



## wwPDB EM Map/Model Validation Report ⓘ

Apr 10, 2016 – 02:19 PM BST

PDB ID : 5GAP  
EMDB ID: : EMD-8014  
Title : Body region of the U4/U6.U5 tri-snRNP  
Authors : Nguyen, T.H.D.; Galej, W.P.; Oubridge, C.; Bai, X.C.; Newman, A.; Scheres, S.; Nagai, K.  
Deposited on : 2015-12-15  
Resolution : 3.60 Å(reported)  
Based on PDB ID : ?

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

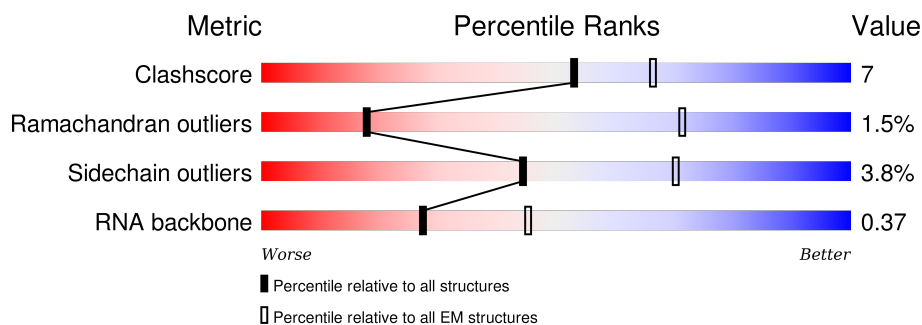
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.60 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
RNA backbone	3027	244

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	V	67	42% 37% 19% .
2	W	112	23% 17% 10% 50%
3	U	214	6% . . 91%
4	x	82	100%
5	A	2413	47% 9% 44%
6	H	465	59% 18% 23%
7	J	899	64% 16% . 19%
8	D	143	76% 21% . .

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	F	494	<div><div></div><div>65%</div><div>18%</div><div>•</div><div>16%</div></div>
10	G	469	<div><div></div><div>59%</div><div>9%</div><div></div><div>32%</div></div>
11	K	126	<div><div></div><div>70%</div><div>25%</div><div>• •</div></div>
12	B	2163	<div><div></div><div>•</div><div>97%</div></div>

## 2 Entry composition

There are 12 unique types of molecules in this entry. The entry contains 31576 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 RNA chain called U4 snRNA, 5' region, nucleotides 1-67.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	V	67	Total	C	N	O	P	0	0
			1426	637	247	475	67		

- Molecule 2 is a RNA chain called U6 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	W	56	Total	C	N	O	P	0	0
			1190	533	210	391	56		

- Molecule 3 is a RNA chain called U5 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	U	20	Total	C	N	O	P	0	0
			414	186	64	144	20		

- Molecule 4 is a protein called unknown protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
4	x	82	Total	C	N	O	0	0
			410	246	82	82		

- Molecule 5 is a protein called Pre-mRNA-splicing factor 8.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	A	1349	Total	C	N	O	S	0	0
			11066	7094	1901	2031	40		

- Molecule 6 is a protein called U4/U6 small nuclear ribonucleoprotein PRP4.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	H	357	Total	C	N	O	S	0	0
			2789	1743	501	532	13		

- Molecule 7 is a protein called Pre-mRNA-splicing factor 6.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	J	729	Total	C	N	O	S	0	0
			5822	3726	992	1079	25		

- Molecule 8 is a protein called Spliceosomal protein DIB1.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	D	140	Total	C	N	O	S	0	0
			1151	728	200	212	11		

- Molecule 9 is a protein called Pre-mRNA-processing factor 31.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	F	415	Total	C	N	O	S	0	0
			3218	2052	575	580	11		

- Molecule 10 is a protein called U4/U6 small nuclear ribonucleoprotein PRP3.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	G	318	Total	C	N	O	S	0	0
			2632	1659	469	488	16		

- Molecule 11 is a protein called 13 kDa ribonucleoprotein-associated protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	K	124	Total	C	N	O	S	0	0
			936	597	161	174	4		

- Molecule 12 is a protein called Pre-mRNA-splicing helicase BRR2.

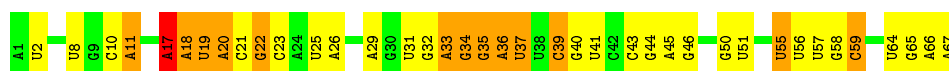
Mol	Chain	Residues	Atoms					AltConf	Trace
12	B	71	Total	C	N	O	S	0	0
			522	326	89	106	1		

### 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.

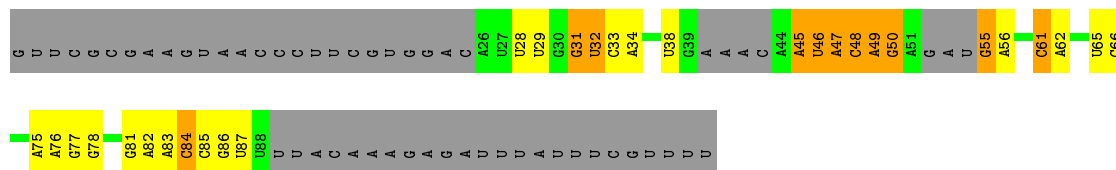
- Molecule 1: U4 snRNA, 5' region, nucleotides 1-67

Chain V: 



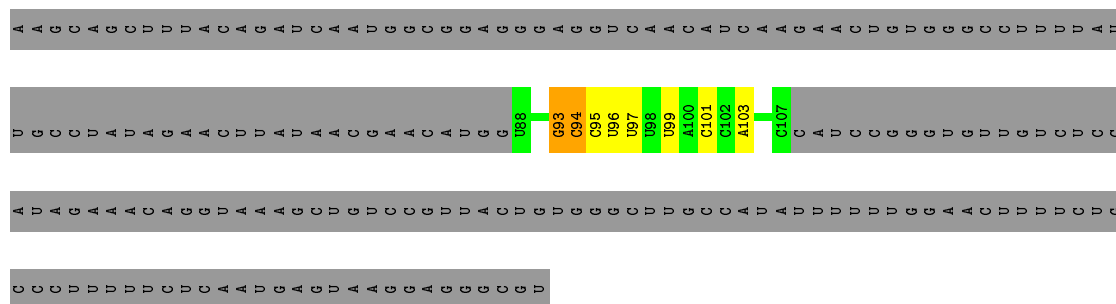
- Molecule 2: U6 snRNA

Chain W: 



- Molecule 3: U5 snRNA

Chain U: 



- Molecule 4: unknown protein

Chain x: 

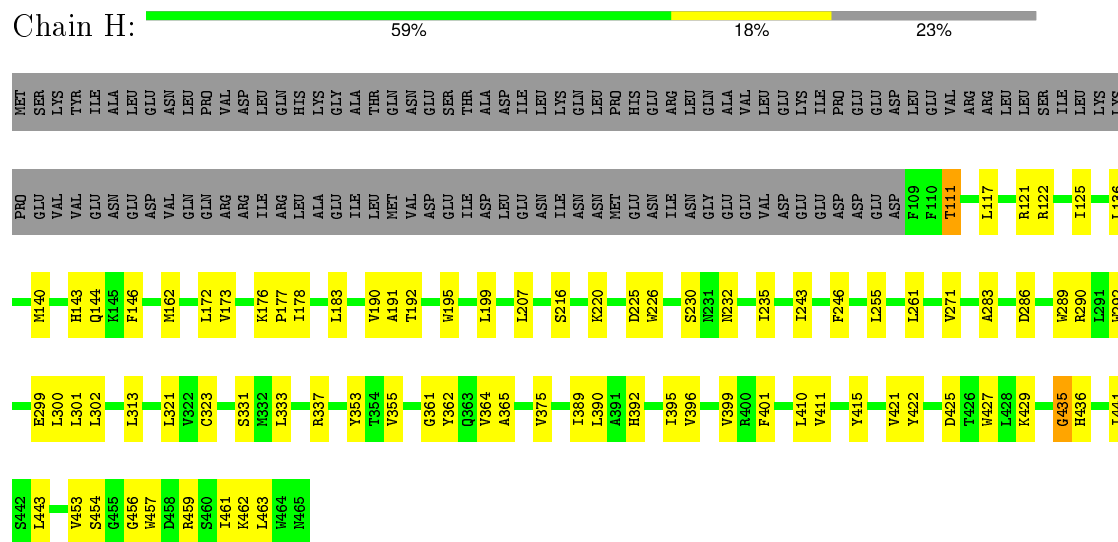
There are no outlier residues recorded for this chain.

- Molecule 5: Pre-mRNA-splicing factor 8

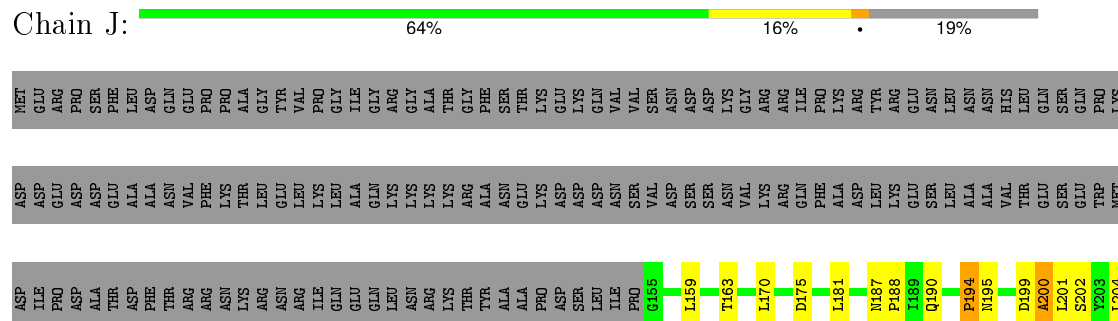
Chain A: 

I1632	I1407	M1262	L1049	L852	LEU	ASN	HIS	ALA	GLU	TRP	PRO	HIS	LYS	ILE	MET
I1638	L1412	R1268	L1050	T853	LEU	HIS	SER	LEU	GLU	SER	PHE	PRO	ALA	ASN	SER
I1646	W1414	R1268	F1063	I857	ARG	LYS	PHE	PHE	TYR	ASP	ASP	TRP	ARG	GLU	GLY
F1649	A1427	P1271	T1064	I874	ILE	LYS	THR	PRO	GLY	LEU	GLU	ALA	ASN	ILE	PRO
W1654	GLY	M1275	L1065	I882	ASN	LYS	LEU	SER	GLY	PRO	PRO	LYS	LEU	VAL	PRO
I1657	MET	E1276	L1066	I914	LEU	THR	THR	ASN	ILE	MET	LEU	VAL	THR	PHE	GLY
I1668	HIS	W1278	L1070	I919	THR	LYS	LEU	ARG	LYS	TYR	GLU	VAL	PRO	PRO	PHE
L1669	GLU	D1281	I1073	I922	ARG	LYS	ASN	THR	VAL	ARG	GLN	THR	ALA	GLU	GLU
V1681	D1433	D1282	I1082	W923	PHE	THR	VAL	ARG	THR	LEU	HIS	HIS	MET	PRO	ASP
Y1743	I1437	W1286	T1083	I939	GLY	LYS	ARG	ILE	PHE	SER	ILE	THR	PRO	PRO	SER
S1749	F1451	K1085	A1084	I943	ARG	PHE	GLN	ASN	VAL	PRO	ILE	GLY	ALA	PRO	LEU
P1768	W1458	M1086	M1087	P947	SER	THR	GLN	ASN	SER	ARG	GLU	THR	LEU	GLY	ALA
I1777	Y1461	F1312	V1088	T950	GLU	HIS	THR	ALA	THR	ASP	PRO	THR	ARG	LEU	LEU
D1778	L1494	G1318	F1092	T960	THR	LYS	LEU	ASN	VAL	ASP	ASN	PRO	ILE	GLY	LEU
L1779	D1505	I1319	K1093	M971	LYS	GLY	GLU	THR	THR	LYS	LYS	PRO	ASN	THR	PRO
F1791	R1512	M1321	D1094	M972	ARG	GLN	GLY	LEU	ALA	TYR	LEU	ASP	GLN	ALA	PRO
L1794	K1535	G1323	M1095	I978	LEU	LYS	LEU	GLN	GLY	PHE	LEU	ASP	GLN	VAL	GLY
K1795	L1536	W1335	I1103	S979	THR	LYS	LEU	GLN	THR	ASP	GLY	ASP	GLN	VAL	ILE
P1796	W1537	I1104	I1107	P980	ASN	ASN	ARG	GLY	THR	GLU	THR	ASP	ASP	GLU	GLU
S1801	Y1542	L1342	I1113	Y832	LYS	GLY	GLN	GLY	LEU	TYR	THR	LYS	GLY	LEU	LEU
A1811	V1546	F1343	F1114	I950	GLY	HIS	VAL	VAL	ASP	VAL	GLY	LYS	THR	VAL	ASN
R1820	I1553	T1344	Q1115	D1004	VAL	ASN	GLN	VAL	ASP	ASP	ASP	TRP	LYS	LYS	PRO
L1823	E1558	E1354	Y1116	L1004	GLY	LEU	THR	PHE	PHE	GLY	TRP	SER	ASP	LYS	MET
Q1824	H1559	P1355	L1125	P1010	PRO	LEU	PHE	ALA	ALA	ALA	TRP	ALA	LYS	LEU	VAL
Q1827	T1560	L1360	L1126	M1011	GLY	ASN	ARG	PRO	PRO	PRO	TYR	THR	ALA	ASP	PRO
L1835	A1578	R1366	Y1161	W1012	GLY	ILE	ASN	LEU	LEU	LEU	ALA	ILE	GLY	GLY	SER
L1836	S1599	I1367	M1170	F1022	PHE	VAL	HIS	GLU	TYR	ASN	ARG	ALA	GLY	LYS	VAL
S1837	Q1600	Q1368	M170	L1023	GLN	ASP	ARG	GLU	ASN	ALA	ARG	ARG	LEU	THR	ASN
L1843	I1601	K1372	T1183	L1024	ALA	ALA	PHE	GLY	PRO	ILE	LEU	ARG	LYS	VAL	PRO
L1849	P1602	L1375	I1216	V1025	TRP	GLN	GLN	LEU	SER	GLY	GLY	ARG	PRO	THR	PHE
L1850	R1605	L1375	I1216	K1027	ARG	LEU	LEU	ILE	TYR	PRO	ASP	ASP	GLY	ARG	LEU
T1881	F1606	F1383	M1221	W1028	VAL	ALA	THR	VAL	ILE	SER	LYS	ASP	HIS	LYS	PRO
W1893	T1607	P1384	I1230	T1029	TRP	ASP	GLY	ASP	ILE	LYS	THR	LYS	ALA	SER	PRO
W1911	L1608	P1385	I1230	L1032	ASN	ILE	ILE	LEU	THR	GLN	ARG	THR	ILE	ARG	PRO
K1912	W1609	A1386	L1238	I1038	PHE	HIS	GLY	LYS	ASN	GLY	GLY	PHE	LYS	ARG	ALA
T1913	W1610	V1387	L1238	W1039	LEU	THR	THR	ASN	GLU	GLY	ASP	THR	LEU	LEU	PRO
	I1613	L1394	V1250	A1047	ARG	ILE	LEU	ALA	TYR	PRO	THR	ARG	GLU	LYS	SER
	I1614	A1402	Y1251	V1048	GLY	THR	ASN	MET	GLY	TYR	SER	THR	GLU	LYS	THR
	F1623		M1254		ILE	THR	THR	TYR	PRO	GLU	LYS	PHE	MET	PRO	GLU

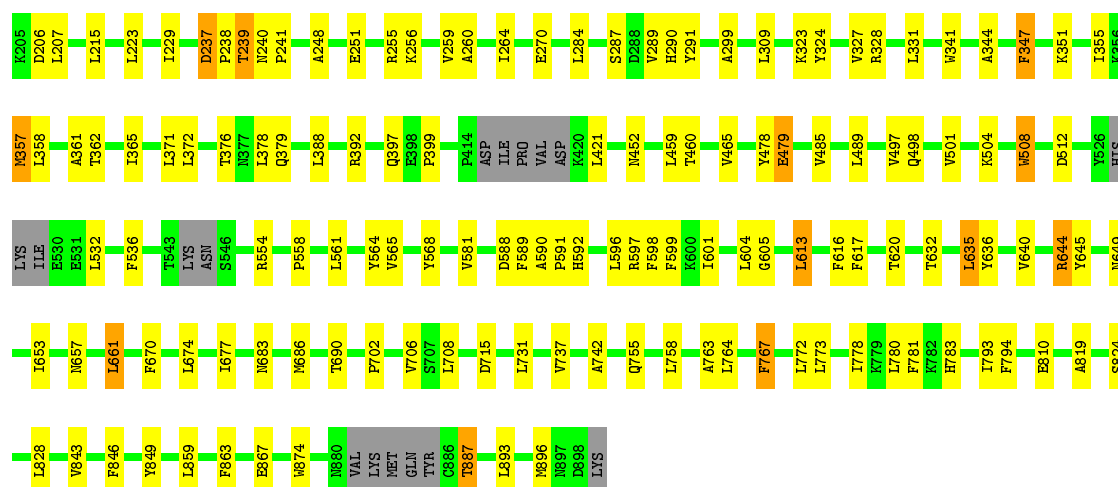
- Molecule 6: U4/U6 small nuclear ribonucleoprotein PRP4



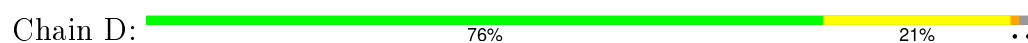
- Molecule 7: Pre-mRNA-splicing factor 6



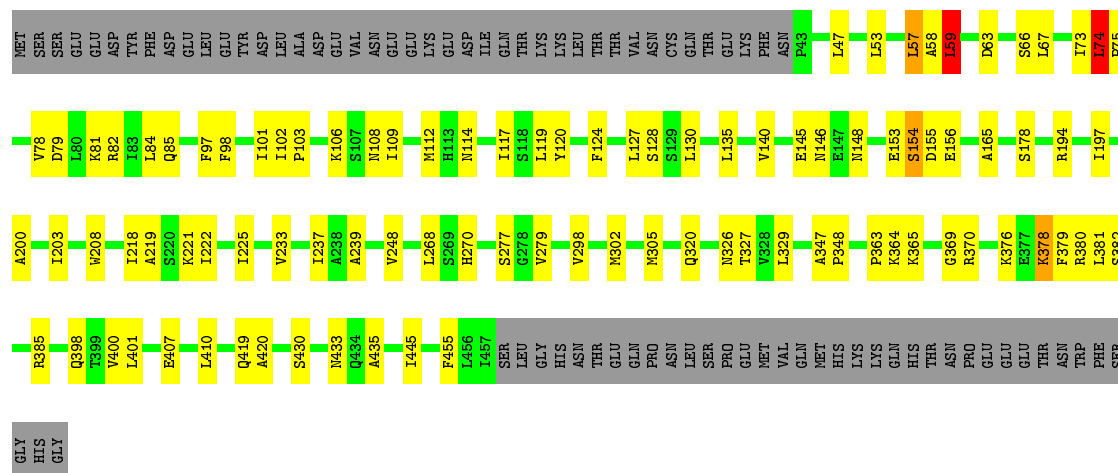




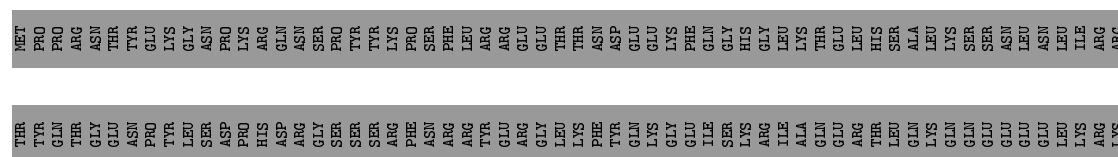
- Molecule 8: Spliceosomal protein DIB1



- Molecule 9: Pre-mRNA-processing factor 31



- Molecule 10: U4/U6 small nuclear ribonucleoprotein PRP3











## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	140155	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	Not provided	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	81000	Depositor
Image detector	Not provided	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	V	0.36	0/1593	0.79	1/2480 (0.0%)
10	G	0.41	0/2687	0.69	0/3611
11	K	0.45	0/949	0.81	0/1292
12	B	0.42	0/529	0.69	0/716
2	W	0.38	0/1328	0.84	2/2061 (0.1%)
3	U	0.32	0/459	0.76	0/710
5	A	0.42	1/11327 (0.0%)	0.73	2/15348 (0.0%)
6	H	0.38	0/2845	0.71	0/3843
7	J	0.44	0/5934	0.80	2/8039 (0.0%)
8	D	0.42	0/1172	0.75	1/1578 (0.1%)
9	F	0.43	0/3273	0.80	2/4413 (0.0%)
All	All	0.42	1/32096 (0.0%)	0.76	10/44091 (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
12	B	0	1
5	A	0	3
9	F	0	1
All	All	0	5

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	825	SER	CB-OG	5.40	1.49	1.42

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	J	613	LEU	CA-CB-CG	7.12	131.69	115.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	979	SER	C-N-CD	-6.95	105.32	120.60
1	V	17	A	C2'-C3'-O3'	6.90	124.74	113.70
2	W	55	G	C2'-C3'-O3'	5.73	122.87	113.70
9	F	74	LEU	CA-CB-CG	5.58	128.14	115.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	A	1278	VAL	Peptide
5	A	1994	ASP	Peptide
5	A	979	SER	Peptide
12	B	423	ILE	Peptide
9	F	154	SER	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	V	1426	0	716	22	0
2	W	1190	0	603	14	0
3	U	414	0	213	4	0
4	x	410	0	89	0	0
5	A	11066	0	11078	138	0
6	H	2789	0	2725	53	0
7	J	5822	0	5792	103	0
8	D	1151	0	1138	17	0
9	F	3218	0	3297	53	0
10	G	2632	0	2599	27	0
11	K	936	0	987	24	0
12	B	522	0	506	5	0
All	All	31576	0	29743	413	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 7.

The worst 5 of 413 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)
5:A:1578:ALA:HB1	5:A:1602:PRO:HB3	1.40	1.02
5:A:1275:MET:HE2	5:A:1281:ASN:ND2	1.92	0.84
10:G:272:VAL:HG12	10:G:280:VAL:HG21	1.59	0.81
5:A:1067:ASN:HB2	5:A:1083:THR:HG21	1.64	0.80
7:J:199:ASP:O	7:J:201:LEU:N	2.14	0.79

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	
5	A	1343/2413 (56%)	1226 (91%)	104 (8%)	13 (1%)	19	66
6	H	355/465 (76%)	301 (85%)	48 (14%)	6 (2%)	11	55
7	J	719/899 (80%)	650 (90%)	56 (8%)	13 (2%)	11	54
8	D	138/143 (96%)	126 (91%)	11 (8%)	1 (1%)	26	72
9	F	413/494 (84%)	364 (88%)	38 (9%)	11 (3%)	6	46
10	G	316/469 (67%)	283 (90%)	32 (10%)	1 (0%)	46	83
11	K	122/126 (97%)	111 (91%)	9 (7%)	2 (2%)	12	56
12	B	69/2163 (3%)	57 (83%)	8 (12%)	4 (6%)	2	25
All	All	3475/7172 (48%)	3118 (90%)	306 (9%)	51 (2%)	18	57

5 of 51 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	A	1093	LYS
5	A	1278	VAL
5	A	2088	ILE
6	H	362	TYR
7	J	479	GLU



### 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	
5	A	1216/2182 (56%)	1194 (98%)	22 (2%)	66	89
6	H	305/410 (74%)	290 (95%)	15 (5%)	31	71
7	J	627/813 (77%)	600 (96%)	27 (4%)	35	75
8	D	129/132 (98%)	121 (94%)	8 (6%)	23	65
9	F	346/445 (78%)	324 (94%)	22 (6%)	22	63
10	G	289/436 (66%)	283 (98%)	6 (2%)	61	87
11	K	102/104 (98%)	90 (88%)	12 (12%)	6	34
12	B	56/1955 (3%)	52 (93%)	4 (7%)	18	59
All	All	3070/6477 (47%)	2954 (96%)	116 (4%)	44	77

5 of 116 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	J	588	ASP
8	D	39	CYS
11	K	69	LEU
7	J	635	LEU
7	J	661	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
6	H	303	GLN
7	J	498	GLN
10	G	300	GLN
6	H	306	HIS
5	A	1281	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	V	66/67 (98%)	29 (43%)	6 (9%)
2	W	54/112 (48%)	20 (37%)	9 (16%)
3	U	19/214 (8%)	8 (42%)	1 (5%)
All	All	139/393 (35%)	57 (41%)	16 (11%)

5 of 57 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	V	2	U
1	V	11	A
1	V	18	A
1	V	19	U
1	V	20	A

5 of 16 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	W	32	U
2	W	45	A
2	W	75	A
2	W	31	G
2	W	83	A

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
4	x	1
5	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	x	62:UNK	C	101:UNK	N	54.04
1	A	1860:VAL	C	1861:THR	N	4.50