



wwPDB EM Validation Summary Report ⓘ

Jun 3, 2024 – 09:22 PM JST

PDB ID : 8H6E
EMDB ID : EMD-34500
Title : Cryo-EM structure of human exon-defined spliceosome in the late pre-B state.
Authors : Zhang, W.; Zhan, X.; Zhang, X.; Bai, R.; Lei, J.; Yan, C.; Shi, Y.
Deposited on : 2022-10-17
Resolution : 3.20 Å (reported)

This is a wwPDB EM 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/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ 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 ⓘ](#)) were used in the production of this report:

EMDB validation analysis : **FAILED**
Mogul : 1.8.5 (274361), CSD as541be (2020)
MolProbity : 4.02b-467
buster-report : 1.1.7 (2018)
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ : **FAILED**
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.36.2

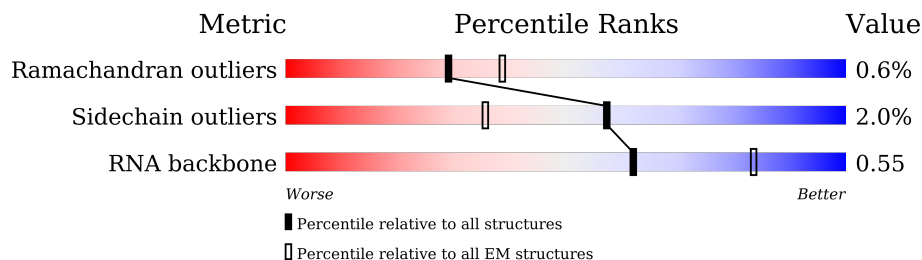
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.20 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)
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826
RNA backbone	4643	859

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	144	15% 25% . 59%
2	6A	107	41% 10% . 45%
3	6a	95	89% . 6%
4	6b	102	70% . 27%
5	6c	139	53% . 47%
6	6d	91	78% . 21%
7	6e	80	85% . 12%
8	6f	103	63% 37%
9	6g	96	63% . 36%






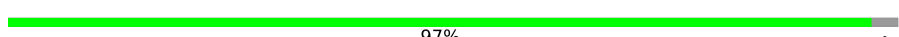
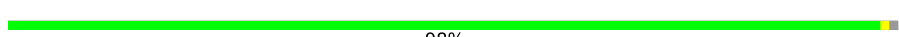



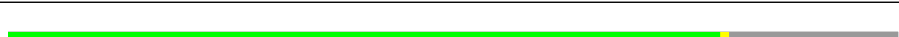


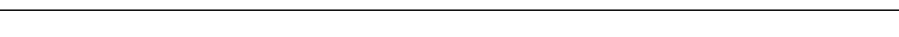
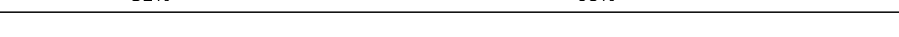
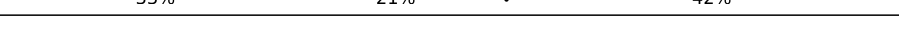



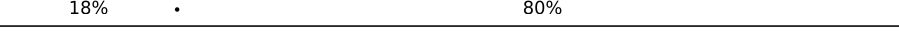



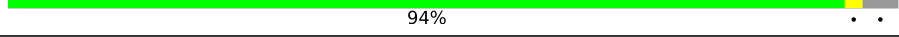

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	5A	117	62% 32%
11	5B	2335	93% 5%
12	5C	972	87% 12%
13	5D	2136	92% 6%
14	5E	357	85% 15%
15	2a	231	37% 63%
15	4a	231	35% 65%
15	5a	231	37% 63%
16	2b	119	69% 31%
16	4b	119	68% 32%
16	5b	119	69% 31%
17	2c	118	71% 28%
17	4c	118	78% 22%
17	5c	118	81% 18%
18	2d	86	86% 14%
18	4d	86	83% 16%
18	5d	86	85% 15%
19	2e	92	86% 14%
19	4e	92	82% 17%
19	5e	92	86% 14%
20	2f	76	88% 12%
20	4f	76	96%
20	5f	76	97%
21	2g	126	63% 37%
21	4g	126	66% 34%



Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
21	5g	126	
22	4A	146	
23	4B	683	
24	4C	522	
25	4D	499	
26	4E	128	
27	4F	142	
28	4G	941	
29	4R	480	
30	4S	800	
31	4T	565	
32	4U	820	
33	4X	155	
34	4Y	1007	
35	2A	188	
36	2B	255	
37	2C	225	
38	2D	793	
39	2E	464	
40	2F	501	
41	2G	1304	
42	2H	895	
43	2I	1217	
44	2J	424	
45	2K	125	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
46	2L	110	 81% 19%
47	2M	86	 74% 23%

2 Entry composition

There are 51 unique types of molecules in this entry. The entry contains 96125 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 RNA chain called pre-mRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
1	A	59	1232	551	193	429	59	0	0

- Molecule 2 is a RNA chain called U6 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	6A	59	1251	558	230	404	59	0	0

- Molecule 3 is a protein called U6 snRNA-associated Sm-like protein LSm2.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
3	6a	89	356	178	89	89	0	0

- Molecule 4 is a protein called U6 snRNA-associated Sm-like protein LSm3.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
4	6b	74	296	148	74	74	0	0

- Molecule 5 is a protein called U6 snRNA-associated Sm-like protein LSm4.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
5	6c	74	296	148	74	74	0	0

- Molecule 6 is a protein called U6 snRNA-associated Sm-like protein LSm5.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
6	6d	72	288	144	72	72	0	0

- Molecule 7 is a protein called U6 snRNA-associated Sm-like protein LSm6.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
7	6e	70	280	140	70	70	0	0

- Molecule 8 is a protein called U6 snRNA-associated Sm-like protein LSm7.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
8	6f	65	260	130	65	65	0	0

- Molecule 9 is a protein called U6 snRNA-associated Sm-like protein LSm8.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
9	6g	61	244	122	61	61	0	0

- Molecule 10 is a RNA chain called U5 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
10	5A	114	2398	1074	398	812	114	0	0

- Molecule 11 is a protein called Pre-mRNA-processing-splicing factor 8.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
11	5B	2214	18281	11782	3177	3243	79	0	0

- Molecule 12 is a protein called 116 kDa U5 small nuclear ribonucleoprotein component.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
12	5C	852	6727	4300	1127	1266	34	0	0

- Molecule 13 is a protein called U5 small nuclear ribonucleoprotein 200 kDa helicase.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
13	5D	2001	16077	10235	2767	2991	84	0	0

- Molecule 14 is a protein called U5 small nuclear ribonucleoprotein 40 kDa protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
14	5E	304	Total	C	N	O	0	0
			1497	888	304	305		

- Molecule 15 is a protein called Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B'.

Mol	Chain	Residues	Atoms				AltConf	Trace
15	5a	85	Total	C	N	O	0	0
			340	170	85	85		
15	4a	82	Total	C	N	O	0	0
			405	241	82	82		
15	2a	85	Total	C	N	O	0	0
			340	170	85	85		

- Molecule 16 is a protein called Small nuclear ribonucleoprotein Sm D1.

Mol	Chain	Residues	Atoms				AltConf	Trace
16	5b	82	Total	C	N	O	0	0
			328	164	82	82		
16	4b	81	Total	C	N	O	0	0
			401	239	81	81		
16	2b	82	Total	C	N	O	0	0
			328	164	82	82		

- Molecule 17 is a protein called Small nuclear ribonucleoprotein Sm D2.

Mol	Chain	Residues	Atoms				AltConf	Trace
17	5c	97	Total	C	N	O	0	0
			388	194	97	97		
17	4c	92	Total	C	N	O	0	0
			455	271	92	92		
17	2c	85	Total	C	N	O	0	0
			340	170	85	85		

- Molecule 18 is a protein called Small nuclear ribonucleoprotein F.

Mol	Chain	Residues	Atoms				AltConf	Trace
18	5d	73	Total	C	N	O	0	0
			292	146	73	73		
18	4d	72	Total	C	N	O	0	0
			351	207	72	72		
18	2d	74	Total	C	N	O	0	0
			296	148	74	74		

- Molecule 19 is a protein called Small nuclear ribonucleoprotein E.

Mol	Chain	Residues	Atoms				AltConf	Trace
19	5e	79	Total	C	N	O	0	0
			316	158	79	79		
19	4e	76	Total	C	N	O	0	0
			376	224	76	76		
19	2e	79	Total	C	N	O	0	0
			316	158	79	79		

- Molecule 20 is a protein called Small nuclear ribonucleoprotein G.

Mol	Chain	Residues	Atoms				AltConf	Trace
20	5f	74	Total	C	N	O	0	0
			296	148	74	74		
20	4f	73	Total	C	N	O	0	0
			358	212	73	73		
20	2f	67	Total	C	N	O	0	0
			268	134	67	67		

- Molecule 21 is a protein called Small nuclear ribonucleoprotein Sm D3.

Mol	Chain	Residues	Atoms				AltConf	Trace
21	5g	77	Total	C	N	O	0	0
			308	154	77	77		
21	4g	83	Total	C	N	O	0	0
			409	243	83	83		
21	2g	80	Total	C	N	O	0	0
			320	160	80	80		

- Molecule 22 is a RNA chain called U4 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
22	4A	125	Total	C	N	O	P	0	0
			2653	1187	471	872	123		

- Molecule 23 is a protein called U4/U6 small nuclear ribonucleoprotein Prp3.

Mol	Chain	Residues	Atoms				AltConf	Trace
23	4B	193	Total	C	N	O	0	0
			954	568	193	193		

- Molecule 24 is a protein called U4/U6 small nuclear ribonucleoprotein Prp4.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
24	4C	359	1765	1047	359	359	0	0

- Molecule 25 is a protein called U4/U6 small nuclear ribonucleoprotein Prp31.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
25	4D	327	1764	1057	357	348	2	0	0

- Molecule 26 is a protein called NHP2-like protein 1.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
26	4E	124	615	367	124	124	0	0

- Molecule 27 is a protein called Thioredoxin-like protein 4A.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
27	4F	141	1169	751	194	214	10	0	0

- Molecule 28 is a protein called Pre-mRNA-processing factor 6.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
28	4G	794	4803	2915	933	943	12	0	0

- Molecule 29 is a protein called RNA-binding protein 42.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
29	4R	106	874	553	160	157	4	0	0

- Molecule 30 is a protein called U4/U6.U5 tri-snRNP-associated protein 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
30	4S	70	549	343	103	100	3	0	0

- Molecule 31 is a protein called U4/U6.U5 tri-snRNP-associated protein 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
31	4T	456	3749	2427	635	673	14	0	0

- Molecule 32 is a protein called Probable ATP-dependent RNA helicase DDX23.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
32	4U	153	1338	831	262	243	2	0	0

- Molecule 33 is a protein called U4/U6.U5 small nuclear ribonucleoprotein 27 kDa protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
33	4X	21	184	115	40	28	1	0	0

- Molecule 34 is a protein called Serine/threonine-protein kinase PRP4 homolog.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
34	4Y	322	1595	951	322	322	0	0

- Molecule 35 is a RNA chain called U2 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
35	2A	109	2311	1032	396	774	109	0	0

- Molecule 36 is a protein called U2 small nuclear ribonucleoprotein A'.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
36	2B	162	648	324	162	162	0	0

- Molecule 37 is a protein called U2 small nuclear ribonucleoprotein B'.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
37	2C	94	376	188	94	94	0	0

- Molecule 38 is a protein called Splicing factor 3A subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
38	2D	181	Total	C	N	O	S	0	0
			969	546	206	216	1		

- Molecule 39 is a protein called Splicing factor 3A subunit 2.

Mol	Chain	Residues	Atoms				AltConf	Trace
39	2E	94	Total	C	N	O	0	0
			376	188	94	94		

- Molecule 40 is a protein called Splicing factor 3A subunit 3.

Mol	Chain	Residues	Atoms				AltConf	Trace
40	2F	423	Total	C	N	O	0	0
			1693	847	423	423		

- Molecule 41 is a protein called Splicing factor 3B subunit 1.

Mol	Chain	Residues	Atoms				AltConf	Trace
41	2G	1048	Total	C	N	O	0	0
			4192	2096	1048	1048		

- Molecule 42 is a protein called Splicing factor 3B subunit 2.

Mol	Chain	Residues	Atoms				AltConf	Trace
42	2H	182	Total	C	N	O	0	0
			728	364	182	182		

- Molecule 43 is a protein called Splicing factor 3B subunit 3.

Mol	Chain	Residues	Atoms				AltConf	Trace
43	2I	1168	Total	C	N	O	0	0
			4672	2336	1168	1168		

- Molecule 44 is a protein called Splicing factor 3B subunit 4.

Mol	Chain	Residues	Atoms				AltConf	Trace
44	2J	78	Total	C	N	O	0	0
			312	156	78	78		

- Molecule 45 is a protein called Splicing factor 3B subunit 6.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
45	2K	108	432	216	108	108	0	0

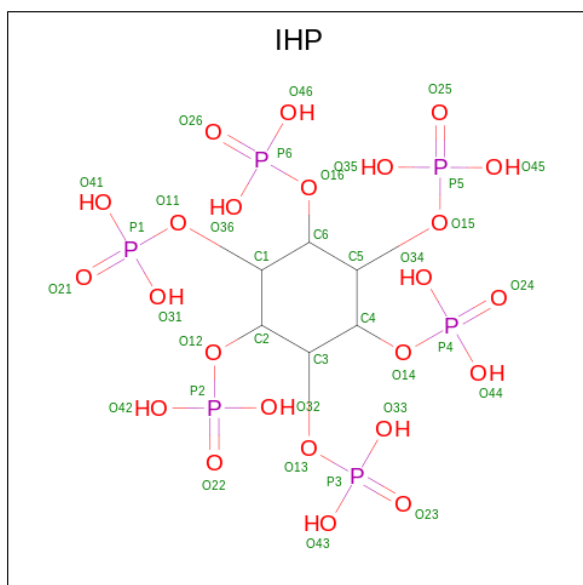
- Molecule 46 is a protein called PHD finger-like domain-containing protein 5A.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
46	2L	89	356	178	89	89	0	0

- Molecule 47 is a protein called Splicing factor 3B subunit 5.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
47	2M	66	264	132	66	66	0	0

- Molecule 48 is INOSITOL HEXAKISPHOSPHATE (three-letter code: IHP) (formula: $C_6H_{18}O_{24}P_6$) (labeled as "Ligand of Interest" by depositor).

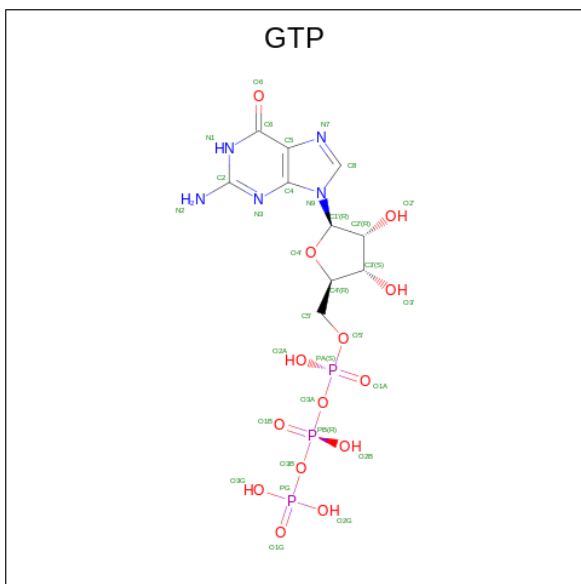


Mol	Chain	Residues	Atoms				AltConf
			Total	C	O	P	
48	5B	1	36	6	24	6	0

- Molecule 49 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	AltConf
49	5C	1	Total Mg 1 1	0

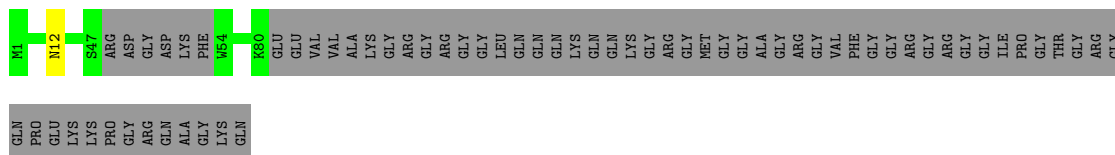
- Molecule 50 is GUANOSINE-5'-TRIPHOSPHATE (three-letter code: GTP) (formula: $C_{10}H_{16}N_5O_{14}P_3$) (labeled as "Ligand of Interest" by depositor).



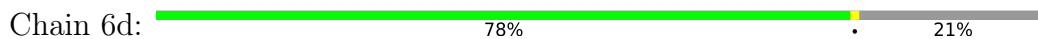
Mol	Chain	Residues	Atoms	AltConf
50	5C	1	Total C N O P 32 10 5 14 3	0

- Molecule 51 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

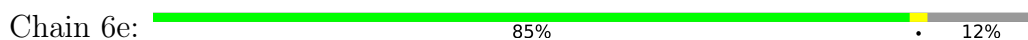
Mol	Chain	Residues	Atoms	AltConf
51	4T	1	Total Zn 1 1	0



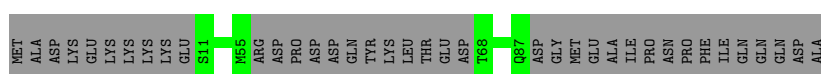
- Molecule 6: U6 snRNA-associated Sm-like protein LSm5



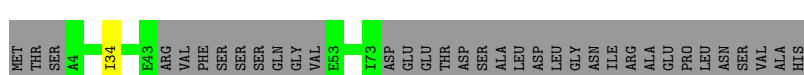
- Molecule 7: U6 snRNA-associated Sm-like protein LSm6



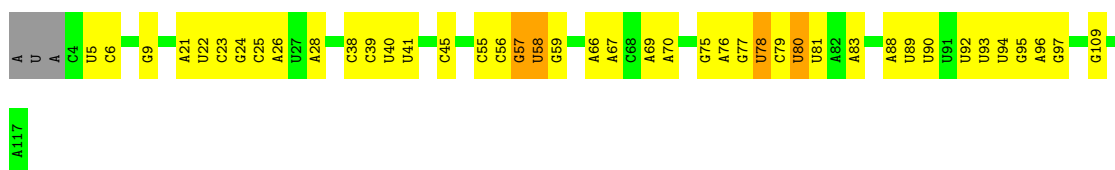
- Molecule 8: U6 snRNA-associated Sm-like protein LSm7



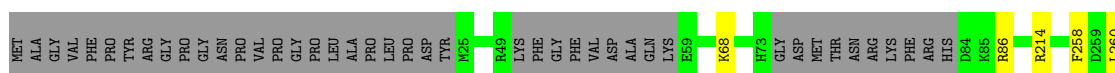
- Molecule 9: U6 snRNA-associated Sm-like protein LSm8

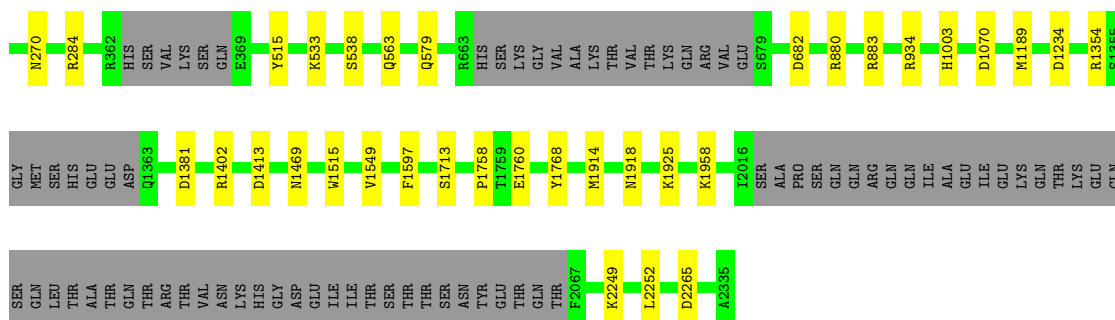


- Molecule 10: U5 snRNA



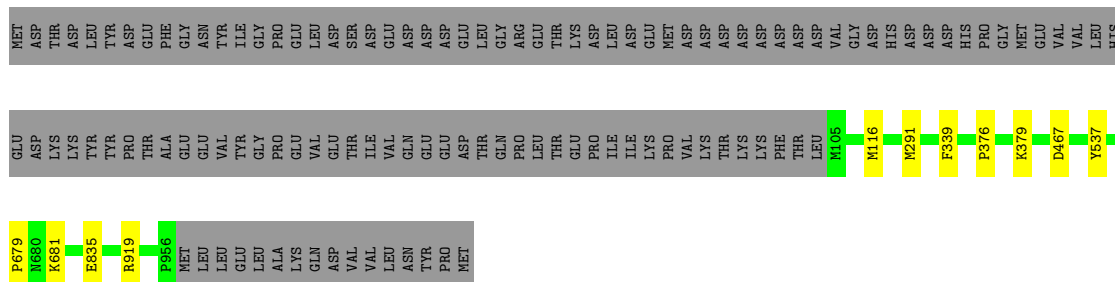
- Molecule 11: Pre-mRNA-processing-splicing factor 8





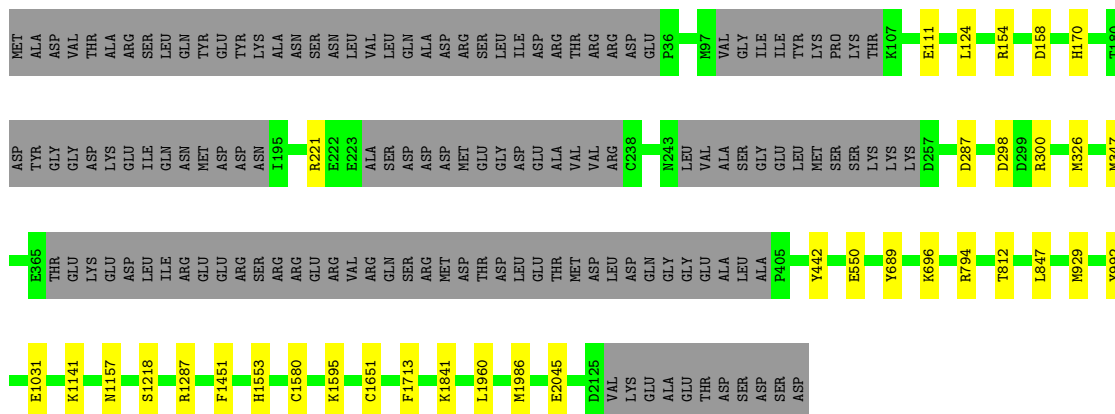
- Molecule 12: 116 kDa U5 small nuclear ribonucleoprotein component

Chain 5C: 87% 12%



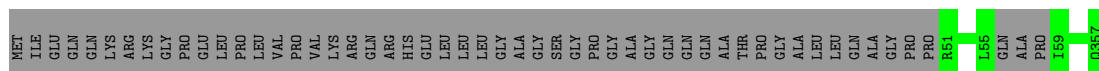
- Molecule 13: U5 small nuclear ribonucleoprotein 200 kDa helicase

Chain 5D: 92% 6%

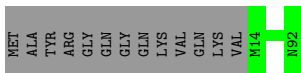
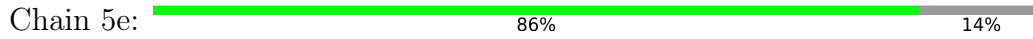


- Molecule 14: U5 small nuclear ribonucleoprotein 40 kDa protein

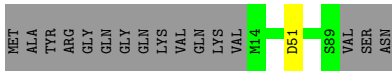
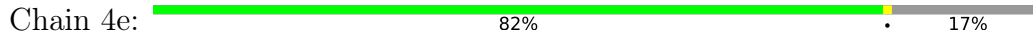
Chain 5E: 85% 15%



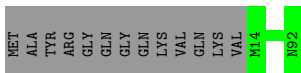
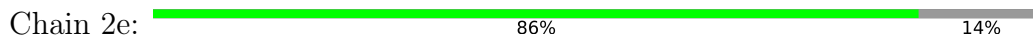
- Molecule 15: Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B'



• Molecule 19: Small nuclear ribonucleoprotein E



• Molecule 19: Small nuclear ribonucleoprotein E



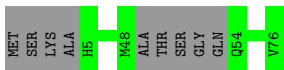
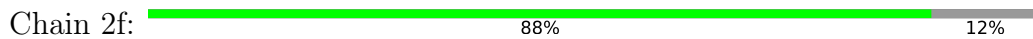
• Molecule 20: Small nuclear ribonucleoprotein G



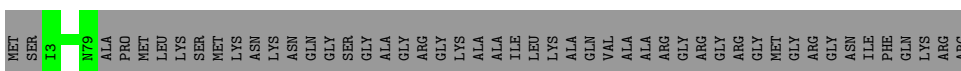
• Molecule 20: Small nuclear ribonucleoprotein G




• Molecule 20: Small nuclear ribonucleoprotein G



• Molecule 21: Small nuclear ribonucleoprotein Sm D3




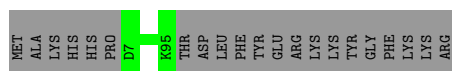
• Molecule 21: Small nuclear ribonucleoprotein Sm D3

Chain 2K:  83% 14%




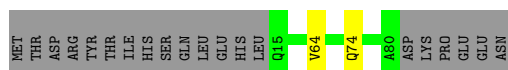
- Molecule 46: PHD finger-like domain-containing protein 5A

Chain 2L:  81% 19%



- Molecule 47: Splicing factor 3B subunit 5

Chain 2M:  74% 23%



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	140426	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	50	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2500	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	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: GTP, MG, IHP, 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	0.66	3/1368 (0.2%)	0.90	1/2122 (0.0%)
2	6A	0.42	1/1398 (0.1%)	0.91	6/2172 (0.3%)
3	6a	0.43	0/355	0.67	0/442
4	6b	0.47	0/294	0.75	0/364
5	6c	0.34	0/294	0.61	0/364
6	6d	0.43	0/286	0.59	0/354
7	6e	0.43	0/279	0.72	0/347
8	6f	0.39	0/258	0.61	0/319
9	6g	0.41	0/242	0.64	0/299
10	5A	0.31	0/2673	0.95	11/4156 (0.3%)
11	5B	0.27	0/18786	0.49	0/25502
12	5C	0.27	0/6879	0.51	1/9344 (0.0%)
13	5D	0.25	0/16393	0.50	0/22174
14	5E	0.28	0/1495	0.57	0/2074
15	2a	0.50	0/339	0.69	0/422
15	4a	0.30	0/404	0.60	0/561
15	5a	0.51	0/339	0.69	0/422
16	2b	0.56	0/327	0.68	0/407
16	4b	0.29	0/400	0.61	0/556
16	5b	0.57	0/327	0.67	0/407
17	2c	0.71	0/338	0.73	0/419
17	4c	0.33	0/454	0.63	0/631
17	5c	0.69	0/387	0.72	0/482
18	2d	0.78	0/295	0.76	0/367
18	4d	0.36	0/350	0.64	0/483
18	5d	0.78	0/291	0.76	0/362
19	2e	0.64	0/315	0.75	0/392
19	4e	0.29	0/375	0.66	0/521
19	5e	0.64	0/315	0.75	0/392
20	2f	0.55	0/266	0.63	0/329
20	4f	0.30	0/357	0.65	0/494
20	5f	0.54	0/295	0.61	0/367

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
21	2g	0.47	0/318	0.56	0/394
21	4g	0.29	0/408	0.58	0/566
21	5g	0.46	0/307	0.55	0/382
22	4A	0.49	2/2960 (0.1%)	0.92	12/4599 (0.3%)
23	4B	0.25	0/949	0.45	0/1314
24	4C	0.24	0/1764	0.48	0/2450
25	4D	0.29	0/1767	0.49	0/2434
26	4E	0.23	0/614	0.43	0/855
27	4F	0.27	0/1198	0.48	0/1620
28	4G	0.25	0/4833	0.48	1/6625 (0.0%)
29	4R	0.28	0/891	0.56	0/1188
30	4S	0.26	0/555	0.53	0/731
31	4T	0.27	0/3845	0.48	0/5208
32	4U	0.24	0/1356	0.53	0/1790
33	4X	0.28	0/187	0.63	0/245
34	4Y	0.32	0/1592	0.53	0/2215
35	2A	0.86	11/2576 (0.4%)	1.43	56/4003 (1.4%)
36	2B	0.63	0/647	1.42	0/807
37	2C	0.61	0/375	1.20	0/467
38	2D	0.24	0/972	0.46	0/1260
39	2E	0.28	0/373	0.52	0/461
40	2F	0.25	0/1688	0.47	0/2102
41	2G	1.04	4/4184 (0.1%)	0.83	2/5216 (0.0%)
42	2H	0.74	0/722	0.72	0/892
43	2I	0.85	0/4664	0.76	0/5816
44	2J	0.62	0/311	0.65	0/387
45	2K	0.79	0/431	0.79	0/537
46	2L	0.74	0/355	0.68	0/442
47	2M	1.01	0/263	0.77	0/327
All	All	0.45	21/98279 (0.0%)	0.66	90/133380 (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
15	4a	0	1
17	2c	0	1
17	5c	0	1
18	4d	0	1
19	4e	0	1

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
22	4A	0	1
24	4C	0	1
40	2F	0	1
41	2G	0	11
42	2H	0	3
43	2I	0	11
45	2K	0	1
47	2M	0	1
All	All	0	35

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
41	2G	407	MET	N-CA	12.40	1.71	1.46
22	4A	87	C	O3'-P	11.54	1.75	1.61
2	6A	59	G	P-O5'	8.69	1.68	1.59
41	2G	406	ALA	C-N	7.96	1.52	1.34
35	2A	142	C	C1'-N1	7.25	1.59	1.48

The worst 5 of 90 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
10	5A	78	U	P-O3'-C3'	-12.65	104.52	119.70
10	5A	57	G	P-O3'-C3'	-12.60	104.58	119.70
35	2A	167	U	C5-C4-O4	11.56	132.84	125.90
41	2G	406	ALA	C-N-CA	10.26	147.35	121.70
10	5A	77	G	P-O3'-C3'	-10.20	107.46	119.70

There are no chirality outliers.

5 of 35 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
22	4A	90	G	Sidechain
24	4C	459	PRO	Peptide
15	4a	53	PRO	Peptide
18	4d	40	MET	Peptide
17	5c	112	ASN	Peptide

5.2 Too-close contacts [i](#)

Due to software issues we are unable to calculate clashes - this section is therefore empty.

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	
3	6a	87/95 (92%)	76 (87%)	7 (8%)	4 (5%)	2	18
4	6b	70/102 (69%)	64 (91%)	3 (4%)	3 (4%)	2	20
5	6c	70/139 (50%)	64 (91%)	5 (7%)	1 (1%)	11	46
6	6d	68/91 (75%)	63 (93%)	4 (6%)	1 (2%)	10	44
7	6e	68/80 (85%)	64 (94%)	2 (3%)	2 (3%)	4	28
8	6f	61/103 (59%)	56 (92%)	5 (8%)	0	100	100
9	6g	57/96 (59%)	52 (91%)	4 (7%)	1 (2%)	8	41
11	5B	2200/2335 (94%)	2105 (96%)	92 (4%)	3 (0%)	51	83
12	5C	850/972 (87%)	819 (96%)	31 (4%)	0	100	100
13	5D	1989/2136 (93%)	1906 (96%)	83 (4%)	0	100	100
14	5E	300/357 (84%)	282 (94%)	18 (6%)	0	100	100
15	2a	83/231 (36%)	81 (98%)	2 (2%)	0	100	100
15	4a	80/231 (35%)	71 (89%)	9 (11%)	0	100	100
15	5a	83/231 (36%)	81 (98%)	2 (2%)	0	100	100
16	2b	80/119 (67%)	77 (96%)	3 (4%)	0	100	100
16	4b	79/119 (66%)	75 (95%)	4 (5%)	0	100	100
16	5b	80/119 (67%)	77 (96%)	3 (4%)	0	100	100
17	2c	81/118 (69%)	78 (96%)	3 (4%)	0	100	100
17	4c	90/118 (76%)	84 (93%)	6 (7%)	0	100	100
17	5c	95/118 (80%)	91 (96%)	4 (4%)	0	100	100
18	2d	72/86 (84%)	68 (94%)	4 (6%)	0	100	100
18	4d	70/86 (81%)	69 (99%)	1 (1%)	0	100	100
18	5d	71/86 (83%)	68 (96%)	3 (4%)	0	100	100
19	2e	77/92 (84%)	76 (99%)	1 (1%)	0	100	100
19	4e	74/92 (80%)	71 (96%)	3 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
19	5e	77/92 (84%)	76 (99%)	1 (1%)	0	100	100
20	2f	63/76 (83%)	61 (97%)	2 (3%)	0	100	100
20	4f	71/76 (93%)	65 (92%)	6 (8%)	0	100	100
20	5f	72/76 (95%)	70 (97%)	2 (3%)	0	100	100
21	2g	76/126 (60%)	74 (97%)	2 (3%)	0	100	100
21	4g	81/126 (64%)	76 (94%)	5 (6%)	0	100	100
21	5g	75/126 (60%)	73 (97%)	2 (3%)	0	100	100
23	4B	183/683 (27%)	177 (97%)	6 (3%)	0	100	100
24	4C	357/522 (68%)	332 (93%)	25 (7%)	0	100	100
25	4D	321/499 (64%)	297 (92%)	19 (6%)	5 (2%)	9	43
26	4E	122/128 (95%)	114 (93%)	8 (7%)	0	100	100
27	4F	139/142 (98%)	134 (96%)	5 (4%)	0	100	100
28	4G	786/941 (84%)	744 (95%)	39 (5%)	3 (0%)	34	69
29	4R	104/480 (22%)	95 (91%)	9 (9%)	0	100	100
30	4S	66/800 (8%)	64 (97%)	2 (3%)	0	100	100
31	4T	454/565 (80%)	419 (92%)	35 (8%)	0	100	100
32	4U	147/820 (18%)	143 (97%)	4 (3%)	0	100	100
33	4X	19/155 (12%)	19 (100%)	0	0	100	100
34	4Y	316/1007 (31%)	294 (93%)	18 (6%)	4 (1%)	12	47
36	2B	160/255 (63%)	146 (91%)	12 (8%)	2 (1%)	12	47
37	2C	92/225 (41%)	90 (98%)	2 (2%)	0	100	100
38	2D	173/793 (22%)	157 (91%)	12 (7%)	4 (2%)	6	34
39	2E	88/464 (19%)	61 (69%)	17 (19%)	10 (11%)	0	2
40	2F	413/501 (82%)	367 (89%)	41 (10%)	5 (1%)	13	49
41	2G	1032/1304 (79%)	845 (82%)	165 (16%)	22 (2%)	7	37
42	2H	170/895 (19%)	151 (89%)	15 (9%)	4 (2%)	6	34
43	2I	1152/1217 (95%)	1053 (91%)	89 (8%)	10 (1%)	17	56
44	2J	76/424 (18%)	75 (99%)	1 (1%)	0	100	100
45	2K	106/125 (85%)	85 (80%)	18 (17%)	3 (3%)	5	29
46	2L	87/110 (79%)	74 (85%)	13 (15%)	0	100	100
47	2M	64/86 (74%)	55 (86%)	8 (12%)	1 (2%)	9	43

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	14077/22191 (63%)	13104 (93%)	885 (6%)	88 (1%)	29	64

5 of 88 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	6a	55	LEU
4	6b	84	MET
6	6d	70	ASP
7	6e	52	VAL
7	6e	55	GLN

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	
11	5B	1980/2108 (94%)	1944 (98%)	36 (2%)	59	82
12	5C	754/866 (87%)	744 (99%)	10 (1%)	69	87
13	5D	1779/1908 (93%)	1744 (98%)	35 (2%)	55	80
25	4D	35/424 (8%)	28 (80%)	7 (20%)	1	6
27	4F	129/130 (99%)	127 (98%)	2 (2%)	62	84
28	4G	246/792 (31%)	237 (96%)	9 (4%)	34	68
29	4R	94/369 (26%)	90 (96%)	4 (4%)	29	64
30	4S	54/681 (8%)	51 (94%)	3 (6%)	21	57
31	4T	418/511 (82%)	414 (99%)	4 (1%)	76	90
32	4U	139/721 (19%)	138 (99%)	1 (1%)	84	94
33	4X	19/144 (13%)	19 (100%)	0	100	100
38	2D	53/709 (8%)	49 (92%)	4 (8%)	13	45
All	All	5700/9363 (61%)	5585 (98%)	115 (2%)	57	80

5 of 115 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
13	5D	442	TYR
32	4U	317	SER
13	5D	1451	PHE
31	4T	377	ASP
29	4R	384	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
30	4S	734	HIS
31	4T	233	ASN
31	4T	362	HIS
11	5B	1823	HIS
13	5D	785	HIS

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	57/144 (39%)	33 (57%)	10 (17%)
10	5A	113/117 (96%)	36 (31%)	5 (4%)
2	6A	56/107 (52%)	13 (23%)	5 (8%)
22	4A	119/146 (81%)	24 (20%)	3 (2%)
35	2A	105/188 (55%)	22 (20%)	3 (2%)
All	All	450/702 (64%)	128 (28%)	26 (5%)

5 of 128 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	8	U
1	A	9	U
1	A	10	C
1	A	11	C
1	A	12	U

5 of 26 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	6A	77	C
10	5A	80	U
35	2A	164	C
10	5A	79	C

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
10	5A	94	U

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, 2 are monoatomic - leaving 2 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
50	GTP	5C	1002	49	26,34,34	1.18	2 (7%)	32,54,54	1.71	7 (21%)
48	IHP	5B	2401	-	36,36,36	0.74	0	54,60,60	1.15	3 (5%)

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
50	GTP	5C	1002	49	-	2/18/38/38	0/3/3/3
48	IHP	5B	2401	-	-	7/30/54/54	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
50	5C	1002	GTP	C5-C6	-4.20	1.38	1.47

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
50	5C	1002	GTP	C2-N3	2.07	1.38	1.33

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
50	5C	1002	GTP	PB-O3B-PG	-4.45	117.55	132.83
50	5C	1002	GTP	PA-O3A-PB	-3.72	120.06	132.83
50	5C	1002	GTP	C5-C6-N1	3.34	119.86	113.95
48	5B	2401	IHP	C5-C4-C3	3.24	117.50	110.41
50	5C	1002	GTP	C8-N7-C5	3.10	108.89	102.99

There are no chirality outliers.

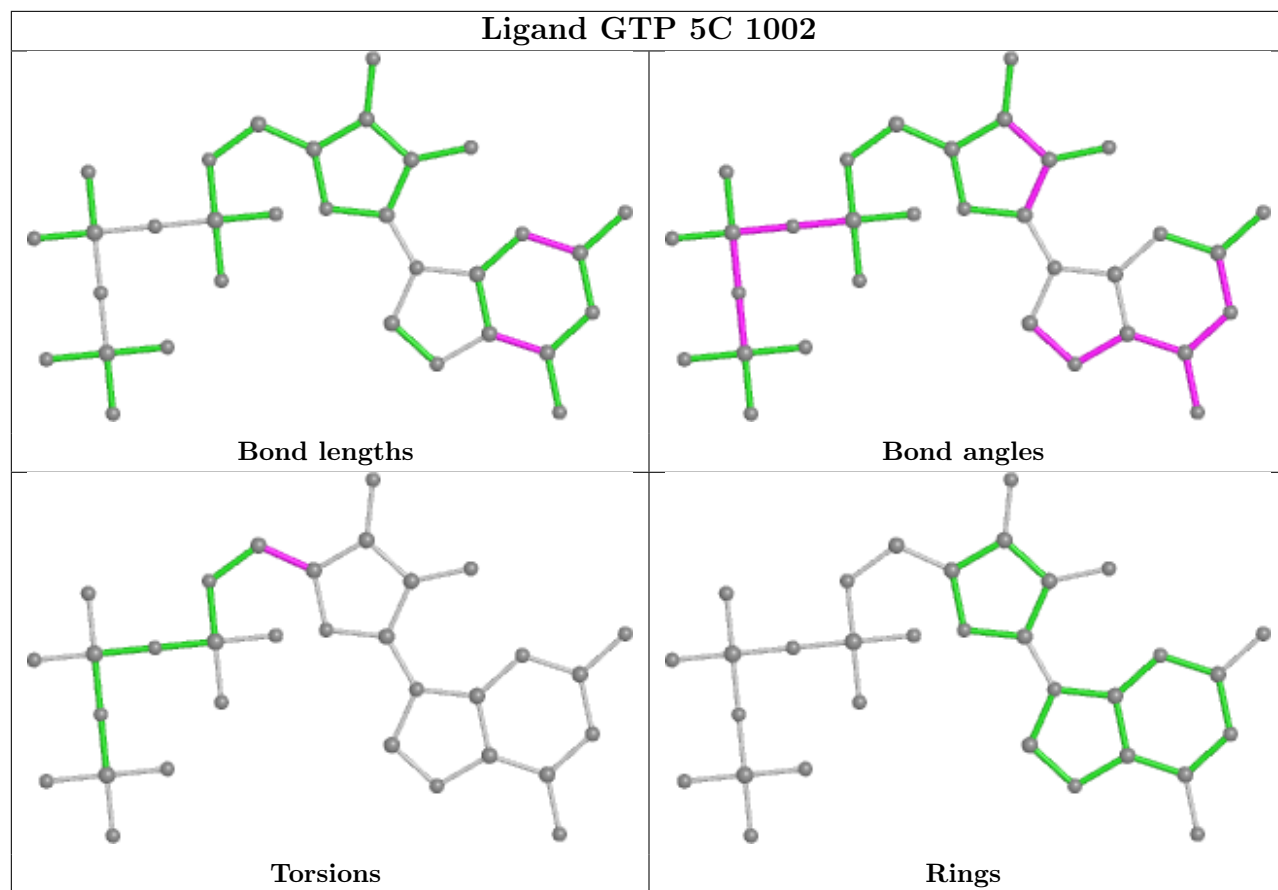
5 of 9 torsion outliers are listed below:

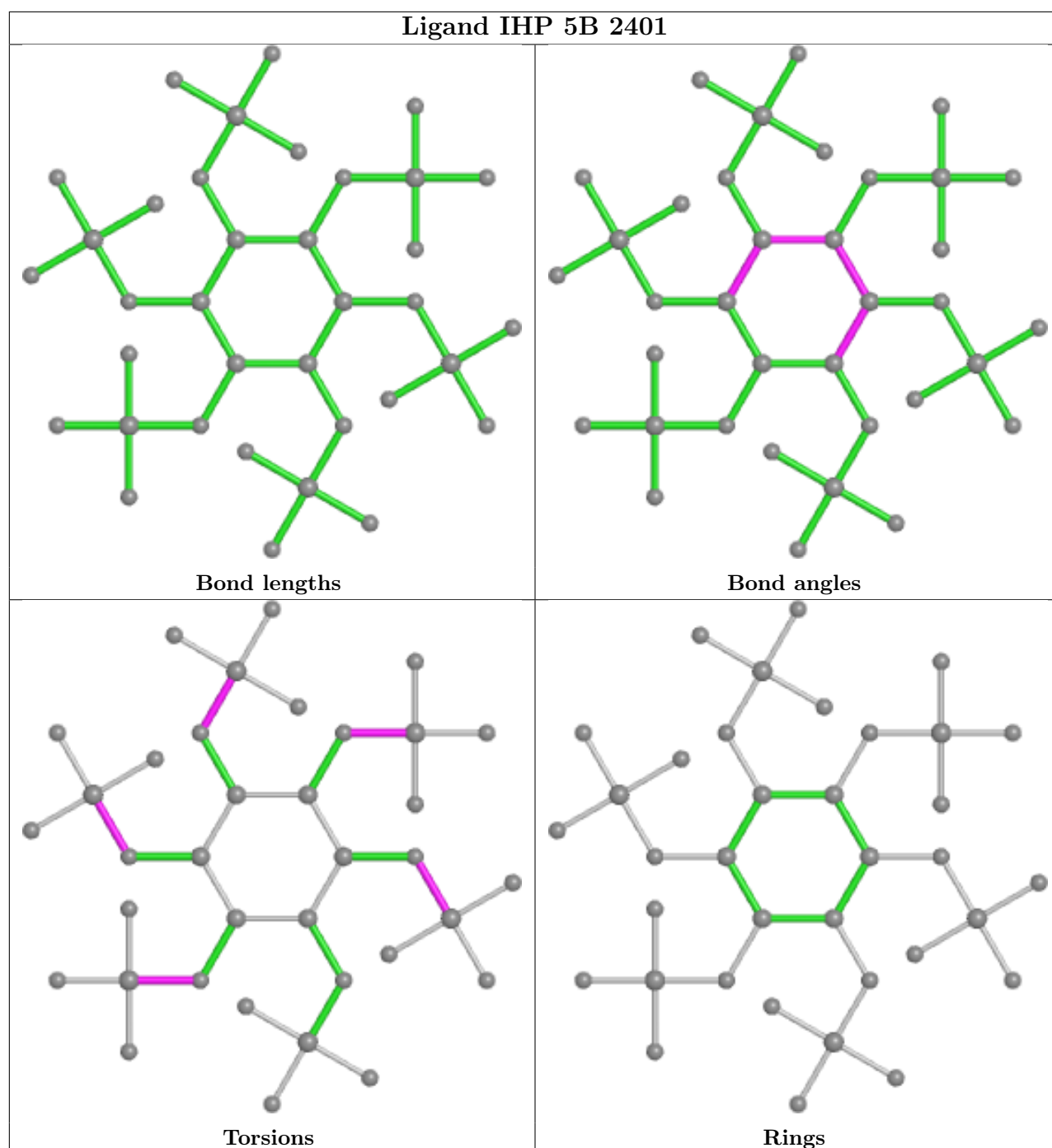
Mol	Chain	Res	Type	Atoms
48	5B	2401	IHP	C2-O12-P2-O32
48	5B	2401	IHP	C3-O13-P3-O43
50	5C	1002	GTP	O4'-C4'-C5'-O5'
50	5C	1002	GTP	C3'-C4'-C5'-O5'
48	5B	2401	IHP	C4-O14-P4-O44

There are no ring outliers.

No monomer is involved in short contacts.

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 than 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.





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.