Full wwPDB X-ray Structure Validation Report ⓘ

Feb 5, 2025 – 12:09 PM EST

PDB ID	:	7S01
Title	:	X-ray structure of the phage AR9 non-virion RNA polymerase holoenzyme in complex with a forked oligonucleotide containing the P077 promoter
Authors	:	Leiman, P.G.; Sokolova, M.L.; Gordeeva, J.; Fraser, A.; Severinov, K.V.
Deposited on	:	2021-08-28
Resolution	:	3.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ 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 ⓘ) were used in the production of this report:

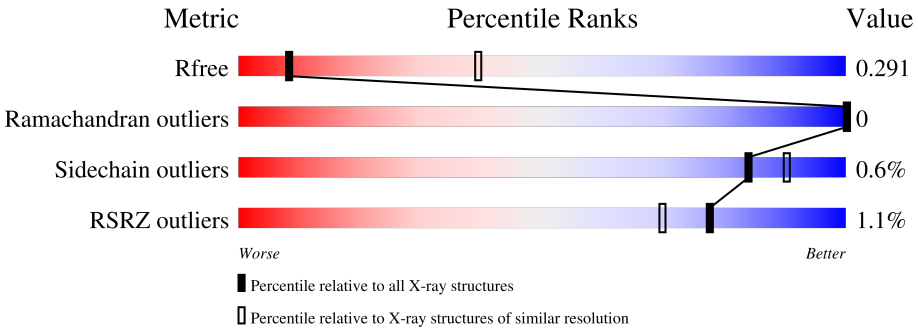
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 3.40 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.



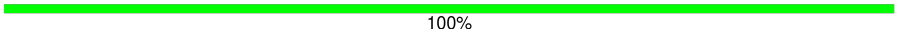
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1140 (3.46-3.34)
Ramachandran outliers	177936	1172 (3.46-3.34)
Sidechain outliers	177891	1172 (3.46-3.34)
RSRZ outliers	164620	1140 (3.46-3.34)

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	464	99%
2	d	426	99%
3	c	496	97%
4	D	631	99%
5	C	665	97%
6	N	14	93% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	t	14	
7	T	32	
8	n	32	

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 23345 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	463	Total	C	N	O	S	0	0	0
			3847	2478	620	729	20			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	d	426	Total	C	N	O	S	0	0	0
			3488	2256	561	663	8			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	c	484	Total	C	N	O	S	0	0	0
			4003	2580	658	754	11			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	631	Total	C	N	O	S	0	0	0
			5140	3276	837	1009	18			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	C	649	Total	C	N	O	S	0	0	0
			5271	3337	868	1041	25			

- Molecule 6 is a DNA chain called Non-template strand of the forked DNA oligonucleotide (downstream copy).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	N	13	Total	C	N	O	P	0	0	0
			267	128	49	77	13			
6	t	14	Total	C	N	O	P	0	0	0
			289	138	54	83	14			

- Molecule 7 is a DNA chain called Template strand of the forked DNA oligonucleotide (downstream copy) containing the P077 AR9 promoter motif.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	T	29	Total	C	N	O	P	0	0	0
			589	283	99	178	29			

- Molecule 8 is a DNA chain called Template strand of the forked DNA oligonucleotide (downstream copy) containing the P077 AR9 promoter motif.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	n	22	Total	C	N	O	P	0	0	0
			450	217	80	131	22			

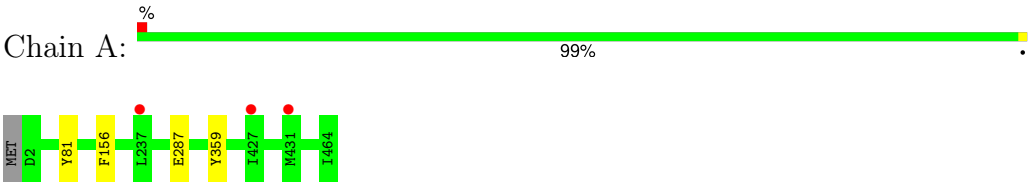
- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	1	Total	Zn	0	0
			1	1		

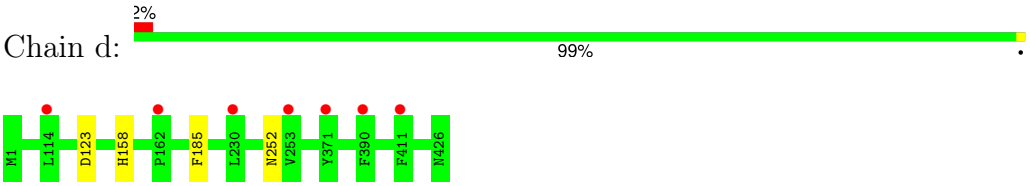
3 Residue-property plots (i)

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.

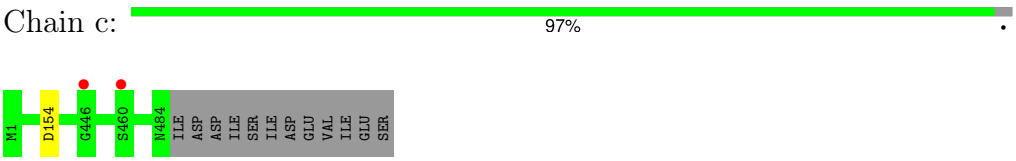
• Molecule 1: DNA-directed RNA polymerase subunit



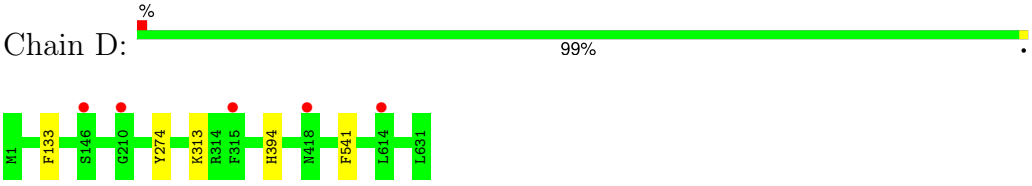
• Molecule 2: DNA-directed RNA polymerase beta' subunit



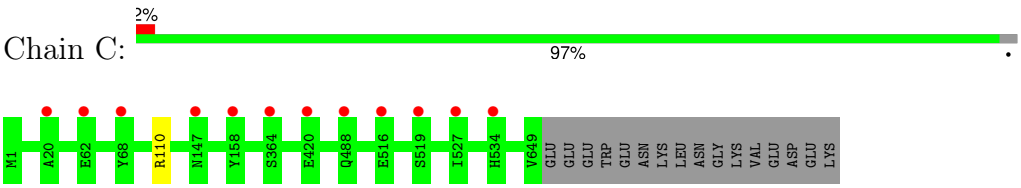
• Molecule 3: DNA-directed RNA polymerase beta subunit



• Molecule 4: DNA-directed RNA polymerase

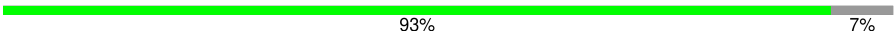


• Molecule 5: DNA-directed RNA polymerase



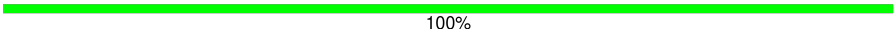
• Molecule 6: Non-template strand of the forked DNA oligonucleotide (downstream copy)

Chain N:



• Molecule 6: Non-template strand of the forked DNA oligonucleotide (downstream copy)

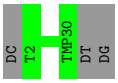
Chain t:



There are no outlier residues recorded for this chain.

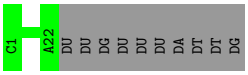
• Molecule 7: Template strand of the forked DNA oligonucleotide (downstream copy) containing the P077 AR9 promoter motif

Chain T:



• Molecule 8: Template strand of the forked DNA oligonucleotide (downstream copy) containing the P077 AR9 promoter motif

Chain n:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	176.93Å 110.45Å 222.38Å 90.00° 98.52° 90.00°	Depositor
Resolution (Å)	49.35 – 3.40 49.35 – 3.40	Depositor EDS
% Data completeness (in resolution range)	96.9 (49.35-3.40) 97.0 (49.35-3.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.13 (at 3.40Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.239 , 0.293 0.239 , 0.291	Depositor DCC
R_{free} test set	2839 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	109.7	Xtriage
Anisotropy	0.316	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 112.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	23345	wwPDB-VP
Average B, all atoms (Å^2)	133.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.32% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, D5M, UMP, TMP, DGP

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	0.24	0/3909	0.40	0/5259
2	d	0.24	0/3558	0.41	0/4804
3	c	0.24	0/4082	0.44	0/5502
4	D	0.25	0/5222	0.44	0/7038
5	C	0.25	0/5370	0.44	0/7241
6	N	0.53	0/299	0.93	0/459
6	t	0.57	0/324	0.91	0/498
7	T	0.54	0/483	0.96	0/743
8	n	0.54	0/504	0.95	0/775
All	All	0.28	0/23751	0.49	0/32319

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

Due to software issues we are unable to calculate clashes - this section is therefore empty.

5.3 Torsion angles

5.3.1 Protein backbone

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	461/464 (99%)	452 (98%)	9 (2%)	0	100	100
2	d	424/426 (100%)	418 (99%)	6 (1%)	0	100	100
3	c	482/496 (97%)	473 (98%)	9 (2%)	0	100	100
4	D	629/631 (100%)	623 (99%)	6 (1%)	0	100	100
5	C	647/665 (97%)	625 (97%)	22 (3%)	0	100	100
All	All	2643/2682 (98%)	2591 (98%)	52 (2%)	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 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	444/445 (100%)	440 (99%)	4 (1%)	75	86
2	d	390/390 (100%)	386 (99%)	4 (1%)	73	83
3	c	457/469 (97%)	456 (100%)	1 (0%)	92	96
4	D	590/590 (100%)	585 (99%)	5 (1%)	79	87
5	C	593/608 (98%)	592 (100%)	1 (0%)	92	96
All	All	2474/2502 (99%)	2459 (99%)	15 (1%)	84	90

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	81	TYR
1	A	156	PHE
1	A	287	GLU
1	A	359	TYR
2	d	123	ASP
2	d	158	HIS
2	d	185	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	d	252	ASN
3	c	154	ASP
4	D	133	PHE
4	D	274	TYR
4	D	313	LYS
4	D	394	HIS
4	D	541	PHE
5	C	110	ARG

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

Mol	Chain	Res	Type
2	d	95	GLN
2	d	252	ASN
3	c	165	ASN
4	D	494	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

Of 1 ligands modelled in this entry, 1 is 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

6.1 Protein, DNA and RNA chains

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^{th} 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	463/464 (99%)	0.16	3 (0%) 85 80	92, 151, 211, 270	0
2	d	426/426 (100%)	0.19	7 (1%) 70 61	77, 116, 183, 248	0
3	c	484/496 (97%)	0.18	2 (0%) 89 86	77, 125, 193, 259	0
4	D	631/631 (100%)	0.23	5 (0%) 82 75	78, 122, 170, 208	0
5	C	649/665 (97%)	0.28	12 (1%) 67 58	66, 101, 144, 190	0
6	N	13/14 (92%)	0.07	0 100 100	210, 230, 349, 352	0
6	t	14/14 (100%)	-0.02	0 100 100	185, 246, 330, 351	0
7	T	21/32 (65%)	0.51	0 100 100	200, 233, 285, 363	0
8	n	22/32 (68%)	-0.22	0 100 100	142, 205, 261, 275	0
All	All	2723/2774 (98%)	0.21	29 (1%) 77 70	66, 121, 197, 363	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	C	488	GLN	3.4
1	A	427	ILE	2.9
2	d	253	VAL	2.9
5	C	534	HIS	2.8
3	c	446	GLY	2.8
2	d	371	TYR	2.7
4	D	418	ASN	2.5
1	A	237	LEU	2.5
5	C	158	TYR	2.4
5	C	519	SER	2.4
4	D	315	PHE	2.4
2	d	411	PHE	2.3
5	C	147	ASN	2.3
2	d	390	PHE	2.3
4	D	210	GLY	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	c	460	SER	2.3
2	d	162	PRO	2.2
4	D	614	LEU	2.2
5	C	420	GLU	2.2
4	D	146	SER	2.2
5	C	62	GLU	2.2
5	C	364	SER	2.1
5	C	516	GLU	2.1
5	C	68	TYR	2.1
5	C	527	ILE	2.1
2	d	114	LEU	2.1
1	A	431	MET	2.1
2	d	230	LEU	2.0
5	C	20	ALA	2.0

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, 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(Å^2)	Q<0.9
9	ZN	D	801	1/1	0.97	0.05	116,116,116,116	0

6.5 Other polymers (i)

There are no such residues in this entry.