

## wwPDB EM Map/Model Validation Report ⓘ

Apr 10, 2016 – 01:54 PM BST

PDB ID : 3J70  
EMDB ID: : EMD-5020  
Title : Model of gp120, including variable regions, in complex with CD4 and 17b  
Authors : Rasheed, M.; Bettadapura, R.; Bajaj, C.  
Deposited on : 2014-04-22  
Resolution : unknown (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

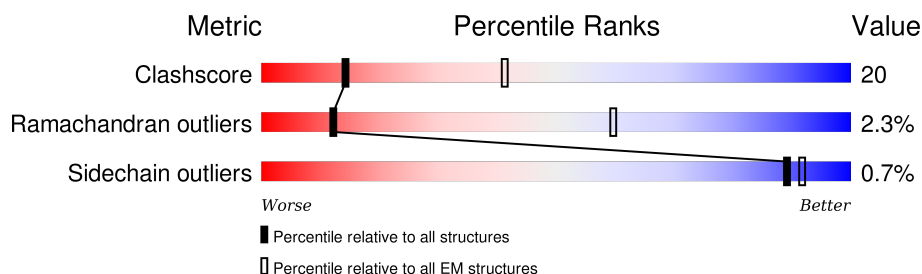
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is unknown.

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.






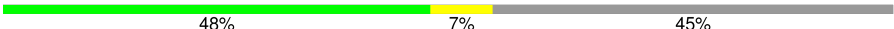


Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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	212	62% 35% .
1	M	212	62% 35% .
1	R	212	62% 35% .
2	B	210	67% 30% .
2	N	210	67% 29% .
2	S	210	67% 29% .
3	C	185	63% 34% ..
3	O	185	64% 34% ..
3	T	185	64% 34% ..

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	470	 65% 32% •
4	P	470	 65% 32% •
4	U	470	 64% 33% •
5	E	140	 48% 7% 45%
5	Q	140	 49% 6% 45%
5	V	140	 49% 6% 45%

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 26016 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 monoclonal antibody 17b light chain.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	212	Total	C	N	O	S	0	0
			1590	1007	255	321	7		
1	M	212	Total	C	N	O	S	0	0
			1590	1007	255	321	7		
1	R	212	Total	C	N	O	S	0	0
			1590	1007	255	321	7		

- Molecule 2 is a protein called monoclonal antibody 17b heavy chain.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	210	Total	C	N	O	S	0	0
			1605	1005	266	329	5		
2	N	210	Total	C	N	O	S	0	0
			1605	1005	266	329	5		
2	S	210	Total	C	N	O	S	0	0
			1605	1005	266	329	5		

- Molecule 3 is a protein called T-cell surface glycoprotein CD4.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	181	Total	C	N	O	S	0	0
			1412	885	247	276	4		
3	O	181	Total	C	N	O	S	0	0
			1412	885	247	276	4		
3	T	181	Total	C	N	O	S	0	0
			1412	885	247	276	4		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	184	ASN	-	EXPRESSION TAG	UNP P01730
C	185	THR	-	EXPRESSION TAG	UNP P01730
O	184	ASN	-	EXPRESSION TAG	UNP P01730

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
O	185	THR	-	EXPRESSION TAG	UNP P01730
T	184	ASN	-	EXPRESSION TAG	UNP P01730
T	185	THR	-	EXPRESSION TAG	UNP P01730

- Molecule 4 is a protein called Envelope glycoprotein gp120.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	470	Total	C	N	O	S	0	0
			3680	2307	643	703	27		
4	P	470	Total	C	N	O	S	0	0
			3680	2307	643	703	27		
4	U	470	Total	C	N	O	S	0	0
			3680	2307	643	703	27		

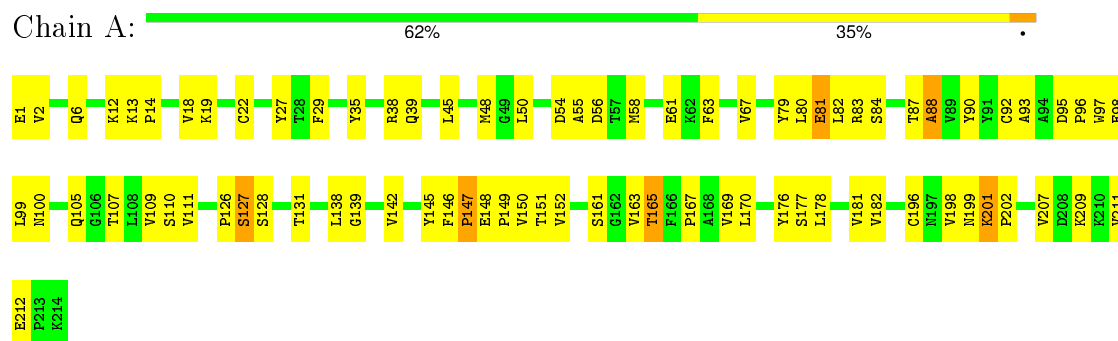
- Molecule 5 is a protein called envelope glycoprotein gp41.

Mol	Chain	Residues	Atoms				AltConf	Trace
5	E	77	Total	C	N	O	0	0
			385	231	77	77		
5	Q	77	Total	C	N	O	0	0
			385	231	77	77		
5	V	77	Total	C	N	O	0	0
			385	231	77	77		

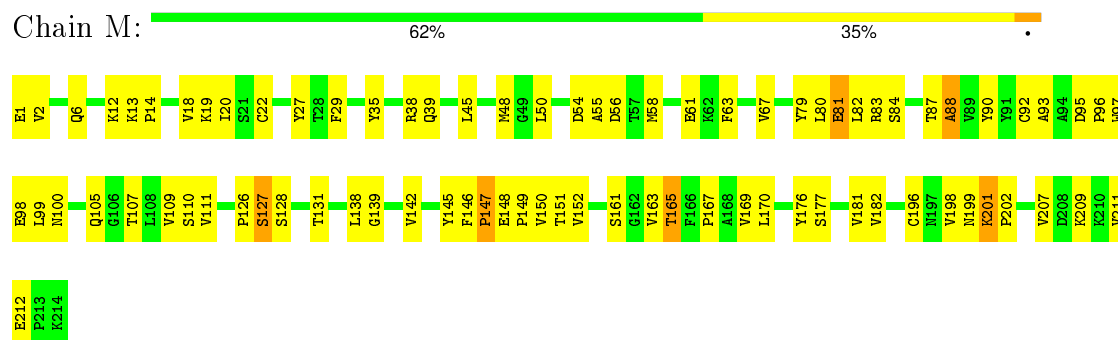
### 3 Residue-property plots

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.

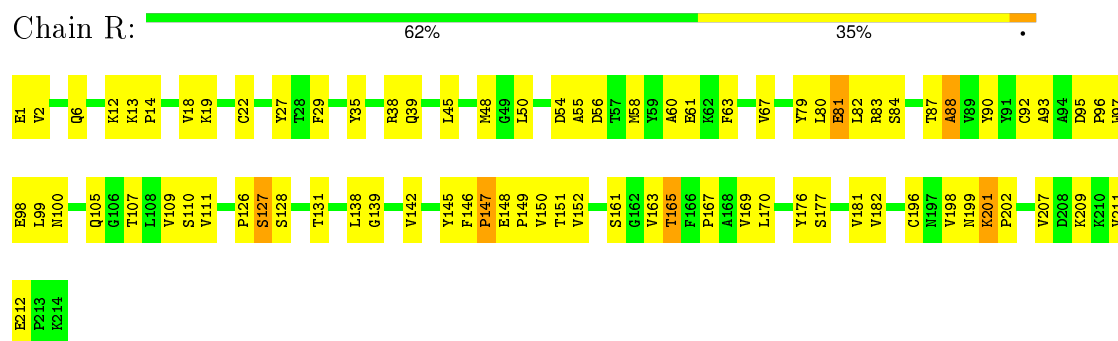
- Molecule 1: monoclonal antibody 17b light chain



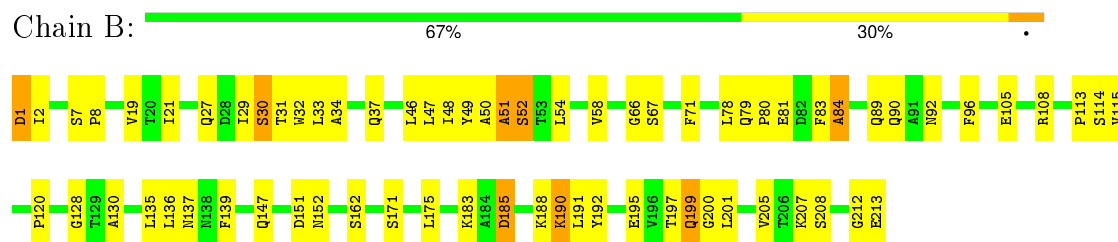
- Molecule 1: monoclonal antibody 17b light chain



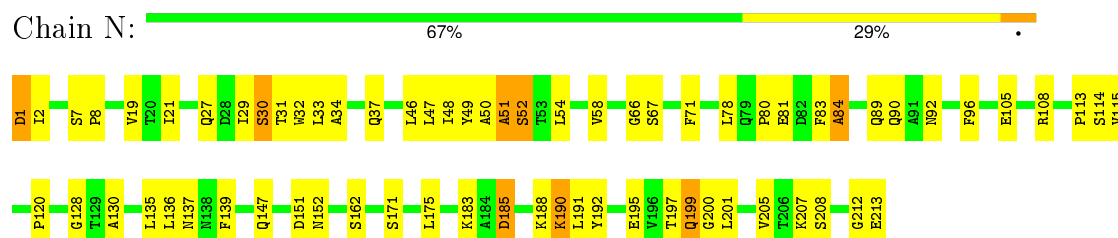
- Molecule 1: monoclonal antibody 17b light chain



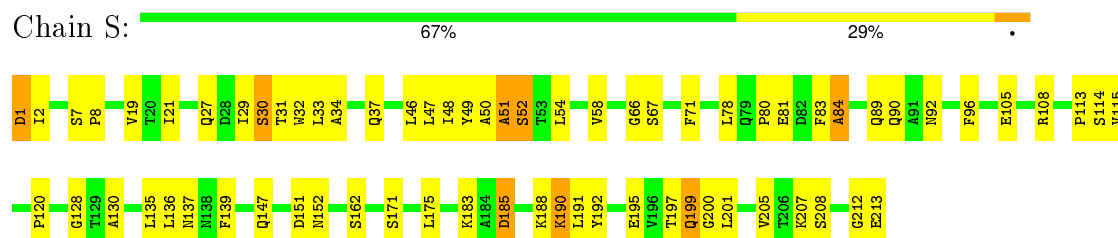
- Molecule 2: monoclonal antibody 17b heavy chain



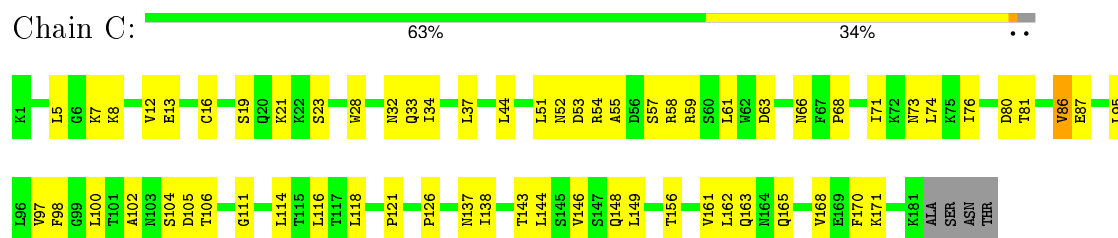
- Molecule 2: monoclonal antibody 17b heavy chain



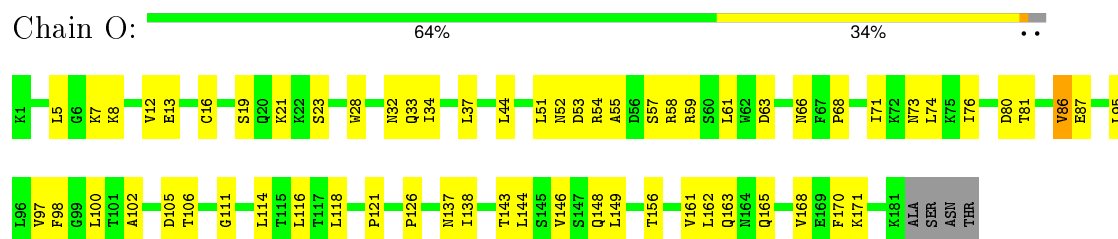
- Molecule 2: monoclonal antibody 17b heavy chain



- Molecule 3: T-cell surface glycoprotein CD4

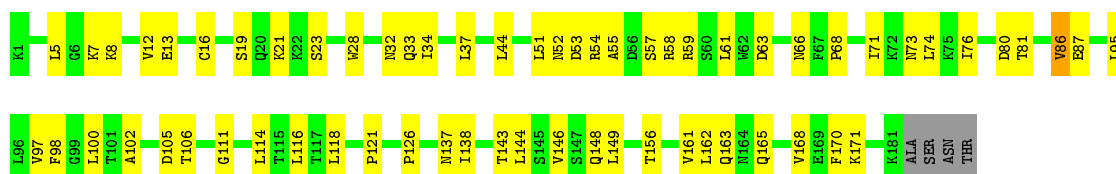


- Molecule 3: T-cell surface glycoprotein CD4



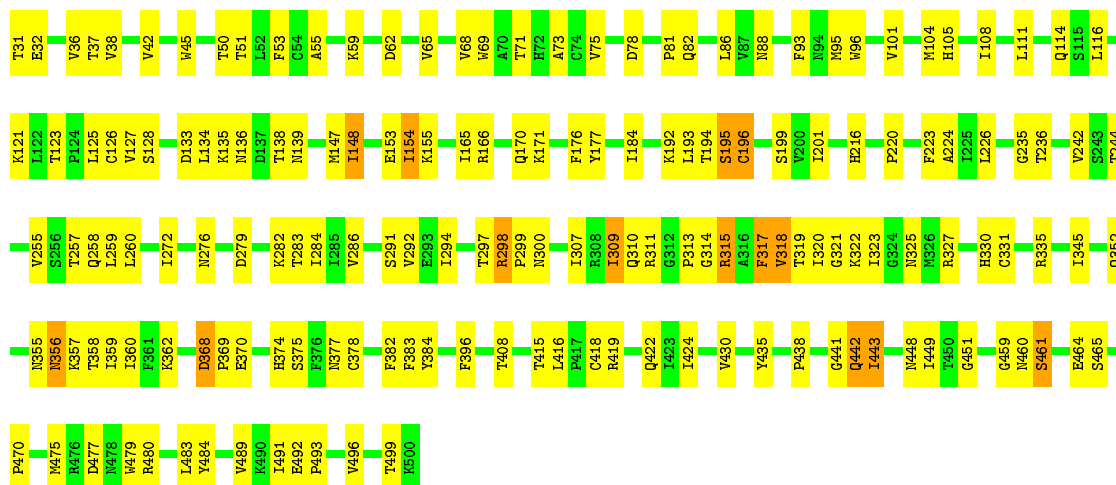
- Molecule 3: T-cell surface glycoprotein CD4





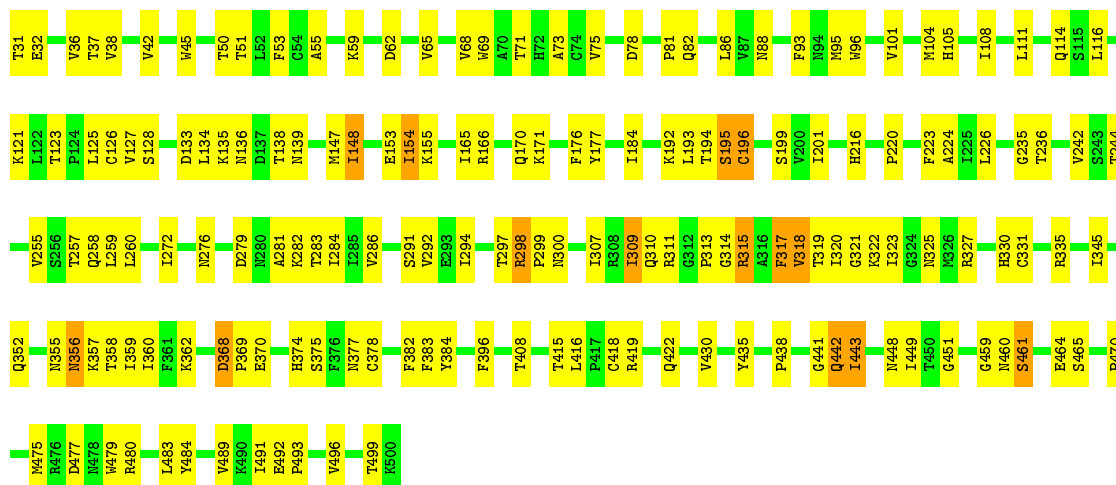
• Molecule 4: Envelope glycoprotein gp120

Chain D: 65% 32%



• Molecule 4: Envelope glycoprotein gp120

Chain P: 65% 32%

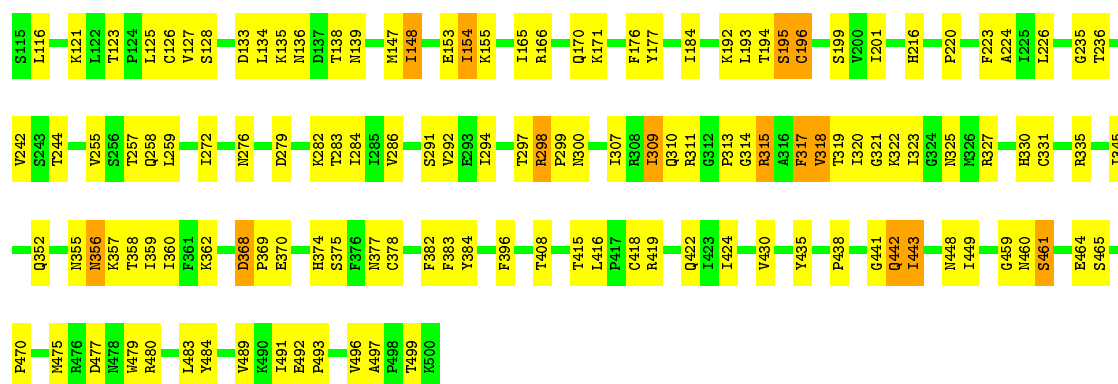


• Molecule 4: Envelope glycoprotein gp120

Chain U: 64% 33%

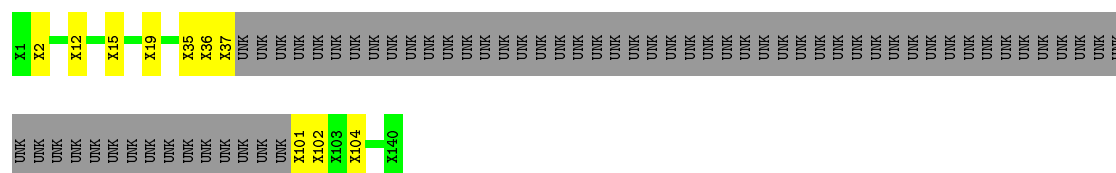






- Molecule 5: envelope glycoprotein gp41

Chain E: 48% 7% 45%



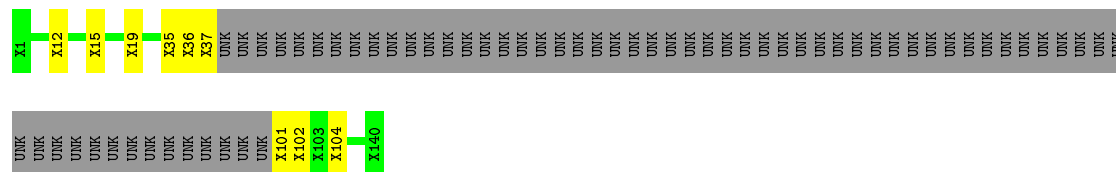
- Molecule 5: envelope glycoprotein gp41

Chain Q: 49% 6% 45%



- Molecule 5: envelope glycoprotein gp41

Chain V: 49% 6% 45%



## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	Not provided	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	Not provided	Depositor
Microscope	Not provided	Depositor
Voltage (kV)	Not provided	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	Not provided	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	Not provided	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	0.24	0/1627	0.41	0/2219
1	M	0.24	0/1627	0.41	0/2219
1	R	0.24	0/1627	0.41	0/2219
2	B	0.25	0/1638	0.38	0/2222
2	N	0.25	0/1638	0.38	0/2222
2	S	0.25	0/1638	0.38	0/2222
3	C	0.24	0/1432	0.37	0/1930
3	O	0.24	0/1432	0.37	0/1930
3	T	0.24	0/1432	0.37	0/1930
4	D	0.24	0/3757	0.36	0/5101
4	P	0.24	0/3757	0.36	0/5101
4	U	0.24	0/3757	0.36	0/5101
All	All	0.24	0/25362	0.37	0/34416

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 [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	1590	0	1557	80	0
1	M	1590	0	1557	81	0
1	R	1590	0	1557	82	0
2	B	1605	0	1560	60	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	N	1605	0	1560	62	0
2	S	1605	0	1560	62	0
3	C	1412	0	1444	53	0
3	O	1412	0	1444	52	0
3	T	1412	0	1444	53	0
4	D	3680	0	3636	171	0
4	P	3680	0	3636	172	0
4	U	3680	0	3636	173	0
5	E	385	0	82	23	0
5	Q	385	0	82	22	0
5	V	385	0	82	23	0
All	All	26016	0	24837	1031	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 20.

The worst 5 of 1031 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)
4:U:42:VAL:HB	5:V:104:UNK:CB	1.14	1.62
4:D:42:VAL:CB	5:E:104:UNK:CB	1.75	1.59
4:D:42:VAL:HB	5:E:104:UNK:CB	1.14	1.58
4:P:42:VAL:HB	5:Q:104:UNK:CB	1.14	1.58
4:P:42:VAL:CB	5:Q:104:UNK:CB	1.75	1.58

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	210/212 (99%)	197 (94%)	7 (3%)	6 (3%)	<b>6</b> <b>6</b>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	M	210/212 (99%)	197 (94%)	7 (3%)	6 (3%)	6	6
1	R	210/212 (99%)	197 (94%)	7 (3%)	6 (3%)	6	6
2	B	208/210 (99%)	198 (95%)	4 (2%)	6 (3%)	6	6
2	N	208/210 (99%)	198 (95%)	4 (2%)	6 (3%)	6	6
2	S	208/210 (99%)	198 (95%)	4 (2%)	6 (3%)	6	6
3	C	179/185 (97%)	172 (96%)	6 (3%)	1 (1%)	30	30
3	O	179/185 (97%)	172 (96%)	6 (3%)	1 (1%)	30	30
3	T	179/185 (97%)	172 (96%)	6 (3%)	1 (1%)	30	30
4	D	468/470 (100%)	434 (93%)	22 (5%)	12 (3%)	7	7
4	P	468/470 (100%)	434 (93%)	22 (5%)	12 (3%)	7	7
4	U	468/470 (100%)	434 (93%)	22 (5%)	12 (3%)	7	7
All	All	3195/3231 (99%)	3003 (94%)	117 (4%)	75 (2%)	12	8

5 of 75 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	81	GLU
1	A	127	SER
1	A	147	PRO
1	A	201	LYS
2	B	51	ALA

### 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	178/178 (100%)	178 (100%)	0	100	100
1	M	178/178 (100%)	178 (100%)	0	100	100
1	R	178/178 (100%)	178 (100%)	0	100	100
2	B	182/182 (100%)	178 (98%)	4 (2%)	60	60
2	N	182/182 (100%)	178 (98%)	4 (2%)	60	60

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	S	182/182 (100%)	178 (98%)	4 (2%)	60	60
3	C	164/167 (98%)	163 (99%)	1 (1%)	90	90
3	O	164/167 (98%)	163 (99%)	1 (1%)	90	90
3	T	164/167 (98%)	163 (99%)	1 (1%)	90	90
4	D	421/421 (100%)	419 (100%)	2 (0%)	92	92
4	P	421/421 (100%)	419 (100%)	2 (0%)	92	92
4	U	421/421 (100%)	419 (100%)	2 (0%)	92	92
All	All	2835/2844 (100%)	2814 (99%)	21 (1%)	89	88

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	N	185	ASP
3	O	137	ASN
2	S	207	LYS
2	N	105	GLU
3	T	137	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
3	O	52	ASN
4	P	160	ASN
4	U	160	ASN
3	O	137	ASN
4	P	325	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.