



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 12:01 PM GMT

PDB ID : 3QQC  
Title : Crystal structure of archaeal Spt4/5 bound to the RNAP clamp domain  
Authors : Martinez-Rucobo, F.W.  
Deposited on : 2011-02-15  
Resolution : 3.30 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB 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/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

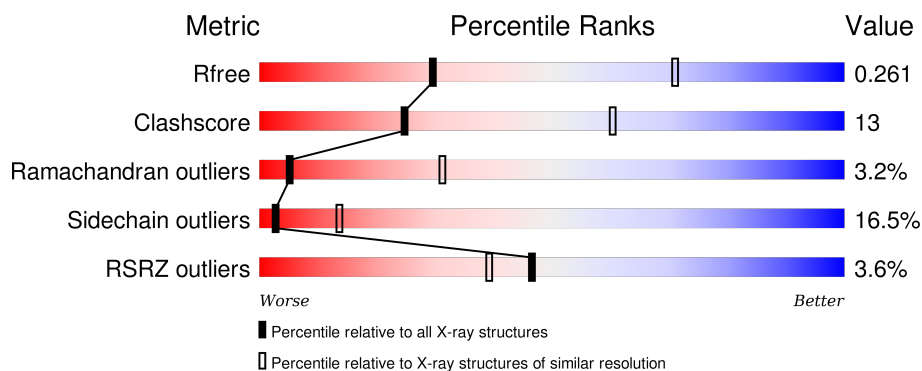
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	2060 (3.40-3.20)
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower 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  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	436	<div> <div>2%</div> <div> <div></div> <div>50%</div> <div>29%</div> <div>8%</div> <div>13%</div> </div> </div>
2	D	163	<div> <div>2%</div> <div> <div></div> <div>34%</div> <div>14%</div> <div>•</div> <div>50%</div> </div> </div>
3	E	61	<div> <div>11%</div> <div> <div></div> <div>72%</div> <div>21%</div> <div>•</div> <div>5%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4187 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, 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 b, DNA-directed RNA polymerase subunit a', DNA-directed RNA polymerase subunit A".

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	381	Total	C	N	O	S	0	0	0
			3067	1947	542	559	19			

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP Q8U0M3
A	65	GLY	-	LINKER	UNP Q8U0M3
A	66	GLY	-	LINKER	UNP Q8U0M3
A	391	GLY	-	LINKER	UNP Q8U0M4
A	392	ALA	-	LINKER	UNP Q8U0M4
A	393	GLY	-	LINKER	UNP Q8U0M4
A	394	SER	-	LINKER	UNP Q8U0M4
A	395	GLY	-	LINKER	UNP Q8U0M4
A	396	ALA	-	LINKER	UNP Q8U0M4
A	397	GLY	-	LINKER	UNP Q8U0M4
A	398	SER	-	LINKER	UNP Q8U0M4
A	399	GLY	-	LINKER	UNP Q8U0M4
A	438	LYS	-	EXPRESSION TAG	UNP Q8U0M5
A	439	HIS	-	EXPRESSION TAG	UNP Q8U0M5
A	440	HIS	-	EXPRESSION TAG	UNP Q8U0M5
A	441	HIS	-	EXPRESSION TAG	UNP Q8U0M5
A	442	HIS	-	EXPRESSION TAG	UNP Q8U0M5
A	443	HIS	-	EXPRESSION TAG	UNP Q8U0M5
A	444	HIS	-	EXPRESSION TAG	UNP Q8U0M5

- Molecule 2 is a protein called Transcription antitermination protein nusG.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	82	Total	C	N	O	0	0	0
			657	427	115	115			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	153	ALA	-	EXPRESSION TAG	UNP Q8TZK1
D	154	ALA	-	EXPRESSION TAG	UNP Q8TZK1
D	155	ALA	-	EXPRESSION TAG	UNP Q8TZK1
D	156	LEU	-	EXPRESSION TAG	UNP Q8TZK1
D	157	GLU	-	EXPRESSION TAG	UNP Q8TZK1
D	158	HIS	-	EXPRESSION TAG	UNP Q8TZK1
D	159	HIS	-	EXPRESSION TAG	UNP Q8TZK1
D	160	HIS	-	EXPRESSION TAG	UNP Q8TZK1
D	161	HIS	-	EXPRESSION TAG	UNP Q8TZK1
D	162	HIS	-	EXPRESSION TAG	UNP Q8TZK1
D	163	HIS	-	EXPRESSION TAG	UNP Q8TZK1

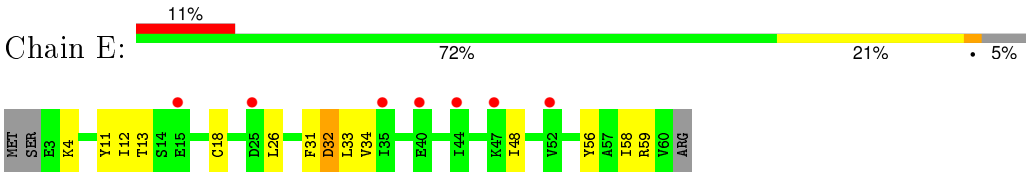
- Molecule 3 is a protein called DNA-directed RNA polymerase, subunit e”.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	58	Total	C	N	O	S	0	0	0
			459	288	81	86	4			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total	Zn	0	0
			3	3		
4	E	1	Total	Zn	0	0
			1	1		





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	168.57Å 107.04Å 51.09Å 90.00° 97.10° 90.00°	Depositor
Resolution (Å)	50.70 – 3.30 50.70 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.8 (50.70-3.30) 99.7 (50.70-3.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.14	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.09 (at 3.33Å)	Xtriage
Refinement program	BUSTER 2.9.2	Depositor
R, $R_{free}$	0.193 , 0.247 0.210 , 0.261	Depositor DCC
$R_{free}$ test set	674 reflections (5.22%)	DCC
Wilson B-factor (Å <sup>2</sup> )	106.3	Xtriage
Anisotropy	0.305	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 107.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtriage
Outliers	0 of 13602 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	4187	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	131.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.45% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 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	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.54	0/3122	0.87	4/4206 (0.1%)
2	D	0.48	0/672	0.77	0/909
3	E	0.44	0/467	0.64	0/629
All	All	0.52	0/4261	0.83	4/5744 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	221	GLU	C-N-CA	6.86	138.84	121.70
1	A	221	GLU	N-CA-C	5.78	126.61	111.00
1	A	427	ASN	N-CA-C	5.71	126.41	111.00
1	A	222	LYS	N-CA-C	5.63	126.21	111.00

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	3067	0	3133	94	0
2	D	657	0	676	12	0
3	E	459	0	453	7	0
4	A	3	0	0	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	1	0	0	0	0
All	All	4187	0	4262	108	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 13.

The worst 5 of 108 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:D:34:TYR:HB2	2:D:49:GLU:HG3	1.38	1.00
1:A:147:VAL:HG21	1:A:341:THR:HA	1.59	0.82
1:A:54:MET:HG3	1:A:369:LEU:HD21	1.64	0.78
1:A:156:ILE:O	1:A:160:THR:HG22	1.85	0.76
1:A:323:GLY:HA3	2:D:82:PHE:CZ	2.22	0.75

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 X-ray entries followed by that with respect to entries of similar resolution.

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	367/436 (84%)	319 (87%)	36 (10%)	12 (3%)	5	30
2	D	80/163 (49%)	69 (86%)	8 (10%)	3 (4%)	4	26
3	E	56/61 (92%)	48 (86%)	7 (12%)	1 (2%)	11	46
All	All	503/660 (76%)	436 (87%)	51 (10%)	16 (3%)	5	31

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	109	ASP
1	A	123	THR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	357	LYS
1	A	358	SER
2	D	26	VAL

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

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	340/375 (91%)	275 (81%)	65 (19%)	2	8
2	D	69/136 (51%)	60 (87%)	9 (13%)	5	22
3	E	51/54 (94%)	49 (96%)	2 (4%)	39	75
All	All	460/565 (81%)	384 (84%)	76 (16%)	3	14

5 of 76 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	219	LYS
1	A	249	ASP
2	D	56	VAL
1	A	221	GLU
1	A	232	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	146	HIS
1	A	367	GLN
1	A	337	GLN
1	A	137	HIS
1	A	313	ASN

### 5.3.3 RNA ⓘ

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 4 ligands modelled in this entry, 4 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 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	381/436 (87%)	0.21	8 (2%) 67 60	68, 108, 175, 203	0
2	D	82/163 (50%)	0.28	4 (4%) 33 27	110, 157, 210, 227	0
3	E	58/61 (95%)	0.72	7 (12%) 6 4	155, 190, 219, 225	1 (1%)
All	All	521/660 (78%)	0.28	19 (3%) 46 39	68, 121, 201, 227	1 (0%)

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	E	40	GLU	5.2
2	D	23	TYR	4.7
1	A	296	ILE	4.4
1	A	291	THR	3.8
3	E	25	ASP	3.3

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	ZN	A	2002	1/1	0.99	0.18	-0.30	124,124,124,124	0
4	ZN	A	2001	1/1	0.97	0.18	-1.00	131,131,131,131	0
4	ZN	A	2000	1/1	0.89	0.10	-1.34	180,180,180,180	0
4	ZN	E	100	1/1	0.95	0.04	-3.34	261,261,261,261	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.