

wwPDB EM Validation Summary Report (i)

Jun 11, 2024 - 08:06 am BST

PDB ID	:	8P2I
EMDB ID	:	EMD-17366
Title	:	Cryo-EM structure of Pyrococcus furiosus apo form RNA polymerase con-
		tracted clamp conformation with Spt4/5
Authors	:	Tarau, D.M.; Reichelt, R.; Heiss, F.B.; Pilsl, M.; Hausner, W.; Engel, C.;
		Grohmann, D.
Deposited on	:	2023-05-16
Resolution	:	3.40 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 (1)) were used in the production of this report:

EMDB validation analysis	:	0.0.1. dev 92
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
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 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)	${ m EM~structures} \ (\#{ m Entries})$
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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain	
1	А	910	83%	15% •
2	В	1117	85%	13% •
3	С	397	77%	21% •
4	D	275	84%	10% 6%
5	Е	189	57% 50%	•
6	F	120	63% 37% 58%	. .
7	Н	82	78%	12% • 9%
8	K	57	86%	7% • 5%

Continued on next page...



Conti	nued from	n previous	page								
Mol	Chain	Length		Quality of chain							
9	L	95	.	16%							
10	Ν	65		129	%						
11	Р	49	-	73%			24%	•			
12	G	152	8%	16%	•	47%					
12	т	61		69%				50/			
19	1	01	51%	6		44%		5%			



2 Entry composition (i)

There are 15 unique types of molecules in this entry. The entry contains 27479 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-directed RNA polymerase subunit Rpo1N.

Mol	Chain	Residues		Α	AltConf	Trace			
1	А	901	Total 7197	C 4541	N 1282	0 1335	S 39	0	0

• Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues		Α	AltConf	Trace			
2	В	1101	Total 8825	C 5592	N 1568	O 1631	S 34	0	0

• Molecule 3 is a protein called DNA-directed RNA polymerase subunit Rpo1C.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	С	390	Total 3057	C 1931	N 528	O 588	S 10	0	0

• Molecule 4 is a protein called DNA-directed RNA polymerase subunit Rpo3.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	258	Total 2086	C 1358	N 330	0 396	$\frac{S}{2}$	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	262	TRP	-	expression tag	UNP Q8U0E4
D	263	SER	-	expression tag	UNP Q8U0E4
D	264	HIS	-	expression tag	UNP Q8U0E4
D	265	PRO	-	expression tag	UNP Q8U0E4
D	266	GLN	-	expression tag	UNP Q8U0E4
D	267	PHE	-	expression tag	UNP Q8U0E4
D	268	GLU	-	expression tag	UNP Q8U0E4
D	269	LYS	-	expression tag	UNP Q8U0E4
D	270	HIS	-	expression tag	UNP Q8U0E4

Continued on next page...



Chain	Residue	Modelled	Actual	Comment	Reference
D	271	HIS	-	expression tag	UNP Q8U0E4
D	272	HIS	-	expression tag	UNP Q8U0E4
D	273	HIS	-	expression tag	UNP Q8U0E4
D	274	HIS	-	expression tag	UNP Q8U0E4
D	275	HIS	-	expression tag	UNP Q8U0E4

Continued from previous page...

• Molecule 5 is a protein called DNA-directed RNA polymerase subunit Rpo7.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	Е	189	Total 1528	C 990	N 253	O 279	S 6	0	0

• Molecule 6 is a protein called DNA-directed RNA polymerase subunit Rpo4.

Mol	Chain	Residues	Atoms				AltConf	Trace	
6	F	116	Total 946	C 602	N 156	0 182	${ m S}{ m 6}$	0	0

• Molecule 7 is a protein called DNA-directed RNA polymerase subunit Rpo5.

Mol	Chain	Residues	Atoms			AltConf	Trace	
7	Н	75	Total 600	C 393	N 99	O 108	0	0

• Molecule 8 is a protein called DNA-directed RNA polymerase subunit Rpo6.

Mol	Chain	Residues	Atoms				AltConf	Trace	
0	K	54	Total	С	Ν	0	S	0	0
0	Γ	- 54	413	272	70	70	1	0	0

• Molecule 9 is a protein called DNA-directed RNA polymerase subunit Rpo11.

Mol	Chain	Residues	Atoms				AltConf	Trace	
9	L	94	Total 777	C 500	N 129	0 146	$\frac{S}{2}$	0	0

• Molecule 10 is a protein called DNA-directed RNA polymerase subunit Rpo10.

Mol	Chain	Residues	Atoms				AltConf	Trace	
10	Ν	65	Total 543	C 345	N 94	O 97	S 7	0	0



• Molecule 11 is a protein called DNA-directed RNA polymerase subunit Rpo12.

Mol	Chain	Residues	Atoms				AltConf	Trace	
11	D	18	Total	С	Ν	Ο	\mathbf{S}	0	0
	I	40	393	251	75	63	4	0	0

• Molecule 12 is a protein called Transcription elongation factor Spt5.

Mol	Chain	Residues		Ato	ms		AltConf	Trace
12	G	81	Total 648	C 422	N 114	O 112	0	0

• Molecule 13 is a protein called Transcription elongation factor Spt4.

Mol	Chain	Residues	Atoms				AltConf	Trace	
13	Ι	58	Total 459	C 288	N 81	O 86	$\begin{array}{c} \mathrm{S} \\ \mathrm{4} \end{array}$	0	0

• Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	AltConf
14	А	2	Total Zn 2 2	0
14	В	1	Total Zn 1 1	0
14	Ν	1	Total Zn 1 1	0
14	Р	1	Total Zn 1 1	0
14	Ι	1	Total Zn 1 1	0

• Molecule 15 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	AltConf
15	А	1	Total Mg 1 1	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 and atom inclusion in map density. 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. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). 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-directed RNA polymerase subunit Rpo1N









MET

H10 Y11 T12

R7 H8 C18 P19 V20 C21 <mark>V60</mark> Arg

157

(46



130

F31 D32 /34

G22 S23 R24 D25 L26 S27 E28 E28



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	107970	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE	Depositor
	CORRECTION	
Microscope	JEOL CRYO ARM 200	Depositor
Voltage (kV)	200	Depositor
Electron dose $(e^-/\text{\AA}^2)$	1	Depositor
Minimum defocus (nm)	600	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.028	Depositor
Minimum map value	-0.009	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.001	Depositor
Recommended contour level	0.007	Depositor
Map size (Å)	251.68, 251.68, 251.68	wwPDB
Map dimensions	260, 260, 260	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.968, 0.968, 0.968	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: MG, 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).

Mal	Chain	Bond	lengths	Bo	ond angles
MOI	Ullalli	RMSZ	# Z > 5	RMSZ	# Z > 5
1	А	0.25	0/7338	0.51	0/9907
2	В	0.25	0/9003	0.51	0/12166
3	С	0.24	0/3098	0.50	0/4180
4	D	0.26	0/2133	0.45	0/2894
5	Е	0.27	0/1556	0.57	1/2091~(0.0%)
6	F	0.28	0/959	0.65	0/1284
7	Н	0.29	0/613	0.54	0/829
8	Κ	0.29	0/421	0.54	0/572
9	L	0.25	0/792	0.51	0/1067
10	Ν	0.29	0/553	0.56	0/740
11	Р	0.27	0/400	0.62	0/534
12	G	0.27	0/663	0.57	0/897
13	Ι	0.26	0/467	0.54	0/629
All	All	0.25	0/27996	0.52	1/37790~(0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
5	Ε	93	MET	CA-CB-CG	6.39	124.17	113.30

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



Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	А	7197	0	7263	91	0
2	В	8825	0	8904	101	0
3	С	3057	0	3166	62	0
4	D	2086	0	2109	19	0
5	Е	1528	0	1593	85	0
6	F	946	0	961	69	0
7	Н	600	0	628	9	0
8	Κ	413	0	451	4	0
9	L	777	0	788	10	0
10	Ν	543	0	541	5	0
11	Р	393	0	423	9	0
12	G	648	0	670	23	0
13	Ι	459	0	453	19	0
14	А	2	0	0	0	0
14	В	1	0	0	0	0
14	Ι	1	0	0	0	0
14	Ν	1	0	0	0	0
14	Р	1	0	0	0	0
15	А	1	0	0	0	0
All	All	27479	0	27950	455	0

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.

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 8.

The worst 5 of 455 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:244:VAL:HG11	2:B:254:MET:HG3	1.59	0.84
5:E:145:ARG:HH12	6:F:97:ALA:HB3	1.46	0.81
1:A:620:ASP:OD2	1:A:621:LYS:N	2.13	0.80
2:B:654:TYR:HB3	2:B:657:HIS:HD2	1.47	0.79
2:B:296:ASP:OD1	2:B:310:ARG:NH2	2.17	0.77

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	Perce	ntiles
1	А	899/910~(99%)	878 (98%)	21 (2%)	0	100	100
2	В	1097/1117 (98%)	1070 (98%)	27 (2%)	0	100	100
3	С	388/397~(98%)	374 (96%)	14 (4%)	0	100	100
4	D	256/275~(93%)	255 (100%)	1 (0%)	0	100	100
5	Е	187/189~(99%)	175 (94%)	12 (6%)	0	100	100
6	F	114/120~(95%)	110 (96%)	4 (4%)	0	100	100
7	Н	73/82~(89%)	72 (99%)	1 (1%)	0	100	100
8	К	52/57~(91%)	52 (100%)	0	0	100	100
9	L	92/95~(97%)	89 (97%)	3 (3%)	0	100	100
10	Ν	63/65~(97%)	63 (100%)	0	0	100	100
11	Р	46/49~(94%)	44 (96%)	2 (4%)	0	100	100
12	G	79/152~(52%)	77 (98%)	2 (2%)	0	100	100
13	Ι	56/61~(92%)	54 (96%)	2 (4%)	0	100	100
All	All	3402/3569~(95%)	3313 (97%)	89 (3%)	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 side chain 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	А	777/785~(99%)	765~(98%)	12 (2%)	65 82

Continued on next page...



Mol	Chain	Analysed	Rotameric	Outliers	Perce	entiles
2	В	949/962~(99%)	936~(99%)	13 (1%)	67	83
3	С	338/345~(98%)	334 (99%)	4 (1%)	71	85
4	D	230/246~(94%)	230 (100%)	0	100	100
5	Е	167/167~(100%)	158 (95%)	9~(5%)	22	52
6	F	99/103~(96%)	93~(94%)	6 (6%)	18	48
7	Н	65/70~(93%)	64 (98%)	1 (2%)	65	82
8	K	44/47~(94%)	43 (98%)	1 (2%)	50	74
9	L	83/84~(99%)	80~(96%)	3~(4%)	35	63
10	Ν	60/60~(100%)	60 (100%)	0	100	100
11	Р	44/45~(98%)	43 (98%)	1 (2%)	50	74
12	G	68/128~(53%)	64 (94%)	4 (6%)	19	49
13	Ι	51/54~(94%)	50 (98%)	1 (2%)	55	77
All	All	2975/3096~(96%)	2920 (98%)	55 (2%)	61	79

Continued from previous page...

 $5~{\rm of}~55$ residues with a non-rotameric side chain are listed below:

Mol	Chain	\mathbf{Res}	Type
3	С	264	ARG
5	Е	134	TYR
13	Ι	11	TYR
11	Р	13	ARG
5	Е	17	MET

Sometimes side chains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such side chains are listed below:

Mol	Chain	Res	Type
2	В	680	ASN
2	В	1037	HIS
2	В	1061	ASN
3	С	390	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)

Of 7 ligands modelled in this entry, 7 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.



6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-17366. These allow visual inspection of the internal detail of the map and identification of artifacts.

Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections (i)

6.1.1 Primary map



6.1.2 Raw map



The images above show the map projected in three orthogonal directions.



6.2 Central slices (i)

6.2.1 Primary map



X Index: 130





Z Index: 130

6.2.2 Raw map



X Index: 130

Y Index: 130



The images above show central slices of the map in three orthogonal directions.



6.3 Largest variance slices (i)

6.3.1 Primary map



X Index: 115



Y Index: 119



Z Index: 141

6.3.2 Raw map



X Index: 115

Y Index: 121



The images above show the largest variance slices of the map in three orthogonal directions.



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

6.4.1 Primary map



6.4.2 Raw map



The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.



6.5 Orthogonal surface views (i)

6.5.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.007. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.5.2 Raw map



These images show the 3D surface of the raw map. The raw map's contour level was selected so that its surface encloses the same volume as the primary map does at its recommended contour level.



Mask visualisation (i) 6.6

This section shows the 3D surface view of the primary map at 50% transparency overlaid with the specified mask at 0% transparency

A mask typically either:

- Encompasses the whole structure
- Separates out a domain, a functional unit, a monomer or an area of interest from a larger structure

$emd_{17366}msk_{1.map}$ (i) 6.6.1





7 Map analysis (i)

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

7.1 Map-value distribution (i)



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.



7.2 Volume estimate (i)



The volume at the recommended contour level is 174 $\rm nm^3;$ this corresponds to an approximate mass of 157 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.



7.3 Rotationally averaged power spectrum (i)



*Reported resolution corresponds to spatial frequency of 0.294 \AA^{-1}



8 Fourier-Shell correlation (i)

Fourier-Shell Correlation (FSC) is the most commonly used method to estimate the resolution of single-particle and subtomogram-averaged maps. The shape of the curve depends on the imposed symmetry, mask and whether or not the two 3D reconstructions used were processed from a common reference. The reported resolution is shown as a black line. A curve is displayed for the half-bit criterion in addition to lines showing the 0.143 gold standard cut-off and 0.5 cut-off.

8.1 FSC (i)



*Reported resolution corresponds to spatial frequency of 0.294 $\mathrm{\AA^{-1}}$



8.2 Resolution estimates (i)

$\mathbf{Bosolution ostimato}(\mathbf{\hat{A}})$	Estimation criterion (FSC cut-off)			
Resolution estimate (A)	0.143	0.5	Half-bit	
Reported by author	3.40	-	-	
Author-provided FSC curve	-	-	-	
Unmasked-calculated*	3.68	4.32	3.80	

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.



9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-17366 and PDB model 8P2I. Per-residue inclusion information can be found in section 3 on page 7.

9.1 Map-model overlay (i)



The images above show the 3D surface view of the map at the recommended contour level 0.007 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.



9.2 Q-score mapped to coordinate model (i)



The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

9.3 Atom inclusion mapped to coordinate model (i)



The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.007).



9.4 Atom inclusion (i)



At the recommended contour level, 89% of all backbone atoms, 74% of all non-hydrogen atoms, are inside the map.



1.0

0.0 <0.0

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (0.007) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7400	0.4610
А	0.8260	0.4960
В	0.8180	0.4900
С	0.6390	0.4100
D	0.8170	0.5050
Ε	0.3380	0.2800
F	0.2920	0.2590
G	0.6030	0.4140
Н	0.8150	0.5050
Ι	0.2450	0.3000
Κ	0.8490	0.5130
L	0.7880	0.5020
Ν	0.8720	0.5200
Р	0.7320	0.4840

