



## wwPDB EM Map/Model Validation Report ⓘ

Apr 10, 2016 – 01:49 PM BST

PDB ID : 3J26  
EMDB ID: : EMD-5495  
Title : The 3.5 Å resolution structure of the Sputnik virophage by cryo-EM  
Authors : Zhang, X.Z.  
Deposited on : 2012-09-18  
Resolution : 3.50 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.  
For rigid body fitted models, validation errors reported here could stem from errors in the original structure(s) used in the fitting.  
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/EMValidationReportHelp>

---

MolProbity : 4.02b-467  
Mogul : unknown  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk27241

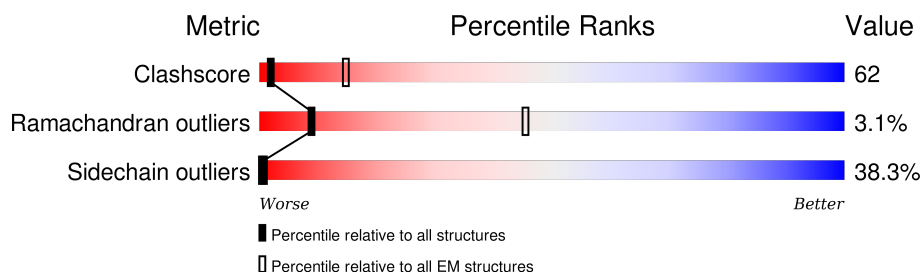
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.50 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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	508	36% 39% 23% .
1	B	508	32% 41% 25% .
1	C	508	30% 43% 27% .
1	D	508	34% 40% 25% .
1	E	508	34% 44% 21% .
1	F	508	36% 41% 22% .
1	G	508	37% 40% 21% .
1	H	508	34% 43% 22% .
1	I	508	32% 40% 27% .

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	J	508	<div><div></div><div>33%</div><div>41%</div><div>24%</div><div></div></div>
1	K	508	<div><div></div><div>37%</div><div>40%</div><div>23%</div><div></div></div>
1	L	508	<div><div></div><div>35%</div><div>41%</div><div>23%</div><div></div></div>
1	M	508	<div><div></div><div>34%</div><div>42%</div><div>23%</div><div></div></div>
2	N	378	<div><div></div><div>26%</div><div>42%</div><div>25%</div><div></div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 54479 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 capsid protein V20.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	B	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	C	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	D	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	E	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	F	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	G	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	H	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	I	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	J	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	K	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	L	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		
1	M	508	Total	C	N	O	S	0	0
			3968	2516	672	769	11		

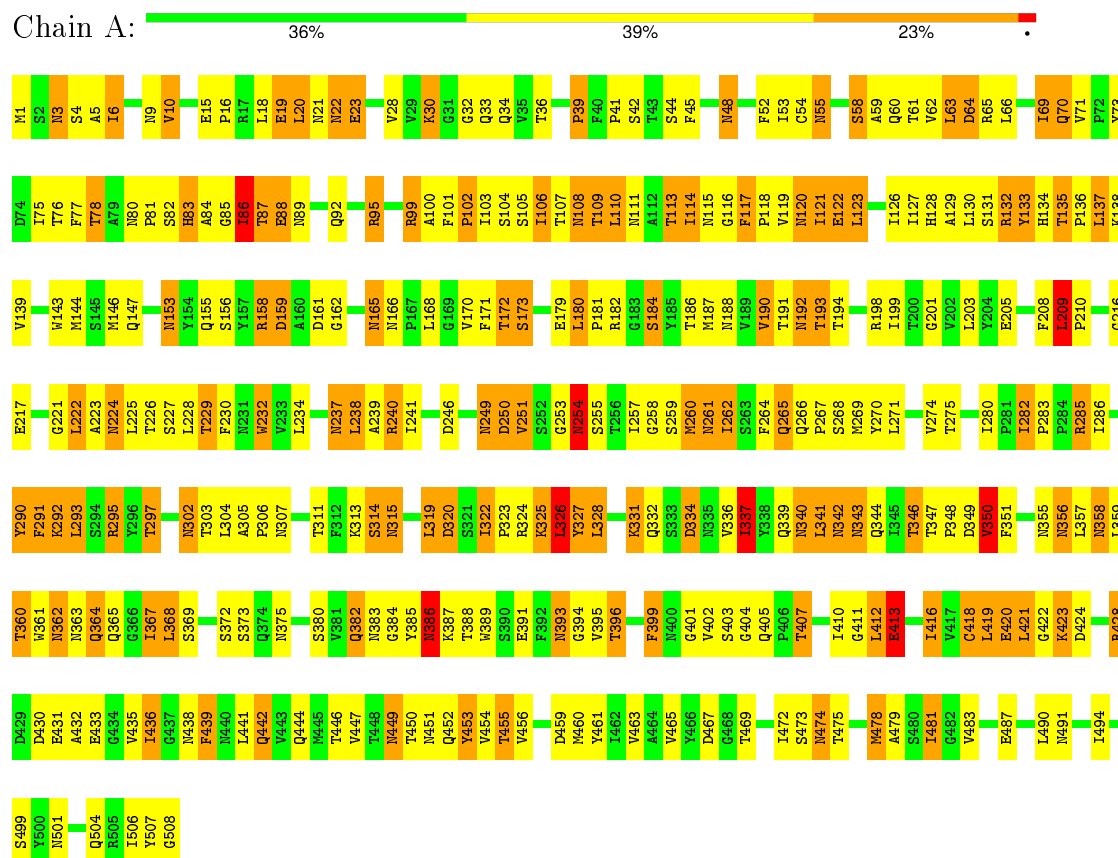
- Molecule 2 is a protein called Minor virion protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	N	362	Total	C	N	O	S	0	0
			2895	1889	451	549	6		

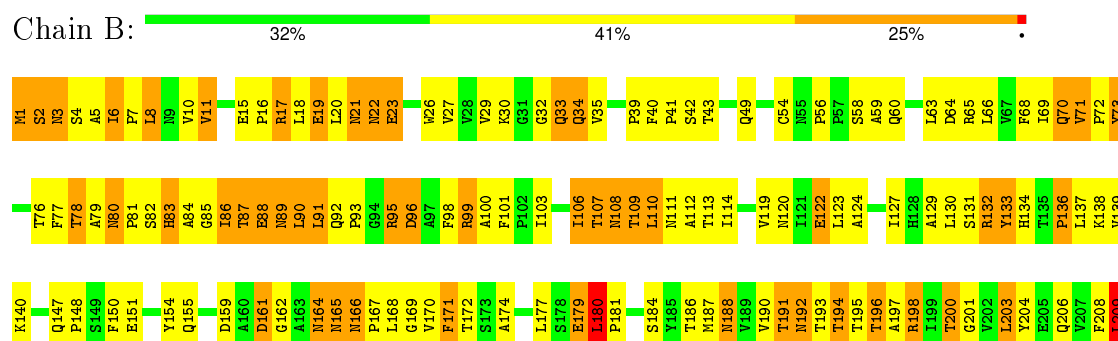
### 3 Residue-property plots

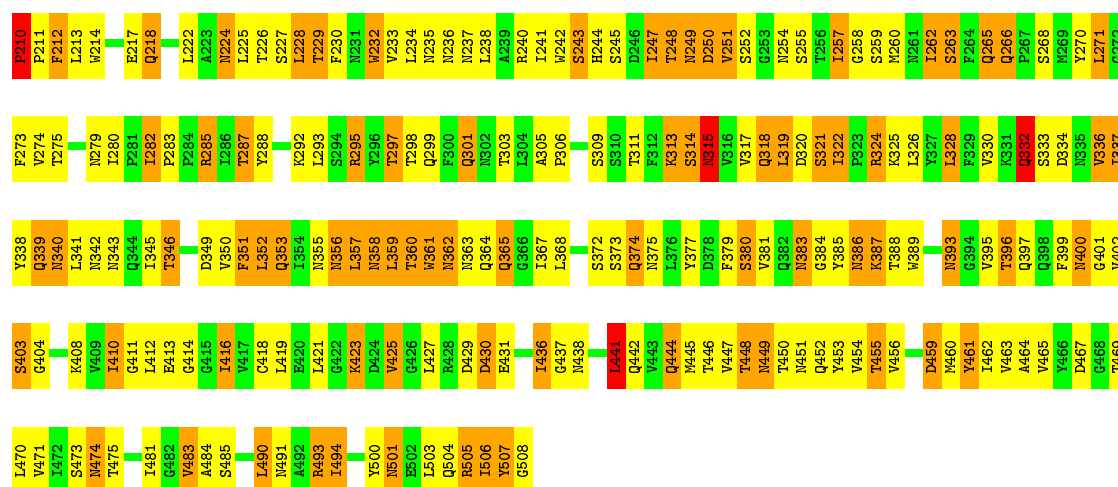
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric 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: capsid protein V20

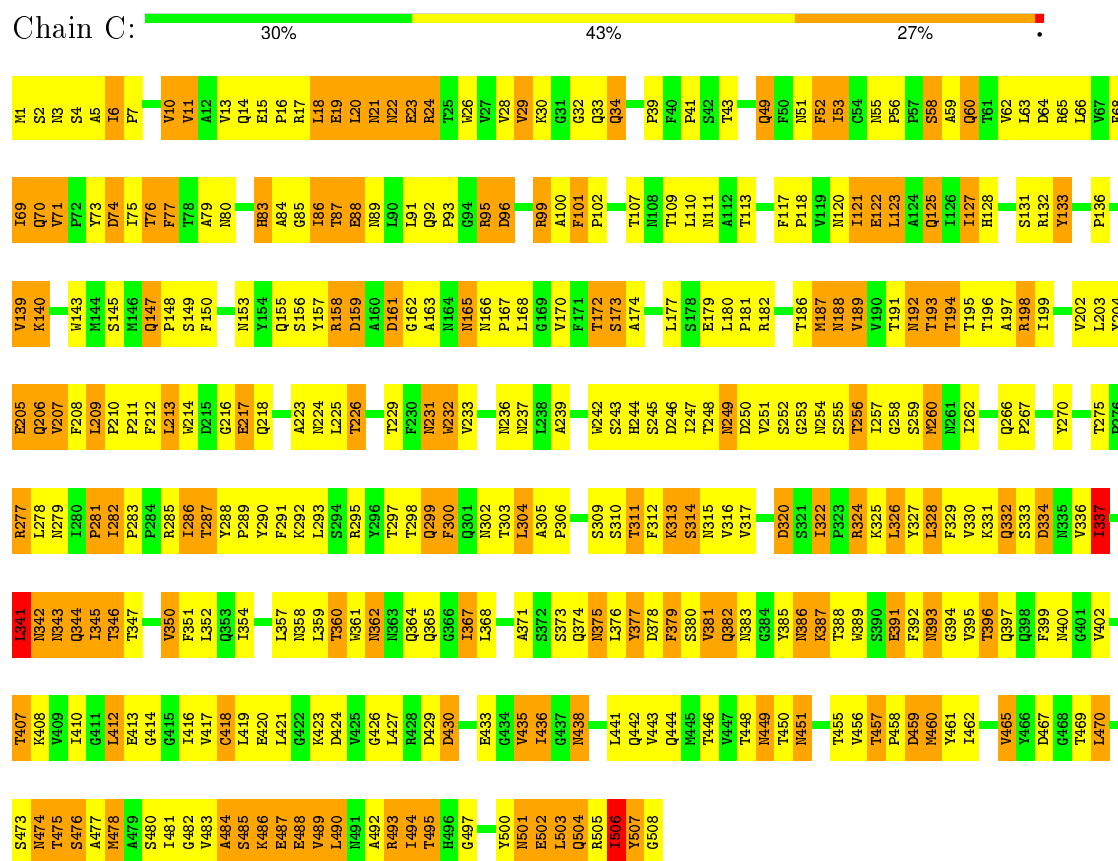


- Molecule 1: capsid protein V20

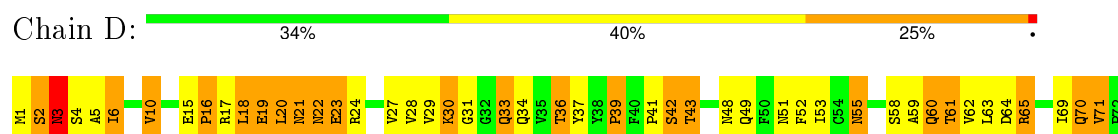


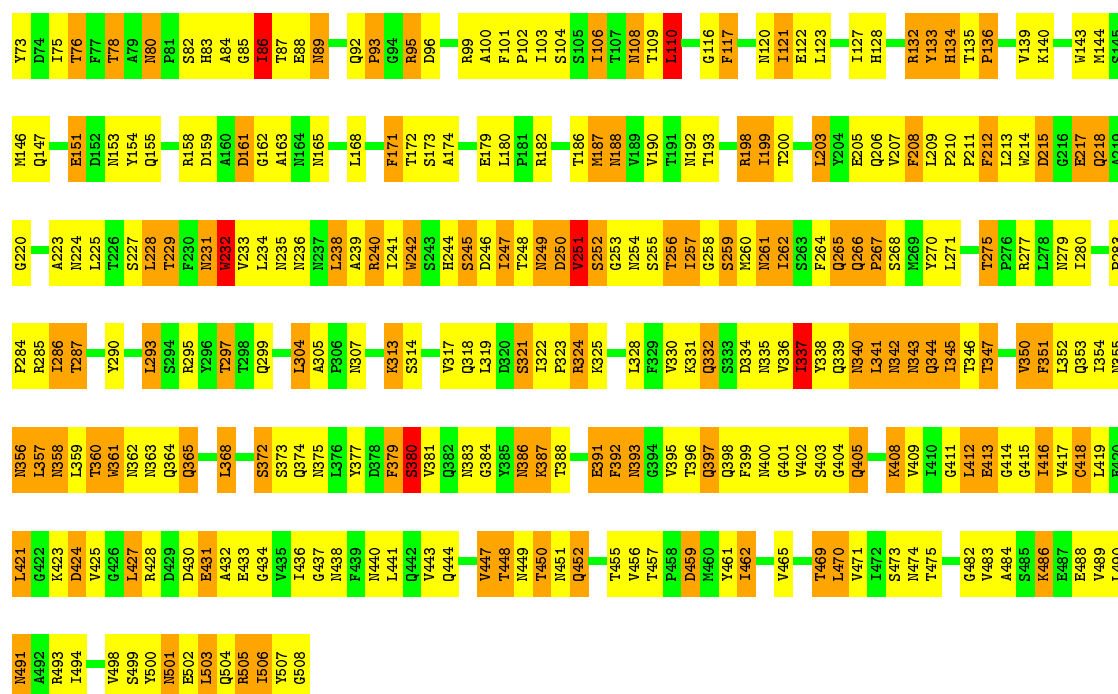


• Molecule 1: capsid protein V20

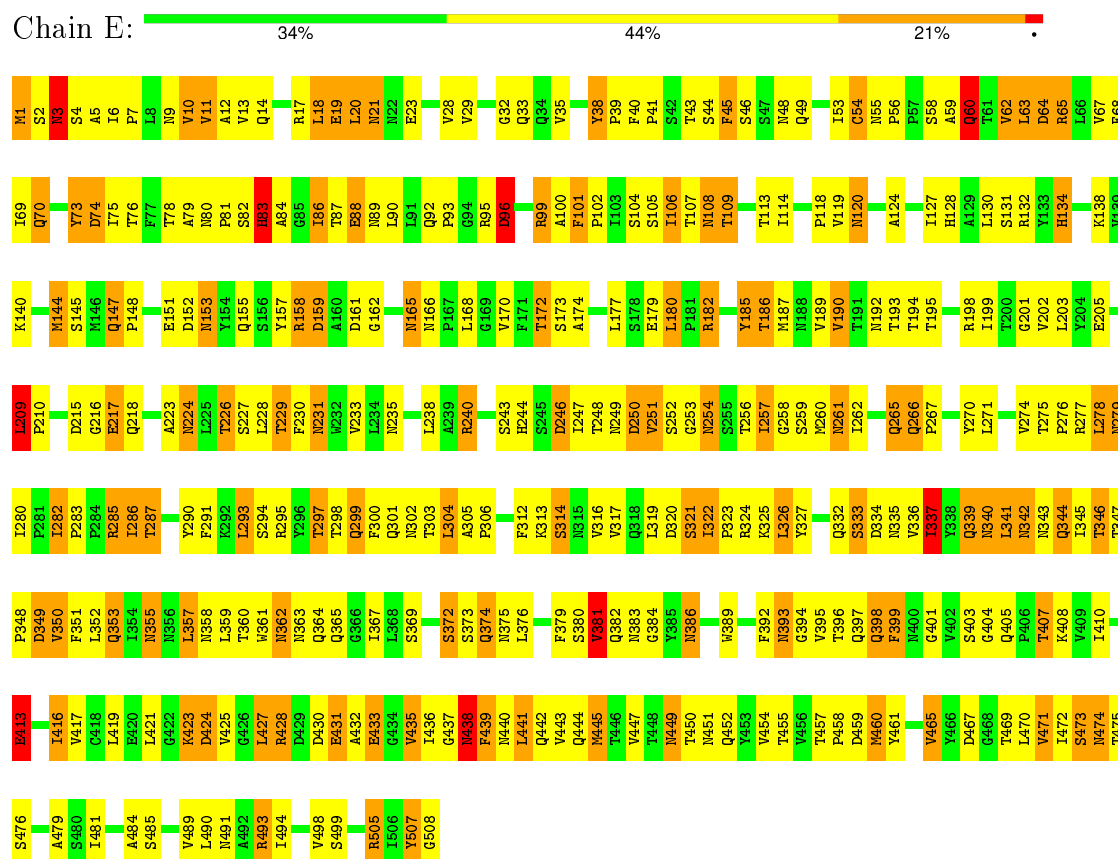


• Molecule 1: capsid protein V20

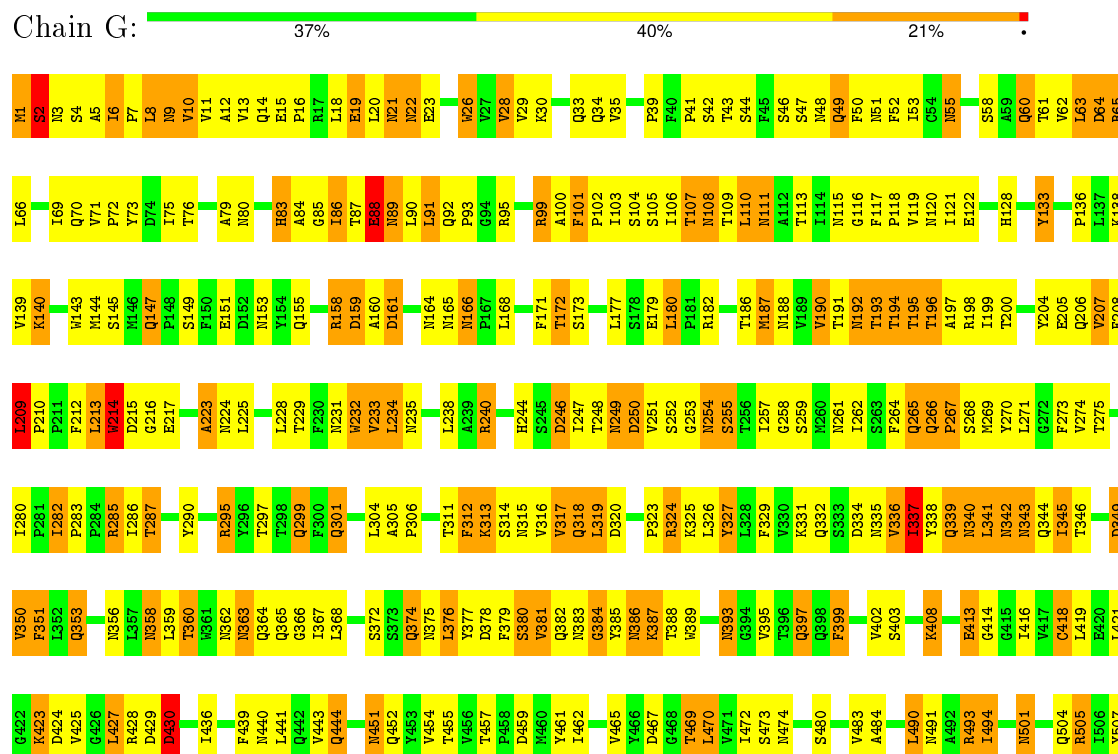




• Molecule 1: capsid protein V20



• Molecule 1: capsid protein V20

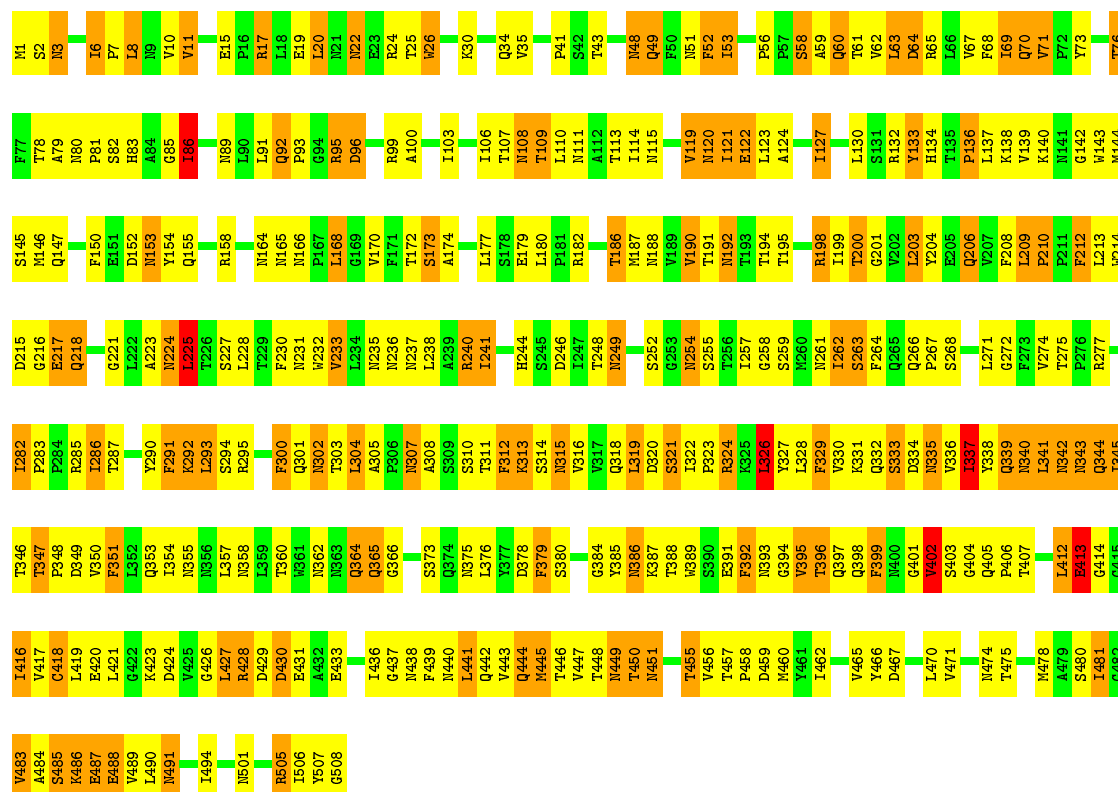




8508

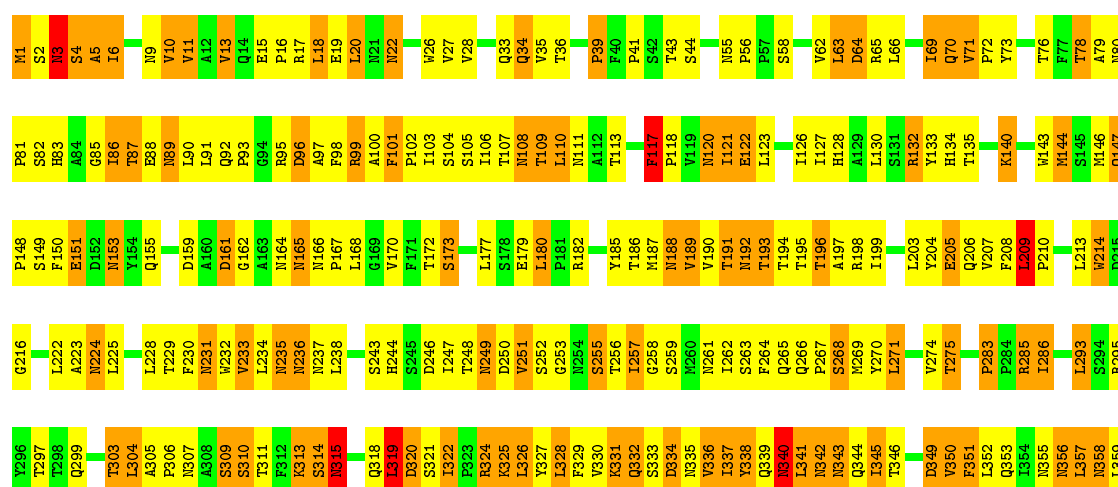
- Molecule 1: capsid protein V20

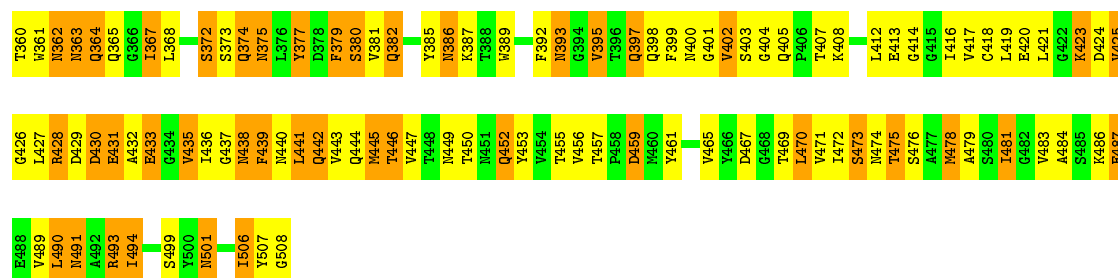
Chain H: 34% 43% 22%



- Molecule 1: capsid protein V20

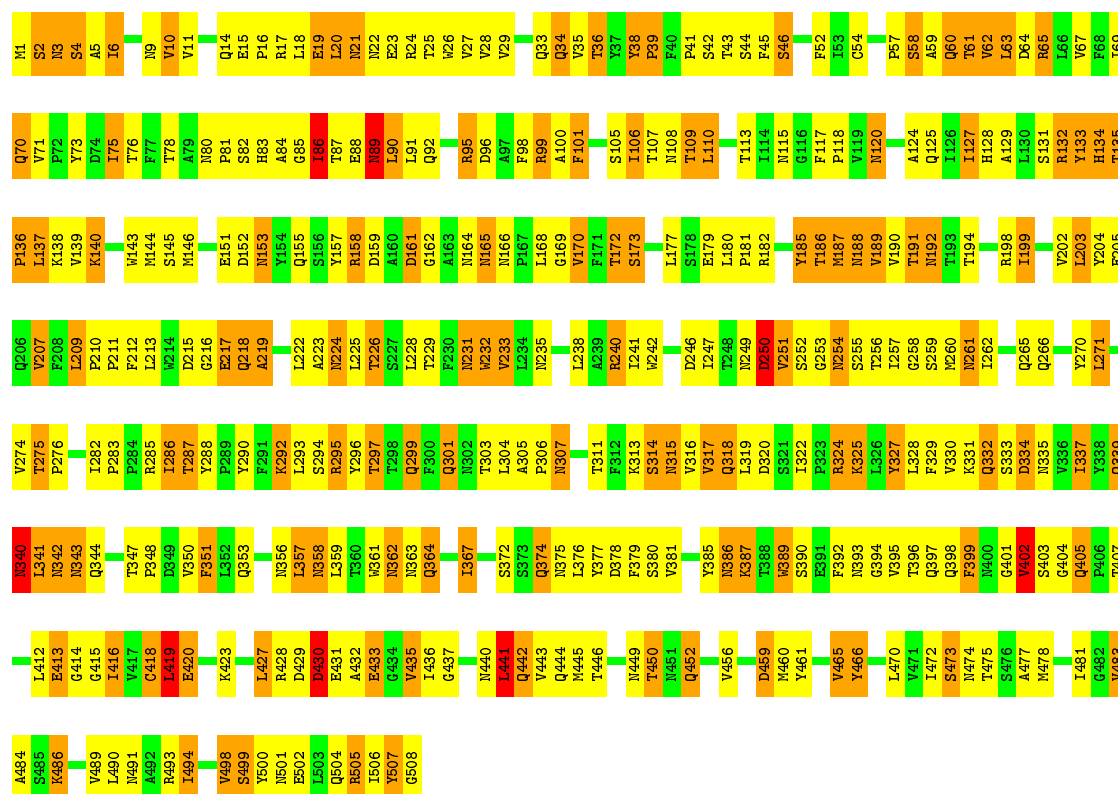
Chain I: 32% 40% 27%





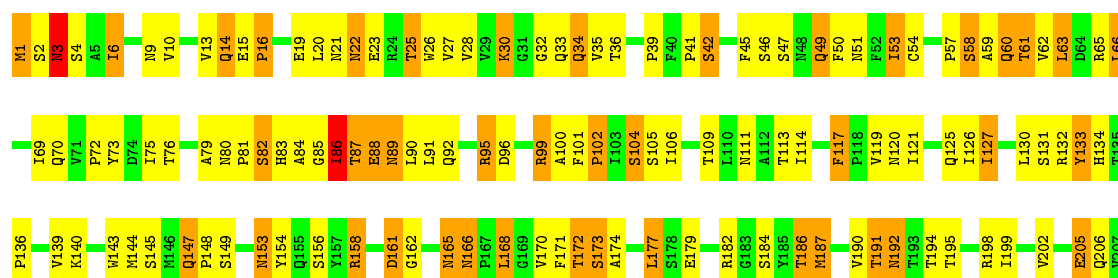
• Molecule 1: capsid protein V20

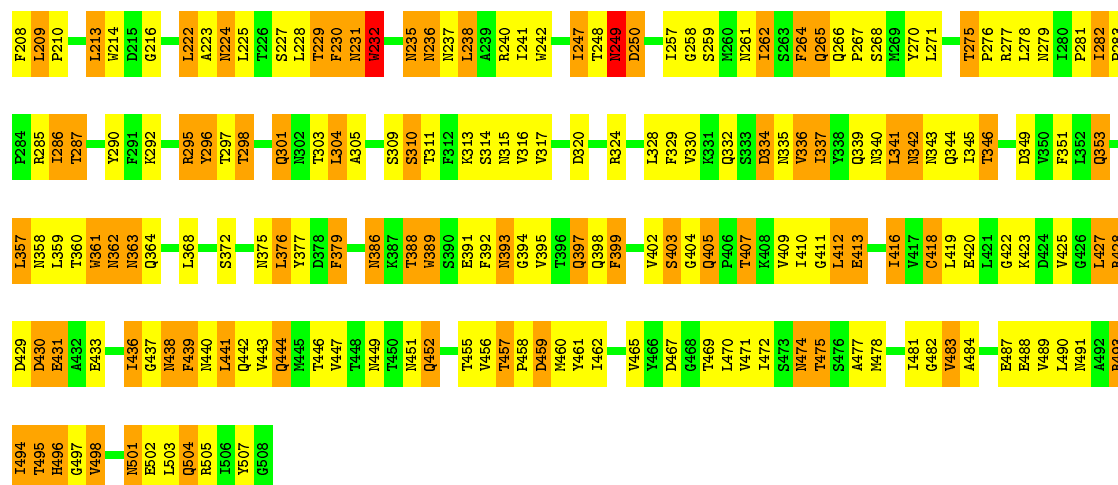
Chain J: 33% 41% 24%



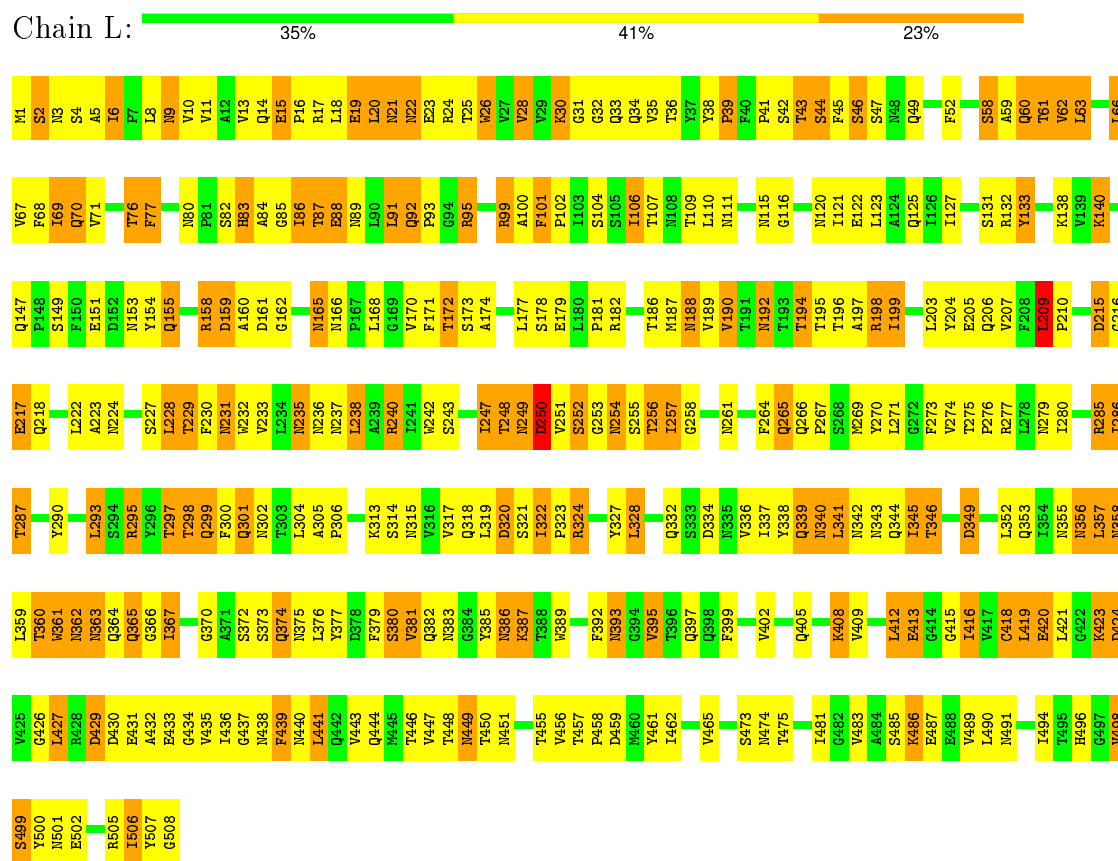
• Molecule 1: capsid protein V20

Chain K: 37% 40% 23%

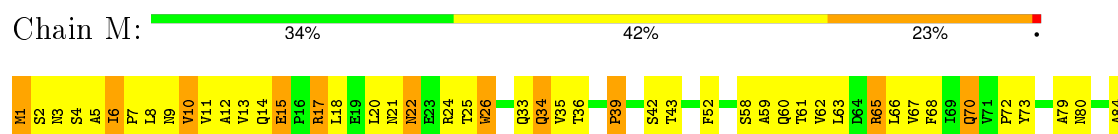


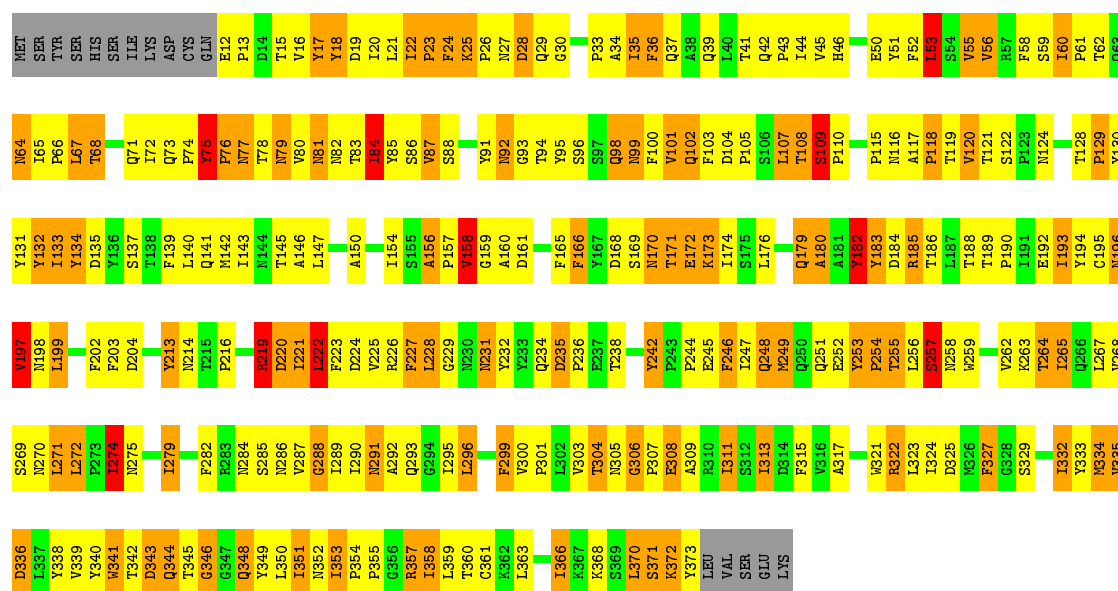


• Molecule 1: capsid protein V20



• Molecule 1: capsid protein V20





## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, 1	Depositor
Number of images	12000	Depositor
Resolution determination method	FSC at 0.143 cut-off	Depositor
CTF correction method	each particle	Depositor
Microscope	FEI Titan Krios	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	22	Depositor
Minimum defocus (nm)	1100	Depositor
Maximum defocus (nm)	2400	Depositor
Magnification	59000	Depositor
Image detector	KODAK SO163 FILM	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	0.57	2/4060 (0.0%)	0.77	3/5551 (0.1%)
1	B	0.57	2/4060 (0.0%)	0.75	5/5551 (0.1%)
1	C	0.57	1/4060 (0.0%)	0.78	2/5551 (0.0%)
1	D	0.56	1/4060 (0.0%)	0.73	1/5551 (0.0%)
1	E	0.58	2/4060 (0.0%)	0.76	3/5551 (0.1%)
1	F	0.57	1/4060 (0.0%)	0.73	3/5551 (0.1%)
1	G	0.57	2/4060 (0.0%)	0.74	3/5551 (0.1%)
1	H	0.59	1/4060 (0.0%)	0.76	5/5551 (0.1%)
1	I	0.61	3/4060 (0.1%)	0.77	5/5551 (0.1%)
1	J	0.60	1/4060 (0.0%)	0.78	5/5551 (0.1%)
1	K	0.56	1/4060 (0.0%)	0.73	1/5551 (0.0%)
1	L	0.57	1/4060 (0.0%)	0.74	3/5551 (0.1%)
1	M	0.56	2/4060 (0.0%)	0.80	7/5551 (0.1%)
2	N	0.58	1/2986 (0.0%)	0.77	3/4103 (0.1%)
All	All	0.57	21/55766 (0.0%)	0.76	49/76266 (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
1	A	0	1
1	B	0	3
1	C	0	1
1	D	0	1
1	E	0	1
1	F	0	1
1	L	0	1
1	M	0	1
All	All	0	10

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	I	117	PHE	CG-CD2	8.70	1.51	1.38
1	E	473	SER	CB-OG	8.12	1.52	1.42
1	I	117	PHE	CE1-CZ	8.06	1.52	1.37
2	N	341	TRP	CD2-CE2	5.93	1.48	1.41
1	M	361	TRP	CD2-CE2	5.72	1.48	1.41

The worst 5 of 49 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	M	302	ASN	N-CA-CB	14.64	136.95	110.60
1	C	506	ILE	N-CA-CB	-9.54	88.85	110.80
1	M	301	GLN	N-CA-C	-8.64	87.68	111.00
1	A	326	LEU	CA-CB-CG	7.07	131.57	115.30
1	L	209	LEU	CA-CB-CG	7.03	131.47	115.30

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	340	ASN	Peptide
1	B	132	ARG	Peptide
1	B	332	GLN	Peptide
1	B	380	SER	Peptide
1	C	341	LEU	Peptide

## 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3968	0	3871	408	0
1	B	3968	0	3871	627	0
1	C	3968	0	3871	631	0
1	D	3968	0	3871	619	0
1	E	3968	0	3871	509	0
1	F	3968	0	3871	429	0
1	G	3968	0	3871	497	0
1	H	3968	0	3871	523	0
1	I	3968	0	3871	521	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	J	3968	0	3871	525	0
1	K	3968	0	3871	401	0
1	L	3968	0	3871	492	0
1	M	3968	0	3871	524	0
2	N	2895	0	2810	508	0
All	All	54479	0	53133	6718	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 62.

The worst 5 of 6718 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:B:396:THR:CA	1:B:410:ILE:HD11	1.25	1.67
2:N:173:LYS:HD3	2:N:254:PRO:CG	1.22	1.62
1:C:376:LEU:HA	1:C:379:PHE:CE2	1.34	1.61
2:N:173:LYS:CD	2:N:254:PRO:HG3	1.24	1.60
2:N:186:THR:CG2	2:N:226:ARG:HD3	1.27	1.59

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	506/508 (100%)	430 (85%)	61 (12%)	15 (3%)	5	42
1	B	506/508 (100%)	456 (90%)	41 (8%)	9 (2%)	11	53
1	C	506/508 (100%)	433 (86%)	65 (13%)	8 (2%)	12	55
1	D	506/508 (100%)	442 (87%)	48 (10%)	16 (3%)	5	40
1	E	506/508 (100%)	443 (88%)	49 (10%)	14 (3%)	6	43

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	F	506/508 (100%)	431 (85%)	59 (12%)	16 (3%)	5	40
1	G	506/508 (100%)	447 (88%)	45 (9%)	14 (3%)	6	43
1	H	506/508 (100%)	436 (86%)	58 (12%)	12 (2%)	7	47
1	I	506/508 (100%)	424 (84%)	64 (13%)	18 (4%)	4	37
1	J	506/508 (100%)	434 (86%)	54 (11%)	18 (4%)	4	37
1	K	506/508 (100%)	436 (86%)	55 (11%)	15 (3%)	5	42
1	L	506/508 (100%)	449 (89%)	47 (9%)	10 (2%)	9	51
1	M	506/508 (100%)	439 (87%)	49 (10%)	18 (4%)	4	37
2	N	360/378 (95%)	280 (78%)	49 (14%)	31 (9%)	1	12
All	All	6938/6982 (99%)	5980 (86%)	744 (11%)	214 (3%)	9	41

5 of 214 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	86	ILE
1	A	173	SER
1	A	251	VAL
1	B	209	LEU
1	D	39	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	447/447 (100%)	274 (61%)	173 (39%)	0	1
1	B	447/447 (100%)	263 (59%)	184 (41%)	0	1
1	C	447/447 (100%)	266 (60%)	181 (40%)	0	1
1	D	447/447 (100%)	278 (62%)	169 (38%)	0	1
1	E	447/447 (100%)	284 (64%)	163 (36%)	0	1
1	F	447/447 (100%)	271 (61%)	176 (39%)	0	1
1	G	447/447 (100%)	280 (63%)	167 (37%)	0	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	H	447/447 (100%)	284 (64%)	163 (36%)	0	1
1	I	447/447 (100%)	266 (60%)	181 (40%)	0	1
1	J	447/447 (100%)	277 (62%)	170 (38%)	0	1
1	K	447/447 (100%)	276 (62%)	171 (38%)	0	1
1	L	447/447 (100%)	283 (63%)	164 (37%)	0	1
1	M	447/447 (100%)	274 (61%)	173 (39%)	0	1
2	N	328/344 (95%)	212 (65%)	116 (35%)	0	2
All	All	6139/6155 (100%)	3788 (62%)	2351 (38%)	1	1

5 of 2351 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	46	SER
1	H	326	LEU
1	M	341	LEU
1	G	140	LYS
1	G	425	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 410 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	236	ASN
1	H	501	ASN
1	M	231	ASN
1	G	339	GLN
1	H	153	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.