691	0	2603	86	0
1	M	2911	0	2819	105	0
1	N	2911	0	2819	70	0
1	O	2911	0	2819	138	0
1	P	2911	0	2819	71	0
2	U	477	0	457	0	0
2	V	477	0	457	0	0
2	W	477	0	457	0	0
2	X	477	0	457	0	0
3	I	1141	0	1125	173	0
3	J	1141	0	1119	147	0
3	K	1141	0	1124	225	0
3	L	1141	0	1123	173	0
4	E	336	0	361	106	0
4	F	336	0	361	121	0
4	G	336	0	361	137	0
4	H	336	0	361	109	0
5	Q	613	0	630	348	0
5	R	613	0	631	368	0
5	S	613	0	629	404	0
5	T	613	0	631	357	0
All	All	32676	0	31979	2466	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 38.

The worst 5 of 2466 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)
3:J:160:TYR:CE2	5:R:398:THR:HG23	1.18	1.67
3:K:160:TYR:CZ	5:S:398:THR:HG23	1.24	1.65
3:I:160:TYR:CZ	5:Q:398:THR:HG23	1.27	1.64
4:H:402:ILE:HB	4:H:407:MET:SD	1.34	1.64
3:K:132:ASP:H	5:S:402:LEU:CD1	1.12	1.63

There are no symmetry-related clashes.

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 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	339/866 (39%)	320 (94%)	18 (5%)	1 (0%)	44	81
1	B	339/866 (39%)	320 (94%)	18 (5%)	1 (0%)	44	81
1	C	339/866 (39%)	320 (94%)	18 (5%)	1 (0%)	44	81
1	D	339/866 (39%)	320 (94%)	18 (5%)	1 (0%)	44	81
1	M	381/866 (44%)	370 (97%)	8 (2%)	3 (1%)	22	66
1	N	381/866 (44%)	372 (98%)	8 (2%)	1 (0%)	44	81
1	O	381/866 (44%)	370 (97%)	10 (3%)	1 (0%)	44	81
1	P	381/866 (44%)	370 (97%)	10 (3%)	1 (0%)	44	81
2	U	58/60 (97%)	57 (98%)	1 (2%)	0	100	100
2	V	58/60 (97%)	57 (98%)	1 (2%)	0	100	100
2	W	58/60 (97%)	57 (98%)	1 (2%)	0	100	100
2	X	58/60 (97%)	57 (98%)	1 (2%)	0	100	100
3	I	147/149 (99%)	131 (89%)	13 (9%)	3 (2%)	9	46
3	J	147/149 (99%)	131 (89%)	13 (9%)	3 (2%)	9	46
3	K	147/149 (99%)	131 (89%)	13 (9%)	3 (2%)	9	46
3	L	147/149 (99%)	131 (89%)	13 (9%)	3 (2%)	9	46
4	E	44/46 (96%)	33 (75%)	5 (11%)	6 (14%)	0	6
4	F	44/46 (96%)	33 (75%)	5 (11%)	6 (14%)	0	6
4	G	44/46 (96%)	33 (75%)	5 (11%)	6 (14%)	0	6
4	H	44/46 (96%)	33 (75%)	5 (11%)	6 (14%)	0	6
5	Q	79/81 (98%)	40 (51%)	15 (19%)	24 (30%)	0	0
5	R	79/81 (98%)	40 (51%)	15 (19%)	24 (30%)	0	0
5	S	79/81 (98%)	41 (52%)	17 (22%)	21 (27%)	0	1
5	T	79/81 (98%)	41 (52%)	20 (25%)	18 (23%)	0	2
All	All	4192/8272 (51%)	3808 (91%)	251 (6%)	133 (3%)	8	36

5 of 133 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	M	126	THR
1	N	126	THR
1	O	126	THR
1	P	126	THR
3	I	177	LYS

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	303/741 (41%)	294 (97%)	9 (3%)	46	72
1	B	303/741 (41%)	294 (97%)	9 (3%)	46	72
1	C	303/741 (41%)	294 (97%)	9 (3%)	46	72
1	D	303/741 (41%)	294 (97%)	9 (3%)	46	72
1	M	320/741 (43%)	316 (99%)	4 (1%)	73	87
1	N	320/741 (43%)	316 (99%)	4 (1%)	73	87
1	O	320/741 (43%)	315 (98%)	5 (2%)	68	85
1	P	320/741 (43%)	316 (99%)	4 (1%)	73	87
2	U	57/57 (100%)	55 (96%)	2 (4%)	41	69
2	V	57/57 (100%)	55 (96%)	2 (4%)	41	69
2	W	57/57 (100%)	55 (96%)	2 (4%)	41	69
2	X	57/57 (100%)	55 (96%)	2 (4%)	41	69
3	I	118/118 (100%)	109 (92%)	9 (8%)	15	47
3	J	118/118 (100%)	109 (92%)	9 (8%)	15	47
3	K	118/118 (100%)	109 (92%)	9 (8%)	15	47
3	L	118/118 (100%)	109 (92%)	9 (8%)	15	47
4	E	38/38 (100%)	37 (97%)	1 (3%)	51	75
4	F	38/38 (100%)	37 (97%)	1 (3%)	51	75
4	G	38/38 (100%)	37 (97%)	1 (3%)	51	75

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	H	38/38 (100%)	37 (97%)	1 (3%)	51 75
5	Q	70/70 (100%)	63 (90%)	7 (10%)	9 33
5	R	70/70 (100%)	61 (87%)	9 (13%)	5 25
5	S	70/70 (100%)	63 (90%)	7 (10%)	9 33
5	T	70/70 (100%)	61 (87%)	9 (13%)	5 25
All	All	3624/7060 (51%)	3491 (96%)	133 (4%)	43 68

5 of 133 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	P	370	CYS
3	J	179	THR
5	S	401	GLU
3	I	113	CYS
3	I	223	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	O	138	GLN
1	D	18	HIS
3	K	233	ASN
1	C	146	GLN
3	K	246	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.