



wwPDB/EMDatabank EM Map/Model Validation Summary Report

Feb 20, 2018 – 10:17 pm GMT


PDB ID : 6EOJ
EMDB ID: : EMD-3908
Title : PolyA polymerase module of the cleavage and polyadenylation factor (CPF)
from *Saccharomyces cerevisiae*
Authors : Casanal, A.; Kumar, A.; Hill, C.H.; Emsley, P.; Passmore, L.
Deposited on : 2017-10-09
Resolution : 3.55 Å(reported)

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

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>

with specific help available everywhere you see the  symbol.

MolProbity : 4.02b-467
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30686

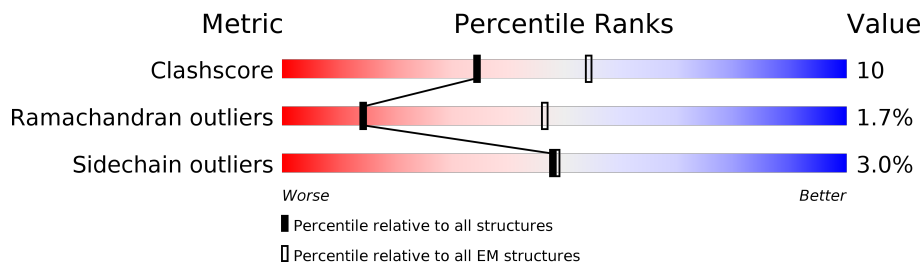
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.55 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	136279	1886
Ramachandran outliers	132675	1663
Sidechain outliers	132484	1531

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	1357	
2	B	208	
3	D	470	

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 13688 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 Protein CFT1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	1233	9801	6270	1638	1854	39	0	0

- Molecule 2 is a protein called mRNA 3'-end-processing protein YTH1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	94	788	509	140	132	7	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	PRO	-	expression tag	UNP Q06102

- Molecule 3 is a protein called Polyadenylation factor subunit 2, Polyadenylation factor subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	D	388	3097	1968	541	568	20	0	0

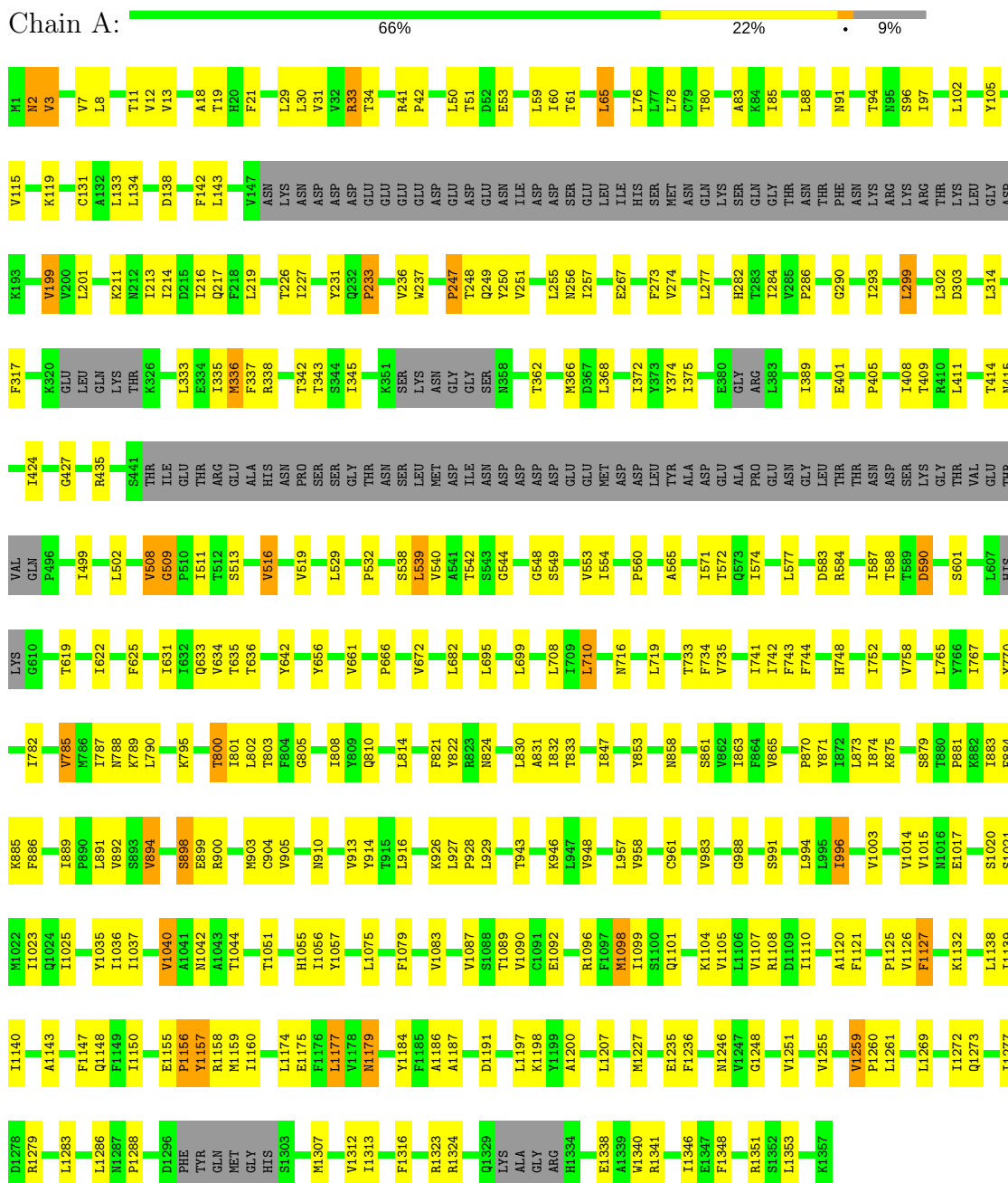
- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
4	B	2	Total	Zn	0
			2	2	

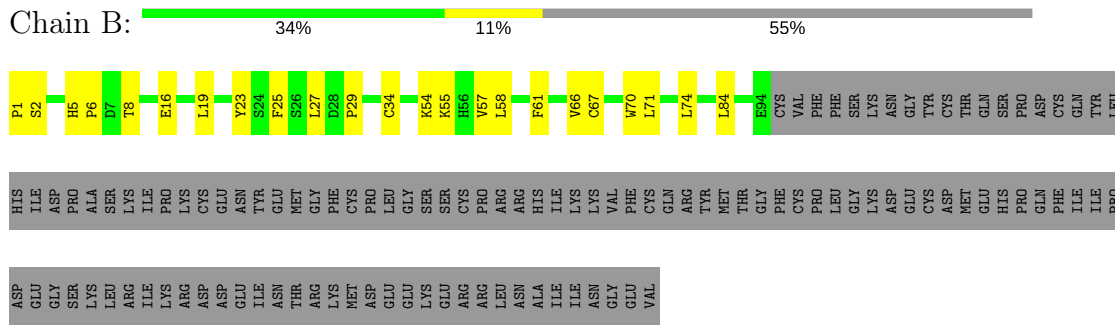
3 Residue-property plots [i](#)

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.

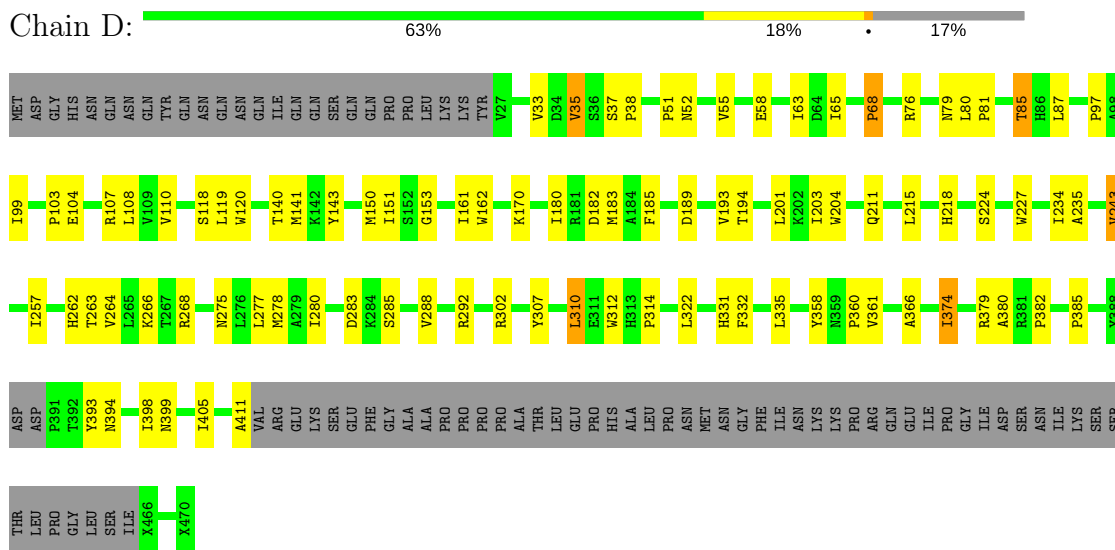
- Molecule 1: Protein CFT1



● Molecule 2: mRNA 3'-end-processing protein YTH1



● Molecule 3: Polyadenylation factor subunit 2, Polyadenylation factor subunit 2



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	77197	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 ($e^-/\text{\AA}^2$)	45	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	81000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality i

5.1 Standard geometry i

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	0.37	0/9986	0.64	1/13520 (0.0%)
2	B	0.41	0/816	0.63	0/1102
3	D	0.38	0/3156	0.61	0/4283
All	All	0.38	0/13958	0.63	1/18905 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	916	LEU	CA-CB-CG	5.14	127.12	115.30

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	9801	0	9889	206	0
2	B	788	0	773	11	0
3	D	3097	0	2986	58	0
4	B	2	0	0	0	0
All	All	13688	0	13648	261	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 261 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)
3:D:280:ILE:HD11	3:D:307:TYR:HB3	1.56	0.87
1:A:236:VAL:HG11	1:A:247:PRO:HG2	1.61	0.83
3:D:99:ILE:HG23	3:D:108:LEU:HD11	1.63	0.81
1:A:408:ILE:HG23	1:A:424:ILE:HG12	1.63	0.80
1:A:587:ILE:HG21	1:A:622:ILE:HD12	1.65	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	1215/1357 (90%)	1074 (88%)	123 (10%)	18 (2%)	11	51
2	B	92/208 (44%)	80 (87%)	9 (10%)	3 (3%)	4	34
3	D	379/470 (81%)	333 (88%)	39 (10%)	7 (2%)	9	47
All	All	1686/2035 (83%)	1487 (88%)	171 (10%)	28 (2%)	14	48

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1126	VAL
1	A	414	THR
1	A	898	SER
1	A	899	GLU
1	A	1127	PHE

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	1115/1230 (91%)	1079 (97%)	36 (3%)	42 74
2	B	89/195 (46%)	86 (97%)	3 (3%)	40 73
3	D	339/420 (81%)	332 (98%)	7 (2%)	56 82
All	All	1543/1845 (84%)	1497 (97%)	46 (3%)	48 76

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	785	VAL
1	A	894	VAL
3	D	243	VAL
1	A	800	THR
1	A	824	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	D	69	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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 2 ligands modelled in this entry, 2 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.