wwPDB X-ray Structure Validation Summary Report (i)

Feb 21, 2024 – 04:57 AM EST

PDB ID	:	4Q5S
Title	:	Thermus thermophilus RNA polymerase initially transcribing complex containing 6-mer RNA
Authors	:	Murakami, K.S.
Deposited on	:	2014-04-17
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
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:

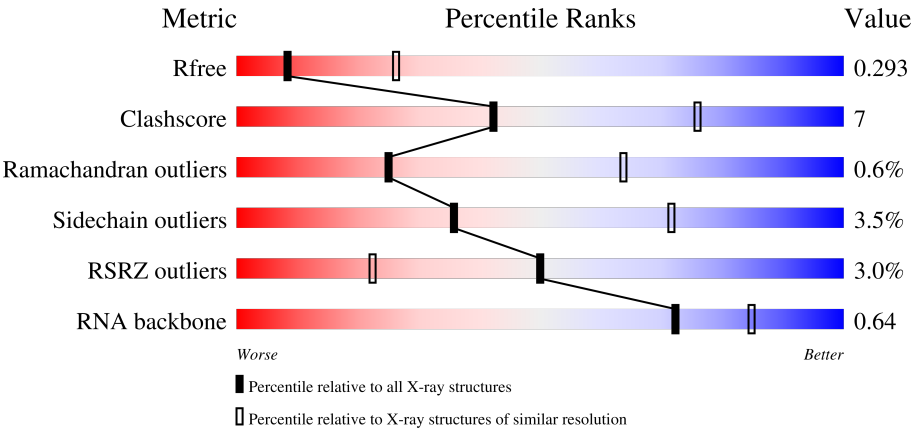
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (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 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.



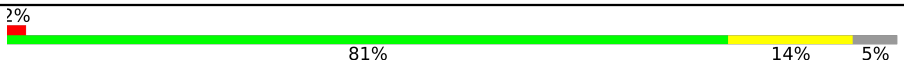
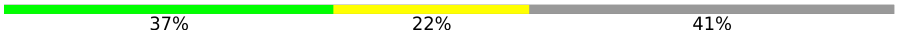
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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	315	59% 12% 28%
1	B	315	60% 9% 30%
2	C	1119	3% 79% 20%
3	D	1524	2% 78% 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	
5	F	423	
6	G	22	
7	H	27	
8	I	5	

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 28290 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	222	Total	C	N	O	S	0	0	0
			1750	1118	304	326	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1111	Total	C	N	O	S	0	0	0
			8770	5548	1564	1634	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	7430	2064	2192	36			

- 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	338	Total	C	N	O	S	0	0	0
			2747	1736	500	507	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	14	Total	C	N	O	P	0	0	0
			289	137	52	86	14			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	16	Total	C	N	O	P	0	0	0
			333	159	66	93	15			

- Molecule 8 is a RNA chain called RNA (5'-R(P*CP*UP*CP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	5	Total	C	N	O	P	0	0	0
			102	46	16	35	5			

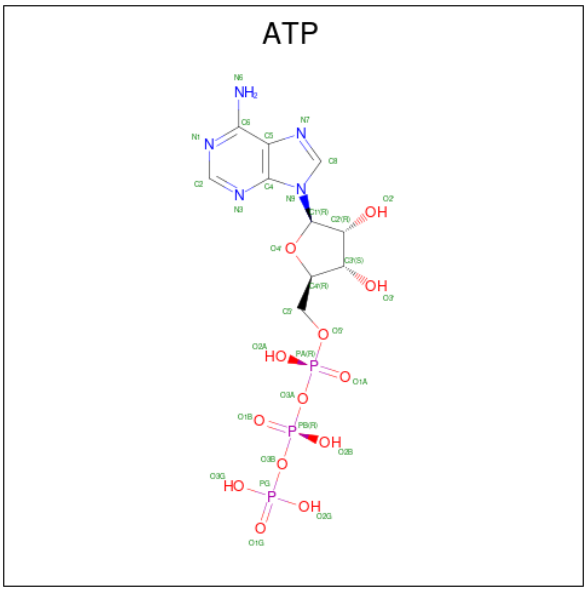
- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 10 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 11 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: $C_{10}H_{16}N_5O_{13}P_3$).



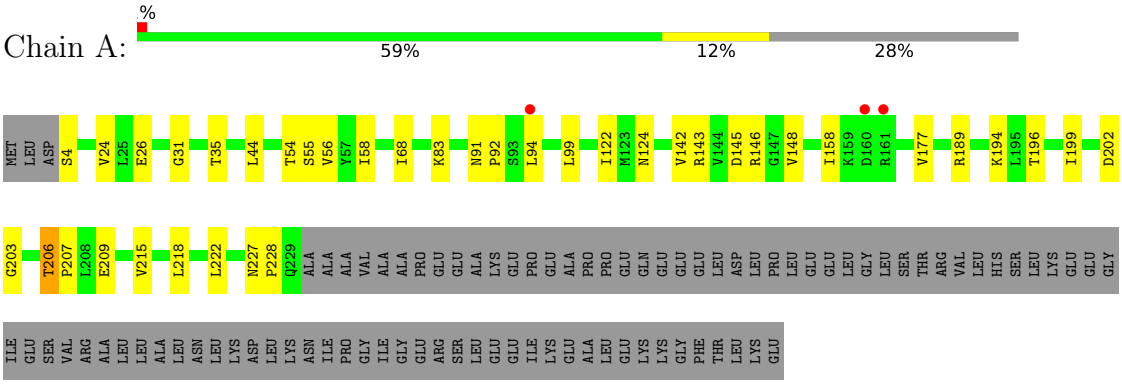
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
11	I	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

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.

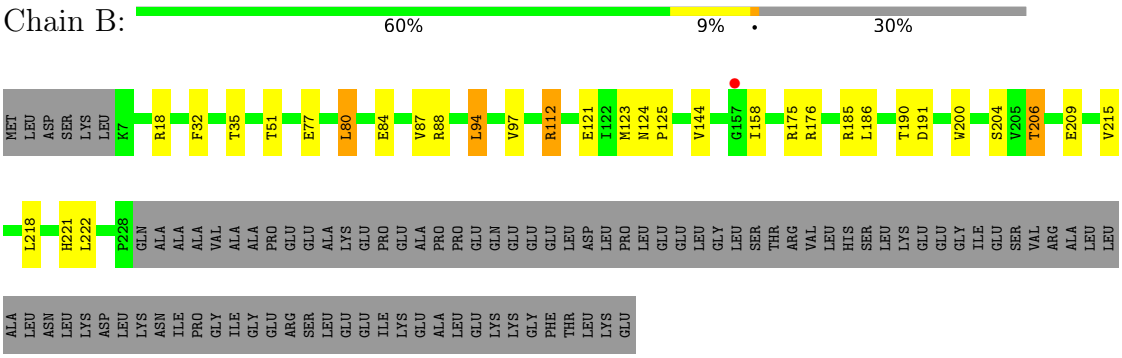
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



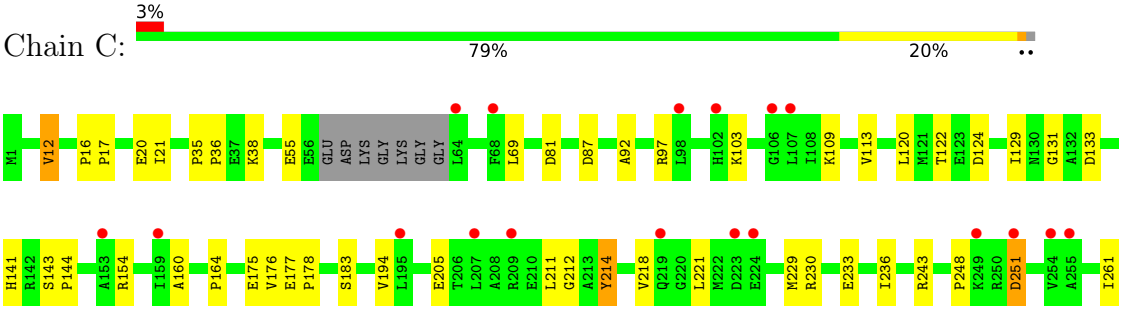
• Molecule 1: DNA-directed RNA polymerase subunit alpha

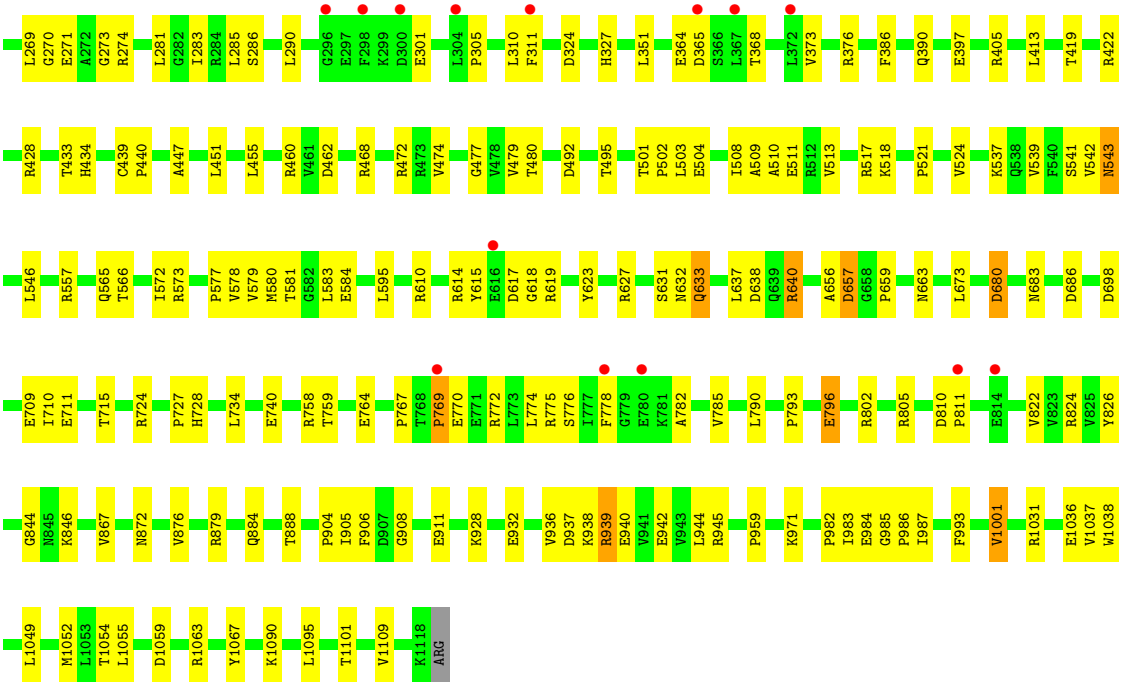
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

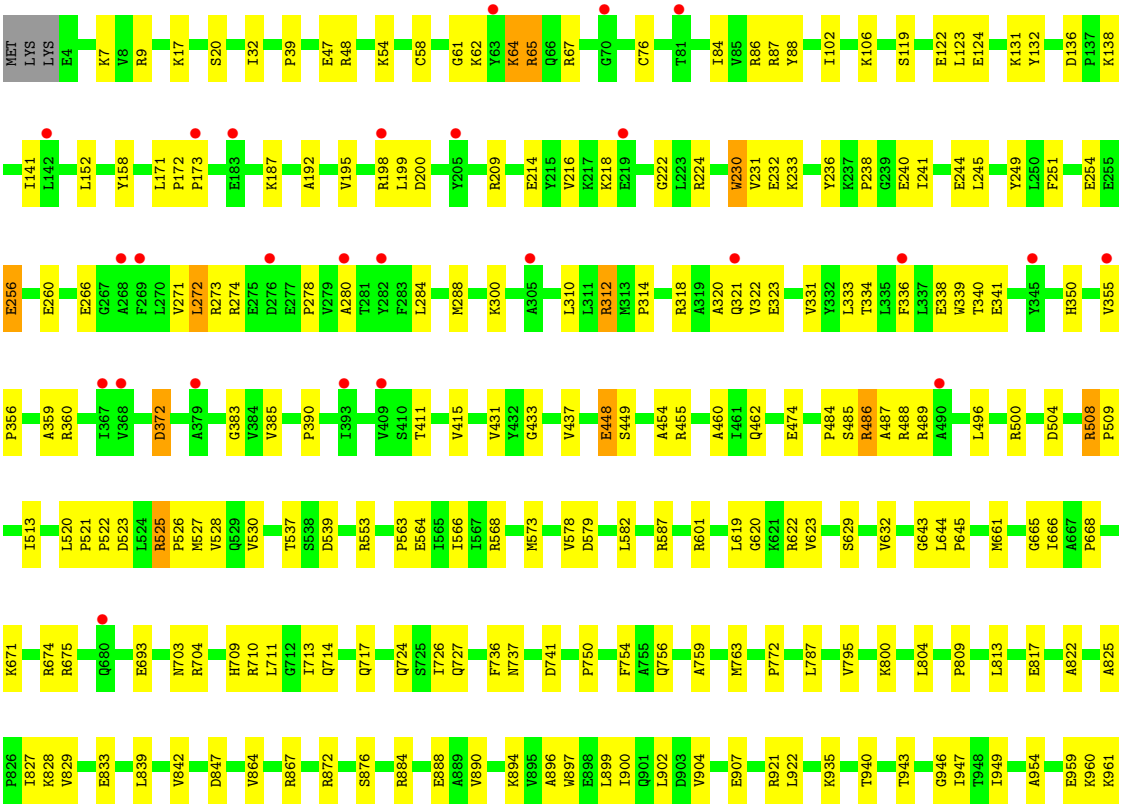
Chain C:

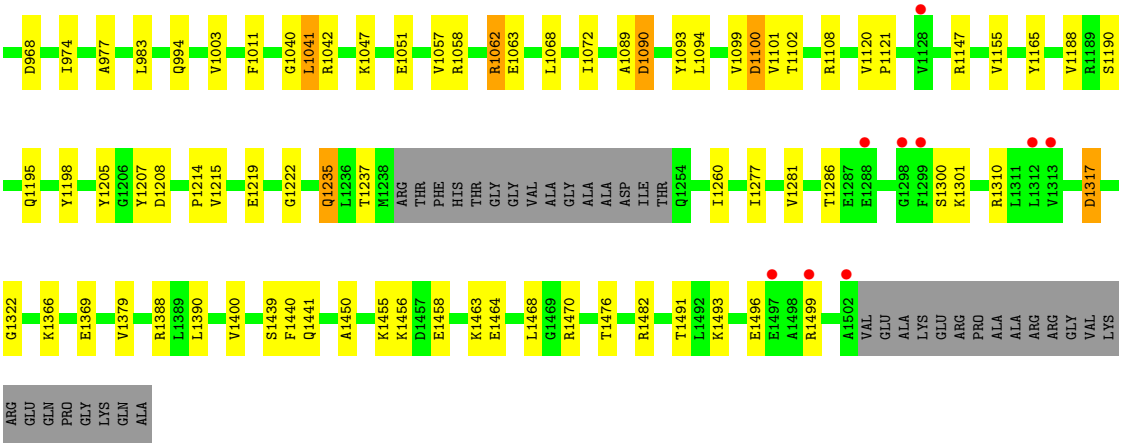




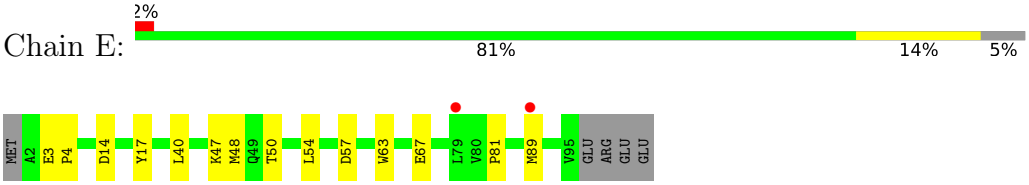
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:

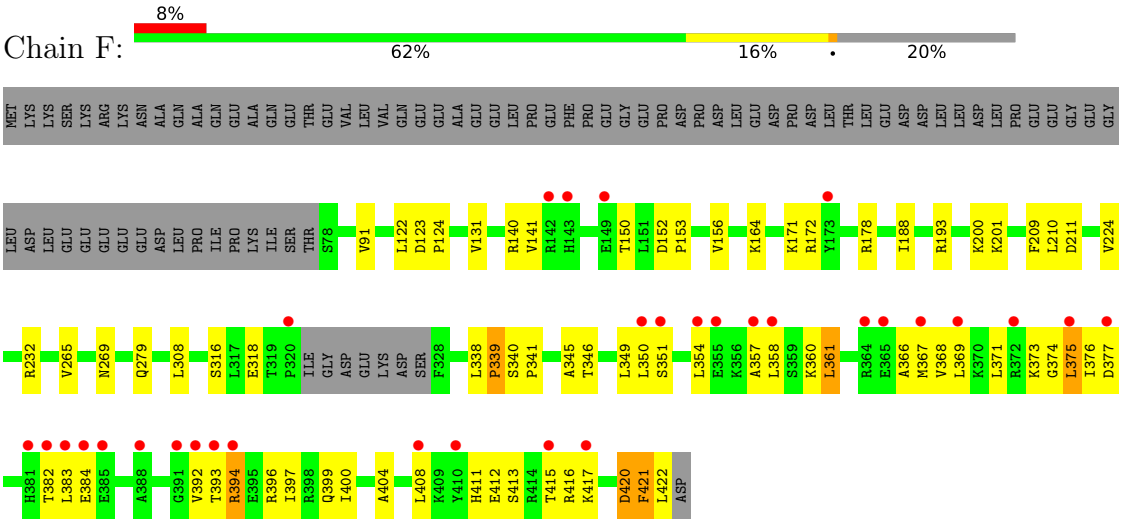




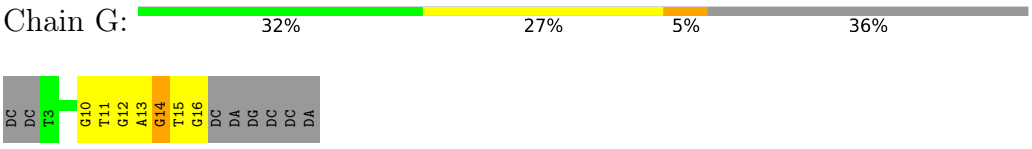
• Molecule 4: DNA-directed RNA polymerase subunit omega



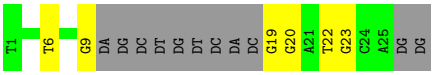
• Molecule 5: RNA polymerase sigma factor SigA



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



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



• Molecule 8: RNA (5'-R(P*CP*UP*CP*AP*C)-3')

Chain I:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	187.09Å 102.08Å 297.26Å 90.00° 98.14° 90.00°	Depositor
Resolution (Å)	41.98 – 3.00 41.98 – 2.97	Depositor EDS
% Data completeness (in resolution range)	95.3 (41.98-3.00) 89.6 (41.98-2.97)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.22 (at 2.95Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1563)	Depositor
R, R_{free}	0.269 , 0.294 0.274 , 0.293	Depositor DCC
R_{free} test set	1883 reflections (1.75%)	wwPDB-VP
Wilson B-factor (Å^2)	59.4	Xtriage
Anisotropy	0.521	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 29.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	28290	wwPDB-VP
Average B, all atoms (Å^2)	108.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.35% 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: ATP, MG, 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	0.21	0/1814	0.40	0/2466
1	B	0.21	0/1782	0.40	0/2424
2	C	0.21	0/8937	0.39	0/12087
3	D	0.21	0/11928	0.39	0/16127
4	E	0.21	0/775	0.38	0/1045
5	F	0.23	1/2791 (0.0%)	0.38	0/3754
6	G	0.55	0/323	1.02	1/497 (0.2%)
7	H	0.46	0/374	1.03	0/575
8	I	1.07	0/112	1.03	0/171
All	All	0.23	1/28836 (0.0%)	0.43	1/39146 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	F	339	PRO	C-N	-5.72	1.20	1.34

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	G	14	DG	O3'-P-O5'	-5.61	93.35	104.00

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	1782	0	1834	24	0
1	B	1750	0	1797	20	0
2	C	8770	0	8874	149	0
3	D	11722	0	11949	160	0
4	E	761	0	778	9	0
5	F	2747	0	2831	79	0
6	G	289	0	159	9	0
7	H	333	0	183	6	0
8	I	102	0	55	3	0
9	D	1	0	0	0	0
10	D	2	0	0	0	0
11	I	31	0	11	2	0
All	All	28290	0	28471	390	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 7.

The worst 5 of 390 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:C:776:SER:CB	5:F:373:LYS:NZ	2.02	1.23
2:C:769:PRO:HG2	5:F:374:GLY:O	1.40	1.19
2:C:776:SER:HB2	5:F:373:LYS:NZ	1.58	1.19
2:C:769:PRO:CG	5:F:374:GLY:O	1.93	1.17
3:D:233:LYS:NZ	3:D:240:GLU:OE2	1.88	1.04

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	224/315 (71%)	209 (93%)	15 (7%)	0	100	100
1	B	220/315 (70%)	207 (94%)	12 (6%)	1 (0%)	29	68
2	C	1107/1119 (99%)	1029 (93%)	70 (6%)	8 (1%)	22	60
3	D	1480/1524 (97%)	1393 (94%)	77 (5%)	10 (1%)	22	60
4	E	92/99 (93%)	89 (97%)	3 (3%)	0	100	100
5	F	334/423 (79%)	316 (95%)	16 (5%)	2 (1%)	25	64
All	All	3457/3795 (91%)	3243 (94%)	193 (6%)	21 (1%)	25	64

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	F	421	PHE
2	C	12	VAL
2	C	211	LEU
3	D	448	GLU
3	D	484	PRO

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%)	197 (99%)	2 (1%)	76	91
1	B	195/273 (71%)	188 (96%)	7 (4%)	35	70
2	C	936/941 (100%)	908 (97%)	28 (3%)	41	75
3	D	1251/1279 (98%)	1196 (96%)	55 (4%)	28	65
4	E	83/88 (94%)	82 (99%)	1 (1%)	71	90
5	F	294/371 (79%)	282 (96%)	12 (4%)	30	67
All	All	2958/3225 (92%)	2853 (96%)	105 (4%)	36	71

5 of 105 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	632	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	D	974	ILE
5	F	369	LEU
3	D	717	GLN
3	D	864	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1441	GLN
3	D	1359	GLN
3	D	350	HIS
3	D	1124	GLN
3	D	294	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	4/5 (80%)	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)

Of 4 ligands modelled in this entry, 3 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	$\#\|Z\| > 2$
11	ATP	I	101	-	26,33,33	1.68	1 (3%)	31,52,52	1.38	3 (9%)

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
11	ATP	I	101	-	-	4/18/38/38	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	I	101	ATP	C5'-C4'	-7.94	1.26	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
11	I	101	ATP	C5'-C4'-C3'	-3.79	100.98	115.18
11	I	101	ATP	PB-O3B-PG	-3.45	120.97	132.83
11	I	101	ATP	PA-O3A-PB	-3.07	122.30	132.83

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
11	I	101	ATP	C5'-O5'-PA-O3A
11	I	101	ATP	O4'-C4'-C5'-O5'
11	I	101	ATP	C3'-C4'-C5'-O5'
11	I	101	ATP	C5'-O5'-PA-O1A

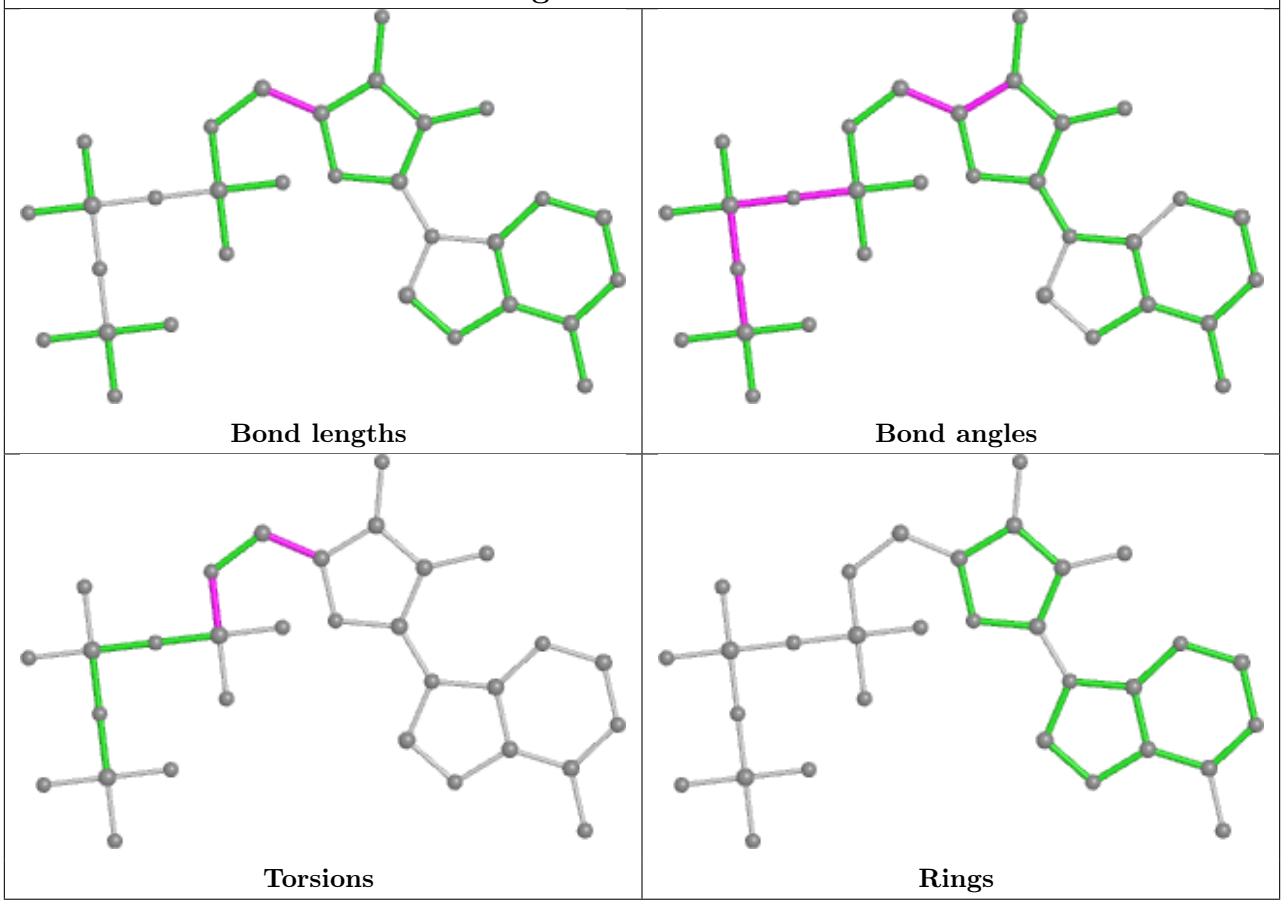
There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
11	I	101	ATP	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand ATP I 101



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	226/315 (71%)	-0.08	3 (1%) 77 51	76, 116, 149, 155	0
1	B	222/315 (70%)	-0.27	1 (0%) 91 75	63, 90, 119, 133	0
2	C	1111/1119 (99%)	0.05	32 (2%) 51 23	28, 107, 187, 229	0
3	D	1484/1524 (97%)	-0.01	35 (2%) 59 30	30, 90, 161, 225	1 (0%)
4	E	94/99 (94%)	-0.07	2 (2%) 63 34	64, 100, 177, 190	0
5	F	338/423 (79%)	0.36	32 (9%) 8 3	94, 135, 226, 265	0
6	G	14/22 (63%)	-0.02	0 100 100	78, 106, 188, 200	0
7	H	16/27 (59%)	-0.29	0 100 100	126, 135, 179, 183	0
8	I	5/5 (100%)	-0.63	0 100 100	60, 65, 75, 76	0
All	All	3510/3849 (91%)	0.02	105 (2%) 50 22	28, 104, 177, 265	1 (0%)

The worst 5 of 105 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	391	GLY	8.7
2	C	300	ASP	5.4
5	F	149	GLU	5.0
2	C	311	PHE	4.7
5	F	392	VAL	4.6

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

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands (i)

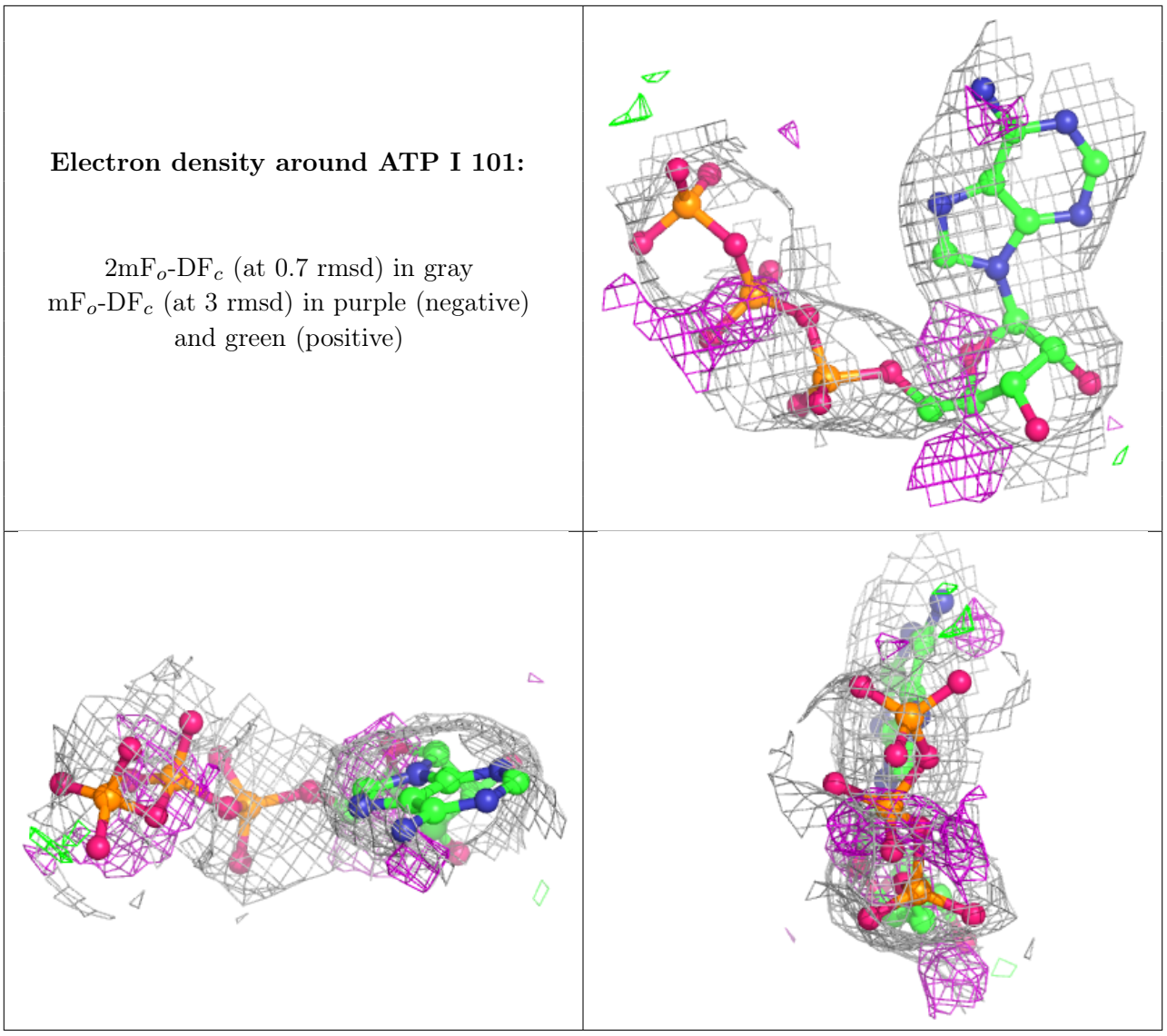
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
11	ATP	I	101	31/31	0.82	0.24	72,98,200,233	0
10	ZN	D	2002	1/1	0.89	0.23	238,238,238,238	0
9	MG	D	2001	1/1	0.93	0.10	24,24,24,24	0
10	ZN	D	2003	1/1	0.98	0.15	47,47,47,47	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around ATP I 101:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.