



## wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Feb 9, 2017 – 01:52 PM EST

PDB ID : 5M1S  
EMDB ID: : EMD-4141  
Title : Cryo-EM structure of the E. coli replicative DNA polymerase-clamp-exonuclease-theta complex bound to DNA in the editing mode  
Authors : Fernandez-Leiro, R.; Conrad, J.; Scheres, S.H.W.; Lamers, M.H.  
Deposited on : 2016-10-10  
Resolution : 6.70 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](mailto: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  
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) : rb-20028442

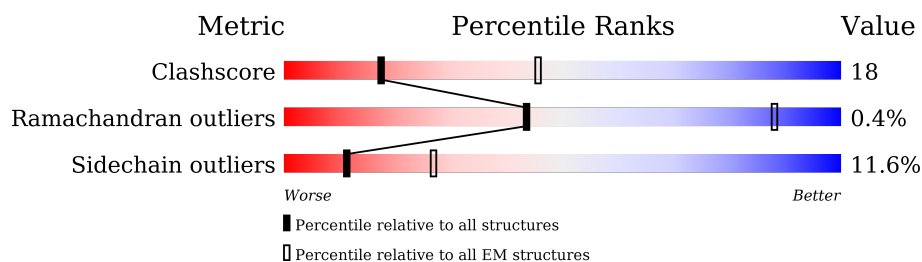
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 6.70 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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	A	927	68% 28% .
2	B	366	71% 26% .
2	C	366	60% 35% 5%
3	D	243	53% 37% 5% 6%
4	P	17	65% 35%
5	T	22	55% 45%
6	F	56	50% 45% 5%

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 16020 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 DNA polymerase III subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	927	Total	C	N	O	S	0	0
			7274	4630	1240	1363	41		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	921	LEU	ALA	engineered mutation	UNP P10443
A	923	LEU	MET	engineered mutation	UNP P10443

- Molecule 2 is a protein called DNA polymerase III subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	366	Total	C	N	O	S	0	0
			2844	1786	498	541	19		
2	C	366	Total	C	N	O	S	0	0
			2844	1786	498	541	19		

- Molecule 3 is a protein called DNA polymerase III subunit epsilon.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	229	Total	C	N	O	S	0	0
			1800	1140	316	335	9		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	183	LEU	THR	engineered mutation	UNP P03007
D	185	LEU	MET	engineered mutation	UNP P03007
D	186	PRO	ALA	engineered mutation	UNP P03007
D	187	LEU	PHE	engineered mutation	UNP P03007

- Molecule 4 is a DNA chain called DNA Primer Strand.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	P	17	Total	C	N	O	P	0	0
			354	168	69	100	17		

- Molecule 5 is a DNA chain called DNA Template Strand.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	T	22	Total	C	N	O	P	0	0
			448	214	77	135	22		

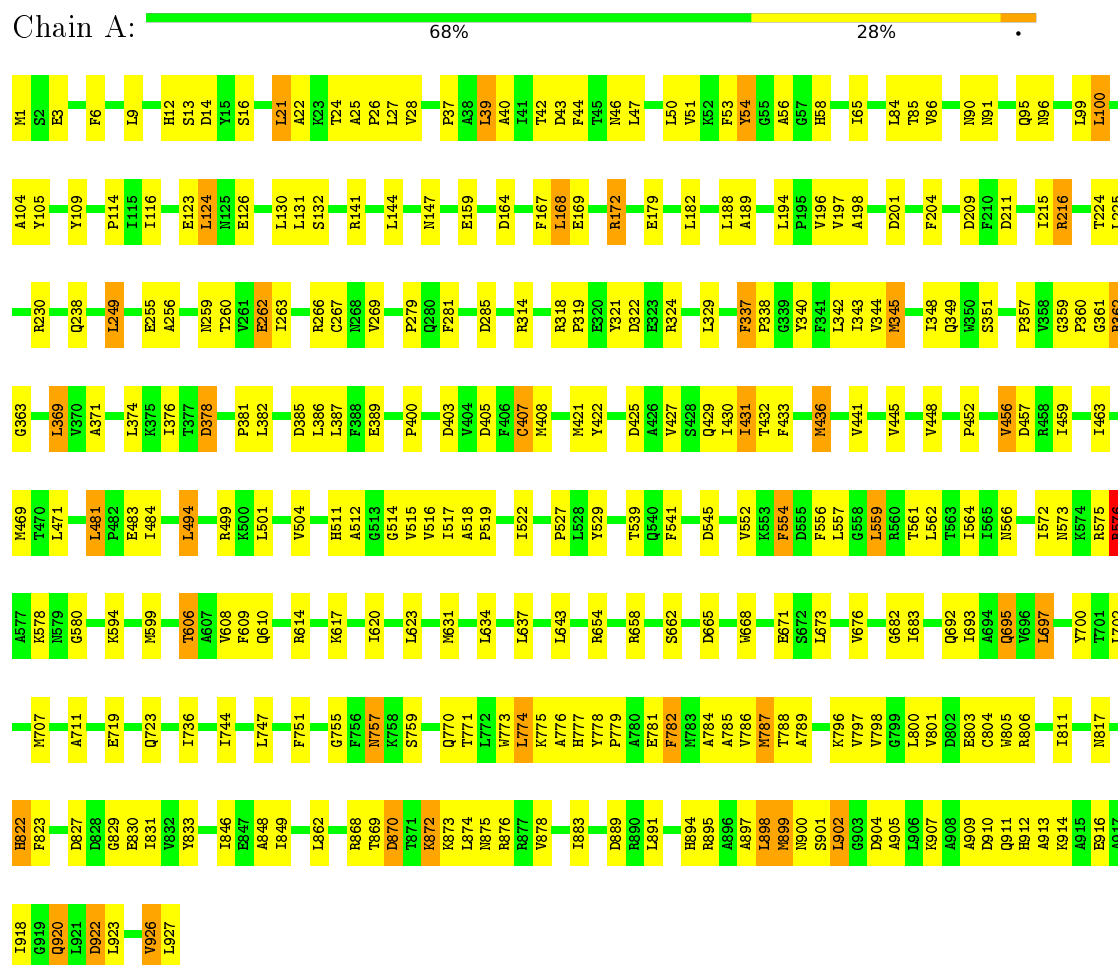
- Molecule 6 is a protein called DNA polymerase III subunit theta.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	56	Total	C	N	O	S	0	0
			456	285	86	83	2		

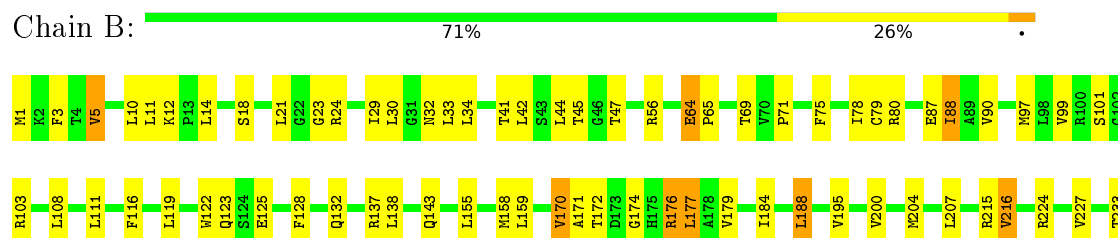
### 3 Residue-property plots

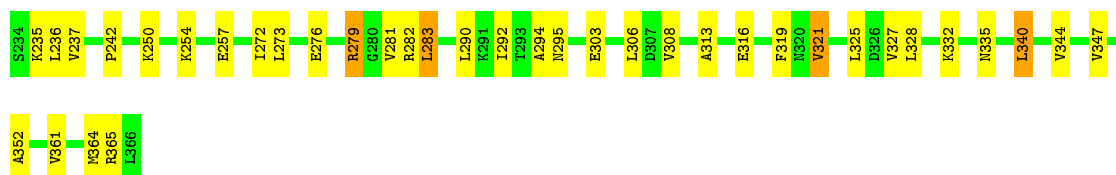
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric 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: DNA polymerase III subunit alpha



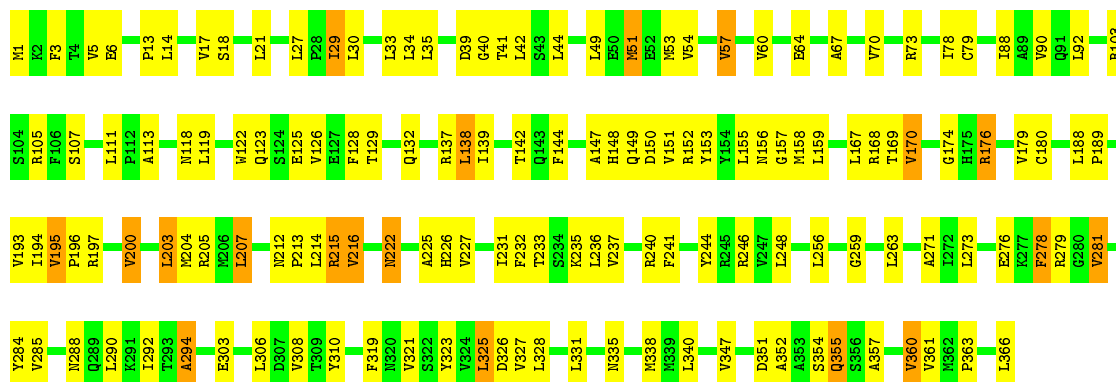
#### • Molecule 2: DNA polymerase III subunit beta





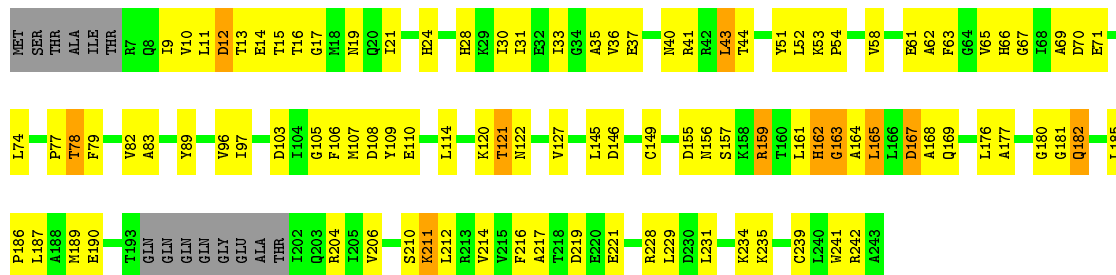
- Molecule 2: DNA polymerase III subunit beta

Chain C: 60% 35% 5%



- Molecule 3: DNA polymerase III subunit epsilon

Chain D: 53% 37% 5% 6%



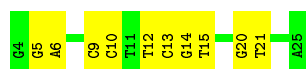
- Molecule 4: DNA Primer Strand

Chain P: 65% 35%

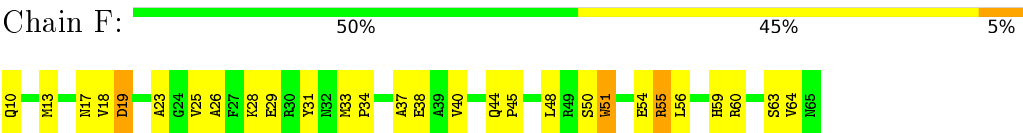


- Molecule 5: DNA Template Strand

Chain T: 55% 45%



- Molecule 6: DNA polymerase III subunit theta



## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	15616	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)	1800	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	79545	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	A	0.53	0/7427	0.81	3/10040 (0.0%)
2	B	0.48	0/2893	0.83	1/3915 (0.0%)
2	C	0.54	1/2893 (0.0%)	0.85	0/3915
3	D	0.62	0/1830	0.88	3/2469 (0.1%)
4	P	0.45	0/398	0.86	0/613
5	T	0.54	0/500	0.86	0/769
6	F	0.98	1/464 (0.2%)	1.04	2/626 (0.3%)
All	All	0.55	2/16405 (0.0%)	0.84	9/22347 (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
1	A	0	1
2	B	0	1
2	C	0	1
3	D	0	1
All	All	0	4

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	276	GLU	CD-OE1	-5.86	1.19	1.25
6	F	51	TRP	CB-CG	5.32	1.59	1.50

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	F	19	ASP	CB-CG-OD1	8.20	125.68	118.30
6	F	19	ASP	CB-CG-OD2	-7.73	111.35	118.30
2	B	215	ARG	NE-CZ-NH1	7.42	124.01	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	167	ASP	CB-CG-OD1	7.33	124.90	118.30
3	D	155	ASP	CB-CG-OD2	5.61	123.34	118.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	556	PHE	Peptide
2	B	71	PRO	Peptide
2	C	294	ALA	Peptide
3	D	163	GLY	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	A	7274	0	7219	286	0
2	B	2844	0	2861	74	0
2	C	2844	0	2861	92	0
3	D	1800	0	1802	106	0
4	P	354	0	192	8	0
5	T	448	0	250	9	0
6	F	456	0	450	20	0
All	All	16020	0	15635	563	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 18.

The worst 5 of 563 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:894:HIS:O	1:A:898:LEU:CD2	1.73	1.37
1:A:899:MET:O	1:A:902:LEU:HG	1.17	1.29
1:A:894:HIS:O	1:A:898:LEU:HD23	1.36	1.14
1:A:899:MET:HA	1:A:902:LEU:HD11	1.23	1.11
1:A:782:PHE:CE2	1:A:786:VAL:CG2	2.32	1.11

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	925/927 (100%)	862 (93%)	59 (6%)	4 (0%)	39	80
2	B	364/366 (100%)	350 (96%)	14 (4%)	0	100	100
2	C	364/366 (100%)	342 (94%)	20 (6%)	2 (0%)	34	77
3	D	225/243 (93%)	211 (94%)	12 (5%)	2 (1%)	21	67
6	F	54/56 (96%)	52 (96%)	2 (4%)	0	100	100
All	All	1932/1958 (99%)	1817 (94%)	107 (6%)	8 (0%)	43	80

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	182	GLN
3	D	186	PRO
1	A	3	GLU
1	A	576	ARG
2	C	189	PRO

#### 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	765/765 (100%)	692 (90%)	73 (10%)	11	41
2	B	313/313 (100%)	271 (87%)	42 (13%)	5	27

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	C	313/313 (100%)	263 (84%)	50 (16%)	3	21
3	D	189/200 (94%)	173 (92%)	16 (8%)	13	48
6	F	47/47 (100%)	39 (83%)	8 (17%)	2	18
All	All	1627/1638 (99%)	1438 (88%)	189 (12%)	11	32

5 of 189 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	138	LEU
2	B	308	VAL
3	D	165	LEU
2	B	176	ARG
2	B	216	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	824	HIS
1	A	912	HIS
2	C	355	GLN
1	A	900	ASN
1	A	920	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 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

There are no ligands in this entry.

## 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.