



wwPDB EM Validation Summary Report ⓘ

Oct 22, 2024 – 12:37 AM JST

PDB ID : 8H6K
EMDB ID : EMD-34507
Title : Cryo-EM structure of human exon-defined spliceosome in the mature B state.
Authors : Zhang, W.; Zhan, X.; Zhang, X.; Bai, R.; Lei, J.; Yan, C.; Shi, Y.
Deposited on : 2022-10-18
Resolution : 2.70 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:

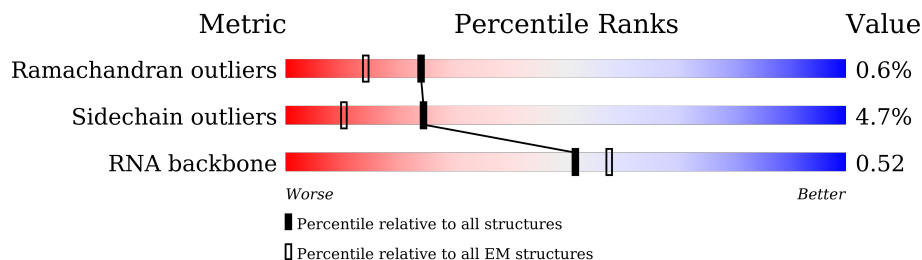
EMDB validation analysis : **FAILED**
Mogul : 1.8.5 (274361), CSD as541be (2020)
MolProbity : 4.02b-467
buster-report : 1.1.7 (2018)
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ : **FAILED**
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.39

1 Overall quality at a glance i

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

The reported resolution of this entry is 2.70 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	207382	16835
Sidechain outliers	206894	16415
RNA backbone	6643	2191

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	144	
2	5A	117	
3	5B	2335	
4	5C	972	
5	5D	2136	
6	5E	357	
7	2a	231	
7	4a	231	
7	5a	231	






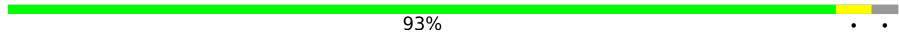
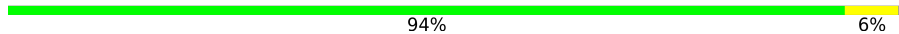

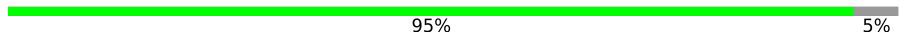















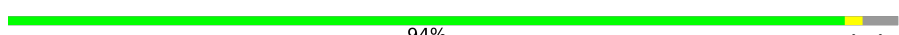
Continued on next page...

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	61%	39%	
14	6A	107	47%	9%	44%
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	 64% 25% 11%
23	4B	683	 36% 63%
24	4C	522	 76% 5% 18%
25	4D	499	 71% 25%
26	4E	128	 93%
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	4K	439	 41% 57%
33	4L	312	 51% 5% 44%
34	4M	73	 92% 8%
35	4N	199	 35% 6% 60%
36	4Z	513	 82% 18%
37	2A	188	 33% 21% 42%
38	2B	255	 63% 36%
39	2C	225	 42% 58%
40	2D	793	 25% 74%
41	2E	464	 18% 80%
42	2F	501	 83% 16%
43	2G	1304	 78% 20%
44	2H	895	 23% 76%
45	2I	1217	 94%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
46	2J	424	 18% 82%
47	2K	125	 83% 14%
48	2L	110	 81% 19%
49	2M	86	 73% 23%

2 Entry composition [i](#)

There are 53 unique types of molecules in this entry. The entry contains 98459 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	64	1334	597	209	464	64	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	18639	11991	3249	3318	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	6430	4108	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	C	N	O	0	0
			336	168	84	84		
7	4a	64	Total	C	N	O	0	0
			256	128	64	64		
7	2a	86	Total	C	N	O	0	0
			344	172	86	86		

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

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

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

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

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

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

- 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	77	Total	C	N	O	0	0
			308	154	77	77		
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	60	Total	C	N	O	P	0	0
			1273	568	235	410	60		

- 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	C	N	O	0	0
			296	148	74	74		

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

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

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

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

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

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

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

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

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

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

- Molecule 22 is a RNA chain called U4 snRNA.

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

- 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	152	Total	C	N	O	S	0	0
			1144	709	204	229	2		

- Molecule 32 is a protein called Microfibrillar-associated protein 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
32	4K	188	Total	C	N	O	S	0	0
			1192	741	219	230	2		

- Molecule 33 is a protein called Pre-mRNA-splicing factor 38A.

Mol	Chain	Residues	Atoms					AltConf	Trace
33	4L	175	Total	C	N	O	S	0	0
			1452	934	244	265	9		

- Molecule 34 is a protein called Ubiquitin-like protein 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
34	4M	73	Total	C	N	O	S	0	0
			599	383	103	109	4		

- Molecule 35 is a protein called Zinc finger matrin-type protein 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
35	4N	80	Total	C	N	O	S	0	0
			640	397	116	120	7		

- Molecule 36 is a protein called WD40 repeat-containing protein SMU1.

Mol	Chain	Residues	Atoms				AltConf	Trace
36	4Z	420	Total	C	N	O	0	0
			2092	1252	420	420		

- Molecule 37 is a RNA chain called U2 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
37	2A	109	2311	1032	396	774	109	0	0

- Molecule 38 is a protein called U2 small nuclear ribonucleoprotein A'.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
38	2B	162	648	324	162	162	0	0

- Molecule 39 is a protein called U2 small nuclear ribonucleoprotein B'.

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

- Molecule 40 is a protein called Splicing factor 3A subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
40	2D	204	1134	644	238	250	2	0	0

- Molecule 41 is a protein called Splicing factor 3A subunit 2.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
41	2E	94	376	188	94	94	0	0

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

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
42	2F	423	1693	847	423	423	0	0

- Molecule 43 is a protein called Splicing factor 3B subunit 1.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
43	2G	1048	4192	2096	1048	1048	0	0

- Molecule 44 is a protein called Splicing factor 3B subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
44	2H	213	Total	C	N	O	S	0	0
			959	510	220	226	3		

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

Mol	Chain	Residues	Atoms				AltConf	Trace
45	2I	1168	Total	C	N	O	0	0
			4672	2336	1168	1168		

- Molecule 46 is a protein called Splicing factor 3B subunit 4.

Mol	Chain	Residues	Atoms				AltConf	Trace
46	2J	78	Total	C	N	O	0	0
			312	156	78	78		

- Molecule 47 is a protein called Splicing factor 3B subunit 6.

Mol	Chain	Residues	Atoms				AltConf	Trace
47	2K	108	Total	C	N	O	0	0
			432	216	108	108		

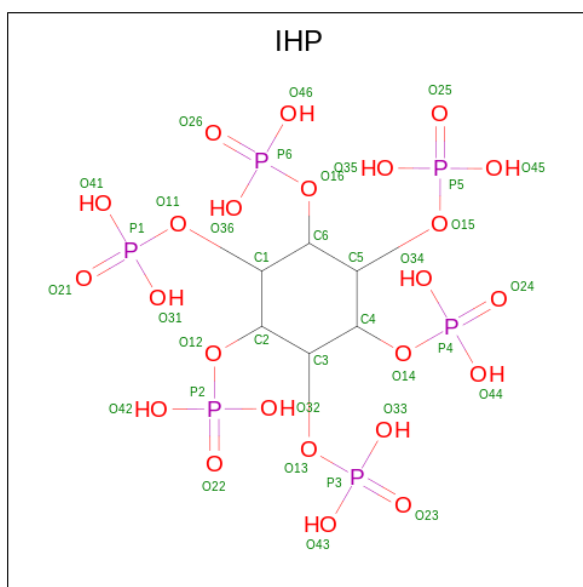
- Molecule 48 is a protein called PHD finger-like domain-containing protein 5A.

Mol	Chain	Residues	Atoms				AltConf	Trace
48	2L	89	Total	C	N	O	0	0
			356	178	89	89		

- Molecule 49 is a protein called Splicing factor 3B subunit 5.

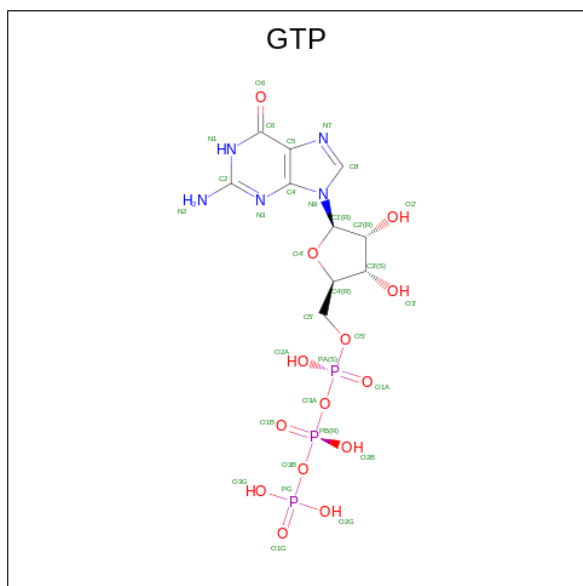
Mol	Chain	Residues	Atoms				AltConf	Trace
49	2M	66	Total	C	N	O	0	0
			264	132	66	66		

- Molecule 50 is INOSITOL HEXAKISPHOSPHATE (three-letter code: IHP) (formula: C₆H₁₈O₂₄P₆).



Mol	Chain	Residues	Atoms				AltConf
			Total	C	O	P	
50	5B	1	36	6	24	6	0

- Molecule 51 is GUANOSINE-5'-TRIPHOSPHATE (three-letter code: GTP) (formula: $C_{10}H_{16}N_5O_{14}P_3$).



Mol	Chain	Residues	Atoms					AltConf
			Total	C	N	O	P	
51	5C	1	32	10	5	14	3	0

- Molecule 52 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
52	5C	1	Total	Mg	0
			1	1	

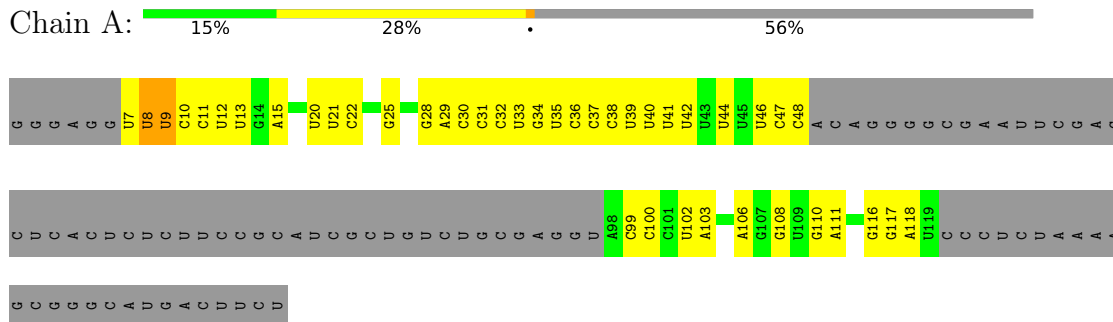
- Molecule 53 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
53	4I	1	Total	Zn	0
			1	1	
53	4N	1	Total	Zn	0
			1	1	

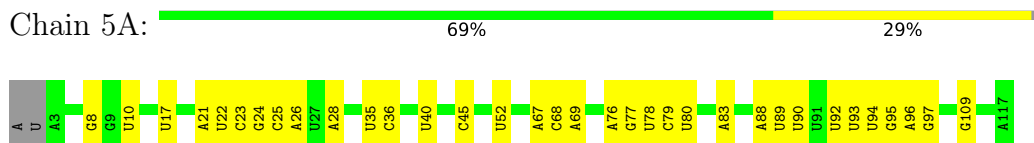
3 Residue-property plots [i](#)

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.

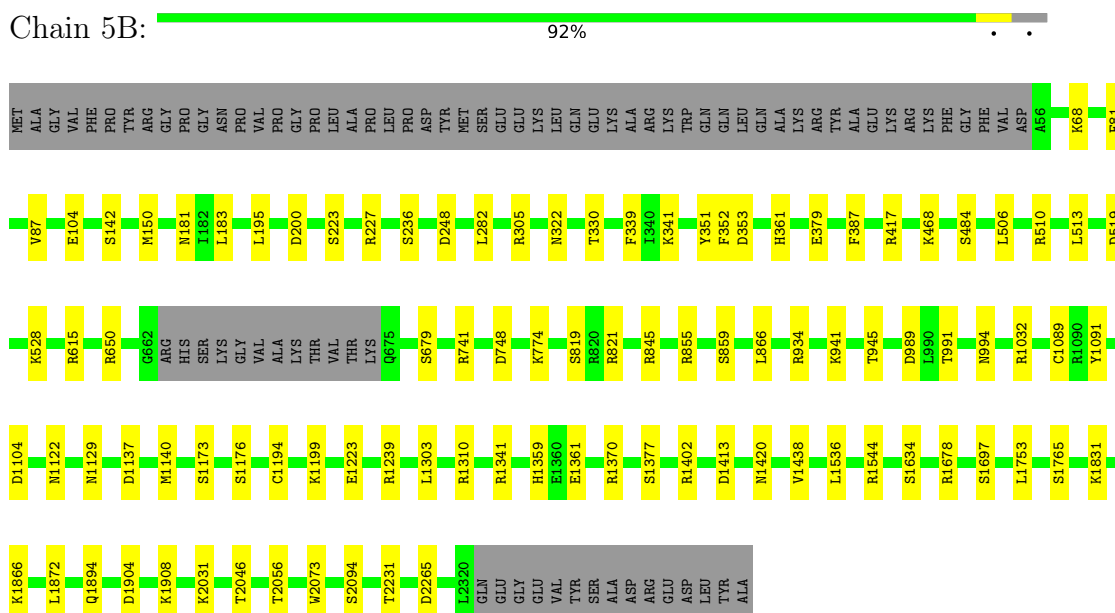
• Molecule 1: pre-mRNA

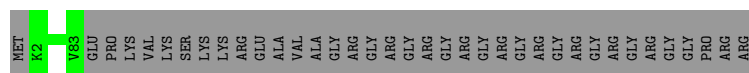


• Molecule 2: U5 snRNA

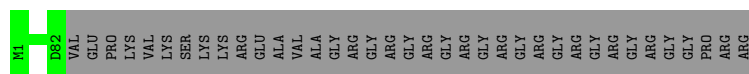


• Molecule 3: Pre-mRNA-processing-splicing factor 8

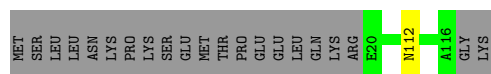
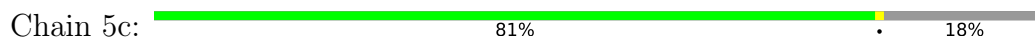




- Molecule 8: Small nuclear ribonucleoprotein Sm D1



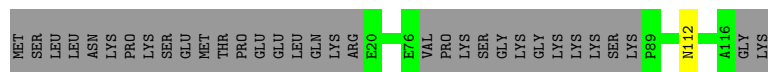
- Molecule 9: Small nuclear ribonucleoprotein Sm D2



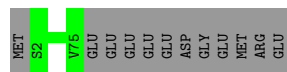
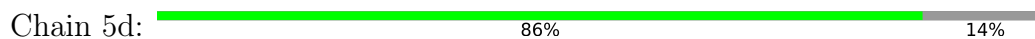
- Molecule 9: Small nuclear ribonucleoprotein Sm D2



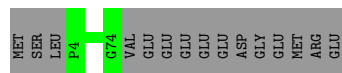
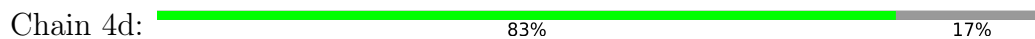
- Molecule 9: Small nuclear ribonucleoprotein Sm D2




- Molecule 10: Small nuclear ribonucleoprotein F

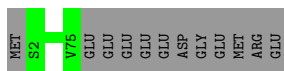


- Molecule 10: Small nuclear ribonucleoprotein F




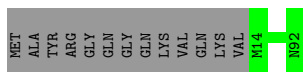
- Molecule 10: Small nuclear ribonucleoprotein F

Chain 2d:  86% 14%




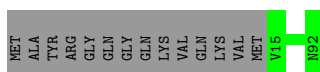
- Molecule 11: Small nuclear ribonucleoprotein E

Chain 5e:  86% 14%




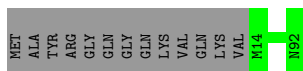
- Molecule 11: Small nuclear ribonucleoprotein E

Chain 4e:  85% 15%



- Molecule 11: Small nuclear ribonucleoprotein E

Chain 2e:  86% 14%



- Molecule 12: Small nuclear ribonucleoprotein G

Chain 5f:  95% 5%




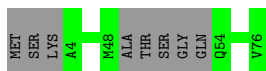
- Molecule 12: Small nuclear ribonucleoprotein G

Chain 4f:  96% .



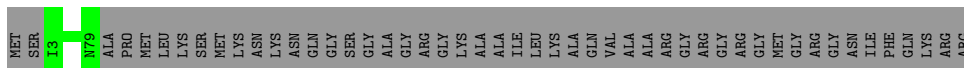
- Molecule 12: Small nuclear ribonucleoprotein G

Chain 2f:  89% 11%



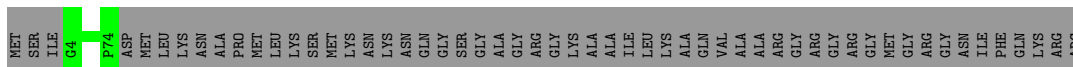
- Molecule 13: Small nuclear ribonucleoprotein Sm D3

Chain 5g:  61% 39%



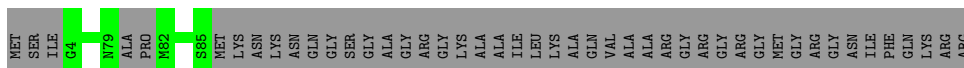
- Molecule 13: Small nuclear ribonucleoprotein Sm D3

Chain 4g:  56% 44%



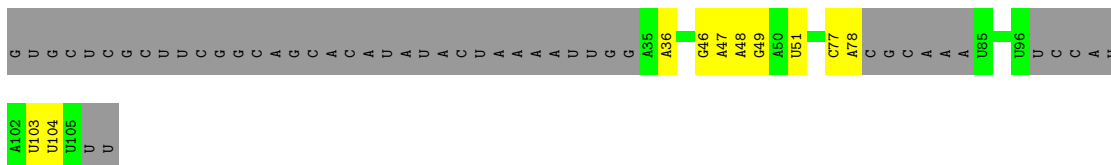
- Molecule 13: Small nuclear ribonucleoprotein Sm D3

Chain 2g:  63% 37%



- Molecule 14: U6 snRNA

Chain 6A:  47% 9% 44%



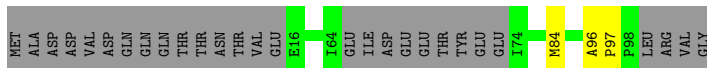
- Molecule 15: U6 snRNA-associated Sm-like protein LSM2

Chain 6a:  91% 5%



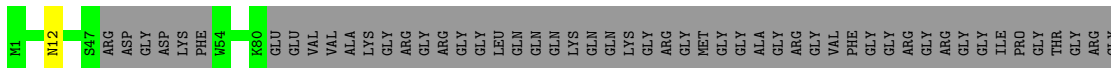
- Molecule 16: U6 snRNA-associated Sm-like protein LSM3

Chain 6b:  70% 27%



- Molecule 17: U6 snRNA-associated Sm-like protein LSM4

Chain 6c:  53% 47%



GLN
PRO
GLU
LYS
LYS
PRO
GLY
GLN
ALA
GLY
LYS
GLN

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

Chain 6d: 78% 21%

MET
ALA
ALA
ASN
ALA
THR
THR
ASN
PRO
S10
I59
THR
PRO
GLU
GLY
R64
D70
G85
GLY
GLY
PRO
GLU
VAL

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

Chain 6e: 85% 12%

MET
SER
LEU
ARG
LYS
GLN
T77
V52
Q55
G76
LYS
ARG
MET

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

Chain 6f: 63% 37%

MET
ALA
ASP
LYS
GLU
LYS
LYS
S11
R56
ARG
ASP
PRO
ASP
ASP
GLN
TYR
LYS
LEU
THR
GLU
ASP
R68
Q87
ASP
GLY
MET
GLU
ALA
ILE
PRO
ASN
PRO
PHE
ILE
GLN
GLN
GLN
ASP
ALA

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

Chain 6g: 63% 36%

MET
THR
A4
I34
E43
ARG
SER
SER
GLN
GLY
VAL
E53
I73
ASP
GLU
GLU
THR
THR
ASP
SER
SER
ALA
LEU
LEU
ASP
GLY
ASN
ILE
ARG
ALA
GLU
PRO
LEU
ASN
VAL
VAL
ALA
HIS

- Molecule 22: U4 snRNA

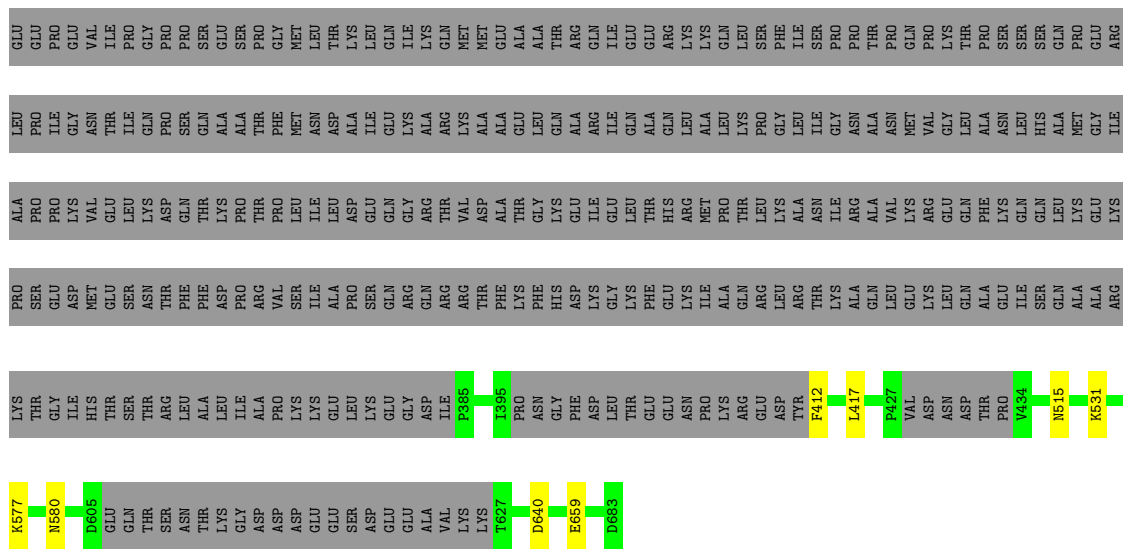
Chain 4A: 64% 25% 11%

M790
A1
G2
A25
C37
U38
A39
U40
G45
U53
A54
U62
U63
G
A
A
A
A
A
C
U70
U71
U72
U73
C74
C75
C76
A
A
U
A
A
C
C
C83
C84
G85
G90
C99
A100
A103
G109
U114
G115
A118
A119
U120
U121
U124
G125
A126
C127
U
A
C
G
G133
G138
G144

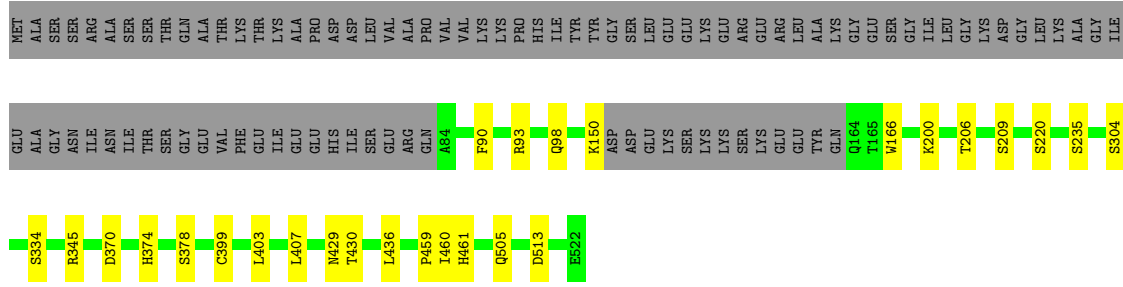
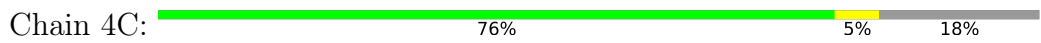
- Molecule 23: U4/U6 small nuclear ribonucleoprotein Prp3

Chain 4B: 36% 63%

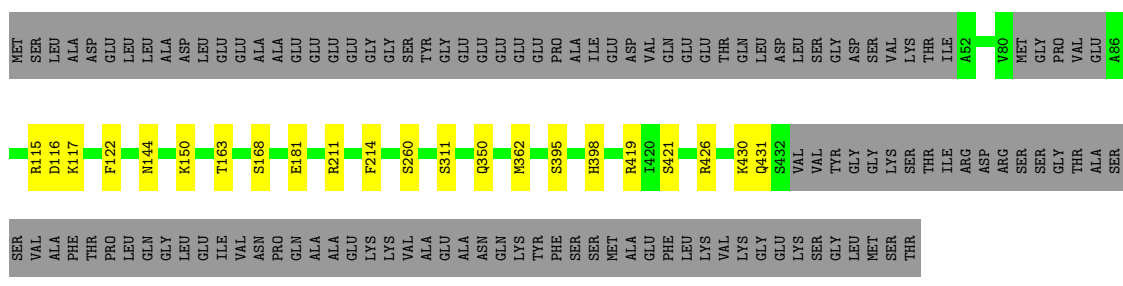
MET
ALA
LEU
SER
LYS
ARG
GLU
LEU
LEU
ASP
ALA
VAL
GLU
LEU
GLY
PRO
TRP
ILE
SER
SER
ARG
HIS
THR
SER
LYS
SER
ARG
SER
VAL
SER
SER
LEU
SER
ASP
PHE
SER
SER
ARG
GLU
PRO
LYS
VAL
VAL
THR
ALA
ALA
PHE
LEU
ASN
GLY
ASP
VAL
SER
GLY
MET
ILE
LYS
GLU
SER
SER
GLY
VAL
VAL
LYS
HIS
ARG
LEU
LEU
PRO
PHE
ARG
PHE
ASP
GLU
SER
VAL
THR
LEU



• Molecule 24: U4/U6 small nuclear ribonucleoprotein Prp4



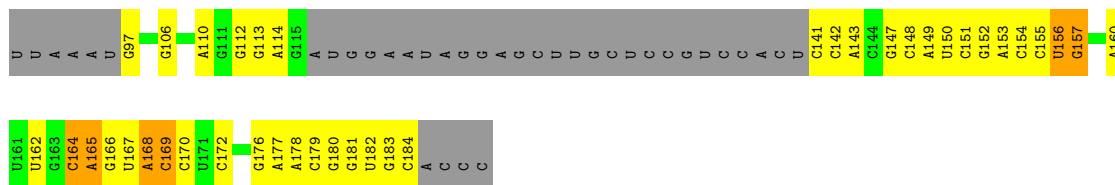
• Molecule 25: U4/U6 small nuclear ribonucleoprotein Prp31



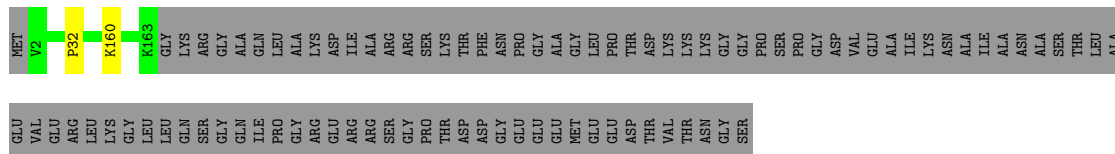
• Molecule 26: NHP2-like protein 1



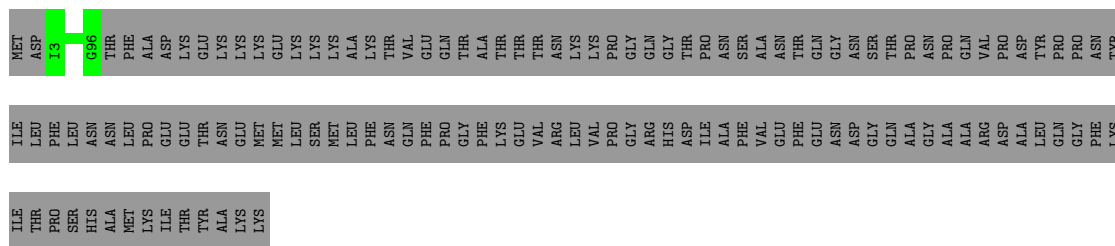
• Molecule 27: Thioredoxin-like protein 4A



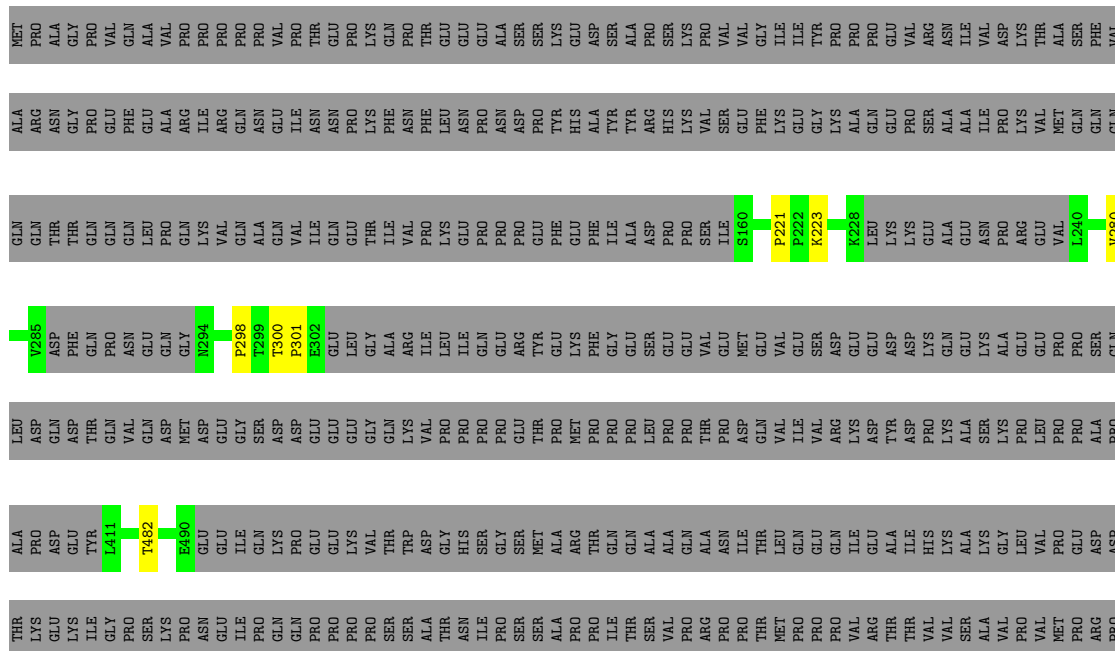
• Molecule 38: U2 small nuclear ribonucleoprotein A'

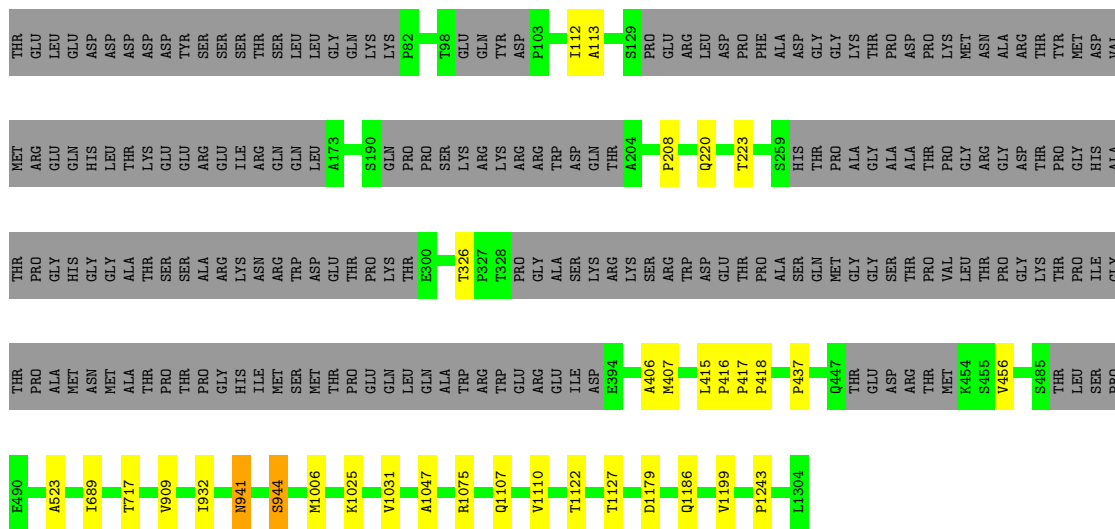


• Molecule 39: U2 small nuclear ribonucleoprotein B'

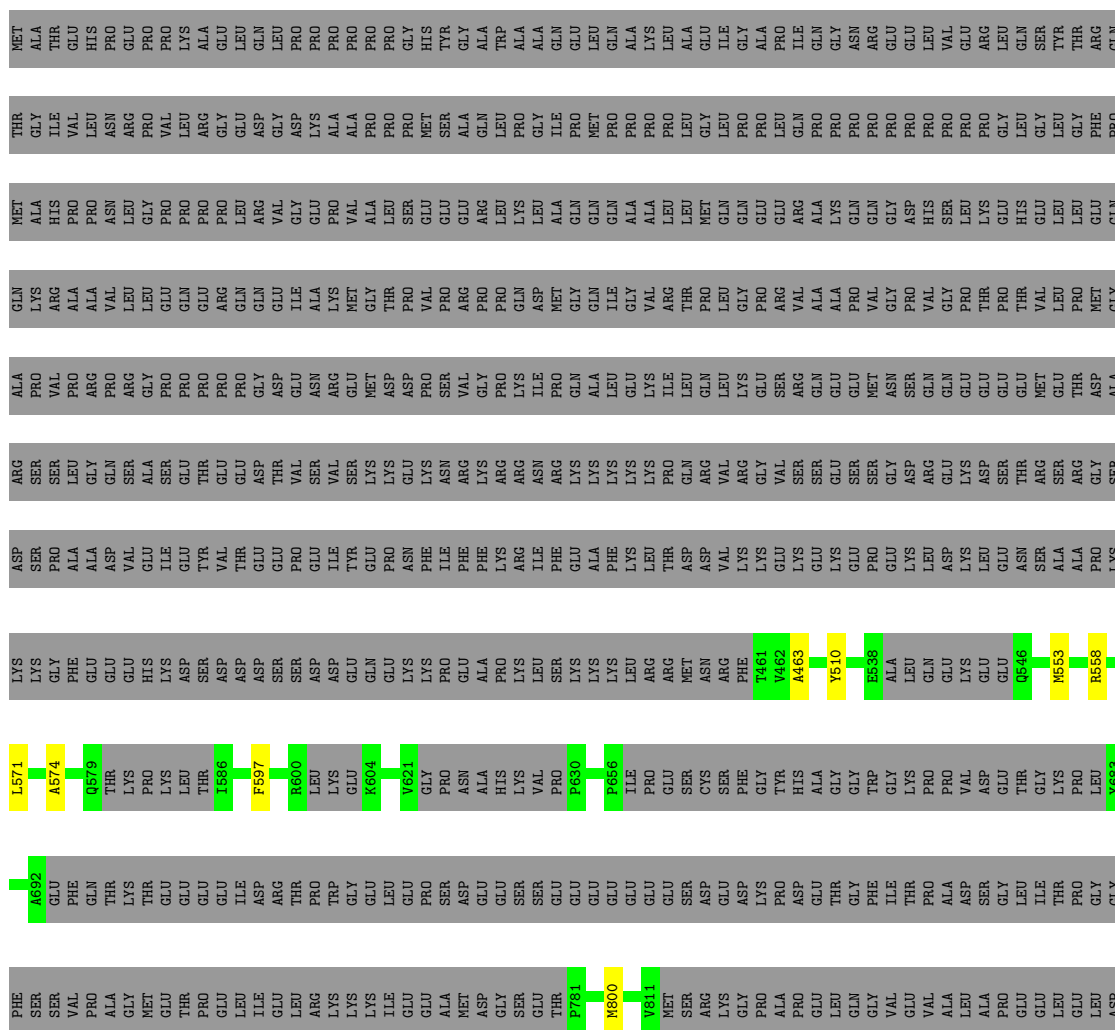


• Molecule 40: Splicing factor 3A subunit 1





● Molecule 44: Splicing factor 3B subunit 2



MET	THR	ASP	ARG	TYR	THR	ILE	HIS	SER	GLN	LEU	GLU	HIS	LEU	Q15	A56	V64	Q74	A80	ASP	LYS	PRO	GLU	GLU	ASN
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	414060	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	50	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2500	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor

5 Model quality i

5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG, IHP, M7M, GTP

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	3/1481 (0.2%)	0.92	0/2297
2	5A	0.29	0/2698	0.86	3/4195 (0.1%)
3	5B	0.26	0/19154	0.51	1/26000 (0.0%)
4	5C	0.26	0/6573	0.52	1/8929 (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	0/307	0.55	0/382
14	6A	0.24	0/1423	0.77	0/2211
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	0.24	0/3025	0.76	2/4702 (0.0%)
23	4B	0.25	0/2114	0.50	0/2836
24	4C	0.25	0/3452	0.52	0/4675
25	4D	0.24	0/2912	0.50	0/3924
26	4E	0.24	0/974	0.47	0/1316
27	4F	0.29	0/1198	0.54	1/1620 (0.1%)
28	4G	0.24	0/5592	0.49	1/7615 (0.0%)
29	4H	0.26	0/853	0.46	0/1188
30	4I	0.25	0/502	0.44	0/683
31	4J	0.25	0/1149	0.52	0/1542
32	4K	0.24	0/1209	0.43	0/1655
33	4L	0.25	0/1481	0.52	0/1995
34	4M	0.23	0/609	0.49	0/819
35	4N	0.27	0/646	0.52	0/859
36	4Z	0.24	0/2100	0.45	0/2926
37	2A	0.86	11/2576 (0.4%)	1.43	55/4003 (1.4%)
38	2B	0.63	0/647	1.42	0/807
39	2C	0.61	0/375	1.20	0/467
40	2D	0.23	0/1139	0.49	0/1477
41	2E	0.22	0/373	0.58	1/461 (0.2%)
42	2F	0.25	0/1688	0.47	0/2102
43	2G	1.04	4/4184 (0.1%)	0.83	2/5216 (0.0%)
44	2H	0.65	0/957	0.67	0/1209
45	2I	0.85	0/4664	0.76	0/5816
46	2J	0.62	0/311	0.64	0/387
47	2K	0.79	0/431	0.79	0/537
48	2L	0.74	0/355	0.68	0/442
49	2M	1.01	0/263	0.77	0/327
All	All	0.44	18/100770 (0.0%)	0.64	68/136218 (0.0%)

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

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
42	2F	0	1
43	2G	0	11
44	2H	0	3
45	2I	0	11
47	2K	0	1
49	2M	0	1
All	All	0	31

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
43	2G	407	MET	N-CA	12.36	1.71	1.46
43	2G	406	ALA	C-N	7.94	1.52	1.34
37	2A	142	C	C1'-N1	7.32	1.59	1.48
43	2G	1243	PRO	N-CA	-7.11	1.35	1.47
37	2A	182	U	C1'-N1	6.94	1.59	1.48

The worst 5 of 68 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
37	2A	167	U	C5-C4-O4	11.61	132.87	125.90
43	2G	406	ALA	C-N-CA	10.28	147.39	121.70
37	2A	164	C	N1-C2-O2	-10.12	112.83	118.90
3	5B	1194	CYS	CA-CB-SG	9.09	130.36	114.00
37	2A	162	U	N3-C2-O2	-8.97	115.92	122.20

There are no chirality outliers.

5 of 31 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
42	2F	443	THR	Peptide
43	2G	220	GLN	Peptide
43	2G	415	LEU	Mainchain
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

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	5B	2249/2335 (96%)	2156 (96%)	93 (4%)	0	100	100
4	5C	814/972 (84%)	755 (93%)	58 (7%)	1 (0%)	48	73
5	5D	1694/2136 (79%)	1618 (96%)	75 (4%)	1 (0%)	48	73
6	5E	297/357 (83%)	272 (92%)	16 (5%)	9 (3%)	3	9
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	75/126 (60%)	73 (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	4
16	6b	70/102 (69%)	64 (91%)	3 (4%)	3 (4%)	2	4
17	6c	70/139 (50%)	63 (90%)	6 (9%)	1 (1%)	9	24
18	6d	68/91 (75%)	63 (93%)	4 (6%)	1 (2%)	8	22
19	6e	68/80 (85%)	64 (94%)	2 (3%)	2 (3%)	3	9
20	6f	61/103 (59%)	56 (92%)	5 (8%)	0	100	100
21	6g	57/96 (59%)	52 (91%)	4 (7%)	1 (2%)	7	18
23	4B	248/683 (36%)	235 (95%)	13 (5%)	0	100	100
24	4C	422/522 (81%)	393 (93%)	28 (7%)	1 (0%)	44	68
25	4D	372/499 (74%)	353 (95%)	19 (5%)	0	100	100
26	4E	122/128 (95%)	115 (94%)	7 (6%)	0	100	100
27	4F	139/142 (98%)	134 (96%)	5 (4%)	0	100	100
28	4G	795/941 (84%)	747 (94%)	48 (6%)	0	100	100
29	4H	167/177 (94%)	158 (95%)	9 (5%)	0	100	100
30	4I	73/376 (19%)	73 (100%)	0	0	100	100
31	4J	142/800 (18%)	136 (96%)	6 (4%)	0	100	100
32	4K	184/439 (42%)	174 (95%)	10 (5%)	0	100	100
33	4L	173/312 (55%)	165 (95%)	6 (4%)	2 (1%)	11	28
34	4M	71/73 (97%)	66 (93%)	5 (7%)	0	100	100
35	4N	78/199 (39%)	77 (99%)	0	1 (1%)	10	26
36	4Z	414/513 (81%)	401 (97%)	12 (3%)	1 (0%)	44	68
38	2B	160/255 (63%)	146 (91%)	12 (8%)	2 (1%)	10	26
39	2C	92/225 (41%)	90 (98%)	2 (2%)	0	100	100
40	2D	196/793 (25%)	180 (92%)	10 (5%)	6 (3%)	3	8
41	2E	88/464 (19%)	63 (72%)	16 (18%)	9 (10%)	0	0
42	2F	413/501 (82%)	367 (89%)	41 (10%)	5 (1%)	11	28
43	2G	1032/1304 (79%)	844 (82%)	166 (16%)	22 (2%)	5	15
44	2H	199/895 (22%)	179 (90%)	16 (8%)	4 (2%)	6	16
45	2I	1152/1217 (95%)	1053 (91%)	89 (8%)	10 (1%)	14	35
46	2J	76/424 (18%)	75 (99%)	1 (1%)	0	100	100
47	2K	106/125 (85%)	85 (80%)	18 (17%)	3 (3%)	4	10

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
48	2L	87/110 (79%)	74 (85%)	13 (15%)	0	100	100
49	2M	64/86 (74%)	55 (86%)	7 (11%)	2 (3%)	3	8
All	All	14181/21253 (67%)	13204 (93%)	886 (6%)	91 (1%)	24	45

5 of 91 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	2033/2108 (96%)	1938 (95%)	95 (5%)	22	49
4	5C	717/866 (83%)	678 (95%)	39 (5%)	18	42
5	5D	1517/1908 (80%)	1492 (98%)	25 (2%)	58	82
23	4B	225/599 (38%)	217 (96%)	8 (4%)	30	59
24	4C	362/442 (82%)	336 (93%)	26 (7%)	12	30
25	4D	299/424 (70%)	277 (93%)	22 (7%)	11	28
26	4E	108/111 (97%)	103 (95%)	5 (5%)	23	49
27	4F	129/130 (99%)	122 (95%)	7 (5%)	18	42
28	4G	417/792 (53%)	389 (93%)	28 (7%)	13	33
29	4H	10/148 (7%)	10 (100%)	0	100	100
30	4I	32/333 (10%)	27 (84%)	5 (16%)	2	6
31	4J	112/681 (16%)	104 (93%)	8 (7%)	12	30
32	4K	66/395 (17%)	60 (91%)	6 (9%)	7	19
33	4L	159/293 (54%)	146 (92%)	13 (8%)	9	23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
34	4M	66/66 (100%)	60 (91%)	6 (9%)	7	19
35	4N	75/181 (41%)	65 (87%)	10 (13%)	3	8
36	4Z	11/450 (2%)	11 (100%)	0	100	100
40	2D	70/709 (10%)	69 (99%)	1 (1%)	62	84
44	2H	26/776 (3%)	25 (96%)	1 (4%)	28	56
All	All	6434/11412 (56%)	6129 (95%)	305 (5%)	24	49

5 of 305 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
28	4G	110	LYS
33	4L	172	GLU
28	4G	356	ASP
31	4J	149	GLU
35	4N	102	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
26	4E	17	HIS
33	4L	81	GLN
26	4E	28	GLN
28	4G	741	HIS
3	5B	1944	HIS

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	62/144 (43%)	37 (59%)	12 (19%)
14	6A	56/107 (52%)	10 (17%)	1 (1%)
2	5A	114/117 (97%)	30 (26%)	5 (4%)
22	4A	124/145 (85%)	33 (26%)	2 (1%)
37	2A	105/188 (55%)	22 (20%)	3 (2%)
All	All	461/701 (65%)	132 (28%)	23 (4%)

5 of 132 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
2	5A	79	C
14	6A	77	C
2	5A	96	A
22	4A	99	C
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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 5 ligands modelled in this entry, 3 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	IHP	5B	3000	-	36,36,36	0.76	0	54,60,60	1.13	2 (3%)
51	GTP	5C	1500	52	26,34,34	1.11	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
50	IHP	5B	3000	-	-	3/30/54/54	0/1/1/1
51	GTP	5C	1500	52	-	8/18/38/38	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
51	5C	1500	GTP	C5-C6	-3.96	1.39	1.47
51	5C	1500	GTP	C2-N3	2.20	1.38	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
51	5C	1500	GTP	C5-C6-N1	3.27	119.73	113.95
51	5C	1500	GTP	PB-O3B-PG	-3.15	122.02	132.83
51	5C	1500	GTP	PA-O3A-PB	-3.11	122.14	132.83
51	5C	1500	GTP	C8-N7-C5	3.09	108.87	102.99
51	5C	1500	GTP	C2-N1-C6	-2.91	119.73	125.10

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

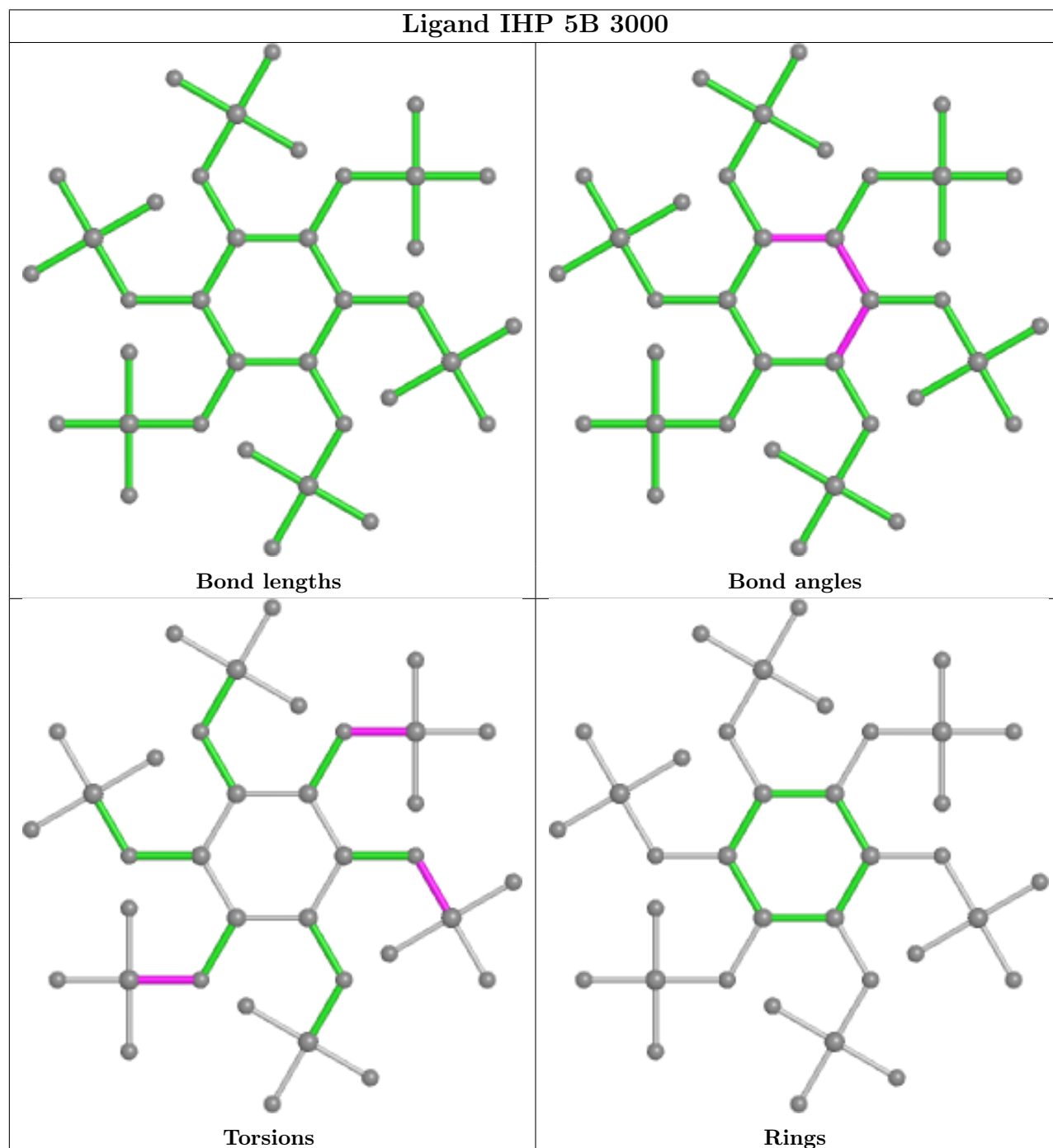
Mol	Chain	Res	Type	Atoms
50	5B	3000	IHP	C4-O14-P4-O44
51	5C	1500	GTP	C5'-O5'-PA-O1A
51	5C	1500	GTP	C5'-O5'-PA-O2A
51	5C	1500	GTP	PB-O3A-PA-O5'
51	5C	1500	GTP	C4'-C5'-O5'-PA

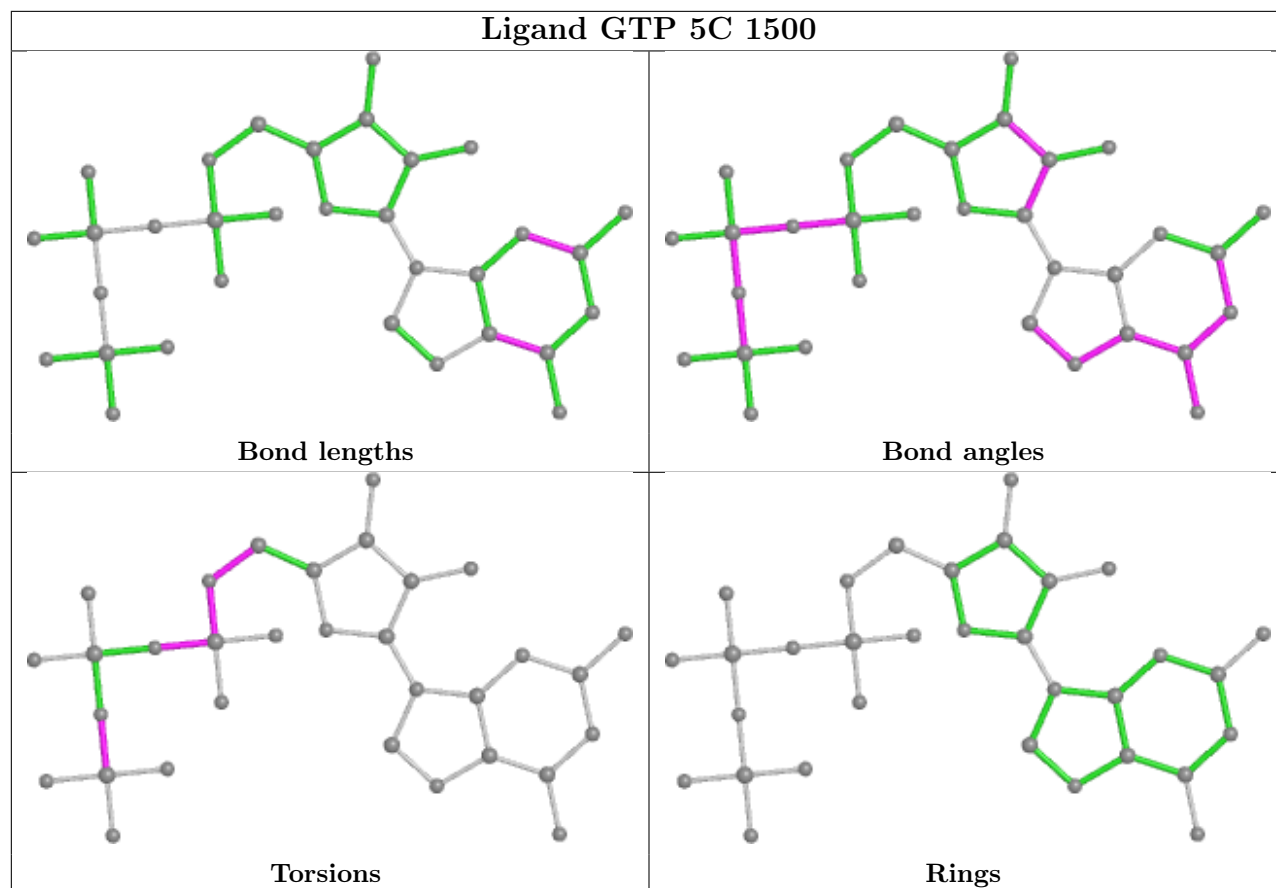
There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the

average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.





5.7 Other polymers [\(i\)](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [\(i\)](#)

There are no chain breaks in this entry.