



wwPDB EM Map/Model Validation Report ⓘ

Jun 14, 2016 – 02:39 PM EDT

PDB ID : 5IYC
EMDB ID: : EMD-8137
Title : Human core-PIC in the initial transcribing state
Authors : He, Y.; Yan, C.; Fang, J.; Inouye, C.; Tjian, R.; Ivanov, I.; Nogales, E.
Deposited on : 2016-03-24
Resolution : 3.90 Å(reported)
Based on PDB ID : ?

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) : rb-20027790

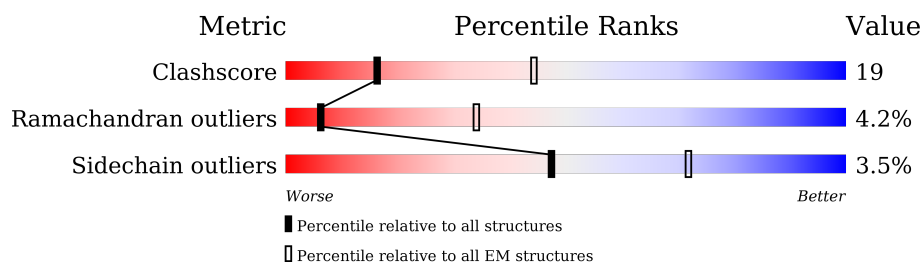
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.90 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)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	1970	
2	B	1174	
3	C	275	
4	D	142	
5	E	210	
6	F	127	
7	G	172	
8	H	150	
9	I	125	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	J	67	 63% 28% 6% .
11	K	117	 79% 19% .
12	L	58	 50% 22% 5% . 21%
13	M	316	 67% 26% . . .
14	N	376	 21% 7% . . 70%
15	O	109	 68% 23% 9%
16	P	339	 39% 14% . 45%
17	Q	439	 25% 14% . 59%
18	R	291	 32% 21% . . 43%
19	S	517	 18% 8% 73%
20	T	249	 62% 23% . . 11%
21	U	301	 40% 13% . . 44%
22	X	80	 55% 45%
23	Y	80	 73% 28%

2 Entry composition

There are 25 unique types of molecules in this entry. The entry contains 47927 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
1	A	1454	Total	C	N	O	S	0	0
			11515	7234	2058	2150	73		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	1165	Total	C	N	O	S	0	0
			9317	5878	1637	1738	64		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	275	Total	C	N	O	S	0	0
			2213	1386	380	440	7		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	129	Total	C	N	O	S	0	0
			1062	665	179	214	4		

- Molecule 5 is a protein called DNA-directed RNA polymerase II subunit RPB5.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	210	Total	C	N	O	S	0	0
			1723	1088	301	325	9		

- Molecule 6 is a protein called DNA-directed RNA polymerase II subunit RPB6.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	86	Total	C	N	O	S	0	0
			689	437	120	127	5		

- 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
			1351	875	219	249	8		

- Molecule 8 is a protein called DNA-directed RNA polymerase II subunit RPB8.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	150	Total	C	N	O	S	0	0
			1205	764	196	239	6		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	125	Total	C	N	O	S	0	0
			1013	626	177	198	12		

- Molecule 10 is a protein called DNA-directed RNA polymerase II subunit RPB10.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	J	67	Total	C	N	O	S	0	0
			533	345	90	92	6		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
11	K	117	Total	C	N	O	S	0	0
			937	604	154	177	2		

- Molecule 12 is a protein called DNA-directed RNA polymerase II subunit RPB12.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	L	46	Total	C	N	O	S	0	0
			388	241	75	66	6		

- Molecule 13 is a protein called Transcription initiation factor IIB.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	M	310	Total	C	N	O	S	0	0
			2391	1490	426	457	18		

- Molecule 14 is a protein called Transcription initiation factor IIA subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	N	113	Total	C	N	O	S	0	0
			930	585	152	189	4		

- Molecule 15 is a protein called Transcription initiation factor IIA subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	O	99	Total	C	N	O	S	0	0
			806	510	142	151	3		

- Molecule 16 is a protein called TATA-box-binding protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	P	185	Total	C	N	O	S	0	0
			1462	946	257	252	7		

- Molecule 17 is a protein called General transcription factor IIE subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
17	Q	180	Total	C	N	O	S	0	0
			1484	938	262	273	11		

- Molecule 18 is a protein called Transcription initiation factor IIE subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
18	R	165	Total	C	N	O	S	0	0
			1357	865	235	253	4		

- Molecule 19 is a protein called General transcription factor IIF subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
19	S	138	Total	C	N	O	S	0	0
			1138	719	208	208	3		

- Molecule 20 is a protein called General transcription factor IIF subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
20	T	222	Total	C	N	O	S	0	0
			1788	1127	320	338	3		

- Molecule 21 is a protein called Transcription elongation factor A protein 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
21	U	170	Total	C	N	O	S	0	0
			1343	818	247	263	15		

- Molecule 22 is a DNA chain called SCP-X.

Mol	Chain	Residues	Atoms					AltConf	Trace
22	X	80	Total	C	N	O	P	0	0
			1645	785	292	489	79		

- Molecule 23 is a DNA chain called SCP-Y.

Mol	Chain	Residues	Atoms					AltConf	Trace
23	Y	80	Total	C	N	O	P	0	0
			1624	771	291	483	79		

- Molecule 24 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
24	A	2	Total	Mg	0
			2	2	

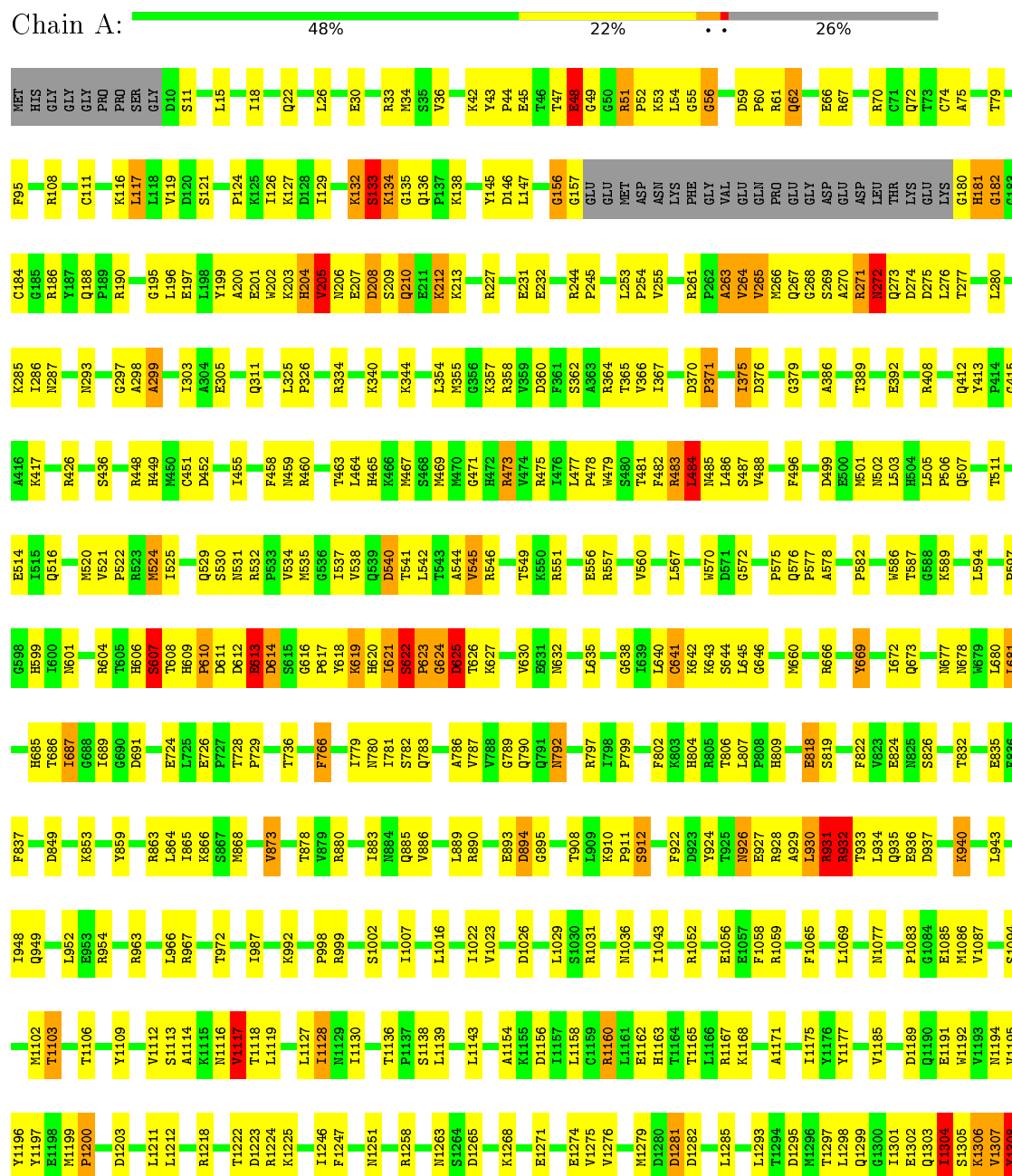
- Molecule 25 is ZINC ION (three-letter code: ZN) (formula: Zn).

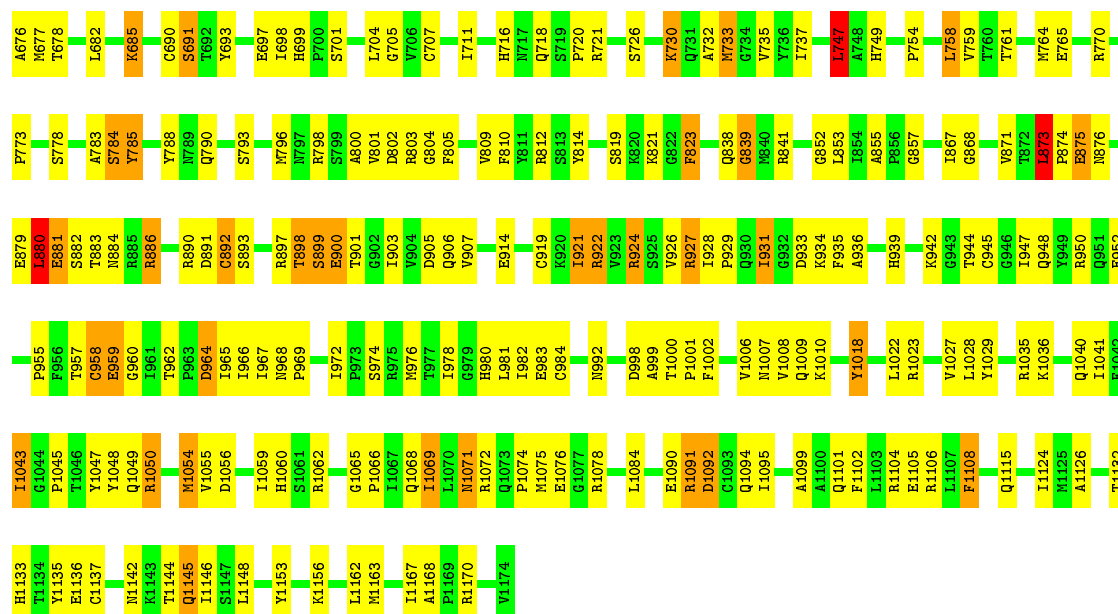
Mol	Chain	Residues	Atoms		AltConf
25	J	1	Total	Zn	0
			1	1	
25	Q	1	Total	Zn	0
			1	1	
25	B	1	Total	Zn	0
			1	1	
25	I	2	Total	Zn	0
			2	2	
25	C	1	Total	Zn	0
			1	1	
25	A	2	Total	Zn	0
			2	2	
25	U	1	Total	Zn	0
			1	1	
25	L	1	Total	Zn	0
			1	1	
25	M	1	Total	Zn	0
			1	1	

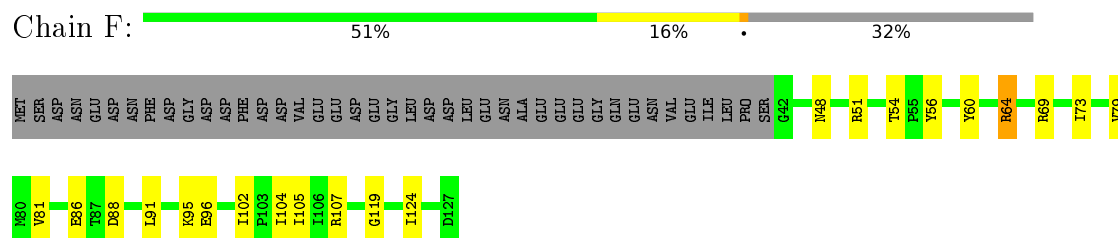
3 Residue-property plots

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

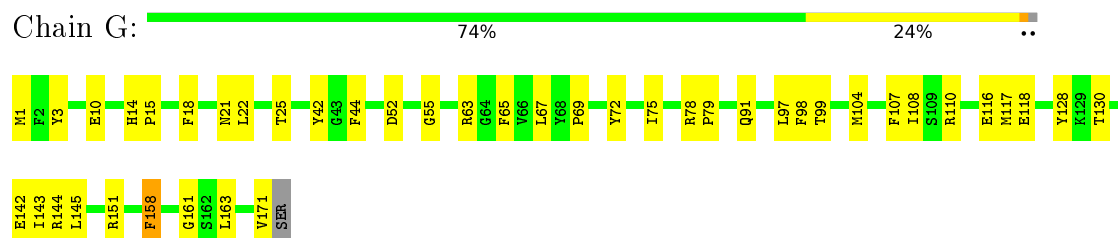
- Molecule 1: DNA-directed RNA polymerase II subunit RPB1



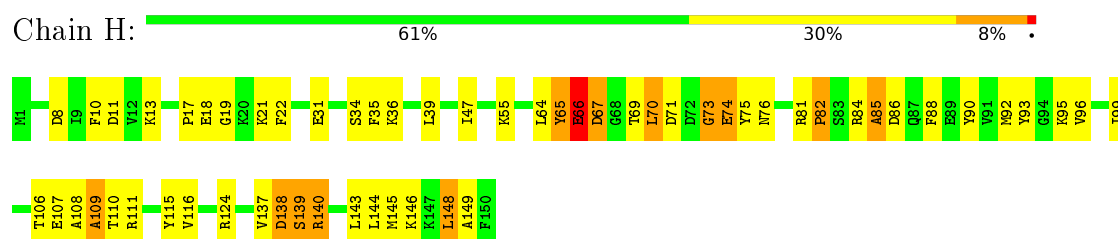




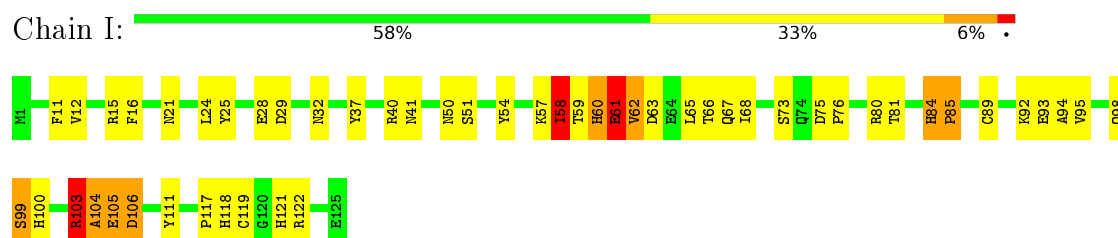
- Molecule 7: DNA-directed RNA polymerase II subunit RPB7



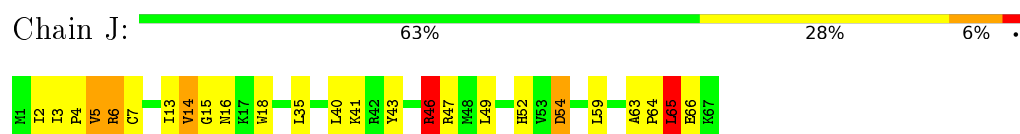
- Molecule 8: DNA-directed RNA polymerase II subunit RPB8



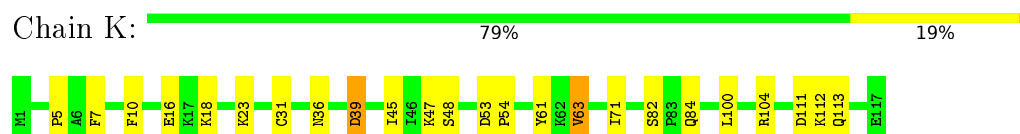
- Molecule 9: DNA-directed RNA polymerase II subunit RPB9



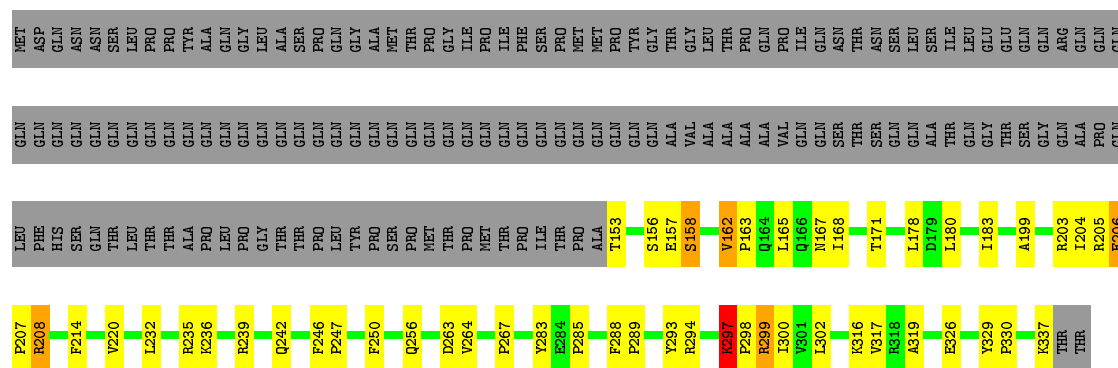
- Molecule 10: DNA-directed RNA polymerase II subunit RPB10



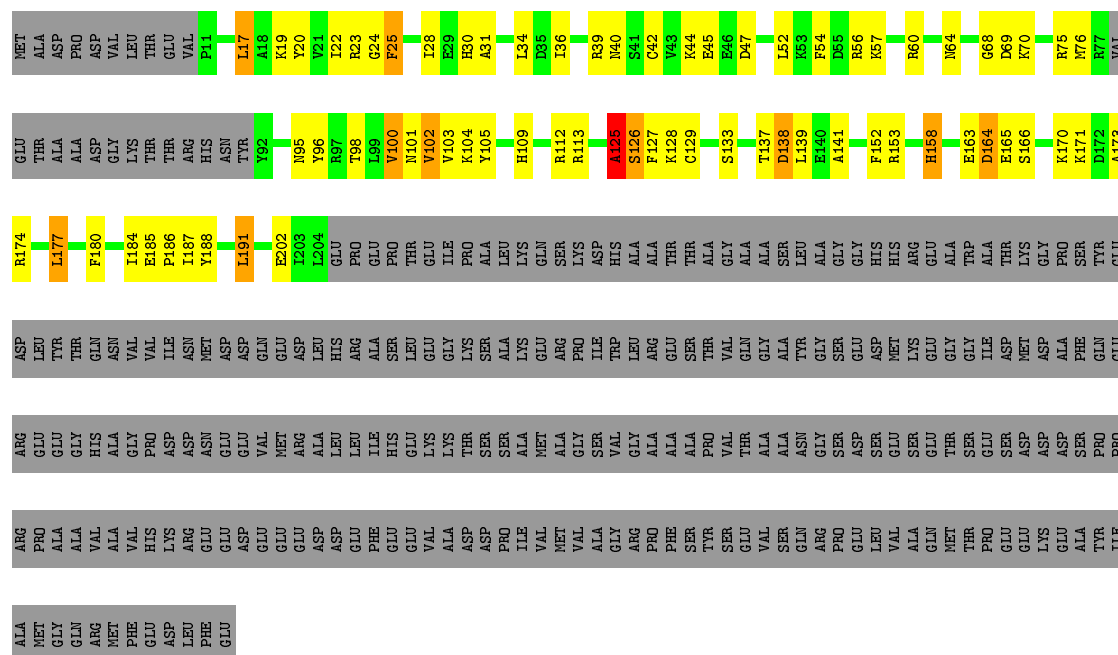
- Molecule 11: DNA-directed RNA polymerase II subunit RPB11-a



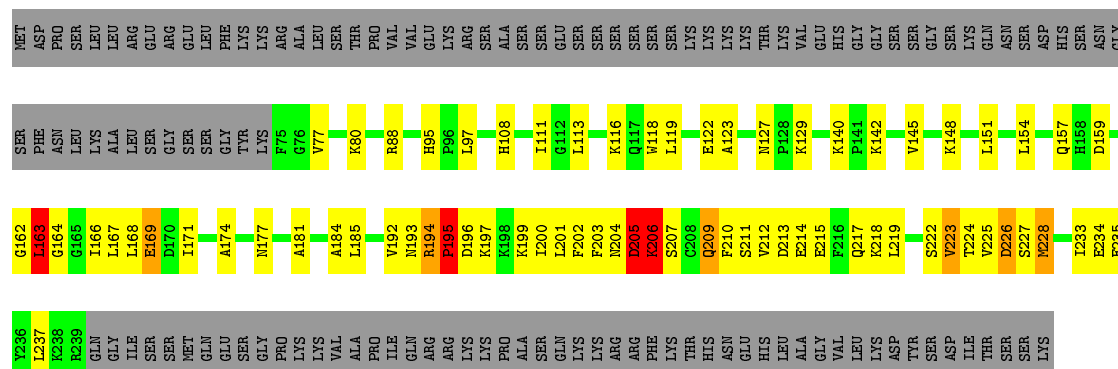
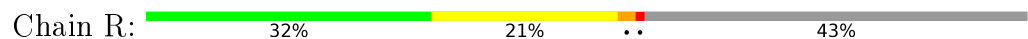
- Chain P:  30% 14% 45%



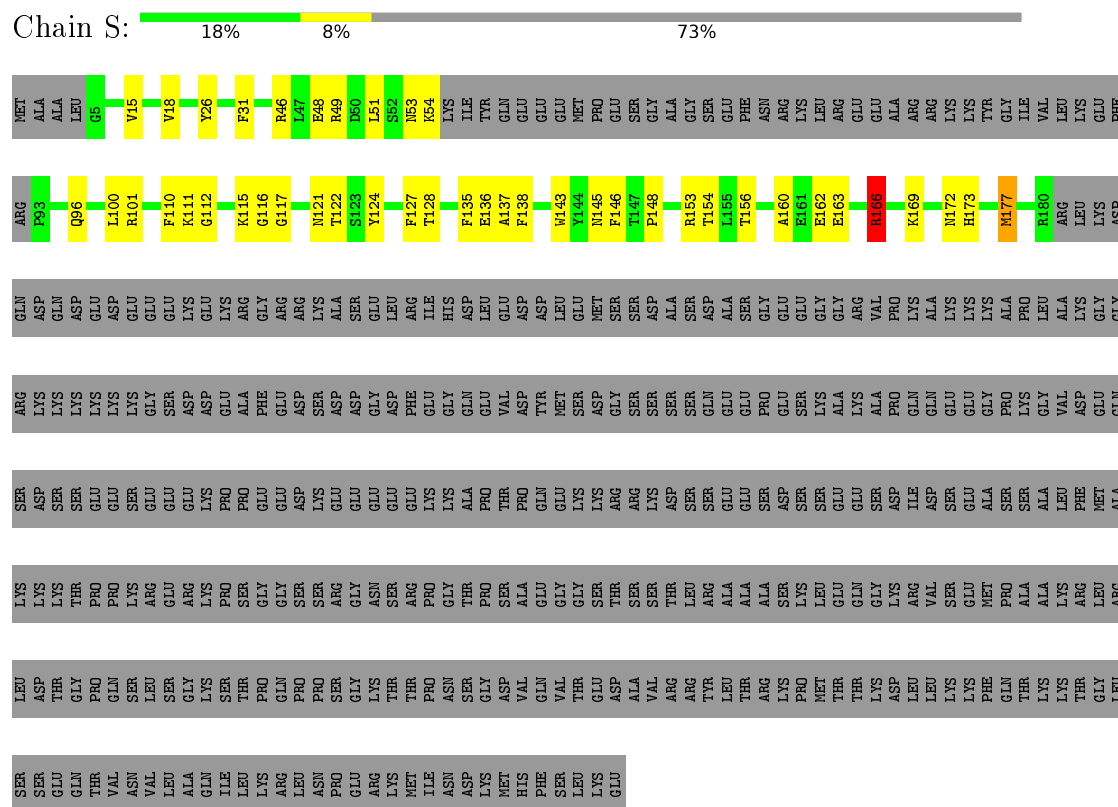
- Molecule 17: General transcription factor IIE subunit 1



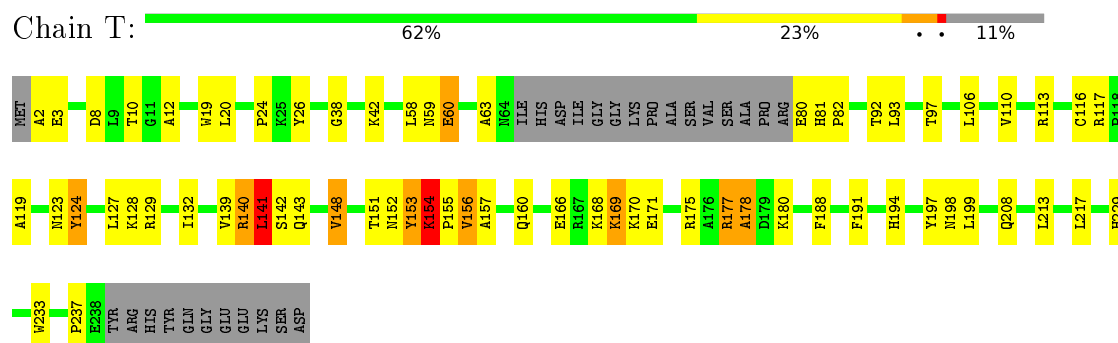
- Molecule 18: Transcription initiation factor IIE subunit beta



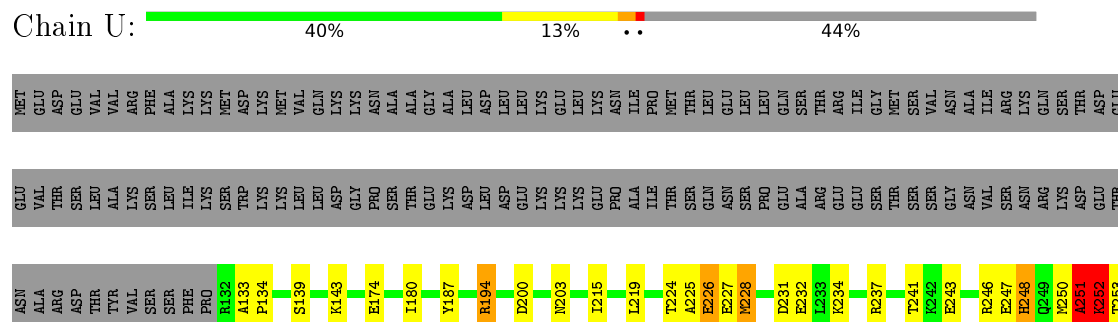
- Molecule 19: General transcription factor IIF subunit 1

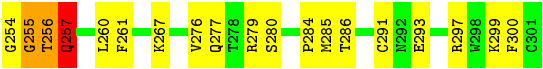


- Molecule 20: General transcription factor IIF subunit 2

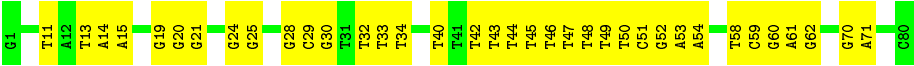


- Molecule 21: Transcription elongation factor A protein 1





• Molecule 22: SCP-X



• Molecule 23: SCP-Y



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	90590	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	Not provided	Depositor
Minimum defocus (nm)	2000	Depositor
Maximum defocus (nm)	4000	Depositor
Magnification	27500	Depositor
Image detector	Not provided	Depositor

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: 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	0.62	4/11727 (0.0%)	0.84	24/15833 (0.2%)
10	J	0.75	0/542	0.96	2/730 (0.3%)
11	K	0.49	0/956	0.64	0/1294
12	L	0.53	0/394	0.70	0/524
13	M	0.40	0/2429	0.86	11/3281 (0.3%)
14	N	0.26	0/945	0.68	3/1274 (0.2%)
15	O	0.24	0/816	0.49	0/1105
16	P	0.29	0/1489	0.54	1/2005 (0.0%)
17	Q	0.30	0/1507	0.62	1/2023 (0.0%)
18	R	0.51	0/1380	1.04	5/1854 (0.3%)
19	S	0.25	0/1167	0.54	1/1576 (0.1%)
2	B	0.75	9/9503 (0.1%)	0.92	27/12831 (0.2%)
20	T	0.37	2/1817 (0.1%)	0.70	2/2445 (0.1%)
21	U	0.28	0/1358	0.64	3/1820 (0.2%)
22	X	0.66	0/1843	1.01	0/2847
23	Y	0.60	0/1817	0.95	0/2800
3	C	0.60	0/2259	0.85	6/3073 (0.2%)
4	D	0.24	0/1077	0.44	0/1446
5	E	0.46	0/1753	0.71	0/2368
6	F	0.64	0/700	0.78	0/946
7	G	0.29	0/1382	0.53	0/1874
8	H	0.44	0/1227	0.76	3/1654 (0.2%)
9	I	0.37	0/1038	1.08	5/1407 (0.4%)
All	All	0.56	15/49126 (0.0%)	0.82	94/67010 (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
13	M	0	1
14	N	0	2
16	P	0	1
17	Q	0	1
18	R	0	1
19	S	0	1
2	B	0	3
20	T	0	2
7	G	0	1
All	All	0	15

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
20	T	60	GLU	CD-OE1	7.39	1.33	1.25
2	B	191	GLU	CG-CD	6.38	1.61	1.51
20	T	60	GLU	CD-OE2	5.92	1.32	1.25
2	B	1048	TYR	CD2-CE2	-5.68	1.30	1.39
2	B	1047	TYR	CD1-CE1	-5.66	1.30	1.39

The worst 5 of 94 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	I	84	HIS	C-N-CD	-28.89	57.05	120.60
18	R	194	ARG	C-N-CD	-24.66	66.35	120.60
13	M	10	LEU	C-N-CD	-22.77	70.51	120.60
3	C	6	GLN	C-N-CD	-20.32	75.89	120.60
1	A	483	ARG	NE-CZ-NH1	11.54	126.07	120.30

There are no chirality outliers.

5 of 15 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1086	MET	Peptide
1	A	1308	TYR	Peptide
2	B	416	ARG	Sidechain
2	B	525	ASN	Peptide
2	B	873	LEU	Mainchain

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	11515	0	11607	525	0
2	B	9317	0	9308	354	0
3	C	2213	0	2153	90	0
4	D	1062	0	1042	11	0
5	E	1723	0	1745	60	0
6	F	689	0	715	18	0
7	G	1351	0	1358	26	0
8	H	1205	0	1168	44	0
9	I	1013	0	932	55	0
10	J	533	0	553	35	0
11	K	937	0	959	32	0
12	L	388	0	393	26	0
13	M	2391	0	2410	126	0
14	N	930	0	888	23	0
15	O	806	0	818	16	0
16	P	1462	0	1549	58	0
17	Q	1484	0	1494	147	0
18	R	1357	0	1379	205	0
19	S	1138	0	1103	35	0
20	T	1788	0	1819	97	0
21	U	1343	0	1338	45	0
22	X	1645	0	908	34	0
23	Y	1624	0	899	27	0
24	A	2	0	0	0	0
25	A	2	0	0	0	0
25	B	1	0	0	0	0
25	C	1	0	0	0	0
25	I	2	0	0	0	0
25	J	1	0	0	0	0
25	L	1	0	0	0	0
25	M	1	0	0	0	0
25	Q	1	0	0	0	0
25	U	1	0	0	0	0
All	All	47927	0	46538	1735	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 19.

The worst 5 of 1735 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:932:ARG:CG	1:A:943:LEU:CD1	1.74	1.56
18:R:195:PRO:CG	18:R:199:LYS:CB	1.86	1.52
5:E:64:HIS:CE1	5:E:68:PRO:HG3	1.44	1.52
13:M:37:CYS:SG	13:M:39:LEU:HD22	1.46	1.52
1:A:932:ARG:CD	1:A:943:LEU:HD21	1.38	1.50

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 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	1450/1970 (74%)	1270 (88%)	118 (8%)	62 (4%)	3	35
2	B	1163/1174 (99%)	1001 (86%)	110 (10%)	52 (4%)	3	33
3	C	273/275 (99%)	242 (89%)	24 (9%)	7 (3%)	7	47
4	D	127/142 (89%)	120 (94%)	7 (6%)	0	100	100
5	E	208/210 (99%)	191 (92%)	8 (4%)	9 (4%)	3	35
6	F	84/127 (66%)	80 (95%)	4 (5%)	0	100	100
7	G	169/172 (98%)	161 (95%)	8 (5%)	0	100	100
8	H	148/150 (99%)	117 (79%)	14 (10%)	17 (12%)	0	9
9	I	123/125 (98%)	90 (73%)	22 (18%)	11 (9%)	1	17
10	J	65/67 (97%)	56 (86%)	5 (8%)	4 (6%)	2	27
11	K	115/117 (98%)	106 (92%)	8 (7%)	1 (1%)	21	65
12	L	44/58 (76%)	32 (73%)	9 (20%)	3 (7%)	1	24
13	M	308/316 (98%)	262 (85%)	28 (9%)	18 (6%)	2	28
14	N	109/376 (29%)	95 (87%)	11 (10%)	3 (3%)	6	46

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
15	O	97/109 (89%)	96 (99%)	1 (1%)	0	100	100
16	P	183/339 (54%)	170 (93%)	7 (4%)	6 (3%)	5	42
17	Q	176/439 (40%)	160 (91%)	8 (4%)	8 (4%)	3	33
18	R	163/291 (56%)	142 (87%)	15 (9%)	6 (4%)	4	39
19	S	134/517 (26%)	119 (89%)	11 (8%)	4 (3%)	5	44
20	T	218/249 (88%)	190 (87%)	16 (7%)	12 (6%)	2	30
21	U	168/301 (56%)	142 (84%)	18 (11%)	8 (5%)	3	32
All	All	5525/7524 (73%)	4842 (88%)	452 (8%)	231 (4%)	6	35

5 of 231 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	51	ARG
1	A	70	ARG
1	A	133	SER
1	A	205	VAL
1	A	208	ASP

5.3.2 Protein sidechains ⓘ

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	1279/1748 (73%)	1224 (96%)	55 (4%)	35	72
2	B	1020/1028 (99%)	982 (96%)	38 (4%)	41	75
3	C	252/252 (100%)	243 (96%)	9 (4%)	42	76
4	D	119/126 (94%)	119 (100%)	0	100	100
5	E	192/192 (100%)	187 (97%)	5 (3%)	54	81
6	F	74/111 (67%)	72 (97%)	2 (3%)	52	80
7	G	152/153 (99%)	151 (99%)	1 (1%)	88	94
8	H	131/131 (100%)	127 (97%)	4 (3%)	47	78
9	I	112/112 (100%)	109 (97%)	3 (3%)	52	80

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
10	J	56/56 (100%)	49 (88%)	7 (12%)	6	32
11	K	106/106 (100%)	103 (97%)	3 (3%)	51	79
12	L	43/55 (78%)	41 (95%)	2 (5%)	32	70
13	M	263/268 (98%)	253 (96%)	10 (4%)	40	74
14	N	105/324 (32%)	101 (96%)	4 (4%)	40	74
15	O	90/98 (92%)	88 (98%)	2 (2%)	60	84
16	P	159/293 (54%)	157 (99%)	2 (1%)	76	89
17	Q	164/373 (44%)	158 (96%)	6 (4%)	41	75
18	R	150/261 (58%)	143 (95%)	7 (5%)	32	70
19	S	121/448 (27%)	117 (97%)	4 (3%)	45	77
20	T	196/218 (90%)	193 (98%)	3 (2%)	72	89
21	U	148/266 (56%)	140 (95%)	8 (5%)	27	67
All	All	4932/6619 (74%)	4757 (96%)	175 (4%)	47	76

5 of 175 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	737	ILE
3	C	70	LEU
19	S	163	GLU
2	B	784	SER
2	B	1035	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	790	GLN
3	C	25	ASN
13	M	102	GLN
2	B	941	GLN
2	B	1101	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 13 ligands modelled in this entry, 13 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 [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.