



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 27, 2020 – 02:57 am BST

PDB ID : 1H38  
Title : Structure of a T7 RNA polymerase elongation complex at 2.9Å resolution  
Authors : Tahirov, T.H.; Temyakov, D.; Anikin, M.; Patlan, V.; McAllister, W.T.; Vassilyev, D.G.; Yokoyama, S.  
Deposited on : 2002-08-24  
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](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11



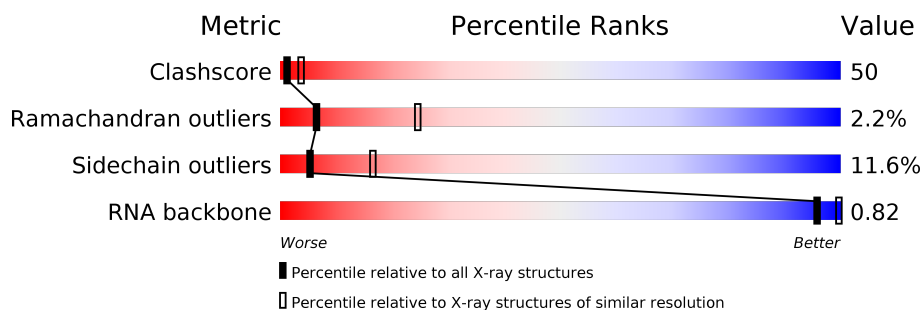
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 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))
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ .

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	883	38% 50% 9% .
1	B	883	38% 49% 9% .
1	C	883	39% 49% 8% .
1	D	883	38% 51% 8% .
2	E	18	61% 28% 6% 6%
2	H	18	17% 39% 39% 6%
2	K	18	56% 39% 6%

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	N	18	
3	F	12	
3	I	12	
3	L	12	
3	O	12	
4	G	10	
4	J	10	
4	M	10	
4	P	10	



## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 30948 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	857	Total	C	N	O	S	0	0	0
			6746	4296	1173	1242	35			
1	B	857	Total	C	N	O	S	0	0	0
			6746	4296	1173	1242	35			
1	C	857	Total	C	N	O	S	0	0	0
			6746	4296	1173	1242	35			
1	D	857	Total	C	N	O	S	0	0	0
			6746	4296	1173	1242	35			

- Molecule 2 is a DNA chain called 5'-D(\*GP\*GP\*GP\*AP\*AP\*TP\*CP\*GP\*AP\*CP \*AP\*TP\*CP\*GP\*CP\*CP\*GP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	17	Total	C	N	O	P	0	0	0
			345	164	67	98	16			
2	H	18	Total	C	N	O	P	0	0	0
			367	174	72	104	17			
2	K	17	Total	C	N	O	P	0	0	0
			345	164	67	98	16			
2	N	17	Total	C	N	O	P	0	0	0
			345	164	67	98	16			

- Molecule 3 is a RNA chain called 5'-R(\*AP\*AP\*CP\*UP\*GP\*CP\*GP\*GP\*CP\*GP \*AP\*U)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	8	Total	C	N	O	P	0	0	0
			171	77	33	54	7			
3	I	8	Total	C	N	O	P	0	0	0
			171	77	33	54	7			
3	L	8	Total	C	N	O	P	0	0	0
			171	77	33	54	7			

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	O	8	Total	C	N	O	P	0	0	0
			171	77	33	54	7			

- Molecule 4 is a DNA chain called 5'-D(\*GP\*TP\*CP\*GP\*AP\*TP\*TP\*CP\*CP\*CP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	G	9	Total	C	N	O	P	0	0	0
			179	87	30	54	8			
4	J	9	Total	C	N	O	P	0	0	0
			179	87	30	54	8			
4	M	9	Total	C	N	O	P	0	0	0
			179	87	30	54	8			
4	P	9	Total	C	N	O	P	0	0	0
			179	87	30	54	8			

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	310	Total	O	0	0
			310	310		
5	B	352	Total	O	0	0
			352	352		
5	C	185	Total	O	0	0
			185	185		
5	D	177	Total	O	0	0
			177	177		
5	E	13	Total	O	0	0
			13	13		
5	F	16	Total	O	0	0
			16	16		
5	G	12	Total	O	0	0
			12	12		
5	H	20	Total	O	0	0
			20	20		
5	I	11	Total	O	0	0
			11	11		
5	J	9	Total	O	0	0
			9	9		
5	K	17	Total	O	0	0
			17	17		
5	L	5	Total	O	0	0
			5	5		

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	M	8	Total 8	O 8	0	0
5	N	12	Total 12	O 12	0	0
5	O	6	Total 6	O 6	0	0
5	P	9	Total 9	O 9	0	0

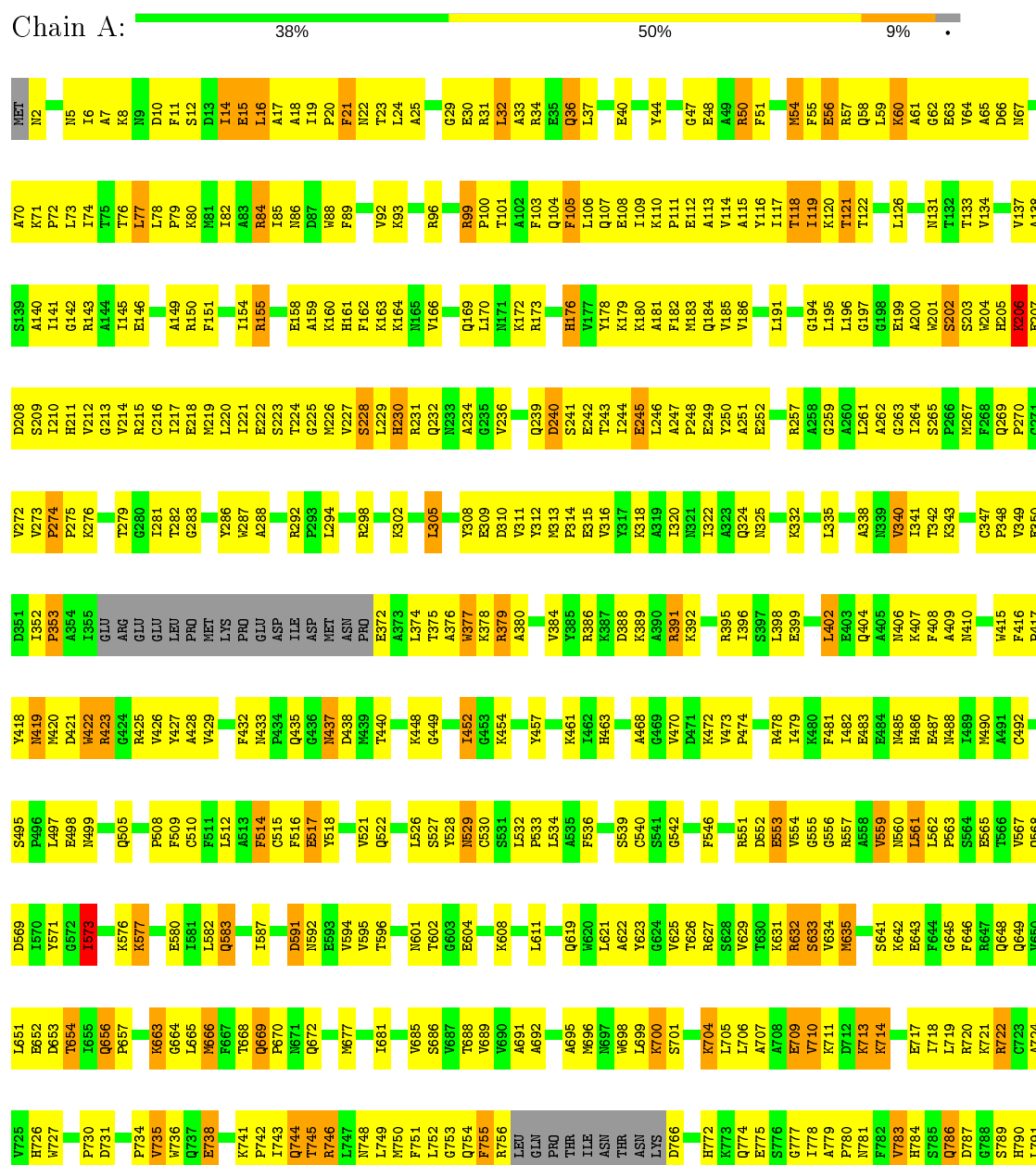


### 3 Residue-property plots

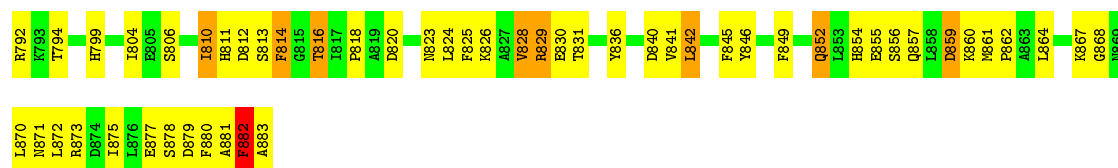
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according 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. 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.

Note EDS was not executed.

#### • Molecule 1: DNA-DIRECTED RNA POLYMERASE

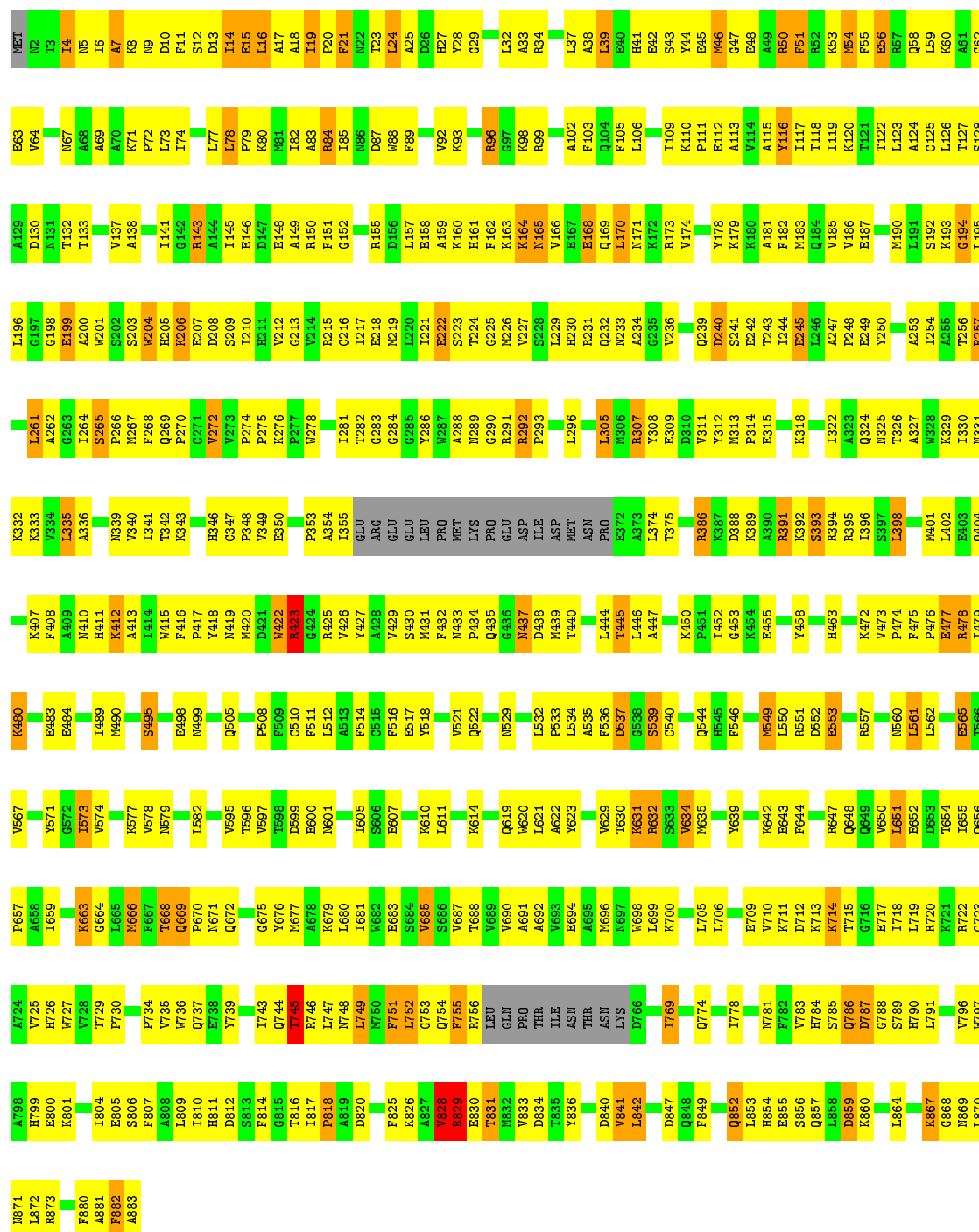






# • Molecule 1: DNA-DIRECTED RNA POLYMERASE

Chain B: 38% 49% 9%

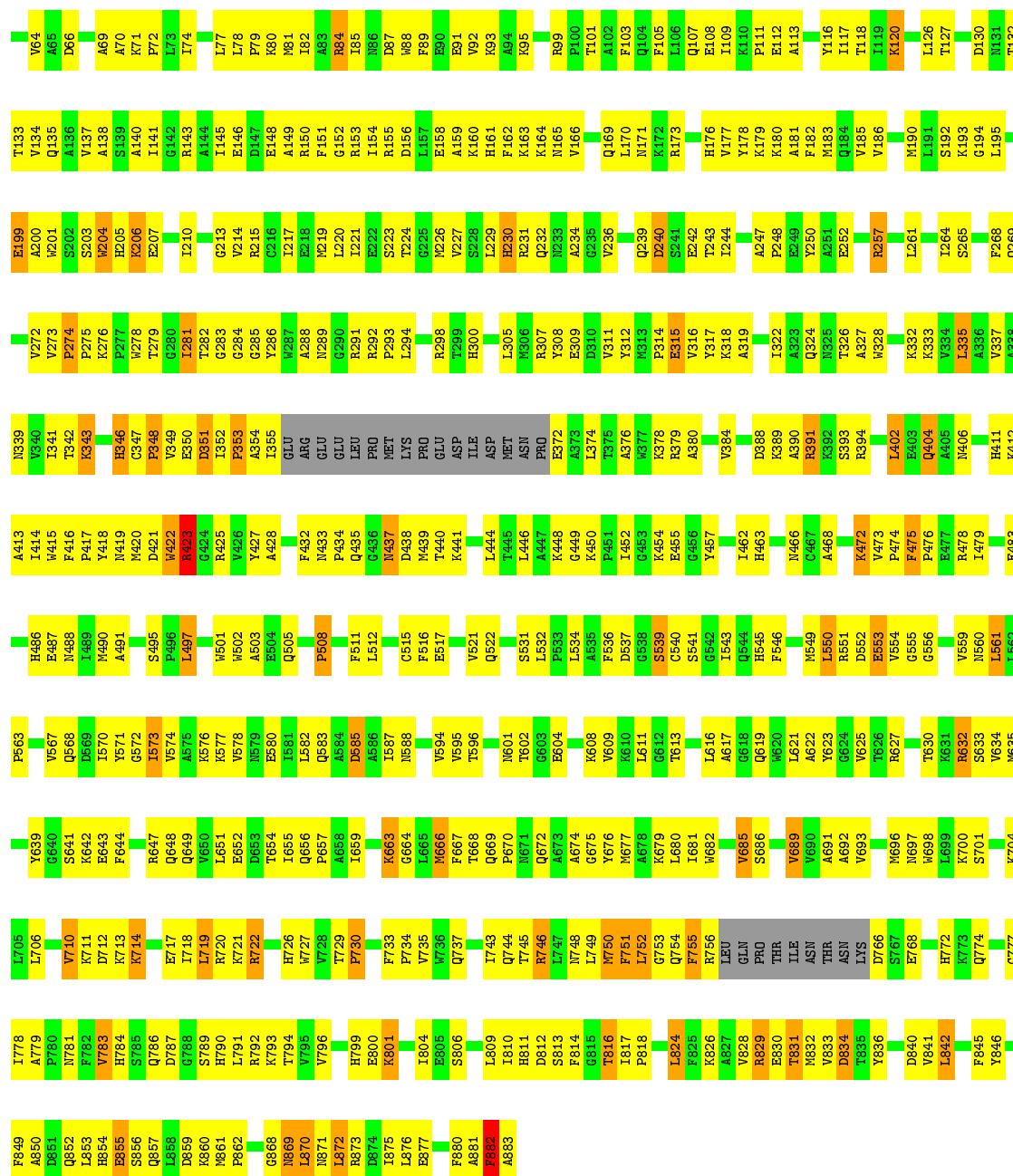




Chain C:  39% 49% 8% .

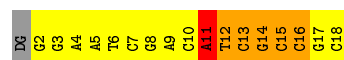






• Molecule 2: 5'-D(\*GP\*GP\*GP\*AP\*AP\*TP\*CP\*GP\*AP\*CP \*AP\*TP\*CP\*GP\*CP\*CP\*GP\*C)-3'

Chain E:



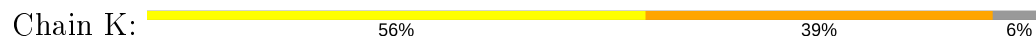
• Molecule 2: 5'-D(\*GP\*GP\*GP\*AP\*AP\*TP\*CP\*GP\*AP\*CP \*AP\*TP\*CP\*GP\*CP\*CP\*GP\*C)-3'

Chain H:

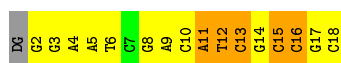




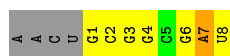
- Molecule 2: 5'-D(\*GP\*GP\*GP\*AP\*AP\*TP\*CP\*GP\*AP\*CP \*AP\*TP\*CP\*GP\*CP\*CP\*GP\*C)-3'



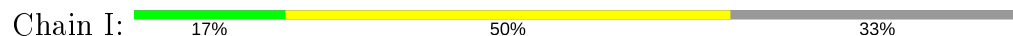
- Molecule 2: 5'-D(\*GP\*GP\*GP\*AP\*AP\*TP\*CP\*GP\*AP\*CP \*AP\*TP\*CP\*GP\*CP\*CP\*GP\*C)-3'



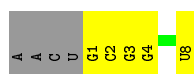
- Molecule 3: 5'-R(\*AP\*AP\*CP\*UP\*GP\*CP\*GP\*GP\*CP\*GP \*AP\*U)-3'



- Molecule 3: 5'-R(\*AP\*AP\*CP\*UP\*GP\*CP\*GP\*GP\*CP\*GP \*AP\*U)-3'



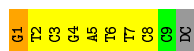
- Molecule 3: 5'-R(\*AP\*AP\*CP\*UP\*GP\*CP\*GP\*GP\*CP\*GP \*AP\*U)-3'



- Molecule 3: 5'-R(\*AP\*AP\*CP\*UP\*GP\*CP\*GP\*GP\*CP\*GP \*AP\*U)-3'



- Molecule 4: 5'-D(\*GP\*TP\*CP\*GP\*AP\*TP\*TP\*CP\*CP\*CP)-3'






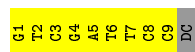
- Molecule 4: 5'-D(\*GP\*TP\*CP\*GP\*AP\*TP\*TP\*CP\*CP\*CP)-3'

Chain J:  20% 70% 10%




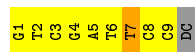
- Molecule 4: 5'-D(\*GP\*TP\*CP\*GP\*AP\*TP\*TP\*CP\*CP\*CP)-3'

Chain M:  90% 10%



- Molecule 4: 5'-D(\*GP\*TP\*CP\*GP\*AP\*TP\*TP\*CP\*CP\*CP)-3'

Chain P:  80% 10% 10%





## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.91 Å   84.97 Å   202.00 Å 90.36°   92.97°   109.94°	Depositor
Resolution (Å)	39.93 – 2.90	Depositor
% Data completeness (in resolution range)	98.0 (39.93-2.90)	Depositor
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.236 , 0.284	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	30948	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	71.0	wwPDB-VP



## 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	0.89	9/6897 (0.1%)	0.85	5/9329 (0.1%)
1	B	0.93	2/6897 (0.0%)	0.87	9/9329 (0.1%)
1	C	0.64	0/6897	0.73	0/9329
1	D	0.58	0/6897	0.70	2/9329 (0.0%)
2	E	0.89	0/387	1.06	1/595 (0.2%)
2	H	0.94	0/412	1.03	1/634 (0.2%)
2	K	0.90	0/387	0.97	0/595
2	N	0.80	0/387	0.95	0/595
3	F	1.09	1/191 (0.5%)	0.85	0/297
3	I	0.94	0/191	0.81	0/297
3	L	0.81	0/191	0.79	0/297
3	O	0.61	0/191	0.74	0/297
4	G	0.88	0/199	0.84	0/305
4	J	0.77	0/199	0.93	0/305
4	M	0.68	0/199	0.85	0/305
4	P	0.74	0/199	1.05	0/305
All	All	0.78	12/30721 (0.0%)	0.81	18/42143 (0.0%)

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	E	0	6
2	H	0	8
2	K	0	7
2	N	0	6
4	G	0	1
4	J	0	1
4	P	0	1
All	All	0	30



The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	510	CYS	CB-SG	-8.42	1.68	1.82
1	A	573	ILE	CA-CB	7.10	1.71	1.54
1	A	654	THR	CA-CB	5.53	1.67	1.53
1	A	783	VAL	CA-CB	-5.40	1.43	1.54
3	F	7	A	C5-C6	-5.31	1.36	1.41

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	537	ASP	CB-CG-OD1	-8.22	110.90	118.30
1	A	557	ARG	NE-CZ-NH2	-8.21	116.20	120.30
1	A	829	ARG	NE-CZ-NH1	-7.68	116.46	120.30
1	B	549	MET	CB-CG-SD	-6.86	91.82	112.40
1	A	787	ASP	CB-CG-OD2	-6.55	112.41	118.30

There are no chirality outliers.

5 of 30 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	E	11	DA	Sidechain
2	E	12	DT	Sidechain
2	E	13	DC	Sidechain
2	E	14	DG	Sidechain
2	E	15	DC	Sidechain

## 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	6746	0	6708	722	0
1	B	6746	0	6708	733	0
1	C	6746	0	6708	680	0
1	D	6746	0	6708	653	0
2	E	345	0	191	30	0
2	H	367	0	202	28	0
2	K	345	0	191	38	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	N	345	0	191	21	0
3	F	171	0	89	7	0
3	I	171	0	89	11	0
3	L	171	0	89	9	0
3	O	171	0	89	5	0
4	G	179	0	104	14	0
4	J	179	0	104	8	0
4	M	179	0	104	18	0
4	P	179	0	104	11	0
5	A	310	0	0	102	0
5	B	352	0	0	133	0
5	C	185	0	0	79	0
5	D	177	0	0	105	0
5	E	13	0	0	2	0
5	F	16	0	0	2	0
5	G	12	0	0	5	0
5	H	20	0	0	6	0
5	I	11	0	0	4	0
5	J	9	0	0	4	0
5	K	17	0	0	7	0
5	L	5	0	0	2	0
5	M	8	0	0	1	0
5	N	12	0	0	4	0
5	O	6	0	0	0	0
5	P	9	0	0	3	0
All	All	30948	0	28379	2933	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 50.

The worst 5 of 2933 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:668:THR:HG22	1:A:669:GLN:NE2	1.51	1.24
1:C:428:ALA:H	1:C:435:GLN:NE2	1.44	1.15
1:B:50:ARG:HG2	1:B:50:ARG:HH11	1.01	1.14
1:C:133:THR:HA	1:C:243:THR:HG22	1.28	1.10
1:A:120:LYS:HE3	1:A:752:LEU:HD21	1.31	1.10

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	851/883 (96%)	740 (87%)	94 (11%)	17 (2%)	7	27
1	B	851/883 (96%)	736 (86%)	98 (12%)	17 (2%)	7	27
1	C	851/883 (96%)	737 (87%)	93 (11%)	21 (2%)	5	21
1	D	851/883 (96%)	735 (86%)	96 (11%)	20 (2%)	6	22
All	All	3404/3532 (96%)	2948 (87%)	381 (11%)	75 (2%)	6	24

5 of 75 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	194	GLY
1	A	199	GLU
1	A	539	SER
1	A	663	LYS
1	A	755	PHE

### 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	703/729 (96%)	615 (88%)	88 (12%)	4	14
1	B	703/729 (96%)	619 (88%)	84 (12%)	5	15
1	C	703/729 (96%)	623 (89%)	80 (11%)	5	17
1	D	703/729 (96%)	628 (89%)	75 (11%)	6	20
All	All	2812/2916 (96%)	2485 (88%)	327 (12%)	5	16



5 of 327 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	668	THR
1	C	120	LYS
1	D	689	VAL
1	B	725	VAL
1	B	829	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 130 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	786	GLN
1	C	289	ASN
1	D	744	GLN
1	B	823	ASN
1	C	171	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	F	7/12 (58%)	0	0
3	I	7/12 (58%)	0	0
3	L	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 ⓘ

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

## 5.5 Carbohydrates ⓘ

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

### 6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.