

wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2023 – 03:53 pm GMT

PDB ID	:	4AYB
Title	:	RNAP at 3.2Ang
Authors	:	Wojtas, M.N.; Mogni, M.; Millet, O.; Bell, S.D.; Abrescia, N.G.A.
Deposited on	:	2012-06-19
Resolution	:	3.20 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
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
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)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: $X\text{-}RAY \, DIFFRACTION$

The reported resolution of this entry is 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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 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.

	01100111	
% 		
55%	39%	5% ••
%		
59%	34%	••
6%		
47%	42%	6% • 5%
%		
70%	28%	••
17%		
54%	39%	•• 5%
70 70 70 70 70 70 70	55% 59% % 47% 6 70% 17% 54%	55% 39% 59% 34% % 47% 47% 42% 6 28% 17% 54%

Continued on next page...



Mol	Chain	Length	(Quality of chain			
6	F	113	40%	42%		11%	7%
7	G	132	3%	43%		6% •	14%
8	Н	84	% 57%		31%	•	10%
9	K	95	% 63%		22%	•	12%
10	L	92	72%			27%	·
11	N	66	50%		41%		8% •
12	Р	48	4%	42%		8%	8%
13	Q	104	8% 17% 25%	5% •	52%		

Continued from previous page...

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
16	SF4	D	1264	-	-	Х	-



4AYB

2 Entry composition (i)

There are 16 unique types of molecules in this entry. The entry contains 54890 atoms, of which 27769 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.

Mol	Chain	Residues			Aton	ns			ZeroOcc	AltConf	Trace
1	А	872	Total 13986	C 4424	Н 7029	N 1225	O 1282	S 26	0	0	0

• Molecule 2 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues			Aton	ns			ZeroOcc	AltConf	Trace
2	В	1103	Total 17664	C 5548	Н 8908	N 1552	O 1627	S 29	0	0	0

• Molecule 3 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues			Atoms	5	ZeroOcc	AltConf	Trace		
3	С	376	Total 5974	C 1840	Н 3068	N 493	O 564	S 9	0	0	0

• Molecule 4 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues			Atom	IS			ZeroOcc	AltConf	Trace
4	D	262	Total 4215	C 1339	Н 2128	N 337	O 398	S 13	0	0	0

• Molecule 5 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues			Atom	IS		ZeroOcc	AltConf	Trace	
5	Е	171	Total 2772	C 874	Н 1413	N 229	0 251	${f S}{5}$	0	0	0

• Molecule 6 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues			Aton	ns	ZeroOcc	AltConf	Trace		
6	F	105	Total 1666	C 519	Н 839	N 134	0 171	${ m S} { m 3}$	0	0	0



• Molecule 7 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
7	G	113	Total 1816	C 572	Н 915	N 152	0 173	$\frac{S}{4}$	0	0	0

• Molecule 8 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
8	Н	76	Total 1284	C 405	Н 660	N 111	O 108	0	0	0

• Molecule 9 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
9	K	84	Total 1390	C 431	Н 717	N 123	0 118	S 1	0	0	0

• Molecule 10 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
10	L	91	Total 1449	C 454	Н 742	N 114	0 137	${ m S} { m 2}$	0	0	0

• Molecule 11 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
11	N	65	Total 1058	C 332	Н 537	N 94	O 88	S 7	0	0	0

• Molecule 12 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
12	Р	44	Total 744	C 236	Н 387	N 62	0 54	${ m S}{ m 5}$	0	0	0

• Molecule 13 is a protein called DNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
13	Q	50	Total 854	C 270	Н 426	N 74	O 83	S 1	0	0	0

• Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
14	А	4	Total Zn 4 4	0	0
14	В	3	Total Zn 3 3	0	0
14	С	1	Total Zn 1 1	0	0
14	Ν	1	Total Zn 1 1	0	0
14	Р	1	Total Zn 1 1	0	0

• Molecule 15 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
15	А	1	Total Mg 1 1	0	0

• Molecule 16 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe₄S₄).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
16	D	1	Total 7	Fe 3	$\frac{S}{4}$	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









W I D E DB



• Molecule 6: DNA-DIRECTED RNA POLYMERASE













4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants	195.72Å 212.41Å 128.76Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Bosolution (Å)	40.10 - 3.20	Depositor
	40.10 - 3.20	EDS
% Data completeness	99.4 (40.10-3.20)	Depositor
(in resolution range)	99.5 (40.10-3.20)	EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$2.41 (at 3.18 \text{\AA})$	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
P. P.	0.242 , 0.300	Depositor
n, n_{free}	0.235 , 0.293	DCC
R_{free} test set	4428 reflections (5.01%)	wwPDB-VP
Wilson B-factor $(Å^2)$	70.6	Xtriage
Anisotropy	0.333	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.34 , 69.7	EDS
L-test for $twinning^2$	$ < L >=0.49, < L^2>=0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	54890	wwPDB-VP
Average B, all atoms $(Å^2)$	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 2.00% of the height of the origin peak. No significant pseudotranslation is detected.

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.



¹Intensities estimated from amplitudes.

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, SF4, 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	Bond	angles
	Unam	RMSZ	# Z > 5	RMSZ	# Z > 5
1	А	0.27	0/7108	0.45	0/9618
2	В	0.25	0/8923	0.44	0/12071
3	С	0.24	0/2930	0.46	0/3944
4	D	0.23	0/2123	0.39	0/2870
5	Ε	0.23	0/1379	0.40	0/1861
6	F	0.23	0/836	0.44	0/1133
7	G	0.24	0/913	0.43	0/1224
8	Н	0.23	0/638	0.42	0/864
9	Κ	0.25	0/682	0.45	0/921
10	L	0.24	0/717	0.40	0/968
11	Ν	0.25	0/532	0.45	0/718
12	Р	0.30	0/365	0.47	0/489
13	Q	0.23	0/434	0.41	0/580
All	All	0.25	0/27580	0.44	0/37261

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 (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	А	6957	7029	7013	415	0

Continued on next page...



4AYB	
------	--

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	В	8756	8908	8888	418	0
3	С	2906	3068	3063	201	0
4	D	2087	2128	2125	59	0
5	Е	1359	1413	1411	75	0
6	F	827	839	839	83	0
7	G	901	915	912	87	0
8	Н	624	660	658	24	0
9	Κ	673	717	716	20	0
10	L	707	742	739	17	0
11	Ν	521	537	535	38	0
12	Р	357	387	387	21	0
13	Q	428	426	426	38	0
14	А	4	0	0	0	0
14	В	3	0	0	0	0
14	С	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
16	D	7	0	0	4	0
All	All	27121	27769	27712	1393	0

Continued from previous page...

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 25.

The worst 5 of 1393 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)
6:F:93:ARG:CB	6:F:94:THR:HA	1.87	1.05
1:A:146:CYS:SG	1:A:154:PHE:CZ	2.55	1.00
2:B:221:PRO:HB2	2:B:222:GLY:HA2	1.41	1.00
6:F:93:ARG:HB3	6:F:94:THR:HA	1.47	0.95
1:A:683:GLU:HA	1:A:684:LEU:HB3	1.51	0.93

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	Perc	entiles
1	А	868/880 (99%)	712 (82%)	124 (14%)	32 (4%)	3	22
2	В	1097/1131~(97%)	928 (85%)	133 (12%)	36 (3%)	4	25
3	С	372/395~(94%)	304 (82%)	49 (13%)	19 (5%)	2	15
4	D	260/265~(98%)	218 (84%)	37 (14%)	5 (2%)	8	39
5	Е	167/180~(93%)	150 (90%)	16 (10%)	1 (1%)	25	64
6	F	103/113 (91%)	78 (76%)	17 (16%)	8 (8%)	1	6
7	G	111/132 (84%)	83 (75%)	20 (18%)	8 (7%)	1	7
8	Н	74/84~(88%)	63~(85%)	8 (11%)	3 (4%)	3	21
9	К	82/95~(86%)	69 (84%)	9 (11%)	4 (5%)	2	17
10	L	89/92~(97%)	83 (93%)	4 (4%)	2 (2%)	6	35
11	N	63/66~(96%)	47 (75%)	10 (16%)	6 (10%)	0	3
12	Р	42/48 (88%)	29~(69%)	9 (21%)	4 (10%)	0	3
13	Q	48/104 (46%)	38 (79%)	5 (10%)	5 (10%)	0	3
All	All	3376/3585~(94%)	2802 (83%)	441 (13%)	133 (4%)	3	22

5 of 133 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	А	34	ASP
1	А	43	SER
1	А	44	VAL
1	А	229	ILE
1	А	291	SER

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.

Mol	Chain	Analysed	Rotameric	Outliers	Perce	entiles
1	А	758/766~(99%)	717 (95%)	41 (5%)	22	58
2	В	951/975~(98%)	905~(95%)	46~(5%)	25	61
3	С	324/341~(95%)	307~(95%)	17 (5%)	23	59
4	D	235/238~(99%)	235~(100%)	0	100	100
5	Ε	150/158~(95%)	143~(95%)	7 (5%)	26	62
6	F	99/107~(92%)	90 (91%)	9~(9%)	9	34
7	G	106/125~(85%)	99~(93%)	7 (7%)	16	51
8	Н	69/75~(92%)	68~(99%)	1 (1%)	67	86
9	Κ	74/83~(89%)	73~(99%)	1 (1%)	67	86
10	L	79/80~(99%)	79~(100%)	0	100	100
11	Ν	59/60~(98%)	57~(97%)	2(3%)	37	70
12	Р	40/43~(93%)	38~(95%)	2(5%)	24	60
13	Q	48/96~(50%)	46 (96%)	2 (4%)	30	65
All	All	2992/3147~(95%)	2857 (96%)	135 (4%)	27	63

The Analysed column shows the number of residues for which the side chain conformation was analysed, and the total number of residues.

5 of 135 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	F	86	ILE
6	F	95	TYR
11	Ν	61	HIS
2	В	217	PHE
2	В	207	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	357	ASN
2	В	640	HIS
2	В	686	ASN
2	В	696	HIS
2	В	926	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 12 ligands modelled in this entry, 11 are monoatomic - leaving 1 for Mogul analysis.

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	Chain	Bos	Link	B	ond leng	gths	E	Bond angles
	Type	nes			Counts	RMSZ	# Z > 2	Counts	RMSZ # Z > 2	
16	SF4	D	1264	4	0,9,12	-	-	-		

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
16	SF4	D	1264	4	-	-	0/3/3/5

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.



1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
16	D	1264	SF4	4	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>	#RSRZ>2	$\mathbf{OWAB}(\mathbf{\AA}^2)$	Q<0.9
1	А	872/880~(99%)	0.05	17 (1%) 66 53	21,75,161,226	0
2	В	1103/1131~(97%)	0.07	23 (2%) 63 49	26, 73, 154, 281	0
3	С	376/395~(95%)	0.38	22 (5%) 22 13	30, 89, 170, 288	0
4	D	262/265~(98%)	0.20	8 (3%) 49 32	54, 92, 150, 180	0
5	Е	171/180~(95%)	1.03	30 (17%) 1 1	49, 169, 233, 275	0
6	F	105/113~(92%)	1.14	24 (22%) 0 0	103, 197, 244, 281	0
7	G	113/132~(85%)	0.56	4 (3%) 44 28	60, 123, 170, 266	0
8	Н	76/84~(90%)	0.05	1 (1%) 77 65	49, 77, 132, 176	0
9	Κ	84/95~(88%)	-0.03	1 (1%) 79 67	35, 58, 149, 195	0
10	L	91/92~(98%)	0.05	0 100 100	48, 86, 132, 218	0
11	Ν	65/66~(98%)	0.08	0 100 100	36, 81, 119, 143	0
12	Р	44/48~(91%)	0.27	2 (4%) 33 21	52, 94, 153, 227	0
13	Q	50/104~(48%)	0.90	8 (16%) 1 1	98, 154, 195, 204	0
All	All	3412/3585~(95%)	0.22	140 (4%) 37 24	21, 85, 188, 288	0

The worst 5 of 140 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	С	229	THR	7.3
6	F	78	ILE	6.4
1	А	40	ILE	5.9
6	F	79	THR	5.6
5	Е	117	THR	5.4

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

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



4AYB

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, 95^{th} 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	$B-factors(A^2)$	Q<0.9
14	ZN	В	2128	1/1	0.83	0.29	81,81,81,81	0
14	ZN	А	1880	1/1	0.89	0.08	81,81,81,81	0
15	MG	А	1884	1/1	0.89	0.23	81,81,81,81	0
14	ZN	А	1882	1/1	0.92	0.34	81,81,81,81	0
14	ZN	А	1883	1/1	0.92	0.16	81,81,81,81	0
14	ZN	В	2127	1/1	0.93	0.34	81,81,81,81	0
14	ZN	С	1395	1/1	0.94	0.06	81,81,81,81	0
14	ZN	Ν	1066	1/1	0.94	0.21	88,88,88,88	0
14	ZN	А	1881	1/1	0.94	0.10	81,81,81,81	0
16	SF4	D	1264	7/8	0.95	0.21	81,81,81,81	0
14	ZN	В	2126	1/1	0.97	0.11	81,81,81,81	0
14	ZN	Р	1049	1/1	0.98	0.10	118,118,118,118	0

6.5 Other polymers (i)

There are no such residues in this entry.

