



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 15, 2024 – 09:36 PM EST

PDB ID : 5VO8
Title : X-ray crystal structure of a bacterial reiterative transcription complex of pyrG promoter
Authors : Murakami, K.S.; Shin, Y.; Turnbough Jr, C.L.; Molodtsov, V.
Deposited on : 2017-05-02
Resolution : 3.30 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 ⓘ 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

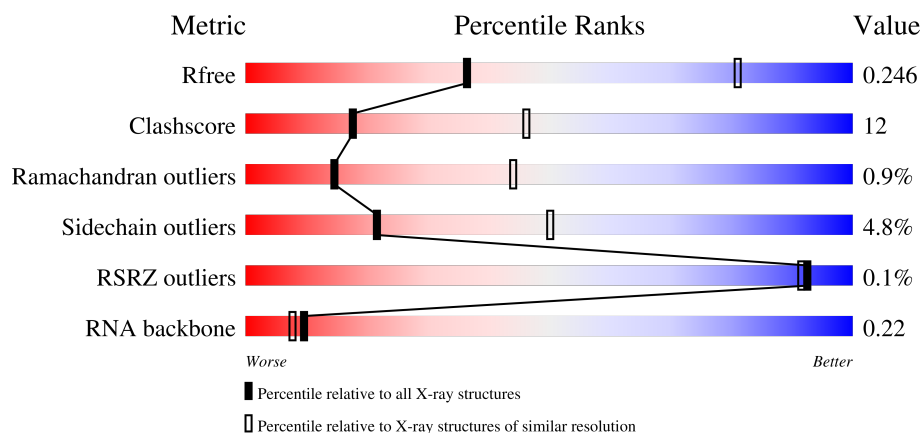
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.30 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.






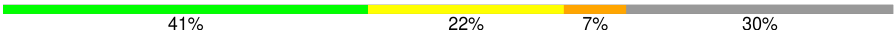

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1085 (3.32-3.28)
Clashscore	180529	1128 (3.32-3.28)
Ramachandran outliers	177936	1125 (3.32-3.28)
Sidechain outliers	177891	1124 (3.32-3.28)
RSRZ outliers	164620	1085 (3.32-3.28)
RNA backbone	3690	1014 (3.64-2.96)

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 $\leq 5\%$. 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	315	
1	B	315	
2	C	1119	
3	D	1524	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	 71% 22% • 5%
5	F	423	 56% 24% • 18%
6	G	22	 32% 23% 32% 14%
7	H	27	 41% 22% 7% 30%
8	I	8	 25% 25% 50%

2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 28534 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 alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	226	Total	C	N	O	S	0	0	0
			1782	1138	310	332	2			
1	B	224	Total	C	N	O	S	0	0	0
			1767	1129	307	329	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1107	Total	C	N	O	S	0	0	0
			8726	5523	1551	1628	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1484	Total	C	N	O	S	0	0	0
			11722	7431	2065	2191	35			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	94	Total	C	N	O	S	0	0	0
			761	486	132	139	4			

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	346	Total	C	N	O	S	0	0	0
			2790	1760	508	518	4			

- Molecule 6 is a DNA chain called DNA (5'-D(P*GP*GP*TP*GP*CP*AP*TP*CP*AP*GP*AP*GP*CP*CP*CP*AP*AP*AP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	19	Total	C	N	O	P	0	0	0
			386	183	75	109	19			

- Molecule 7 is a DNA chain called DNA (5'-D(*TP*AP*TP*AP*AP*TP*GP*GP*GP*CP*TP*GP*AP*TP*GP*CP*AP*CP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	19	Total	C	N	O	P	0	0	0
			394	188	76	112	18			

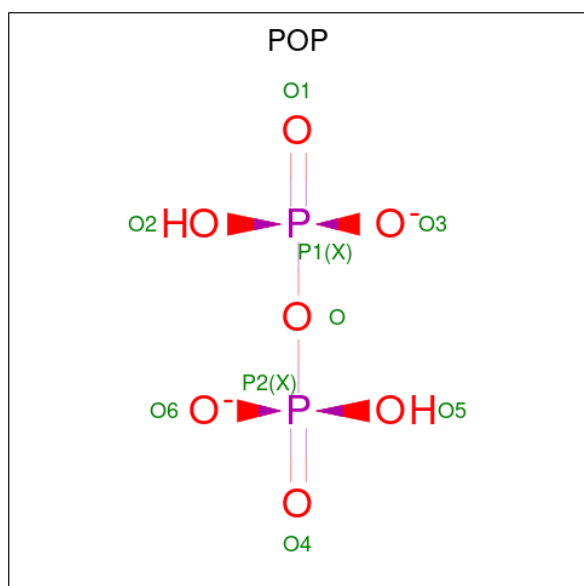
- Molecule 8 is a RNA chain called RNA (5'-D(*(GTP))-R(P*GP*GP*GP*GP*GP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	8	Total	C	N	O	P	0	0	0
			193	80	40	63	10			

- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	B	1	Total	Mg	0	0
			1	1		
9	D	1	Total	Mg	0	0
			1	1		

- Molecule 10 is PYROPHOSPHATE 2- (three-letter code: POP) (formula: H₂O₇P₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
10	C	1	Total	O	P	0	0
			9	7	2		

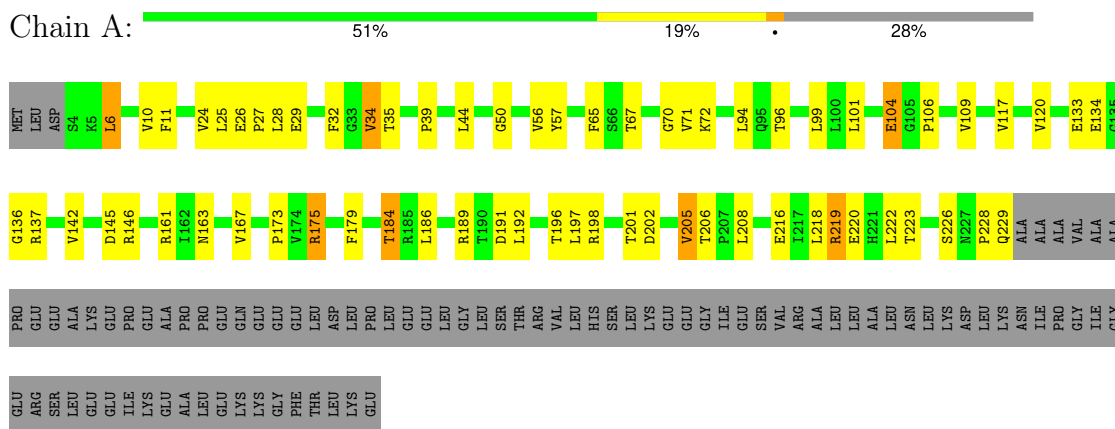
- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	D	2	Total	Zn	0	0
			2	2		

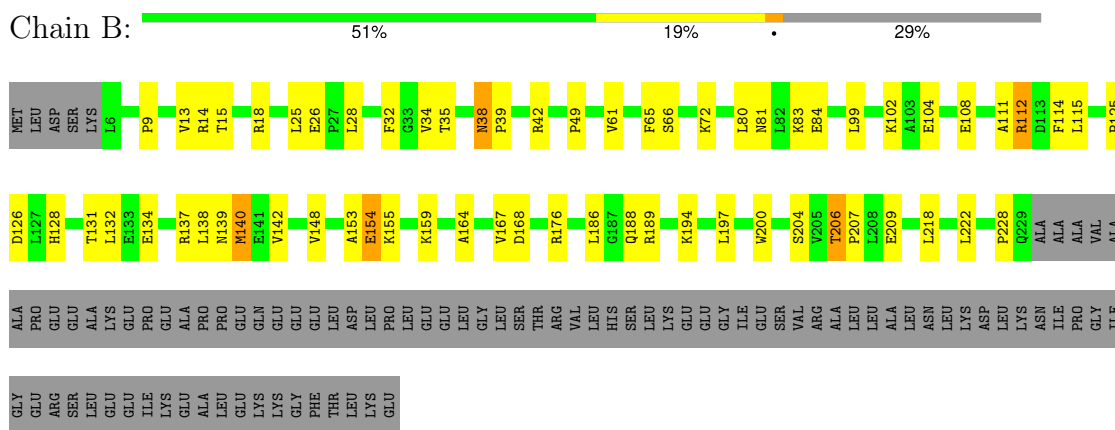
3 Residue-property plots

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.

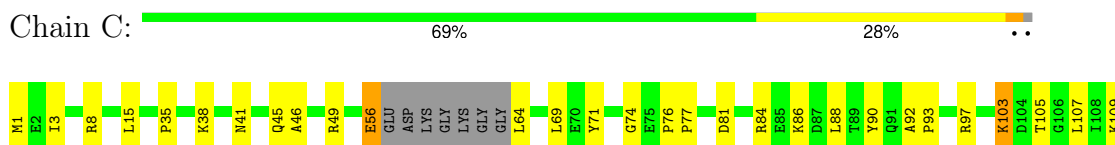
• Molecule 1: DNA-directed RNA polymerase subunit alpha

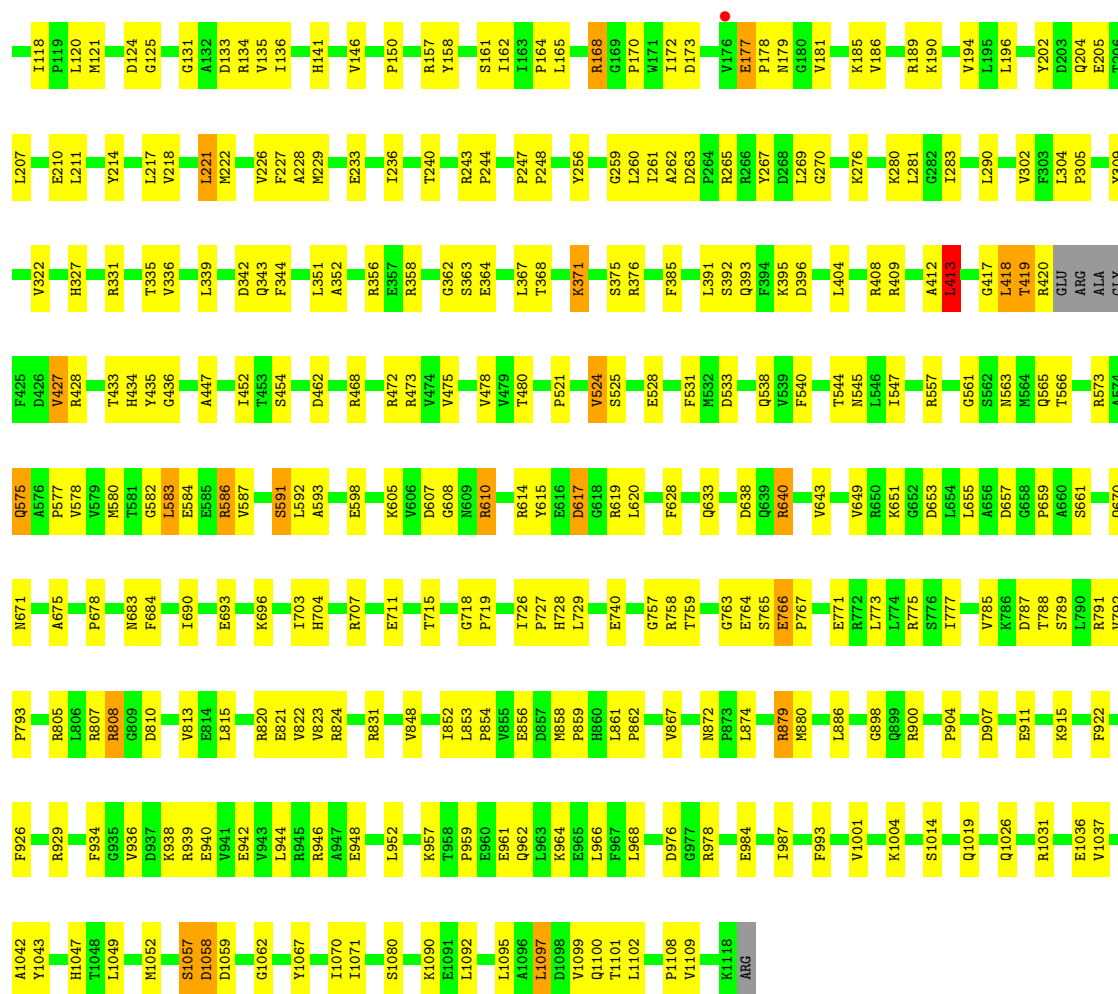


• Molecule 1: DNA-directed RNA polymerase subunit alpha



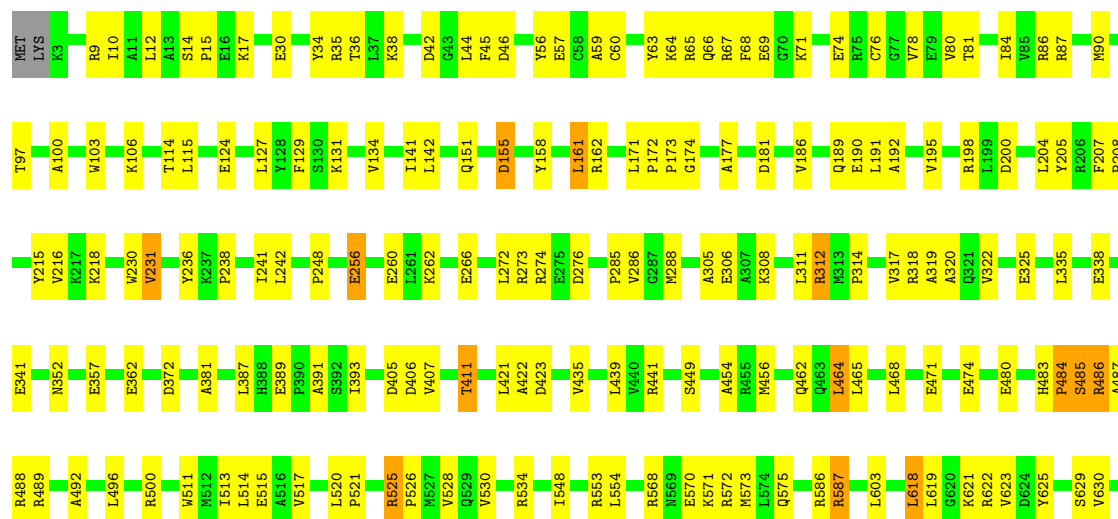
• Molecule 2: DNA-directed RNA polymerase subunit beta

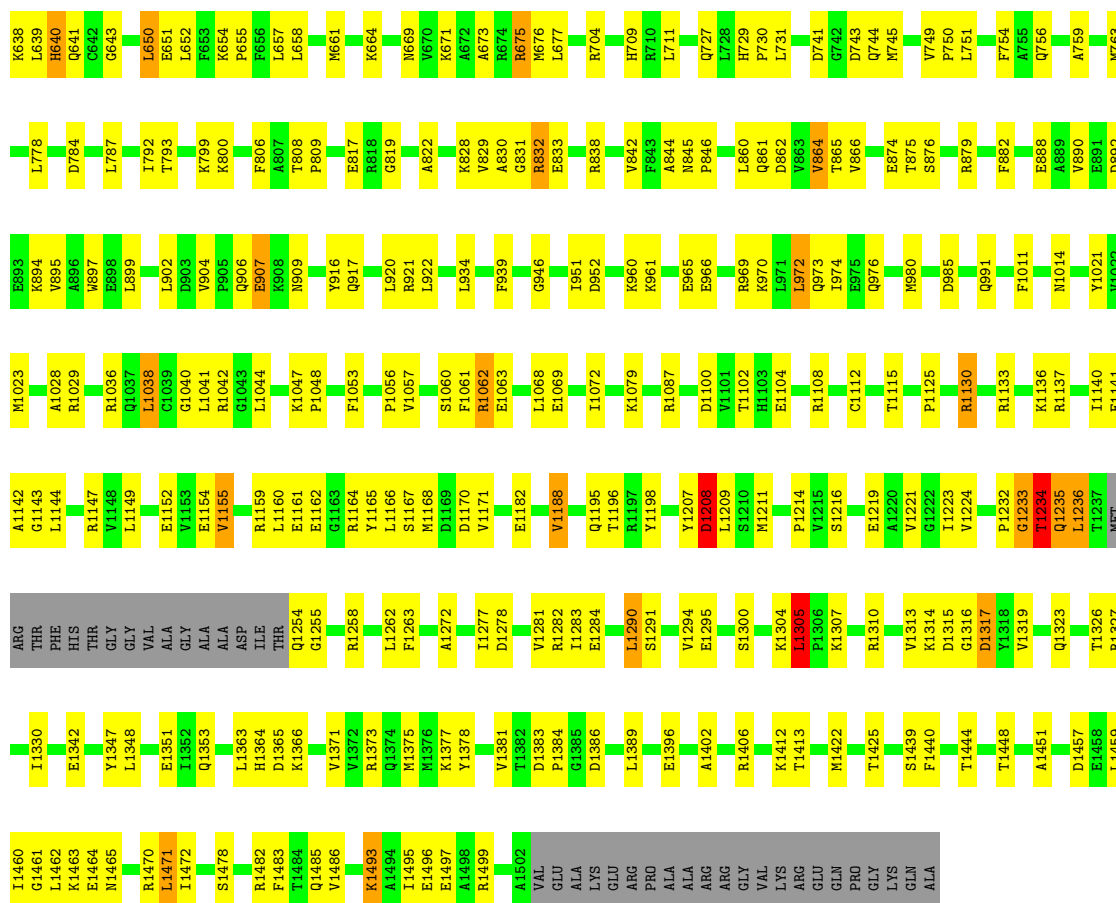




• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D: 68% 27%





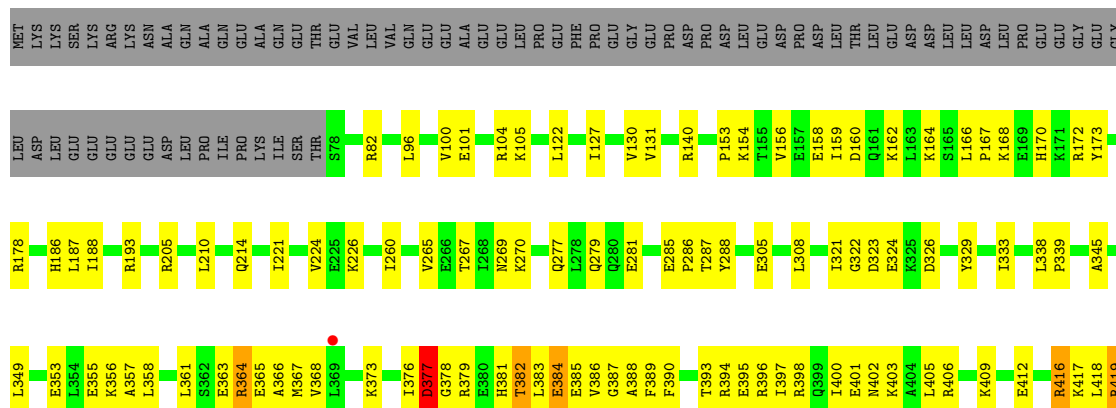
• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E: 71% 22% 5%



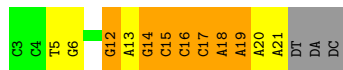
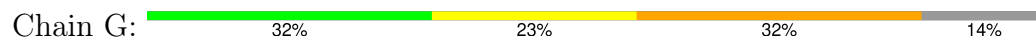
• Molecule 5: RNA polymerase sigma factor SigA

Chain F: 56% 24% 18%

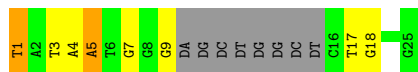




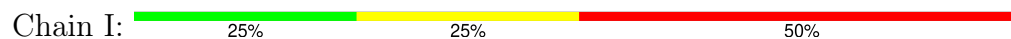
- Molecule 6: DNA (5'-D(P*GP*GP*TP*GP*CP*AP*TP*CP*AP*GP*AP*GP*CP*CP*CP*A P*AP*AP*A)-3')



- Molecule 7: DNA (5'-D(*TP*AP*TP*AP*AP*TP*GP*GP*GP*CP*TP*GP*AP*TP*GP*CP *AP*CP*C)-3')



- Molecule 8: RNA (5'-D(*(GTP))-R(P*GP*GP*GP*GP*GP*G)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	186.54Å 101.98Å 296.11Å 90.00° 98.56° 90.00°	Depositor
Resolution (Å)	29.88 – 3.30 29.88 – 3.30	Depositor EDS
% Data completeness (in resolution range)	94.2 (29.88-3.30) 94.1 (29.88-3.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.75 (at 3.31Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, R_{free}	0.185 , 0.246 0.186 , 0.246	Depositor DCC
R_{free} test set	81007 reflections (2.56%)	wwPDB-VP
Wilson B-factor (Å ²)	89.9	Xtriage
Anisotropy	0.773	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 90.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	28534	wwPDB-VP
Average B, all atoms (Å ²)	123.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.36% 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: GTP, ZN, MG, POP

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	0.58	1/1814 (0.1%)	0.82	1/2466 (0.0%)
1	B	0.60	2/1799 (0.1%)	0.90	1/2447 (0.0%)
2	C	0.53	0/8892	0.81	4/12028 (0.0%)
3	D	0.59	0/11928	0.86	13/16127 (0.1%)
4	E	0.50	0/775	0.76	0/1045
5	F	0.50	0/2835	0.81	1/3816 (0.0%)
6	G	1.35	4/433 (0.9%)	1.30	8/664 (1.2%)
7	H	1.03	1/442 (0.2%)	1.03	3/680 (0.4%)
8	I	1.06	0/181	2.00	10/283 (3.5%)
All	All	0.60	8/29099 (0.0%)	0.86	41/39556 (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	C	0	1
5	F	0	1
All	All	0	2

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	154	GLU	CB-CG	7.78	1.67	1.52
1	B	154	GLU	CG-CD	6.08	1.61	1.51
6	G	19	DA	P-O5'	6.06	1.65	1.59
7	H	1	DT	C1'-N1	5.55	1.56	1.49
6	G	12	DG	C3'-O3'	-5.36	1.36	1.44

The worst 5 of 41 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	G	17	DC	O4'-C4'-C3'	-9.24	100.46	106.00
6	G	19	DA	O4'-C4'-C3'	-9.14	100.52	106.00
6	G	16	DC	O5'-P-OP2	-8.29	98.23	105.70
2	C	879	ARG	NE-CZ-NH2	-7.70	116.45	120.30
3	D	1363	LEU	CB-CG-CD2	-7.02	99.06	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	766	GLU	Peptide
5	F	389	PHE	Peptide

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	1782	0	1834	44	0
1	B	1767	0	1816	45	0
2	C	8726	0	8814	243	0
3	D	11722	0	11950	309	1
4	E	761	0	778	21	0
5	F	2790	0	2854	79	0
6	G	386	0	212	14	0
7	H	394	0	217	8	0
8	I	193	0	88	8	0
9	B	1	0	0	0	0
9	D	1	0	0	0	0
10	C	9	0	0	0	0
11	D	2	0	0	0	0
All	All	28534	0	28563	694	1

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 12.

The worst 5 of 694 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)
3:D:218:LYS:HG2	3:D:338:GLU:HG2	1.48	0.92
2:C:367:LEU:HA	2:C:371:LYS:HE2	1.49	0.92
2:C:418:LEU:O	2:C:419:THR:OG1	1.88	0.91
5:F:358:LEU:HB3	5:F:366:ALA:HB1	1.49	0.90
3:D:238:PRO:HD3	3:D:318:ARG:HG3	1.53	0.89

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:985:ASP:N	3:D:1497:GLU:OE1[1_545]	2.11	0.09

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	224/315 (71%)	219 (98%)	5 (2%)	0	100	100
1	B	222/315 (70%)	210 (95%)	11 (5%)	1 (0%)	25	56
2	C	1101/1119 (98%)	1065 (97%)	29 (3%)	7 (1%)	22	53
3	D	1480/1524 (97%)	1439 (97%)	30 (2%)	11 (1%)	19	50
4	E	92/99 (93%)	89 (97%)	2 (2%)	1 (1%)	12	40
5	F	344/423 (81%)	316 (92%)	18 (5%)	10 (3%)	3	22
All	All	3463/3795 (91%)	3338 (96%)	95 (3%)	30 (1%)	14	44

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	363	SER
2	C	418	LEU
2	C	419	THR
3	D	484	PRO

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	D	486	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	199/273 (73%)	188 (94%)	11 (6%)	18	45
1	B	197/273 (72%)	187 (95%)	10 (5%)	20	48
2	C	931/941 (99%)	888 (95%)	43 (5%)	23	52
3	D	1250/1279 (98%)	1182 (95%)	68 (5%)	18	46
4	E	83/88 (94%)	80 (96%)	3 (4%)	30	57
5	F	296/371 (80%)	290 (98%)	6 (2%)	50	71
All	All	2956/3225 (92%)	2815 (95%)	141 (5%)	21	50

5 of 141 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1130	ARG
3	D	1195	GLN
3	D	1493	LYS
2	C	617	ASP
2	C	610	ARG

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

Mol	Chain	Res	Type
3	D	640	HIS
3	D	744	GLN
5	F	411	HIS
2	C	683	ASN
2	C	390	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	8/8 (100%)	3 (37%)	3 (37%)

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	I	3	G
8	I	4	G
8	I	5	G

All (3) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
8	I	1	GTP
8	I	2	G
8	I	4	G

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 5 ligands modelled in this entry, 4 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	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
10	POP	C	1201	-	6,8,8	0.62	0	12,13,13	1.00	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
10	POP	C	1201	-	-	0/6/6/6	-

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 [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	226/315 (71%)	-0.85	0 100 100	86, 127, 163, 192	0
1	B	224/315 (71%)	-0.87	0 100 100	73, 114, 162, 206	0
2	C	1107/1119 (98%)	-0.84	1 (0%) 92 91	47, 116, 201, 261	0
3	D	1484/1524 (97%)	-0.91	0 100 100	46, 105, 183, 273	0
4	E	94/99 (94%)	-0.90	0 100 100	67, 121, 193, 216	0
5	F	346/423 (81%)	-0.66	1 (0%) 90 85	82, 136, 247, 282	0
6	G	19/22 (86%)	-0.67	0 100 100	68, 117, 257, 289	0
7	H	19/27 (70%)	-0.92	0 100 100	119, 141, 268, 271	0
8	I	7/8 (87%)	-0.46	0 100 100	88, 104, 219, 238	0
All	All	3526/3852 (91%)	-0.85	2 (0%) 92 91	46, 117, 199, 289	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	369	LEU	2.4
2	C	176	VAL	2.3

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, 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	B-factors(\AA^2)	Q<0.9
10	POP	C	1201	9/9	0.81	0.07	94,111,146,148	0
9	MG	B	1001	1/1	0.86	0.11	80,80,80,80	0
9	MG	D	2001	1/1	0.99	0.05	46,46,46,46	0
11	ZN	D	2003	1/1	0.99	0.11	123,123,123,123	0
11	ZN	D	2002	1/1	1.00	0.02	134,134,134,134	0

6.5 Other polymers [i](#)

There are no such residues in this entry.