



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Feb 19, 2018 – 05:52 PM EST

PDB ID : 6BLL
EMDB ID: : EMD-7112
Title : Cryo-EM structure of human CPSF-160-WDR33-CPSF-30-PAS RNA complex
at 3.4 Å resolution
Authors : Sun, Y.; Zhang, Y.; Hamilton, K.; Walz, T.; Tong, L.
Deposited on : 2017-11-10
Resolution : 3.40 Å(reported)
Based on PDB ID : 2RHK

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

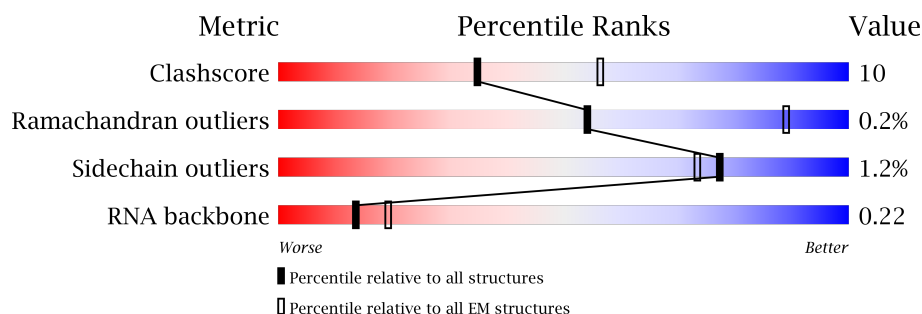
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.40 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
RNA backbone	3398	335

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	1443	65% 16% 18%
2	B	587	49% 15% 36%
3	C	245	40% 7% 52%
4	E	17	18% 24% 59%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 13435 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 Cleavage and polyadenylation specificity factor subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1177	Total	C	N	O	S	0	0
			9337	6003	1600	1679	55		

- Molecule 2 is a protein called pre-mRNA 3' end processing protein WDR33.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	375	Total	C	N	O	S	0	0
			3028	1912	556	540	20		

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-14	MET	-	expression tag	UNP Q9C0J8
B	-13	GLY	-	expression tag	UNP Q9C0J8
B	-12	SER	-	expression tag	UNP Q9C0J8
B	-11	SER	-	expression tag	UNP Q9C0J8
B	-10	HIS	-	expression tag	UNP Q9C0J8
B	-9	HIS	-	expression tag	UNP Q9C0J8
B	-8	HIS	-	expression tag	UNP Q9C0J8
B	-7	HIS	-	expression tag	UNP Q9C0J8
B	-6	HIS	-	expression tag	UNP Q9C0J8
B	-5	HIS	-	expression tag	UNP Q9C0J8
B	-4	SER	-	expression tag	UNP Q9C0J8
B	-3	SER	-	expression tag	UNP Q9C0J8
B	-2	GLY	-	expression tag	UNP Q9C0J8
B	-1	LEU	-	expression tag	UNP Q9C0J8
B	0	VAL	-	expression tag	UNP Q9C0J8

- Molecule 3 is a protein called Cleavage and polyadenylation specificity factor subunit 4.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	117	Total	C	N	O	S	0	0
			917	587	151	165	14		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	0	GLY	-	expression tag	UNP O95639

- Molecule 4 is a RNA chain called RNA (5'-R(P*AP*AP*UP*AP*AP*AP*C)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	7	Total	C	N	O	P	0	0
			150	68	30	45	7		

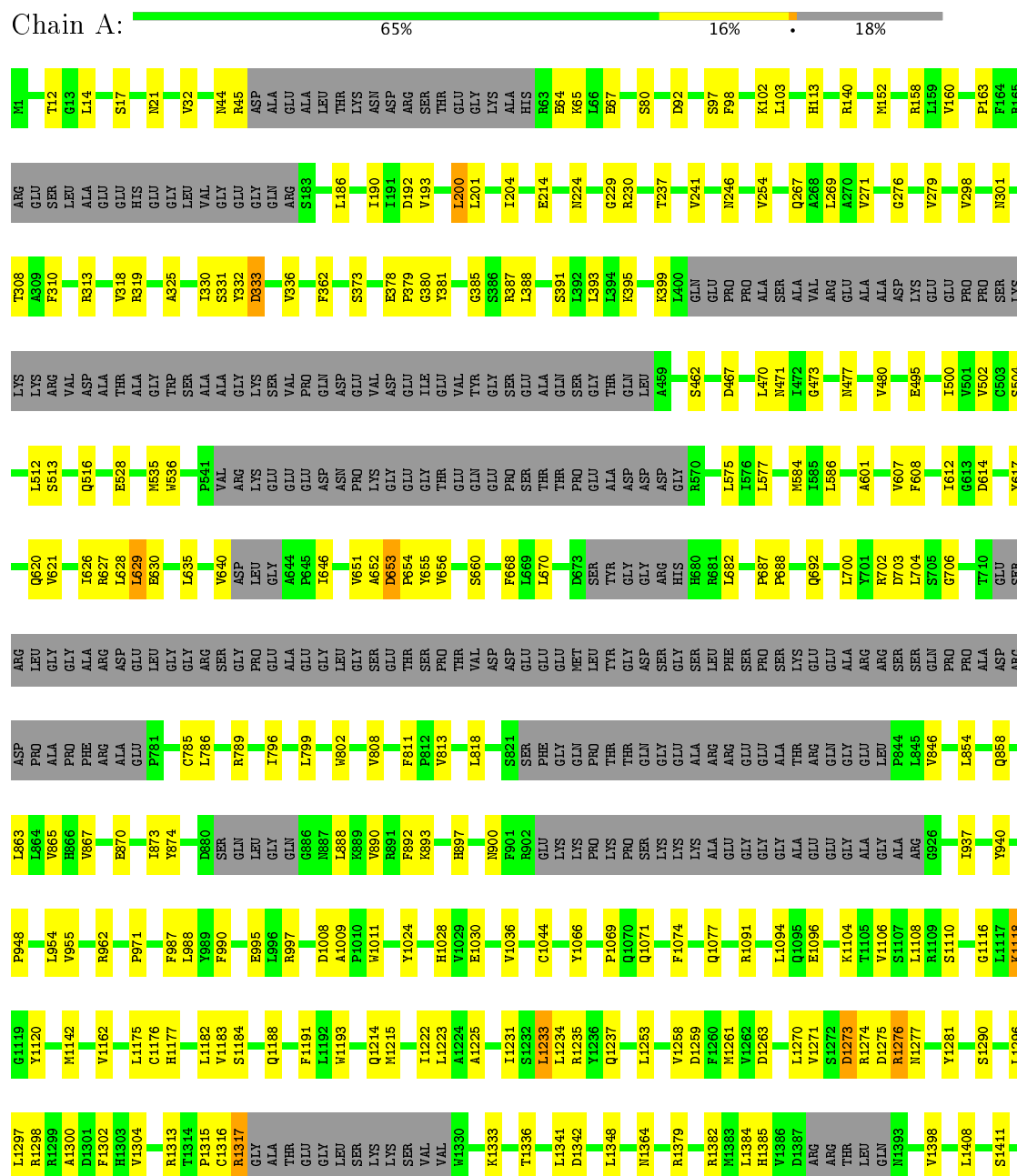
- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
5	C	3	Total	Zn	0
			3	3	

3 Residue-property plots

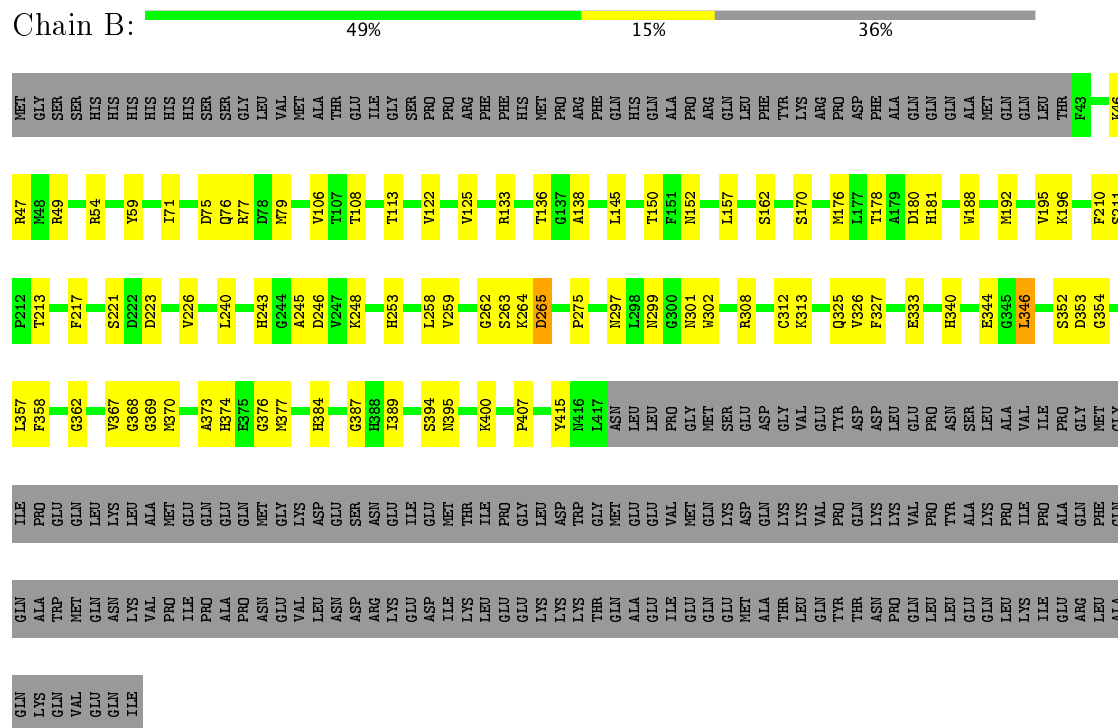
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according 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: Cleavage and polyadenylation specificity factor subunit 1

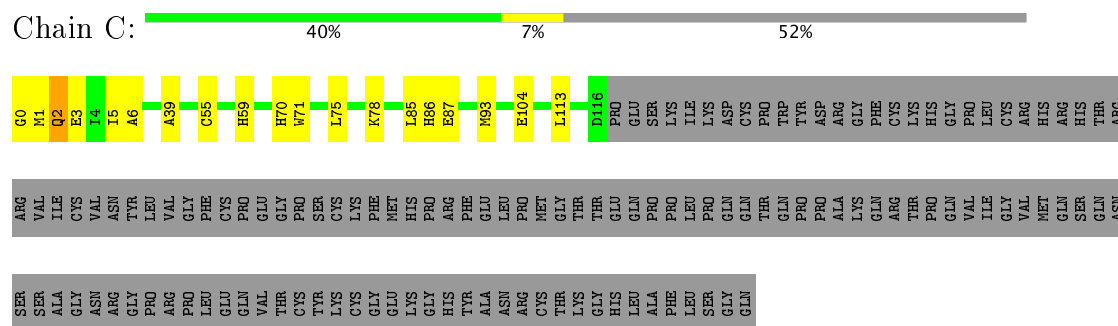




- Molecule 2: pre-mRNA 3' end processing protein WDR33



- Molecule 3: Cleavage and polyadenylation specificity factor subunit 4



- Molecule 4: RNA (5'-R(P*AP*AP*UP*AP*AP*AP*C)-3')



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	173632	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 ($e^-/\text{\AA}^2$)	70	Depositor
Minimum defocus (nm)	900	Depositor
Maximum defocus (nm)	2800	Depositor
Magnification	46729	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.47	0/9553	0.61	0/12967
2	B	0.46	0/3113	0.63	0/4214
3	C	0.40	0/941	0.57	0/1261
4	E	0.46	0/168	0.79	0/259
All	All	0.46	0/13775	0.62	0/18701

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	9337	0	9390	186	0
2	B	3028	0	2942	69	0
3	C	917	0	889	19	0
4	E	150	0	77	3	0
5	C	3	0	0	0	0
All	All	13435	0	13298	267	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 10.

The worst 5 of 267 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:A:608:PHE:HZ	1:A:651:VAL:CG1	1.44	1.29
1:A:1253:LEU:HD11	1:A:1273:ASP:OD2	1.32	1.26
1:A:608:PHE:CZ	1:A:651:VAL:HG12	1.72	1.24
1:A:608:PHE:CZ	1:A:651:VAL:CG1	2.21	1.22
2:B:245:ALA:HB2	2:B:265:ASP:HB2	1.27	1.16

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	1151/1443 (80%)	1056 (92%)	93 (8%)	2 (0%)	51	84
2	B	373/587 (64%)	339 (91%)	33 (9%)	1 (0%)	44	79
3	C	115/245 (47%)	98 (85%)	16 (14%)	1 (1%)	20	61
All	All	1639/2275 (72%)	1493 (91%)	142 (9%)	4 (0%)	54	84

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	653	ASP
3	C	2	GLN
1	A	201	LEU
2	B	152	ASN

5.3.2 Protein sidechains [i](#)

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	1030/1235 (83%)	1015 (98%)	15 (2%)	70	87
2	B	325/514 (63%)	322 (99%)	3 (1%)	82	92
3	C	99/211 (47%)	99 (100%)	0	100	100
All	All	1454/1960 (74%)	1436 (99%)	18 (1%)	77	89

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1235	ARG
1	A	1273	ASP
1	A	1317	ARG
1	A	1176	CYS
1	A	1233	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1332	ASN
2	B	416	ASN
2	B	325	GLN
1	A	783	HIS
2	B	343	HIS

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	E	6/17 (35%)	2 (33%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	E	4	A
4	E	5	A

There are no RNA pucker outliers to report.

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 ⓘ

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.