



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 18, 2017 – 05:08 PM EDT

PDB ID : 3E2E
Title : Crystal Structure of an Intermediate Complex of T7 RNAP and 7nt of RNA
Authors : Durniak, K.J.; Bailey, S.; Steitz, T.A.
Deposited on : unknown
Resolution : 3.00 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
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-20030345

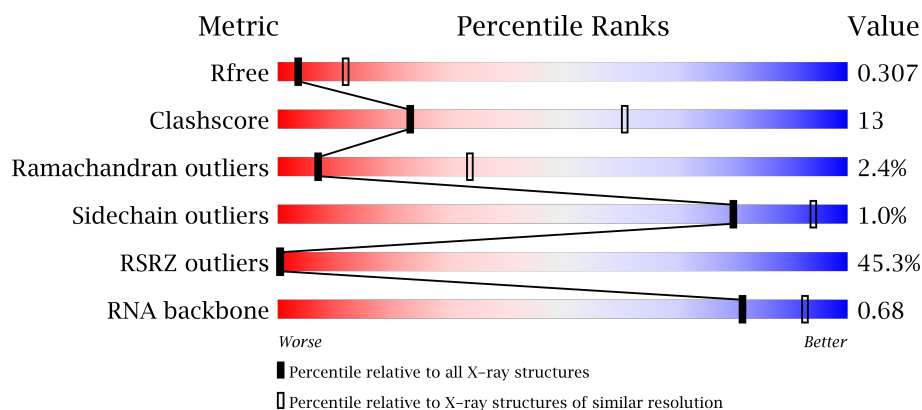
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)
RNA backbone	2435	1007 (3.34-2.66)

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	889	<div> <div>39%</div> <div> <div>72%</div> <div>19%</div> <div>• 6%</div> </div> </div>
2	R	7	<div> <div>100%</div> <div> <div>86%</div> <div>14%</div> </div> </div>
3	T	33	<div> <div>88%</div> <div> <div>52%</div> <div>39%</div> <div>• 6%</div> </div> </div>
4	N	33	<div> <div>82%</div> <div> <div>55%</div> <div>24%</div> <div>6%</div> <div>15%</div> </div> </div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8013 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	832	Total	C	N	O	S	0	0	0
			6589	4206	1139	1207	37			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	EXPRESSION TAG	UNP P00573
A	-4	HIS	-	EXPRESSION TAG	UNP P00573
A	-3	HIS	-	EXPRESSION TAG	UNP P00573
A	-2	HIS	-	EXPRESSION TAG	UNP P00573
A	-1	HIS	-	EXPRESSION TAG	UNP P00573
A	0	HIS	-	EXPRESSION TAG	UNP P00573
A	266	LEU	PRO	ENGINEERED	UNP P00573

- Molecule 2 is a RNA chain called RNA (5'-R(*GP*GP*GP*AP*GP*UP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	R	7	Total	C	N	O	P	0	0	0
			153	69	32	46	6			

- Molecule 3 is a DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	31	Total	C	N	O	P	0	0	0
			633	301	113	188	31			

- Molecule 4 is a DNA chain called DNA (28-MER).

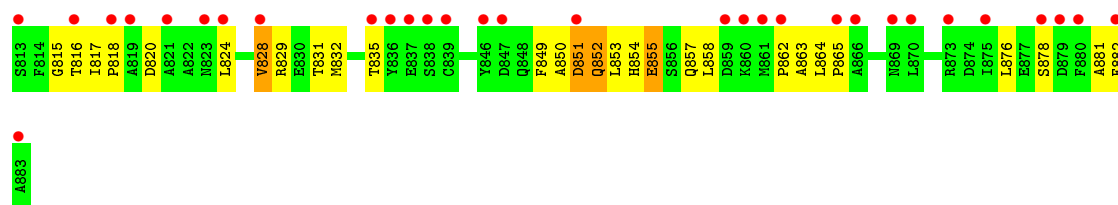
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	N	28	Total	C	N	O	P	0	0	0
			565	271	101	166	27			

- Molecule 5 is water.

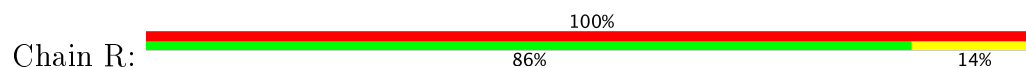
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	64	Total 64	O 64	0	0
5	R	1	Total 1	O 1	0	0
5	T	3	Total 3	O 3	0	0
5	N	5	Total 5	O 5	0	0

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 ($\text{RSRZ} > 2$). 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.

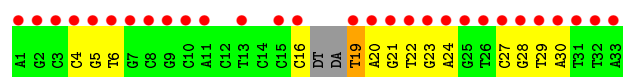
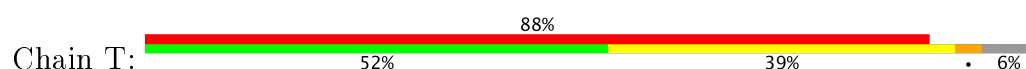
Chain A:



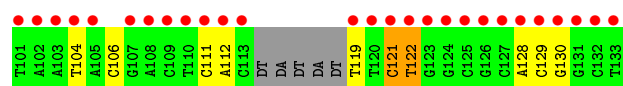
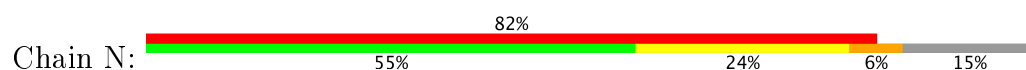
- Molecule 2: RNA (5'-R(*GP*GP*GP*AP*GP*UP*G)-3')



- Molecule 3: DNA (31-MER)



- Molecule 4: DNA (28-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α , β , γ	81.01Å 81.01Å 358.79Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.80 – 3.00 37.79 – 3.00	Depositor EDS
% Data completeness (in resolution range)	98.5 (37.80-3.00) 98.5 (37.79-3.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.31 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.240 , 0.292 0.251 , 0.307	Depositor DCC
R_{free} test set	1330 reflections (5.37%)	DCC
Wilson B-factor (Å ²)	37.9	Xtriage
Anisotropy	0.104	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 85.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.079 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.84	EDS
Total number of atoms	8013	wwPDB-VP
Average B, all atoms (Å ²)	99.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.89% 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

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	0.34	0/6738	0.58	4/9109 (0.0%)
2	R	0.65	0/172	1.12	0/269
3	T	0.63	0/707	1.24	1/1086 (0.1%)
4	N	0.65	0/631	1.33	4/969 (0.4%)
All	All	0.41	0/8248	0.77	9/11433 (0.1%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	T	19	DT	O4'-C1'-N1	8.94	114.26	108.00
4	N	122	DT	O4'-C1'-N1	7.05	112.93	108.00
1	A	43	SER	N-CA-C	6.99	129.88	111.00
4	N	121	DC	P-O3'-C3'	6.39	127.37	119.70
4	N	104	DT	C1'-O4'-C4'	-6.28	103.82	110.10

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	6589	0	6558	176	0
2	R	153	0	75	4	0
3	T	633	0	351	19	0
4	N	565	0	315	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	64	0	0	9	0
5	N	5	0	0	0	0
5	R	1	0	0	0	0
5	T	3	0	0	1	0
All	All	8013	0	7299	195	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 195 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:A:376:ALA:HB3	1:A:377:TRP:CB	1.72	1.19
1:A:553:GLU:CG	1:A:554:VAL:HG23	1.87	1.04
1:A:354:ALA:CA	1:A:355:ILE:HG12	1.88	1.04
1:A:131:ASN:HB3	1:A:132:THR:HA	1.06	1.03
1:A:354:ALA:HA	1:A:355:ILE:CG1	1.87	1.03

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	820/889 (92%)	738 (90%)	62 (8%)	20 (2%)	7 34

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	292	ARG
1	A	371	PRO
1	A	452	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	554	VAL
1	A	557	ARG

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	693/735 (94%)	686 (99%)	7 (1%)	80 94

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	452	ILE
1	A	858	LEU
1	A	735	VAL
1	A	133	THR
1	A	828	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	419	ASN
1	A	522	GLN
1	A	811	HIS
1	A	410	ASN
1	A	786	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	R	6/7 (85%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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](#)

There are no ligands in this entry.

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	832/889 (93%)	2.08	344 (41%) 0 0	30, 100, 101, 102	0
2	R	7/7 (100%)	5.48	7 (100%) 0 0	99, 100, 100, 101	0
3	T	31/33 (93%)	3.78	29 (93%) 0 0	97, 100, 101, 102	0
4	N	28/33 (84%)	4.27	27 (96%) 0 0	99, 100, 101, 101	0
All	All	898/962 (93%)	2.23	407 (45%) 0 0	30, 100, 101, 102	0

The worst 5 of 407 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	R	1	G	13.0
1	A	131	ASN	8.6
4	N	120	DT	8.3
1	A	716	GLY	8.0
1	A	598	THR	7.9

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.