



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Aug 27, 2017 – 08:07 AM EDT

PDB ID : 5MQF
EMDB ID: : EMD-3545
Title : Cryo-EM structure of a human spliceosome activated for step 2 of splicing (C* complex)
Authors : Bertram, K.; Hartmuth, K.; Kastner, B.
Deposited on : unknown
Resolution : 5.90 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

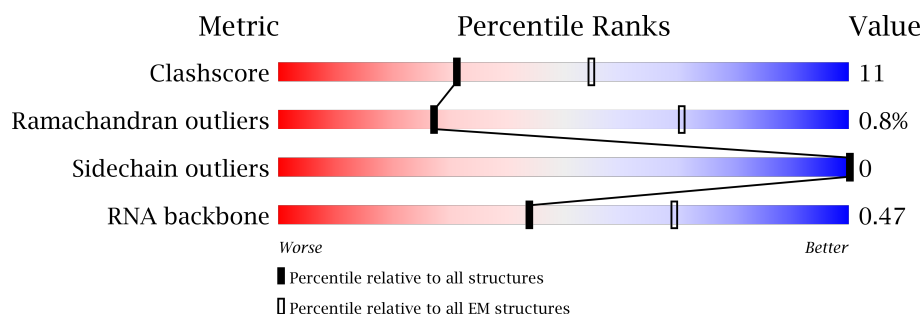
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 5.90 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.











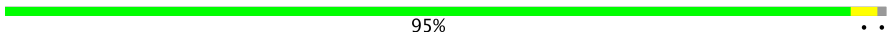
















Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026
RNA backbone	3398	335

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	2335	72% 12% 16%
2	B	972	77% 15% 7%
3	C	536	40% 55%
4	D	515	45% 16% 39%
5	E	579	43% 14% 43%
6	F	357	65% 21% 14%
7	G	504	24% 74%
7	H	504	23% 73%







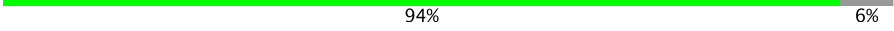


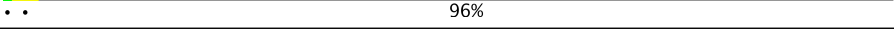



Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	I	504	
7	J	504	
8	K	225	
9	L	802	
10	M	855	
11	N	243	
12	O	848	
13	P	420	
14	Q	144	
15	R	229	
16	S	2752	
17	T	908	
18	U	1485	
19	V	166	
20	W	255	
21	X	225	
22	a	118	
22	h	118	
23	b	86	
23	i	86	
24	c	92	
24	j	92	
25	d	76	
25	k	76	
26	e	126	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
26	l	126	 66% 34%
27	f	240	 30% 70%
27	m	240	 30% 70%
28	g	119	 68% 31%
28	n	119	 69% 31%
29	o	301	 74% 26%
30	p	411	 94% 6%
31	q	1218	 51% 49%
32	Y	324	 6% 90%
32	Z	324	 96%
33	2	188	 47% 20% 7% 26%
34	5	116	 42% 45% 11%
35	6	106	 36% 38% 10% 16%

2 Entry composition

There are 35 unique types of molecules in this entry. The entry contains 55236 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 protein called Pre-mRNA-processing-splicing factor 8.

Mol	Chain	Residues	Atoms				AltConf	Trace
1	A	1965	Total	C	N	O	13	0
			9955	6025	1965	1965		

- Molecule 2 is a protein called 116 kDa U5 small nuclear ribonucleoprotein component.

Mol	Chain	Residues	Atoms				AltConf	Trace
2	B	902	Total	C	N	O	0	0
			4566	2762	902	902		

- Molecule 3 is a protein called SNW domain-containing protein 1.

Mol	Chain	Residues	Atoms				AltConf	Trace
3	C	240	Total	C	N	O	0	0
			1231	751	240	240		

- Molecule 4 is a protein called Pleiotropic regulator 1.

Mol	Chain	Residues	Atoms				AltConf	Trace
4	D	315	Total	C	N	O	8	0
			1576	946	315	315		

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	454	ALA	VAL	conflict	UNP O43660
D	455	VAL	GLN	conflict	UNP O43660
D	456	GLN	PRO	conflict	UNP O43660
D	457	PRO	GLY	conflict	UNP O43660
D	458	GLY	SER	conflict	UNP O43660
D	459	SER	LEU	conflict	UNP O43660
D	460	LEU	ASP	conflict	UNP O43660
D	461	ASP	SER	conflict	UNP O43660
D	492	ASP	-	insertion	UNP O43660

- Molecule 5 is a protein called Pre-mRNA-processing factor 17.

Mol	Chain	Residues	Atoms				AltConf	Trace
5	E	329	Total	C	N	O	0	0
			1646	988	329	329		

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

Mol	Chain	Residues	Atoms				AltConf	Trace
6	F	307	Total	C	N	O	0	0
			1531	917	307	307		

- Molecule 7 is a protein called Pre-mRNA-processing factor 19.

Mol	Chain	Residues	Atoms				AltConf	Trace
7	G	132	Total	C	N	O	0	0
			679	415	132	132		
7	H	135	Total	C	N	O	0	0
			696	426	135	135		
7	I	134	Total	C	N	O	0	0
			691	423	134	134		
7	J	135	Total	C	N	O	0	0
			696	426	135	135		

- Molecule 8 is a protein called Pre-mRNA-splicing factor SPF27.

Mol	Chain	Residues	Atoms				AltConf	Trace
8	K	191	Total	C	N	O	0	0
			960	578	191	191		

- Molecule 9 is a protein called Cell division cycle 5-like protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
9	L	336	Total	C	N	O	0	0
			1680	1008	336	336		

- Molecule 10 is a protein called Pre-mRNA-splicing factor SYF1.

Mol	Chain	Residues	Atoms				AltConf	Trace
10	M	706	Total	C	N	O	0	0
			3553	2141	706	706		

- Molecule 11 is a protein called Pre-mRNA-splicing factor SYF2.

Mol	Chain	Residues	Atoms				AltConf	Trace
11	N	115	Total	C	N	O	0	0
			579	349	115	115		

- Molecule 12 is a protein called Crooked neck-like protein 1.

Mol	Chain	Residues	Atoms				AltConf	Trace
12	O	564	Total	C	N	O	0	0
			2842	1715	564	563		

- Molecule 13 is a protein called Pre-mRNA-splicing factor RBM22.

Mol	Chain	Residues	Atoms				AltConf	Trace
13	P	277	Total	C	N	O	0	0
			1396	842	277	277		

- Molecule 14 is a protein called Protein BUD31 homolog.

Mol	Chain	Residues	Atoms				AltConf	Trace
14	Q	142	Total	C	N	O	0	0
			713	429	142	142		

- Molecule 15 is a protein called Spliceosome-associated protein CWC15 homolog.

Mol	Chain	Residues	Atoms				AltConf	Trace
15	R	98	Total	C	N	O	0	0
			493	297	98	98		

- Molecule 16 is a protein called Serine/arginine repetitive matrix protein 2.

Mol	Chain	Residues	Atoms				AltConf	Trace
16	S	30	Total	C	N	O	0	0
			148	88	30	30		

- Molecule 17 is a protein called Pre-mRNA-splicing factor CWC22 homolog.

Mol	Chain	Residues	Atoms				AltConf	Trace
17	T	395	Total	C	N	O	2	197
			1191	795	198	198		

- Molecule 18 is a protein called Intron-binding protein aquarius.

Mol	Chain	Residues	Atoms		AltConf	Trace
18	U	1287	Total	C		
			1287	1287	6	1287

- Molecule 19 is a protein called Peptidyl-prolyl cis-trans isomerase-like 1.

Mol	Chain	Residues	Atoms				AltConf	Trace
19	V	155	Total	C	N	O	0	0
			774	464	155	155		

- Molecule 20 is a protein called U2 small nuclear ribonucleoprotein A'.

Mol	Chain	Residues	Atoms				AltConf	Trace
20	W	162	Total	C	N	O	0	0
			816	492	162	162		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
W	89	ASP	CYS	conflict	UNP P09661
W	119	CYS	SER	conflict	UNP P09661

- Molecule 21 is a protein called U2 small nuclear ribonucleoprotein B'.

Mol	Chain	Residues	Atoms				AltConf	Trace
21	X	94	Total	C	N	O	0	0
			470	282	94	94		

- Molecule 22 is a protein called Small nuclear ribonucleoprotein Sm D2.

Mol	Chain	Residues	Atoms				AltConf	Trace
22	a	99	Total	C	N	O	0	0
			497	299	99	99		
22	h	98	Total	C	N	O	0	0
			493	297	98	98		

- Molecule 23 is a protein called Small nuclear ribonucleoprotein F.

Mol	Chain	Residues	Atoms				AltConf	Trace
23	b	76	Total	C	N	O	0	0
			379	227	76	76		
23	i	74	Total	C	N	O	0	0
			369	221	74	74		

- Molecule 24 is a protein called Small nuclear ribonucleoprotein E.

Mol	Chain	Residues	Atoms				AltConf	Trace
24	c	79	Total	C	N	O	0	0
			393	235	79	79		
24	j	79	Total	C	N	O	0	0
			393	235	79	79		

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

Mol	Chain	Residues	Atoms				AltConf	Trace
25	d	74	Total	C	N	O	0	0
			369	221	74	74		
25	k	74	Total	C	N	O	0	0
			369	221	74	74		

- Molecule 26 is a protein called Small nuclear ribonucleoprotein Sm D3.

Mol	Chain	Residues	Atoms				AltConf	Trace
26	e	84	Total	C	N	O	0	0
			420	252	84	84		
26	l	83	Total	C	N	O	0	0
			415	249	83	83		

- Molecule 27 is a protein called Small nuclear ribonucleoprotein-associated proteins B and B'.

Mol	Chain	Residues	Atoms				AltConf	Trace
27	f	71	Total	C	N	O	0	0
			355	213	71	71		
27	m	71	Total	C	N	O	0	0
			355	213	71	71		

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

Mol	Chain	Residues	Atoms				AltConf	Trace
28	g	82	Total	C	N	O	0	0
			412	248	82	82		
28	n	82	Total	C	N	O	0	0
			412	248	82	82		

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

Mol	Chain	Residues	Atoms		AltConf	Trace
29	o	223	Total	C	0	223
			223	223		

- Molecule 30 is a protein called Eukaryotic initiation factor 4A-III.

Mol	Chain	Residues	Atoms		AltConf	Trace
30	p	385	Total	C	8	385
			385	385		

- Molecule 31 is a protein called ATP-dependent RNA helicase DHX8.

Mol	Chain	Residues	Atoms		AltConf	Trace
31	q	627	Total	C	5	627
			627	627		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
q	?	-	MET	deletion	UNP Q14562
q	?	-	ASP	deletion	UNP Q14562
q	948	GLY	ALA	conflict	UNP Q14562
q	949	SER	PRO	conflict	UNP Q14562

- Molecule 32 is a RNA chain called MINX pre-mRNA (intron).

Mol	Chain	Residues	Atoms					AltConf	Trace
32	Y	32	Total	C	N	O	P	0	0
			677	303	115	227	32		
32	Z	12	Total	C	N	O	P	0	0
			256	114	47	83	12		

- Molecule 33 is a RNA chain called Human gene for small nuclear RNA U2 (snRNA U2).

Mol	Chain	Residues	Atoms					AltConf	Trace
33	2	140	Total	C	N	O	P	0	0
			2968	1327	510	991	140		

- Molecule 34 is a RNA chain called Homo sapiens U5 A small nuclear RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
34	5	114	Total	C	N	O	P	0	0
			2397	1074	399	810	114		


- Molecule 35 is a RNA chain called Homo sapiens RNA, U6 small nuclear 1 (RNU6-1), small nuclear RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
35	6	89	Total	C	N	O	P	0	0
			1903	851	350	613	89		

[illegible]

PHE MET GLY VAL ARG HIS ASP PRO PRO MET TYR TYR GLU LEU LEU GLN ALA ASN PRO PRO LYS GLU PHE TYR HIS HIS HIS ARG PRO PRO SER SER PHE LEU LEU LEU LEU LEU LEU GLN GLU GLY GLY VAL TYR TYR SER ALA ASP ARG ASP GLU ASP LEU LEU TYR

- Molecule 2: 116 kDa U5 small nuclear ribonucleoprotein component

Chain B:  77% 15% 7%

P927	T720	T500	D833	M116	ASP	MET
F933	K721	I501	I347	D117	THR	
T937	P742	P506	Y348	F118	LEU	
V947	I743	V507	F349	I119	LEU	
	V746	K508	R352	A120	T774	
	D764	S522	F356	D121	ASP	
LEU	Q768	Q823	T357	L122	GLU	
LEU	G777	I524	F369	L134	GLY	
ALA	E782	L530	V370	D147	ASN	
LVS	E783	W531	F372	E151	ILE	
GLN	W786	I532	I373	E151	LEU	
ASP	V787	H538	I374	P155	ASP	
VAL	F789	I539	E375	R177	SSR	
VAL	K788	E540	V385	G178	GLU	
LEU	F789	V541	L392	V179	ASP	
ASN	K790	N542	L396	G180	ASP	
TYR	H828	R543	L396	I181	ASP	
PRO	E829	L551	C420	S183	GLU	
MET	T831	L565	K421	T187	GLY	
	R832	S566	G425	P191	ARG	
	F833	E567	G425	S197	GLU	
	V834	P568	G425	S197	THR	
	E835	L588	F430	S197	LVS	
	R836	K589	M433	N201	ASP	
	A838	V592	C434	P206	LEU	
	H856	I629	P441	G207	GLU	
	Y857	L630	A445	H208	MET	
	T858	C639	V457	S212	ASP	
	I863	H642	D467	V225	ASP	
	S866	D643	C468	V226	ASP	
	F867	K646	D469	V253	ASP	
	L868	M647	P470	T254	VAL	
	Y869	V655	D471	V255	GLY	
	T870	C663	L474	D261	ASP	
	A873	C663	M475	L265	ASP	
	F874	A689	H477	G287	ASP	
	A895	L898	T479	R291	HIS	
	S897	S899	G494	V307	PRO	
	L898	V704	R495	V307	GLY	
	S899	G715	R495	A328	MET	
	V900	F715	G494	V507	ASP	
		F716	R495	V307	ASP	
		F717	R495	V307	HIS	
		F718	S498	A328	PRO	
		F718	S498	A328	GLY	
		F718	S498	A328	MET	
		F718	S498	A328	THR	

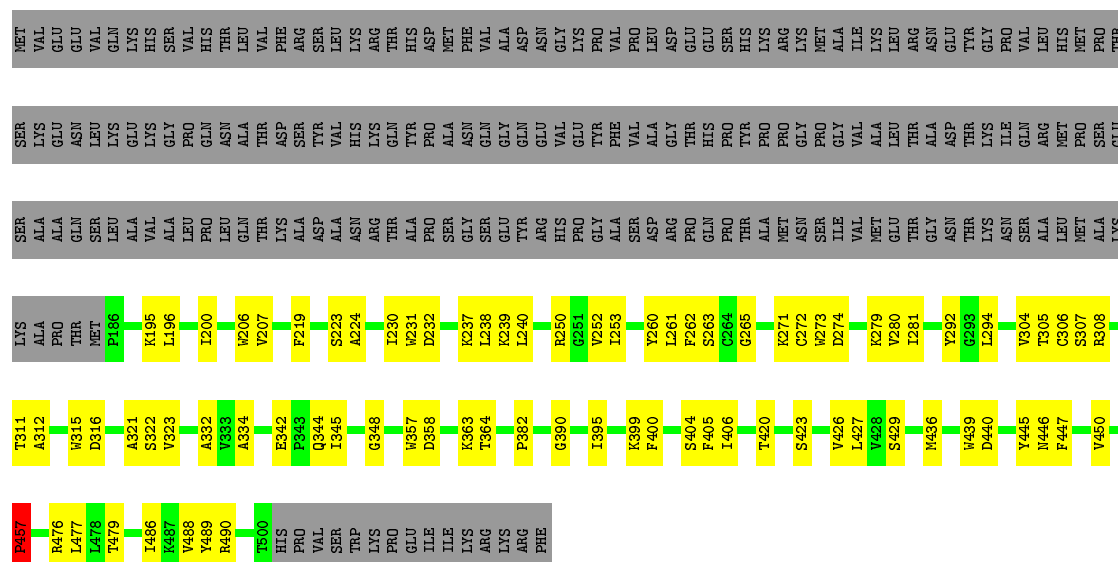
- Molecule 3: SNW domain-containing protein 1

Chain C: 

ASP	TRP	ARG	P209	GLY	MET
PRO	ARG	SER	P210	GLY	LEU
PHE	GLY	LYS	R220	ALA	LEU
GLY	GLY	LEU	R221	PHE	THR
LEU	LYS	GLN	P222	PRO	SER
ASP	ASP	ARG		GLU	PHE
LYS	MET	ASN		ILE	LEU
PHE	ALA	GLU	Y229	VAL	PRO
LEU	GLN	ASN		VAL	ALA
GLU	SER	ARG	T238	ALA	PRO
GLU	ILE	ASP		GLN	THR
ALA	TYR	ILE	T262	TYR	GLN
LYS	ARG	SER	P263	PRO	LEU
GLN	PRO	GLU	L264	LEU	SER
HIS	SER	VAL	D265	ASP	GLN
GLY	LYS	ILE	K266	MET	ASP
GLY	ASN	ALA		GLY	GLN
SER	LEU	LEU	R334	ARG	LEU
LYS	ASP	GLY	ARG	LYS	GLU
ARG	LYS	VAL	ALA	LYS	ALA
PRO	PRO	PRO	GLY	LYS	GLU
SER	MET	ASN	ILE	MET	GLU
SER	TYR	PRO	LYS	SER	LYS
SER	GLY	ARG	THR	ASN	ALA
ASP	ASP	THR	HIS	ASN	ARG
ARG	ASP	SER	VAL	LEU	SER
PRO	LEU	ASN	GLU	GLU	GLN
LYS	GLU	GLU	LYS	ILE	ARG
GLU	ALA	VAL	GLU	GLN	SER
HIS	ARG	GLN	ASP	VAL	ARG
GLU	ILE	TYR	GLY	ASP	GLN
HIS	LYS	ASP	GLU	GLY	THR
GLY	THR	GLN	ALA	GLU	SER
GLU	ASN	ARG	ARG	GLY	LEU
LYS	ARG	LEU	GLU	ARG	VAL
LYS	PHE	PHE	ARG	R95	SER
ARG	VAL	ASN	ASP	ARG	SER
ARG	PRO	GLN	GLU	T112	ARG
LYS	ASP	SER	ILE	R148	ARG
GLU	LYS	LYS	ARG		GLU
	GLU	GLY	HIS	L151	PRO
	PHE	MET	ASP	E152	PRO
	SER	ASP	ARG		PRO
	GLY	SER	ARG	V155	TYR
	SER	SER	LYS	S156	GLY
	ASP	PHE	GLU		TYR
	ARG	ALA	ARG	V159	ARG
	ARG	GLY	GLN	A160	LYS
	GLN	GLY	HIS		GLY
	ARG	GLU	ASP	M163	TRP
	GLY	ASP	ARG	P164	ILE
	ARG	GLU	ASN	V165	PRO
	GLU	ILE	LEU		ARG
	GLY	TYR	SER	A174	LEU
	PRO	ASN	ARG	Q175	LEU
	VAL	VAL	ALA		GLU
	GLN	TYR	ALA	A192	ASP
	PHE	ASP	PRO		PHE
	GLU	GLN	ASP	H199	GLY

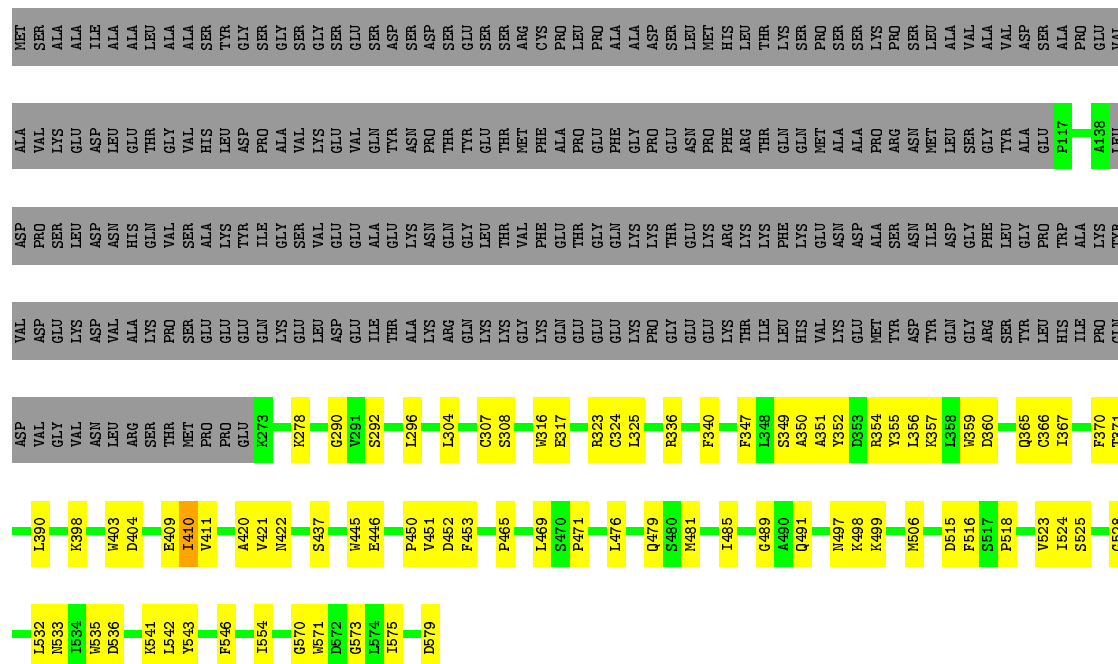
- Molecule 4: Pleiotropic regulator 1

Chain D:  45% 16% 39%



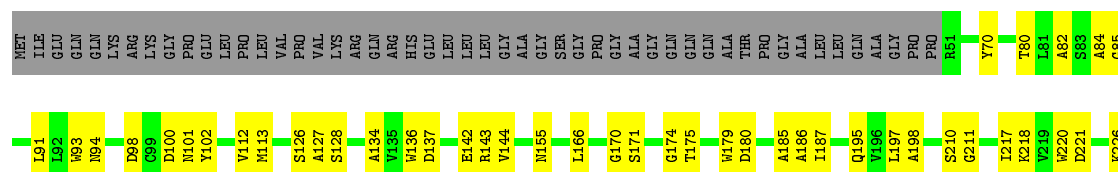
• Molecule 5: Pre-mRNA-processing factor 17

Chain E: 43% 14% 43%



• Molecule 6: U5 small nuclear ribonucleoprotein 40 kDa protein

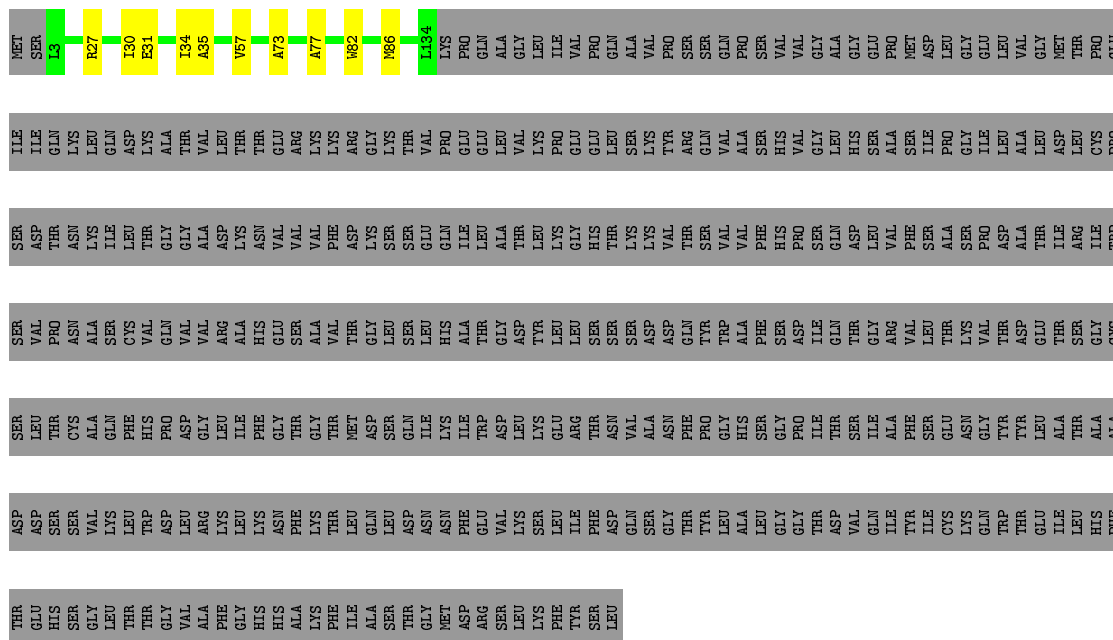
Chain F: 65% 21% 14%





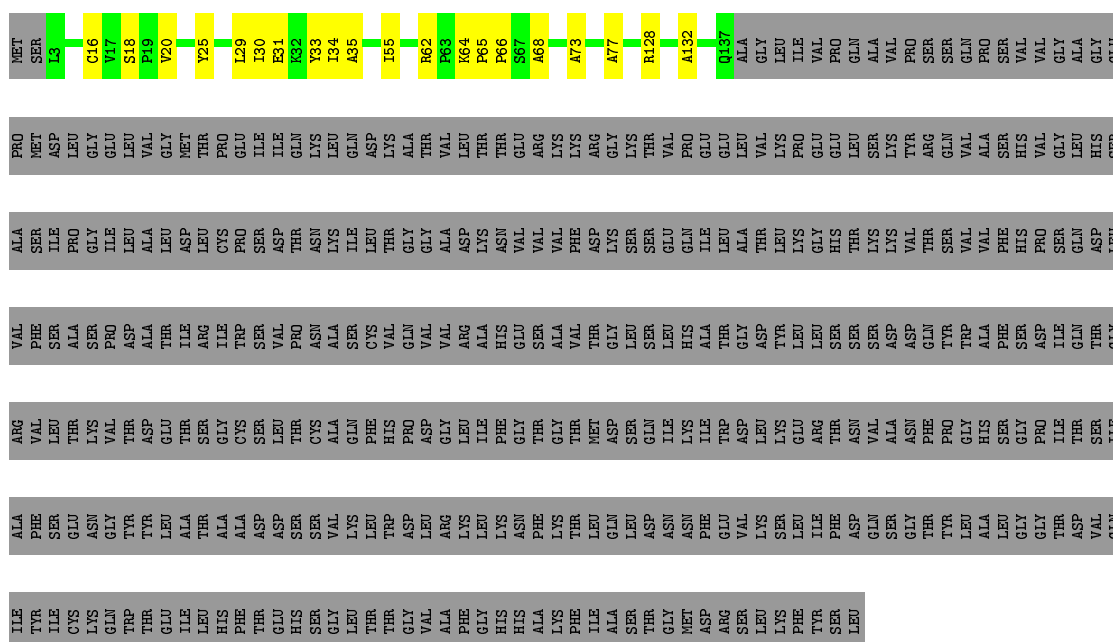
- Molecule 7: Pre-mRNA-processing factor 19

Chain G:  24% . 74%

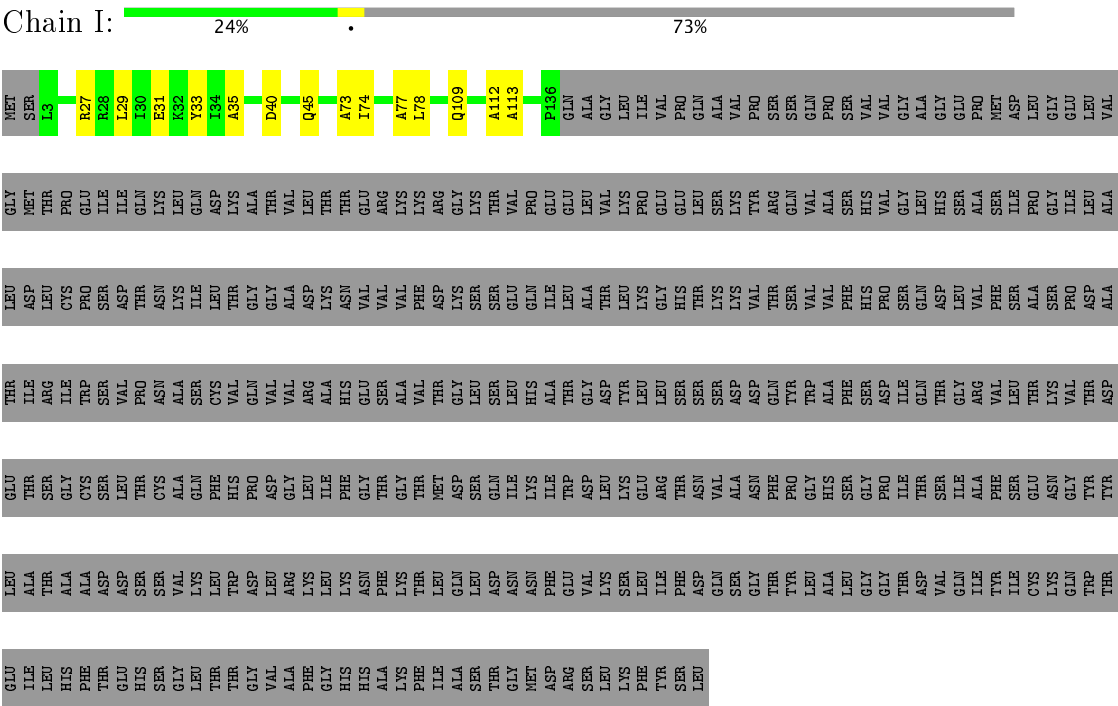


- Molecule 7: Pre-mRNA-processing factor 19

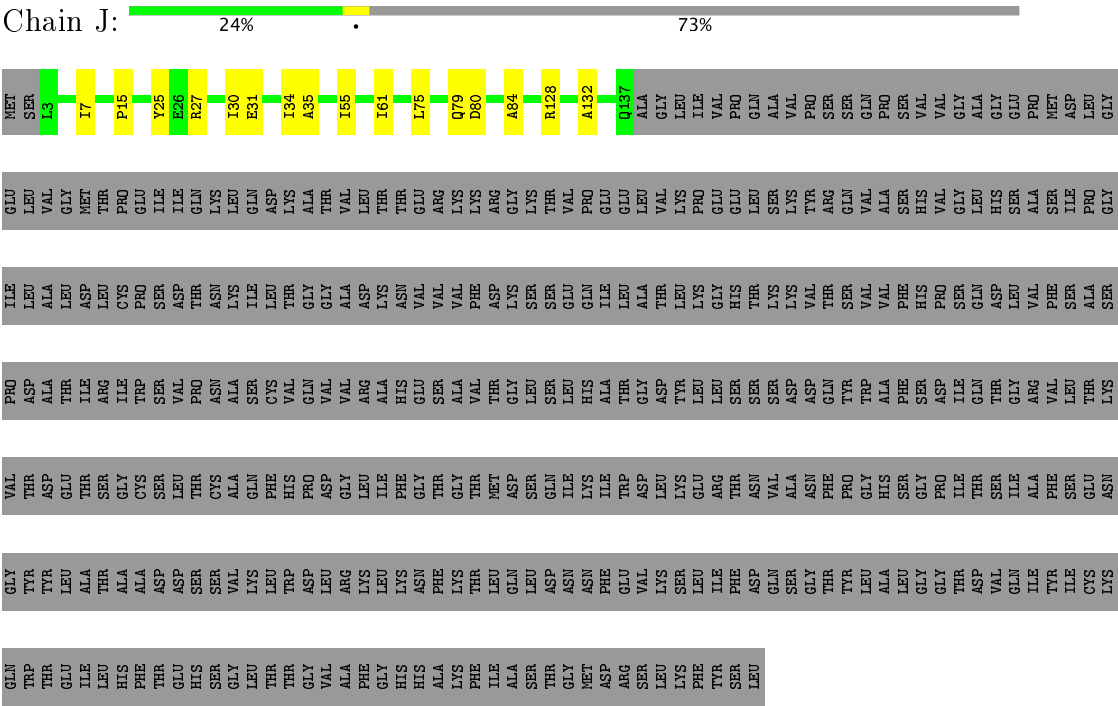
Chain H:  23% 1% 73%



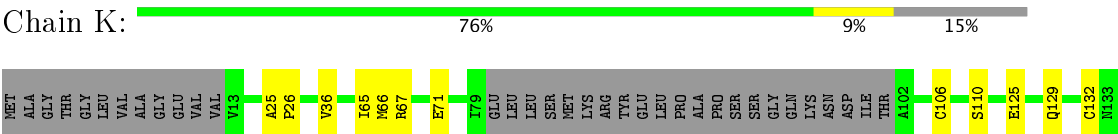
- Molecule 7: Pre-mRNA-processing factor 19

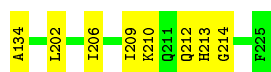


● Molecule 7: Pre-mRNA-processing factor 19



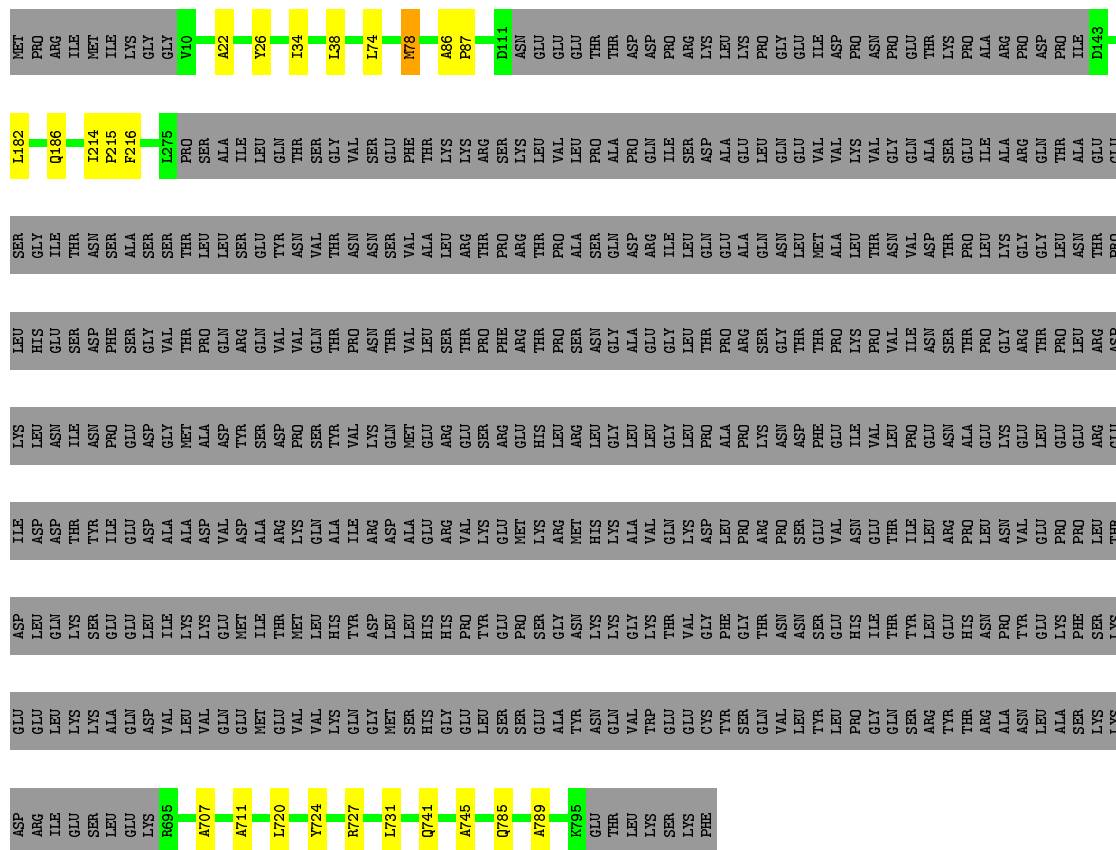
● Molecule 8: Pre-mRNA-splicing factor SPF27





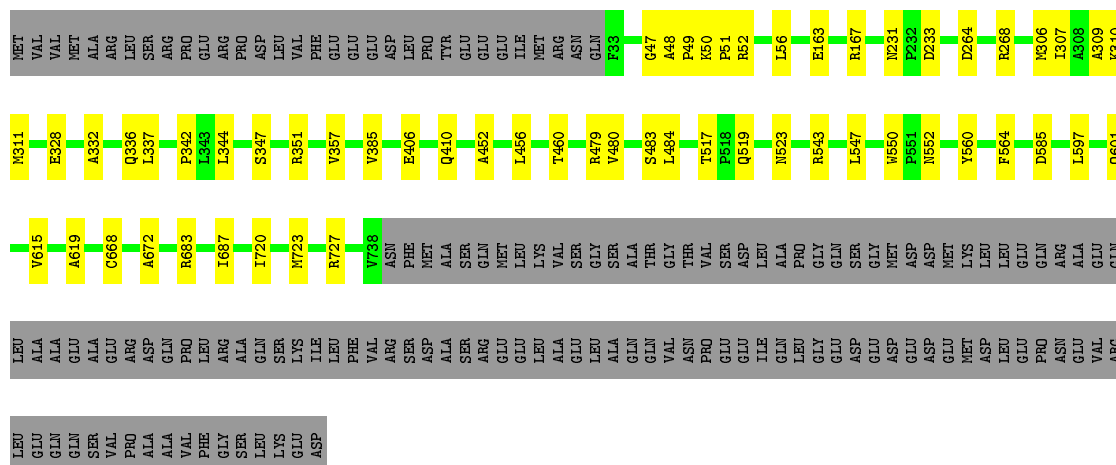
• Molecule 9: Cell division cycle 5-like protein

Chain L: 39% 58%



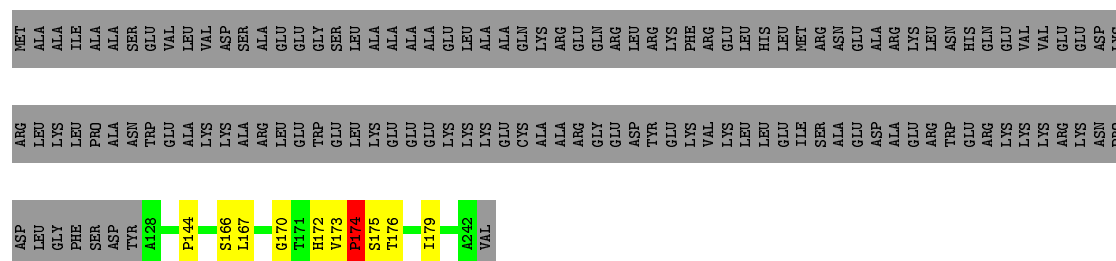
• Molecule 10: Pre-mRNA-splicing factor SYF1

Chain M: 76% 7% 17%



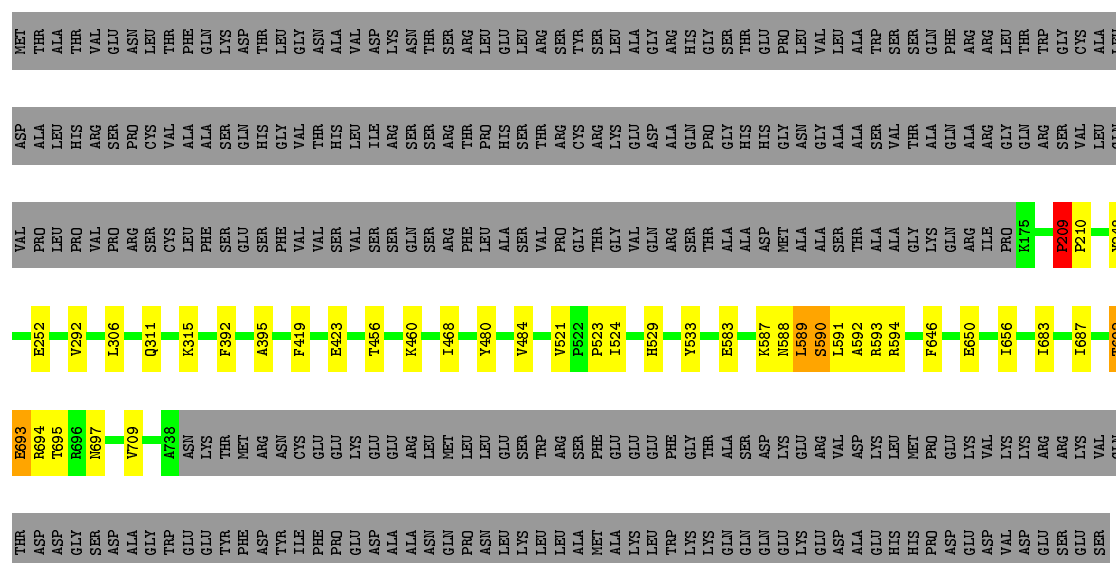
- Molecule 11: Pre-mRNA-splicing factor SYF2

Chain N:  43% . 53%



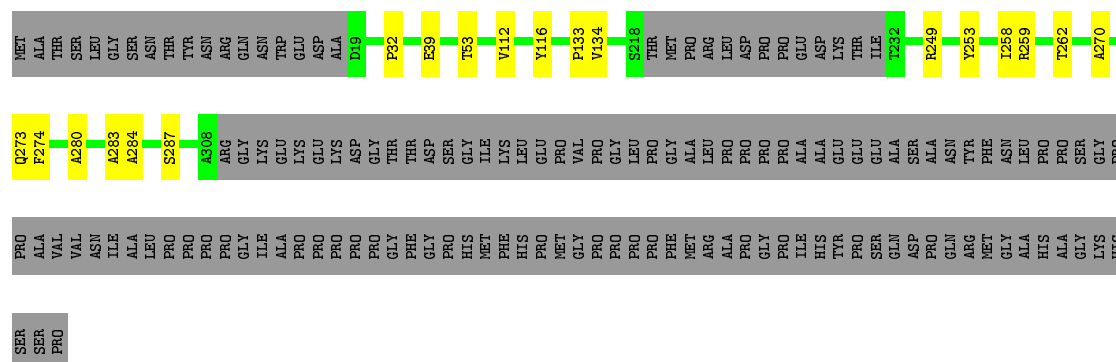
- Molecule 12: Crooked neck-like protein 1

Chain 0:  62% 0 33%



- Molecule 13: Pre-mRNA-splicing factor RBM22

Chain P:  61% 5% 34%



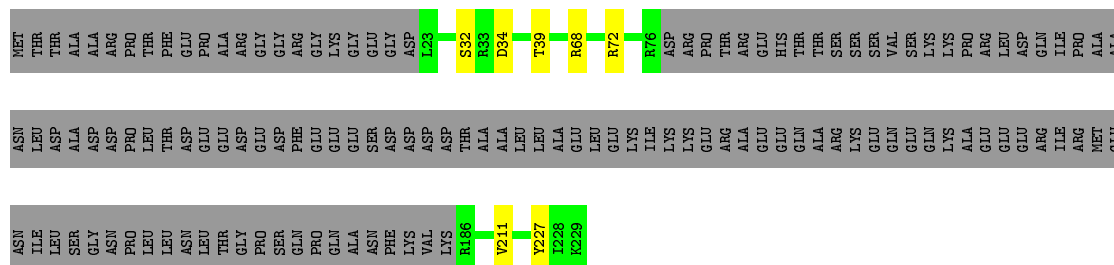
- Molecule 14: Protein BUD31 homolog

Chain Q: 95%



- Molecule 15: Spliceosome-associated protein CWC15 homolog

Chain R: 40% 1% 57%



- Molecule 16: Serine/arginine repetitive matrix protein 2

Chain S:  99%



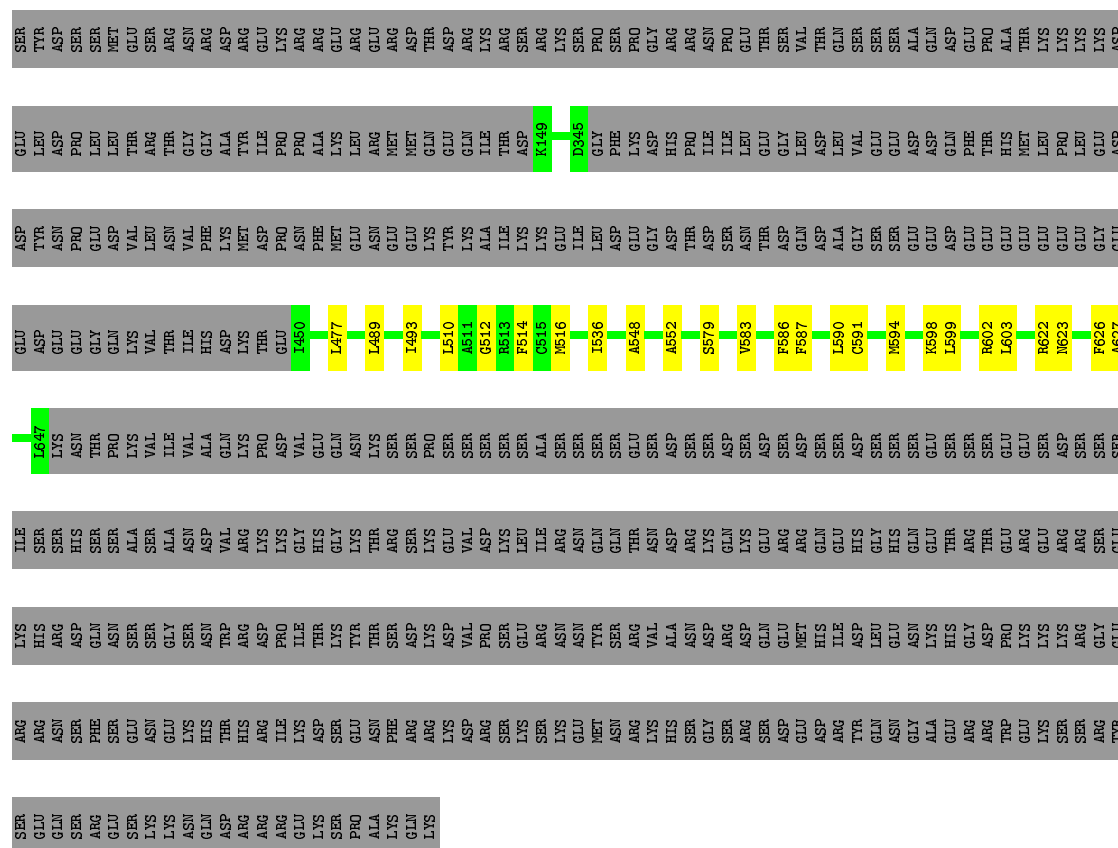



[illegible]

- Molecule 17: Pre-mRNA-splicing factor CWC22 homolog

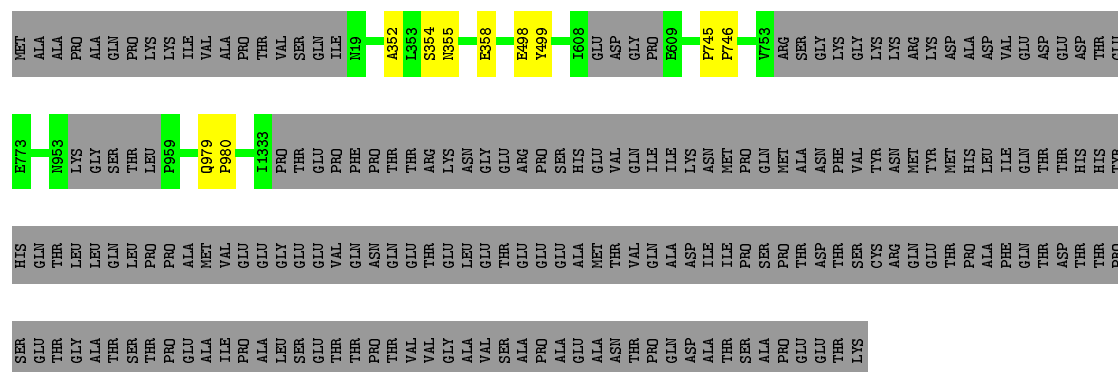
Chain T:  41% . 56%

MET	LYS	SER	SER	SER	VAL	ALA	GLN	ILE	GLU	LYS	PRO	SER	SER	GLY	HIS	ASP	ARG	ARG	GLU	ASN	GLU	ARG	GLN	TYR	SER	LEU	ASN	ASP	HIS	SER	SER	ASN	ARG	PRO	GLU	GLU	ASP	ARG	TYR	GLU	GLU	GLN	GLN	GLU	GLU	GLU	ASP	ASP	ASP	PHE	TYR	ARG	SER	PRO	ASP	ASP	ASP	ARG	TYR	GLU	GLU	GLU	HIS	SER	GLU	GLY	ARG	ARG	ARG	GLY	ASP	ARG
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----



• Molecule 18: Intron-binding protein aquarius

Chain U: 86% 13%



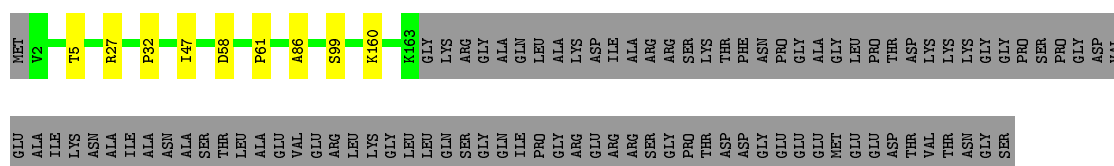
• Molecule 19: Peptidyl-prolyl cis-trans isomerase-like 1

Chain V: 82% 11% 7%

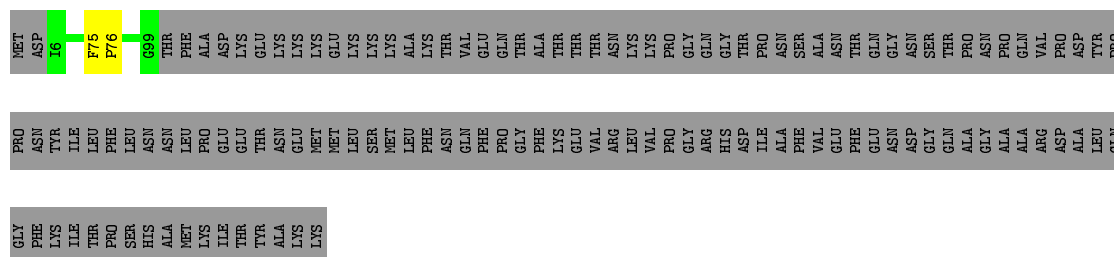


• Molecule 20: U2 small nuclear ribonucleoprotein A'

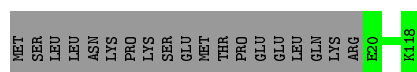
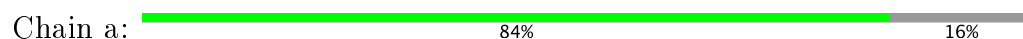
Chain W: 60% 36%



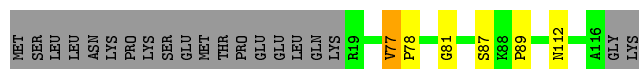
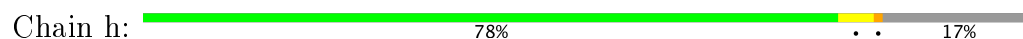
- Molecule 21: U2 small nuclear ribonucleoprotein B''



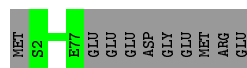
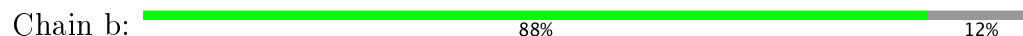
- Molecule 22: Small nuclear ribonucleoprotein Sm D2



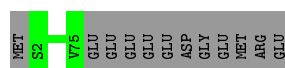
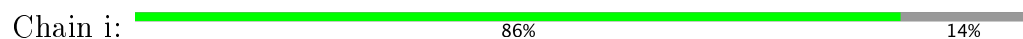
- Molecule 22: Small nuclear ribonucleoprotein Sm D2



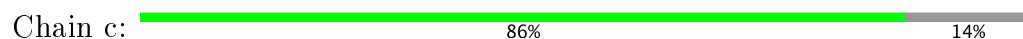
- Molecule 23: Small nuclear ribonucleoprotein F

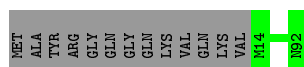


- Molecule 23: Small nuclear ribonucleoprotein F

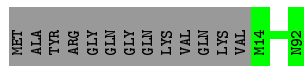
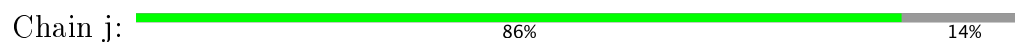


- Molecule 24: Small nuclear ribonucleoprotein E





- Molecule 24: Small nuclear ribonucleoprotein E



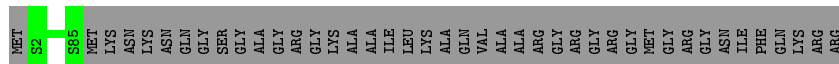
- Molecule 25: Small nuclear ribonucleoprotein G



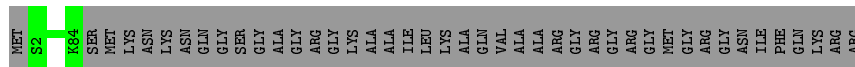
- Molecule 25: Small nuclear ribonucleoprotein G



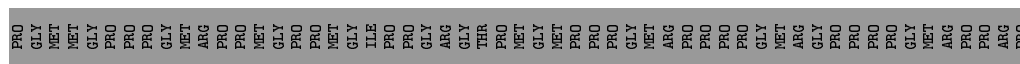
- Molecule 26: Small nuclear ribonucleoprotein Sm D3



- Molecule 26: Small nuclear ribonucleoprotein Sm D3

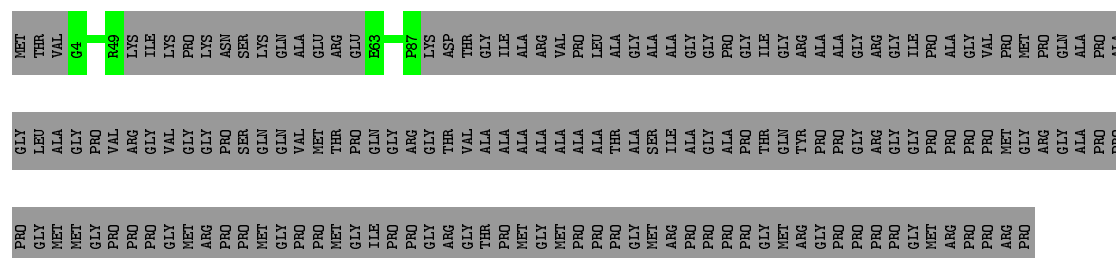


- Molecule 27: Small nuclear ribonucleoprotein-associated proteins B and B'



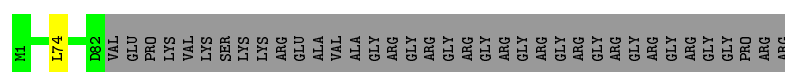
- Molecule 27: Small nuclear ribonucleoprotein-associated proteins B and B'

Chain m:  30% 70%



- Molecule 28: Small nuclear ribonucleoprotein Sm D1

Chain g: 68% 31%



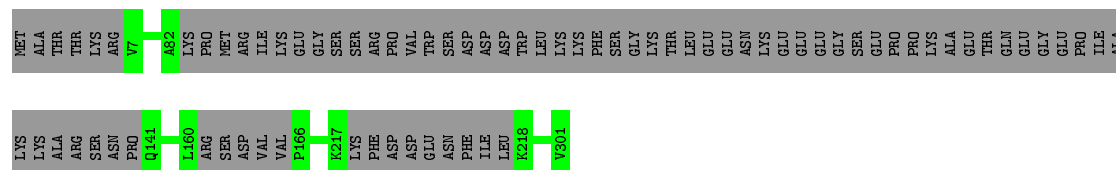
- Molecule 28: Small nuclear ribonucleoprotein Sm D1

Chain n: 69% 31%



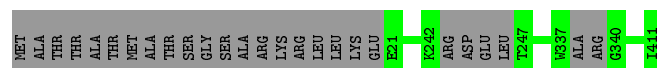
- Molecule 29: Peptidyl-prolyl cis-trans isomerase E

Chain o: 74% 26%



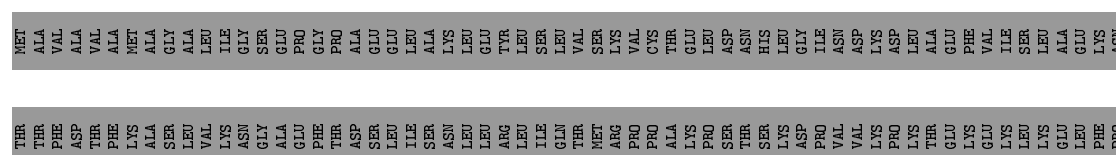
- Molecule 30: Eukaryotic initiation factor 4A-III

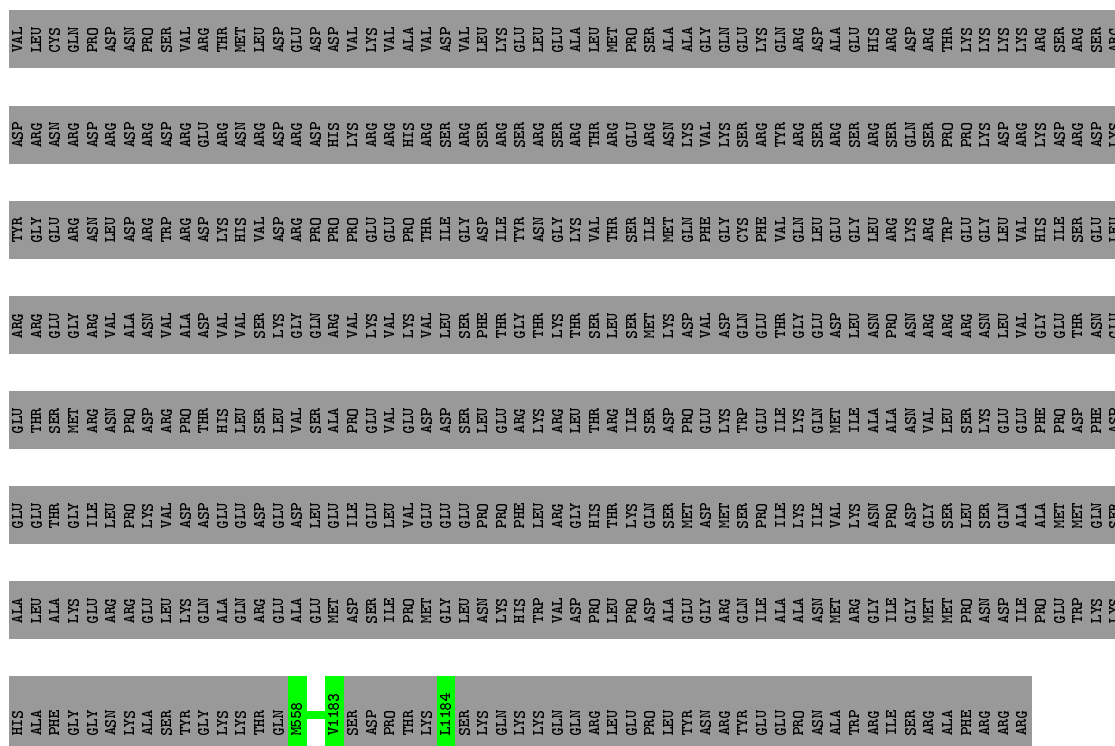
Chain p:  94% 6%



- Molecule 31: ATP-dependent RNA helicase DHX8

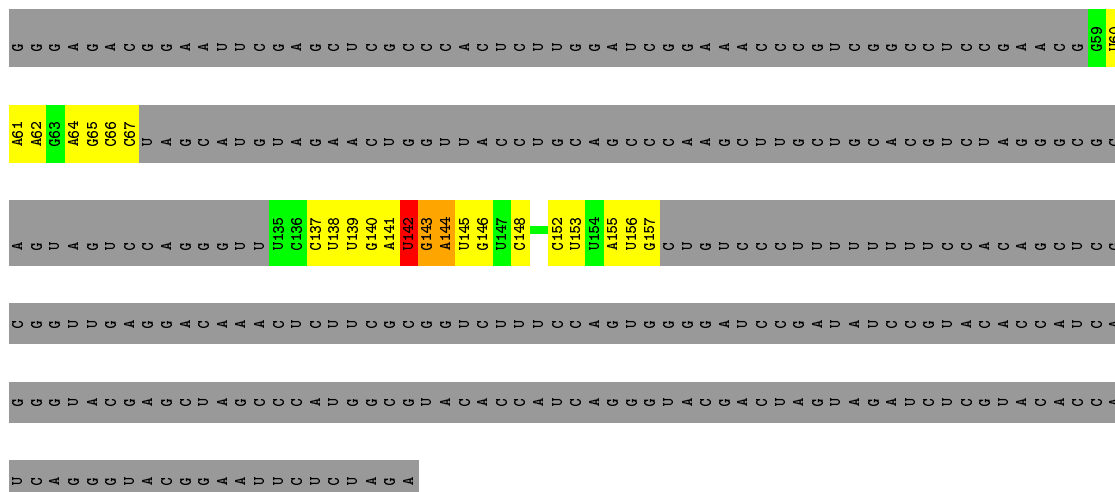
Chain q: 51% 49%





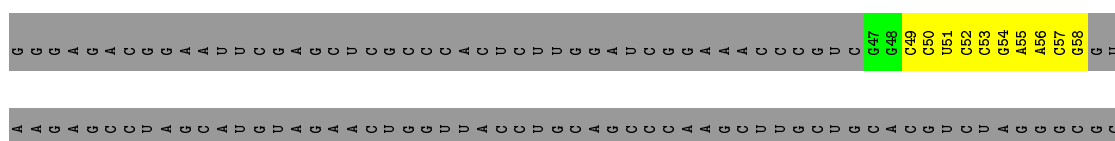
- Molecule 32: MINX pre-mRNA (intron)

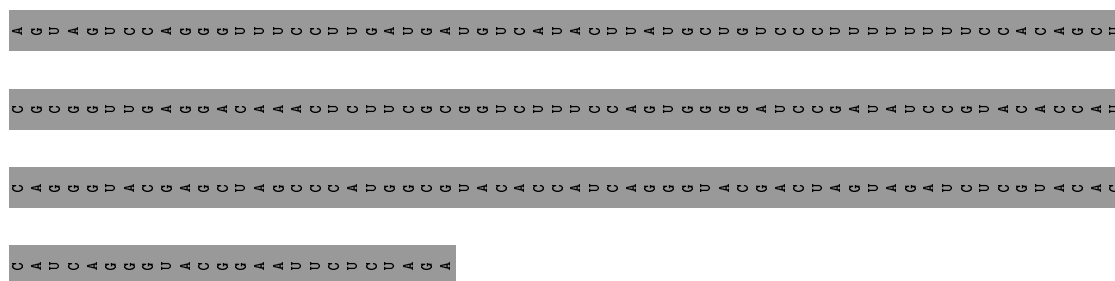
Chain Y: 6% 90%



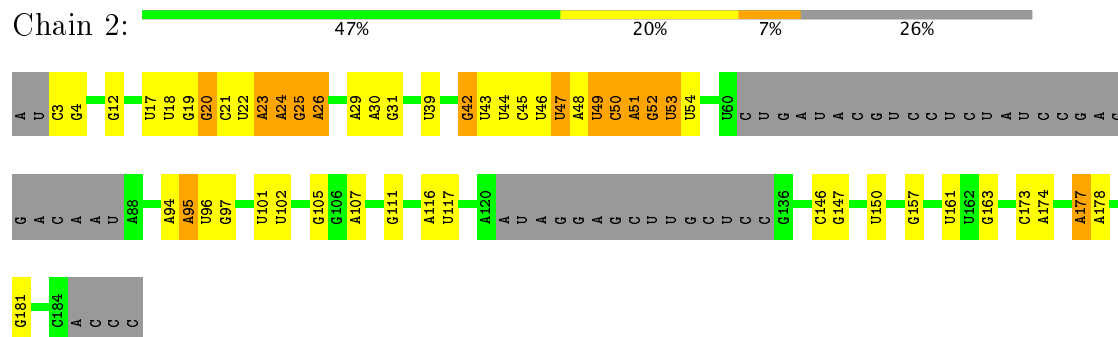
- Molecule 32: MINX pre-mRNA (intron)

Chain Z: 96%

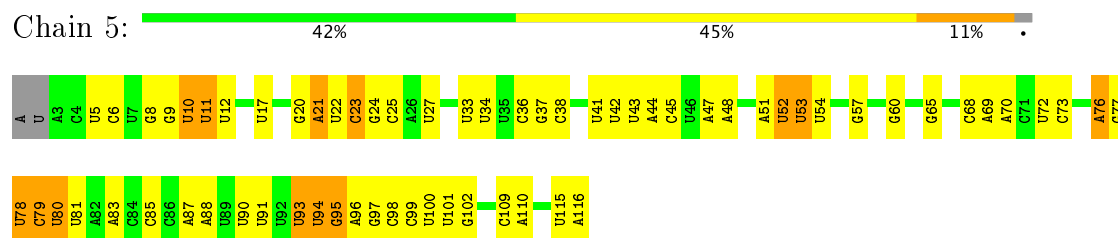




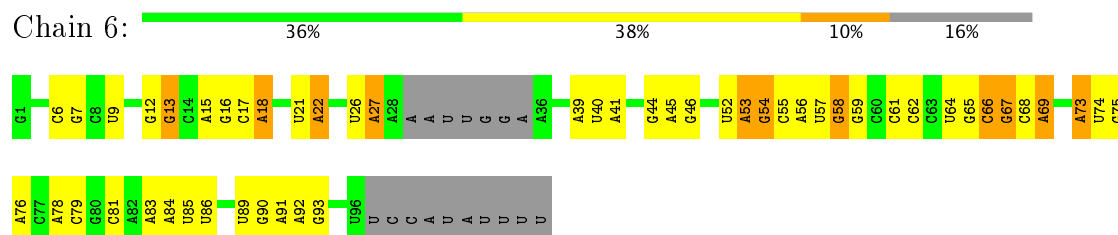
- Molecule 33: Human gene for small nuclear RNA U2 (snRNA U2)



- Molecule 34: Homo sapiens U5 A small nuclear RNA



- Molecule 35: Homo sapiens RNA, U6 small nuclear 1 (RNU6-1), small nuclear RNA



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	136534	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	2.1	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	FEI FALCON II (4k x 4k)	Depositor

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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 >2	RMSZ	# Z >2
1	A	0.34	5/10059 (0.0%)	0.46	2/14115 (0.0%)
10	M	0.23	0/3575	0.43	0/5006
11	N	0.27	0/582	0.52	0/814
12	O	0.50	1/2858 (0.0%)	0.42	0/4007
13	P	0.24	0/1408	0.44	0/1969
14	Q	0.23	0/718	0.39	0/1003
15	R	0.22	0/493	0.40	0/688
16	S	0.28	0/149	0.52	0/205
17	T	0.22	0/999	0.39	0/1397
19	V	0.25	0/782	0.45	0/1088
2	B	0.24	0/4619	0.44	1/6475 (0.0%)
20	W	0.64	0/821	1.48	5/1149 (0.4%)
21	X	0.61	0/472	1.26	0/658
22	a	0.23	0/500	0.47	0/698
22	h	0.62	0/496	1.00	1/693 (0.1%)
23	b	0.23	0/382	0.44	0/531
23	i	0.67	0/372	0.78	0/517
24	c	0.22	0/393	0.43	0/547
24	j	0.64	0/393	0.81	0/547
25	d	0.23	0/371	0.47	0/516
25	k	0.50	0/371	0.69	0/516
26	e	0.23	0/422	0.50	0/588
26	l	0.46	0/417	0.60	0/581
27	f	0.23	0/356	0.47	0/494
27	m	0.46	0/356	0.63	0/494
28	g	0.41	1/414 (0.2%)	0.48	0/578
28	n	0.55	0/414	0.71	0/578
3	C	0.27	0/1249	0.52	0/1758
32	Y	0.45	1/754 (0.1%)	0.83	0/1169
32	Z	0.16	0/285	0.71	0/442
33	2	0.15	0/3312	0.68	0/5151
34	5	0.16	0/2672	0.76	0/4154
35	6	0.19	0/2129	0.78	0/3314
4	D	0.25	0/1587	0.55	1/2215 (0.0%)

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >2	RMSZ	# Z >2
5	E	0.25	0/1655	0.49	0/2309
6	F	0.25	0/1540	0.49	0/2148
7	G	0.23	0/688	0.38	0/969
7	H	0.26	0/706	0.40	0/995
7	I	0.23	0/701	0.38	0/988
7	J	0.25	0/706	0.39	0/995
8	K	0.23	0/963	0.36	0/1346
9	L	0.23	0/1683	0.40	0/2349
All	All	0.32	8/53822 (0.0%)	0.57	10/76754 (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
1	A	0	4
10	M	0	2
11	N	0	2
12	O	0	2
2	B	0	1
22	h	0	1
4	D	0	1
5	E	0	1
6	F	0	1
9	L	0	1
All	All	0	16

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
12	O	209	PRO	C-N	23.47	1.78	1.34
1	A	1134	TRP	C-N	15.33	1.63	1.34
1	A	323	LEU	C-N	10.94	1.55	1.34
1	A	369	GLU	C-N	9.57	1.52	1.34
1	A	174	PRO	C-N	9.19	1.51	1.34

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
22	h	77	VAL	C-N-CD	-17.59	81.91	120.60
2	B	349	PHE	C-N-CA	6.40	137.71	121.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
20	W	5	THR	N-CA-CB	-6.30	98.32	110.30
20	W	27	ARG	CB-CA-C	-5.85	98.70	110.40
1	A	1945[A]	VAL	C-N-CA	-5.79	107.23	121.70

There are no chirality outliers.

5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	172	GLU	Peptide
1	A	1852	LEU	Peptide
1	A	1899	VAL	Peptide
1	A	1946	ASN	Peptide
2	B	352	LYS	Peptide

5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	9955	0	4850	175	0
2	B	4566	0	2269	89	0
3	C	1231	0	665	18	0
4	D	1576	0	774	78	0
5	E	1646	0	787	82	0
6	F	1531	0	747	71	0
7	G	679	0	356	7	0
7	H	696	0	367	13	0
7	I	691	0	365	10	0
7	J	696	0	367	11	0
8	K	960	0	461	14	0
9	L	1680	0	807	12	0
10	M	3553	0	1738	42	0
11	N	579	0	277	5	0
12	O	2842	0	1355	20	0
13	P	1396	0	684	11	0
14	Q	713	0	337	3	0
15	R	493	0	220	5	0
16	S	148	0	78	29	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
17	T	1191	0	457	15	0
18	U	1287	0	0	20	0
19	V	774	0	404	17	0
20	W	816	0	386	1	0
21	X	470	0	223	1	0
22	a	497	0	229	0	0
22	h	493	0	226	0	0
23	b	379	0	186	0	0
23	i	369	0	182	0	0
24	c	393	0	169	0	0
24	j	393	0	169	0	0
25	d	369	0	178	0	0
25	k	369	0	178	0	0
26	e	420	0	200	0	0
26	l	415	0	196	0	0
27	f	355	0	164	0	0
27	m	355	0	164	0	0
28	g	412	0	188	0	0
28	n	412	0	188	0	0
29	o	223	0	0	0	0
30	p	385	0	0	0	0
31	q	627	0	0	0	0
32	Y	677	0	342	8	0
32	Z	256	0	132	11	0
33	2	2968	0	1494	41	0
34	5	2397	0	1216	32	0
35	6	1903	0	964	28	0
All	All	55236	0	25739	813	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 11.

The worst 5 of 813 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1355:SER:CB	16:S:17:GLY:HA3	1.26	1.57
1:A:1355:SER:CB	16:S:17:GLY:CA	1.88	1.49
10:M:309:ALA:HB3	18:U:358:GLU:CA	1.42	1.48
10:M:309:ALA:CB	18:U:358:GLU:CA	1.91	1.45
1:A:1355:SER:CA	16:S:17:GLY:HA3	1.43	1.43

There are no symmetry-related clashes.

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	
1	A	1961/2335 (84%)	1618 (82%)	328 (17%)	15 (1%)	22	66
2	B	898/972 (92%)	778 (87%)	108 (12%)	12 (1%)	14	56
3	C	238/536 (44%)	193 (81%)	44 (18%)	1 (0%)	38	77
4	D	313/515 (61%)	282 (90%)	30 (10%)	1 (0%)	44	81
5	E	325/579 (56%)	287 (88%)	37 (11%)	1 (0%)	44	81
6	F	305/357 (85%)	274 (90%)	30 (10%)	1 (0%)	44	81
7	G	130/504 (26%)	120 (92%)	9 (7%)	1 (1%)	22	66
7	H	133/504 (26%)	120 (90%)	12 (9%)	1 (1%)	22	66
7	I	132/504 (26%)	122 (92%)	10 (8%)	0	100	100
7	J	133/504 (26%)	116 (87%)	14 (10%)	3 (2%)	7	43
8	K	187/225 (83%)	167 (89%)	19 (10%)	1 (0%)	32	74
9	L	330/802 (41%)	297 (90%)	32 (10%)	1 (0%)	44	81
10	M	703/855 (82%)	600 (85%)	97 (14%)	6 (1%)	20	63
11	N	113/243 (46%)	87 (77%)	23 (20%)	3 (3%)	6	40
12	O	562/848 (66%)	476 (85%)	78 (14%)	8 (1%)	13	53
13	P	273/420 (65%)	232 (85%)	39 (14%)	2 (1%)	25	68
14	Q	140/144 (97%)	127 (91%)	12 (9%)	1 (1%)	25	68
15	R	94/229 (41%)	84 (89%)	10 (11%)	0	100	100
16	S	28/2752 (1%)	20 (71%)	5 (18%)	3 (11%)	0	10
17	T	196/908 (22%)	177 (90%)	18 (9%)	1 (0%)	32	74
19	V	153/166 (92%)	145 (95%)	7 (5%)	1 (1%)	25	68
20	W	160/255 (63%)	146 (91%)	12 (8%)	2 (1%)	14	56

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
21	X	92/225 (41%)	90 (98%)	2 (2%)	0	100	100
22	a	97/118 (82%)	95 (98%)	2 (2%)	0	100	100
22	h	96/118 (81%)	83 (86%)	8 (8%)	5 (5%)	2	26
23	b	74/86 (86%)	72 (97%)	2 (3%)	0	100	100
23	i	72/86 (84%)	69 (96%)	3 (4%)	0	100	100
24	c	77/92 (84%)	75 (97%)	2 (3%)	0	100	100
24	j	77/92 (84%)	75 (97%)	2 (3%)	0	100	100
25	d	72/76 (95%)	71 (99%)	1 (1%)	0	100	100
25	k	72/76 (95%)	70 (97%)	2 (3%)	0	100	100
26	e	82/126 (65%)	81 (99%)	1 (1%)	0	100	100
26	l	81/126 (64%)	80 (99%)	1 (1%)	0	100	100
27	f	67/240 (28%)	63 (94%)	4 (6%)	0	100	100
27	m	67/240 (28%)	66 (98%)	1 (2%)	0	100	100
28	g	80/119 (67%)	76 (95%)	4 (5%)	0	100	100
28	n	80/119 (67%)	77 (96%)	3 (4%)	0	100	100
All	All	8693/17096 (51%)	7611 (88%)	1012 (12%)	70 (1%)	27	66

5 of 70 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	988	ILE
1	A	1777	ILE
2	B	182	LYS
2	B	947	VAL
10	M	720	ILE

5.3.2 Protein sidechains ⓘ

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	
1	A	106/2108 (5%)	106 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	55/866 (6%)	55 (100%)	0	100	100
3	C	19/459 (4%)	19 (100%)	0	100	100
4	D	12/441 (3%)	12 (100%)	0	100	100
5	E	11/502 (2%)	11 (100%)	0	100	100
6	F	10/300 (3%)	10 (100%)	0	100	100
7	G	10/435 (2%)	10 (100%)	0	100	100
7	H	11/435 (2%)	11 (100%)	0	100	100
7	I	11/435 (2%)	11 (100%)	0	100	100
7	J	11/435 (2%)	11 (100%)	0	100	100
8	K	5/196 (3%)	5 (100%)	0	100	100
9	L	6/709 (1%)	6 (100%)	0	100	100
10	M	24/749 (3%)	24 (100%)	0	100	100
11	N	4/209 (2%)	4 (100%)	0	100	100
12	O	17/751 (2%)	17 (100%)	0	100	100
13	P	14/361 (4%)	14 (100%)	0	100	100
14	Q	6/130 (5%)	6 (100%)	0	100	100
15	R	2/203 (1%)	2 (100%)	0	100	100
16	S	2/2432 (0%)	2 (100%)	0	100	100
17	T	6/838 (1%)	6 (100%)	0	100	100
19	V	9/134 (7%)	9 (100%)	0	100	100
20	W	6/218 (3%)	6 (100%)	0	100	100
21	X	3/195 (2%)	3 (100%)	0	100	100
22	a	4/110 (4%)	4 (100%)	0	100	100
22	h	4/110 (4%)	4 (100%)	0	100	100
23	b	4/74 (5%)	4 (100%)	0	100	100
23	i	4/74 (5%)	4 (100%)	0	100	100
24	c	1/84 (1%)	1 (100%)	0	100	100
24	j	1/84 (1%)	1 (100%)	0	100	100
25	d	3/66 (4%)	3 (100%)	0	100	100
25	k	3/66 (4%)	3 (100%)	0	100	100
26	e	3/101 (3%)	3 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
26	l	3/101 (3%)	3 (100%)	0	100	100
27	f	3/177 (2%)	3 (100%)	0	100	100
27	m	3/177 (2%)	3 (100%)	0	100	100
28	g	3/101 (3%)	3 (100%)	0	100	100
28	n	3/101 (3%)	3 (100%)	0	100	100
All	All	402/14967 (3%)	402 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
32	Y	30/324 (9%)	21 (70%)	1 (3%)
32	Z	11/324 (3%)	9 (81%)	0
33	2	137/188 (72%)	34 (24%)	2 (1%)
34	5	113/116 (97%)	42 (37%)	2 (1%)
35	6	87/106 (82%)	34 (39%)	1 (1%)
All	All	378/1058 (35%)	140 (37%)	6 (1%)

5 of 140 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
32	Y	60	U
32	Y	61	A
32	Y	62	A
32	Y	64	A
32	Y	65	G

5 of 6 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
33	2	52	G
35	6	53	A
34	5	78	U
33	2	23	A
34	5	93	U

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	2
12	O	1
2	B	1
10	M	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	M	33:PHE	C	34:SER	N	4.52
1	B	797:ALA	C	798:GLN	N	3.21
1	A	1914:MET	C	1915:VAL	N	3.11
1	O	209:PRO	C	210:PRO	N	1.78
1	A	1134:TRP	C	1135:PRO	N	1.63