wwPDB X-ray Structure Validation Summary Report (i)

Jun 23, 2024 – 05:43 PM EDT

PDB ID	:	5C44
Title	:	Crystal structure of a transcribing RNA Polymerase II complex reveals a complete transcription bubble
Authors	:	Barnes, C.O.; Calero, M.; Malik, I.; Spahr, H.; Zhang, Q.; Pullara, F.; Kaplan, C.D.; Calero, G.
Deposited on	:	2015-06-17
Resolution	:	3.95 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:

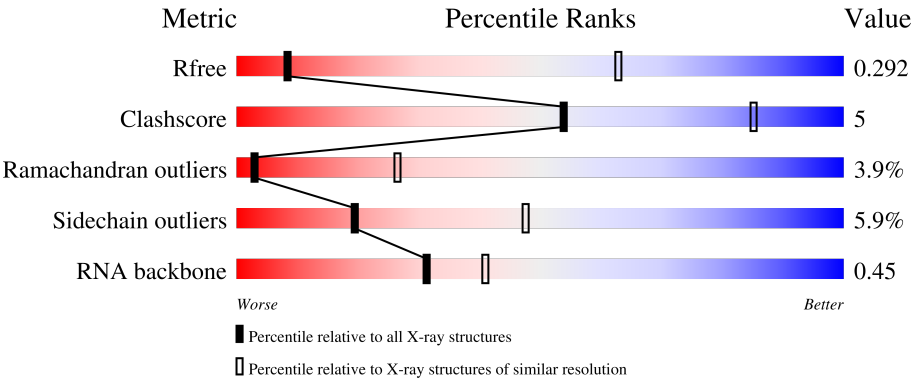
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.95 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.



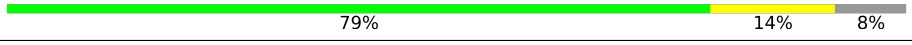
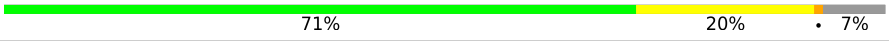
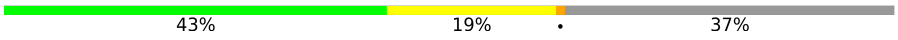
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1025 (4.22-3.70)
Clashscore	141614	1085 (4.22-3.70)
Ramachandran outliers	138981	1047 (4.22-3.70)
Sidechain outliers	138945	1039 (4.22-3.70)
RNA backbone	3102	1041 (4.84-3.00)

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%

Mol	Chain	Length	Quality of chain
1	A	1733	66% 15% • 17%
2	B	1224	73% 20% • 5%
3	C	318	66% 15% • 17%
4	D	221	67% 13% • 19%
5	E	215	86% 13% •
6	F	155	43% 14% 44%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	G	179	
8	H	146	
9	I	120	
10	J	70	
11	K	120	
12	L	70	
13	R	9	
14	S	53	
15	U	53	

2 Entry composition (i)

There are 15 unique types of molecules in this entry. The entry contains 32540 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 II subunit RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1434	11249	7083	1967	2137	62	0	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase II subunit RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1158	9175	5795	1603	1721	56	0	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	265	2086	1312	347	414	13	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase II subunit RPB4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	178	1417	875	254	286	2	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	214	1752	1111	309	321	11	0	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	87	705	451	119	132	3	0	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	171	Total	C	N	O	S	0	0	0
			1339	861	222	248	8			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	172	LEU	-	expression tag	UNP P34087
G	173	GLU	-	expression tag	UNP P34087
G	174	HIS	-	expression tag	UNP P34087
G	175	HIS	-	expression tag	UNP P34087
G	176	HIS	-	expression tag	UNP P34087
G	177	HIS	-	expression tag	UNP P34087
G	178	HIS	-	expression tag	UNP P34087
G	179	HIS	-	expression tag	UNP P34087

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	135	Total	C	N	O	S	0	0	0
			1080	679	182	214	5			

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	I	114	Total	C	N	O	S	0	0	0
			921	568	165	178	10			

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	65	Total	C	N	O	S	0	0	0
			532	339	93	94	6			

- Molecule 11 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	115	Total	C	N	O	S	0	0	0
			924	593	157	172	2			

- Molecule 12 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	L	44	Total	C	N	O	S	0	0	0
			346	214	67	61	4			

- Molecule 13 is a RNA chain called Synthetic RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	R	9	Total	C	N	O	P	0	0	0
			197	88	40	60	9			

- Molecule 14 is a DNA chain called Synthetic DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	S	13	Total	C	N	O	P	0	0	0
			268	128	46	81	13			

- Molecule 15 is a DNA chain called Synthetic DNA.

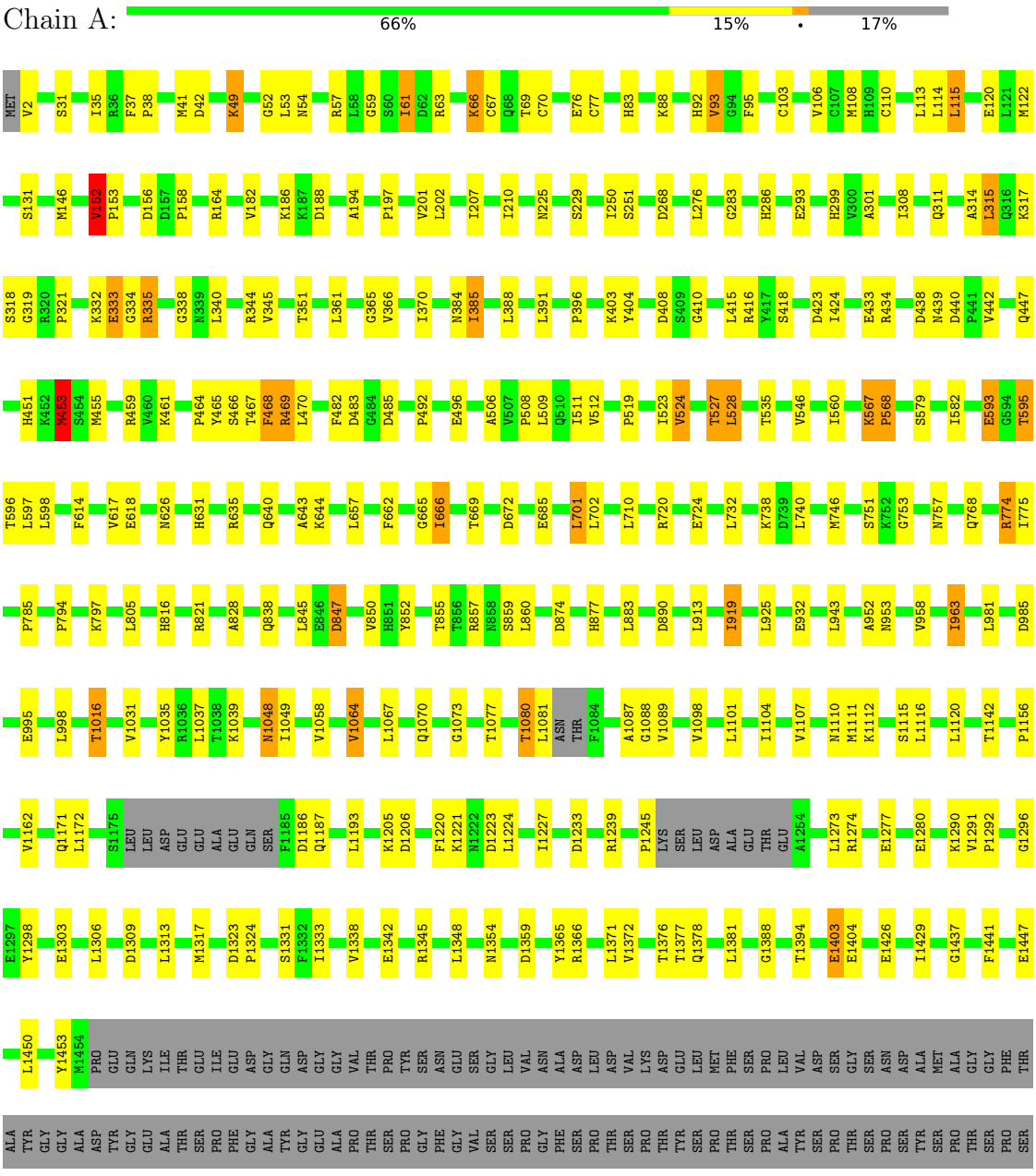
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
15	U	27	Total	C	N	O	P	0	0	0
			549	261	102	159	27			

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. 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.

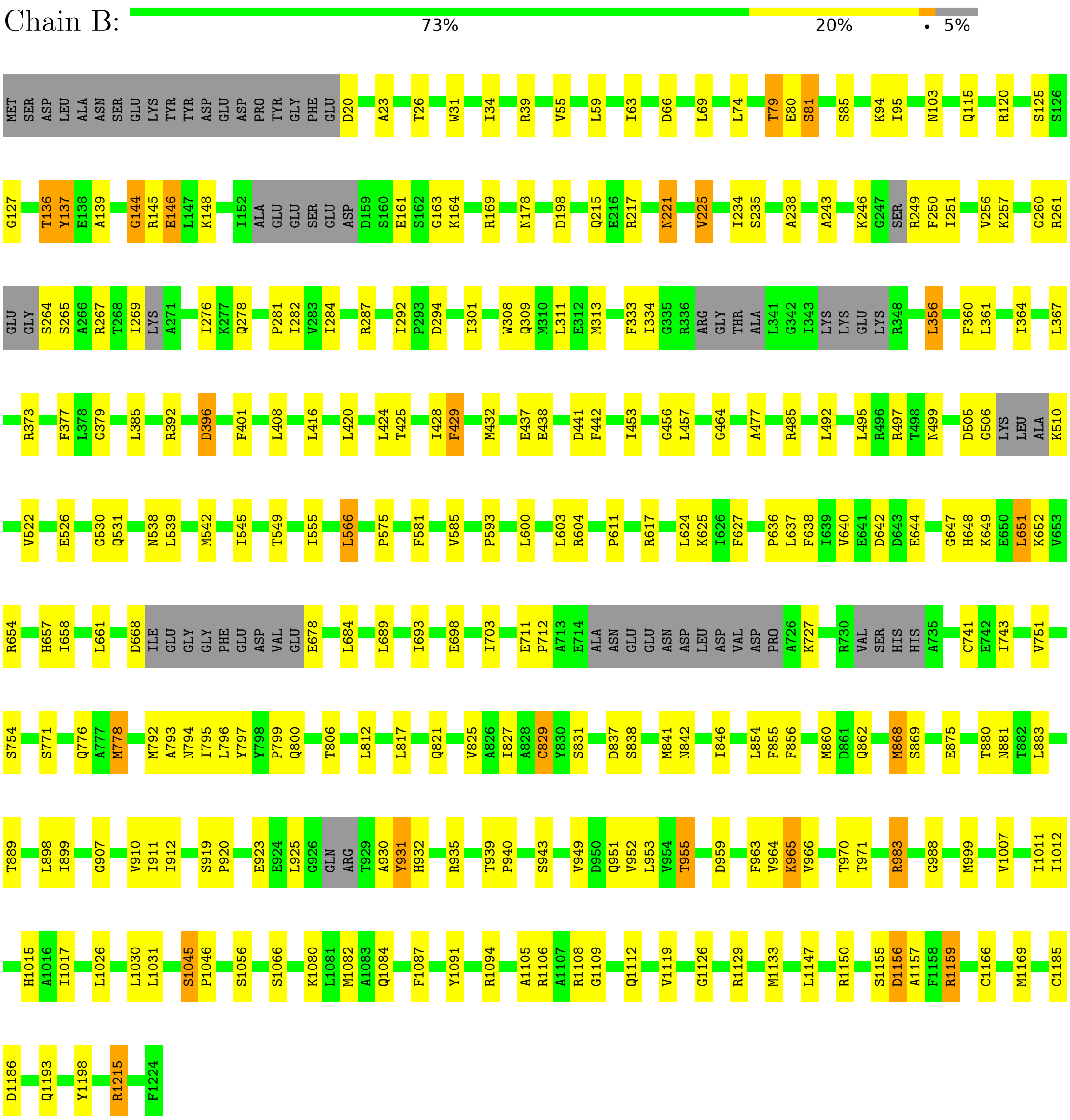
- Molecule 1: DNA-directed RNA polymerase II subunit RPB1

Chain A:



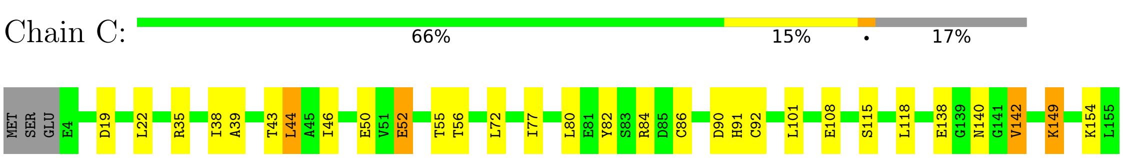
• Molecule 2: DNA-directed RNA polymerase II subunit RPB2

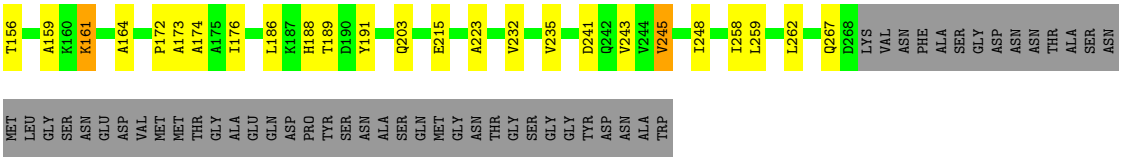
Chain B: 73% 20% • 5%



• Molecule 3: DNA-directed RNA polymerase II subunit RPB3

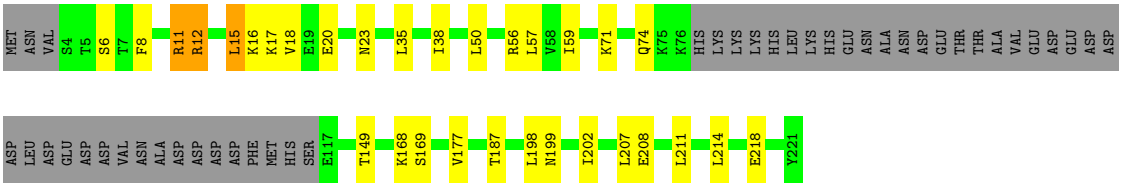
Chain C: 66% 15% • 17%





• Molecule 4: DNA-directed RNA polymerase II subunit RPB4

Chain D:

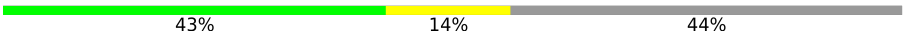


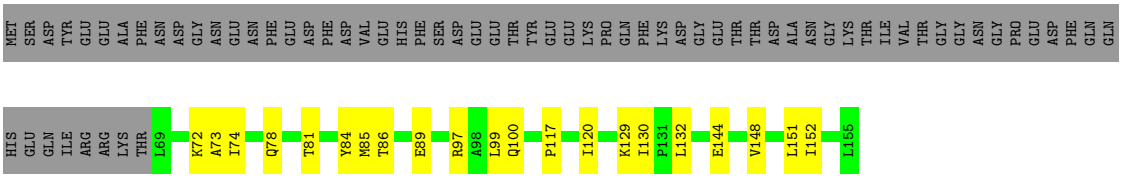
• Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC1

Chain E:

• Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:



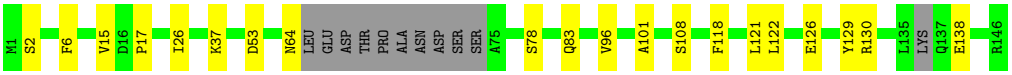


• Molecule 7: DNA-directed RNA polymerase II subunit RPB7

Chain G:

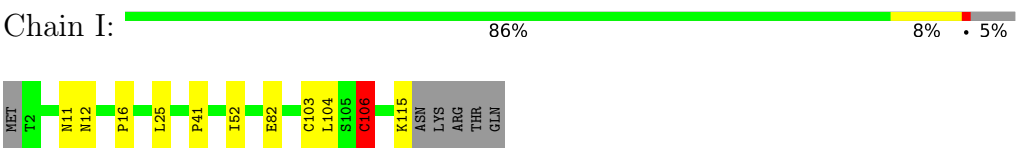
• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H:



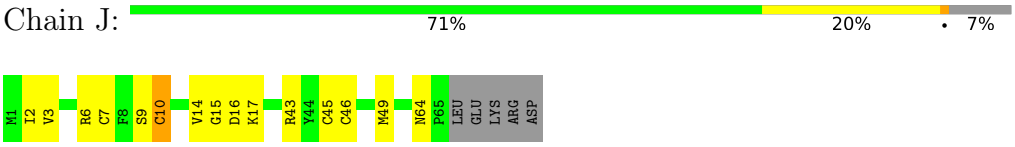
• Molecule 9: DNA-directed RNA polymerase II subunit RPB9

Chain I:



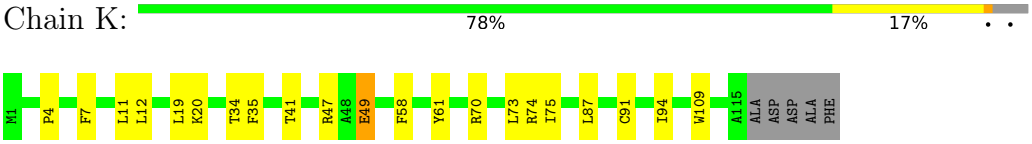
• Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J:



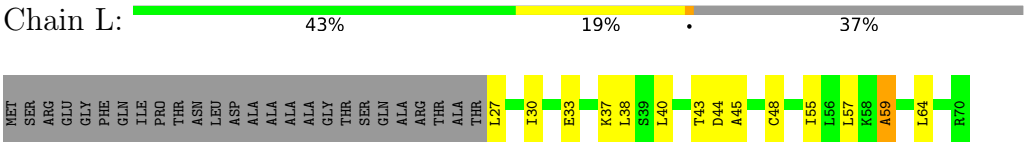
• Molecule 11: DNA-directed RNA polymerase II subunit RPB11

Chain K:



• Molecule 12: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L:

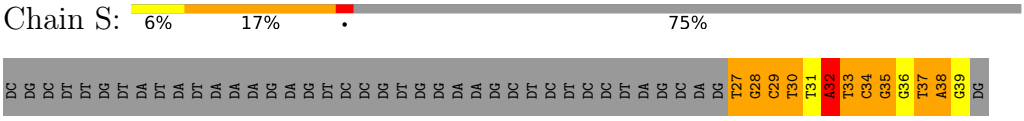


• Molecule 13: Synthetic RNA

Chain R:

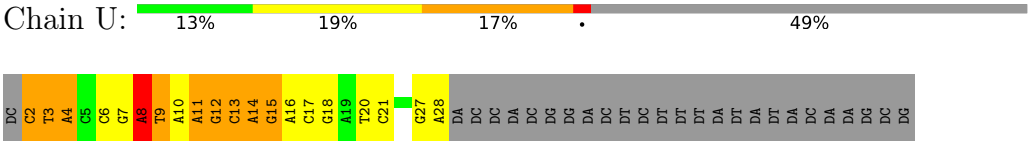
• Molecule 14: Synthetic DNA

Chain S:



• Molecule 15: Synthetic DNA

Chain U:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	220.22Å 391.84Å 282.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	79.00 – 3.95 90.88 – 3.95	Depositor EDS
% Data completeness (in resolution range)	94.3 (79.00-3.95) 94.6 (90.88-3.95)	Depositor EDS
R_{merge}	0.23	Depositor
R_{sym}	0.23	Depositor
$< I/\sigma(I) >$ [1]	1.62 (at 4.01Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.215 , 0.236 0.271 , 0.292	Depositor DCC
R_{free} test set	3013 reflections (2.99%)	wwPDB-VP
Wilson B-factor (Å^2)	114.2	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 214.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.32$, $< L^2 > = 0.15$	Xtriage
Estimated twinning fraction	0.178 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.186 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	32540	wwPDB-VP
Average B, all atoms (Å^2)	197.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.15% 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

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	0.41	0/11452	0.64	0/15492
2	B	0.41	1/9347 (0.0%)	0.63	6/12601 (0.0%)
3	C	0.40	0/2124	0.60	0/2879
4	D	0.40	0/1427	0.59	0/1911
5	E	0.38	0/1788	0.57	0/2406
6	F	0.40	0/717	0.63	0/967
7	G	0.39	0/1367	0.62	0/1844
8	H	0.37	0/1097	0.57	0/1484
9	I	0.38	0/939	0.59	0/1266
10	J	0.38	0/541	0.58	0/727
11	K	0.37	0/942	0.56	0/1272
12	L	0.40	0/348	0.64	0/461
13	R	0.82	0/221	0.90	1/343 (0.3%)
14	S	3.47	54/299 (18.1%)	1.36	1/460 (0.2%)
15	U	2.78	61/615 (9.9%)	1.12	0/945
All	All	0.64	116/33224 (0.3%)	0.65	8/45058 (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
14	S	0	1
15	U	0	1
All	All	0	2

The worst 5 of 116 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
15	U	4	DA	N3-C4	12.86	1.42	1.34
15	U	14	DA	N3-C4	12.21	1.42	1.34
15	U	11	DA	N3-C4	11.55	1.41	1.34

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
14	S	27	DT	P-O5'	11.27	1.71	1.59
15	U	11	DA	N9-C4	11.21	1.44	1.37

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	260	GLY	O-C-N	-8.44	109.19	122.70
2	B	260	GLY	CA-C-N	5.98	130.37	117.20
13	R	2	U	P-O3'-C3'	-5.93	112.58	119.70
2	B	79	THR	C-N-CA	5.90	136.46	121.70
14	S	27	DT	O4'-C1'-N1	5.64	111.95	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
14	S	32	DA	Sidechain
15	U	8	DA	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	11249	0	11277	114	0
2	B	9175	0	9138	123	0
3	C	2086	0	2049	25	0
4	D	1417	0	1428	9	0
5	E	1752	0	1776	11	0
6	F	705	0	731	13	0
7	G	1339	0	1357	12	0
8	H	1080	0	1049	9	0
9	I	921	0	877	9	0
10	J	532	0	546	5	0
11	K	924	0	934	11	0
12	L	346	0	367	3	0
13	R	197	0	97	4	0
14	S	268	0	149	25	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
15	U	549	0	303	33	0
All	All	32540	0	32078	342	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 5.

The worst 5 of 342 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)
14:S:38:DA:N6	15:U:2:DC:N4	2.15	0.93
14:S:38:DA:N6	15:U:2:DC:C4	2.43	0.86
1:A:567:LYS:HB3	1:A:568:PRO:CD	2.07	0.85
1:A:67:CYS:HG	1:A:77:CYS:HG	1.12	0.85
14:S:38:DA:H62	15:U:2:DC:N4	1.74	0.83

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	1426/1733 (82%)	1227 (86%)	135 (10%)	64 (4%)	2 24
2	B	1134/1224 (93%)	967 (85%)	121 (11%)	46 (4%)	3 25
3	C	263/318 (83%)	228 (87%)	26 (10%)	9 (3%)	3 30
4	D	174/221 (79%)	152 (87%)	12 (7%)	10 (6%)	1 19
5	E	212/215 (99%)	198 (93%)	10 (5%)	4 (2%)	8 40
6	F	85/155 (55%)	79 (93%)	3 (4%)	3 (4%)	3 29
7	G	169/179 (94%)	143 (85%)	21 (12%)	5 (3%)	4 32
8	H	129/146 (88%)	111 (86%)	14 (11%)	4 (3%)	4 31

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
9	I	112/120 (93%)	98 (88%)	13 (12%)	1 (1%)	17 54
10	J	63/70 (90%)	56 (89%)	3 (5%)	4 (6%)	1 18
11	K	113/120 (94%)	110 (97%)	3 (3%)	0	100 100
12	L	42/70 (60%)	31 (74%)	8 (19%)	3 (7%)	1 16
All	All	3922/4571 (86%)	3400 (87%)	369 (9%)	153 (4%)	3 26

5 of 153 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	49	LYS
1	A	66	LYS
1	A	319	GLY
1	A	385	ILE
1	A	567	LYS

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	1244/1520 (82%)	1156 (93%)	88 (7%)	14 42
2	B	996/1061 (94%)	938 (94%)	58 (6%)	20 48
3	C	233/274 (85%)	219 (94%)	14 (6%)	19 47
4	D	156/200 (78%)	147 (94%)	9 (6%)	20 48
5	E	196/197 (100%)	186 (95%)	10 (5%)	24 51
6	F	77/137 (56%)	77 (100%)	0	100 100
7	G	152/160 (95%)	145 (95%)	7 (5%)	27 54
8	H	118/128 (92%)	117 (99%)	1 (1%)	81 88
9	I	107/114 (94%)	105 (98%)	2 (2%)	57 75
10	J	60/65 (92%)	57 (95%)	3 (5%)	24 52
11	K	99/102 (97%)	93 (94%)	6 (6%)	18 47
12	L	38/57 (67%)	32 (84%)	6 (16%)	2 16

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	3476/4015 (87%)	3272 (94%)	204 (6%)	19 47

5 of 204 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	624	LEU
2	B	1159	ARG
12	L	33	GLU
2	B	698	GLU
2	B	931	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	9	GLN
9	I	11	ASN
9	I	89	GLN
9	I	46	HIS
1	A	1171	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
13	R	7/9 (77%)	2 (28%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
13	R	3	C
13	R	6	G

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)

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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.