wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Feb 11, 2018 – 12:36 AM EST

PDB ID	:	6FBS
EMDB ID:	:	EMD-4225
Title	:	Cryo-EM structure of the human CPSF160-WDR33-CPSF30 complex bound to the PAS AAUAAA motif at 3.1 Angstrom resolution
Authors	:	Clerici, M.; Faini, M.; Jinek, M.
Deposited on	:	2017-12-19
Resolution	:	3.07 Å(reported)
Based on PDB ID	:	6F9N

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
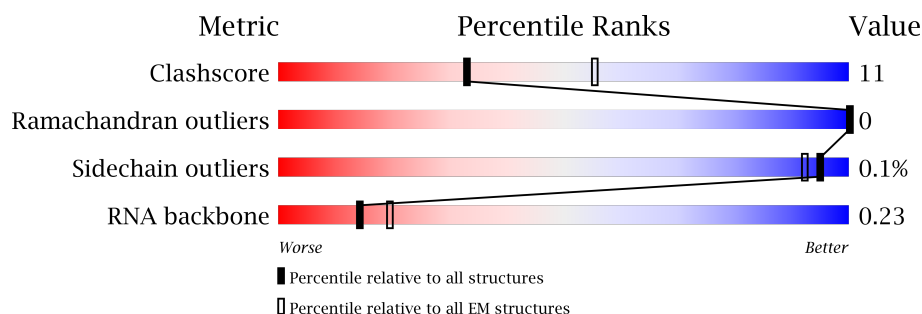with specific help available everywhere you see the ⓘ symbol.

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 3.07 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	Percentile Ranks	Value
Clashscore		11
Ramachandran outliers		0
Sidechain outliers		0.1%
RNA backbone		0.23

Worse .... *Better*

■ Percentile relative to all structures
▯ Percentile relative to all EM structures

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	1446	61% · 21% · 18%
2	B	413	65% · 25% · 10%
3	C	178	47% · 16% · 36%
4	D	10	50% · 20% · 10% · 20%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13421 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	1181	Total 9367	C 6019	N 1608	O 1685	S 55	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP Q10570
A	-1	ASN	-	expression tag	UNP Q10570
A	0	ALA	-	expression tag	UNP Q10570

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	370	Total 2978	C 1882	N 544	O 533	S 19	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	SER	-	expression tag	UNP Q9C0J8
B	-1	ASN	-	expression tag	UNP Q9C0J8
B	0	ALA	-	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	114	Total 897	C 577	N 148	O 158	S 14	0	0

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

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
4	D	8	176	79	37	52	8	0	0

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

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

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

Chain A:

F1269 L1270 V1271 S1272 R1276 M1277 L1278 M1279 V1280 Y1281 E1286 D1301 F1302 H1307 V1308 F1311 W1312 R1313 R1317 GLY ALA THR GLU GLY LEU SER LYS SER VAL V1329 N1332 K1333 H1334 I1335 F1338 A1339 T1340 L1347 L1348 L1349 P1350 M1351 Q1363 N1364 A1365 L1370

R1379 A1380 F1381 R1382 M1383 L1384 H1385 VAL ASP ARG ARG THR LEU GLN M1393 D1400 G1401 E1402 L1403 R1406 Y1409 L1418 A1441 H1442 F1443

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

Chain B: 65%   25%   10%

SER ASN ALA MET ALA THR GLU ILE GLY SER PRO PRO ARG PHE PHE HIS MET PRO ARG PHE GLN HIS GLN ALA PRO GLN LEU PHE TYR LYS ARG PRO ASP PHE ALA GLN GLN GLN ALA MET GLN GLN L41 T42 F43 R49 K50 I57 D75 Q76 R77 R80 N100

N104 A105 K117 V125 R133 L134 V135 T136 L145 W146 N147 G148 E154 T155 H160 D161 S162 P163 A166 H171 W175 M176 A179 D180 I181 G182 K186 Y187 W188 N191 M192 N193 N194 V195 K196 M197 H201 A204 I205 S211 P212 T213 D214 N215 K216

S221 D222 D223 G224 T225 V226 R227 D230 F231 L232 L240 H243 H253 P254 V259 G262 S263 Q267 Q268 P269 I270 W273 D274 P275 T283 L284 H285 V294 N297 L298 N299 W302 L303 L304 T305 C311 L314 F315 E323 L324 Q325 V326 F327 K313 C312

R328 K332 E333 V337 E344 G345 L356 L357 V361 K365 G369 H374 C391 S394 N395 D396 R406 K410

• Molecule 3: Cleavage and polyadenylation specificity factor subunit 4

Chain C: 47%   16% · 36%

M1 A6 S7 V8 D9 H10 I11 L15 L23 P27 L28 M33 ASP K35 A39 K51 C55 P56 F57 K64 V67 W71 C76 K77 K78 C82 L85 H86 K92 M93 P94 C110 P111 F112 L113 H114 I115 ASP PRO GLU SER LYS ILE LYS

ASP CYS PRO TRP TYR ASP ARG GLY PHE CYS LYS HIS GLY PRO LEU CYS ARG HIS ARG HIS THR ARG ARG VAL ILE CYS VAL ASN TYR LEU VAL GLY PHE CYS PRO GLU GLY PRO SER CYS LYS PHE MET HIS PRO ARG PHE GLU LEU PRO MET GLY THR THR GLU GLN GLN

• Molecule 4: RNA (5'-R(P*AP*AP*UP*AP*AP*AP*GP*G)-3')

Chain D: 50%   20%   10%   20%

A C A1 A2 U3 G7 G8

4   Experimental information ⓘ

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	137000	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 $(e^-/\text{Å}^2)$	80	Depositor
Minimum defocus (nm)	1100	Depositor
Maximum defocus (nm)	2200	Depositor
Magnification	47259	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	0.51	0/9581	0.59	0/13000
2	B	0.51	0/3062	0.61	1/4146 (0.0%)
3	C	0.39	0/920	0.53	1/1231 (0.1%)
4	D	0.73	0/198	0.87	0/307
All	All	0.51	0/13761	0.60	2/18684 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	92	LYS	C-N-CA	5.73	136.01	121.70
2	B	356	LEU	CA-CB-CG	5.38	127.67	115.30

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	9367	0	9420	188	0
2	B	2978	0	2895	78	0
3	C	897	0	872	25	0
4	D	176	0	88	1	0
5	C	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	13421	0	13275	281	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 11.

The worst 5 of 281 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:1365:ALA:HB1	1:A:1418:LEU:HD11	1.65	0.79
1:A:575:LEU:HD23	1:A:586:LEU:HD12	1.66	0.78
1:A:481:GLY:O	1:A:499:GLU:HB3	1.85	0.77
1:A:648:GLN:HB3	1:A:659:MET:HB3	1.68	0.76
1:A:1184:SER:HG	1:A:1193:TRP:HE1	1.27	0.76

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	1153/1446 (80%)	1010 (88%)	143 (12%)	0	100	100
2	B	368/413 (89%)	336 (91%)	32 (9%)	0	100	100
3	C	110/178 (62%)	92 (84%)	18 (16%)	0	100	100
All	All	1631/2037 (80%)	1438 (88%)	193 (12%)	0	100	100

There are no Ramachandran outliers to report.

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	1032/1237 (83%)	1031 (100%)	1 (0%)	94	98
2	B	320/357 (90%)	319 (100%)	1 (0%)	94	97
3	C	97/156 (62%)	97 (100%)	0	100	100
All	All	1449/1750 (83%)	1447 (100%)	2 (0%)	95	98

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	901	PHE
2	B	49	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	572	HIS
1	A	1237	GLN
2	B	343	HIS
1	A	315	GLN
2	B	194	ASN

5.3.3   RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	D	7/10 (70%)	2 (28%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	D	2	A
4	D	7	G

There are no RNA pucker outliers to report.

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)

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 (i)

There are no such residues in this entry.

5.8   Polymer linkage issues (i)

There are no chain breaks in this entry.