



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 11, 2018 – 05:50 PM EST

PDB ID : 6BOG
Title : Crystal structure of RapA, a Swi2/Snf2 protein that recycles RNA polymerase during transcription
Authors : Shaw, G.X.; Gan, J.; Zhou, Y.N.; Zhang, R.; Joachimiak, A.; Jin, D.J.; Ji, X.
Deposited on : 2017-11-20
Resolution : 3.21 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

<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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030736
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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) : rb-20030736

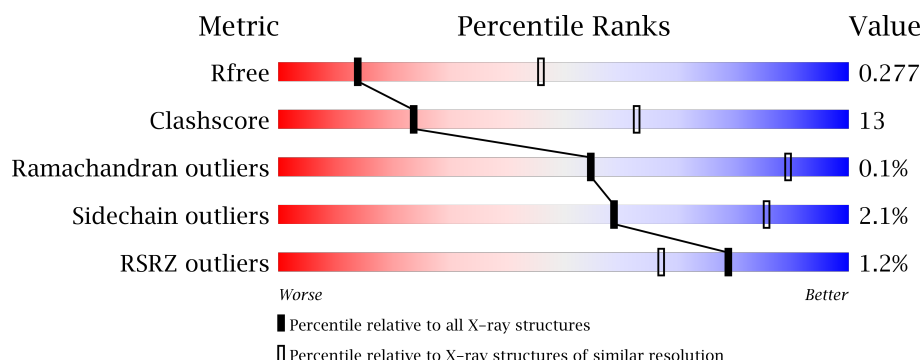
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.21 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}	100719	1015 (3.22-3.18)
Clashscore	112137	1009 (3.20-3.20)
Ramachandran outliers	110173	1118 (3.22-3.18)
Sidechain outliers	110143	1117 (3.22-3.18)
RSRZ outliers	101464	1020 (3.22-3.18)

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 ≥ 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	968	<div> <div></div> <div>70%</div> <div>29%</div> <div>.</div> </div>
1	B	968	<div> <div></div> <div>73%</div> <div>26%</div> <div>.</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	SO4	A	1003	-	-	X	-
2	SO4	B	1002	-	-	X	-
2	SO4	B	1003	-	-	X	-

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 15473 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 RNA polymerase-associated protein RapA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	967	Total	C	N	O	S	Se	0	0	0
			7719	4831	1382	1477	6	23			
1	B	967	Total	C	N	O	S	Se	0	0	0
			7719	4831	1382	1477	6	23			

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		

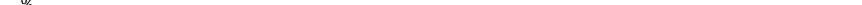
Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		

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

- [illegible]

- Chain B: 

V927	Y83	L205	R350	W484	L608	R886	V927
N928	L84	L206	R354	W485	L613	R807	N928
R932	T86	V207	D355	M486	B614	A808	R932
D933	T87	W215	A356	P489	K615	L809	D933
D934	L88	R221	V357	W493	T616	T813	D934
E935	D89	R222	R358	W493	Y625	L814	E935
I939	T90	F223	L359	R505	H626	L815	I939
V946	E92	D232	L360	A511	B627	L819	V946
N947	V95	F246	L370	A514	B640	L830	N947
D951	A96	L256	L386	A514	V646	R834	D951
R956	L97	D257	D392	Q518	V647	P838	R956
L957	K105	L260	R400	V522	L650	R842	L957
D958	L106	R261	V404	R526	B671	D846	D958
A959	V107	S262	L407	R533	R685	N850	A959
L962	F108	R263	M408	H535	D686	D851	L962
I963	S109	E267	R410	M538	R687	L852	I963
V965	K110	C270	G412	S539	L688	V856	V965
T966	Q112	E271	R415	E542	L689	E857	T966
H967	D113	A272	W274	R543	B690	F858	H967
Q968	F116	D275	N420	D544	D711	E859	Q968
		V278	H434	F550	I716	T860	
		A132	T435	C562	M720	R863	
		H283	L436	S563	N721	N866	
		H305	L437	E564	I725	A867	
		V309	P441	I565	I726	R870	
		L310	Y444	E568	I727	G873	
		L311	Q445	G569	I728	N729	
		L312	R449	R570	Q730	V877	
		T313	M578	M578	M737	E896	
		A314	R457	D582	I738	R900	
		E317	K458	L583	T741	I903	
		Q318	E461	F584	D744	D904	
		H324	D462	F585	H745	R907	
		K183	R463	N586	M746	D911	
		T184	P687	P687	R765	E912	
		L185	D588	D588	R771	K913	
		M189	L589	L589	E772	L918	
		I190	Q592	Q592	I788	L924	
		Q193	R593	R593	T801	R925	
		R202	D480	D597	R805	A926	
			N481	D598			
			A482	R599			
			T483				

4 Data and refinement statistics

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α , β , γ	123.86Å 123.86Å 187.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.19 – 3.21 29.19 – 3.20	Depositor EDS
% Data completeness (in resolution range)	98.9 (29.19-3.21) 98.9 (29.19-3.20)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.38 (at 3.18Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.244 , 0.279 0.244 , 0.277	Depositor DCC
R_{free} test set	995 reflections (2.17%)	DCC
Wilson B-factor (Å ²)	91.6	Xtriage
Anisotropy	0.222	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 38.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	0.449 for h,-k,-l	Xtriage
Reported twinning fraction	0.500 for h,-k,-l	Depositor
Outliers	0 of 45884 reflections	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	15473	wwPDB-VP
Average B, all atoms (Å ²)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.02% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.31	0/7840	0.53	0/10581
1	B	0.29	0/7840	0.53	0/10581
All	All	0.30	0/15680	0.53	0/21162

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

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	7719	0	7583	208	0
1	B	7719	0	7583	183	0
2	A	15	0	0	3	0
2	B	20	0	0	5	0
All	All	15473	0	15166	386	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 386 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:B:533:VAL:O	1:B:538:MSE:HE1	1.35	1.22
1:B:737:MSE:HE1	1:B:771:ARG:NH1	1.90	0.85
1:A:540:ILE:HD13	1:A:543:ARG:NH2	1.95	0.81
1:B:70:GLN:HB3	1:B:86:THR:HB	1.62	0.80
1:A:540:ILE:CD1	1:A:543:ARG:NH2	2.47	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 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	965/968 (100%)	930 (96%)	35 (4%)	0	100	100
1	B	965/968 (100%)	928 (96%)	36 (4%)	1 (0%)	55	89
All	All	1930/1936 (100%)	1858 (96%)	71 (4%)	1 (0%)	55	89

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	564	GLU

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	828/805 (103%)	810 (98%)	18 (2%)	57	85

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	828/805 (103%)	812 (98%)	16 (2%)	62	86
All	All	1656/1610 (103%)	1622 (98%)	34 (2%)	59	85

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	896	GLU
1	B	19	LEU
1	B	896	GLU
1	A	934	ASP
1	A	410	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	694	ASN
1	B	475	GLN
1	A	942	ASN
1	A	193	GLN
1	B	163	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

7 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and

the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	SO4	A	1001	-	4,4,4	0.15	0	6,6,6	0.17	0
2	SO4	A	1002	-	4,4,4	0.15	0	6,6,6	0.13	0
2	SO4	A	1003	-	4,4,4	0.15	0	6,6,6	0.14	0
2	SO4	B	1001	-	4,4,4	0.14	0	6,6,6	0.12	0
2	SO4	B	1002	-	4,4,4	0.14	0	6,6,6	0.10	0
2	SO4	B	1003	-	4,4,4	0.14	0	6,6,6	0.07	0
2	SO4	B	1004	-	4,4,4	0.15	0	6,6,6	0.05	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	SO4	A	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	A	1002	-	-	0/0/0/0	0/0/0/0
2	SO4	A	1003	-	-	0/0/0/0	0/0/0/0
2	SO4	B	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	B	1002	-	-	0/0/0/0	0/0/0/0
2	SO4	B	1003	-	-	0/0/0/0	0/0/0/0
2	SO4	B	1004	-	-	0/0/0/0	0/0/0/0

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.

5 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1002	SO4	1	0
2	A	1003	SO4	2	0
2	B	1001	SO4	1	0
2	B	1002	SO4	2	0
2	B	1003	SO4	2	0

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	944/968 (97%)	-0.03	12 (1%) 77 65	58, 105, 147, 170	0
1	B	944/968 (97%)	-0.05	10 (1%) 80 68	43, 99, 151, 199	0
All	All	1888/1936 (97%)	-0.04	22 (1%) 79 67	43, 102, 149, 199	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	49	PRO	5.4
1	A	813	THR	4.6
1	B	3	PHE	4.4
1	B	74	VAL	4.0
1	A	480	ASP	3.9

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, 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	LLDF	B-factors(\AA^2)	Q<0.9
2	SO4	B	1002	5/5	0.94	0.21	-0.40	74,75,82,103	0
2	SO4	A	1001	5/5	0.99	0.17	-0.59	56,57,70,88	0
2	SO4	B	1003	5/5	0.90	0.18	-0.80	113,118,134,134	0
2	SO4	A	1002	5/5	0.91	0.18	-0.90	74,91,95,99	0
2	SO4	A	1003	5/5	0.95	0.14	-1.08	95,102,125,133	0
2	SO4	B	1001	5/5	0.98	0.12	-1.94	64,64,69,78	0
2	SO4	B	1004	5/5	0.97	0.21	-	84,98,114,122	0

6.5 Other polymers [i](#)

There are no such residues in this entry.