



## wwPDB EM Validation Summary Report ⓘ

Apr 6, 2025 – 02:06 AM JST

PDB ID : 9L5R / pdb\_0000915r  
EMDB ID : EMD-62841  
Title : Cryo-EM structure of the thermophile spliceosome (state ILS)  
Authors : Li, Y.; Fischer, P.; Wang, M.; Yuan, R.; Meng, W.; Luehrmann, R.; Lau, B.;  
Hurt, E.; Cheng, J.  
Deposited on : 2024-12-23  
Resolution : 2.80 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](mailto: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:

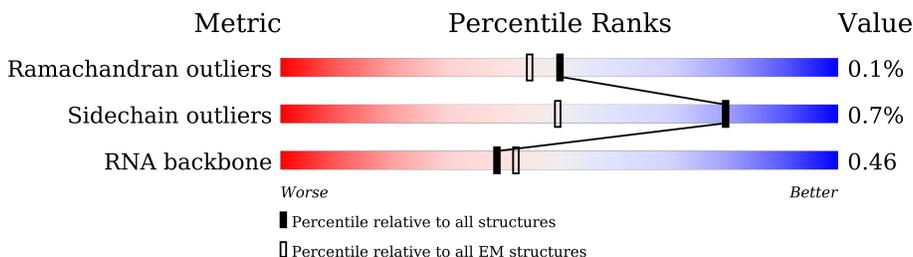
EMDB validation analysis : **FAILED**  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
MolProbity : 4.02b-467  
buster-report : 1.1.7 (2018)  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
MapQ : **FAILED**  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.42

# 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 2.80 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  $\geq 3$ , 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	2	193	
2	5	116	
3	6	101	
4	A	2463	
5	B	326	
6	C	1011	
7	D	325	
8	E	352	
9	F	233	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	I	839	87% 12%
11	J	687	88% 11%
12	L	768	82% 17%
13	K	231	99%
14	q	480	29% 71%
14	r	480	29% 70%
14	s	480	29% 71%
14	t	480	29% 71%
15	N	148	97%
16	S	167	93% 6%
17	T	496	64% 34%
18	U	757	45% 54%
19	M	395	62% 37%
20	0	408	65% 32%
21	R	578	58% 41%
22	W	547	72% 27%
23	P	260	43% 56%
24	Y	1416	95% 5%
25	a	98	88% 12%
25	j	98	90% 10%
26	b	94	76% 24%
26	l	94	85% 14%
27	c	592	14% 86%
27	m	592	14% 85%
28	d	118	69% 31%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
28	o	118	 72% 26%
29	e	211	 37% 63%
29	p	211	 48% 51%
30	f	114	 74% 26%
30	u	114	 79% 21%
31	g	82	 88% 12%
31	k	82	 88% 11%
32	h	242	 50% 50%
33	i	201	 51% 49%
34	CY	510	 18% 82%
35	Cb	975	 52% 48%
36	Cc	764	 93% 7%
37	7	101	 21% 74%
38	Ci	153	 55% 45%
39	H	22	 100%

## 2 Entry composition

There are 44 unique types of molecules in this entry. The entry contains 104144 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 U2 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
1	2	106	2224	994	357	767	106	0	0

- Molecule 2 is a RNA chain called U5 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	5	111	2343	1048	398	786	111	0	0

- Molecule 3 is a RNA chain called U6 snRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
3	6	98	2090	935	379	678	98	0	0

- Molecule 4 is a protein called PRP8.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	A	2006	16540	10628	2874	2976	62	0	0

- Molecule 5 is a protein called Pre-mRNA-splicing factor SYF2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	B	258	2071	1273	388	405	5	0	0

- Molecule 6 is a protein called SNU114.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	C	922	7301	4668	1229	1368	36	0	0

- Molecule 7 is a protein called SDE2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	D	95	786	477	150	154	5	0	0

- Molecule 8 is a protein called Anaphase-promoting complex subunit 4-like WD40 domain-containing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
8	E	310	2379	1493	414	462	10	0	0

- Molecule 9 is a protein called CCDC12.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
9	F	110	879	544	166	167	2	0	0

- Molecule 10 is a protein called Putative pre-mRNA splicing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
10	I	738	6135	3924	1071	1112	28	0	0

- Molecule 11 is a protein called Suppressor of forked domain-containing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
11	J	608	5176	3297	937	929	13	0	0

- Molecule 12 is a protein called Putative pre-mRNA splicing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
12	L	637	5052	3111	952	974	15	0	0

- Molecule 13 is a protein called Pre-mRNA-splicing factor SPF27.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
13	K	231	1797	1122	317	354	4	0	0

- Molecule 14 is a protein called Pre-mRNA-processing factor 19.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	q	139	Total	C	N	O	S	0	0
			1110	699	195	214	2		
14	t	141	Total	C	N	O	S	0	0
			1122	707	197	216	2		
14	r	143	Total	C	N	O	S	0	0
			1132	713	199	218	2		
14	s	140	Total	C	N	O	S	0	0
			1115	702	196	215	2		

- Molecule 15 is a protein called Putative bud site selection protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	N	148	Total	C	N	O	S	0	0
			1200	755	213	220	12		

- Molecule 16 is a protein called Peptidyl-prolyl cis-trans isomerase.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	S	157	Total	C	N	O	S	0	0
			1209	763	217	223	6		

- Molecule 17 is a protein called Pre-mRNA-splicing factor PRP46.

Mol	Chain	Residues	Atoms					AltConf	Trace
17	T	328	Total	C	N	O	S	0	0
			2565	1621	461	469	14		

- Molecule 18 is a protein called Cell cycle control protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
18	U	348	Total	C	N	O	S	0	0
			2821	1770	516	524	11		

- Molecule 19 is a protein called Putative pre-mRNA splicing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
19	M	248	Total	C	N	O	S	0	0
			1964	1238	355	354	17		

- Molecule 20 is a protein called Putative pre-mRNA splicing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
20	0	276	Total	C	N	O	S	0	0
			2224	1381	424	412	7		

- Molecule 21 is a protein called Pre-mRNA-processing protein 45.

Mol	Chain	Residues	Atoms					AltConf	Trace	
21	R	340	Total	C	N	O	P	S	0	0
			2686	1664	509	503	2	8		

- Molecule 22 is a protein called PRP17.

Mol	Chain	Residues	Atoms					AltConf	Trace
22	W	397	Total	C	N	O	S	0	0
			2224	1336	444	440	4		

- Molecule 23 is a protein called Putative pre-mRNA splicing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
23	P	114	Total	C	N	O	S	0	0
			925	577	182	165	1		

- Molecule 24 is a protein called Pre-mRNA-splicing factor.

Mol	Chain	Residues	Atoms					AltConf	Trace
24	Y	1344	Total	C	N	O	S	0	0
			10819	6892	1900	2002	25		

- Molecule 25 is a protein called Small nuclear ribonucleoprotein E.

Mol	Chain	Residues	Atoms					AltConf	Trace
25	a	86	Total	C	N	O		0	0
			426	253	86	87			
25	j	88	Total	C	N	O	S	0	0
			704	456	121	126	1		

- Molecule 26 is a protein called Sm protein F.

Mol	Chain	Residues	Atoms					AltConf	Trace
26	b	71	Total	C	N	O		0	0
			351	209	71	71			
26	l	81	Total	C	N	O	S	0	0
			649	412	114	120	3		

- Molecule 27 is a protein called Delta(14)-sterol reductase.

Mol	Chain	Residues	Atoms					AltConf	Trace
27	c	84	Total	C	N	O	0	0	
			416	247	84	85			
27	m	86	Total	C	N	O	S	0	0
			678	427	129	118	4		

- Molecule 28 is a protein called Small nuclear ribonucleoprotein Sm D1.

Mol	Chain	Residues	Atoms					AltConf	Trace
28	d	81	Total	C	N	O	0	0	
			401	239	81	81			
28	o	87	Total	C	N	O	S	0	0
			679	433	114	128	4		

- Molecule 29 is a protein called Sm protein B.

Mol	Chain	Residues	Atoms					AltConf	Trace
29	e	79	Total	C	N	O	0	0	
			388	230	79	79			
29	p	103	Total	C	N	O	S	0	0
			788	490	148	145	5		

- Molecule 30 is a protein called Small nuclear ribonucleoprotein Sm D3.

Mol	Chain	Residues	Atoms					AltConf	Trace
30	f	84	Total	C	N	O	0	0	
			414	246	84	84			
30	u	90	Total	C	N	O	S	0	0
			716	449	132	131	4		

- Molecule 31 is a protein called Small nuclear ribonucleoprotein G.

Mol	Chain	Residues	Atoms					AltConf	Trace
31	g	72	Total	C	N	O	0	0	
			354	210	72	72			
31	k	73	Total	C	N	O	S	0	0
			577	369	101	105	2		

- Molecule 32 is a protein called U2 small nuclear ribonucleoprotein A'.

Mol	Chain	Residues	Atoms				AltConf	Trace
32	h	122	Total	C	N	O	0	0
			605	361	122	122		

- Molecule 33 is a protein called U2 small nuclear ribonucleoprotein B'-like protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
33	i	102	Total	C	N	O	0	0
			504	300	102	102		

- Molecule 34 is a protein called Nineteen complex-related protein 2-domain-containing protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
34	CY	91	Total	C	N	O	0	0
			444	262	91	91		

- Molecule 35 is a protein called G-patch domain-containing protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
35	Cb	505	Total	C	N	O	0	0
			2496	1486	505	505		

- Molecule 36 is a protein called RNA helicase.

Mol	Chain	Residues	Atoms				AltConf	Trace
36	Cc	714	Total	C	N	O	0	0
			3541	2112	714	715		

- Molecule 37 is a RNA chain called Unknown mRNA.

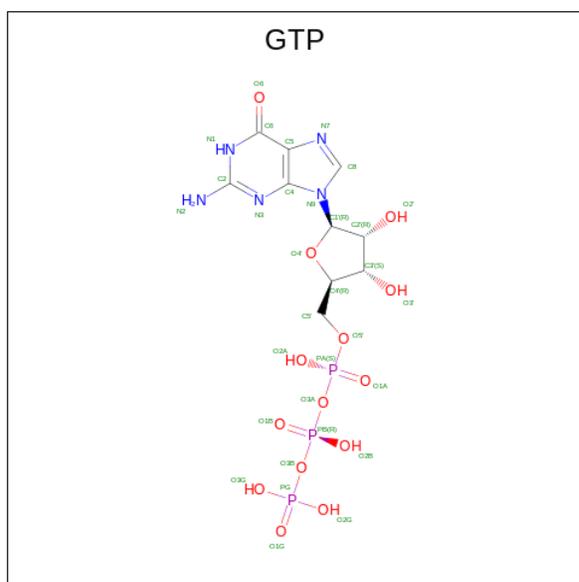
Mol	Chain	Residues	Atoms					AltConf	Trace
37	7	26	Total	C	N	O	P	0	0
			494	241	68	159	26		

- Molecule 38 is a protein called Putative cyclophilin protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
38	Ci	84	Total	C	N	O	0	0
			415	247	84	84		

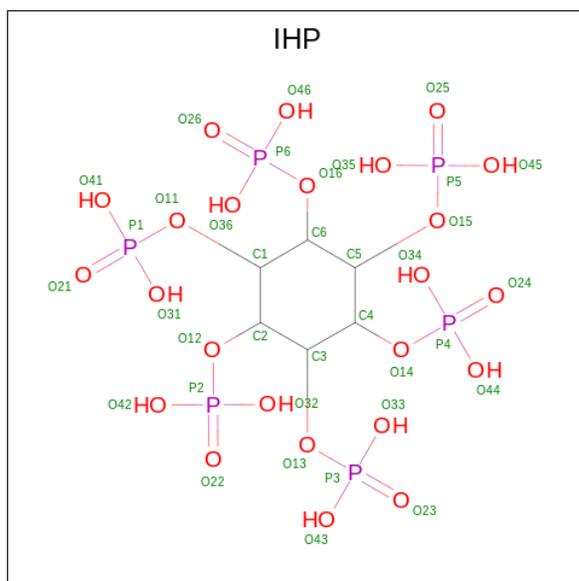
- Molecule 39 is a protein called Unknown protein.





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

- Molecule 43 is INOSITOL HEXAKISPHOSPHATE (CCD ID: IHP) (formula:  $C_6H_{18}O_{24}P_6$ ).



Mol	Chain	Residues	Atoms			AltConf	
			Total	C	O		P
43	J	1	36	6	24	6	0

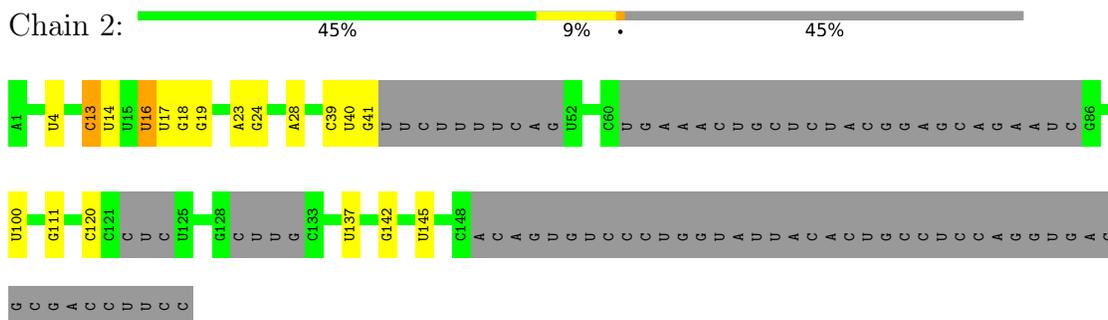
- Molecule 44 is ZINC ION (CCD ID: ZN) (formula: Zn).

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>AltConf</b>
44	N	3	Total 3	Zn 3	0
44	U	1	Total 1	Zn 1	0
44	M	2	Total 2	Zn 2	0

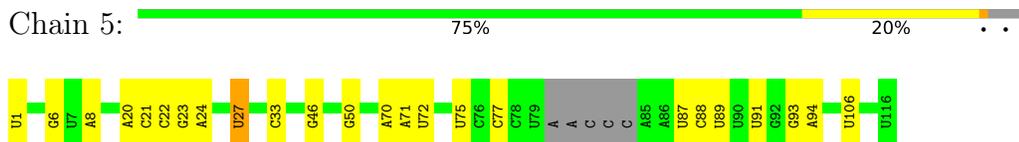
### 3 Residue-property plots [i](#)

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.

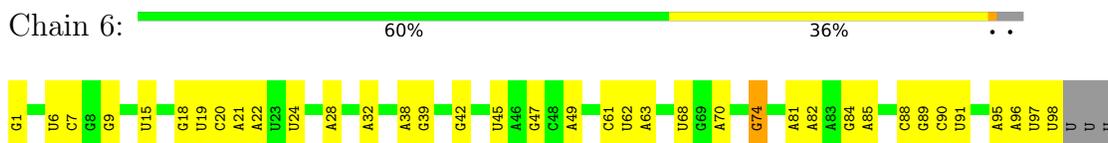
- Molecule 1: U2 snRNA



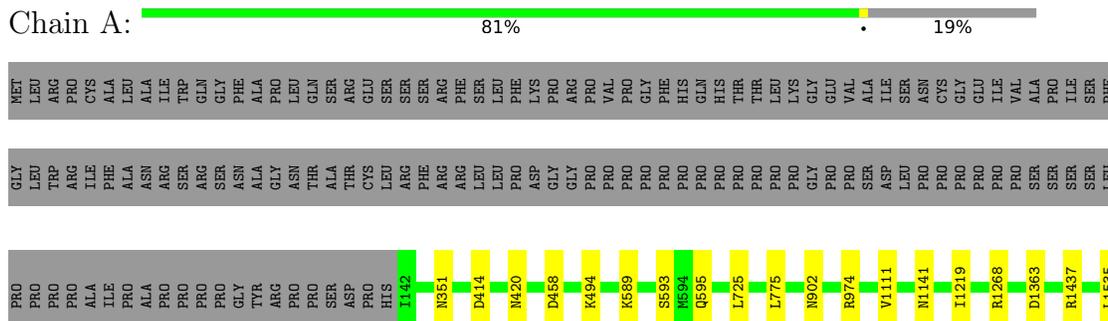
- Molecule 2: U5 snRNA

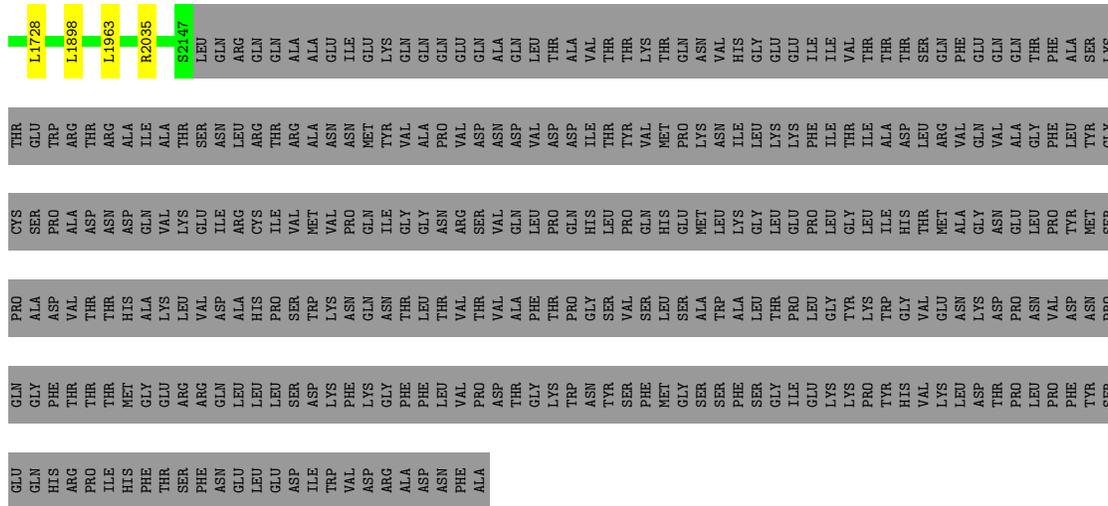


- Molecule 3: U6 snRNA

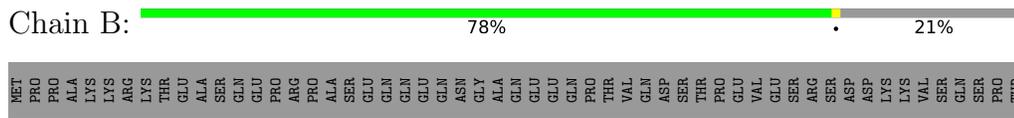


- Molecule 4: PRP8

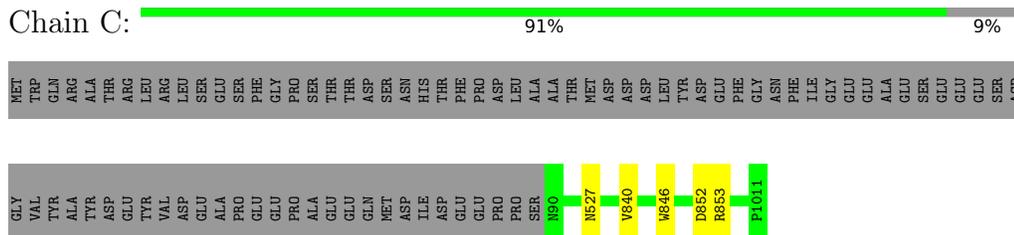




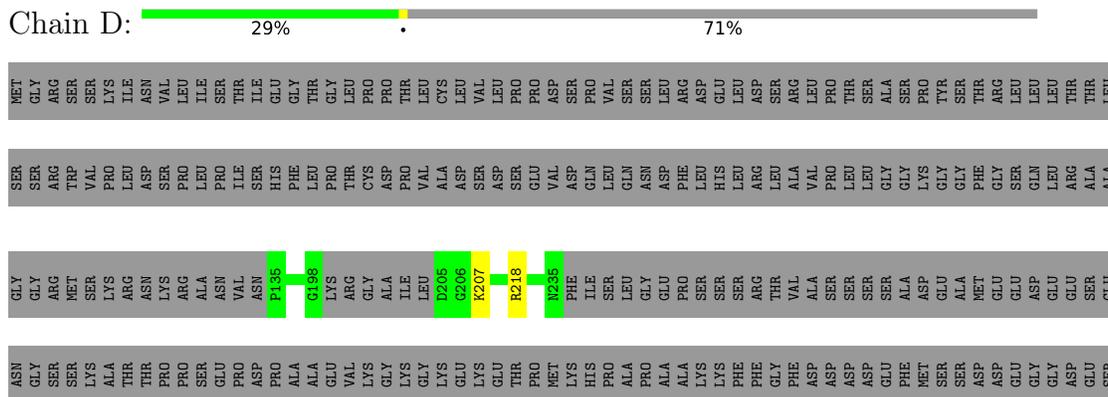
• Molecule 5: Pre-mRNA-splicing factor SYF2



• Molecule 6: SNU114

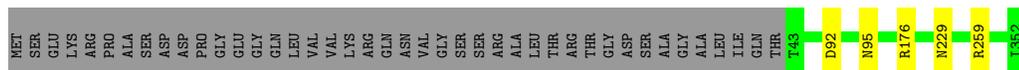


• Molecule 7: SDE2



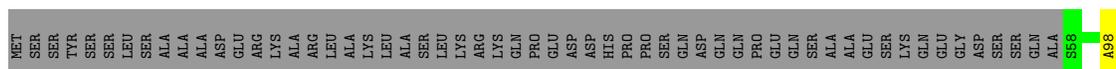
• Molecule 8: Anaphase-promoting complex subunit 4-like WD40 domain-containing protein

Chain E: 87% 12%



• Molecule 9: CCDC12

Chain F: 46% 53%



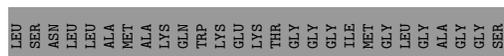
• Molecule 10: Putative pre-mRNA splicing protein

Chain I: 87% 12%



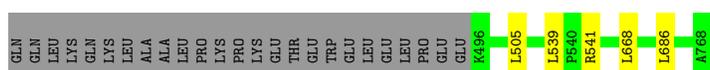
• Molecule 11: Suppressor of forked domain-containing protein

Chain J: 88% 11%



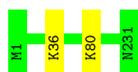
• Molecule 12: Putative pre-mRNA splicing protein

Chain L: 82% 17%



• Molecule 13: Pre-mRNA-splicing factor SPF27

Chain K:  99%



● Molecule 14: Pre-mRNA-processing factor 19

Chain q:  29% 71%

Table of amino acid sequences for Chain q, with highlighted residues R4, R25, and A139. The table is organized into four columns of residues.

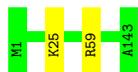
● Molecule 14: Pre-mRNA-processing factor 19

Chain t:  29% 71%

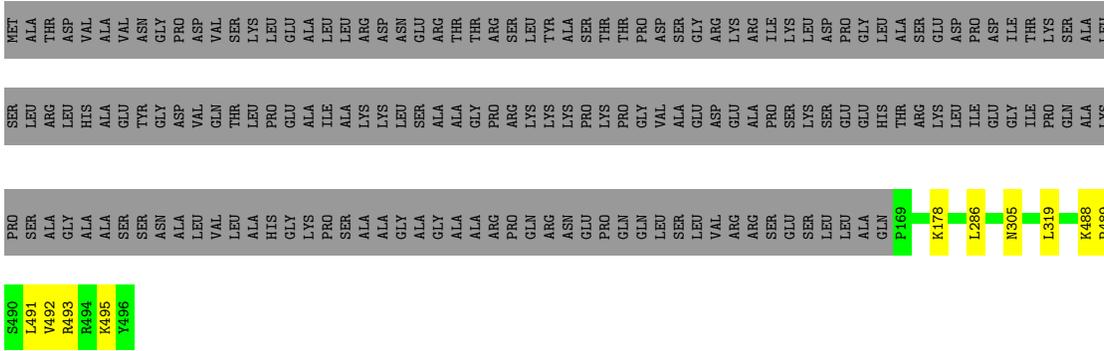
Table of amino acid sequences for Chain t, with highlighted residues H1 and P141. The table is organized into four columns of residues.

● Molecule 14: Pre-mRNA-processing factor 19

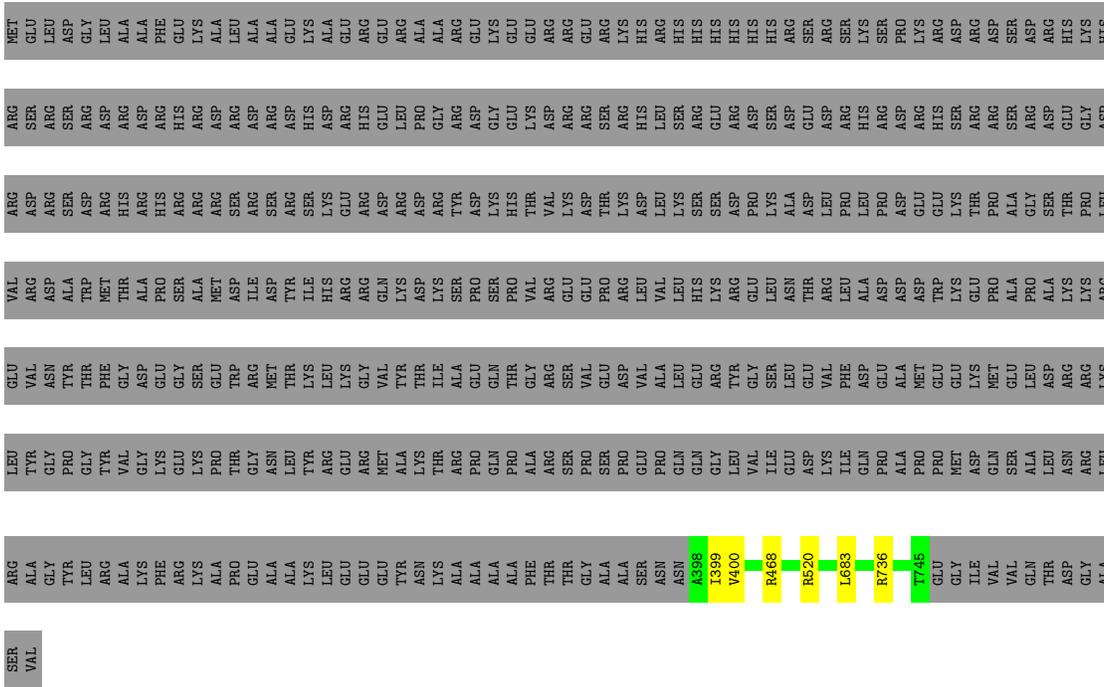
Chain r:  29% 70%



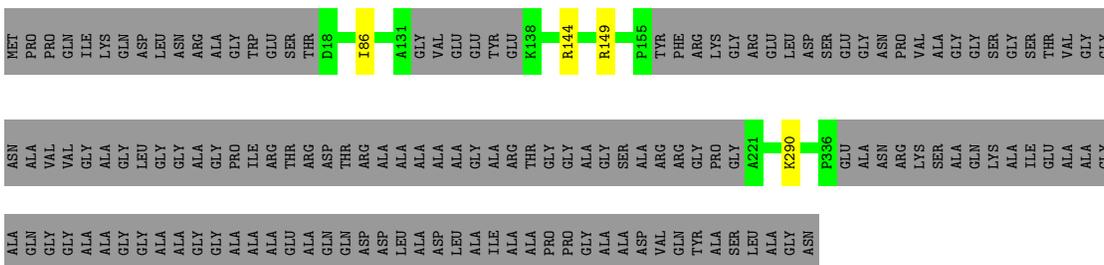




• Molecule 18: Cell cycle control protein



• Molecule 19: Putative pre-mRNA splicing protein



• Molecule 20: Putative pre-mRNA splicing protein













MET SER ASP ALA ALA R6 R89 ALA LYS PRO PRO LYS SER ALA ALA GLN GLN LEU GLY SER LYS THR VAL TRP GLN GLN GLY TRP LEU ALA HIS ALA VAL SER GLU ASP ARG LEU ALA ALA ARG ALA LYS ASN ALA ARG ASN ASN ASP PRO MET GLN GLY

LEU GLU ASP LEU VAL ALA GLY PRO LYS PRO GLU

- Molecule 39: Unknown protein

Chain H:  100%

There are no outlier residues recorded for this chain.

## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	77668	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE; Relion	Depositor
Microscope	TFS 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	FEI FALCON IV (4k 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: Y5P, SEP, P5P, ZN, M7M, IHP, MG, 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	2	0.43	0/2471	0.82	2/3832 (0.1%)
2	5	0.73	1/2612 (0.0%)	0.90	3/4059 (0.1%)
3	6	0.86	1/2338 (0.0%)	0.96	3/3639 (0.1%)
4	A	0.46	1/16969 (0.0%)	0.62	11/23006 (0.0%)
5	B	0.41	0/2099	0.63	1/2806 (0.0%)
6	C	0.40	0/7464	0.59	1/10117 (0.0%)
7	D	0.33	0/793	0.66	0/1051
8	E	0.35	0/2428	0.66	0/3295
9	F	0.31	0/891	0.66	0/1201
10	I	0.32	0/6282	0.59	1/8491 (0.0%)
11	J	0.42	0/5301	0.60	1/7149 (0.0%)
12	L	0.35	0/5120	0.63	2/6882 (0.0%)
13	K	0.29	0/1833	0.57	0/2493
14	q	0.27	0/1128	0.61	0/1533
14	r	0.27	0/1151	0.58	0/1566
14	s	0.26	0/1133	0.58	1/1540 (0.1%)
14	t	0.25	0/1141	0.56	0/1552
15	N	0.52	0/1227	0.66	1/1655 (0.1%)
16	S	0.36	0/1235	0.66	0/1671
17	T	0.56	0/2635	0.76	3/3582 (0.1%)
18	U	0.34	0/2883	0.62	1/3895 (0.0%)
19	M	0.39	0/2006	0.67	1/2703 (0.0%)
20	0	0.39	0/2278	0.61	0/3081
21	R	0.44	0/2724	0.63	0/3675
22	W	0.32	0/2237	0.57	1/3068 (0.0%)
23	P	0.46	0/945	0.66	0/1264
24	Y	0.25	0/11057	0.49	0/14995
25	a	0.24	0/425	0.43	0/589
25	j	0.26	0/716	0.60	0/969
26	b	0.24	0/350	0.47	0/486
26	l	0.26	0/661	0.63	0/898
27	c	0.24	0/415	0.45	0/575

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
27	m	0.27	0/687	0.61	0/922
28	d	0.24	0/400	0.44	0/556
28	o	0.29	0/691	0.61	0/940
29	e	0.25	0/386	0.47	0/533
29	p	0.29	0/799	0.72	0/1079
30	f	0.24	0/413	0.48	0/573
30	u	0.30	0/726	0.72	0/979
31	g	0.24	0/353	0.47	0/489
31	k	0.28	0/584	0.76	1/787 (0.1%)
32	h	0.24	0/604	0.44	0/841
33	i	0.24	0/503	0.40	0/699
34	CY	0.24	0/443	0.34	0/612
35	Cb	0.24	0/2494	0.35	0/3471
36	Cc	0.24	0/3540	0.39	0/4935
37	7	0.73	0/71	1.48	1/106 (0.9%)
38	Ci	0.25	0/414	0.42	0/575
All	All	0.40	3/106056 (0.0%)	0.62	35/145415 (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
4	A	0	1
8	E	0	1
10	I	0	1
12	L	0	2
16	S	0	1
All	All	0	6

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	6	1	G	OP3-P	-10.80	1.48	1.61
2	5	1	U	OP3-P	-10.61	1.48	1.61
4	A	1111	VAL	CB-CG1	-6.05	1.40	1.52

The worst 5 of 35 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	725	LEU	CB-CG-CD1	-8.57	96.42	111.00

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
22	W	108	MET	CA-CB-CG	8.11	127.08	113.30
10	I	424	ILE	CG1-CB-CG2	-7.69	94.49	111.40
3	6	74	G	C4-C5-N7	7.67	113.87	110.80
15	N	144	CYS	CB-CA-C	-7.66	95.08	110.40

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	A	1437	ARG	Peptide
8	E	92	ASP	Peptide
10	I	416	GLY	Peptide
12	L	505	LEU	Peptide
12	L	539	LEU	Peptide

## 5.2 Too-close contacts [i](#)

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

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	A	2004/2463 (81%)	1911 (95%)	90 (4%)	3 (0%)	48	77
5	B	256/326 (78%)	249 (97%)	6 (2%)	1 (0%)	30	61
6	C	920/1011 (91%)	889 (97%)	29 (3%)	2 (0%)	44	73
7	D	91/325 (28%)	90 (99%)	1 (1%)	0	100	100
8	E	308/352 (88%)	285 (92%)	23 (8%)	0	100	100
9	F	108/233 (46%)	104 (96%)	3 (3%)	1 (1%)	14	42
10	I	736/839 (88%)	705 (96%)	29 (4%)	2 (0%)	37	67

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
11	J	604/687 (88%)	592 (98%)	12 (2%)	0	100	100
12	L	629/768 (82%)	602 (96%)	26 (4%)	1 (0%)	44	73
13	K	229/231 (99%)	224 (98%)	5 (2%)	0	100	100
14	q	137/480 (28%)	135 (98%)	2 (2%)	0	100	100
14	r	141/480 (29%)	140 (99%)	1 (1%)	0	100	100
14	s	138/480 (29%)	137 (99%)	1 (1%)	0	100	100
14	t	139/480 (29%)	136 (98%)	3 (2%)	0	100	100
15	N	146/148 (99%)	136 (93%)	8 (6%)	2 (1%)	9	30
16	S	155/167 (93%)	148 (96%)	7 (4%)	0	100	100
17	T	326/496 (66%)	311 (95%)	14 (4%)	1 (0%)	37	67
18	U	346/757 (46%)	321 (93%)	23 (7%)	2 (1%)	22	51
19	M	242/395 (61%)	225 (93%)	17 (7%)	0	100	100
20	0	274/408 (67%)	261 (95%)	13 (5%)	0	100	100
21	R	334/578 (58%)	312 (93%)	21 (6%)	1 (0%)	37	67
22	W	391/547 (72%)	376 (96%)	14 (4%)	1 (0%)	37	67
23	P	110/260 (42%)	102 (93%)	8 (7%)	0	100	100
24	Y	1342/1416 (95%)	1315 (98%)	27 (2%)	0	100	100
25	a	84/98 (86%)	83 (99%)	1 (1%)	0	100	100
25	j	86/98 (88%)	84 (98%)	2 (2%)	0	100	100
26	b	69/94 (73%)	69 (100%)	0	0	100	100
26	l	79/94 (84%)	77 (98%)	2 (2%)	0	100	100
27	c	82/592 (14%)	80 (98%)	2 (2%)	0	100	100
27	m	84/592 (14%)	80 (95%)	4 (5%)	0	100	100
28	d	79/118 (67%)	76 (96%)	3 (4%)	0	100	100
28	o	85/118 (72%)	78 (92%)	7 (8%)	0	100	100
29	e	75/211 (36%)	73 (97%)	2 (3%)	0	100	100
29	p	99/211 (47%)	92 (93%)	7 (7%)	0	100	100
30	f	82/114 (72%)	79 (96%)	3 (4%)	0	100	100
30	u	88/114 (77%)	83 (94%)	5 (6%)	0	100	100
31	g	70/82 (85%)	69 (99%)	1 (1%)	0	100	100
31	k	71/82 (87%)	68 (96%)	3 (4%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
32	h	120/242 (50%)	117 (98%)	3 (2%)	0	100	100
33	i	100/201 (50%)	100 (100%)	0	0	100	100
34	CY	89/510 (18%)	88 (99%)	1 (1%)	0	100	100
35	Cb	501/975 (51%)	495 (99%)	6 (1%)	0	100	100
36	Cc	712/764 (93%)	694 (98%)	18 (2%)	0	100	100
38	Ci	82/153 (54%)	82 (100%)	0	0	100	100
All	All	12843/19790 (65%)	12373 (96%)	453 (4%)	17 (0%)	50	77

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	494	LYS
4	A	1219	ILE
5	B	111	GLN
6	C	840	VAL
9	F	98	ALA

### 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	
4	A	1813/2212 (82%)	1804 (100%)	9 (0%)	86	95
5	B	208/270 (77%)	207 (100%)	1 (0%)	86	95
6	C	809/884 (92%)	807 (100%)	2 (0%)	92	97
7	D	80/276 (29%)	78 (98%)	2 (2%)	42	75
8	E	254/287 (88%)	250 (98%)	4 (2%)	58	85
9	F	92/179 (51%)	91 (99%)	1 (1%)	70	90
10	I	646/729 (89%)	644 (100%)	2 (0%)	91	97
11	J	525/592 (89%)	525 (100%)	0	100	100
12	L	520/635 (82%)	517 (99%)	3 (1%)	84	95
13	K	186/186 (100%)	184 (99%)	2 (1%)	70	90

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
14	q	122/385 (32%)	121 (99%)	1 (1%)	79	93
14	r	123/385 (32%)	121 (98%)	2 (2%)	58	85
14	s	122/385 (32%)	121 (99%)	1 (1%)	79	93
14	t	123/385 (32%)	123 (100%)	0	100	100
15	N	131/131 (100%)	129 (98%)	2 (2%)	60	86
16	S	126/135 (93%)	126 (100%)	0	100	100
17	T	276/408 (68%)	269 (98%)	7 (2%)	42	75
18	U	294/649 (45%)	291 (99%)	3 (1%)	73	91
19	M	210/293 (72%)	207 (99%)	3 (1%)	62	87
20	0	227/335 (68%)	216 (95%)	11 (5%)	21	53
21	R	278/476 (58%)	277 (100%)	1 (0%)	89	96
22	W	73/459 (16%)	73 (100%)	0	100	100
23	P	91/213 (43%)	90 (99%)	1 (1%)	70	90
24	Y	1173/1231 (95%)	1169 (100%)	4 (0%)	91	97
25	j	79/85 (93%)	79 (100%)	0	100	100
26	l	72/84 (86%)	71 (99%)	1 (1%)	62	87
27	m	77/497 (16%)	76 (99%)	1 (1%)	65	88
28	o	78/95 (82%)	76 (97%)	2 (3%)	41	75
29	p	84/152 (55%)	83 (99%)	1 (1%)	67	89
30	u	80/94 (85%)	80 (100%)	0	100	100
31	k	64/71 (90%)	64 (100%)	0	100	100
All	All	9036/13198 (68%)	8969 (99%)	67 (1%)	80	94

5 of 67 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
23	P	238	ARG
24	Y	395	ARG
28	o	44	ARG
13	K	36	LYS
12	L	541	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
20	0	212	ASN
24	Y	683	GLN
4	A	1737	GLN
14	q	110	HIS
17	T	220	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	2	101/193 (52%)	18 (17%)	2 (1%)
2	5	109/116 (93%)	21 (19%)	1 (0%)
3	6	97/101 (96%)	36 (37%)	2 (2%)
37	7	2/101 (1%)	2 (100%)	0
All	All	309/511 (60%)	77 (24%)	5 (1%)

5 of 77 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	2	4	U
1	2	14	U
1	2	16	U
1	2	17	U
1	2	18	G

All (5) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	2	13	C
1	2	23	A
2	5	70	A
3	6	15	U
3	6	38	A

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

25 non-standard protein/DNA/RNA residues are modelled in this entry.

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$
37	Y5P	7	14	37	14,19,20	2.35	1 (7%)	18,26,29	1.02	1 (5%)
37	Y5P	7	4	37	14,19,20	2.42	1 (7%)	18,26,29	1.11	1 (5%)
37	Y5P	7	13	37	14,19,20	2.41	1 (7%)	18,26,29	1.00	1 (5%)
37	P5P	7	3	37,3	16,23,24	0.80	0	14,33,36	0.75	0
37	Y5P	7	95	37	14,19,20	2.34	1 (7%)	18,26,29	0.99	1 (5%)
37	Y5P	7	8	37	14,19,20	3.67	1 (7%)	18,26,29	0.81	1 (5%)
21	SEP	R	242	21	8,9,10	0.68	0	8,12,14	1.20	1 (12%)
37	Y5P	7	101	37	14,19,20	2.30	1 (7%)	18,26,29	1.03	1 (5%)
37	Y5P	7	99	37	14,19,20	3.84	1 (7%)	18,26,29	0.79	0
37	P5P	7	93	37,1	16,23,24	0.79	0	14,33,36	0.76	0
37	Y5P	7	98	37	14,19,20	2.43	1 (7%)	18,26,29	0.98	1 (5%)
37	Y5P	7	90	37	14,19,20	3.73	1 (7%)	18,26,29	0.78	1 (5%)
37	Y5P	7	6	37	14,19,20	2.37	1 (7%)	18,26,29	0.94	1 (5%)
37	P5P	7	7	37,3	16,23,24	1.28	2 (12%)	14,33,36	1.96	2 (14%)
37	Y5P	7	96	37	14,19,20	2.48	1 (7%)	18,26,29	1.02	1 (5%)
37	Y5P	7	94	37	14,19,20	3.78	1 (7%)	18,26,29	0.77	1 (5%)
37	Y5P	7	92	37	14,19,20	2.35	1 (7%)	18,26,29	0.96	1 (5%)
37	Y5P	7	100	37	14,19,20	3.75	1 (7%)	18,26,29	0.79	1 (5%)
21	SEP	R	234	21	8,9,10	1.37	1 (12%)	8,12,14	0.95	0
37	P5P	7	91	37,1	16,23,24	0.77	0	14,33,36	0.73	0
37	P5P	7	5	37	16,23,24	1.33	2 (12%)	14,33,36	1.95	2 (14%)
37	Y5P	7	12	37	14,19,20	2.43	1 (7%)	18,26,29	1.01	1 (5%)
37	Y5P	7	11	37	14,19,20	2.39	1 (7%)	18,26,29	0.99	1 (5%)
37	Y5P	7	89	37	14,19,20	2.36	1 (7%)	18,26,29	0.96	1 (5%)
37	Y5P	7	10	37	14,19,20	2.41	1 (7%)	18,26,29	0.99	1 (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
37	Y5P	7	14	37	-	2/7/33/34	0/2/2/2
37	Y5P	7	4	37	-	1/7/33/34	0/2/2/2

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
37	Y5P	7	13	37	-	1/7/33/34	0/2/2/2
37	P5P	7	3	37,3	-	2/3/25/26	0/3/3/3
37	Y5P	7	95	37	-	4/7/33/34	0/2/2/2
37	Y5P	7	8	37	-	1/7/33/34	0/2/2/2
21	SEP	R	242	21	-	2/5/8/10	-
37	Y5P	7	101	37	-	1/7/33/34	0/2/2/2
37	Y5P	7	99	37	-	1/7/33/34	0/2/2/2
37	P5P	7	93	37,1	-	1/3/25/26	0/3/3/3
37	Y5P	7	98	37	-	1/7/33/34	0/2/2/2
37	Y5P	7	90	37	-	2/7/33/34	0/2/2/2
37	Y5P	7	6	37	-	1/7/33/34	0/2/2/2
37	P5P	7	7	37,3	-	2/3/25/26	0/3/3/3
37	Y5P	7	96	37	-	1/7/33/34	0/2/2/2
37	Y5P	7	94	37	-	1/7/33/34	0/2/2/2
37	Y5P	7	92	37	-	1/7/33/34	0/2/2/2
37	Y5P	7	100	37	-	1/7/33/34	0/2/2/2
21	SEP	R	234	21	-	1/5/8/10	-
37	P5P	7	91	37,1	-	0/3/25/26	0/3/3/3
37	P5P	7	5	37	-	0/3/25/26	0/3/3/3
37	Y5P	7	12	37	-	4/7/33/34	0/2/2/2
37	Y5P	7	11	37	-	4/7/33/34	0/2/2/2
37	Y5P	7	89	37	-	3/7/33/34	0/2/2/2
37	Y5P	7	10	37	-	3/7/33/34	0/2/2/2

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
37	7	99	Y5P	C4-N3	-14.20	1.33	1.46
37	7	94	Y5P	C4-N3	-13.97	1.33	1.46
37	7	100	Y5P	C4-N3	-13.85	1.33	1.46
37	7	90	Y5P	C4-N3	-13.72	1.33	1.46
37	7	8	Y5P	C4-N3	-13.53	1.33	1.46

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
37	7	7	P5P	C6-N1-C2	6.43	125.06	115.84
37	7	5	P5P	C6-N1-C2	6.36	124.94	115.84
37	7	101	Y5P	N1-C2-N3	-3.67	114.56	125.33
37	7	10	Y5P	N1-C2-N3	-3.60	114.76	125.33

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
37	7	11	Y5P	N1-C2-N3	-3.59	114.79	125.33

There are no chirality outliers.

5 of 41 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
21	R	234	SEP	N-CA-CB-OG
37	7	6	Y5P	O4'-C1'-N1-C2
37	7	11	Y5P	O4'-C1'-N1-C2
37	7	89	Y5P	O4'-C1'-N1-C2
37	7	94	Y5P	O4'-C1'-N1-C2

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 7 are monoatomic - leaving 3 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
43	IHP	J	1001	-	36,36,36	1.47	6 (16%)	54,60,60	0.98	3 (5%)
42	GTP	C	1102	41	26,34,34	1.38	3 (11%)	32,54,54	1.54	6 (18%)
40	M7M	B	401	-	27,32,33	4.11	15 (55%)	33,49,52	1.31	4 (12%)

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
43	IHP	J	1001	-	-	9/30/54/54	0/1/1/1
42	GTP	C	1102	41	-	6/18/38/38	0/3/3/3
40	M7M	B	401	-	-	2/17/47/48	0/3/3/3

The worst 5 of 24 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
40	B	401	M7M	CBI-NBP	8.02	1.50	1.46
40	B	401	M7M	CBG-NBH	7.81	1.44	1.35
40	B	401	M7M	CBY-CBS	-7.10	1.34	1.53
40	B	401	M7M	OBR-CBS	6.91	1.60	1.45
40	B	401	M7M	CBO-NBP	6.25	1.44	1.35

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
40	B	401	M7M	CBO-CBG-NBH	4.35	109.30	106.71
43	J	1001	IHP	C5-C4-C3	-3.73	102.25	110.41
42	C	1102	GTP	PB-O3B-PG	-3.69	120.15	132.83
42	C	1102	GTP	C5-C6-N1	3.38	119.91	113.95
42	C	1102	GTP	C8-N7-C5	3.07	108.85	102.99

There are no chirality outliers.

5 of 17 torsion outliers are listed below:

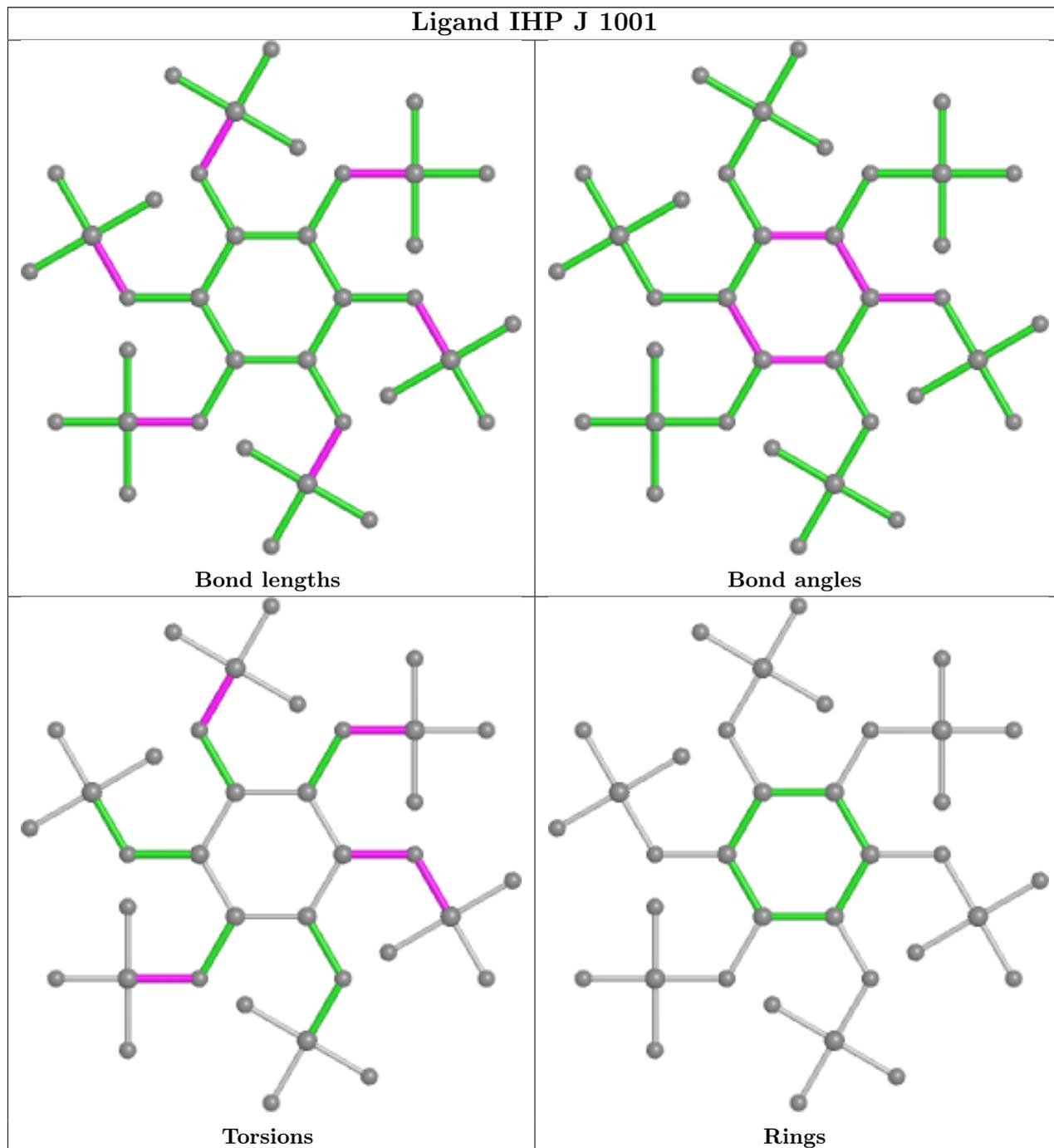
Mol	Chain	Res	Type	Atoms
42	C	1102	GTP	C5'-O5'-PA-O1A
42	C	1102	GTP	C5'-O5'-PA-O2A
43	J	1001	IHP	C4-C5-O15-P5
43	J	1001	IHP	C3-O13-P3-O23
43	J	1001	IHP	C5-O15-P5-O35

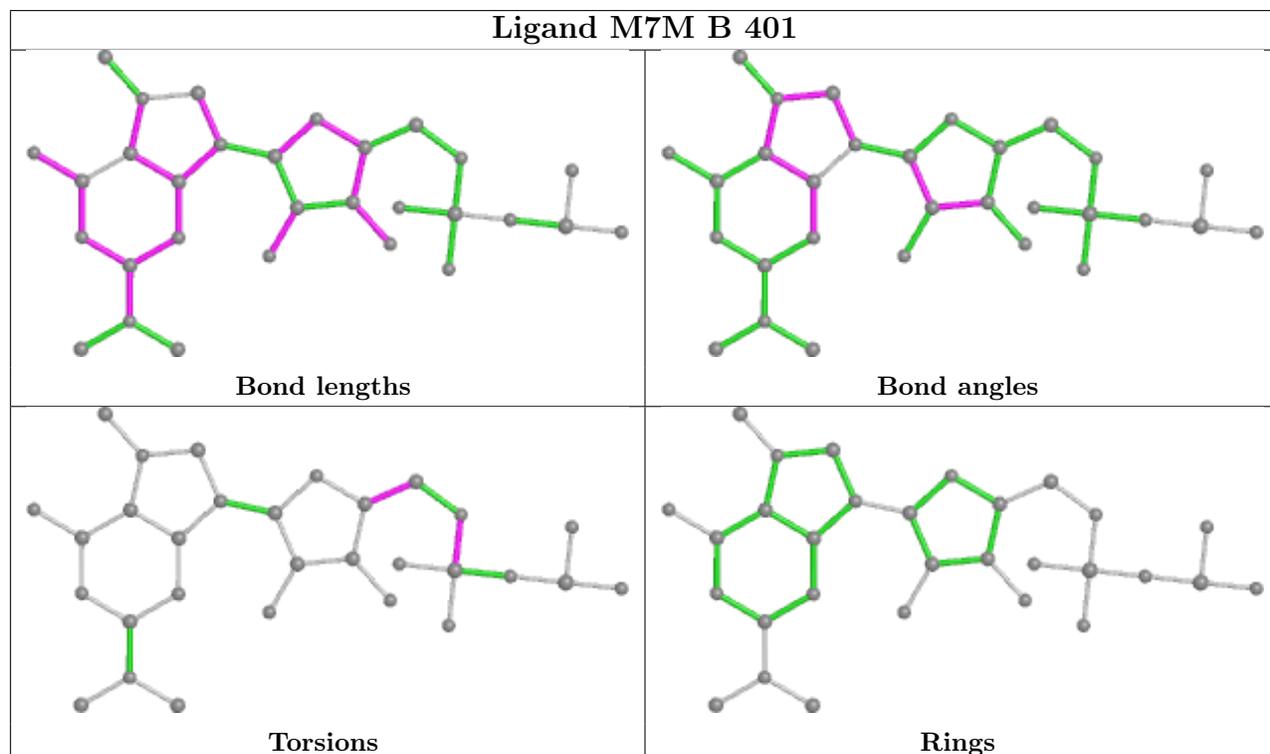
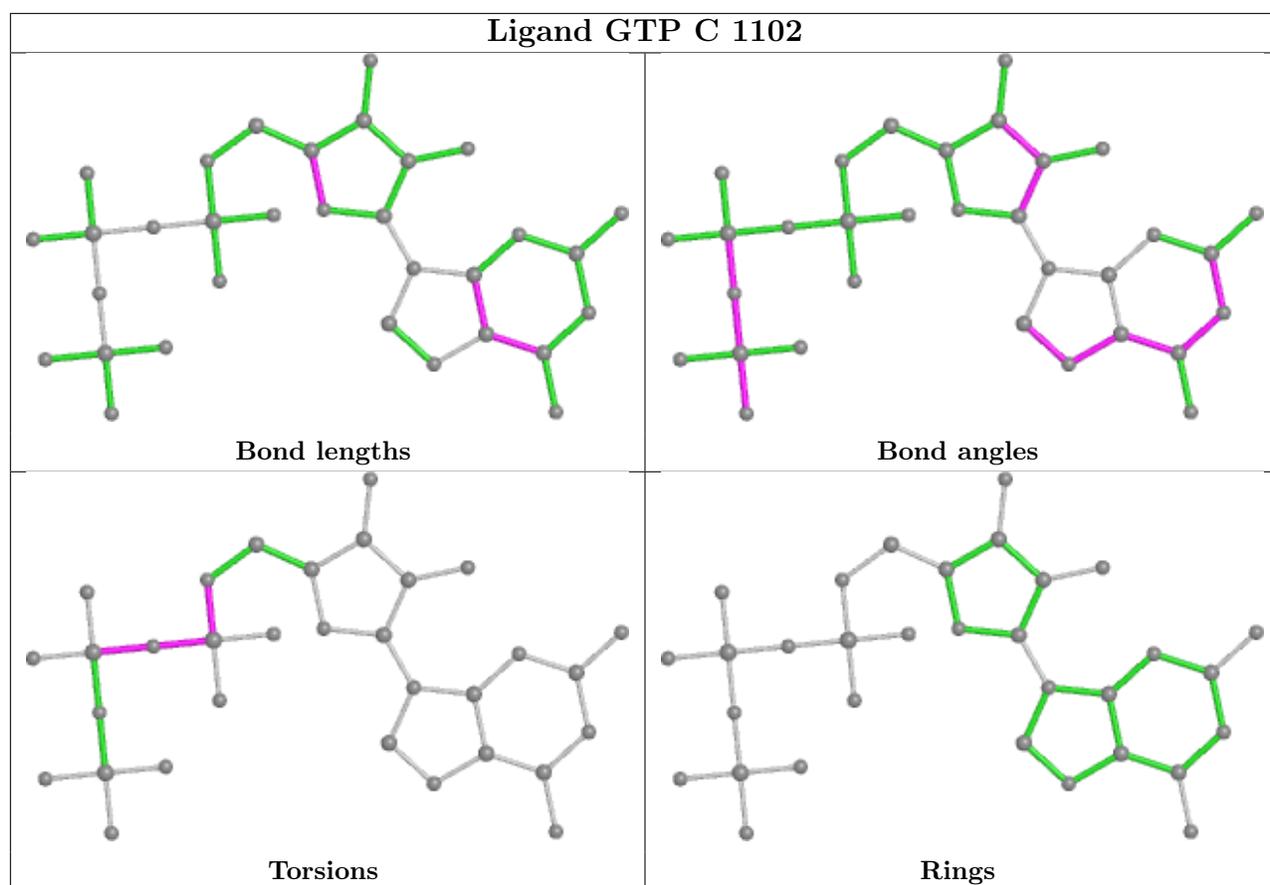
There are no ring outliers.

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.