wwPDB X-ray Structure Validation Summary Report (i)

Sep 2, 2023 – 10:35 AM EDT

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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
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:

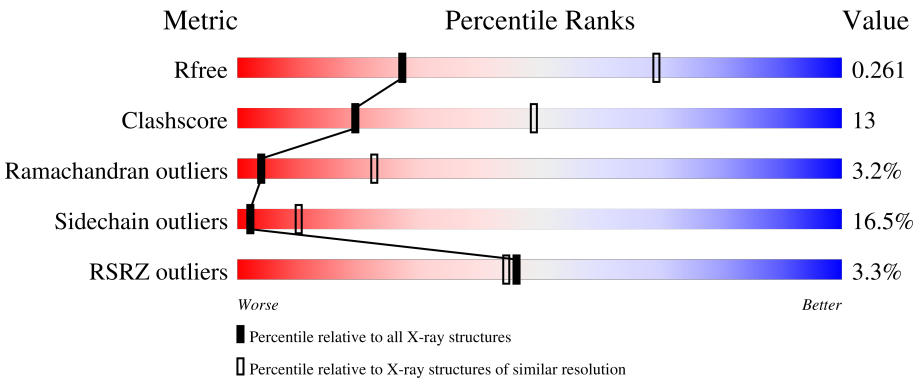
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

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}	130704	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 of the lower 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 electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	436	2%, 50%, 29%, 8%, 13%
2	D	163	2%, 34%, 14%, •, 50%
3	E	61	10%, 72%, 21%, •, 5%

2 Entry composition (i)

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
			Total	C	N	O	S			
1	A	381	3067	1947	542	559	19	0	0	0

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
			Total	C	N	O			
2	D	82	657	427	115	115	0	0	0

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
			Total	C	N	O	S			
3	E	58	459	288	81	86	4	0	0	0

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

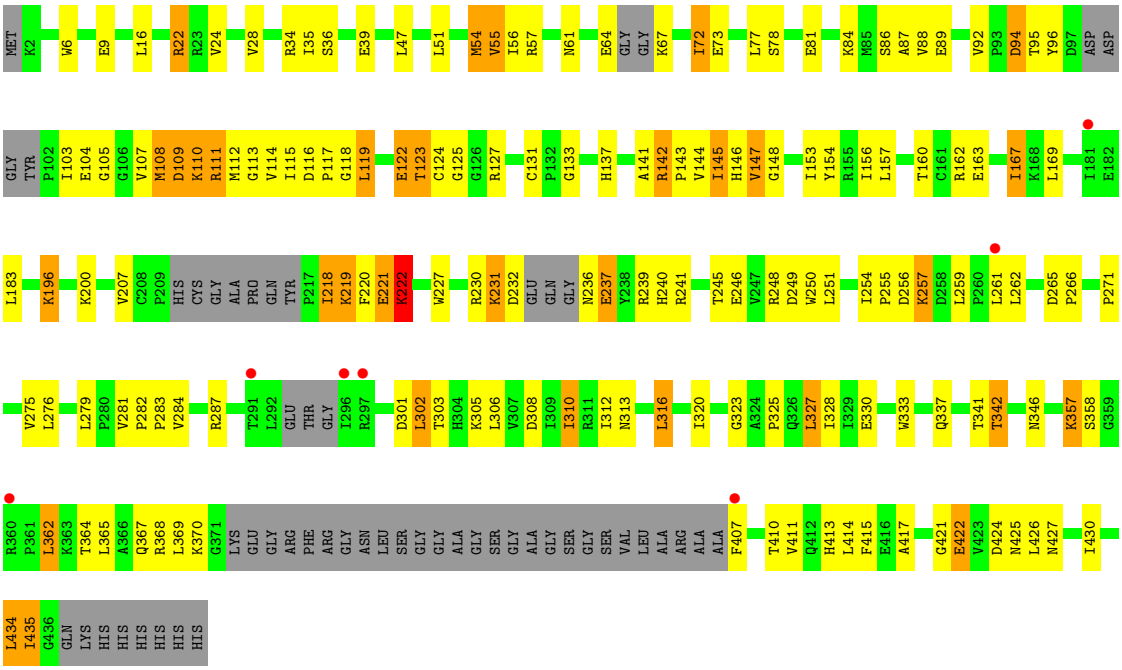
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
4	A	3	3	3	0	0
			Total	Zn		
4	E	1	1	1	0	0

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 and electron 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 dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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.

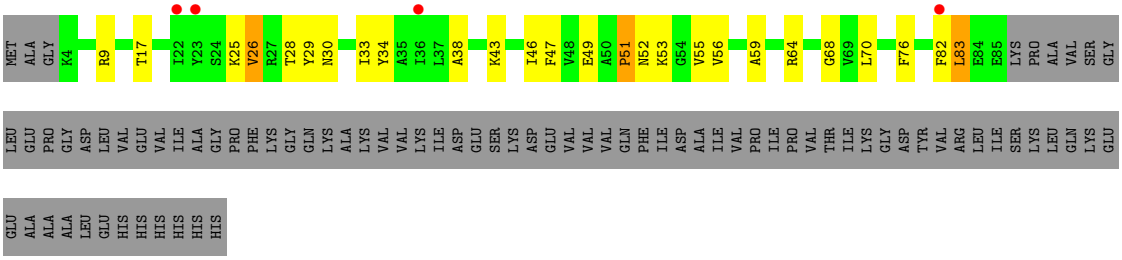
• Molecule 1: DNA-directed RNA polymerase subunit b, DNA-directed RNA polymerase subunit a', DNA-directed RNA polymerase subunit A"

Chain A:



• Molecule 2: Transcription antitermination protein nusG

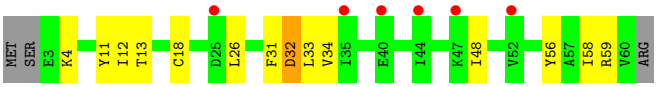
Chain D:



• Molecule 3: DNA-directed RNA polymerase, subunit e"

Chain E:

10%
72%
21%
5%


MET
SER
E3
K4
Y11
I12
T13
C18
D25
L26
F31
D32
L33
V34
I35
E40
I44
K47
I48
V52
Y56
A57
I58
R59
V60
ARG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	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
$< I/\sigma(I) >$ [1]	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 (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	106.3	Xtriage
Anisotropy	0.305	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 107.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	4187	wwPDB-VP
Average B, all atoms (Å^2)	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.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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(^{o})	Ideal(^{o})
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%)	4 22
2	D	80/163 (49%)	69 (86%)	8 (10%)	3 (4%)	3 19
3	E	56/61 (92%)	48 (86%)	7 (12%)	1 (2%)	8 35
All	All	503/660 (76%)	436 (87%)	51 (10%)	16 (3%)	4 22

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 (i)

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%)	1 6
2	D	69/136 (51%)	60 (87%)	9 (13%)	4 17
3	E	51/54 (94%)	49 (96%)	2 (4%)	32 62
All	All	460/565 (81%)	384 (84%)	76 (16%)	2 10

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

Mol	Chain	Res	Type
1	A	411	VAL
2	D	64	ARG
1	A	414	LEU
2	D	17	THR
3	E	34	VAL

Sometimes 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	367	GLN
1	A	339	HIS
1	A	313	ASN
1	A	146	HIS
1	A	337	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 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, 95th 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^2)	Q<0.9
1	A	381/436 (87%)	0.19	7 (1%)	68	67	68, 108, 175, 203	0
2	D	82/163 (50%)	0.25	4 (4%)	29	27	110, 157, 210, 227	0
3	E	58/61 (95%)	0.69	6 (10%)	6	6	155, 190, 219, 224	1 (1%)
All	All	521/660 (78%)	0.25	17 (3%)	46	44	68, 121, 201, 227	1 (0%)

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	E	40	GLU	5.1
2	D	23	TYR	4.5
1	A	296	ILE	4.3
1	A	291	THR	3.6
3	E	25	ASP	3.2

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 monosaccharides 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. The B-factors column lists the minimum, median, 95th 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	B-factors(Å^2)	Q<0.9
4	ZN	A	2000	1/1	0.89	0.10	180,180,180,180	0
4	ZN	E	100	1/1	0.95	0.04	261,261,261,261	0
4	ZN	A	2001	1/1	0.97	0.18	131,131,131,131	0
4	ZN	A	2002	1/1	0.99	0.18	124,124,124,124	0

6.5 Other polymers (i)

There are no such residues in this entry.