wwPDB X-ray Structure Validation Summary Report (i)

Oct 15, 2023 – 05:00 AM EDT

PDB ID	:	8DH0
Title	:	T7 RNA polymerase elongation complex with unnatural base dDs
Authors	:	Oh, J.; Wang, D.
Deposited on	:	2022-06-24
Resolution	:	2.90 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 (i)) were used in the production of this report:

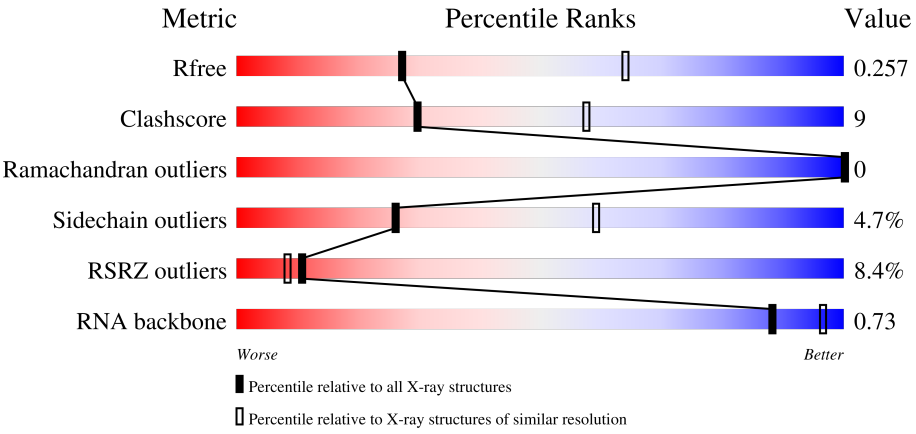
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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.



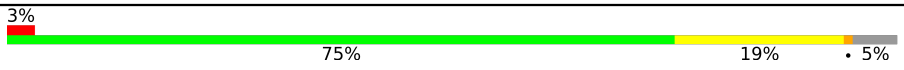
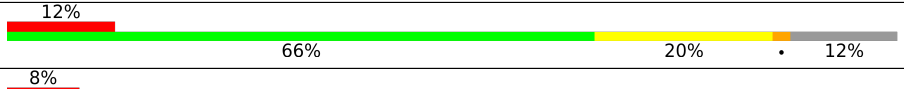
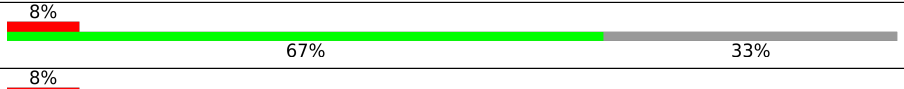
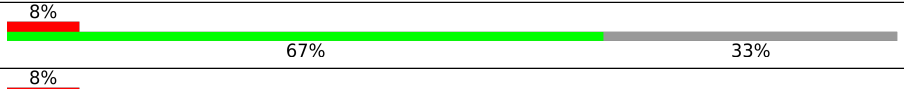
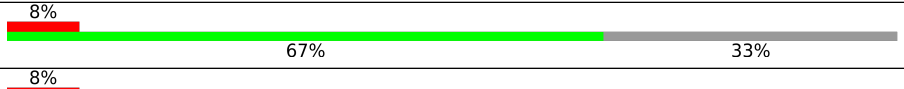
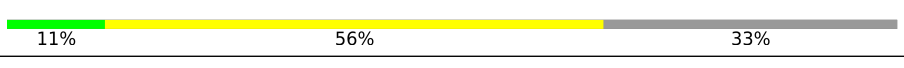
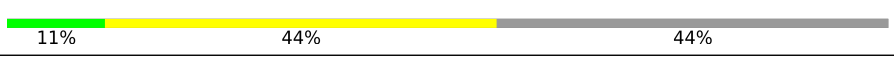
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)
RNA backbone	3102	1007 (3.16-2.64)

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	18	44% 28% 28%
1	E	18	6% 56% 28% 17%
1	I	18	56% 11% 33%
1	M	18	67% 11% 22%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	B	883	
2	F	883	
2	J	883	
2	N	883	
3	C	12	
3	G	12	
3	K	12	
3	O	12	
4	H	9	
4	P	9	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 26201 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 DNA chain called Template strand DNA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	13	Total	C	N	O	P	S	0	0	0
			268	130	47	77	13	1			
1	E	15	Total	C	N	O	P	S	0	0	0
			310	150	57	87	15	1			
1	I	12	Total	C	N	O	P	S	0	0	0
			248	120	45	70	12	1			
1	M	14	Total	C	N	O	P	S	0	0	0
			289	140	52	82	14	1			

- Molecule 2 is a protein called T7 RNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	840	Total	C	N	O	S	0	0	0
			6503	4142	1129	1196	36			
2	F	796	Total	C	N	O	S	0	0	0
			5875	3734	1028	1082	31			
2	J	805	Total	C	N	O	S	0	0	0
			6058	3859	1051	1114	34			
2	N	773	Total	C	N	O	S	0	0	0
			5706	3622	1006	1050	28			

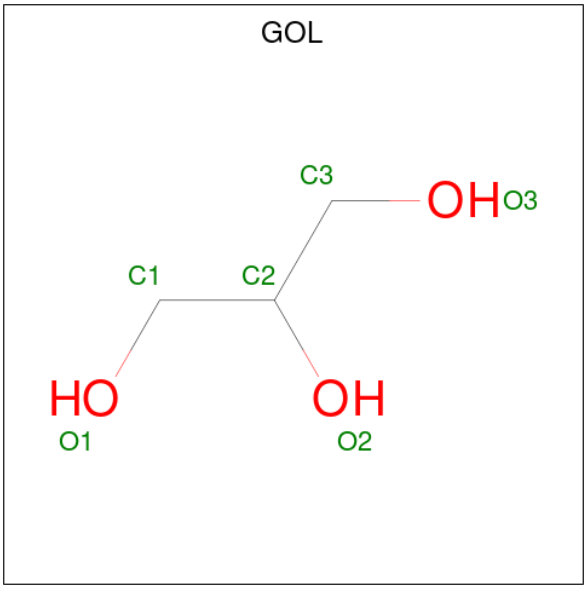
- Molecule 3 is a RNA chain called RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	8	Total	C	N	O	P	0	0	0
			174	77	33	56	8			
3	G	8	Total	C	N	O	P	0	0	0
			174	77	33	56	8			
3	K	8	Total	C	N	O	P	0	0	0
			174	77	33	56	8			
3	O	8	Total	C	N	O	P	0	0	0
			174	77	33	56	8			

- Molecule 4 is a DNA chain called Non-template strand DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	H	6	Total	C	N	O	P	0	0	0
			122	59	19	38	6			
4	P	5	Total	C	N	O	P	0	0	0
			102	49	17	31	5			

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



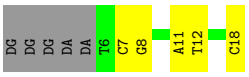
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	B	1	Total	C	O		0	0
			6	3	3			
5	F	1	Total	C	O		0	0
			6	3	3			
5	J	1	Total	C	O		0	0
			6	3	3			
5	N	1	Total	C	O		0	0
			6	3	3			

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.

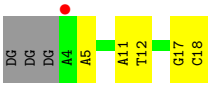
• Molecule 1: Template strand DNA

Chain A:



• Molecule 1: Template strand DNA

Chain E:

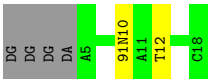


• Molecule 1: Template strand DNA

Chain I:

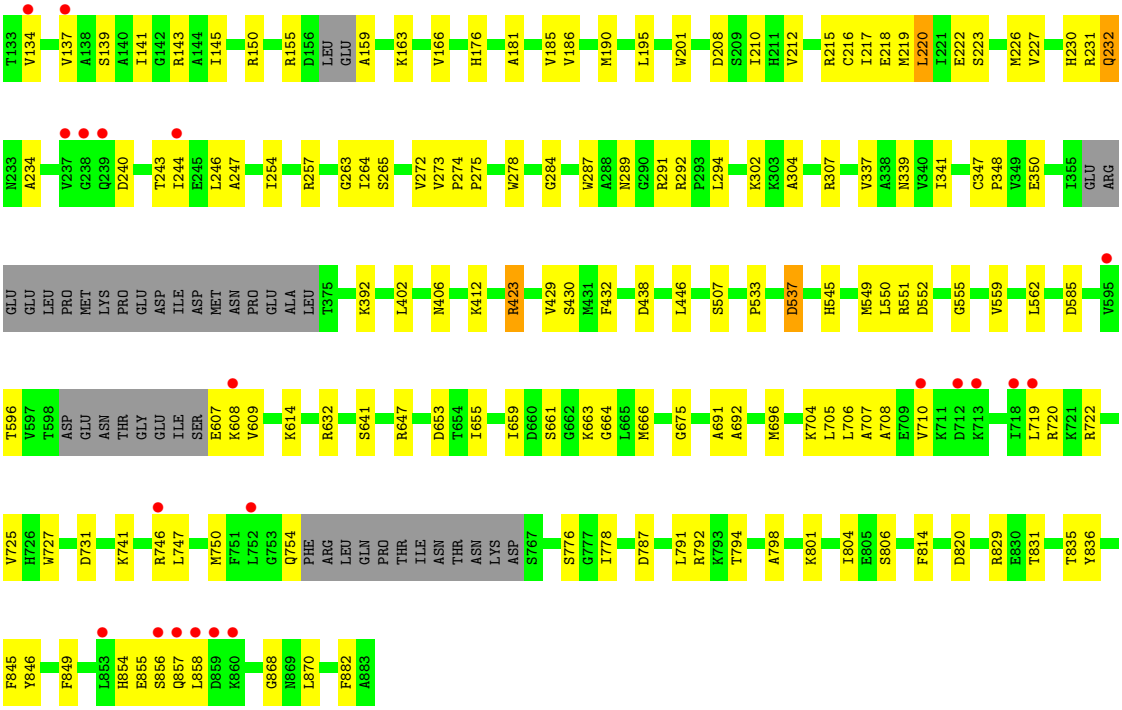
• Molecule 1: Template strand DNA

Chain M:



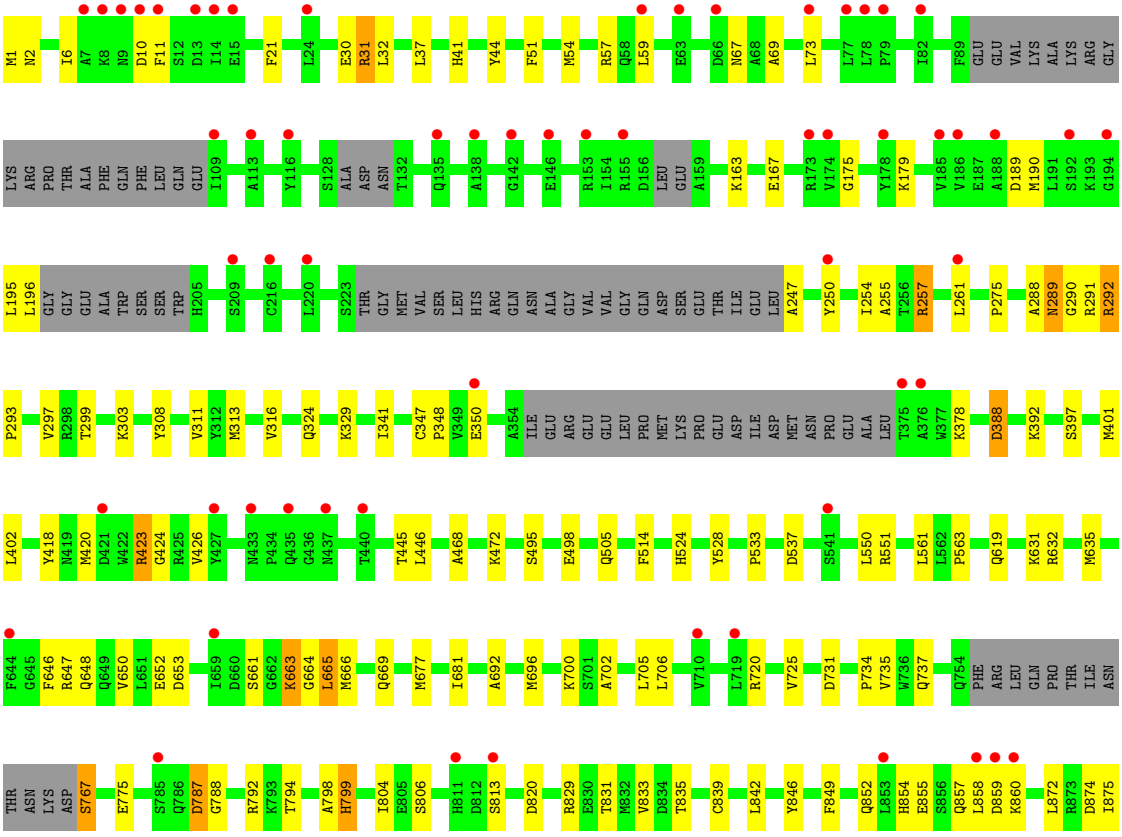
• Molecule 2: T7 RNA polymerase

Chain B:



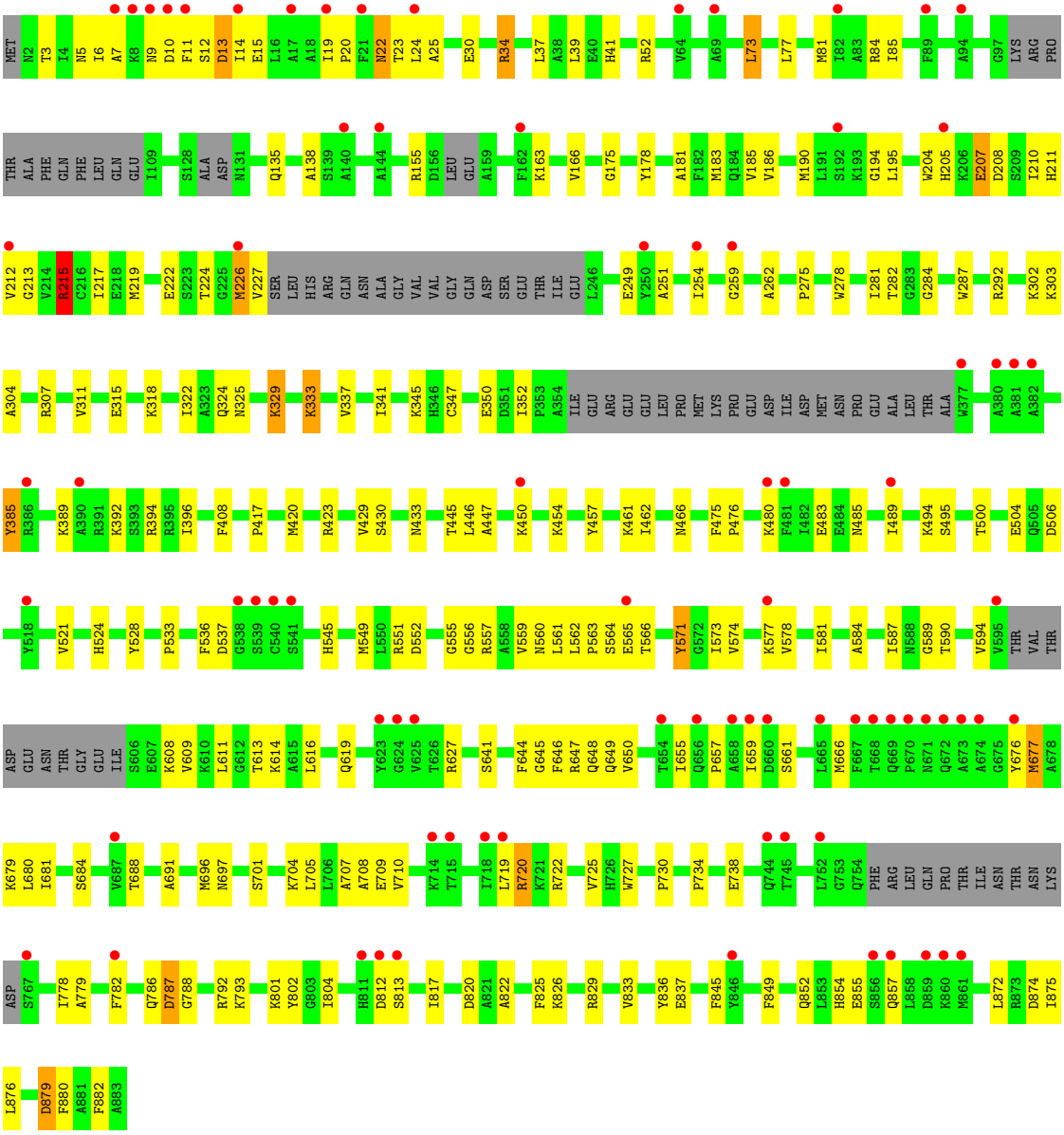
• Molecule 2: T7 RNA polymerase

Chain F:



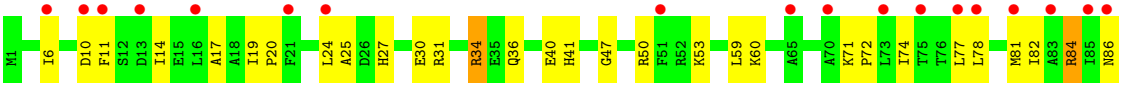
• Molecule 2: T7 RNA polymerase

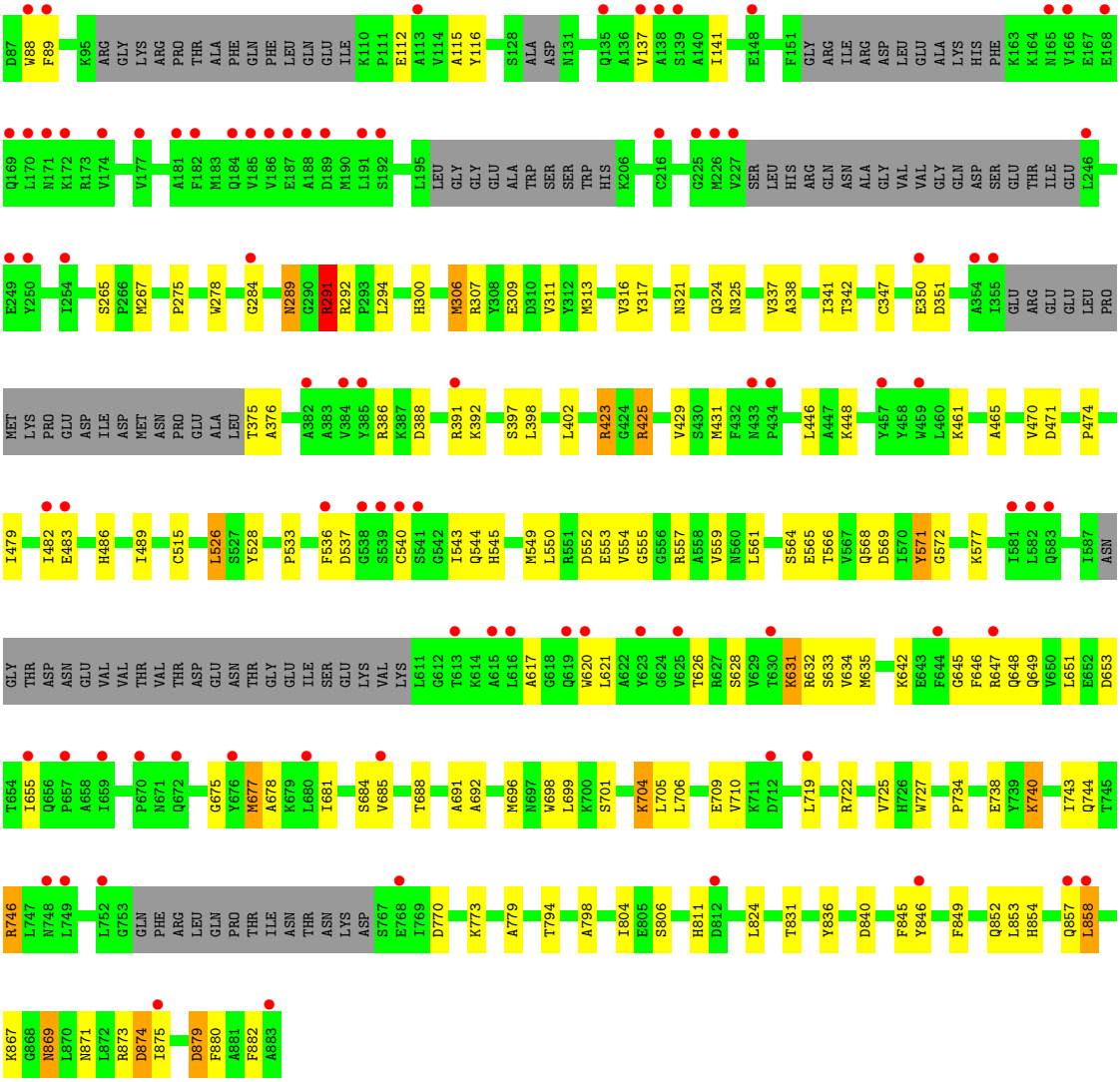
Chain J:



• Molecule 2: T7 RNA polymerase

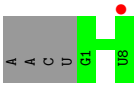
Chain N:





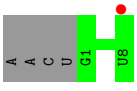
• Molecule 3: RNA

Chain C:



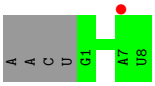
• Molecule 3: RNA

Chain G:



• Molecule 3: RNA

Chain K:

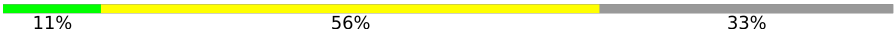


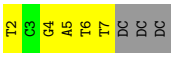
• Molecule 3: RNA

Chain O:

• Molecule 4: Non-template strand DNA

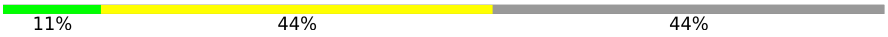
Chain H:





• Molecule 4: Non-template strand DNA

Chain P:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	78.71Å 86.23Å 201.22Å 89.91° 85.23° 69.50°	Depositor
Resolution (Å)	42.03 – 2.90 42.02 – 2.90	Depositor EDS
% Data completeness (in resolution range)	96.9 (42.03-2.90) 96.9 (42.02-2.90)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.06 (at 2.90Å)	Xtriage
Refinement program	PHENIX 1.19	Depositor
R, R_{free}	0.221 , 0.261 0.216 , 0.257	Depositor DCC
R_{free} test set	1783 reflections (1.68%)	wwPDB-VP
Wilson B-factor (Å^2)	82.5	Xtriage
Anisotropy	0.244	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 75.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	26201	wwPDB-VP
Average B, all atoms (Å^2)	101.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.80% 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: GOL, 91N

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	0.52	0/270	0.93	0/411
1	E	0.53	0/318	0.98	1/485 (0.2%)
1	I	0.53	0/248	0.88	0/377
1	M	0.58	0/294	0.96	0/448
2	B	0.27	0/6646	0.52	4/9000 (0.0%)
2	F	0.28	0/5999	0.51	4/8142 (0.0%)
2	J	0.31	0/6192	0.55	4/8399 (0.0%)
2	N	0.30	0/5824	0.56	6/7906 (0.1%)
3	C	0.21	0/194	0.79	0/301
3	G	0.19	0/194	0.77	0/301
3	K	0.18	0/194	0.75	0/301
3	O	0.18	0/194	0.72	0/301
4	H	0.54	0/135	1.06	0/206
4	P	0.53	0/113	0.97	0/172
All	All	0.30	0/26815	0.58	19/36750 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	J	0	3
2	N	0	1
All	All	0	4

There are no bond length outliers.

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	F	59	LEU	CA-CB-CG	7.43	132.39	115.30
2	N	537	ASP	CB-CG-OD1	7.28	124.86	118.30
2	J	13	ASP	CB-CG-OD1	7.22	124.80	118.30
2	F	787	ASP	CB-CG-OD1	6.17	123.85	118.30
2	N	526	LEU	CA-CB-CG	6.17	129.48	115.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	J	215	ARG	Sidechain
2	J	34	ARG	Sidechain
2	J	394	ARG	Sidechain
2	N	291	ARG	Sidechain

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	268	0	135	4	0
1	E	310	0	157	5	0
1	I	248	0	123	3	0
1	M	289	0	146	2	0
2	B	6503	0	6377	104	0
2	F	5875	0	5480	82	0
2	J	6058	0	5747	146	0
2	N	5706	0	5328	125	0
3	C	174	0	88	0	0
3	G	174	0	88	0	0
3	K	174	0	88	0	0
3	O	174	0	88	2	0
4	H	122	0	70	4	0
4	P	102	0	58	3	0
5	B	6	0	8	0	0
5	F	6	0	8	0	0
5	J	6	0	8	0	0
5	N	6	0	8	0	0
All	All	26201	0	24005	468	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 9.

The worst 5 of 468 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:J:30:GLU:OE1	2:J:34:ARG:NH2	2.14	0.80
2:N:317:TYR:O	2:N:321:ASN:ND2	2.17	0.77
2:N:19:ILE:HG13	2:N:20:PRO:HD3	1.67	0.76
2:J:30:GLU:HB2	2:J:34:ARG:HH21	1.51	0.76
2:J:561:LEU:HG	2:J:875:ILE:HG12	1.68	0.76

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	
2	B	828/883 (94%)	811 (98%)	17 (2%)	0	100	100
2	F	780/883 (88%)	770 (99%)	10 (1%)	0	100	100
2	J	789/883 (89%)	774 (98%)	15 (2%)	0	100	100
2	N	755/883 (86%)	745 (99%)	10 (1%)	0	100	100
All	All	3152/3532 (89%)	3100 (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	
2	B	662/729 (91%)	638 (96%)	24 (4%)	35	69
2	F	542/729 (74%)	521 (96%)	21 (4%)	32	66
2	J	583/729 (80%)	556 (95%)	27 (5%)	27	60
2	N	526/729 (72%)	490 (93%)	36 (7%)	16	42
All	All	2313/2916 (79%)	2205 (95%)	108 (5%)	26	59

5 of 108 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	J	454	LYS
2	J	879	ASP
2	N	704	LYS
2	J	495	SER
2	J	666	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
2	J	466	ASN
2	J	619	GLN
2	N	744	GLN
2	J	786	GLN
2	N	737	GLN

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	C	7/12 (58%)	0	0
3	G	7/12 (58%)	0	0
3	K	7/12 (58%)	0	0
3	O	7/12 (58%)	0	0
All	All	28/48 (58%)	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

4 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	$\#\|Z\| > 2$
5	GOL	F	901	-	5,5,5	0.84	0	5,5,5	1.03	0
5	GOL	B	901	-	5,5,5	0.92	0	5,5,5	0.98	0
5	GOL	N	901	-	5,5,5	0.88	0	5,5,5	1.03	0
5	GOL	J	901	-	5,5,5	0.88	0	5,5,5	1.04	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
5	GOL	F	901	-	-	0/4/4/4	-
5	GOL	B	901	-	-	0/4/4/4	-
5	GOL	N	901	-	-	0/4/4/4	-
5	GOL	J	901	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	J	901	GOL	C1-C2-C3-O3
5	J	901	GOL	O2-C2-C3-O3

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	12/18 (66%)	0.20	0 100 100	49, 60, 153, 172	0
1	E	14/18 (77%)	0.49	1 (7%) 16 12	65, 99, 229, 260	0
1	I	11/18 (61%)	0.26	0 100 100	68, 80, 208, 218	0
1	M	13/18 (72%)	0.58	0 100 100	85, 108, 198, 228	0
2	B	840/883 (95%)	0.18	28 (3%) 46 41	40, 69, 134, 191	0
2	F	796/883 (90%)	0.41	60 (7%) 14 11	44, 79, 169, 209	0
2	J	805/883 (91%)	0.50	80 (9%) 7 5	58, 106, 169, 246	0
2	N	773/883 (87%)	0.71	105 (13%) 3 2	72, 118, 192, 258	0
3	C	8/12 (66%)	0.64	1 (12%) 3 3	52, 60, 85, 100	0
3	G	8/12 (66%)	0.68	1 (12%) 3 3	56, 80, 124, 138	0
3	K	8/12 (66%)	0.62	1 (12%) 3 3	77, 96, 122, 152	0
3	O	8/12 (66%)	0.90	1 (12%) 3 3	90, 113, 177, 193	0
4	H	6/9 (66%)	0.68	0 100 100	199, 205, 233, 253	0
4	P	5/9 (55%)	0.53	0 100 100	203, 206, 221, 252	0
All	All	3307/3670 (90%)	0.45	278 (8%) 11 8	40, 96, 175, 260	0

The worst 5 of 278 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	N	188	ALA	6.8
2	J	8	LYS	5.7
2	J	671	ASN	5.7
2	J	625	VAL	5.5
2	N	148	GLU	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.

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
5	GOL	N	901	6/6	0.82	0.30	86,98,101,102	0
5	GOL	J	901	6/6	0.84	0.28	86,93,105,109	0
5	GOL	F	901	6/6	0.94	0.21	66,75,81,85	0
5	GOL	B	901	6/6	0.96	0.23	52,61,68,87	0

6.5 Other polymers (i)

There are no such residues in this entry.