wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 02:12 PM BST

PDB ID	:	4V1N
EMDB ID:	:	EMD-2785
Title	:	Architecture of the RNA polymerase II-Mediator core transcription initiation complex
Authors	:	Plaschka, C.; Lariviere, L.; Wenzeck, L.; Hemann, M.; Tegunov, D.; Petrotchenko, E.V.; Borchers, C.H.; Baumeister, W.; Herzog, F.; Villa, E.; Cramer, P.
Deposited on	:	2014-09-29
Resolution	:	7.80 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

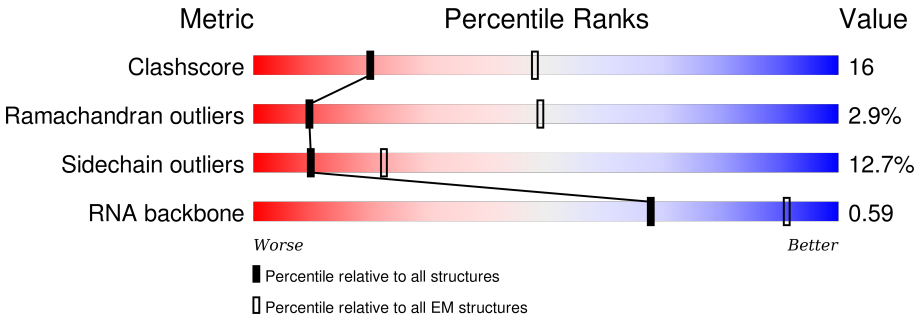
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 7.80 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.



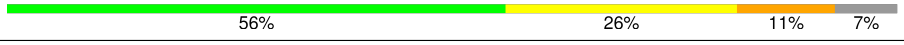
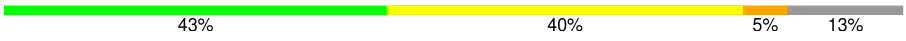
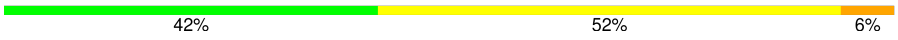
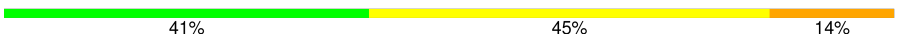
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686
RNA backbone	3027	244

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	55% 23% • • 18%
2	B	1224	62% 28% • 6%
3	C	318	61% 18% • 16%
4	D	221	55% 20% 5% 19%
5	E	215	74% 22% •
6	F	155	36% 16% • 46%
7	G	171	72% 25% • •
8	H	146	58% 28% 5% 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	I	122	
10	J	70	
11	K	120	
12	L	70	
13	M	345	
14	N	50	
15	O	181	
16	P	6	
17	Q	734	
18	R	331	
19	T	58	

2 Entry composition

There are 21 unique types of molecules in this entry. The entry contains 38446 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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					AltConf	Trace
			Total	C	N	O	S		
1	A	1422	11174	7036	1954	2122	62	0	0

- Molecule 2 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	1156	9140	5781	1606	1697	56	0	0

- Molecule 3 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	266	2095	1317	348	417	13	0	0

- Molecule 4 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	178	1434	887	257	288	2	0	0

- Molecule 5 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 1.

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

- Molecule 6 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	84	Total	C	N	O	S	0	0
			679	434	115	127	3		

- Molecule 7 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB7.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	171	Total	C	N	O	S	0	0
			1340	861	222	249	8		

- Molecule 8 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	133	Total	C	N	O	S	0	0
			1068	673	180	211	4		

- Molecule 9 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB9.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	119	Total	C	N	O	S	0	0
			971	596	179	186	10		

- Molecule 10 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 5.

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

- Molecule 11 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB11.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	K	115	Total	C	N	O	S	0	1
			920	590	157	171	2		

- Molecule 12 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 4.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	L	46	Total	C	N	O	S	0	0
			363	224	72	63	4		

- Molecule 13 is a protein called TRANSCRIPTION INITIATION FACTOR IIB.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	M	300	Total	C	N	O	S	0	1
			2202	1384	380	423	15		

- Molecule 14 is a DNA chain called NONTEMPLATE DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	N	50	Total	C	N	O	P	0	0
			975	490	191	246	48		

- Molecule 15 is a protein called TATA-BOX-BINDING PROTEIN.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	O	180	Total	C	N	O	S	0	0
			1416	921	242	247	6		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
O	60	MET	-	EXPRESSION TAG	UNP P13393

- Molecule 16 is a RNA chain called RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	P	6	Total	C	N	O	P	0	0
			123	57	22	39	5		

- Molecule 17 is a protein called TRANSCRIPTION INITIATION FACTOR IIF SUBUNIT ALPHA.

Mol	Chain	Residues	Atoms				AltConf	Trace
17	Q	122	Total	C	N	O	0	0
			606	362	122	122		

- Molecule 18 is a protein called TRANSCRIPTION INITIATION FACTOR IIF SUBUNIT BETA.

Mol	Chain	Residues	Atoms				AltConf	Trace
18	R	105	Total	C	N	O	0	0
			521	311	105	105		

- Molecule 19 is a DNA chain called TEMPLATE DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
19	T	58	Total	C	N	O	P	0	0
			1125	568	206	294	57		

- Molecule 20 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
20	J	1	Total	Zn	0
			1	1	
20	B	1	Total	Zn	0
			1	1	
20	I	2	Total	Zn	0
			2	2	
20	C	1	Total	Zn	0
			1	1	
20	A	2	Total	Zn	0
			2	2	
20	L	1	Total	Zn	0
			1	1	
20	M	1	Total	Zn	0
			1	1	

- Molecule 21 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

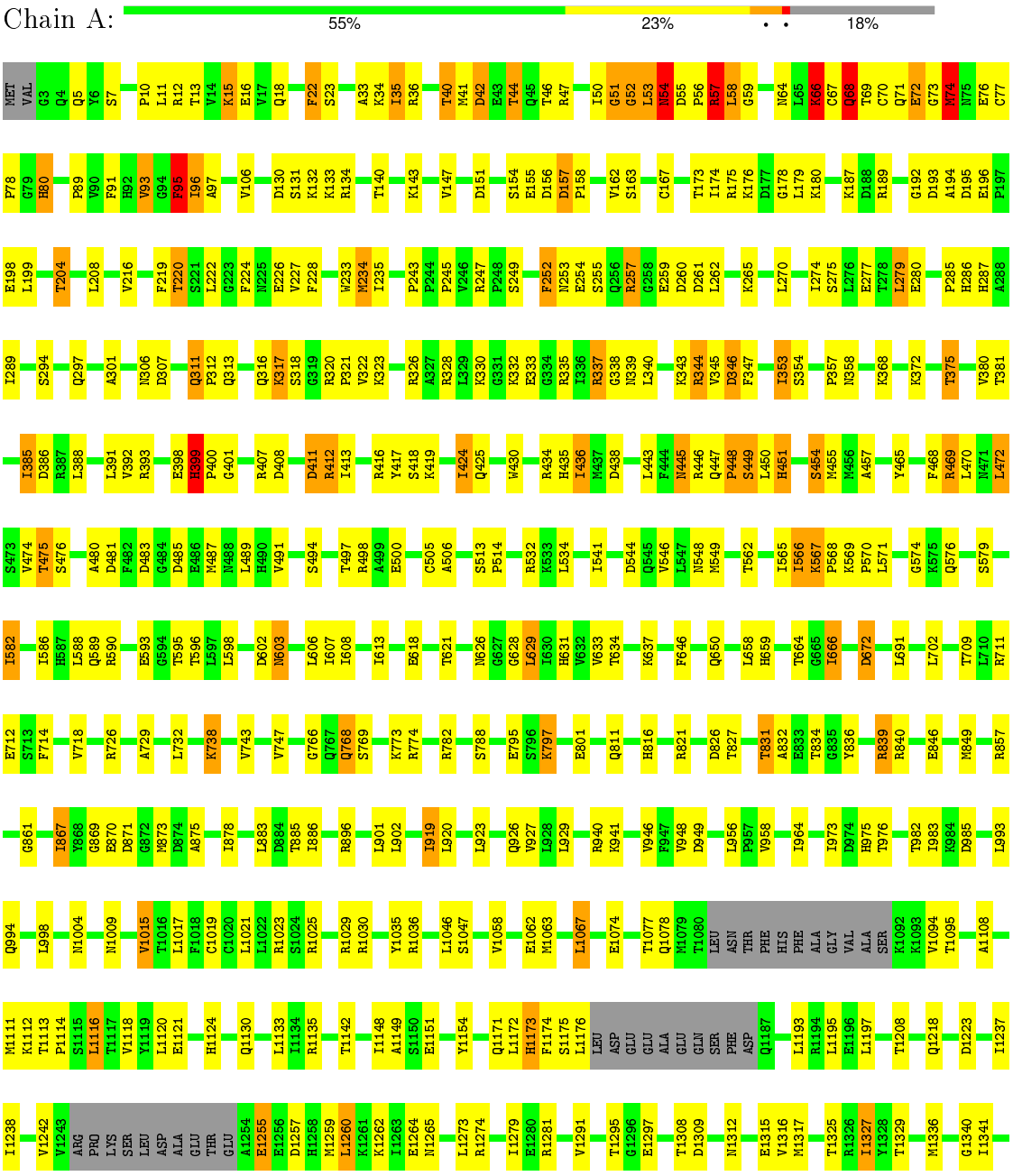
Mol	Chain	Residues	Atoms		AltConf
21	A	1	Total	Mg	0
			1	1	

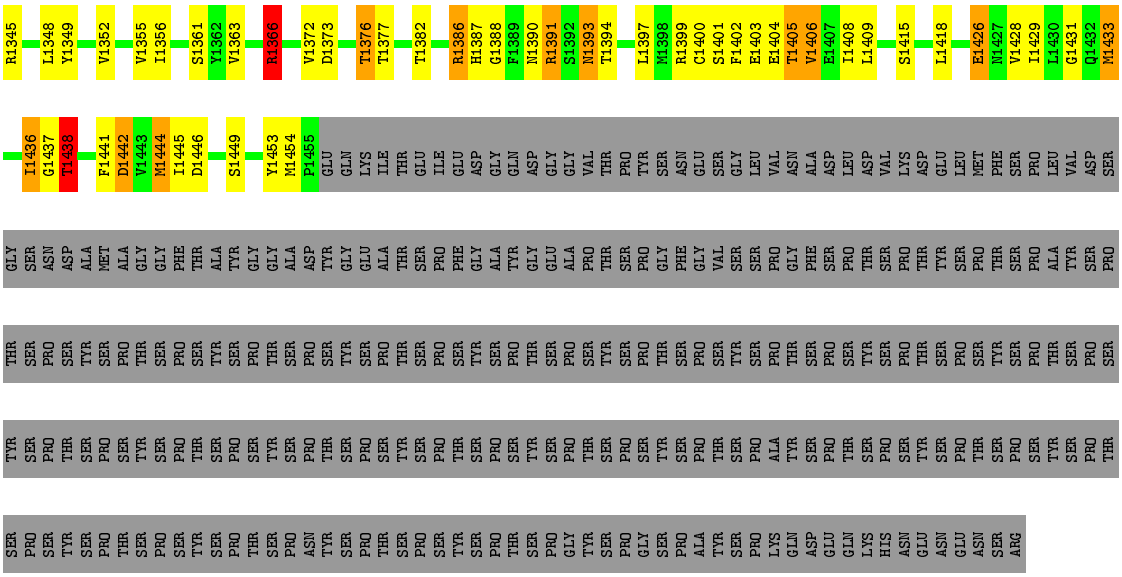
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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

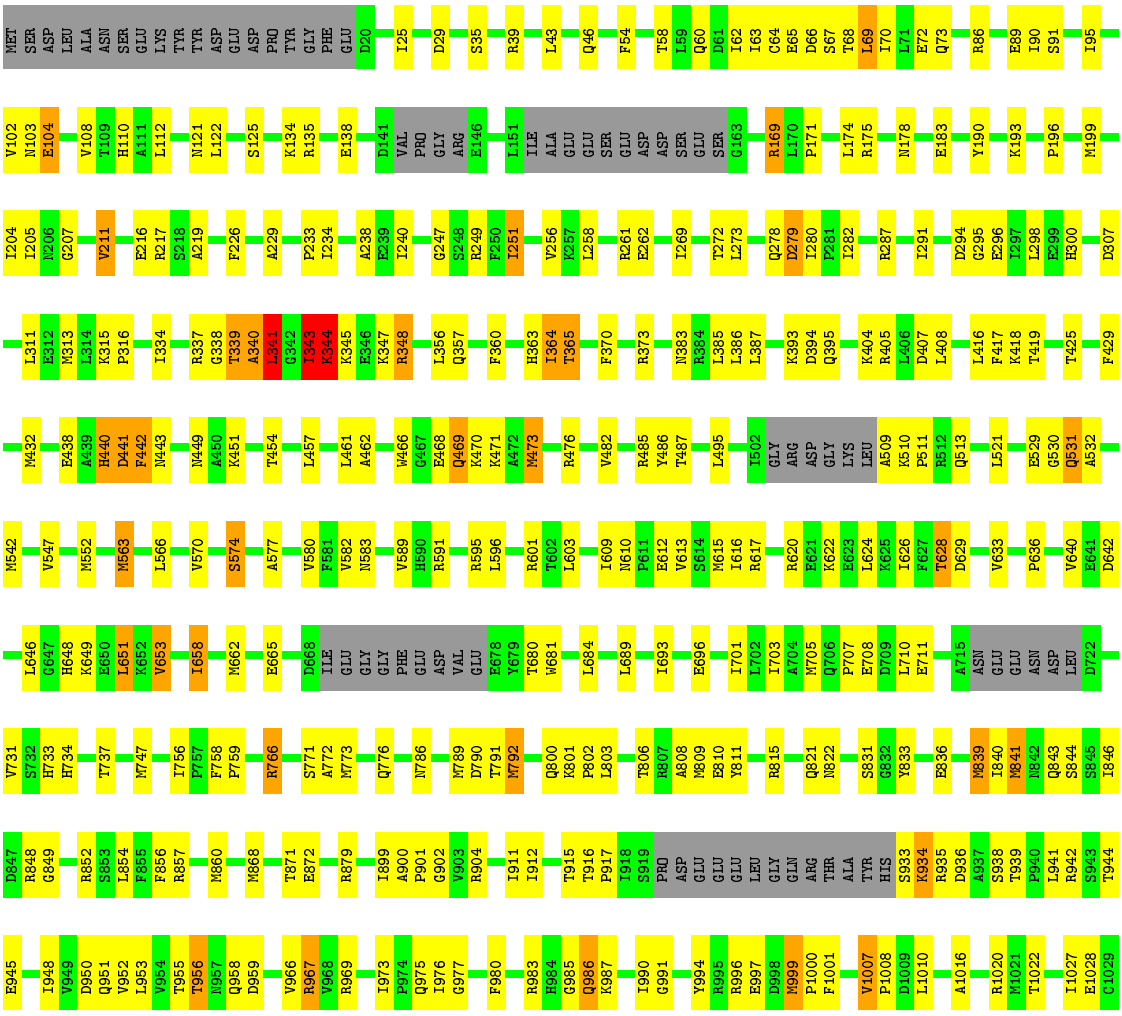
Chain A:

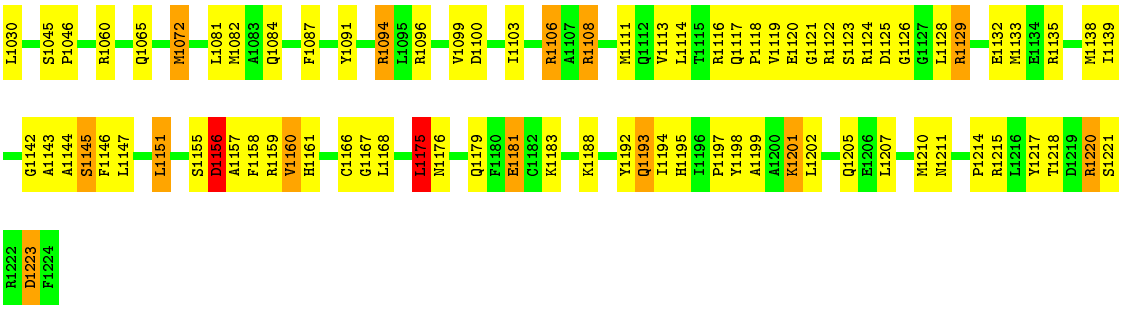




• Molecule 2: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB2

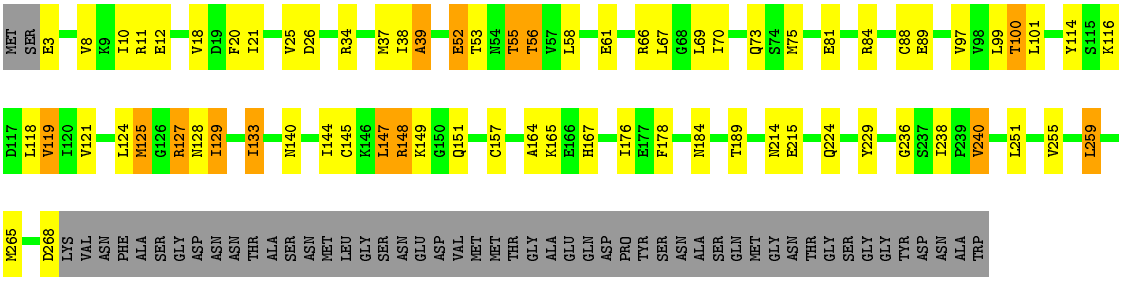
Chain B:





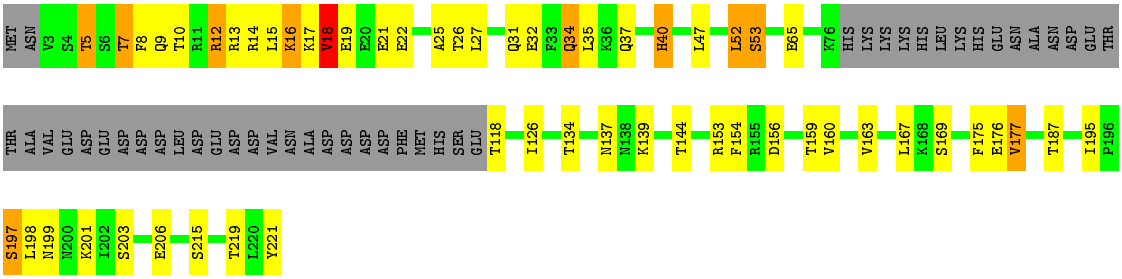
• Molecule 3: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB3

Chain C:



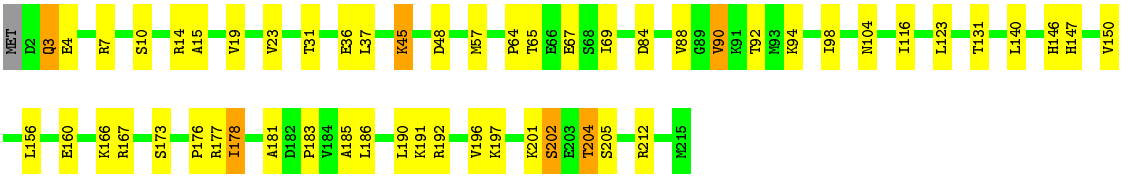
• Molecule 4: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB4

Chain D:



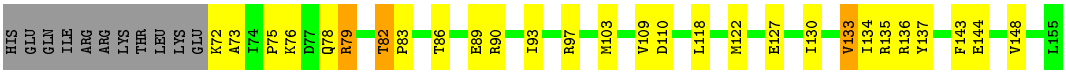
• Molecule 5: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 1

Chain E:



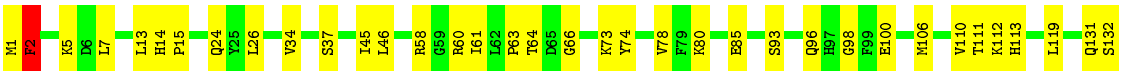
• Molecule 6: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 2

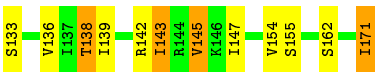
Chain F:



• Molecule 7: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB7

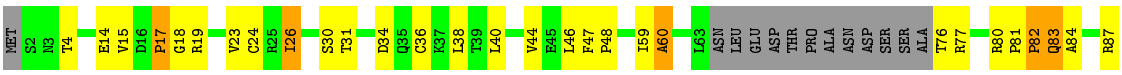
Chain G:

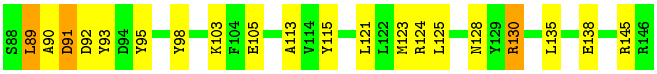




• Molecule 8: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 3

Chain H:





• Molecule 9: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB9

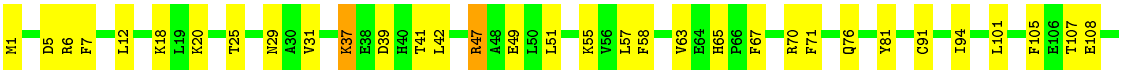
Chain I:

• Molecule 10: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 5

Chain J:

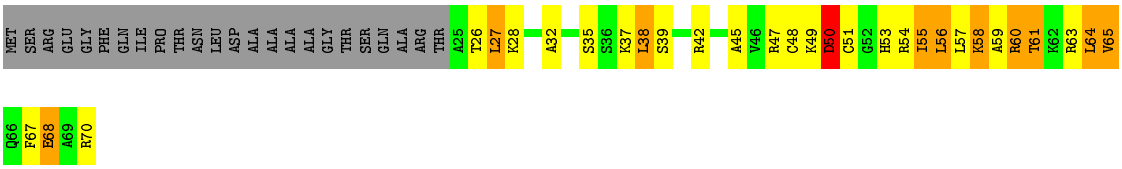
• Molecule 11: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB11

Chain K:



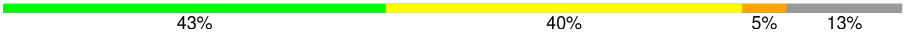
• Molecule 12: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 4

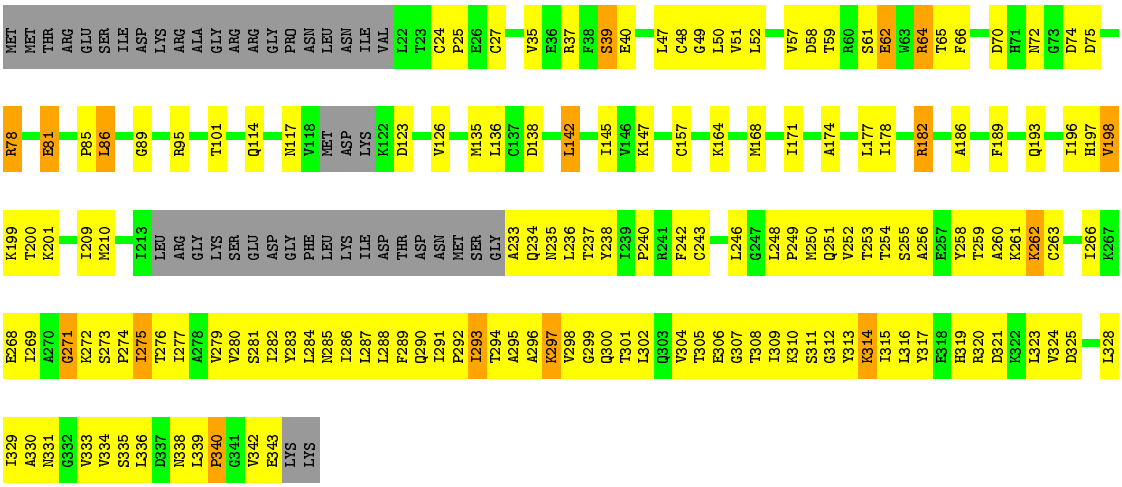
Chain L:



• Molecule 13: TRANSCRIPTION INITIATION FACTOR IIB

Chain M:



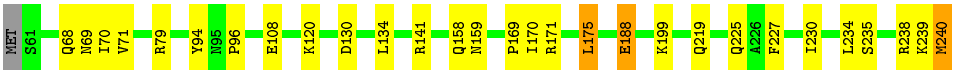


• Molecule 14: NONTEMPLATE DNA

Chain N:

• Molecule 15: TATA-BOX-BINDING PROTEIN

Chain O:

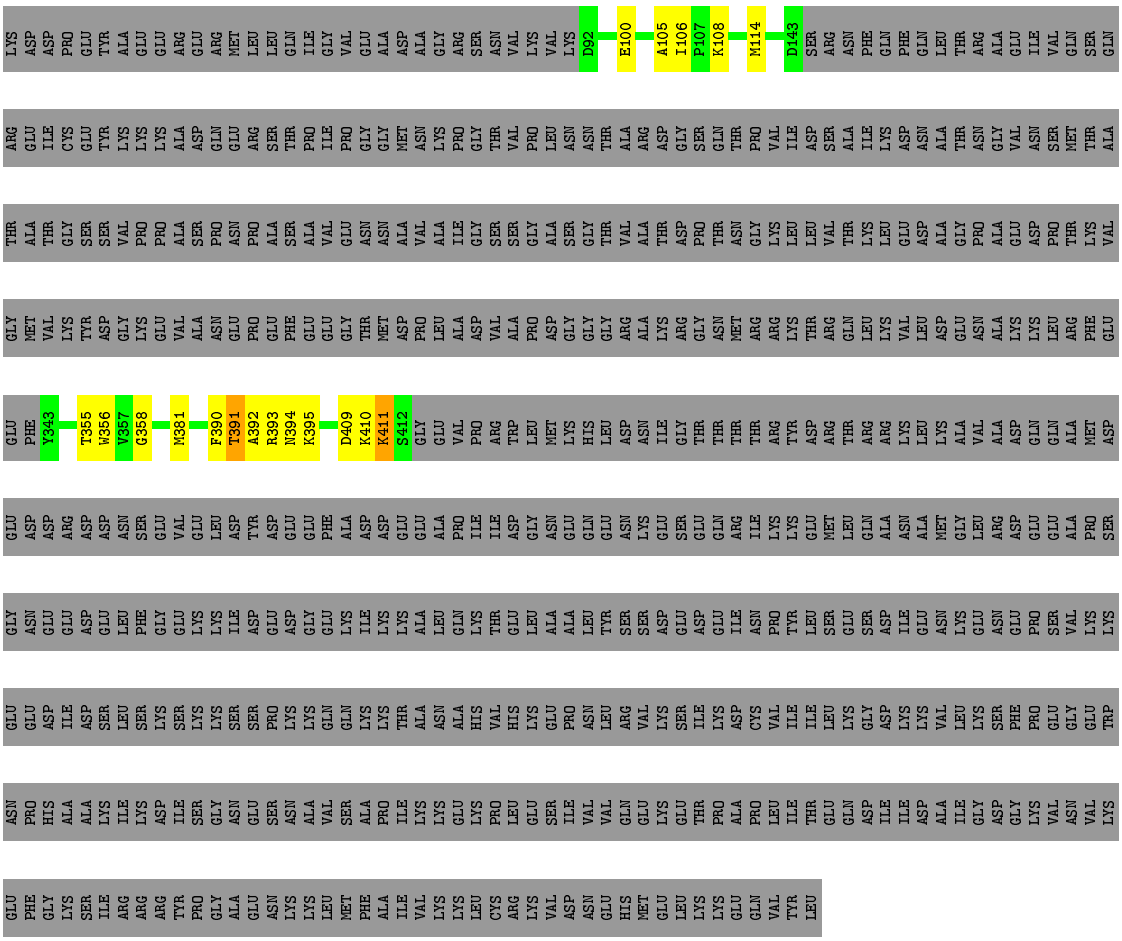


• Molecule 16: RNA

Chain P:

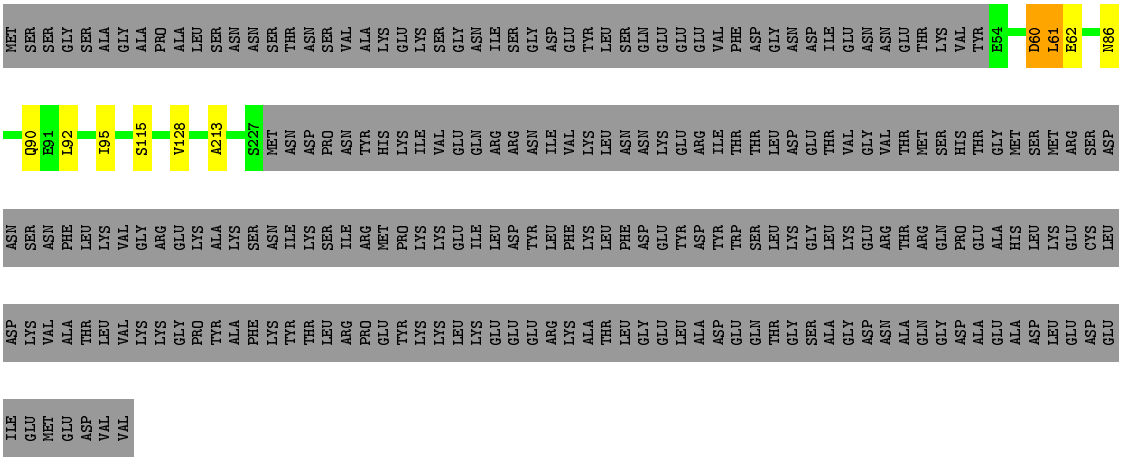
• Molecule 17: TRANSCRIPTION INITIATION FACTOR IIF SUBUNIT ALPHA

Chain Q:



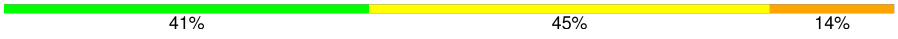
• Molecule 18: TRANSCRIPTION INITIATION FACTOR IIF SUBUNIT BETA

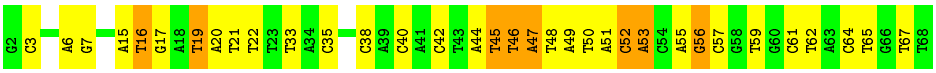
Chain R:



• Molecule 19: TEMPLATE DNA

Chain T:




G2
C3
A6
G7
A15
T16
G17
A18
T19
A20
T21
T22
T23
T33
A34
C35
C38
A39
C40
A41
C42
T43
A44
T45
T46
A47
T48
A49
T50
A51
C52
A53
C54
A55
G56
C57
G58
T59
G60
C61
T62
A63
C64
T65
G66
T67
T68

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	Not provided	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	EACH PARTICLE	Depositor
Microscope	OTHER	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	25	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2500	Depositor
Magnification	37000	Depositor
Image detector	GATAN K2 (4K X 4K)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG

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	#$\|Z\|$ >2	RMSZ	#$\|Z\|$ >2
1	A	0.56	4/11374 (0.0%)	0.81	11/15384 (0.1%)
10	J	0.54	0/541	0.88	1/727 (0.1%)
11	K	0.47	0/938	0.71	0/1267
12	L	0.54	0/365	0.95	0/485
13	M	0.61	0/2232	0.77	1/3031 (0.0%)
14	N	1.13	14/1100 (1.3%)	1.31	5/1625 (0.3%)
15	O	0.58	0/1443	0.78	1/1942 (0.1%)
16	P	0.34	0/137	0.80	0/211
17	Q	0.95	0/604	1.19	3/840 (0.4%)
18	R	0.92	0/520	1.21	2/724 (0.3%)
19	T	1.22	18/1265 (1.4%)	1.44	15/1866 (0.8%)
2	B	0.49	1/9316 (0.0%)	0.74	4/12564 (0.0%)
3	C	0.49	0/2133	0.78	2/2891 (0.1%)
4	D	0.51	0/1444	0.83	2/1935 (0.1%)
5	E	0.48	0/1788	0.71	0/2406
6	F	0.62	0/691	0.81	0/933
7	G	0.52	0/1368	0.81	0/1844
8	H	0.50	0/1086	0.80	0/1470
9	I	0.47	0/989	0.78	0/1331
All	All	0.60	37/39334 (0.1%)	0.85	47/53476 (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
1	A	0	2
17	Q	0	1
2	B	0	1
All	All	0	4

The worst 5 of 37 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1436	ILE	C-N	14.99	1.60	1.33
1	A	95	PHE	C-N	-14.47	1.00	1.34
19	T	53	DA	P-O5'	-11.71	1.48	1.59
1	A	234	MET	C-N	-11.62	1.07	1.34
19	T	47	DA	C1'-N9	-10.46	1.32	1.47

The worst 5 of 47 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	234	MET	O-C-N	-14.63	99.29	122.70
19	T	46	DT	O4'-C4'-C3'	-11.79	98.93	106.00
19	T	6	DA	O3'-P-O5'	-9.12	86.68	104.00
19	T	42	DC	O5'-P-OP1	-9.08	97.53	105.70
14	N	26	DT	O4'-C4'-C3'	-9.01	100.59	106.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	234	MET	Mainchain
1	A	95	PHE	Mainchain
2	B	43	LEU	Mainchain
17	Q	410	LYS	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	11174	0	11223	407	0
2	B	9140	0	9111	313	0
3	C	2095	0	2051	43	0
4	D	1434	0	1460	79	0
5	E	1752	0	1776	28	0
6	F	679	0	701	26	0
7	G	1340	0	1357	33	0
8	H	1068	0	1040	21	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
9	I	971	0	927	15	0
10	J	532	0	542	14	0
11	K	920	0	929	19	0
12	L	363	0	386	26	0
13	M	2202	0	2154	391	0
14	N	975	0	567	22	0
15	O	1416	0	1491	31	0
16	P	123	0	66	1	0
17	Q	606	0	256	30	0
18	R	521	0	216	5	0
19	T	1125	0	661	64	0
20	A	2	0	0	0	0
20	B	1	0	0	0	0
20	C	1	0	0	0	0
20	I	2	0	0	0	0
20	J	1	0	0	0	0
20	L	1	0	0	0	0
20	M	1	0	0	0	0
21	A	1	0	0	0	0
All	All	38446	0	36914	1219	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 16.

The worst 5 of 1219 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:B:868:MET:CE	13:M:182:ARG:HG2	1.34	1.54
1:A:867:ILE:CG1	1:A:867:ILE:CD1	1.83	1.53
17:Q:356:TRP:CA	17:Q:392:ALA:HA	1.06	1.52
2:B:1215:ARG:NH1	4:D:15:LEU:HD13	1.27	1.49
17:Q:356:TRP:HA	17:Q:392:ALA:CA	0.93	1.40

There are no symmetry-related clashes.

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 PDB entries followed by that with respect to all EM entries.

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	1414/1733 (82%)	1248 (88%)	115 (8%)	51 (4%)	4 38
2	B	1140/1224 (93%)	1019 (89%)	85 (8%)	36 (3%)	5 41
3	C	264/318 (83%)	242 (92%)	20 (8%)	2 (1%)	24 69
4	D	174/221 (79%)	148 (85%)	18 (10%)	8 (5%)	3 33
5	E	212/215 (99%)	195 (92%)	13 (6%)	4 (2%)	10 52
6	F	82/155 (53%)	75 (92%)	7 (8%)	0	100 100
7	G	169/171 (99%)	158 (94%)	8 (5%)	3 (2%)	11 53
8	H	129/146 (88%)	106 (82%)	14 (11%)	9 (7%)	1 22
9	I	117/122 (96%)	98 (84%)	16 (14%)	3 (3%)	7 45
10	J	63/70 (90%)	51 (81%)	9 (14%)	3 (5%)	3 32
11	K	113/120 (94%)	109 (96%)	4 (4%)	0	100 100
12	L	44/70 (63%)	27 (61%)	9 (20%)	8 (18%)	0 4
13	M	294/345 (85%)	267 (91%)	24 (8%)	3 (1%)	19 65
15	O	178/181 (98%)	174 (98%)	4 (2%)	0	100 100
17	Q	118/734 (16%)	106 (90%)	11 (9%)	1 (1%)	24 69
18	R	103/331 (31%)	96 (93%)	3 (3%)	4 (4%)	4 36
All	All	4614/6156 (75%)	4119 (89%)	360 (8%)	135 (3%)	9 43

5 of 135 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	74	MET
1	A	96	ILE
1	A	189	ARG
1	A	195	ASP
1	A	286	HIS

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 PDB entries followed by that with respect to all EM entries.

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	1240/1520 (82%)	1065 (86%)	175 (14%)	4 26
2	B	985/1061 (93%)	868 (88%)	117 (12%)	6 31
3	C	234/274 (85%)	206 (88%)	28 (12%)	6 31
4	D	160/200 (80%)	129 (81%)	31 (19%)	2 12
5	E	196/197 (100%)	175 (89%)	21 (11%)	8 36
6	F	74/137 (54%)	67 (90%)	7 (10%)	11 41
7	G	152/152 (100%)	135 (89%)	17 (11%)	7 33
8	H	117/128 (91%)	103 (88%)	14 (12%)	6 31
9	I	113/116 (97%)	106 (94%)	7 (6%)	23 60
10	J	60/65 (92%)	49 (82%)	11 (18%)	2 14
11	K	99/102 (97%)	87 (88%)	12 (12%)	6 31
12	L	40/57 (70%)	27 (68%)	13 (32%)	0 2
13	M	231/299 (77%)	208 (90%)	23 (10%)	9 38
15	O	152/153 (99%)	140 (92%)	12 (8%)	15 51
All	All	3853/4461 (86%)	3365 (87%)	488 (13%)	10 29

5 of 488 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	570	VAL
2	B	1123	SER
12	L	65	VAL
2	B	612	GLU
2	B	839	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	842	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	1195	HIS
13	M	300	GLN
2	B	975	GLN
2	B	1025	HIS

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
16	P	5/6 (83%)	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

Of 10 ligands modelled in this entry, 10 are monoatomic - leaving 0 for Mogul analysis.

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

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.