wwPDB EM Validation Summary Report ⓘ

Jun 3, 2024 – 09:29 PM JST

PDB ID	:	8H6J
EMDB ID	:	EMD-34505
Title	:	Cryo-EM structure of human exon-defined spliceosome in the mature pre-B state.
Authors	:	Zhang, W.; Zhan, X.; Zhang, X.; Lei, J.; Yan, C.; Shi, Y.
Deposited on	:	2022-10-18
Resolution	:	3.25 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:

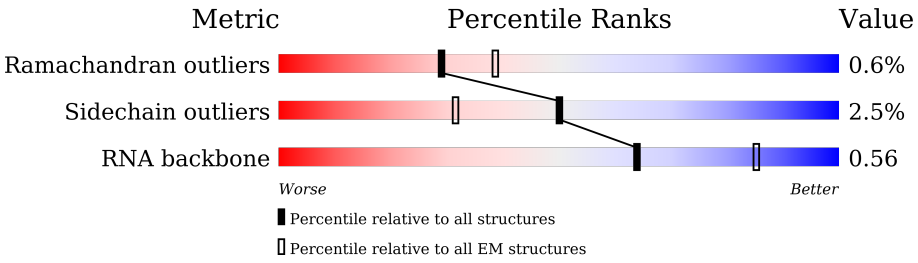
EMDB validation analysis	:	FAILED
Mogul	:	1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

1 Overall quality at a glance

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

The reported resolution of this entry is 3.25 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.



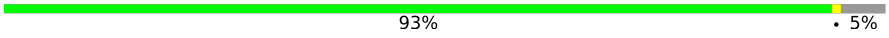
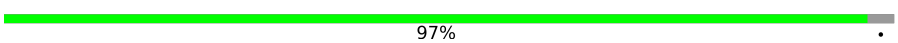
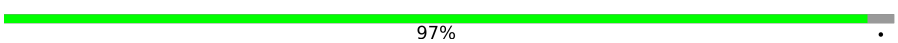
Metric	Whole archive (#Entries)	EM structures (#Entries)
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826
RNA backbone	4643	859

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	8% 20% 71%
2	6A	107	40% 9% 50%
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	
11	5B	2335	
12	5C	972	
13	5D	2136	
14	5E	357	
15	2a	231	
15	4a	231	
15	5a	231	
16	2b	119	
16	4b	119	
16	5b	119	
17	2c	118	
17	4c	118	
17	5c	118	
18	2d	86	
18	4d	86	
18	5d	86	
19	2e	92	
19	4e	92	
19	5e	92	
20	2f	76	
20	4f	76	
20	5f	76	
21	2g	126	
21	4g	126	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
21	5g	126	61% 39%
22	4A	144	67% 19% 13%
23	4B	683	28% 72%
24	4C	522	69% 31%
25	4D	499	54% 46%
26	4E	128	97%
27	4F	142	96%
28	4G	941	81% 17%
29	4R	480	21% 78%
30	4S	800	8% 92%
31	4T	565	80% 19%
32	4U	820	70% 30%
33	4X	155	13% 86%
34	4Y	1007	32% 68%
35	2A	188	33% 21% 42%
36	2B	255	63% 36%
37	2C	225	42% 58%
38	2D	793	15% 84%
39	2E	464	18% 80%
40	2F	501	83% 16%
41	2G	1304	78% 20%
42	2H	895	20% 80%
43	2I	1217	94%
44	2J	424	18% 82%
45	2K	125	83% 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
46	2L	110	
47	2M	86	

2 Entry composition (i)

There are 51 unique types of molecules in this entry. The entry contains 96473 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	42	864	387	124	311	42	0	0

- Molecule 2 is a RNA chain called U6 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	6A	54	1142	509	206	373	54	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
7	6e	70	Total	C	N	O	0	0
			280	140	70	70		

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

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

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

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

- Molecule 10 is a RNA chain called U5 snRNA.

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

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

Mol	Chain	Residues	Atoms					AltConf	Trace
11	5B	2209	Total	C	N	O	S	0	0
			18244	11760	3170	3235	79		

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

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

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

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

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

Mol	Chain	Residues	Atoms				AltConf	Trace
14	5E	301	Total	C	N	O	0	0
			1481	879	301	301		

- 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	86	Total	C	N	O	0	0
			344	172	86	86		
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	74	Total	C	N	O	0	0
			363	215	74	74		
20	2f	66	Total	C	N	O	0	0
			264	132	66	66		

- 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
			2656	1188	468	876	124		

- 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
			953	567	193	193		

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

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

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

Mol	Chain	Residues	Atoms				AltConf	Trace
25	4D	270	Total 1340	C 800	N 270	O 270	0	0

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

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

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

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

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

Mol	Chain	Residues	Atoms					AltConf	Trace
28	4G	784	Total 4539	C 2745	N 884	O 901	S 9	0	0

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

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

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

Mol	Chain	Residues	Atoms					AltConf	Trace
30	4S	61	Total 505	C 317	N 94	O 91	S 3	0	0

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

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

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

Mol	Chain	Residues	Atoms					AltConf	Trace
32	4U	578	Total	C	N	O	S	1	0
			3414	2063	682	667	2		

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

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

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

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

- Molecule 35 is a RNA chain called U2 snRNA.

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

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

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

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

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

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

Mol	Chain	Residues	Atoms				AltConf	Trace
38	2D	124	Total 496	C 248	N 124	O 124	0	0

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

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

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

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

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

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

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

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

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

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

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

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

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

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

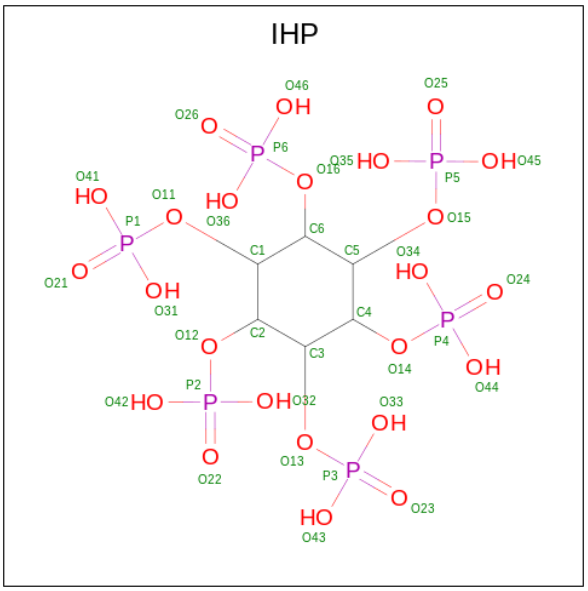
- Molecule 46 is a protein called PHD finger-like domain-containing protein 5A.

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

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

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

- Molecule 48 is INOSITOL HEXAKISPHOSPHATE (three-letter code: IHP) (formula: $C_6H_{18}O_{24}P_6$).

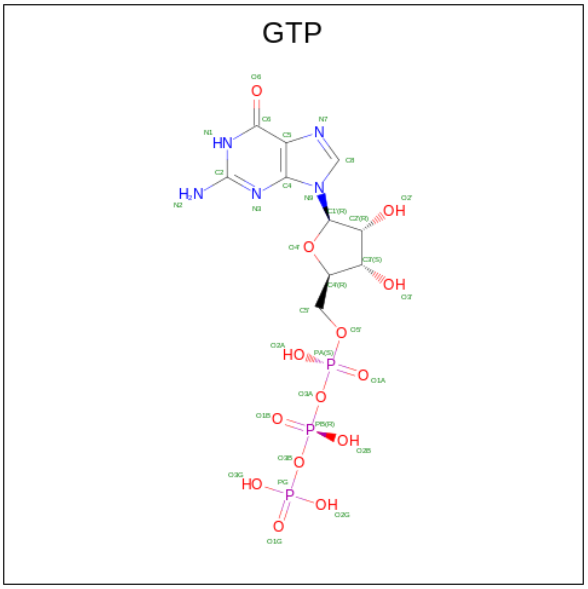


Mol	Chain	Residues	Atoms				AltConf
48	5B	1	Total 36	C 6	O 24	P 6	0

- Molecule 49 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
49	5C	1	Total 1	Mg 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	0
			32	10	5	14	3	

- Molecule 51 is ZINC ION (three-letter code: ZN) (formula: Zn).

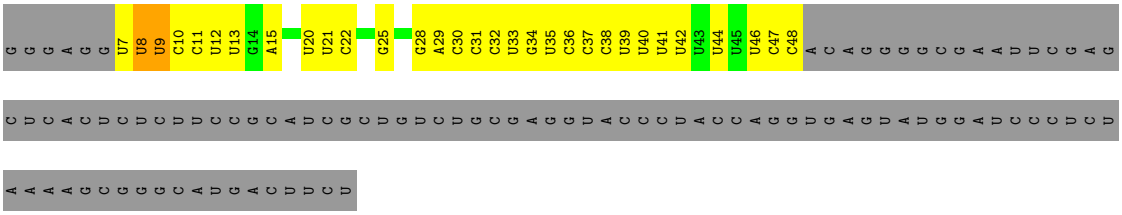
Mol	Chain	Residues	Atoms		AltConf
51	4T	1	Total	Zn	0
			1	1	

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.

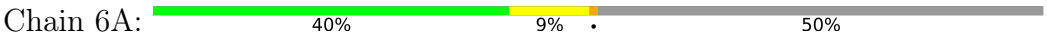
- Molecule 1: pre-mRNA

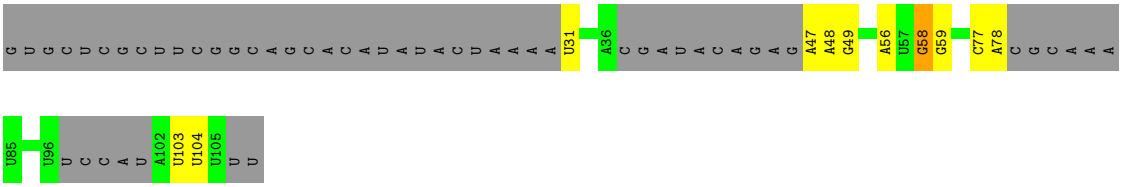
Chain A:



- Molecule 2: U6 snRNA

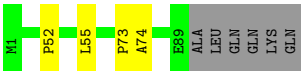
Chain 6A:





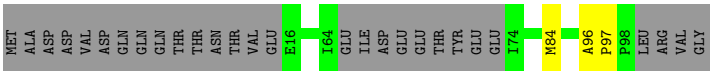
- Molecule 3: U6 snRNA-associated Sm-like protein LSm2

Chain 6a:



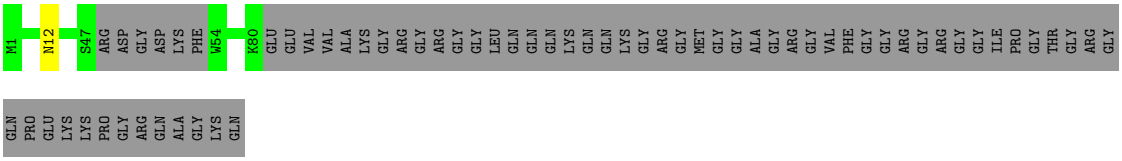
- Molecule 4: U6 snRNA-associated Sm-like protein LSm3

Chain 6b:



- Molecule 5: U6 snRNA-associated Sm-like protein LSm4

Chain 6c:



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

Chain 6d:

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

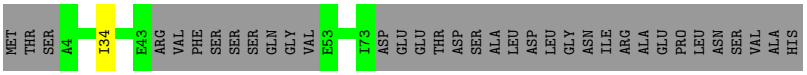
Chain 6e:

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

Chain 6f:

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

Chain 6g:

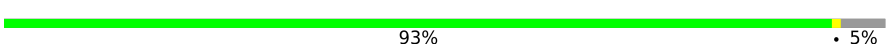


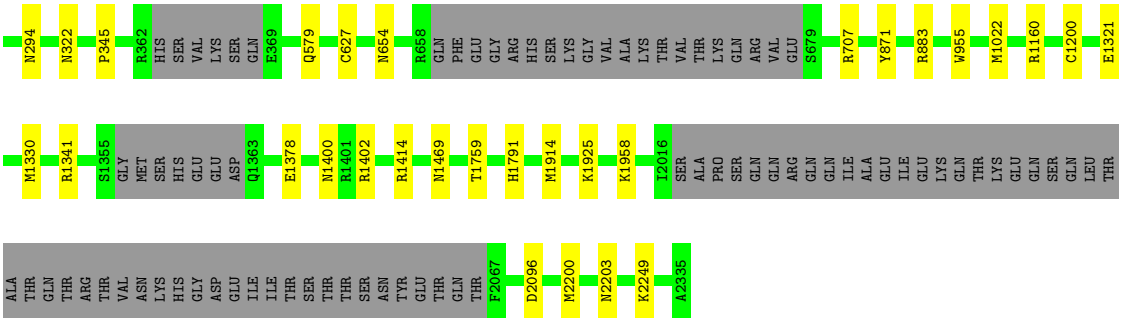
• Molecule 10: U5 snRNA

Chain 5A:

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

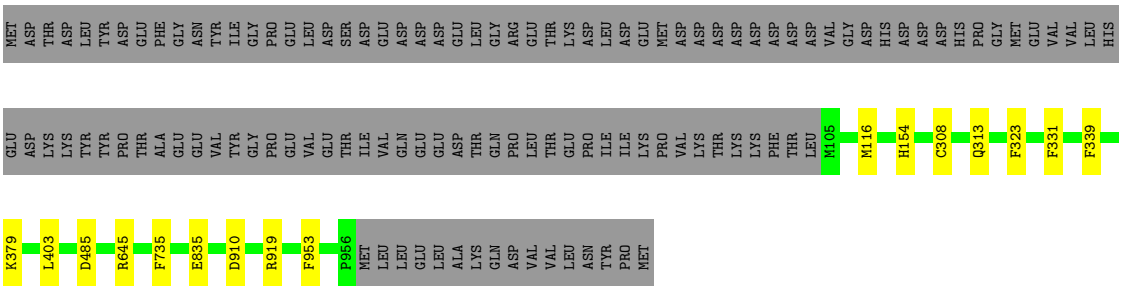
Chain 5B:





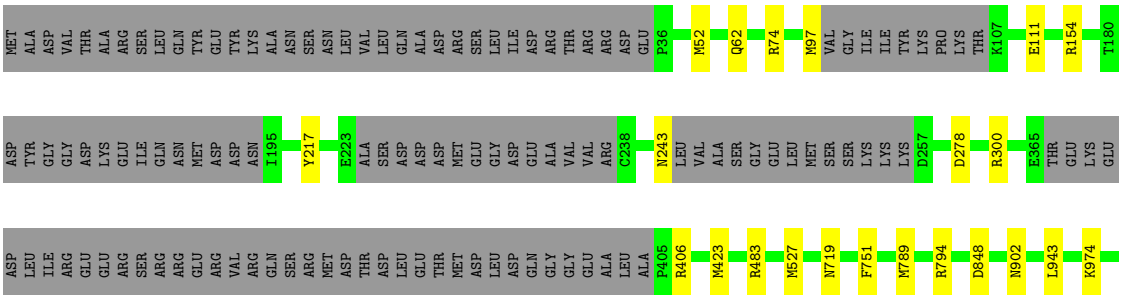
• Molecule 12: 116 kDa U5 small nuclear ribonucleoprotein component

Chain 5C:



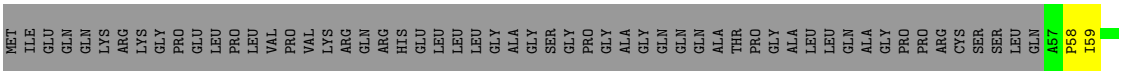
• Molecule 13: U5 small nuclear ribonucleoprotein 200 kDa helicase

Chain 5D:



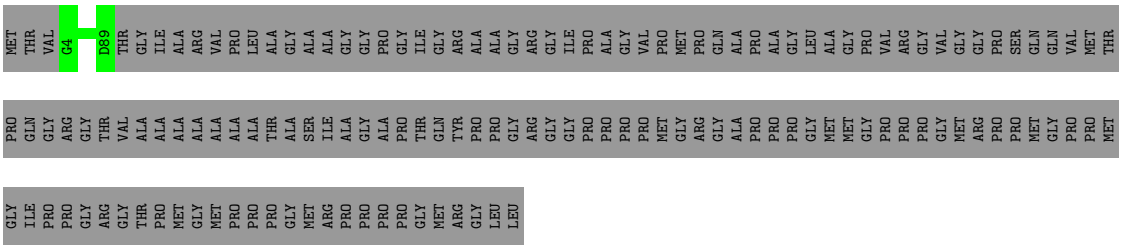
• Molecule 14: U5 small nuclear ribonucleoprotein 40 kDa protein

Chain 5E:



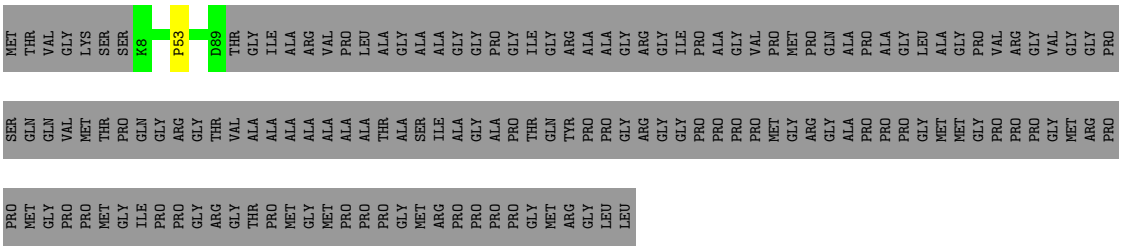
• Molecule 15: Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B'

Chain 5a:



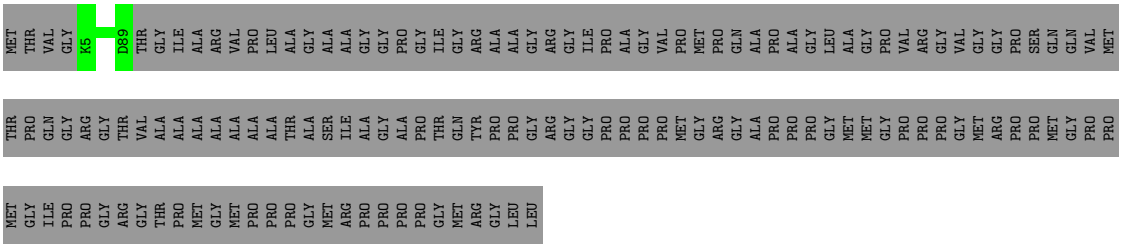
• Molecule 15: Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B'

Chain 4a:



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

Chain 2a:

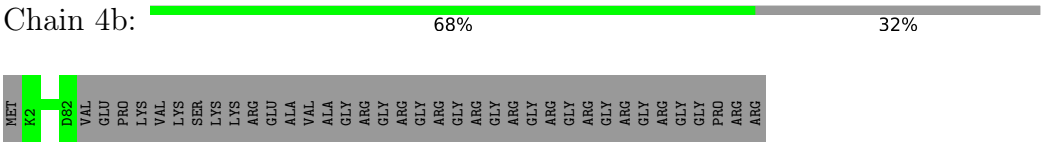


• Molecule 16: Small nuclear ribonucleoprotein Sm D1

Chain 5b:

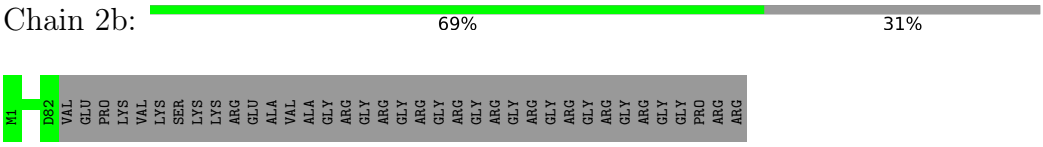
• Molecule 16: Small nuclear ribonucleoprotein Sm D1

Chain 4b:



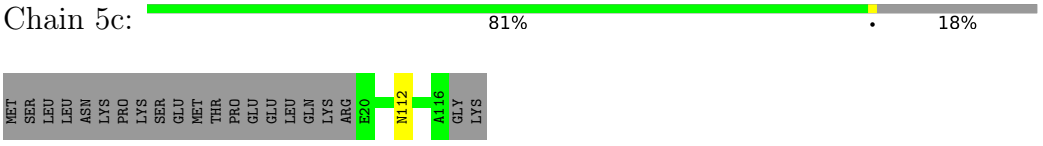
• Molecule 16: Small nuclear ribonucleoprotein Sm D1

Chain 2b:



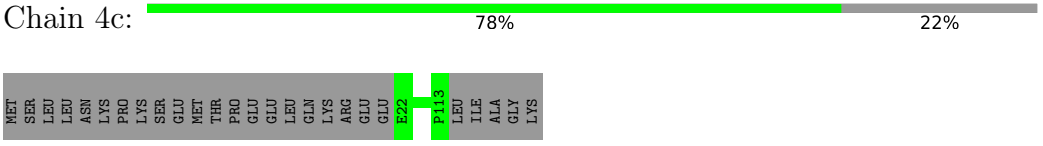
• Molecule 17: Small nuclear ribonucleoprotein Sm D2

Chain 5c:



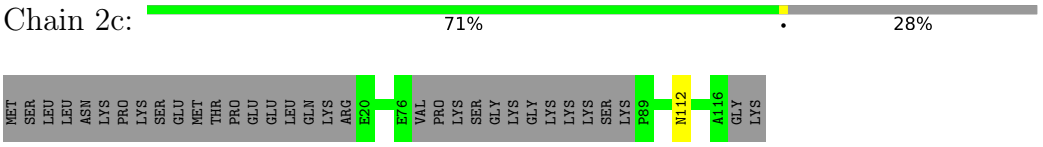
• Molecule 17: Small nuclear ribonucleoprotein Sm D2

Chain 4c:



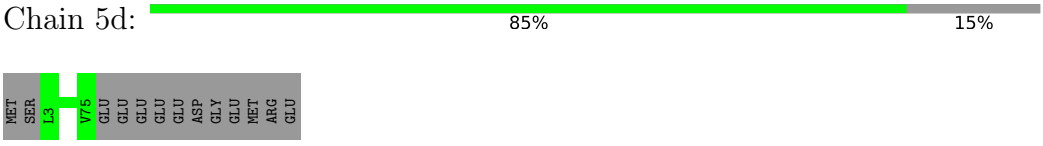
• Molecule 17: Small nuclear ribonucleoprotein Sm D2

Chain 2c:



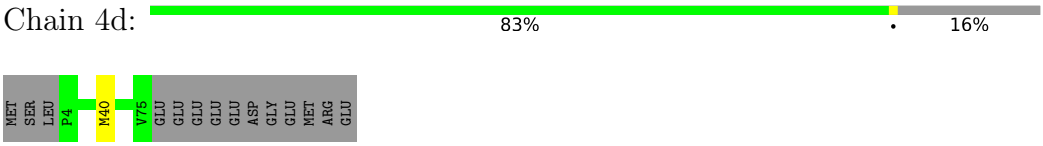
• Molecule 18: Small nuclear ribonucleoprotein F

Chain 5d:



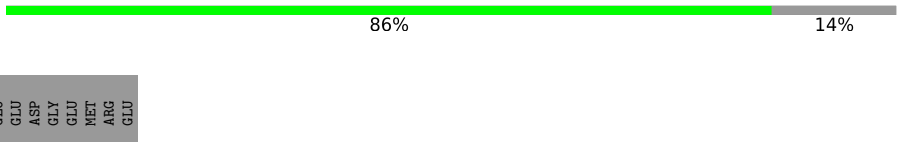
• Molecule 18: Small nuclear ribonucleoprotein F

Chain 4d:



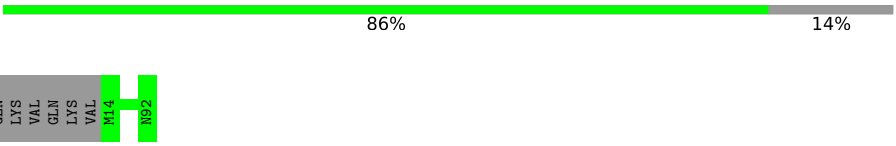
• Molecule 18: Small nuclear ribonucleoprotein F

Chain 2d:



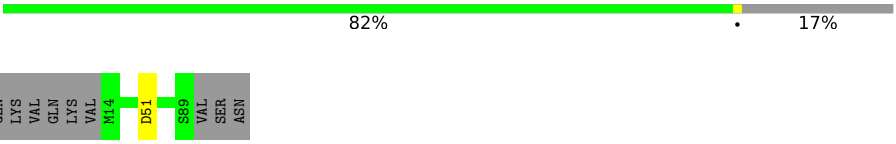
• Molecule 19: Small nuclear ribonucleoprotein E

Chain 5e:



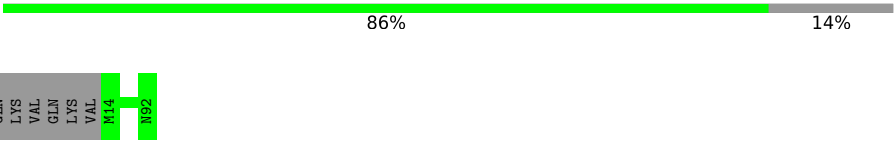
• Molecule 19: Small nuclear ribonucleoprotein E

Chain 4e:



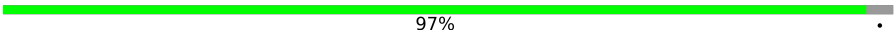
• Molecule 19: Small nuclear ribonucleoprotein E

Chain 2e:



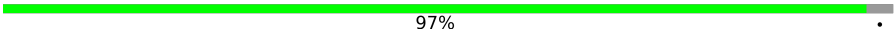
• Molecule 20: Small nuclear ribonucleoprotein G

Chain 5f:



• Molecule 20: Small nuclear ribonucleoprotein G

Chain 4f:

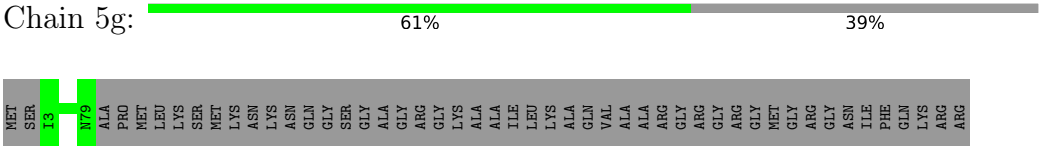


• Molecule 20: Small nuclear ribonucleoprotein G

Chain 2f:

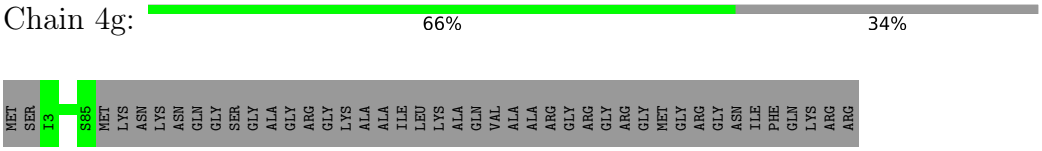
• Molecule 21: Small nuclear ribonucleoprotein Sm D3

Chain 5g:



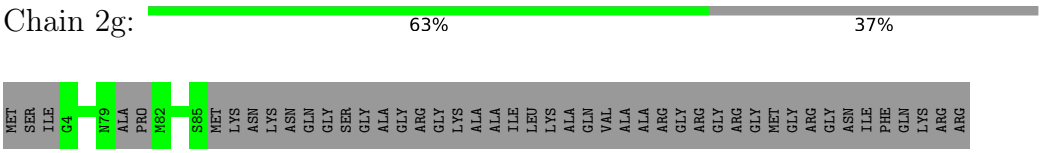
• Molecule 21: Small nuclear ribonucleoprotein Sm D3

Chain 4g:



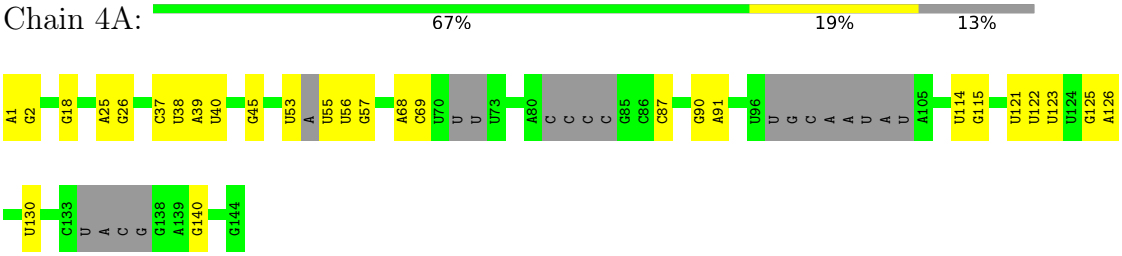
• Molecule 21: Small nuclear ribonucleoprotein Sm D3

Chain 2g:



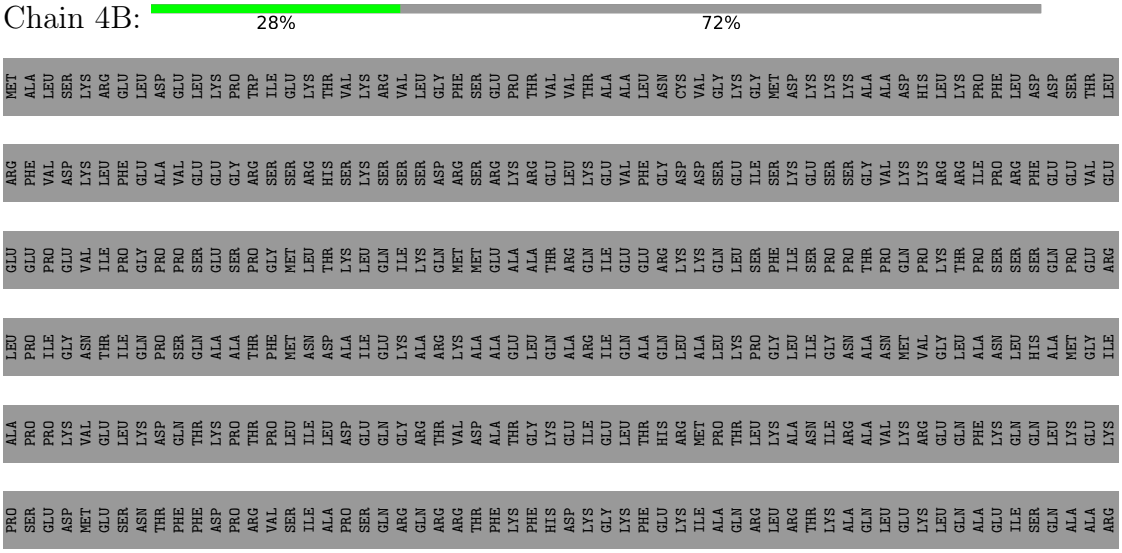
• Molecule 22: U4 snRNA

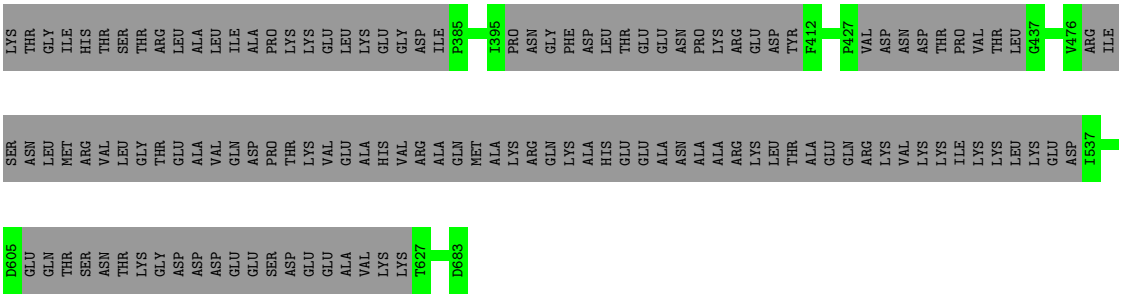
Chain 4A:



• Molecule 23: U4/U6 small nuclear ribonucleoprotein Prp3

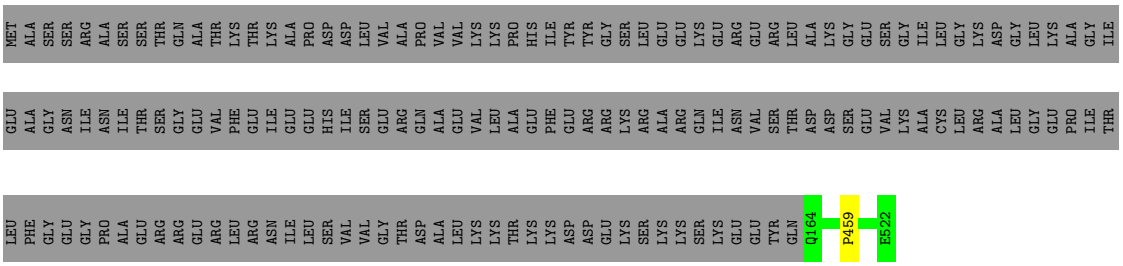
Chain 4B:





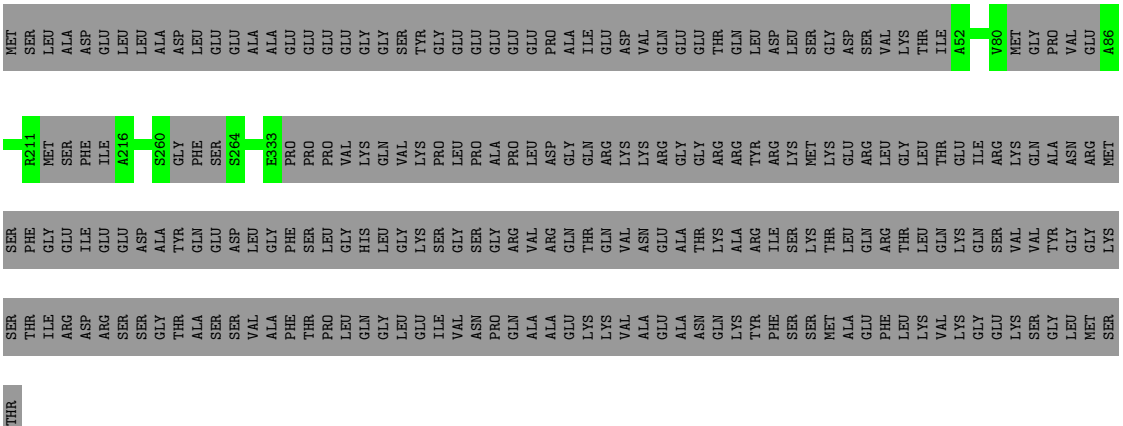
• Molecule 24: U4/U6 small nuclear ribonucleoprotein Prp4

Chain 4C:



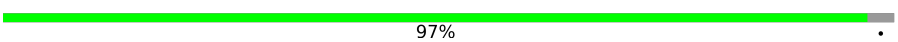
• Molecule 25: U4/U6 small nuclear ribonucleoprotein Prp31

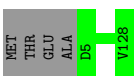
Chain 4D:



• Molecule 26: NHP2-like protein 1

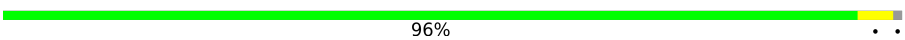
Chain 4E:

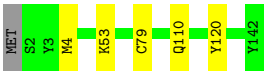




• Molecule 27: Thioredoxin-like protein 4A

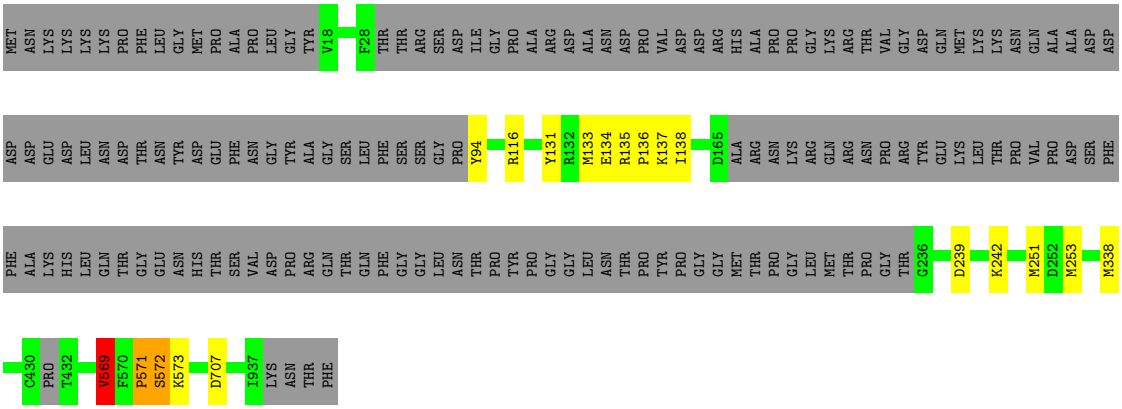
Chain 4F:





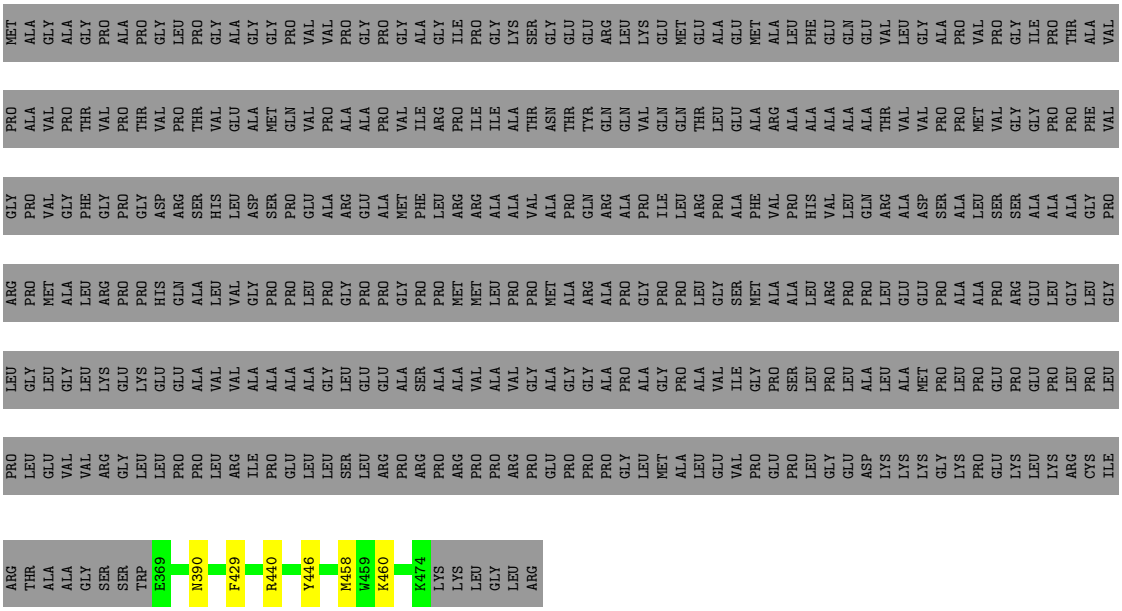
• Molecule 28: Pre-mRNA-processing factor 6

Chain 4G:



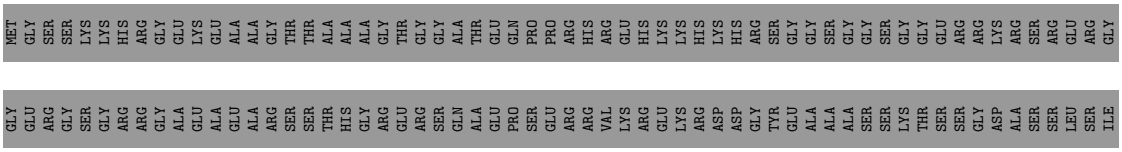
• Molecule 29: RNA-binding protein 42

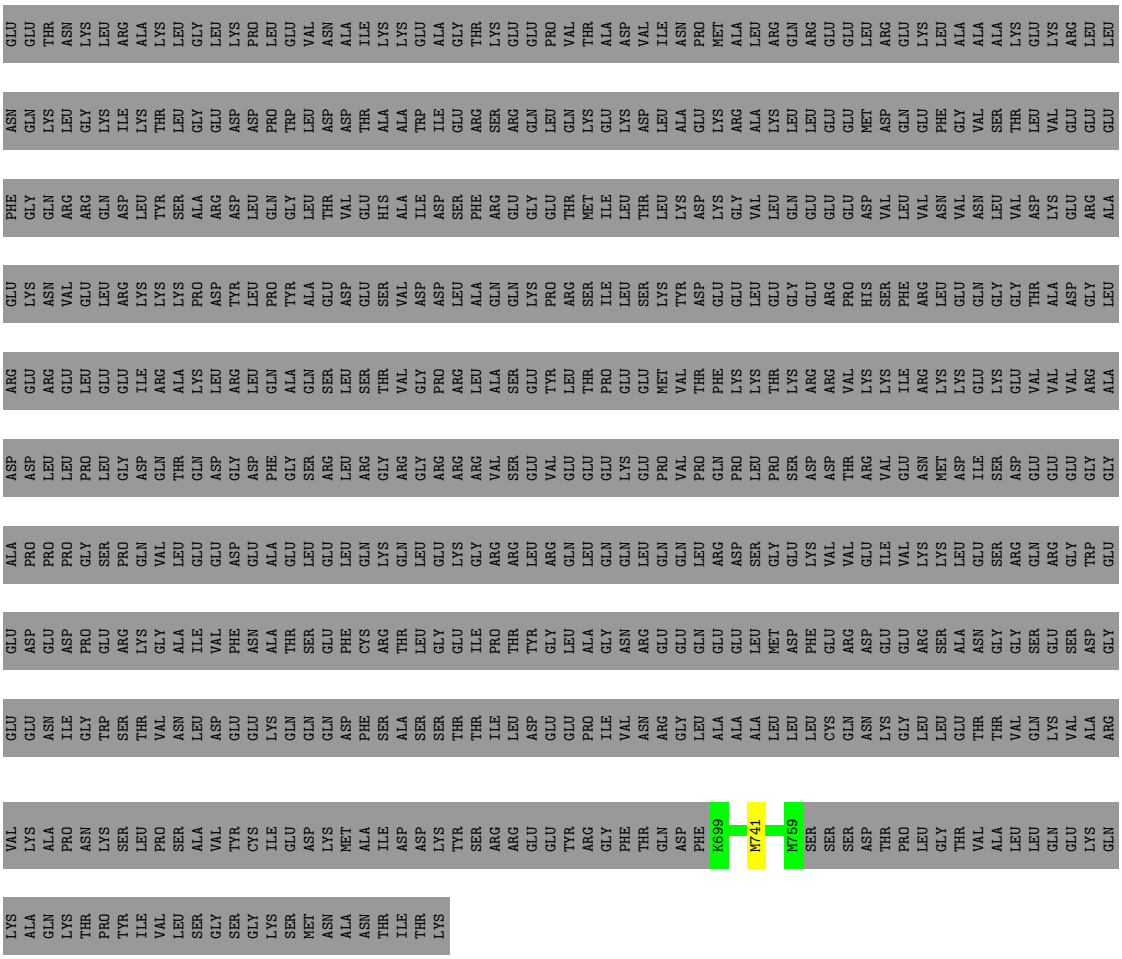
Chain 4R:



• Molecule 30: U4/U6.U5 tri-snRNP-associated protein 1

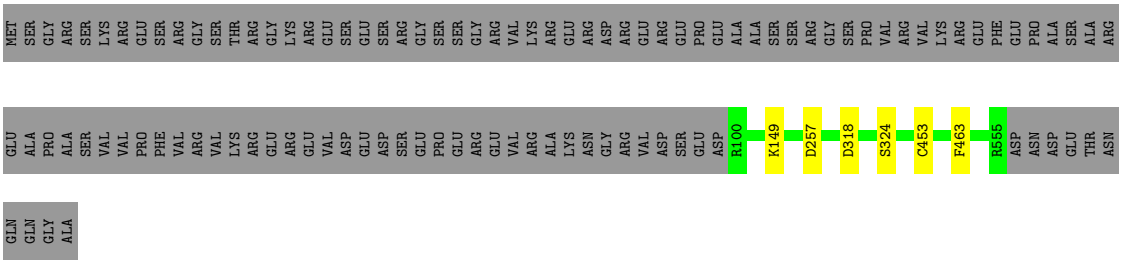
Chain 4S:





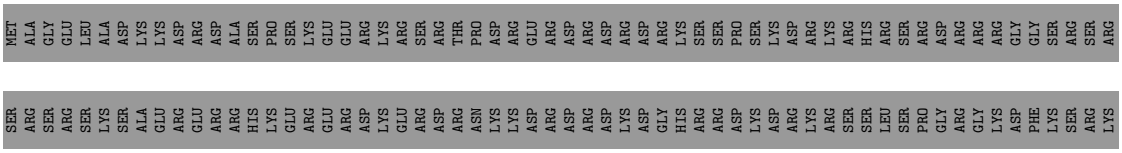
• Molecule 31: U4/U6.U5 tri-snRNP-associated protein 2

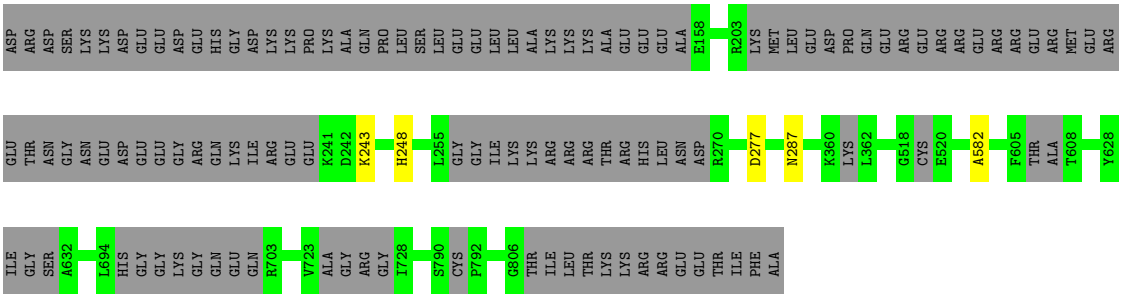
Chain 4T:



• Molecule 32: Probable ATP-dependent RNA helicase DDX23

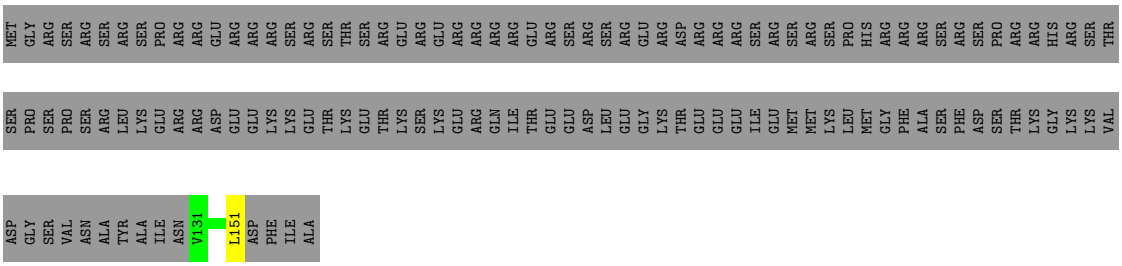
Chain 4U:





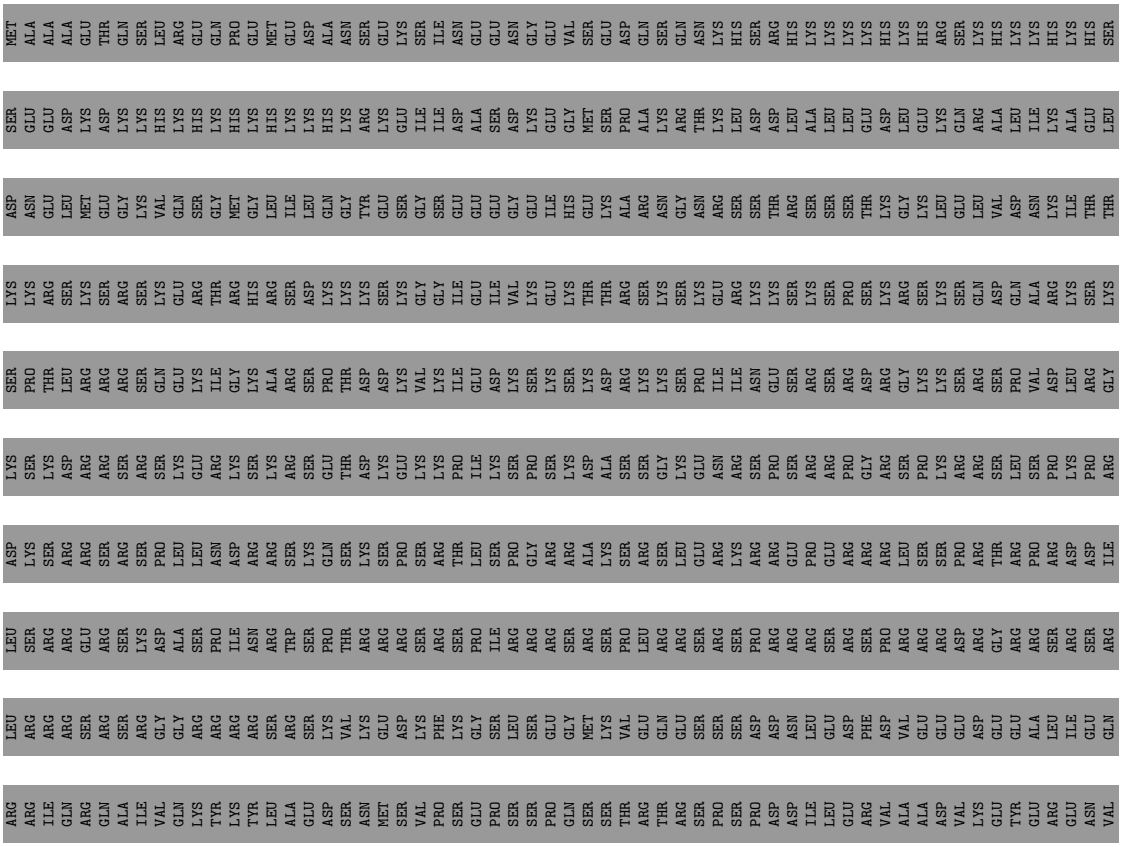
• Molecule 33: U4/U6.U5 small nuclear ribonucleoprotein 27 kDa protein

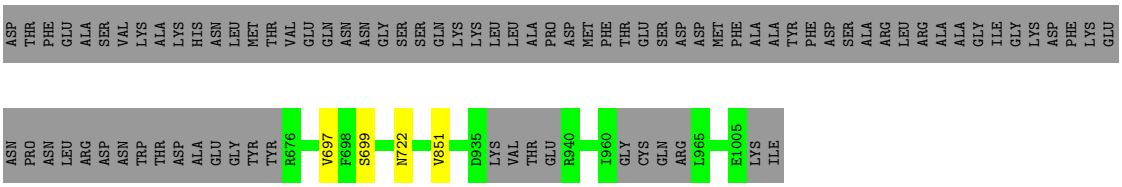
Chain 4X:



• Molecule 34: Serine/threonine-protein kinase PRP4 homolog

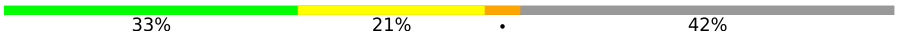
Chain 4Y:

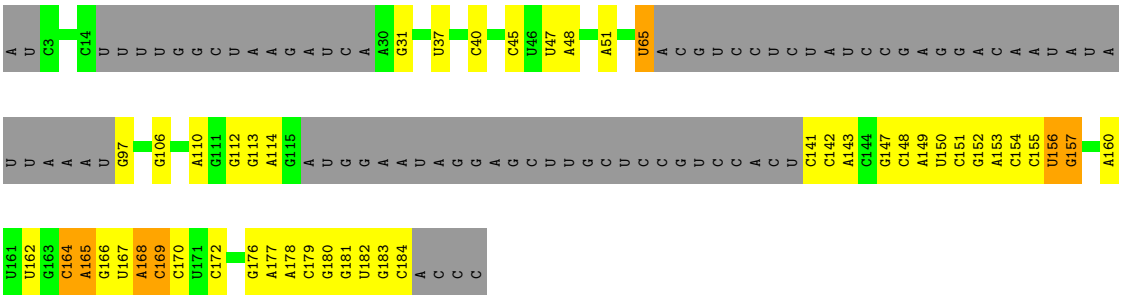




• Molecule 35: U2 snRNA

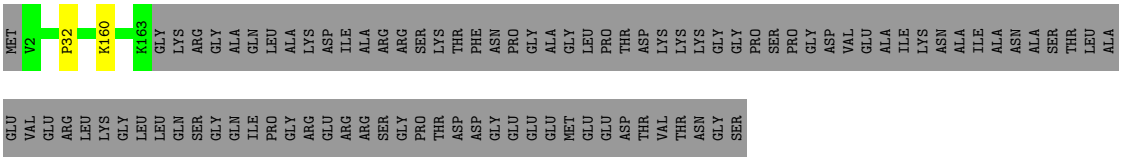
Chain 2A:





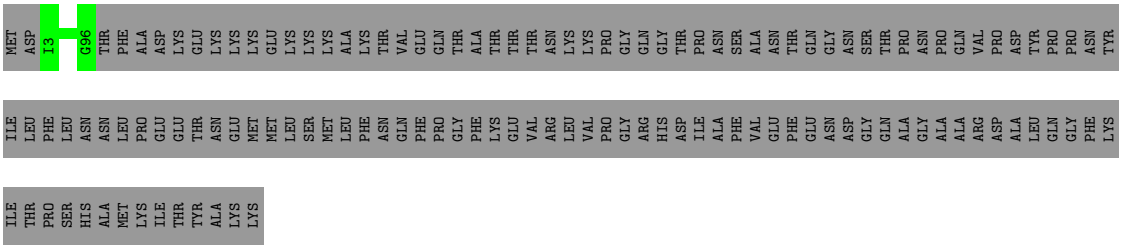
• Molecule 36: U2 small nuclear ribonucleoprotein A'

Chain 2B:



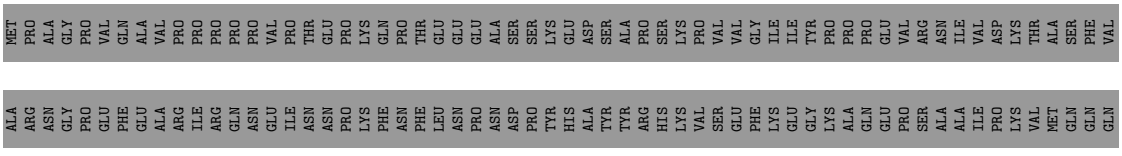
• Molecule 37: U2 small nuclear ribonucleoprotein B"

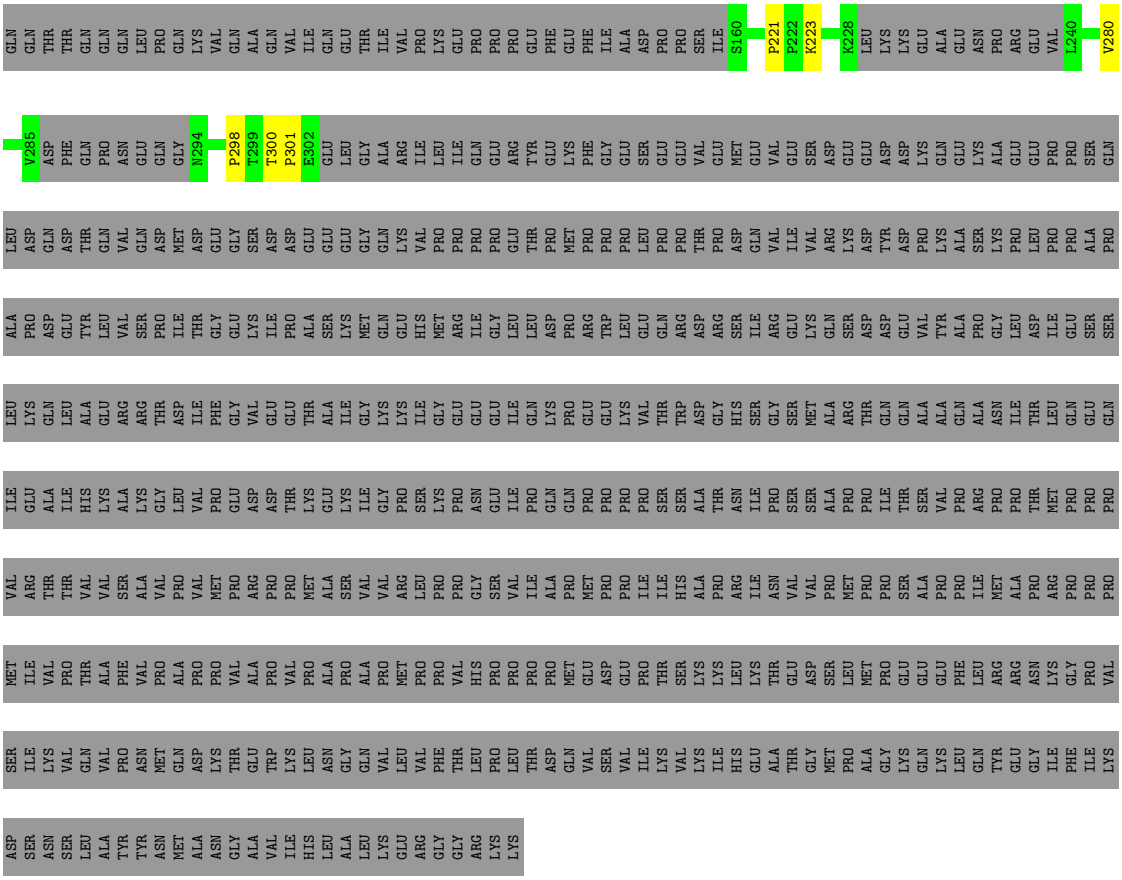
Chain 2C:



• Molecule 38: Splicing factor 3A subunit 1

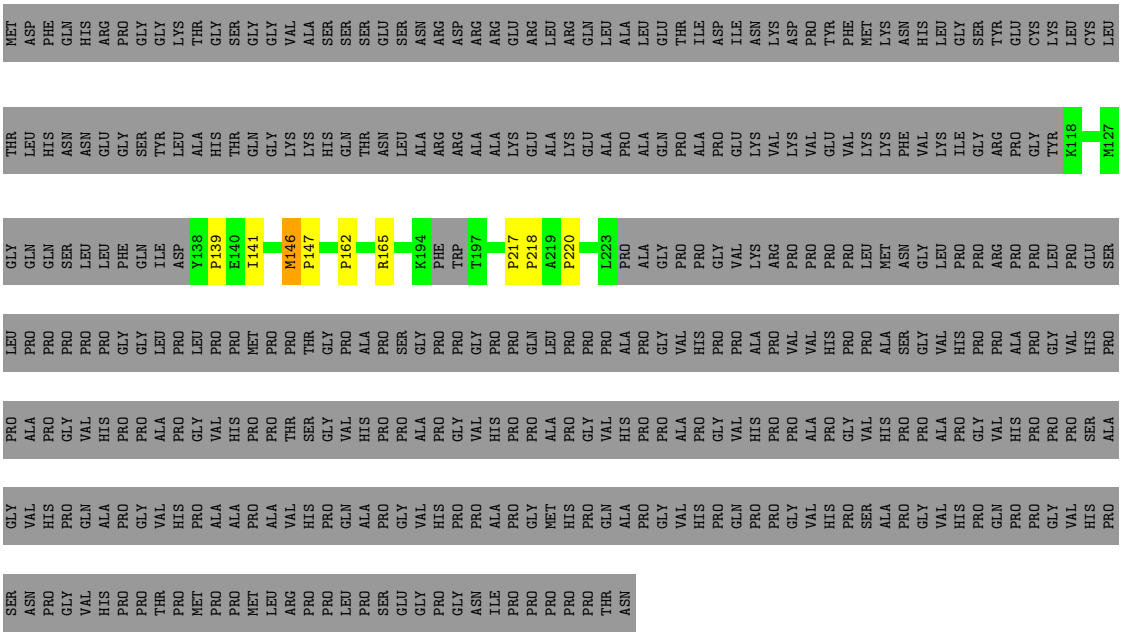
Chain 2D:





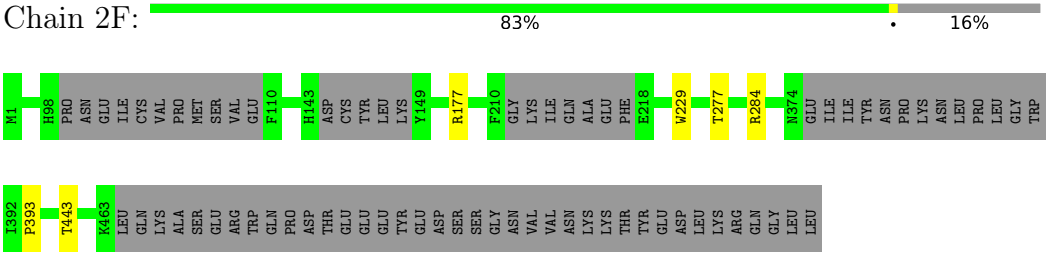
• Molecule 39: Splicing factor 3A subunit 2

Chain 2E:



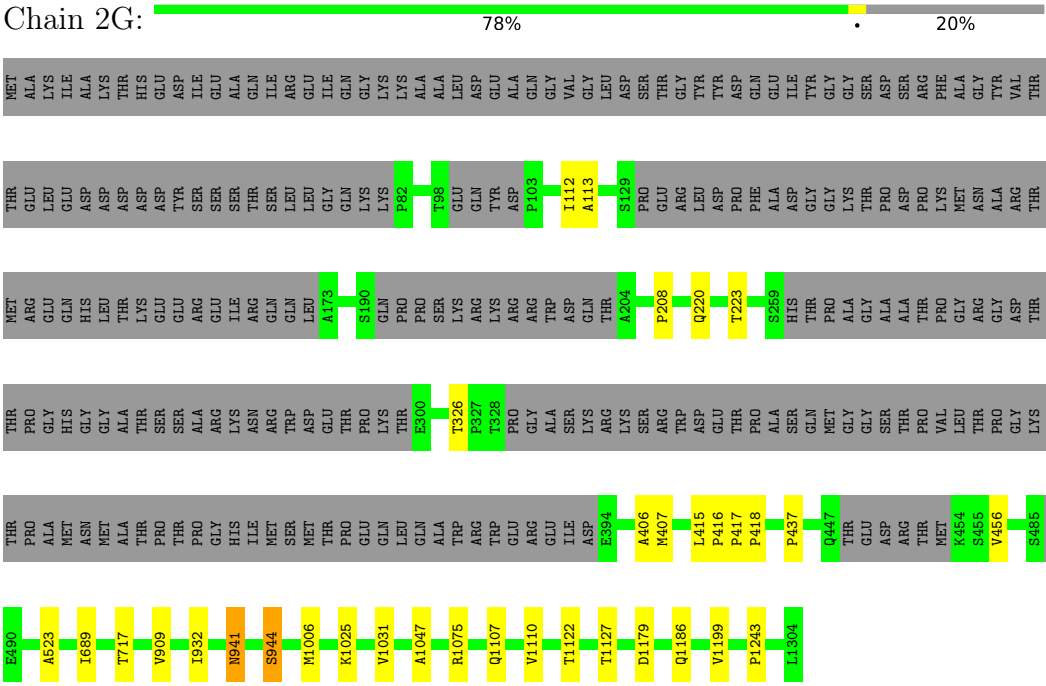
• Molecule 40: Splicing factor 3A subunit 3

Chain 2F:



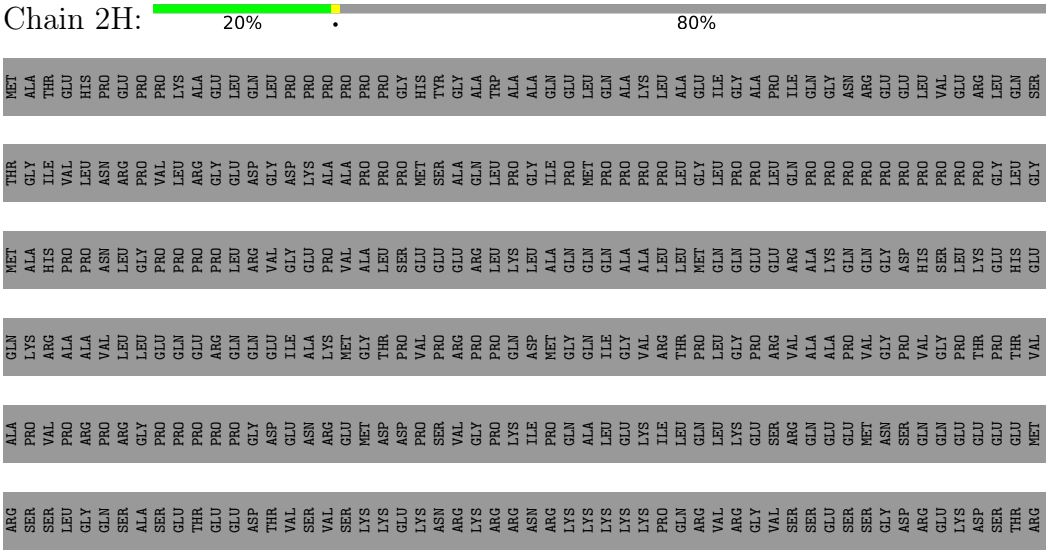
• Molecule 41: Splicing factor 3B subunit 1

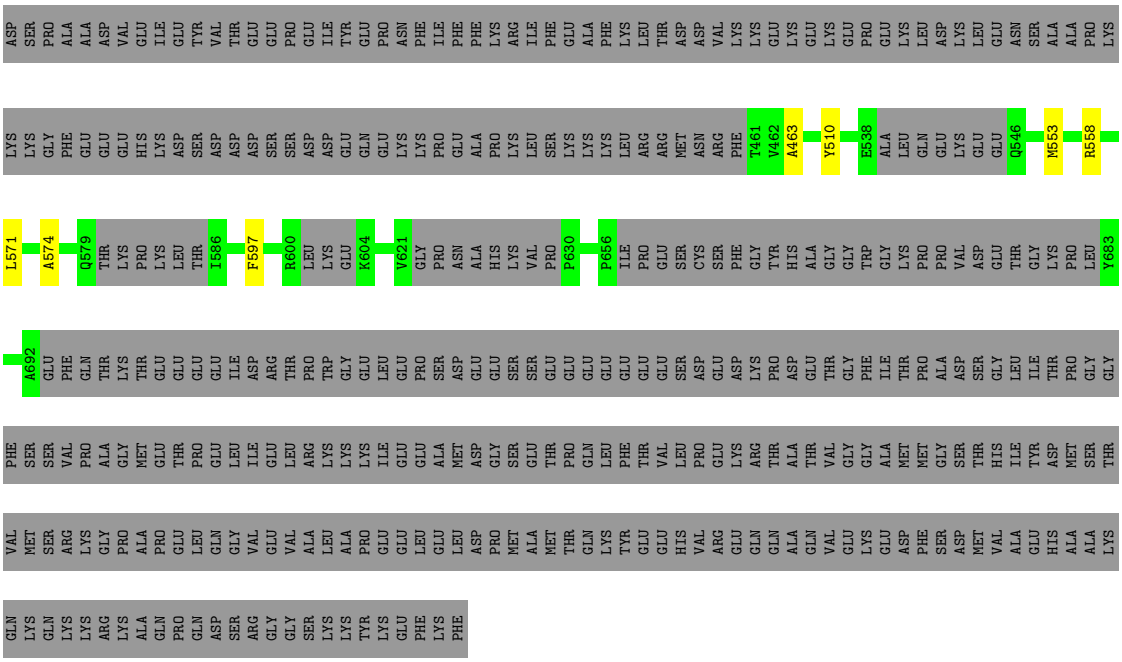
Chain 2G:



• Molecule 42: Splicing factor 3B subunit 2

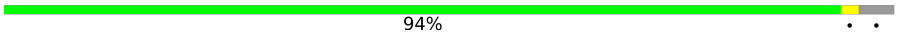
Chain 2H:

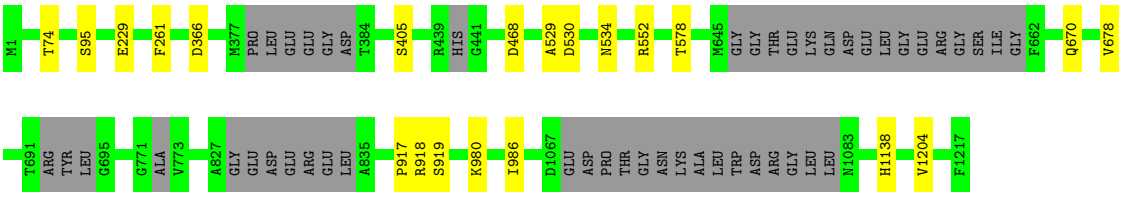




• Molecule 43: Splicing factor 3B subunit 3

Chain 2I:



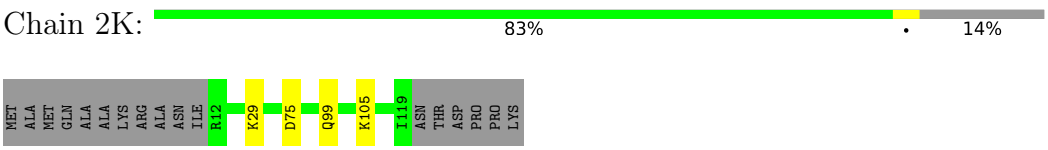


• Molecule 44: Splicing factor 3B subunit 4

Chain 2J:

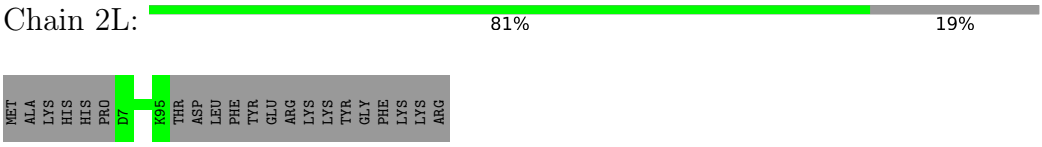
• Molecule 45: Splicing factor 3B subunit 6

Chain 2K:



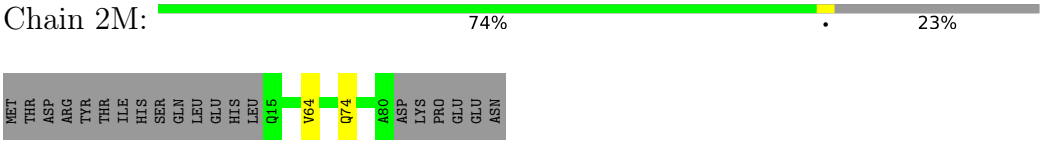
• Molecule 46: PHD finger-like domain-containing protein 5A

Chain 2L:



• Molecule 47: 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	116801	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: MG, GTP, 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	0.73	3/956 (0.3%)	0.92	0/1481
2	6A	0.36	0/1274	0.90	5/1976 (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	0.27	0/2673	0.94	11/4156 (0.3%)
11	5B	0.25	0/18748	0.49	0/25452
12	5C	0.25	0/6879	0.50	0/9344
13	5D	0.25	0/16393	0.49	0/22174
14	5E	0.28	0/1480	0.57	0/2056
15	2a	0.50	0/339	0.69	0/422
15	4a	0.30	0/404	0.60	0/561
15	5a	0.51	0/343	0.69	0/427
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	0.32	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	0.30	0/375	0.66	0/521
19	5e	0.64	0/315	0.74	0/392
20	2f	0.55	0/262	0.63	0/324
20	4f	0.31	0/362	0.65	0/501
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	0.47	2/2963 (0.1%)	0.89	7/4603 (0.2%)
23	4B	0.25	0/948	0.45	0/1312
24	4C	0.24	0/1764	0.48	0/2450
25	4D	0.24	0/1336	0.40	0/1858
26	4E	0.23	0/614	0.43	0/855
27	4F	0.25	0/1198	0.50	0/1620
28	4G	0.29	1/4561 (0.0%)	0.52	5/6271 (0.1%)
29	4R	0.25	0/891	0.55	0/1188
30	4S	0.26	0/512	0.55	0/673
31	4T	0.25	0/3845	0.49	0/5208
32	4U	0.40	0/3422	0.59	0/4659
33	4X	0.27	0/187	0.65	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	0.23	0/493	0.42	0/611
39	2E	0.22	0/373	0.58	1/461 (0.2%)
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	21/98538 (0.0%)	0.65	87/133716 (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
28	4G	0	2
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	37

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	11.49	1.75	1.61
41	2G	406	ALA	C-N	7.96	1.52	1.34
35	2A	142	C	C1'-N1	7.25	1.59	1.48
41	2G	1243	PRO	N-CA	-7.10	1.35	1.47

The worst 5 of 87 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
10	5A	78	U	P-O3'-C3'	-14.15	102.72	119.70
10	5A	58	U	P-O3'-C3'	-12.85	104.29	119.70
35	2A	167	U	C5-C4-O4	11.56	132.84	125.90
10	5A	57	G	P-O3'-C3'	-11.33	106.11	119.70
41	2G	406	ALA	C-N-CA	10.26	147.35	121.70

There are no chirality outliers.

5 of 37 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
22	4A	90	G	Sidechain
24	4C	459	PRO	Peptide
28	4G	569	VAL	Mainchain
28	4G	572	SER	Mainchain
17	5c	112	ASN	Peptide

5.2 Too-close contacts

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 15
4	6b	70/102 (69%)	64 (91%)	3 (4%)	3 (4%)	2 16
5	6c	70/139 (50%)	64 (91%)	5 (7%)	1 (1%)	11 40
6	6d	68/91 (75%)	63 (93%)	4 (6%)	1 (2%)	10 39
7	6e	68/80 (85%)	64 (94%)	2 (3%)	2 (3%)	4 25
8	6f	61/103 (59%)	56 (92%)	5 (8%)	0	100 100
9	6g	57/96 (59%)	52 (91%)	4 (7%)	1 (2%)	8 35
11	5B	2195/2335 (94%)	2099 (96%)	96 (4%)	0	100 100
12	5C	850/972 (87%)	823 (97%)	27 (3%)	0	100 100
13	5D	1989/2136 (93%)	1912 (96%)	77 (4%)	0	100 100
14	5E	299/357 (84%)	280 (94%)	17 (6%)	2 (1%)	22 56
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	84/231 (36%)	82 (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

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
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
19	5e	77/92 (84%)	76 (99%)	1 (1%)	0	100	100
20	2f	62/76 (82%)	60 (97%)	2 (3%)	0	100	100
20	4f	72/76 (95%)	66 (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%)	333 (93%)	24 (7%)	0	100	100
25	4D	262/499 (52%)	252 (96%)	10 (4%)	0	100	100
26	4E	122/128 (95%)	114 (93%)	8 (7%)	0	100	100
27	4F	139/142 (98%)	136 (98%)	3 (2%)	0	100	100
28	4G	776/941 (82%)	739 (95%)	33 (4%)	4 (0%)	29	62
29	4R	104/480 (22%)	93 (89%)	11 (11%)	0	100	100
30	4S	59/800 (7%)	57 (97%)	2 (3%)	0	100	100
31	4T	454/565 (80%)	425 (94%)	29 (6%)	0	100	100
32	4U	559/820 (68%)	541 (97%)	17 (3%)	1 (0%)	47	77
33	4X	19/155 (12%)	19 (100%)	0	0	100	100
34	4Y	316/1007 (31%)	294 (93%)	18 (6%)	4 (1%)	12	41
36	2B	160/255 (63%)	146 (91%)	12 (8%)	2 (1%)	12	41
37	2C	92/225 (41%)	90 (98%)	2 (2%)	0	100	100
38	2D	118/793 (15%)	106 (90%)	6 (5%)	6 (5%)	2	13
39	2E	88/464 (19%)	63 (72%)	16 (18%)	9 (10%)	0	3
40	2F	413/501 (82%)	367 (89%)	41 (10%)	5 (1%)	13	43
41	2G	1032/1304 (79%)	845 (82%)	165 (16%)	22 (2%)	7	32
42	2H	170/895 (19%)	151 (89%)	15 (9%)	4 (2%)	6	28
43	2I	1152/1217 (95%)	1053 (91%)	89 (8%)	10 (1%)	17	50

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
44	2J	76/424 (18%)	75 (99%)	1 (1%)	0	100 100
45	2K	106/125 (85%)	85 (80%)	18 (17%)	3 (3%)	5 25
46	2L	87/110 (79%)	75 (86%)	12 (14%)	0	100 100
47	2M	64/86 (74%)	55 (86%)	8 (12%)	1 (2%)	9 37
All	All	14353/22191 (65%)	13407 (93%)	861 (6%)	85 (1%)	29 59

5 of 85 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	1978/2108 (94%)	1943 (98%)	35 (2%)	59 77
12	5C	754/866 (87%)	738 (98%)	16 (2%)	53 75
13	5D	1779/1908 (93%)	1730 (97%)	49 (3%)	43 69
27	4F	129/130 (99%)	124 (96%)	5 (4%)	32 61
28	4G	185/792 (23%)	172 (93%)	13 (7%)	15 43
29	4R	94/369 (26%)	88 (94%)	6 (6%)	17 47
30	4S	54/681 (8%)	53 (98%)	1 (2%)	57 76
31	4T	418/511 (82%)	412 (99%)	6 (1%)	67 81
32	4U	133/721 (18%)	129 (97%)	4 (3%)	41 67
33	4X	19/144 (13%)	18 (95%)	1 (5%)	22 53
All	All	5543/8230 (67%)	5407 (98%)	136 (2%)	50 71

5 of 136 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
28	4G	251	MET
29	4R	429	PHE
31	4T	463	PHE
12	5C	910	ASP
12	5C	835	GLU

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

Mol	Chain	Res	Type
28	4G	139	GLN
31	4T	362	HIS
31	4T	540	GLN
31	4T	520	HIS
12	5C	903	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	41/144 (28%)	27 (65%)	9 (21%)
10	5A	113/117 (96%)	38 (33%)	6 (5%)
2	6A	50/107 (46%)	8 (16%)	3 (6%)
22	4A	119/144 (82%)	21 (17%)	2 (1%)
35	2A	105/188 (55%)	22 (20%)	3 (2%)
All	All	428/700 (61%)	116 (27%)	23 (5%)

5 of 116 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 23 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
10	5A	78	U
10	5A	96	A

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
10	5A	94	U
22	4A	68	A
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	$\#\|Z\| > 2$
50	GTP	5C	1002	49	26,34,34	1.14	2 (7%)	32,54,54	1.65	6 (18%)
48	IHP	5B	2401	-	36,36,36	0.74	0	54,60,60	1.19	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	-	5/18/38/38	0/3/3/3
48	IHP	5B	2401	-	-	4/30/54/54	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	-4.08	1.39	1.47
50	5C	1002	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})
50	5C	1002	GTP	PA-O3A-PB	-4.22	118.36	132.83
50	5C	1002	GTP	PB-O3B-PG	-4.15	118.59	132.83
50	5C	1002	GTP	C5-C6-N1	3.44	120.02	113.95
48	5B	2401	IHP	C5-C4-C3	3.36	117.76	110.41
50	5C	1002	GTP	C2-N1-C6	-2.97	119.64	125.10

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

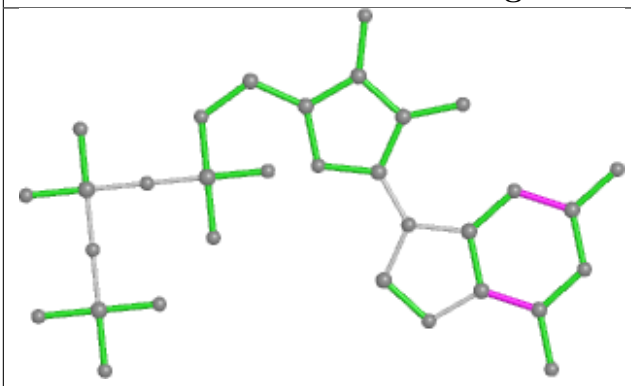
Mol	Chain	Res	Type	Atoms
48	5B	2401	IHP	C1-O11-P1-O41
50	5C	1002	GTP	C5'-O5'-PA-O3A
50	5C	1002	GTP	O4'-C4'-C5'-O5'
50	5C	1002	GTP	C3'-C4'-C5'-O5'
48	5B	2401	IHP	C3-O13-P3-O43

There are no ring outliers.

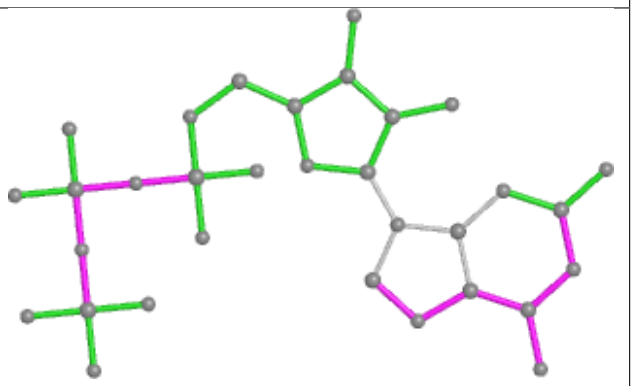
No monomer is involved in short contacts.

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.

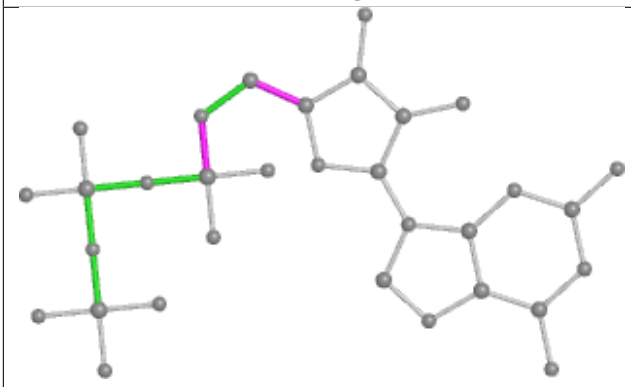
Ligand GTP 5C 1002



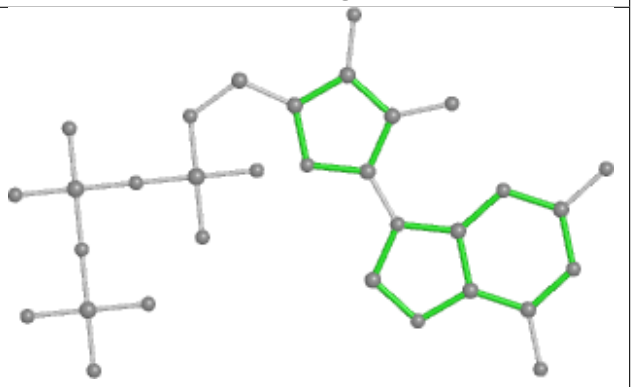
Bond lengths



Bond angles

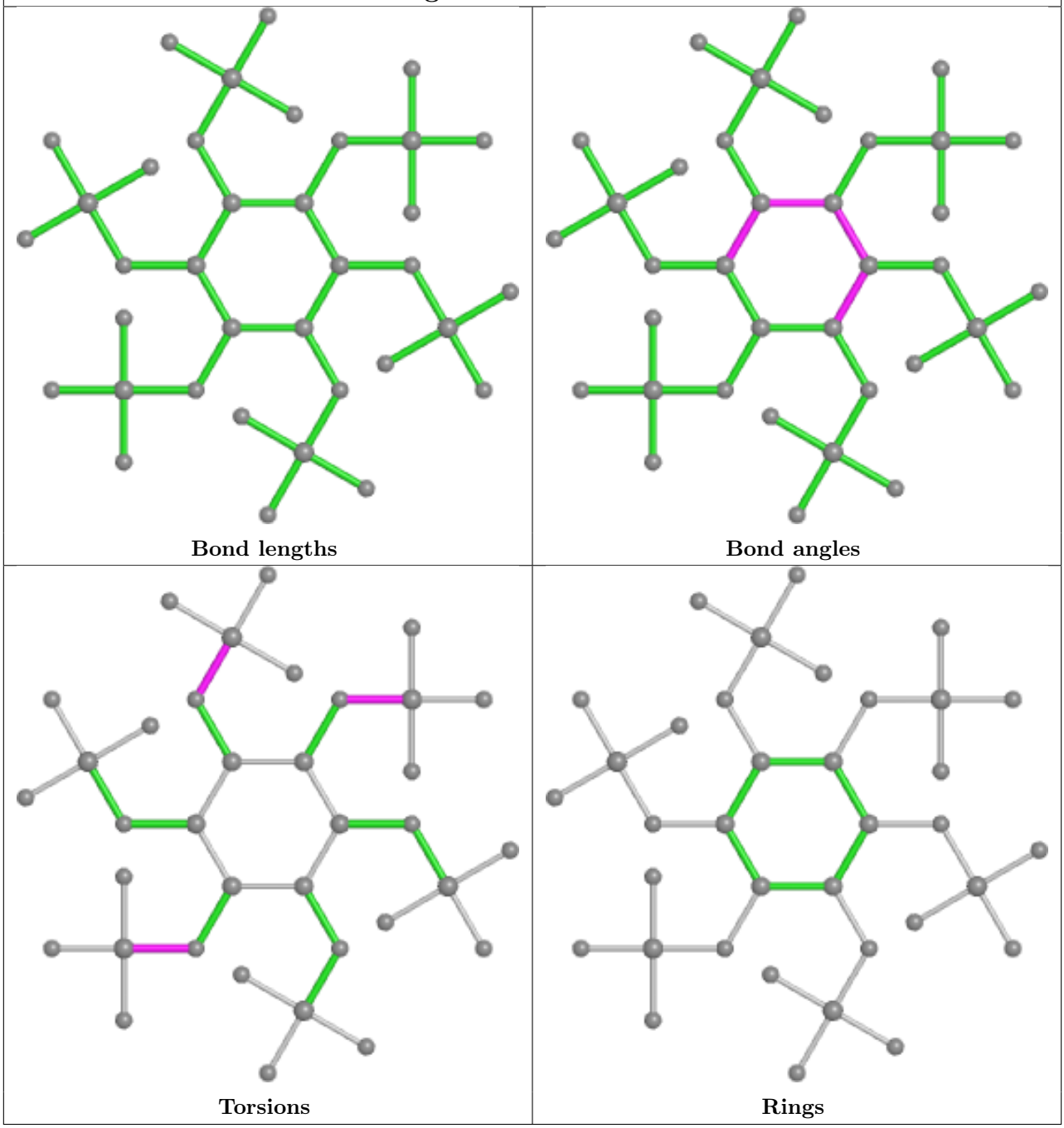


Torsions



Rings

Ligand IHP 5B 2401



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.