wwPDB EM Validation Summary Report (i)

May 4, 2024 – 12:58 pm BST

PDB ID	:	6I52
EMDB ID	:	EMD-4410
Title	:	Yeast RPA bound to ssDNA
Authors	:	Yates, L.A.; Aramayo, R.J.; Zhang, X.
Deposited on	:	2018-11-12
Resolution	:	4.70 Å(reported)

This is a wwPDB EM Validation Summary Report for a publicly released PDB 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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

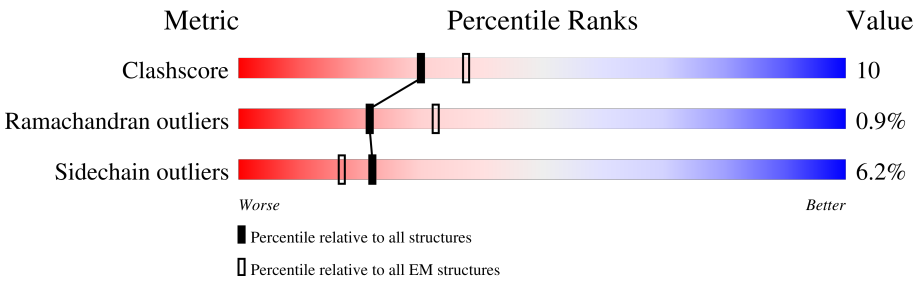
EMDB validation analysis	:	0.0.1.dev92
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 4.70 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	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	122	61% 37% •
2	B	132	67% 30% •
3	C	178	79% 21%
4	D	20	75% 25%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3856 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 Replication factor A protein 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	122	966	610	160	190	6	0	0

- Molecule 2 is a protein called Replication factor A protein 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	132	1041	661	181	195	4	0	0

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	ASN	deletion	UNP P26754
B	?	-	ASP	deletion	UNP P26754
B	?	-	LEU	deletion	UNP P26754
B	?	-	ALA	deletion	UNP P26754
B	?	-	ALA	deletion	UNP P26754
B	?	-	GLY	deletion	UNP P26754
B	?	-	ASN	deletion	UNP P26754
B	?	-	ASP	deletion	UNP P26754
B	?	-	ASP	deletion	UNP P26754
B	?	-	SER	deletion	UNP P26754
B	?	-	SER	deletion	UNP P26754
B	?	-	GLY	deletion	UNP P26754
B	?	-	LYS	deletion	UNP P26754
B	?	-	GLY	deletion	UNP P26754
B	?	-	TYR	deletion	UNP P26754
B	?	-	GLY	deletion	UNP P26754
B	?	-	SER	deletion	UNP P26754
B	?	-	GLN	deletion	UNP P26754
B	?	-	VAL	deletion	UNP P26754

- Molecule 3 is a protein called Replication factor A protein 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	178	1449	906	254	283	6	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(P*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*T)-3').

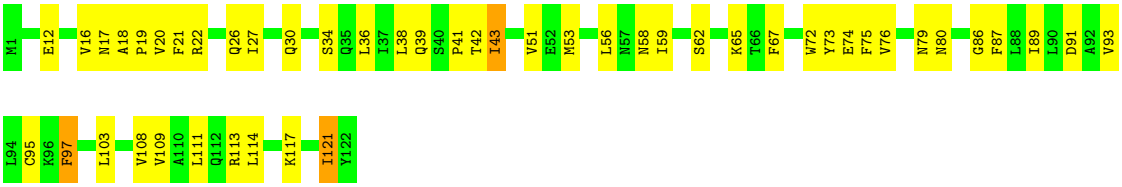
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
4	D	20	400	200	40	140	20	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

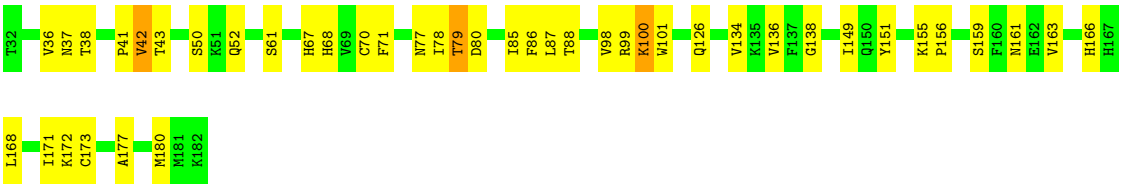
• Molecule 1: Replication factor A protein 3

Chain A:



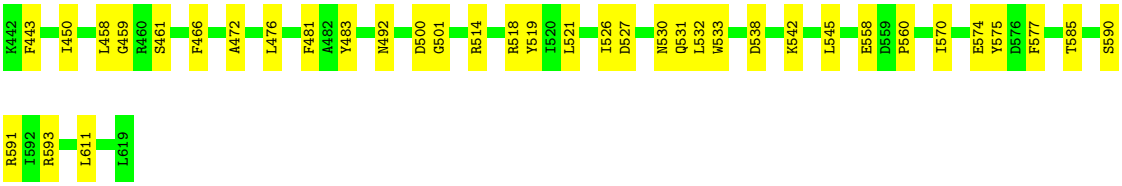
• Molecule 2: Replication factor A protein 2

Chain B:



• Molecule 3: Replication factor A protein 1

Chain C:



• Molecule 4: DNA (5'-D(P*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*T)-3')

Chain D:

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	340864	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 ($e^-/Å^2$)	2.1	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	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	$\#\|Z\| >5$	RMSZ	$\#\|Z\| >5$
1	A	0.24	0/981	0.41	0/1329
2	B	0.25	0/1060	0.45	0/1431
3	C	0.24	0/1474	0.43	0/1993
4	D	0.50	0/439	1.25	0/676
All	All	0.28	0/3954	0.60	0/5429

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	966	0	971	31	0
2	B	1041	0	1037	29	0
3	C	1449	0	1409	17	0
4	D	400	0	241	5	0
All	All	3856	0	3658	75	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 75 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)
2:B:71:PHE:O	2:B:136:VAL:HB	1.59	0.99
1:A:22:ARG:HA	1:A:75:PHE:O	1.67	0.92
2:B:41:PRO:HA	2:B:70:CYS:O	1.71	0.89
1:A:42:THR:HG22	1:A:43:ILE:HG12	1.71	0.70
1:A:108:VAL:O	1:A:111:LEU:HB3	1.96	0.66

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	120/122 (98%)	109 (91%)	9 (8%)	2 (2%)	9 43
2	B	128/132 (97%)	105 (82%)	23 (18%)	0	100 100
3	C	176/178 (99%)	157 (89%)	17 (10%)	2 (1%)	14 52
All	All	424/432 (98%)	371 (88%)	49 (12%)	4 (1%)	21 56

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	59	ILE
1	A	43	ILE
3	C	461	SER
3	C	466	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	113/113 (100%)	107 (95%)	6 (5%)	22	49
2	B	116/116 (100%)	106 (91%)	10 (9%)	10	35
3	C	157/157 (100%)	149 (95%)	8 (5%)	24	50
All	All	386/386 (100%)	362 (94%)	24 (6%)	22	45

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	126	GLN
3	C	514	ARG
3	C	443	PHE
3	C	527	ASP
2	B	38	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	531	GLN
3	C	543	GLN
3	C	571	GLN
1	A	39	GLN
2	B	52	GLN

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 monosaccharides 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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	105:ALA	C	125:ALA	N	11.94

6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-4410. These allow visual inspection of the internal detail of the map and identification of artifacts.

No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

This section was not generated.

6.2 Central slices (i)

This section was not generated.

6.3 Largest variance slices (i)

This section was not generated.

6.4 Orthogonal standard-deviation projections (False-color) (i)

This section was not generated.

6.5 Orthogonal surface views (i)

This section was not generated.

6.6 Mask visualisation (i)

This section was not generated. No masks/segmentation were deposited.

7 Map analysis (i)

This section contains the results of statistical analysis of the map.

7.1 Map-value distribution (i)

This section was not generated.

7.2 Volume estimate versus contour level (i)

This section was not generated.

7.3 Rotationally averaged power spectrum (i)

This section was not generated. The rotationally averaged power spectrum had issues being displayed.

8 Fourier-Shell correlation (i)

This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit (i)

This section was not generated.