



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 4, 2023 – 02:09 AM EDT

PDB ID : 6OY7
Title : X-ray crystal structure of a bacterial reiterative transcription complex of pyrG promoter at 7 min
Authors : Shin, Y.; Murakami, K.S.
Deposited on : 2019-05-14
Resolution : 3.04 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 : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.35.1
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) : 2.35.1

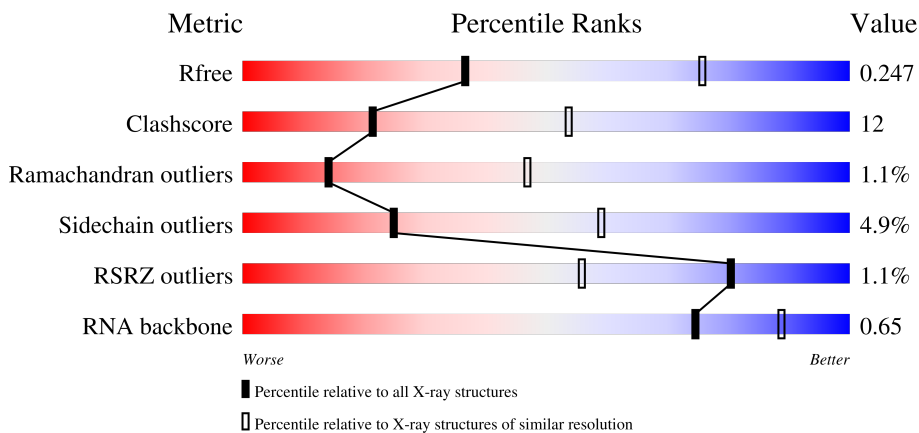
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.04 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	2752 (3.08-3.00)
Clashscore	141614	3096 (3.08-3.00)
Ramachandran outliers	138981	2986 (3.08-3.00)
Sidechain outliers	138945	2988 (3.08-3.00)
RSRZ outliers	127900	2636 (3.08-3.00)
RNA backbone	3102	1034 (3.30-2.78)

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	 52% 17% 28%
1	B	315	 51% 19% 29%
2	C	1119	 69% 28%
3	D	1524	 71% 25%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	<p>%</p> <p>77% 17% 5%</p>
5	F	423	<p>4%</p> <p>56% 23% 18%</p>
6	G	22	<p>5%</p> <p>23% 50% 14% 14%</p>
7	H	27	<p>4%</p> <p>22% 48% 26%</p>
8	I	6	<p>17%</p> <p>17% 33% 33% 17%</p>

2 Entry composition [i](#)

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
			Total	C	N	O	S			
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
			Total	C	N	O	S			
2	C	1111	Total	C	N	O	S	0	0	0
			8745	5534	1555	1632	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	1485	Total	C	N	O	S	0	0	0
			11729	7435	2066	2193	35			

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

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

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	46	THR	ALA	conflict	UNP Q72L95

- Molecule 6 is a DNA chain called DNA (5'-D(P*CP*CP*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
			Total	C	N	O	P			
6	G	19	386	183	75	109	19	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
7	H	20	414	198	78	119	19	0	0	0

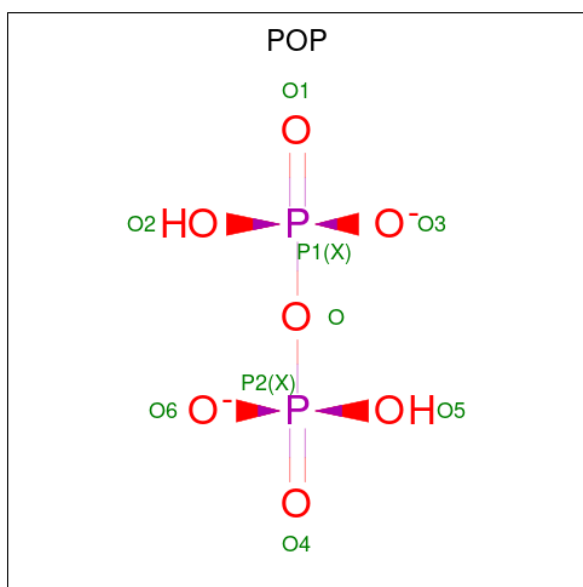
- Molecule 8 is a RNA chain called RNA (5'-D(*(GTP))-R(P*GP*GP*GP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
8	I	6	147	60	30	49	8	0	0	0

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

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

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



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

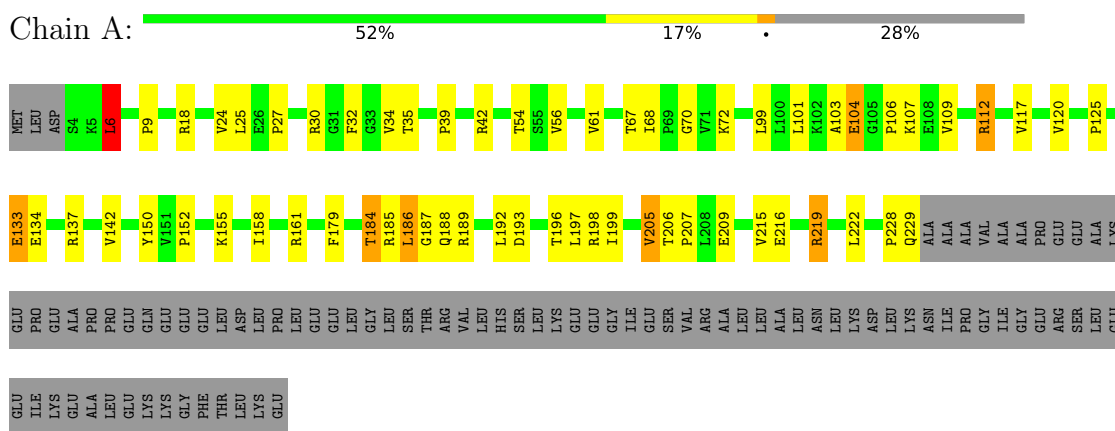
- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

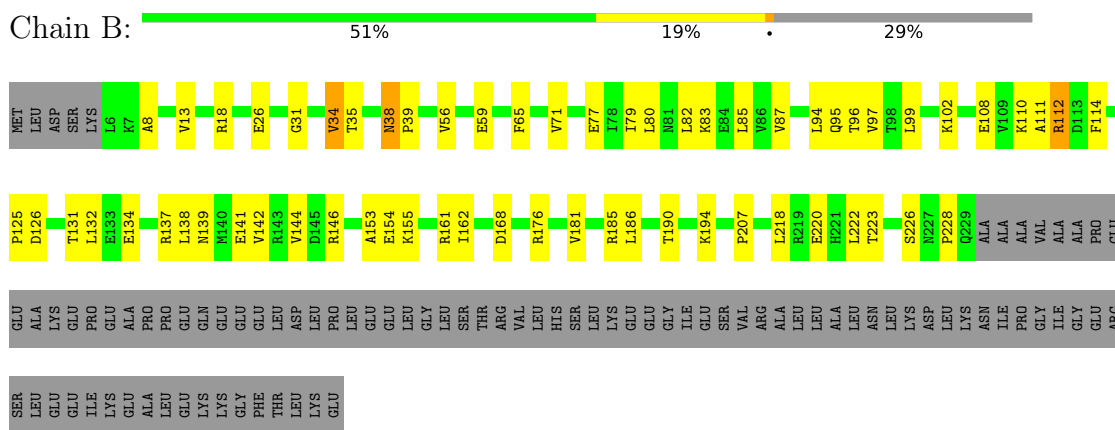
3 Residue-property plots [i](#)

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.

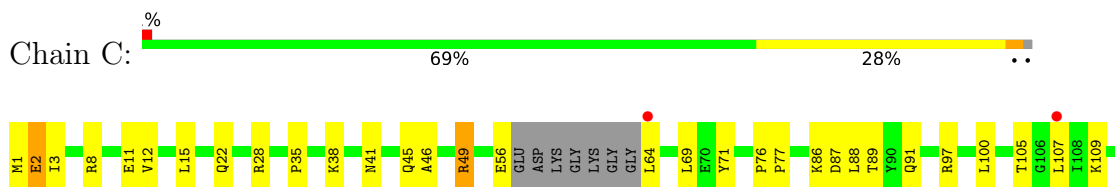
- Molecule 1: DNA-directed RNA polymerase subunit alpha

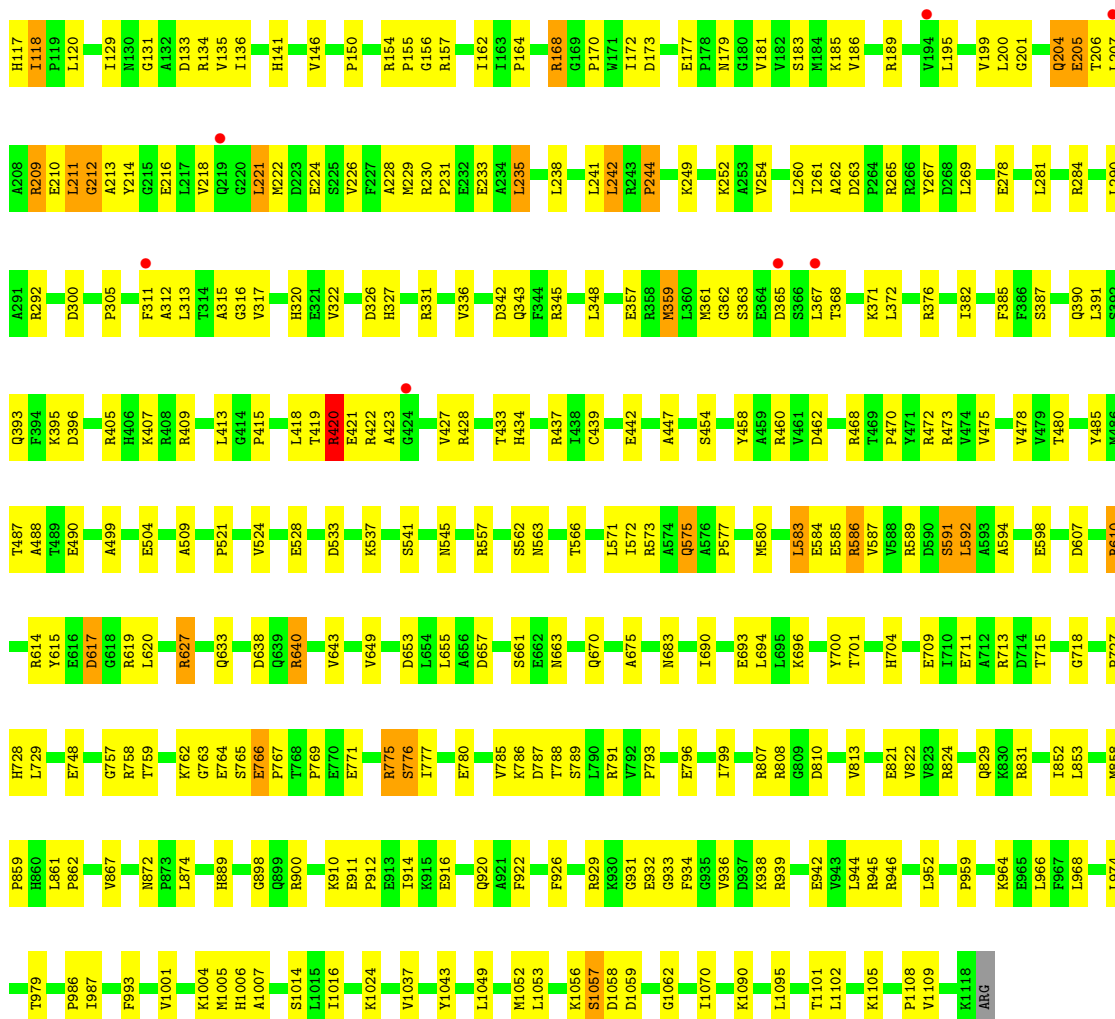


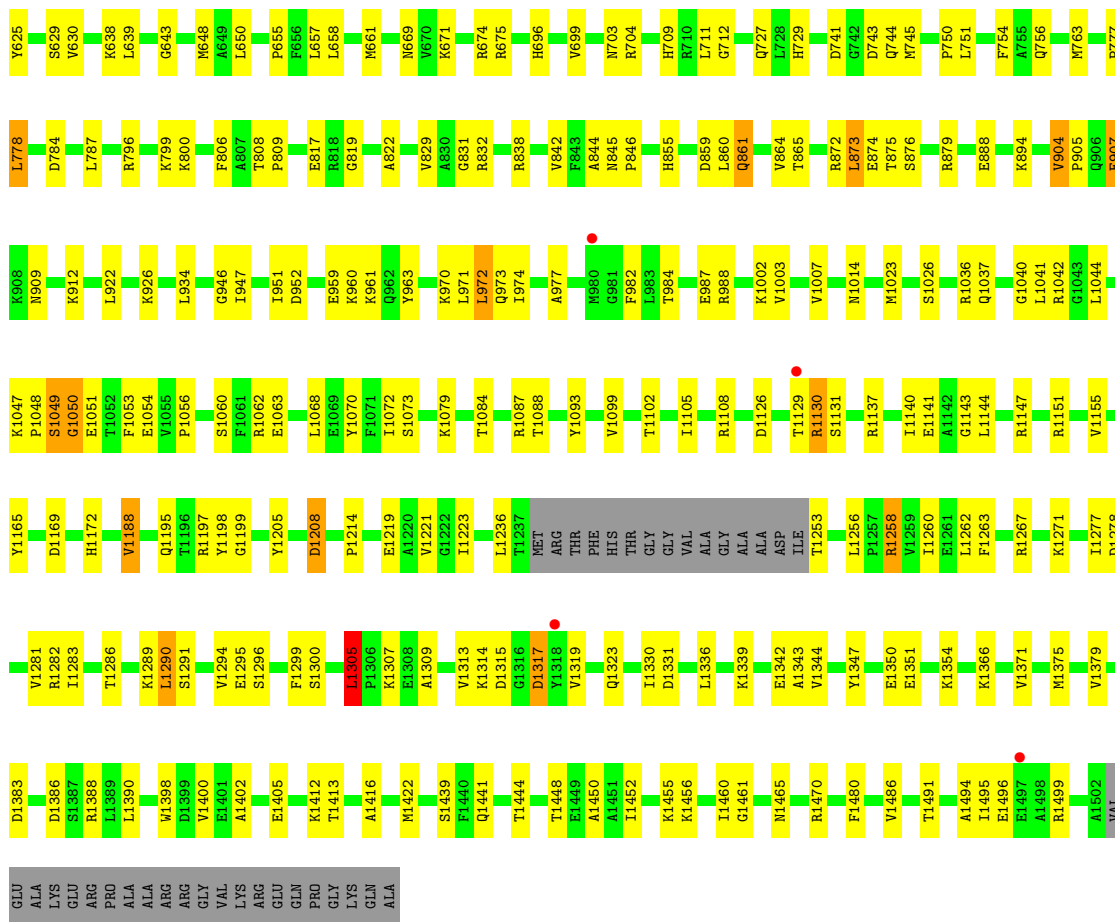
- Molecule 1: DNA-directed RNA polymerase subunit alpha



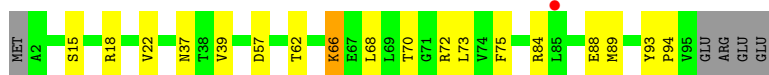
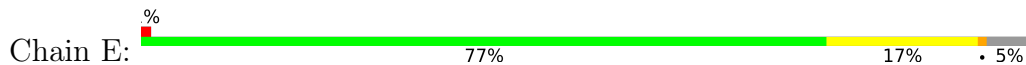
- Molecule 2: DNA-directed RNA polymerase subunit beta



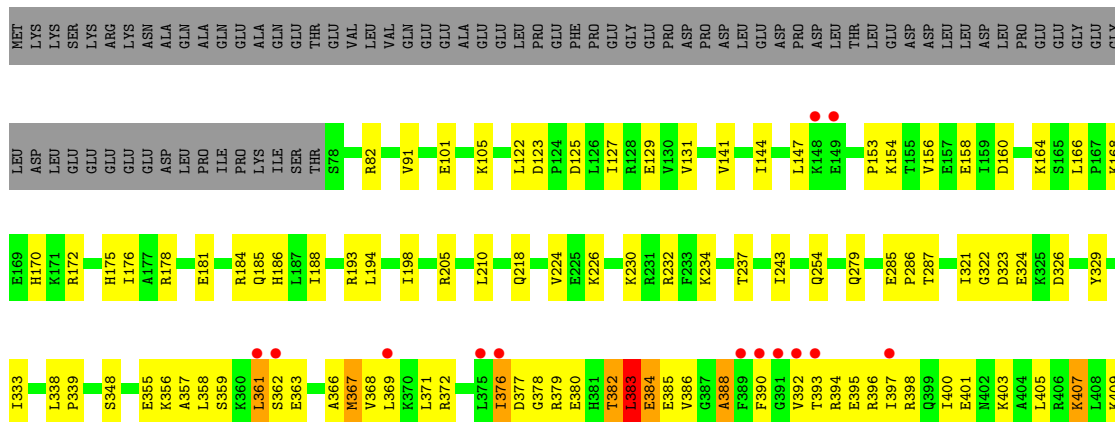


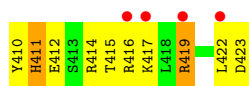


• Molecule 4: DNA-directed RNA polymerase subunit omega

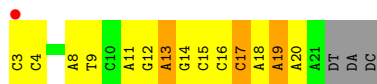
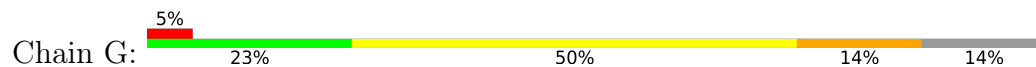


• Molecule 5: RNA polymerase sigma factor SigA

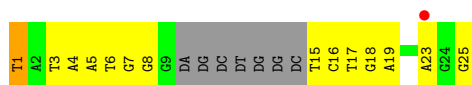
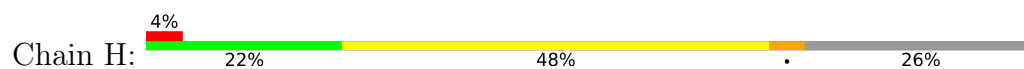




- Molecule 6: DNA (5'-D(P*CP*CP*TP*GP*CP*AP*TP*CP*AP*GP*AP*GP*CP*CP*CP*AP*AP*AP*A)-3')



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



- Molecule 8: RNA (5'-D*(GTP))-R(P*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	185.68Å 101.61Å 295.80Å 90.00° 98.76° 90.00°	Depositor
Resolution (Å)	47.99 – 3.04 47.99 – 3.04	Depositor EDS
% Data completeness (in resolution range)	94.9 (47.99-3.04) 94.9 (47.99-3.04)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.38 (at 3.07Å)	Xtrriage
Refinement program	PHENIX (1.14_3260)	Depositor
R, R_{free}	0.207 , 0.247 0.207 , 0.247	Depositor DCC
R_{free} test set	1993 reflections (1.99%)	wwPDB-VP
Wilson B-factor (Å ²)	88.5	Xtrriage
Anisotropy	0.572	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 42.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	28534	wwPDB-VP
Average B, all atoms (Å ²)	94.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.20% 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 i

5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: MG, POP, GTP, 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).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.58	0/1814	0.89	2/2466 (0.1%)
1	B	0.55	0/1799	0.82	0/2447
2	C	0.55	0/8912	0.82	5/12057 (0.0%)
3	D	0.58	0/11935	0.86	6/16137 (0.0%)
4	E	0.55	0/775	0.80	0/1045
5	F	0.55	0/2835	0.92	6/3816 (0.2%)
6	G	1.45	2/433 (0.5%)	1.33	8/664 (1.2%)
7	H	1.13	1/464 (0.2%)	1.24	2/714 (0.3%)
8	I	1.14	0/129	1.67	4/201 (2.0%)
All	All	0.60	3/29096 (0.0%)	0.88	33/39547 (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	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	G	19	DA	N9-C4	6.14	1.41	1.37
6	G	19	DA	P-O5'	5.88	1.65	1.59
7	H	1	DT	C1'-N1	5.49	1.56	1.49

The worst 5 of 33 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	F	378	GLY	N-CA-C	-10.14	87.75	113.10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	G	19	DA	O4'-C1'-N9	9.01	114.30	108.00
5	F	383	LEU	CA-CB-CG	8.78	135.49	115.30
6	G	17	DC	O4'-C4'-C3'	-8.73	100.76	106.00
2	C	235	LEU	CA-CB-CG	8.10	133.94	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	420	ARG	Peptide
2	C	766	GLU	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	46	0
1	B	1767	0	1816	48	0
2	C	8745	0	8828	235	0
3	D	11729	0	11957	287	1
4	E	761	0	778	15	0
5	F	2790	0	2854	103	0
6	G	386	0	212	15	0
7	H	414	0	229	13	0
8	I	147	0	66	24	0
9	B	1	0	0	0	0
9	D	1	0	0	0	0
10	D	9	0	0	0	0
11	D	2	0	0	0	0
All	All	28534	0	28574	682	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 682 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)
8:I:3:GTP:N3	8:I:4:G:N2	1.81	1.28
8:I:3:GTP:C2	8:I:4:G:N2	2.16	1.12
3:D:218:LYS:HG2	3:D:338:GLU:HG2	1.45	0.97
2:C:946:ARG:HH12	3:D:861:GLN:HE22	1.07	0.92
2:C:12:VAL:HG21	2:C:472:ARG:HD3	1.53	0.90

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:34:TYR:OH	3:D:327:GLU:OE1[4_1359]	2.13	0.07

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	224/315 (71%)	218 (97%)	6 (3%)	0	100	100
1	B	222/315 (70%)	207 (93%)	13 (6%)	2 (1%)	17	52
2	C	1107/1119 (99%)	1055 (95%)	36 (3%)	16 (1%)	11	40
3	D	1481/1524 (97%)	1431 (97%)	40 (3%)	10 (1%)	22	57
4	E	92/99 (93%)	89 (97%)	2 (2%)	1 (1%)	14	47
5	F	344/423 (81%)	311 (90%)	24 (7%)	9 (3%)	5	24
All	All	3470/3795 (91%)	3311 (95%)	121 (4%)	38 (1%)	14	47

5 of 38 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	212	GLY
2	C	316	GLY
2	C	363	SER
2	C	421	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	C	932	GLU

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	199/273 (73%)	185 (93%)	14 (7%)	15	44
1	B	197/273 (72%)	190 (96%)	7 (4%)	35	68
2	C	931/941 (99%)	889 (96%)	42 (4%)	27	62
3	D	1251/1279 (98%)	1183 (95%)	68 (5%)	22	55
4	E	83/88 (94%)	80 (96%)	3 (4%)	35	68
5	F	296/371 (80%)	286 (97%)	10 (3%)	37	70
All	All	2957/3225 (92%)	2813 (95%)	144 (5%)	25	59

5 of 144 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1151	ARG
5	F	419	ARG
3	D	1195	GLN
3	D	1496	GLU
2	C	610	ARG

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

Mol	Chain	Res	Type
3	D	861	GLN
3	D	1037	GLN
5	F	269	ASN
5	F	83	GLN
5	F	218	GLN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	5/6 (83%)	1 (20%)	1 (20%)

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	I	5	G

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
8	I	4	G

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 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	D	1601	-	6,8,8	0.81	0	13,13,13	1.15	1 (7%)

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	D	1601	-	-	0/6/6/6	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
10	D	1601	POP	P2-O-P1	-2.99	122.58	132.83

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%)	-0.35	0 100 100	66, 89, 110, 118	0
1	B	224/315 (71%)	-0.39	0 100 100	65, 93, 117, 134	0
2	C	1111/1119 (99%)	-0.22	9 (0%) 86 65	51, 90, 144, 167	0
3	D	1485/1524 (97%)	-0.29	8 (0%) 91 75	47, 83, 135, 170	0
4	E	94/99 (94%)	-0.34	1 (1%) 80 56	62, 98, 130, 138	0
5	F	346/423 (81%)	-0.10	17 (4%) 29 11	60, 98, 175, 188	0
6	G	19/22 (86%)	-0.15	1 (5%) 26 9	59, 107, 212, 218	0
7	H	20/27 (74%)	-0.34	1 (5%) 28 10	86, 122, 223, 225	0
8	I	5/6 (83%)	0.36	1 (20%) 1 0	62, 68, 90, 154	0
All	All	3530/3850 (91%)	-0.26	38 (1%) 80 56	47, 89, 146, 225	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	392	VAL	6.7
5	F	391	GLY	4.9
5	F	389	PHE	4.4
5	F	393	THR	4.0
5	F	390	PHE	4.0

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, 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	D	1601	9/9	0.95	0.29	73,86,104,112	0
9	MG	B	1001	1/1	0.97	0.41	66,66,66,66	0
11	ZN	D	1603	1/1	0.98	0.12	111,111,111,111	0
11	ZN	D	1604	1/1	0.98	0.22	84,84,84,84	0
9	MG	D	1602	1/1	0.99	0.29	52,52,52,52	0

6.5 Other polymers [i](#)

There are no such residues in this entry.