wwPDB EM Validation Summary Report (i)

Jun 3, 2024 – 09:17 PM JST

PDB ID	:	8H6L
EMDB ID	:	EMD-34508
Title	:	Cryo-EM structure of human exon-defined spliceosome in the early B state.
Authors	:	Zhang, W.; Zhan, X.; Zhang, X.; Bai, R.; Lei, J.; Yan, C.; Shi, Y.
Deposited on	:	2022-10-18
Resolution	:	2.60 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 (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:

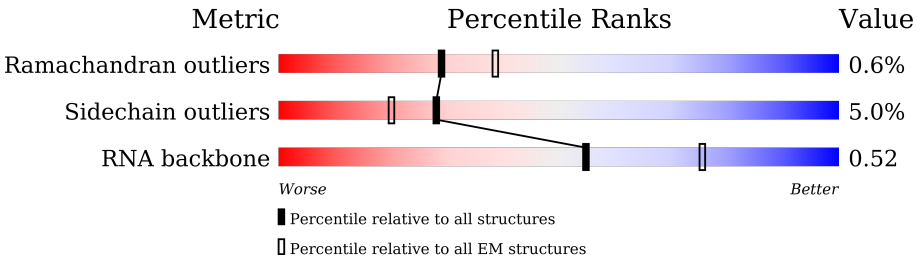
EMDB validation analysis	:	FAILED
Mogul	:	1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

1 Overall quality at a glance

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

The reported resolution of this entry is 2.60 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 <=5%

Mol	Chain	Length	Quality of chain
1	A	144	12% 26% 60%
2	5A	117	72% 26%
3	5B	2335	92%
4	5C	972	80% 5% 16%
5	5D	2136	78% 21%
6	5E	357	81% 16%
7	2a	231	37% 63%
7	4a	231	28% 72%
7	5a	231	36% 64%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
8	2b	119	69% 31%
8	4b	119	69% 31%
8	5b	119	69% 31%
9	2c	118	71% 28%
9	4c	118	63% 37%
9	5c	118	81% 18%
10	2d	86	86% 14%
10	4d	86	83% 17%
10	5d	86	86% 14%
11	2e	92	86% 14%
11	4e	92	85% 15%
11	5e	92	86% 14%
12	2f	76	89% 11%
12	4f	76	96%
12	5f	76	95% 5%
13	2g	126	63% 37%
13	4g	126	56% 44%
13	5g	126	60% 40%
14	6A	107	42% 13% 45%
15	6a	95	91% 5%
16	6b	102	70% 27%
17	6c	139	53% 47%
18	6d	91	78% 21%
19	6e	80	85% 12%
20	6f	103	63% 37%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
21	6g	96	63% 36%
22	4A	145	63% 26% 11%
23	4B	683	35% 63%
24	4C	522	75% 6% 18%
25	4D	499	71% 25%
26	4E	128	94%
27	4F	142	94% 6%
28	4G	941	82% 15%
29	4H	177	95% 5%
30	4I	376	19% 80%
31	4J	800	18% 81%
32	4Z	513	82% 18%
33	2A	188	33% 21% 42%
34	2B	255	63% 36%
35	2C	225	42% 58%
36	2D	793	28% 70%
37	2E	464	18% 80%
38	2F	501	83% 16%
39	2G	1304	78% 20%
40	2H	895	23% 76%
41	2I	1217	94%
42	2J	424	18% 82%
43	2K	125	83% 14%
44	2L	110	81% 19%
45	2M	86	73% 23%

2 Entry composition (i)

There are 49 unique types of molecules in this entry. The entry contains 94667 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	57	1187	531	183	416	57	0	0

- Molecule 2 is a RNA chain called U5 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	5A	115	2420	1084	403	818	115	0	0

- Molecule 3 is a protein called Pre-mRNA-processing-splicing factor 8.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	5B	2253	18642	11992	3250	3319	81	0	0

- Molecule 4 is a protein called 116 kDa U5 small nuclear ribonucleoprotein component.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	5C	818	6436	4114	1085	1205	32	0	0

- Molecule 5 is a protein called U5 small nuclear ribonucleoprotein 200 kDa helicase.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	5D	1696	13633	8715	2329	2519	70	0	0

- Molecule 6 is a protein called U5 small nuclear ribonucleoprotein 40 kDa protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
6	5E	299	1196	598	299	299	0	0

- Molecule 7 is a protein called Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B'.

Mol	Chain	Residues	Atoms				AltConf	Trace
7	5a	84	Total 336	C 168	N 84	O 84	0	0
7	4a	64	Total 256	C 128	N 64	O 64	0	0
7	2a	86	Total 344	C 172	N 86	O 86	0	0

- Molecule 8 is a protein called Small nuclear ribonucleoprotein Sm D1.

Mol	Chain	Residues	Atoms				AltConf	Trace
8	5b	82	Total 328	C 164	N 82	O 82	0	0
8	4b	82	Total 334	C 170	N 82	O 82	0	0
8	2b	82	Total 328	C 164	N 82	O 82	0	0

- Molecule 9 is a protein called Small nuclear ribonucleoprotein Sm D2.

Mol	Chain	Residues	Atoms				AltConf	Trace
9	5c	97	Total 388	C 194	N 97	O 97	0	0
9	4c	74	Total 300	C 152	N 74	O 74	0	0
9	2c	85	Total 340	C 170	N 85	O 85	0	0

- Molecule 10 is a protein called Small nuclear ribonucleoprotein F.

Mol	Chain	Residues	Atoms				AltConf	Trace
10	5d	74	Total 296	C 148	N 74	O 74	0	0
10	4d	71	Total 292	C 150	N 71	O 71	0	0
10	2d	74	Total 296	C 148	N 74	O 74	0	0

- Molecule 11 is a protein called Small nuclear ribonucleoprotein E.

Mol	Chain	Residues	Atoms				AltConf	Trace
11	5e	79	Total	C	N	O	0	0
			316	158	79	79		
11	4e	78	Total	C	N	O	0	0
			314	158	78	78		
11	2e	79	Total	C	N	O	0	0
			316	158	79	79		

- Molecule 12 is a protein called Small nuclear ribonucleoprotein G.

Mol	Chain	Residues	Atoms				AltConf	Trace
12	5f	72	Total	C	N	O	0	0
			288	144	72	72		
12	4f	73	Total	C	N	O	0	0
			298	152	73	73		
12	2f	68	Total	C	N	O	0	0
			272	136	68	68		

- Molecule 13 is a protein called Small nuclear ribonucleoprotein Sm D3.

Mol	Chain	Residues	Atoms				AltConf	Trace
13	5g	76	Total	C	N	O	0	0
			304	152	76	76		
13	4g	71	Total	C	N	O	0	0
			288	146	71	71		
13	2g	80	Total	C	N	O	0	0
			320	160	80	80		

- Molecule 14 is a RNA chain called U6 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	6A	59	Total	C	N	O	P	0	0
			1251	558	230	404	59		

- Molecule 15 is a protein called U6 snRNA-associated Sm-like protein LSm2.

Mol	Chain	Residues	Atoms				AltConf	Trace
15	6a	90	Total	C	N	O	0	0
			360	180	90	90		

- Molecule 16 is a protein called U6 snRNA-associated Sm-like protein LSm3.

Mol	Chain	Residues	Atoms				AltConf	Trace
16	6b	74	Total 296	C 148	N 74	O 74	0	0

- Molecule 17 is a protein called U6 snRNA-associated Sm-like protein LSm4.

Mol	Chain	Residues	Atoms				AltConf	Trace
17	6c	74	Total 296	C 148	N 74	O 74	0	0

- Molecule 18 is a protein called U6 snRNA-associated Sm-like protein LSm5.

Mol	Chain	Residues	Atoms				AltConf	Trace
18	6d	72	Total 288	C 144	N 72	O 72	0	0

- Molecule 19 is a protein called U6 snRNA-associated Sm-like protein LSm6.

Mol	Chain	Residues	Atoms				AltConf	Trace
19	6e	70	Total 280	C 140	N 70	O 70	0	0

- Molecule 20 is a protein called U6 snRNA-associated Sm-like protein LSm7.

Mol	Chain	Residues	Atoms				AltConf	Trace
20	6f	65	Total 260	C 130	N 65	O 65	0	0

- Molecule 21 is a protein called U6 snRNA-associated Sm-like protein LSm8.

Mol	Chain	Residues	Atoms				AltConf	Trace
21	6g	61	Total 244	C 122	N 61	O 61	0	0

- Molecule 22 is a RNA chain called U4 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
22	4A	129	Total 2744	C 1225	N 472	O 917	P 130	0	0

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

Mol	Chain	Residues	Atoms					AltConf	Trace
23	4B	256	Total	C	N	O	S	0	0
			2076	1316	385	367	8		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
24	4C	426	Total	C	N	O	S	0	0
			3370	2118	612	620	20		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
25	4D	376	Total	C	N	O	S	0	0
			2874	1788	524	550	12		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
26	4E	124	Total	C	N	O	S	0	0
			962	608	171	178	5		

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

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

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

Mol	Chain	Residues	Atoms					AltConf	Trace
28	4G	801	Total	C	N	O	S	0	0
			5504	3419	1043	1026	16		

- Molecule 29 is a protein called Peptidyl-prolyl cis-trans isomerase H.

Mol	Chain	Residues	Atoms				AltConf	Trace
29	4H	169	Total	C	N	O	0	0
			844	506	169	169		

- Molecule 30 is a protein called WW domain-binding protein 4.

Mol	Chain	Residues	Atoms					AltConf	Trace
30	4I	75	Total	C	N	O	S	0	0
			494	304	96	91	3		

- Molecule 31 is a protein called U4/U6.U5 tri-snRNP-associated protein 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
31	4J	153	Total	C	N	O	S	0	0
			1153	715	206	230	2		

- Molecule 32 is a protein called WD40 repeat-containing protein SMU1.

Mol	Chain	Residues	Atoms				AltConf	Trace
32	4Z	420	Total	C	N	O	0	0
			2093	1253	420	420		

- Molecule 33 is a RNA chain called U2 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
33	2A	109	Total	C	N	O	P	0	0
			2311	1032	396	774	109		

- Molecule 34 is a protein called U2 small nuclear ribonucleoprotein A'.

Mol	Chain	Residues	Atoms				AltConf	Trace
34	2B	162	Total	C	N	O	0	0
			648	324	162	162		

- Molecule 35 is a protein called U2 small nuclear ribonucleoprotein B".

Mol	Chain	Residues	Atoms				AltConf	Trace
35	2C	94	Total	C	N	O	0	0
			376	188	94	94		

- Molecule 36 is a protein called Splicing factor 3A subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
36	2D	236	Total	C	N	O	S	0	0
			1380	793	285	299	3		

- Molecule 37 is a protein called Splicing factor 3A subunit 2.

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

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

Mol	Chain	Residues	Atoms				AltConf	Trace
38	2F	423	Total 1693	C 847	N 423	O 423	0	0

- Molecule 39 is a protein called Splicing factor 3B subunit 1.

Mol	Chain	Residues	Atoms				AltConf	Trace
39	2G	1048	Total 4192	C 2096	N 1048	O 1048	0	0

- Molecule 40 is a protein called Splicing factor 3B subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
40	2H	213	Total 959	C 510	N 220	O 226	S 3	0	0

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

Mol	Chain	Residues	Atoms				AltConf	Trace
41	2I	1168	Total 4672	C 2336	N 1168	O 1168	0	0

- Molecule 42 is a protein called Splicing factor 3B subunit 4.

Mol	Chain	Residues	Atoms				AltConf	Trace
42	2J	78	Total 312	C 156	N 78	O 78	0	0

- Molecule 43 is a protein called Splicing factor 3B subunit 6.

Mol	Chain	Residues	Atoms				AltConf	Trace
43	2K	108	Total 432	C 216	N 108	O 108	0	0

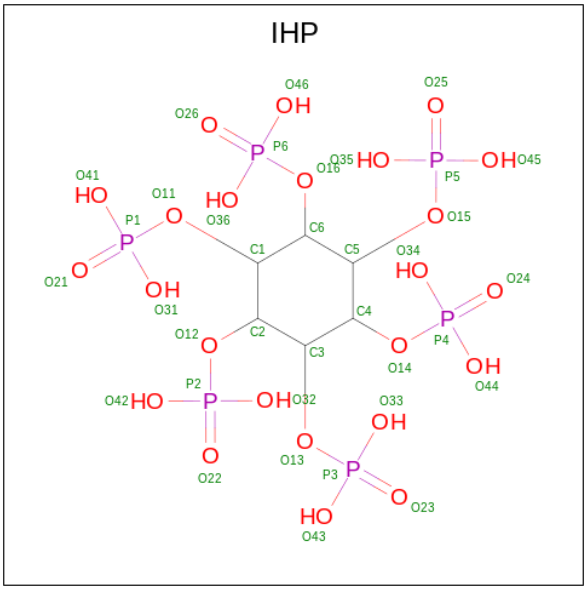
- Molecule 44 is a protein called PHD finger-like domain-containing protein 5A.

Mol	Chain	Residues	Atoms				AltConf	Trace
44	2L	89	Total	C	N	O	0	0
			356	178	89	89		

- Molecule 45 is a protein called Splicing factor 3B subunit 5.

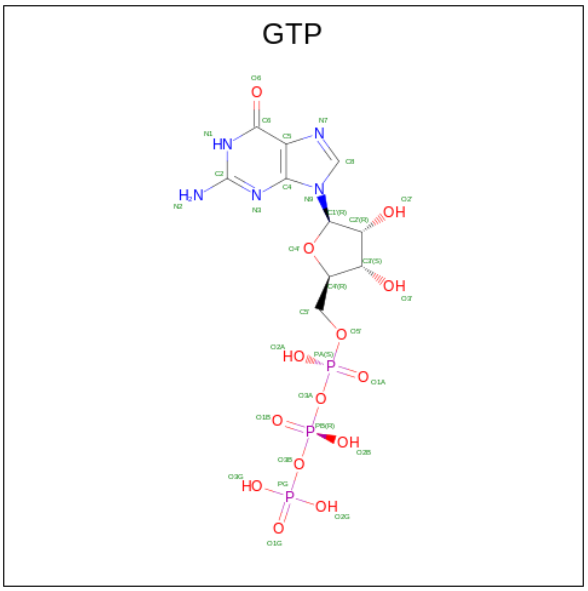
Mol	Chain	Residues	Atoms				AltConf	Trace
45	2M	66	Total	C	N	O	0	0
			264	132	66	66		

- Molecule 46 is INOSITOL HEXAKISPHOSPHATE (three-letter code: IHP) (formula: $C_6H_{18}O_{24}P_6$).



Mol	Chain	Residues	Atoms				AltConf
46	5B	1	Total	C	O	P	0
			36	6	24	6	

- Molecule 47 is GUANOSINE-5'-TRIPHOSPHATE (three-letter code: GTP) (formula: $C_{10}H_{16}N_5O_{14}P_3$).



Mol	Chain	Residues	Atoms					AltConf
47	5C	1	Total	C	N	O	P	0
			32	10	5	14	3	

- Molecule 48 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
48	5C	1	Total	Mg	0
			1	1	

- Molecule 49 is ZINC ION (three-letter code: ZN) (formula: Zn).

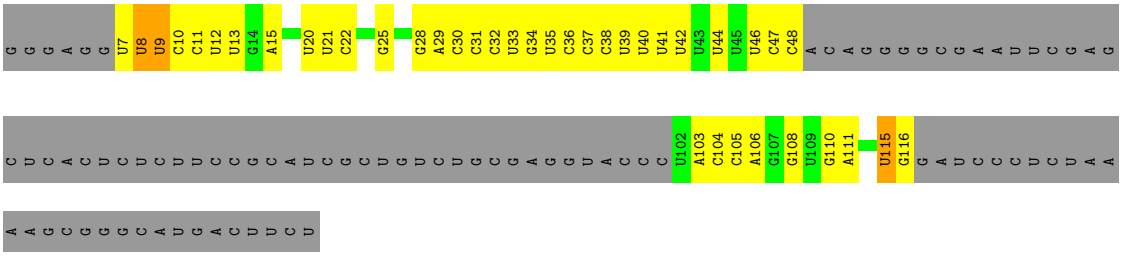
Mol	Chain	Residues	Atoms		AltConf
49	4I	1	Total	Zn	0
			1	1	

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: pre-mRNA

Chain A:

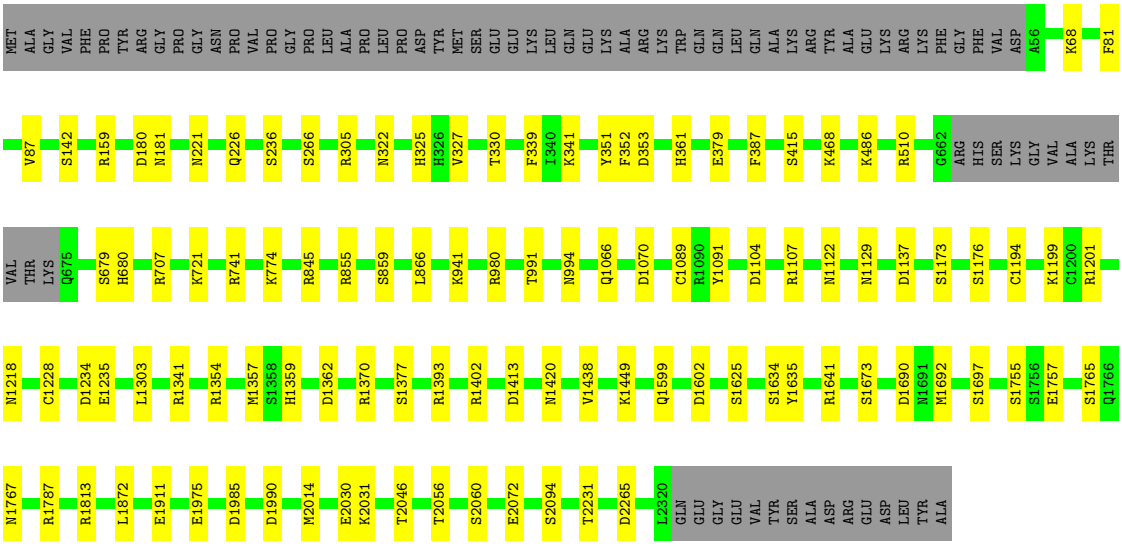


- Molecule 2: U5 snRNA

Chain 5A:

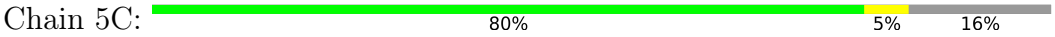
- Molecule 3: Pre-mRNA-processing-splicing factor 8

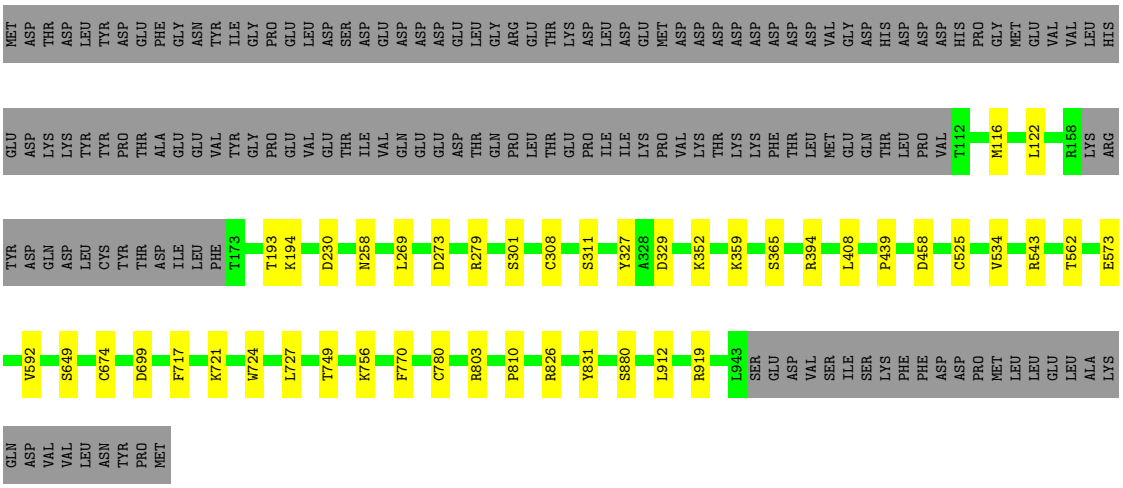
Chain 5B:



• Molecule 4: 116 kDa U5 small nuclear ribonucleoprotein component

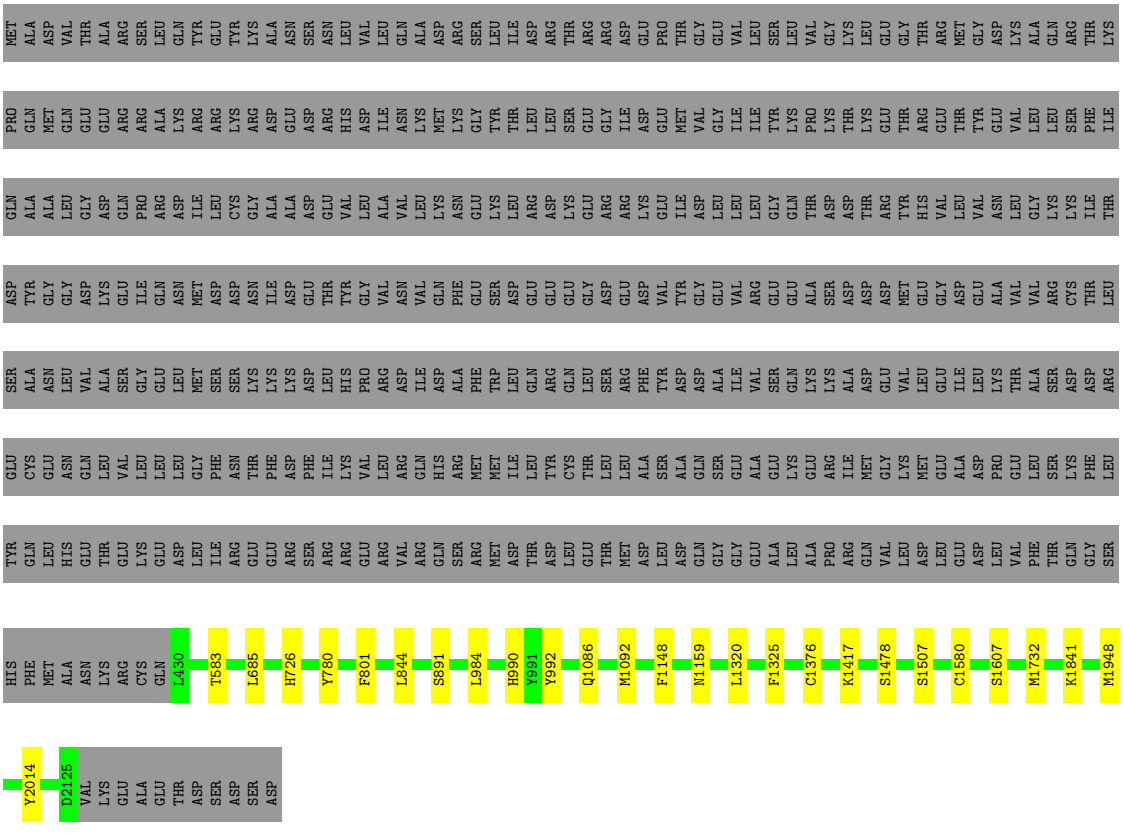
Chain 5C:





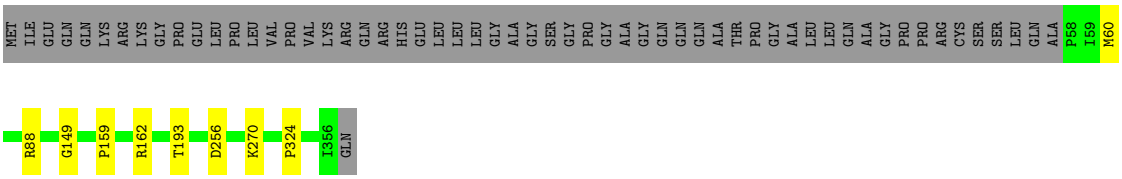
• Molecule 5: U5 small nuclear ribonucleoprotein 200 kDa helicase

Chain 5D:



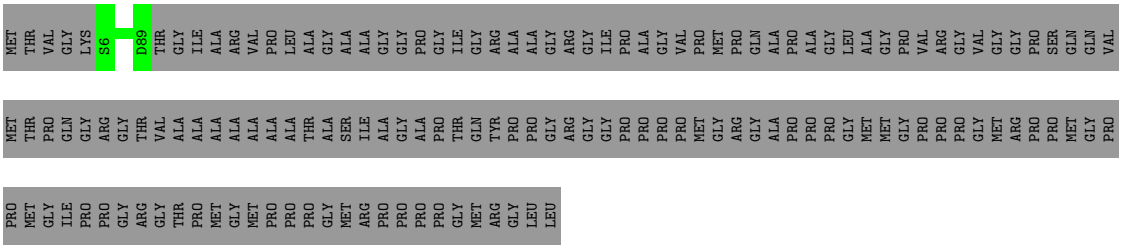
• Molecule 6: U5 small nuclear ribonucleoprotein 40 kDa protein

Chain 5E:



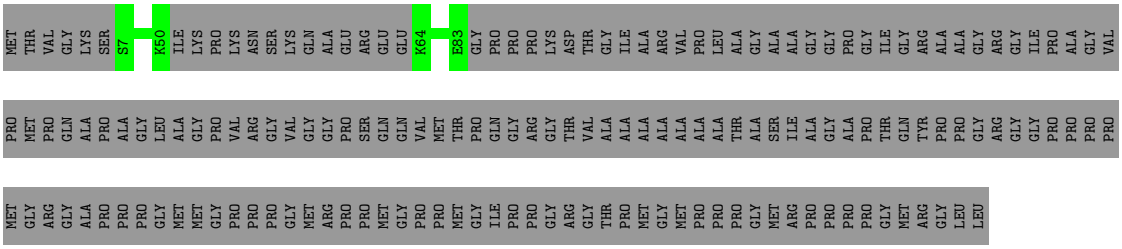
• Molecule 7: Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B'

Chain 5a:



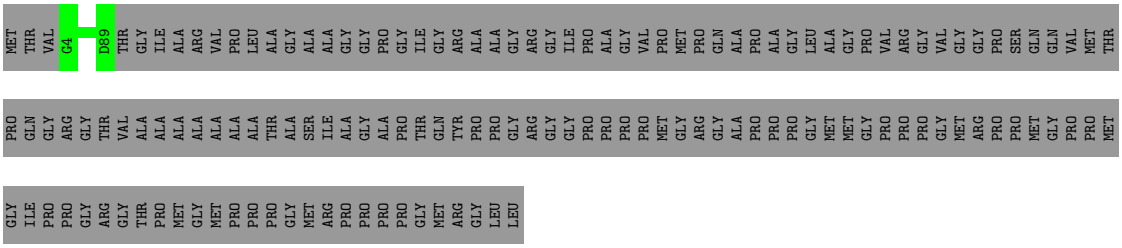
• Molecule 7: Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B'

Chain 4a:



• Molecule 7: Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B'

Chain 2a:

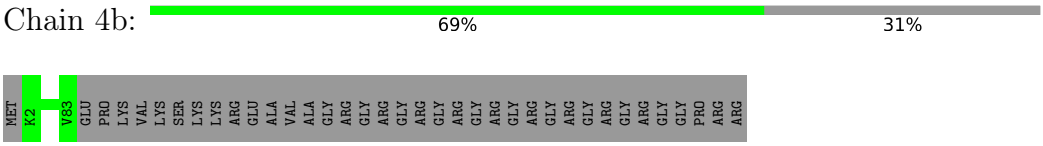


• Molecule 8: Small nuclear ribonucleoprotein Sm D1

Chain 5b:

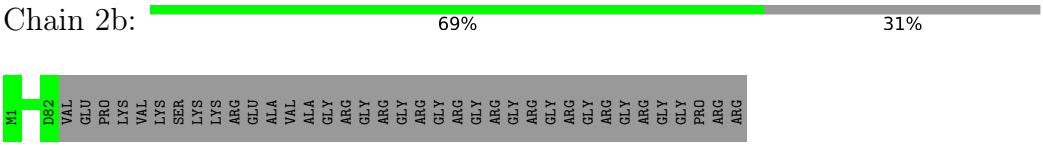
• Molecule 8: Small nuclear ribonucleoprotein Sm D1

Chain 4b:



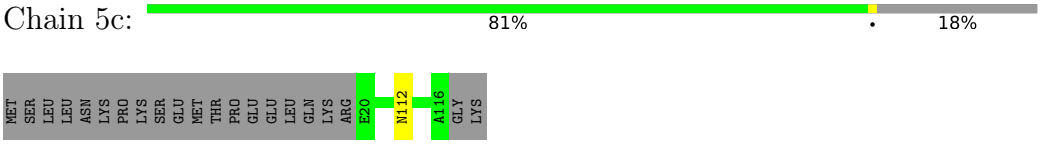
• Molecule 8: Small nuclear ribonucleoprotein Sm D1

Chain 2b:



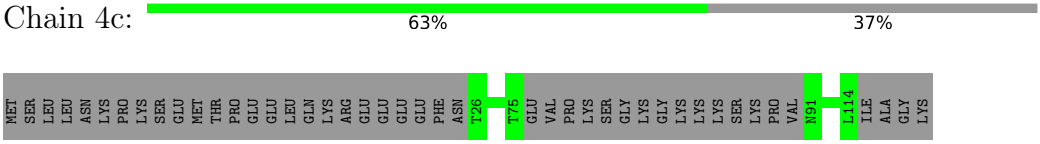
• Molecule 9: Small nuclear ribonucleoprotein Sm D2

Chain 5c:



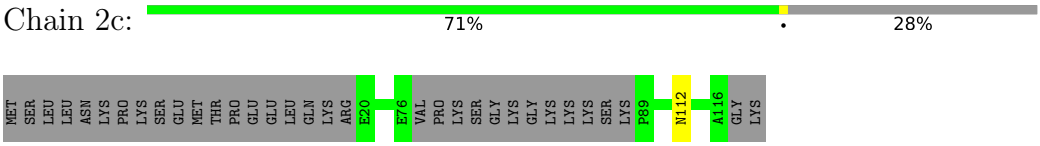
• Molecule 9: Small nuclear ribonucleoprotein Sm D2

Chain 4c:



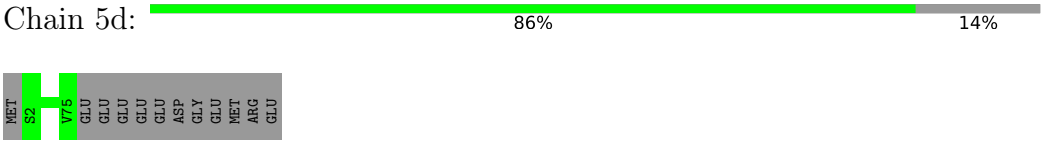
• Molecule 9: Small nuclear ribonucleoprotein Sm D2

Chain 2c:



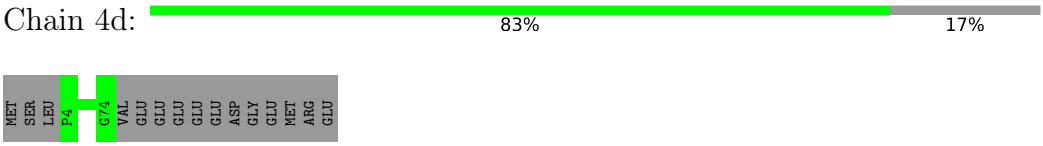
• Molecule 10: Small nuclear ribonucleoprotein F

Chain 5d:



• Molecule 10: Small nuclear ribonucleoprotein F

Chain 4d:

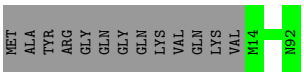


• Molecule 10: Small nuclear ribonucleoprotein F

Chain 2d:

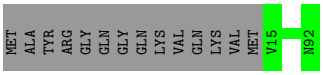
• Molecule 11: Small nuclear ribonucleoprotein E

Chain 5e:



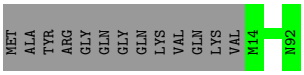
• Molecule 11: Small nuclear ribonucleoprotein E

Chain 4e:



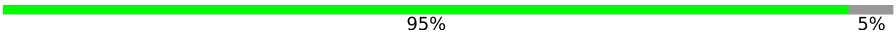
• Molecule 11: Small nuclear ribonucleoprotein E

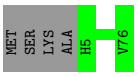
Chain 2e:



• Molecule 12: Small nuclear ribonucleoprotein G

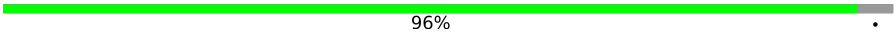
Chain 5f:

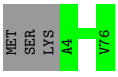




• Molecule 12: Small nuclear ribonucleoprotein G

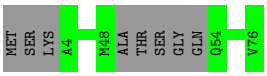
Chain 4f:





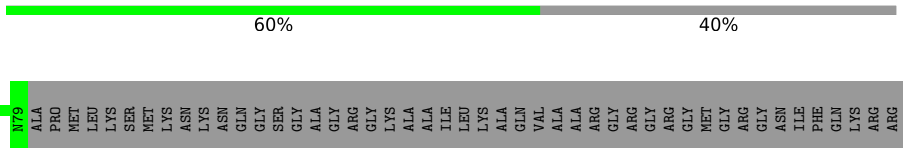
• Molecule 12: Small nuclear ribonucleoprotein G

Chain 2f:



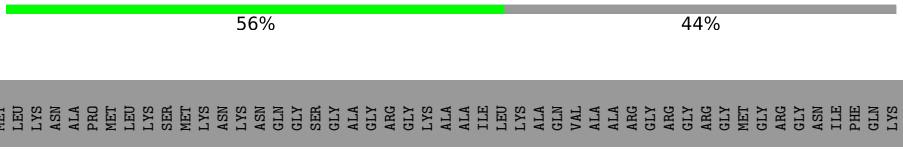
• Molecule 13: Small nuclear ribonucleoprotein Sm D3

Chain 5g:



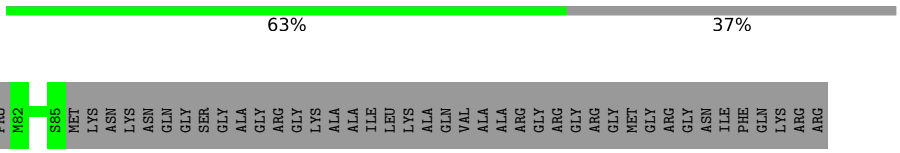
• Molecule 13: Small nuclear ribonucleoprotein Sm D3

Chain 4g:



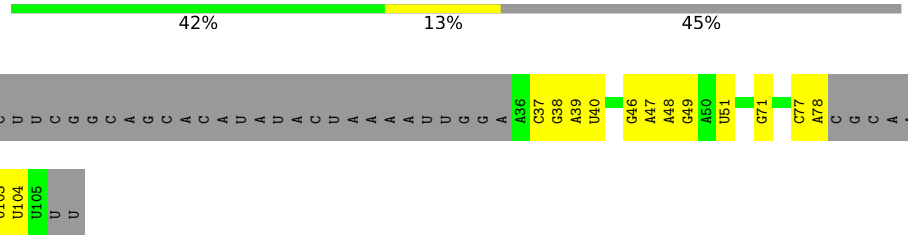
• Molecule 13: Small nuclear ribonucleoprotein Sm D3

Chain 2g:



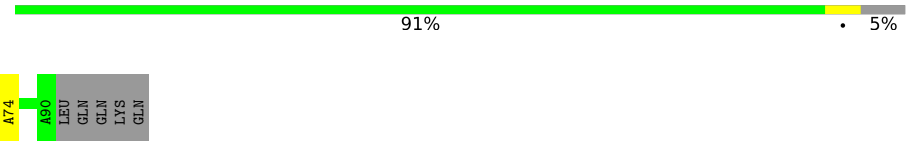
• Molecule 14: U6 snRNA

Chain 6A:



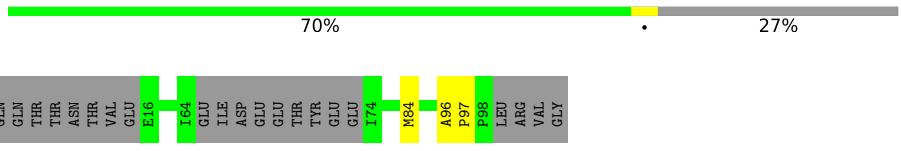
• Molecule 15: U6 snRNA-associated Sm-like protein LSm2

Chain 6a:



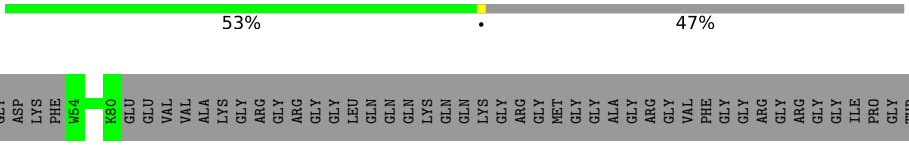
• Molecule 16: U6 snRNA-associated Sm-like protein LSm3

Chain 6b:



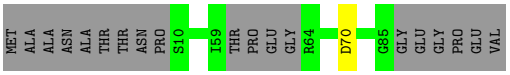
• Molecule 17: U6 snRNA-associated Sm-like protein LSm4

Chain 6c:



• Molecule 18: U6 snRNA-associated Sm-like protein LSm5

Chain 6d:



• Molecule 19: U6 snRNA-associated Sm-like protein LSm6

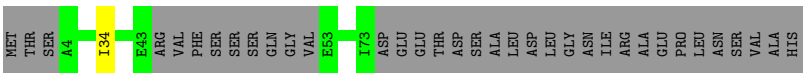
Chain 6e:

• Molecule 20: U6 snRNA-associated Sm-like protein LSm7

Chain 6f:

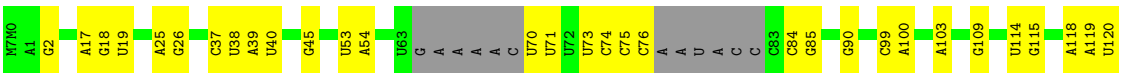
• Molecule 21: U6 snRNA-associated Sm-like protein LSm8

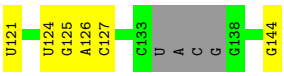
Chain 6g:



• Molecule 22: U4 snRNA

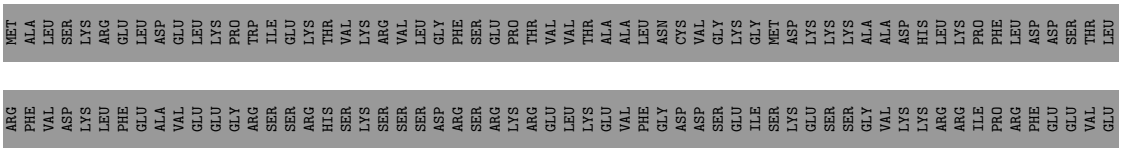
Chain 4A:

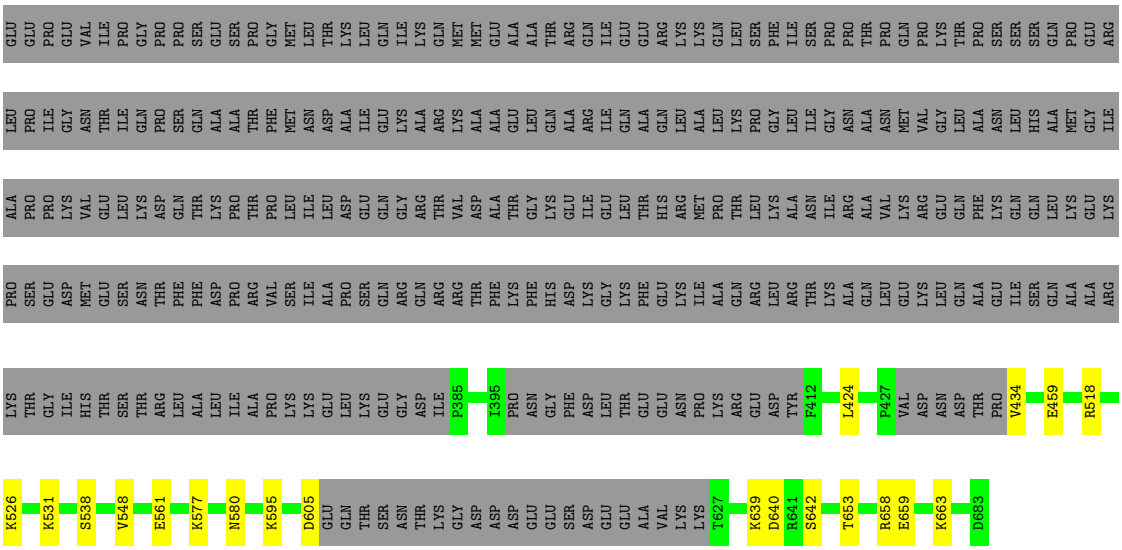




• Molecule 23: U4/U6 small nuclear ribonucleoprotein Prp3

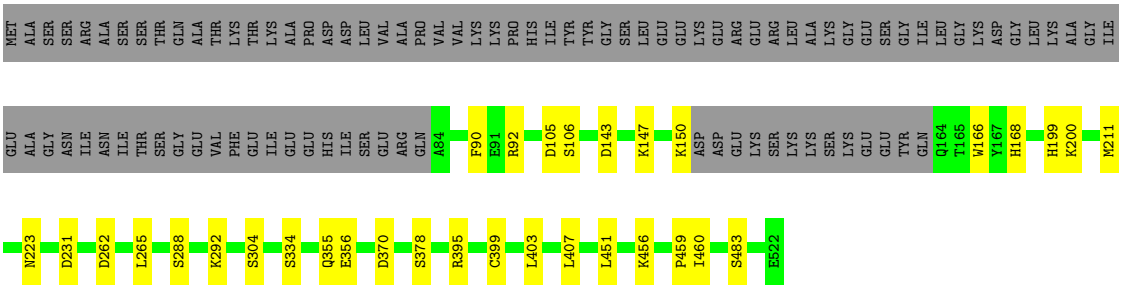
Chain 4B:





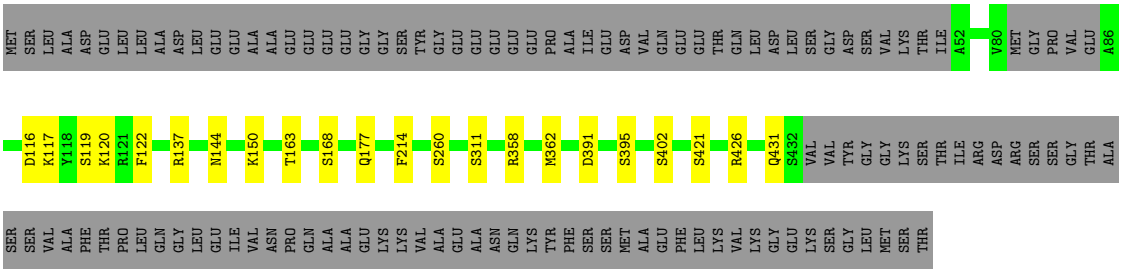
• Molecule 24: U4/U6 small nuclear ribonucleoprotein Prp4

Chain 4C:



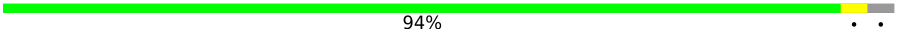
• Molecule 25: U4/U6 small nuclear ribonucleoprotein Prp31

Chain 4D:



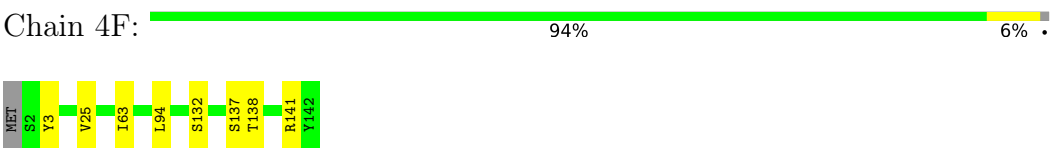
• Molecule 26: NHP2-like protein 1

Chain 4E:



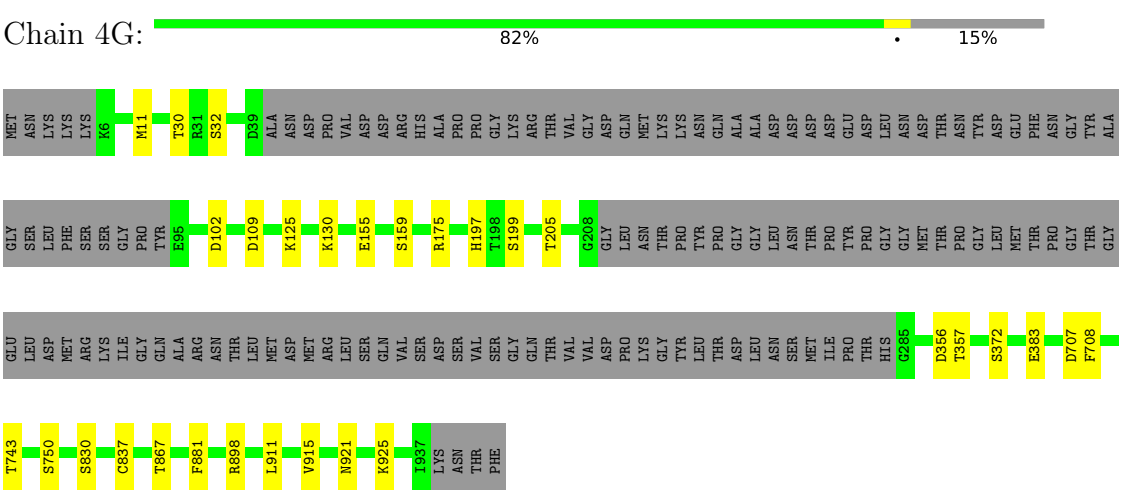
• Molecule 27: Thioredoxin-like protein 4A

Chain 4F:



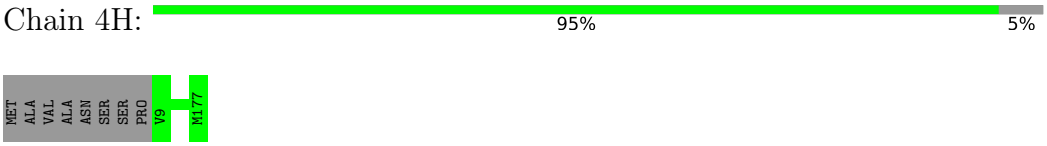
• Molecule 28: Pre-mRNA-processing factor 6

Chain 4G:



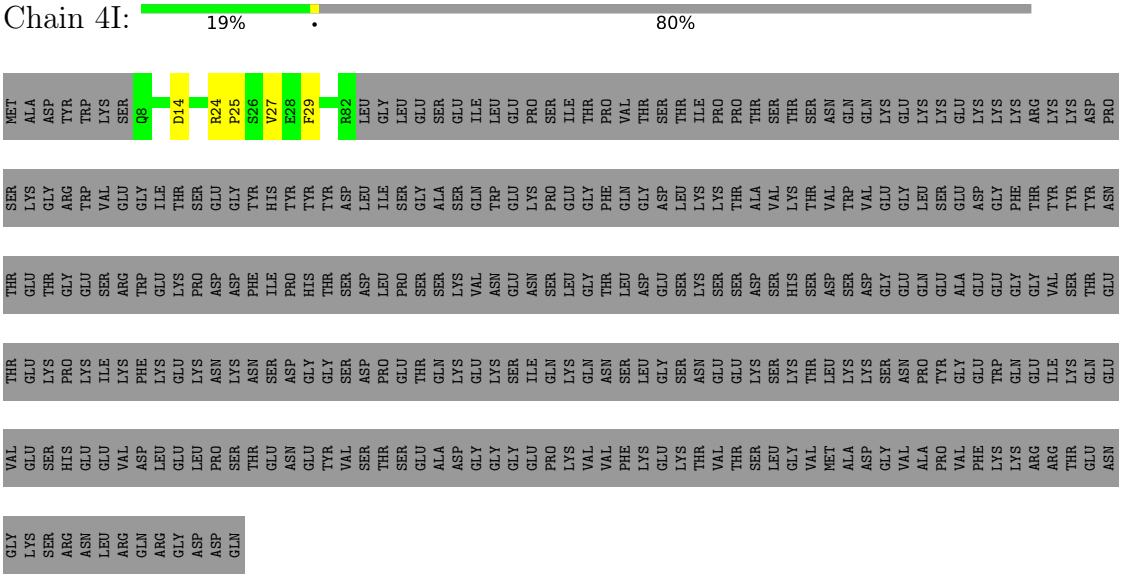
• Molecule 29: Peptidyl-prolyl cis-trans isomerase H

Chain 4H:



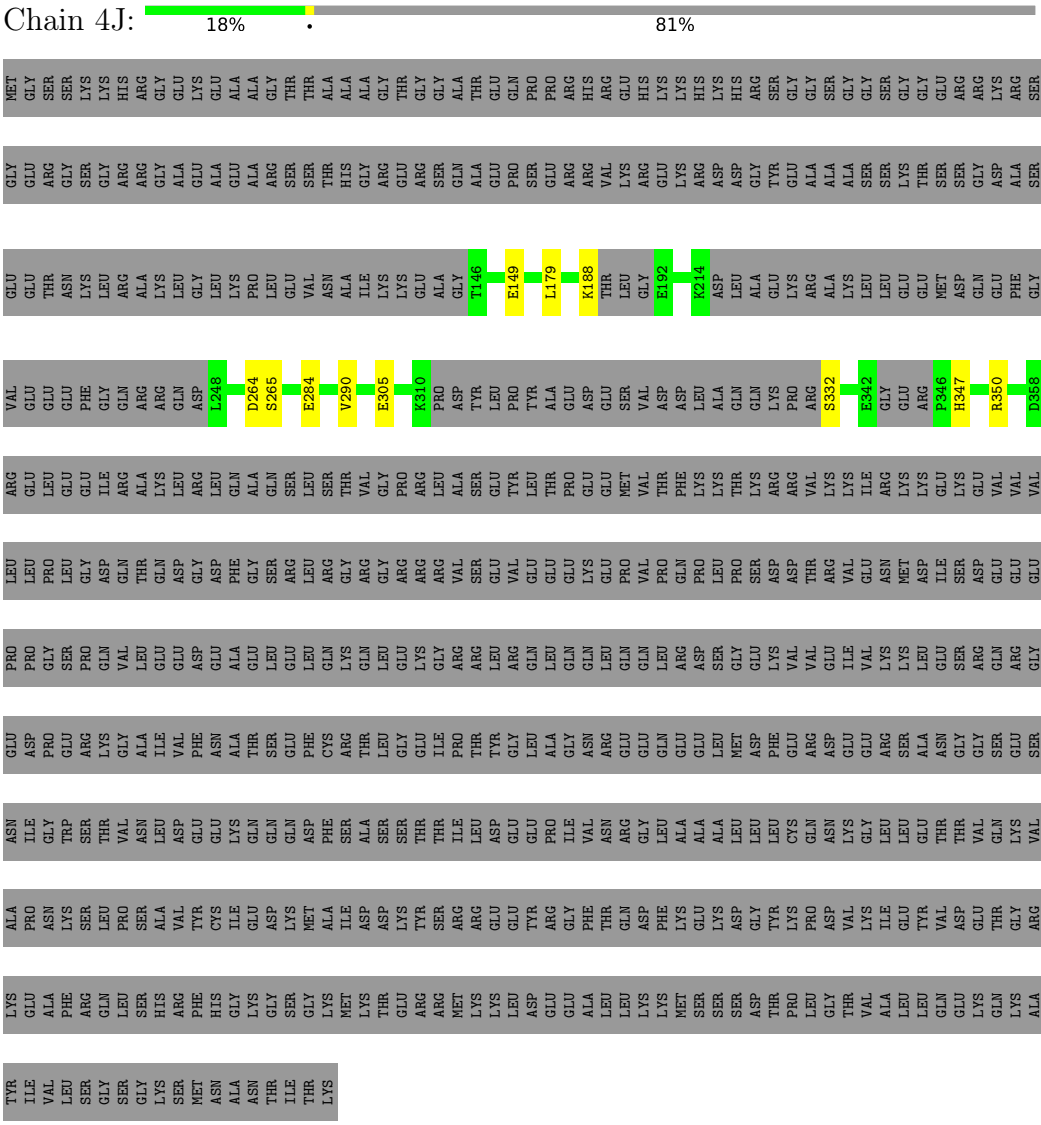
• Molecule 30: WW domain-binding protein 4

Chain 4I:



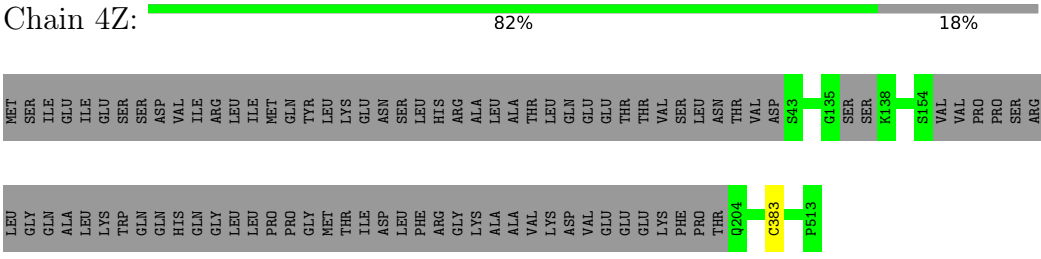
• Molecule 31: U4/U6.U5 tri-snRNP-associated protein 1

Chain 4J:



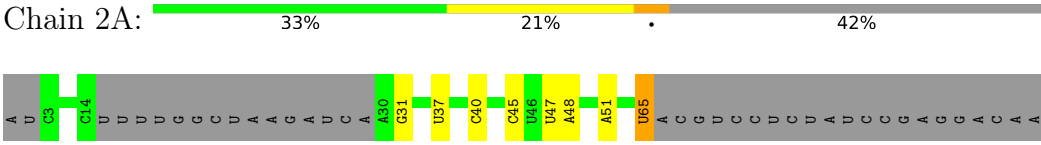
• Molecule 32: WD40 repeat-containing protein SMU1

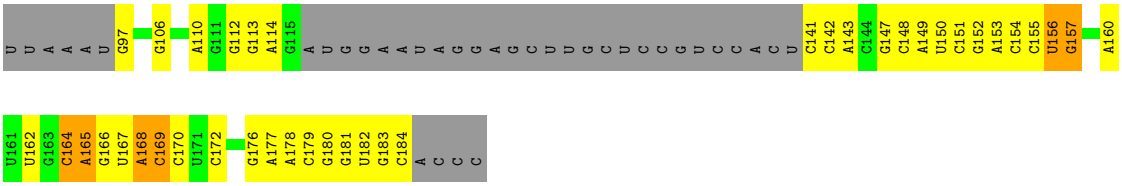
Chain 4Z:



• Molecule 33: U2 snRNA

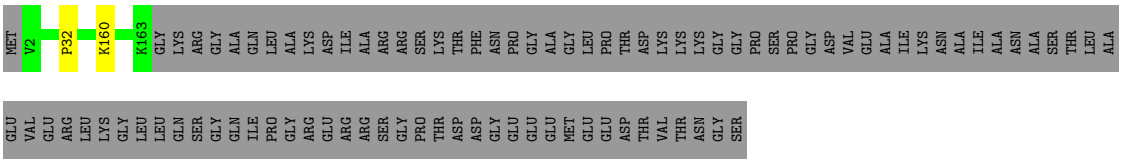
Chain 2A:





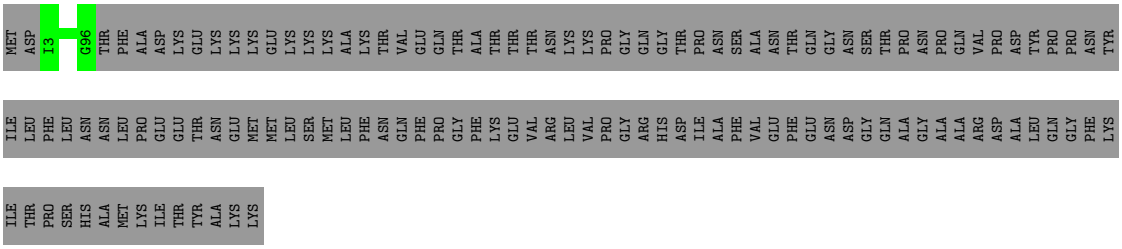
• Molecule 34: U2 small nuclear ribonucleoprotein A'

Chain 2B:



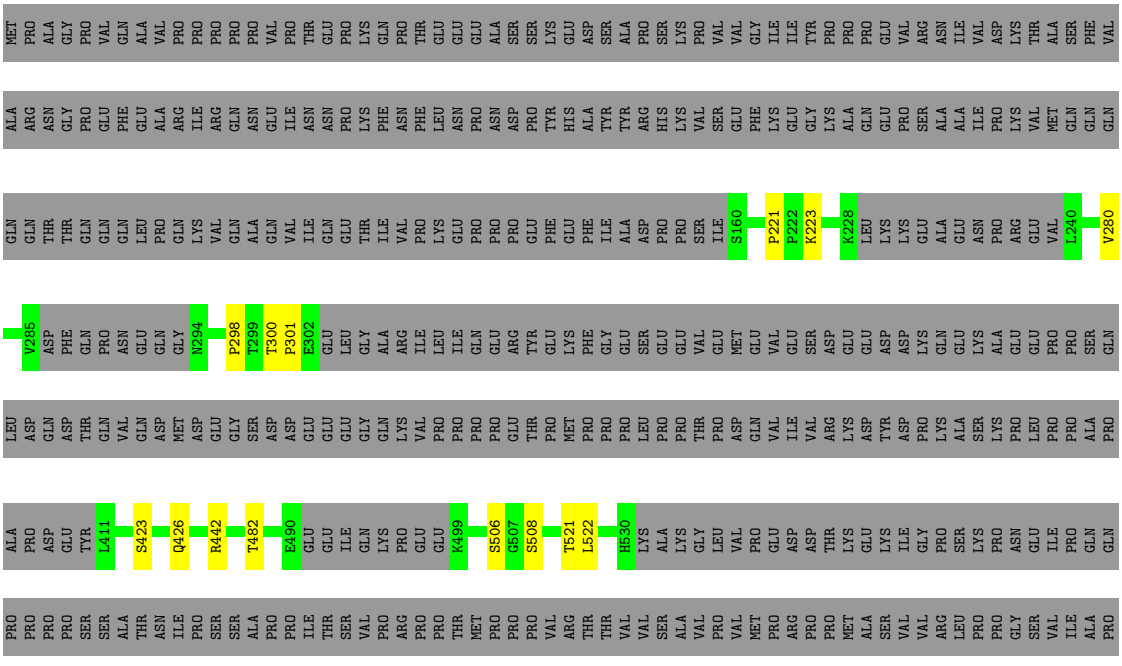
• Molecule 35: U2 small nuclear ribonucleoprotein B"

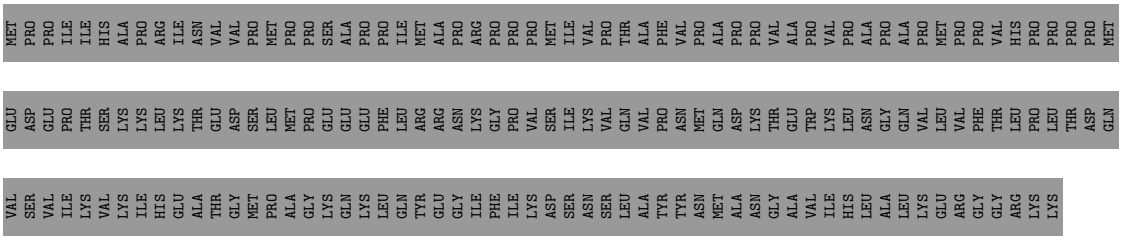
Chain 2C:



• Molecule 36: Splicing factor 3A subunit 1

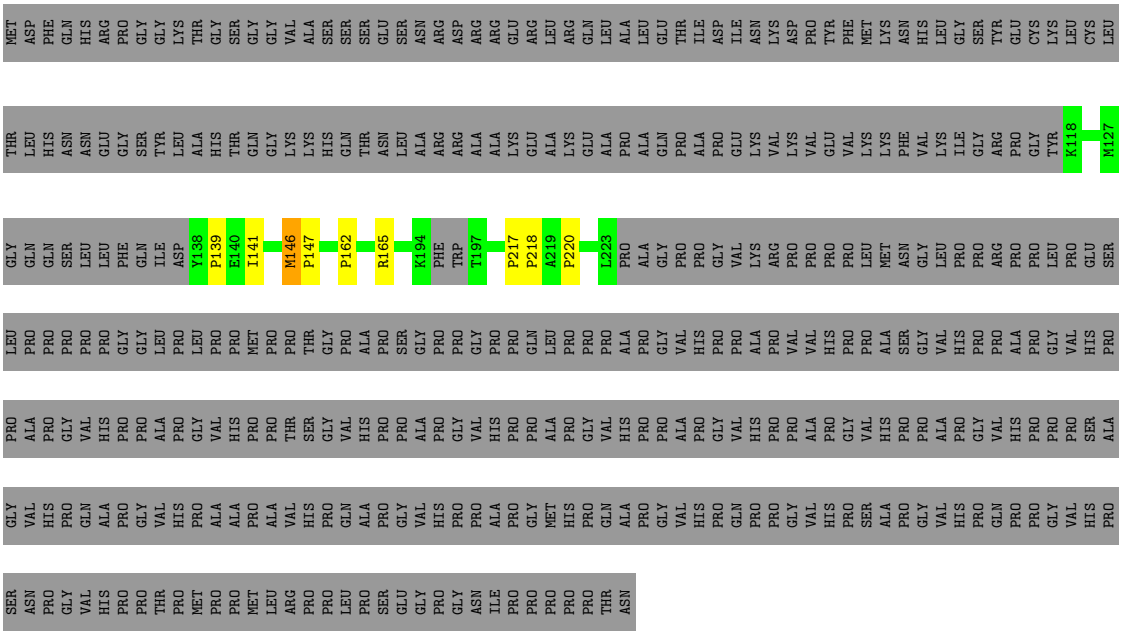
Chain 2D:





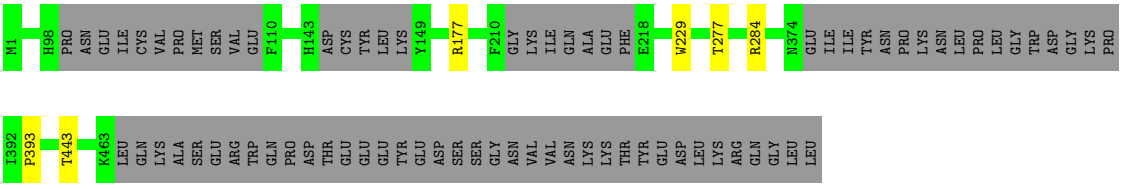
• Molecule 37: Splicing factor 3A subunit 2

Chain 2E:



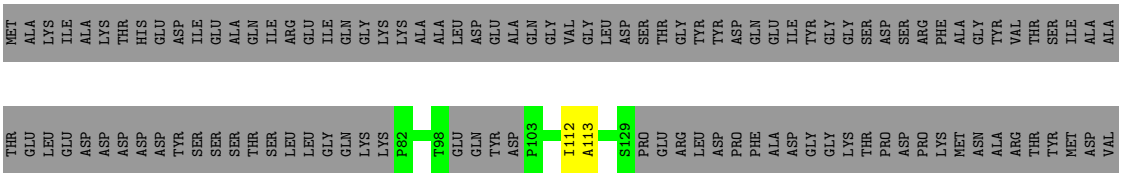
• Molecule 38: Splicing factor 3A subunit 3

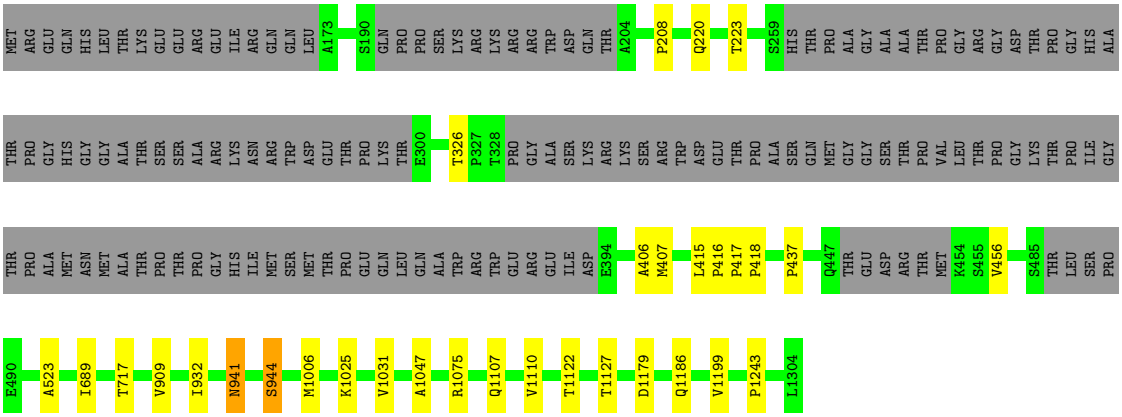
Chain 2F:



• Molecule 39: Splicing factor 3B subunit 1

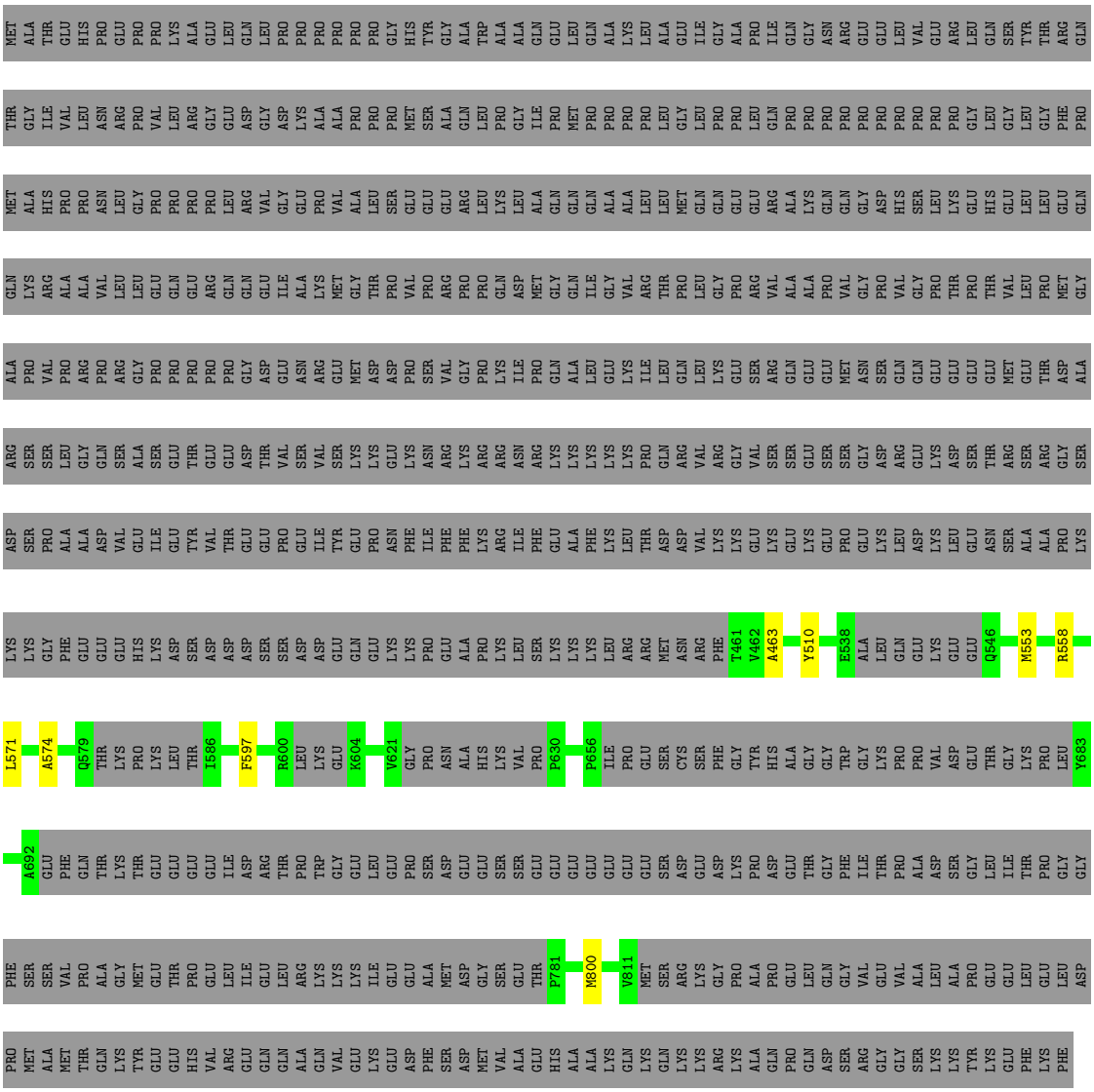
Chain 2G:





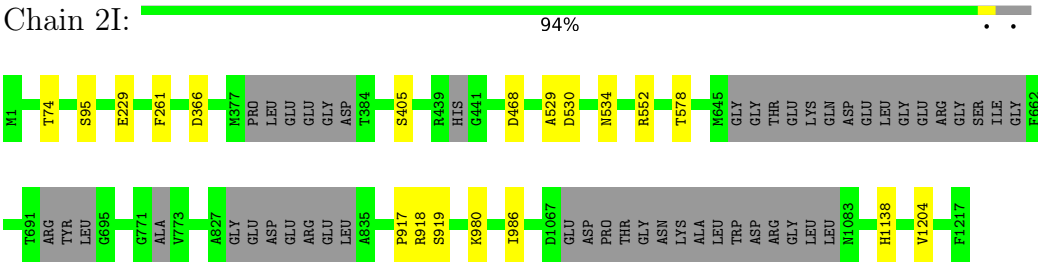
• Molecule 40: Splicing factor 3B subunit 2

Chain 2H:



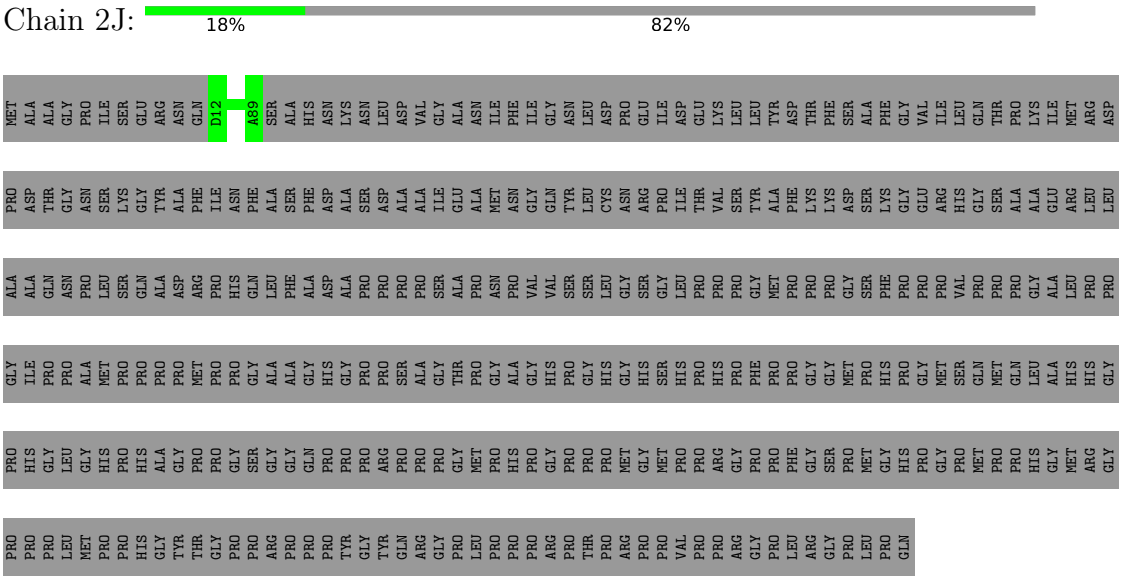
• Molecule 41: Splicing factor 3B subunit 3

Chain 2I:



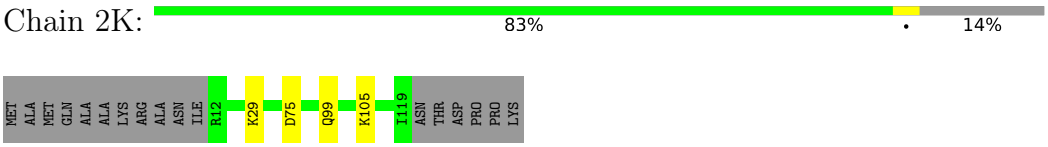
• Molecule 42: Splicing factor 3B subunit 4

Chain 2J:



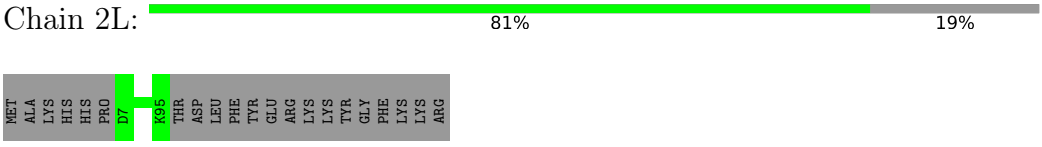
• Molecule 43: Splicing factor 3B subunit 6

Chain 2K:



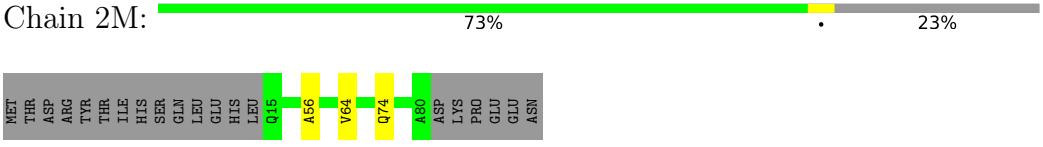
• Molecule 44: PHD finger-like domain-containing protein 5A

Chain 2L:



• Molecule 45: Splicing factor 3B subunit 5

Chain 2M:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	716083	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 ($e^-/Å^2$)	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, IHP, M7M, MG, 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	0.65	3/1317 (0.2%)	0.90	1/2042 (0.0%)
2	5A	0.29	0/2698	0.82	0/4195
3	5B	0.28	1/19157 (0.0%)	0.51	2/26004 (0.0%)
4	5C	0.27	1/6580 (0.0%)	0.56	3/8938 (0.0%)
5	5D	0.26	0/13923	0.49	1/18868 (0.0%)
6	5E	0.67	0/1195	0.71	0/1492
7	2a	0.50	0/343	0.69	0/427
7	4a	0.22	0/254	0.48	0/314
7	5a	0.50	0/335	0.68	0/417
8	2b	0.56	0/327	0.68	0/407
8	4b	0.22	0/333	0.48	0/416
8	5b	0.57	0/327	0.67	0/407
9	2c	0.70	0/338	0.73	0/419
9	4c	0.23	0/298	0.48	0/370
9	5c	0.69	0/387	0.72	0/482
10	2d	0.77	0/295	0.76	0/367
10	4d	0.24	0/291	0.49	0/363
10	5d	0.77	0/295	0.76	0/367
11	2e	0.64	0/315	0.75	0/392
11	4e	0.22	0/313	0.49	0/390
11	5e	0.65	0/315	0.74	0/392
12	2f	0.55	0/270	0.63	0/334
12	4f	0.24	0/297	0.51	0/371
12	5f	0.54	0/287	0.61	0/357
13	2g	0.47	0/318	0.56	0/394
13	4g	0.23	0/287	0.49	0/358
13	5g	0.46	0/302	0.56	0/374
14	6A	0.30	0/1398	0.81	0/2172
15	6a	0.43	0/359	0.67	0/447
16	6b	0.46	0/294	0.75	0/364
17	6c	0.34	0/294	0.61	0/364
18	6d	0.43	0/286	0.59	0/354

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#|Z| >5	RMSZ	#|Z| >5
19	6e	0.43	0/279	0.72	0/347
20	6f	0.38	0/258	0.61	0/319
21	6g	0.41	0/242	0.64	0/299
22	4A	0.31	0/3025	0.77	1/4702 (0.0%)
23	4B	0.25	0/2114	0.50	0/2836
24	4C	0.25	0/3452	0.53	0/4675
25	4D	0.25	0/2912	0.50	0/3924
26	4E	0.25	0/974	0.47	0/1316
27	4F	0.28	0/1198	0.50	0/1620
28	4G	0.24	0/5592	0.48	1/7615 (0.0%)
29	4H	0.24	0/853	0.45	0/1188
30	4I	0.28	0/502	0.62	2/683 (0.3%)
31	4J	0.25	0/1158	0.52	0/1553
32	4Z	0.24	0/2101	0.45	0/2928
33	2A	0.86	11/2576 (0.4%)	1.43	55/4003 (1.4%)
34	2B	0.63	0/647	1.42	0/807
35	2C	0.61	0/375	1.20	0/467
36	2D	0.23	0/1388	0.48	0/1813
37	2E	0.22	0/373	0.58	1/461 (0.2%)
38	2F	0.25	0/1688	0.47	0/2102
39	2G	1.04	4/4184 (0.1%)	0.83	2/5216 (0.0%)
40	2H	0.65	0/957	0.67	0/1209
41	2I	0.85	0/4664	0.76	0/5816
42	2J	0.62	0/311	0.64	0/387
43	2K	0.79	0/431	0.79	0/537
44	2L	0.74	0/355	0.68	0/442
45	2M	1.01	0/263	0.77	0/327
All	All	0.45	20/96900 (0.0%)	0.65	69/130950 (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
3	5B	0	1
9	2c	0	1
9	5c	0	1
25	4D	0	1
38	2F	0	1
39	2G	0	11
40	2H	0	3

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
41	2I	0	11
43	2K	0	1
45	2M	0	1
All	All	0	32

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
39	2G	407	MET	N-CA	12.36	1.71	1.46
39	2G	406	ALA	C-N	7.94	1.52	1.34
33	2A	142	C	C1'-N1	7.32	1.59	1.48
39	2G	1243	PRO	N-CA	-7.11	1.35	1.47
33	2A	182	U	C1'-N1	6.94	1.59	1.48

The worst 5 of 69 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	5C	810	PRO	CA-N-CD	-14.23	91.58	111.50
33	2A	167	U	C5-C4-O4	11.61	132.87	125.90
39	2G	406	ALA	C-N-CA	10.28	147.39	121.70
33	2A	164	C	N1-C2-O2	-10.12	112.83	118.90
3	5B	1194	CYS	CA-CB-SG	9.56	131.21	114.00

There are no chirality outliers.

5 of 32 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
38	2F	443	THR	Peptide
39	2G	220	GLN	Peptide
25	4D	358	ARG	Sidechain
3	5B	941	LYS	Peptide
9	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 (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	
3	5B	2249/2335 (96%)	2145 (95%)	104 (5%)	0	100	100
4	5C	814/972 (84%)	745 (92%)	68 (8%)	1 (0%)	51	75
5	5D	1694/2136 (79%)	1618 (96%)	75 (4%)	1 (0%)	51	75
6	5E	297/357 (83%)	272 (92%)	16 (5%)	9 (3%)	4	7
7	2a	84/231 (36%)	82 (98%)	2 (2%)	0	100	100
7	4a	60/231 (26%)	57 (95%)	3 (5%)	0	100	100
7	5a	82/231 (36%)	80 (98%)	2 (2%)	0	100	100
8	2b	80/119 (67%)	77 (96%)	3 (4%)	0	100	100
8	4b	80/119 (67%)	76 (95%)	4 (5%)	0	100	100
8	5b	80/119 (67%)	77 (96%)	3 (4%)	0	100	100
9	2c	81/118 (69%)	78 (96%)	3 (4%)	0	100	100
9	4c	70/118 (59%)	68 (97%)	2 (3%)	0	100	100
9	5c	95/118 (80%)	91 (96%)	4 (4%)	0	100	100
10	2d	72/86 (84%)	68 (94%)	4 (6%)	0	100	100
10	4d	69/86 (80%)	67 (97%)	2 (3%)	0	100	100
10	5d	72/86 (84%)	69 (96%)	3 (4%)	0	100	100
11	2e	77/92 (84%)	76 (99%)	1 (1%)	0	100	100
11	4e	76/92 (83%)	70 (92%)	6 (8%)	0	100	100
11	5e	77/92 (84%)	76 (99%)	1 (1%)	0	100	100
12	2f	64/76 (84%)	62 (97%)	2 (3%)	0	100	100
12	4f	71/76 (93%)	67 (94%)	4 (6%)	0	100	100
12	5f	70/76 (92%)	68 (97%)	2 (3%)	0	100	100
13	2g	76/126 (60%)	75 (99%)	1 (1%)	0	100	100
13	4g	69/126 (55%)	69 (100%)	0	0	100	100
13	5g	72/126 (57%)	70 (97%)	2 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
15	6a	88/95 (93%)	77 (88%)	7 (8%)	4 (4%)	2 3
16	6b	70/102 (69%)	64 (91%)	3 (4%)	3 (4%)	2 3
17	6c	70/139 (50%)	63 (90%)	6 (9%)	1 (1%)	11 22
18	6d	68/91 (75%)	63 (93%)	4 (6%)	1 (2%)	10 21
19	6e	68/80 (85%)	64 (94%)	2 (3%)	2 (3%)	4 7
20	6f	61/103 (59%)	56 (92%)	5 (8%)	0	100 100
21	6g	57/96 (59%)	52 (91%)	4 (7%)	1 (2%)	8 16
23	4B	248/683 (36%)	229 (92%)	19 (8%)	0	100 100
24	4C	422/522 (81%)	388 (92%)	33 (8%)	1 (0%)	47 71
25	4D	372/499 (74%)	354 (95%)	18 (5%)	0	100 100
26	4E	122/128 (95%)	112 (92%)	10 (8%)	0	100 100
27	4F	139/142 (98%)	134 (96%)	5 (4%)	0	100 100
28	4G	795/941 (84%)	745 (94%)	50 (6%)	0	100 100
29	4H	167/177 (94%)	156 (93%)	11 (7%)	0	100 100
30	4I	73/376 (19%)	71 (97%)	2 (3%)	0	100 100
31	4J	143/800 (18%)	136 (95%)	7 (5%)	0	100 100
32	4Z	414/513 (81%)	401 (97%)	12 (3%)	1 (0%)	47 71
34	2B	160/255 (63%)	146 (91%)	12 (8%)	2 (1%)	12 24
35	2C	92/225 (41%)	90 (98%)	2 (2%)	0	100 100
36	2D	226/793 (28%)	208 (92%)	12 (5%)	6 (3%)	5 8
37	2E	88/464 (19%)	63 (72%)	16 (18%)	9 (10%)	0 0
38	2F	413/501 (82%)	367 (89%)	41 (10%)	5 (1%)	13 27
39	2G	1032/1304 (79%)	844 (82%)	166 (16%)	22 (2%)	7 13
40	2H	199/895 (22%)	179 (90%)	16 (8%)	4 (2%)	7 14
41	2I	1152/1217 (95%)	1053 (91%)	89 (8%)	10 (1%)	17 35
42	2J	76/424 (18%)	75 (99%)	1 (1%)	0	100 100
43	2K	106/125 (85%)	85 (80%)	18 (17%)	3 (3%)	5 7
44	2L	87/110 (79%)	74 (85%)	13 (15%)	0	100 100
45	2M	64/86 (74%)	55 (86%)	7 (11%)	2 (3%)	4 6
All	All	13703/20230 (68%)	12707 (93%)	908 (7%)	88 (1%)	29 47

5 of 88 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	5D	1086	GLN
6	5E	193	THR
15	6a	55	LEU
16	6b	84	MET
18	6d	70	ASP

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
3	5B	2034/2108 (96%)	1932 (95%)	102 (5%)	24 47
4	5C	718/866 (83%)	676 (94%)	42 (6%)	20 40
5	5D	1517/1908 (80%)	1493 (98%)	24 (2%)	62 82
23	4B	225/599 (38%)	205 (91%)	20 (9%)	9 19
24	4C	362/442 (82%)	330 (91%)	32 (9%)	10 19
25	4D	299/424 (70%)	278 (93%)	21 (7%)	15 30
26	4E	108/111 (97%)	104 (96%)	4 (4%)	34 60
27	4F	129/130 (99%)	121 (94%)	8 (6%)	18 37
28	4G	417/792 (53%)	388 (93%)	29 (7%)	15 30
29	4H	10/148 (7%)	10 (100%)	0	100 100
30	4I	32/333 (10%)	28 (88%)	4 (12%)	4 8
31	4J	113/681 (17%)	102 (90%)	11 (10%)	8 15
32	4Z	11/450 (2%)	11 (100%)	0	100 100
36	2D	95/709 (13%)	87 (92%)	8 (8%)	11 21
40	2H	26/776 (3%)	25 (96%)	1 (4%)	33 59
All	All	6096/10477 (58%)	5790 (95%)	306 (5%)	28 47

5 of 306 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
25	4D	214	PHE
31	4J	149	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
25	4D	421	SER
28	4G	155	GLU
36	2D	482	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
5	5D	785	HIS
28	4G	741	HIS
23	4B	480	ASN
31	4J	261	HIS
25	4D	270	HIS

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	55/144 (38%)	35 (63%)	10 (18%)
14	6A	55/107 (51%)	14 (25%)	2 (3%)
2	5A	114/117 (97%)	30 (26%)	5 (4%)
22	4A	124/145 (85%)	35 (28%)	4 (3%)
33	2A	105/188 (55%)	22 (20%)	3 (2%)
All	All	453/701 (64%)	136 (30%)	24 (5%)

5 of 136 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 24 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	5A	96	A
22	4A	18	G
14	6A	77	C
22	4A	38	U
1	A	38	C

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	$\#\|Z\| > 2$
46	IHP	5B	3000	-	36,36,36	0.73	0	54,60,60	1.07	2 (3%)
47	GTP	5C	1500	48	26,34,34	1.13	2 (7%)	32,54,54	1.53	7 (21%)

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
46	IHP	5B	3000	-	-	3/30/54/54	0/1/1/1
47	GTP	5C	1500	48	-	6/18/38/38	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
47	5C	1500	GTP	C5-C6	-4.03	1.39	1.47
47	5C	1500	GTP	C2-N3	2.08	1.38	1.33

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
47	5C	1500	GTP	PB-O3B-PG	-3.28	121.57	132.83
47	5C	1500	GTP	C5-C6-N1	3.26	119.72	113.95
47	5C	1500	GTP	PA-O3A-PB	-3.15	122.01	132.83
47	5C	1500	GTP	C8-N7-C5	3.01	108.72	102.99
47	5C	1500	GTP	C2-N1-C6	-2.88	119.80	125.10

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

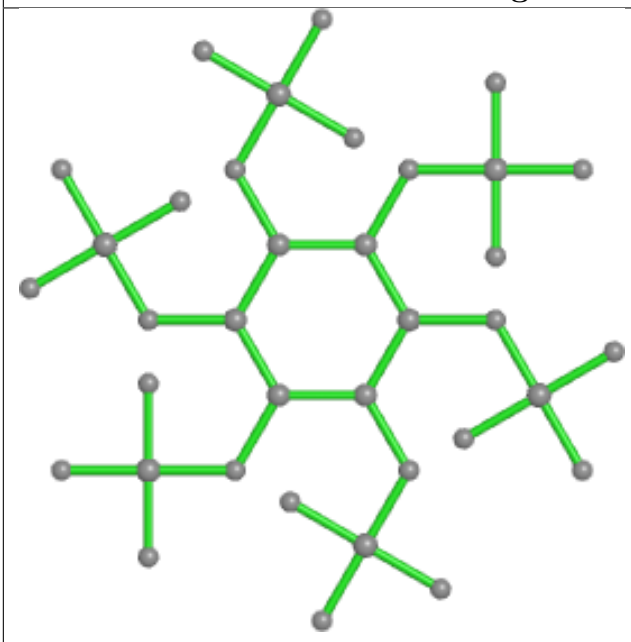
Mol	Chain	Res	Type	Atoms
46	5B	3000	IHP	C2-O12-P2-O42
46	5B	3000	IHP	C4-O14-P4-O44
47	5C	1500	GTP	PB-O3B-PG-O3G
47	5C	1500	GTP	C5'-O5'-PA-O1A
47	5C	1500	GTP	PB-O3A-PA-O5'

There are no ring outliers.

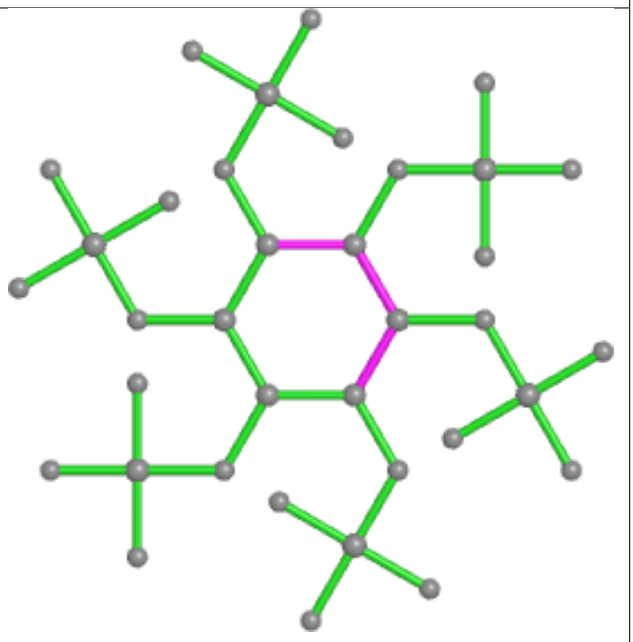
No monomer is involved in short contacts.

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 then 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.

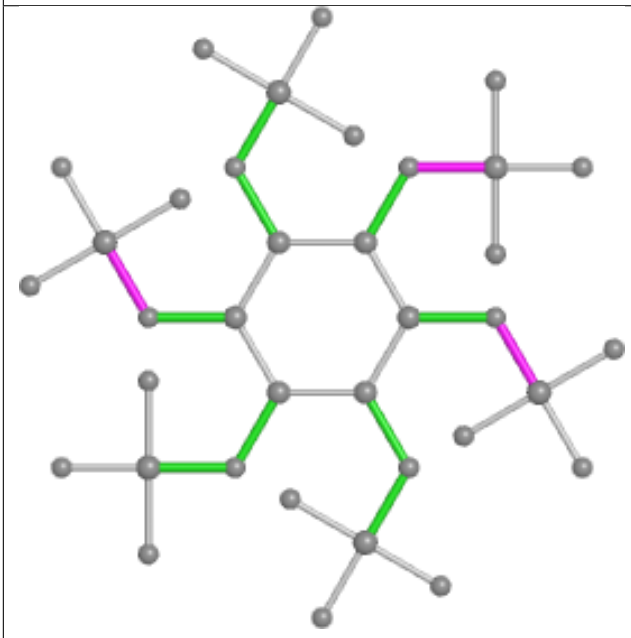
Ligand IHP 5B 3000



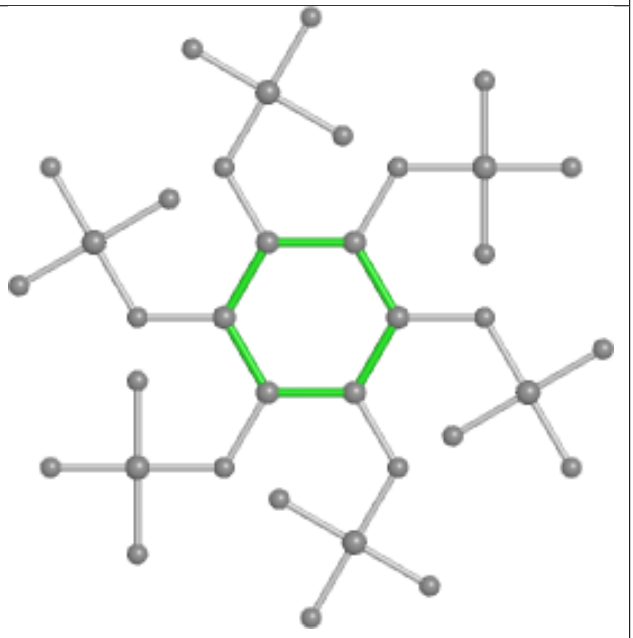
Bond lengths



Bond angles

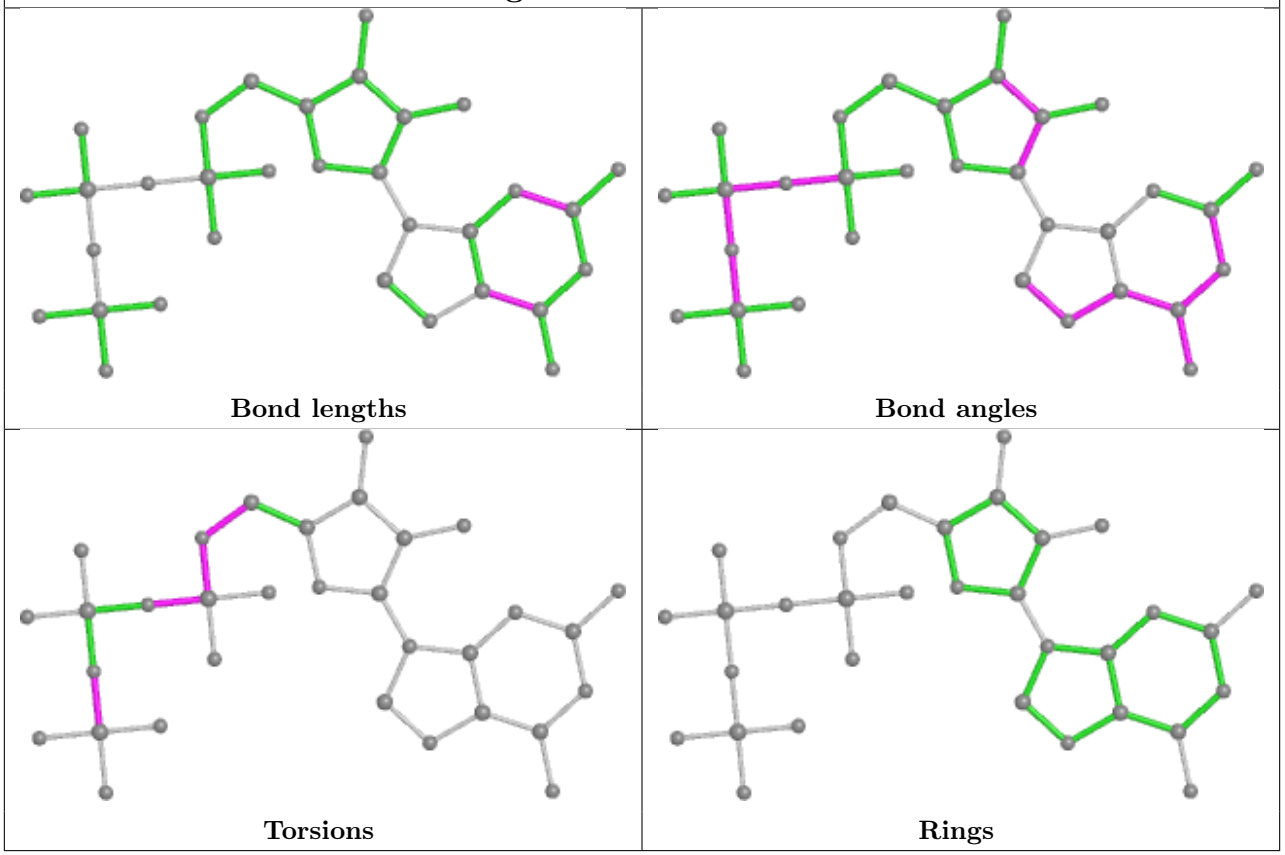


Torsions



Rings

Ligand GTP 5C 1500



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.