wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Apr 15, 2018 – 01:18 AM EDT

PDB ID	:	6FSZ
EMDB ID:	:	EMD-4301
Title	:	Structure of the nuclear RNA exosome
Authors	:	Schuller, J.M.; Falk, S.; Conti, E.
Deposited on	:	2018-02-20
Resolution	:	4.60 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
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

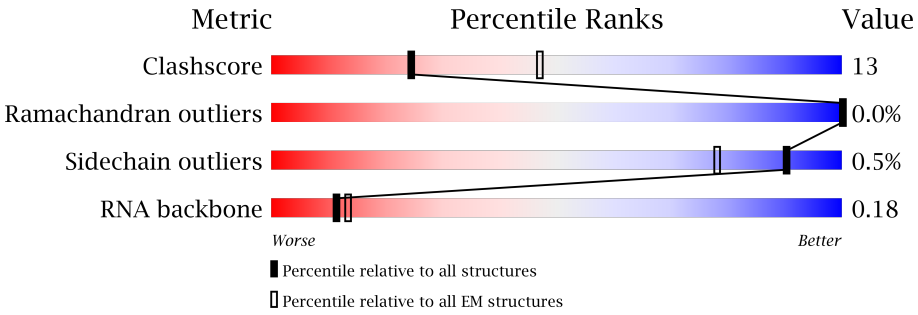
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031021

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 4.60 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531
RNA backbone	3747	458

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 <=5%

Mol	Chain	Length	Quality of chain
1	2	23	35% 43% 22%
2	AA	303	68% 30% .
3	BB	248	68% 30% . .
4	CC	393	60% 25% . 14%
5	DD	245	60% 31% 9%
6	EE	267	69% 31%
7	FF	250	62% 23% 14%
8	GG	242	71% 26% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	HH	361	59% 22% 19%
10	II	301	53% 21% 26%
11	JJ	1003	62% 33% 5%
12	KK	733	47% 10% 44%
13	LL	184	45% 16% 39%
14	MM	1073	59% 32% 9%
15	NN	40	83% 18%

2 Entry composition (i)

There are 15 unique types of molecules in this entry. The entry contains 36972 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 RNA (5'-R(P*AP*AP*AP*AP*UP*UP*UP*AP*AP*AP*UP*UP*UP*UP*UP*UP*UP*UP*UP*UP*UP*UP*U)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
1	2	23	378	166	43	146	23	0	0

- Molecule 2 is a protein called Exosome complex component RRP45.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	AA	299	2304	1444	393	451	16	0	0

- Molecule 3 is a protein called Exosome complex component SKI6.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	BB	244	1886	1177	335	366	8	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
BB	-1	GLY	-	expression tag	UNP P46948
BB	0	HIS	-	expression tag	UNP P46948

- Molecule 4 is a protein called Exosome complex component RRP43.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	CC	339	2589	1640	441	497	11	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
CC	102	SER	ALA	conflict	UNP P25359
CC	363	MET	VAL	conflict	UNP P25359

- Molecule 5 is a protein called Exosome complex component RRP46.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	DD	223	Total	C	N	O	S	1	0
			1701	1072	285	334	10		

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
DD	-21	GLY	-	expression tag	UNP P53256
DD	-20	HIS	-	expression tag	UNP P53256
DD	-19	GLY	-	expression tag	UNP P53256
DD	-18	ASN	-	expression tag	UNP P53256
DD	-17	ASN	-	expression tag	UNP P53256
DD	-16	LYS	-	expression tag	UNP P53256
DD	-15	GLU	-	expression tag	UNP P53256
DD	-14	PRO	-	expression tag	UNP P53256
DD	-13	ASN	-	expression tag	UNP P53256
DD	-12	THR	-	expression tag	UNP P53256
DD	-11	LYS	-	expression tag	UNP P53256
DD	-10	ASN	-	expression tag	UNP P53256
DD	-9	ARG	-	expression tag	UNP P53256
DD	-8	LEU	-	expression tag	UNP P53256
DD	-7	ASP	-	expression tag	UNP P53256
DD	-6	SER	-	expression tag	UNP P53256
DD	-5	ALA	-	expression tag	UNP P53256
DD	-4	GLU	-	expression tag	UNP P53256
DD	-3	LYS	-	expression tag	UNP P53256
DD	-2	LYS	-	expression tag	UNP P53256
DD	-1	LYS	-	expression tag	UNP P53256
DD	0	LYS	-	expression tag	UNP P53256

- Molecule 6 is a protein called Exosome complex component RRP42.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	EE	267	Total	C	N	O	S	1	0
			2050	1308	338	399	5		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
EE	-1	GLY	-	expression tag	UNP Q12277
EE	0	HIS	-	expression tag	UNP Q12277
EE	138	ILE	VAL	conflict	UNP Q12277

- Molecule 7 is a protein called Exosome complex component MTR3.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	FF	215	Total	C	N	O	S	0	0
			1638	1023	273	332	10		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
FF	75	SER	THR	conflict	UNP P48240
FF	161	THR	MET	conflict	UNP P48240

- Molecule 8 is a protein called Exosome complex component RRP40.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	GG	237	Total	C	N	O	S	0	0
			1792	1143	295	344	10		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
GG	-1	GLY	-	expression tag	UNP Q08285
GG	0	HIS	-	expression tag	UNP Q08285

- Molecule 9 is a protein called Exosome complex component RRP4.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	HH	293	Total	C	N	O	S	0	0
			2236	1393	403	428	12		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
HH	-1	ARG	-	expression tag	UNP P38792
HH	0	SER	-	expression tag	UNP P38792

- Molecule 10 is a protein called Exosome complex component CSL4.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	II	222	Total	C	N	O	S	0	0
			1653	1034	287	325	7		

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
II	-8	MET	-	initiating methionine	UNP P53859
II	-7	LYS	-	expression tag	UNP P53859
II	-6	HIS	-	expression tag	UNP P53859
II	-5	HIS	-	expression tag	UNP P53859
II	-4	HIS	-	expression tag	UNP P53859
II	-3	HIS	-	expression tag	UNP P53859
II	-2	HIS	-	expression tag	UNP P53859
II	-1	HIS	-	expression tag	UNP P53859
II	0	PRO	-	expression tag	UNP P53859

- Molecule 11 is a protein called Exosome complex exonuclease DIS3.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	JJ	948	Total	C	N	O	S	0	0
			7430	4693	1310	1392	35		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
JJ	-1	GLY	-	expression tag	UNP Q08162
JJ	0	ALA	-	expression tag	UNP Q08162
JJ	171	ASN	ASP	conflict	UNP Q08162
JJ	551	ASN	ASP	conflict	UNP Q08162

- Molecule 12 is a protein called Exosome complex exonuclease RRP6.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	KK	414	Total	C	N	O	S	0	0
			2517	1544	469	498	6		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
KK	296	ASN	ASP	engineered mutation	UNP Q12149

- Molecule 13 is a protein called Exosome complex protein LRP1.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	LL	113	Total	C	N	O	S	0	0
			894	565	151	174	4		

- Molecule 14 is a protein called ATP-dependent RNA helicase DOB1.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	MM	978	Total	C	N	O	S	0	0
			7627	4871	1294	1419	43		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
MM	80	MET	VAL	conflict	UNP P47047

- Molecule 15 is a protein called M-phase phosphoprotein 6 homolog,M-phase phosphoprotein 6 homolog,Nuclear exosome-associated RNA binding protein,M-phase phosphoprotein 6 homolog.

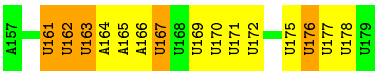
Mol	Chain	Residues	Atoms				AltConf	Trace
15	NN	40	Total	C	N	O	0	0
			277	172	51	54		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

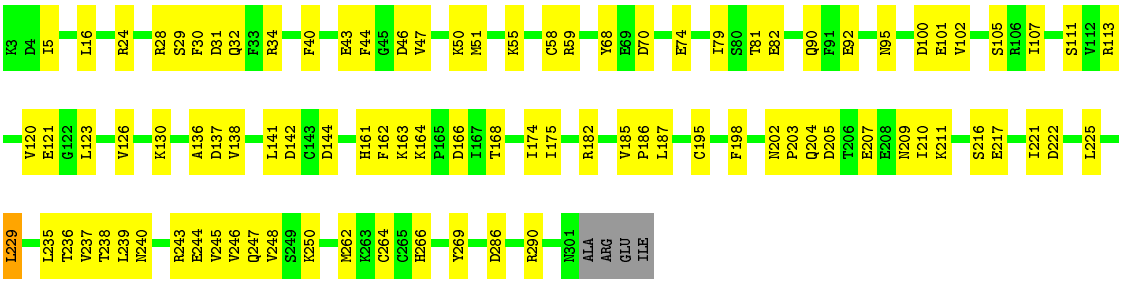
• Molecule 1: RNA (5'-R(P*AP*AP*AP*AP*UP*UP*UP*AP*AP*AP*UP*UP*UP*UP*UP*UP*UP*UP*UP*UP*UP*UP*UP*U)-3')

Chain 2:



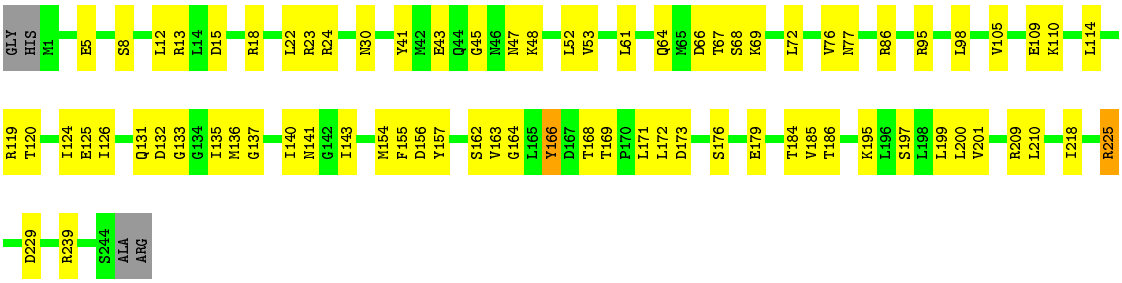
• Molecule 2: Exosome complex component RRP45

Chain AA:



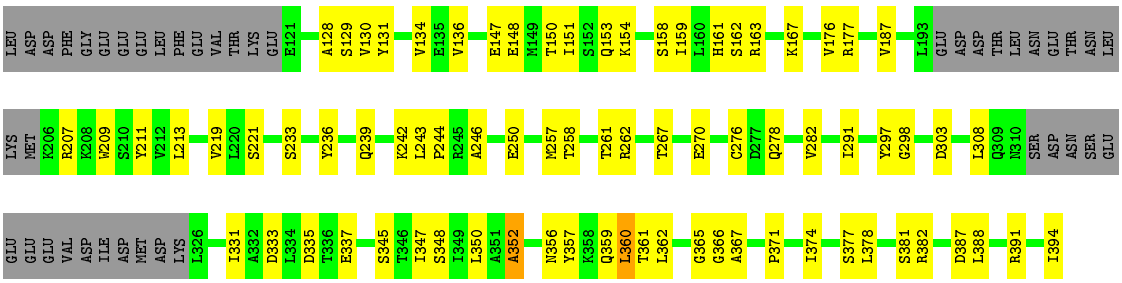
• Molecule 3: Exosome complex component SKI6

Chain BB:

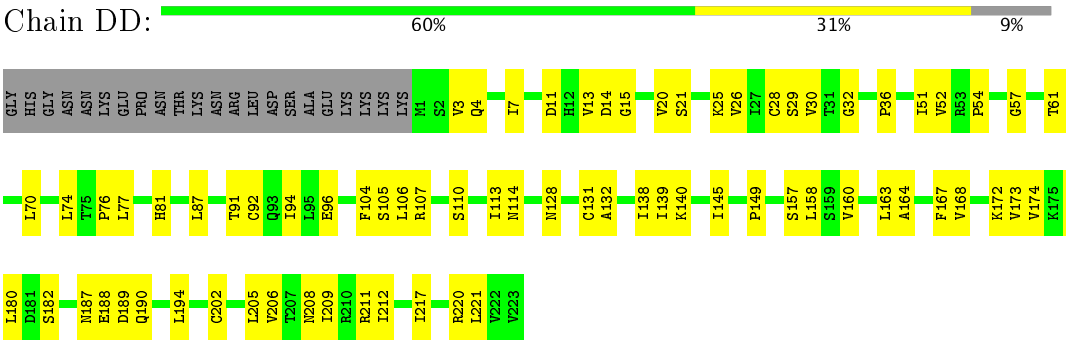


• Molecule 4: Exosome complex component RRP43

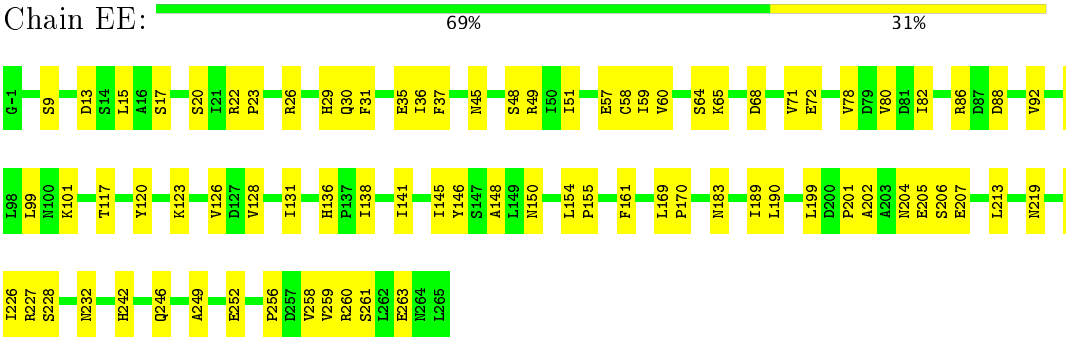
Chain CC:



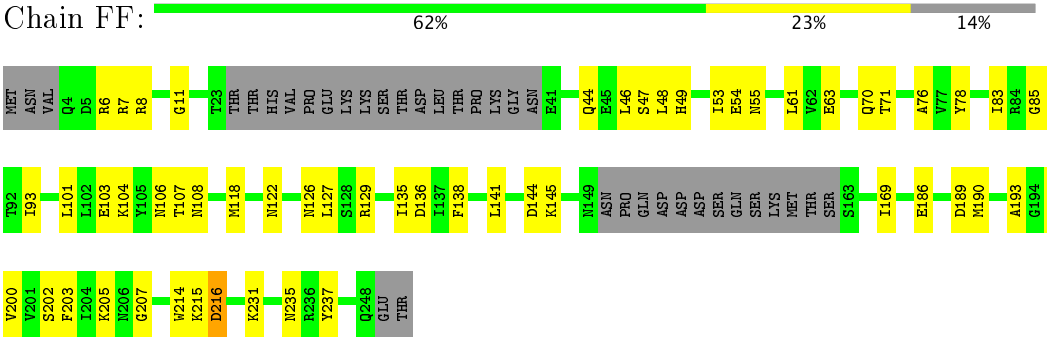
• Molecule 5: Exosome complex component RRP46



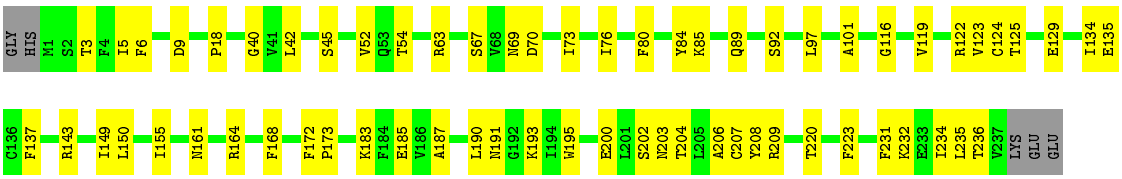
• Molecule 6: Exosome complex component RRP42



• Molecule 7: Exosome complex component MTR3

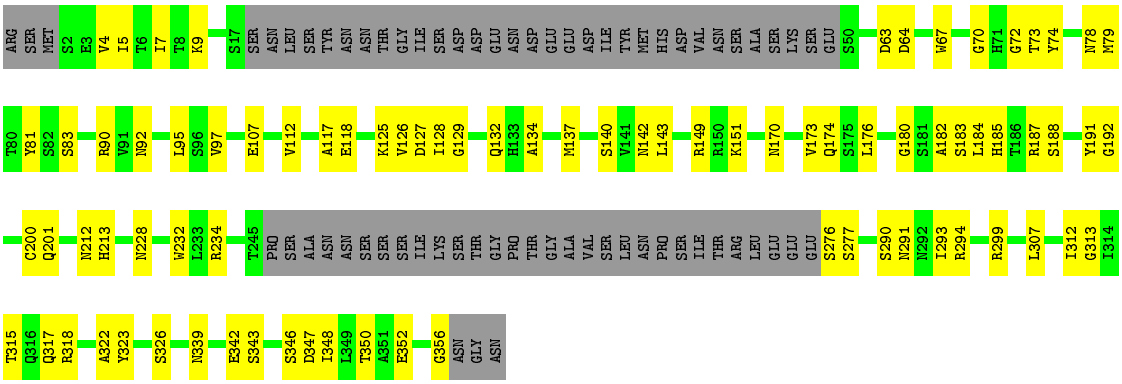


• Molecule 8: Exosome complex component RRP40



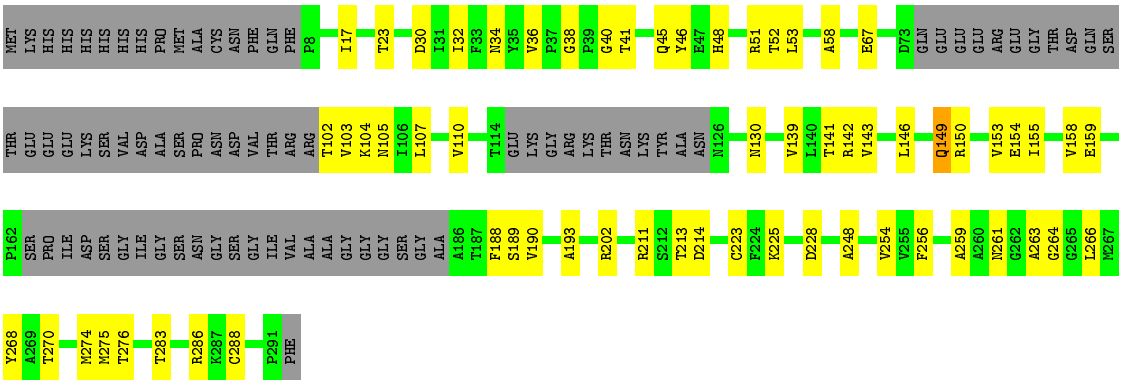
• Molecule 9: Exosome complex component RRP4

Chain HH:



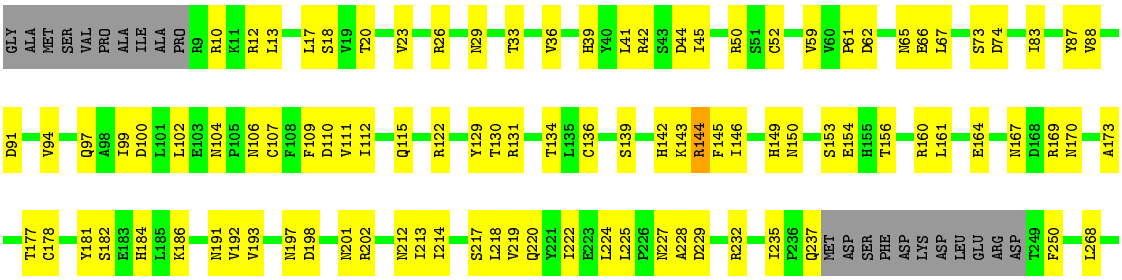
• Molecule 10: Exosome complex component CSL4

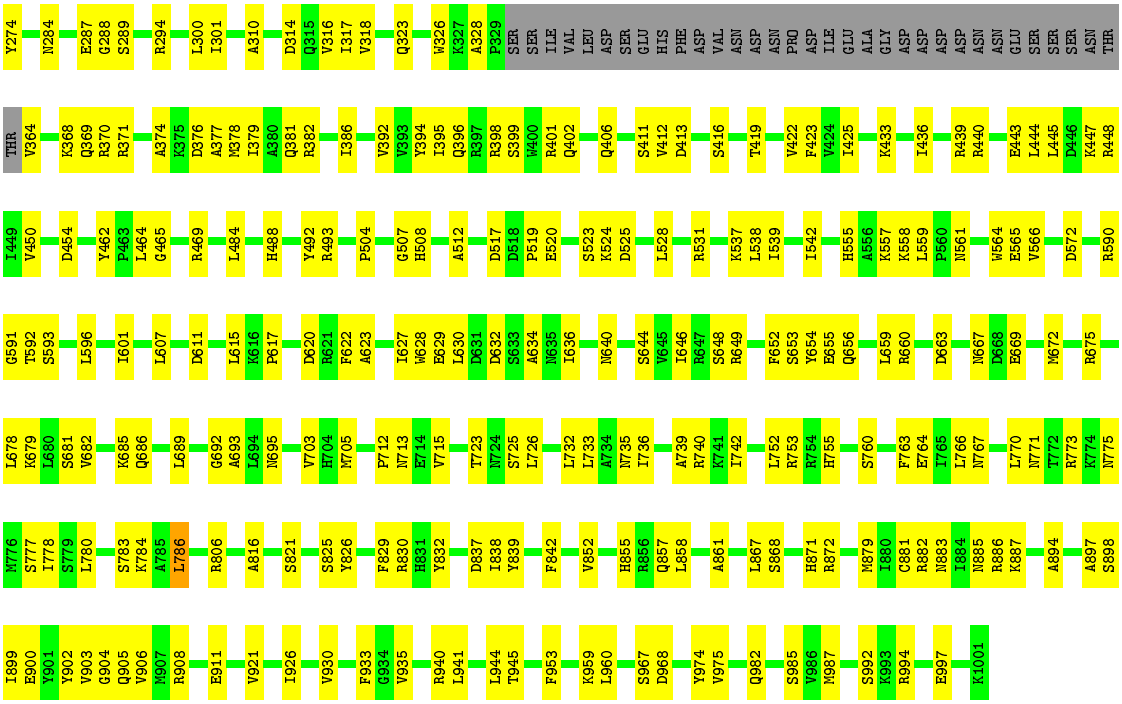
Chain II:



• Molecule 11: Exosome complex exonuclease DIS3

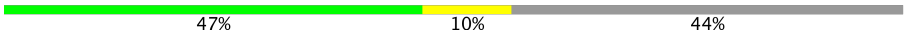
Chain JJ:

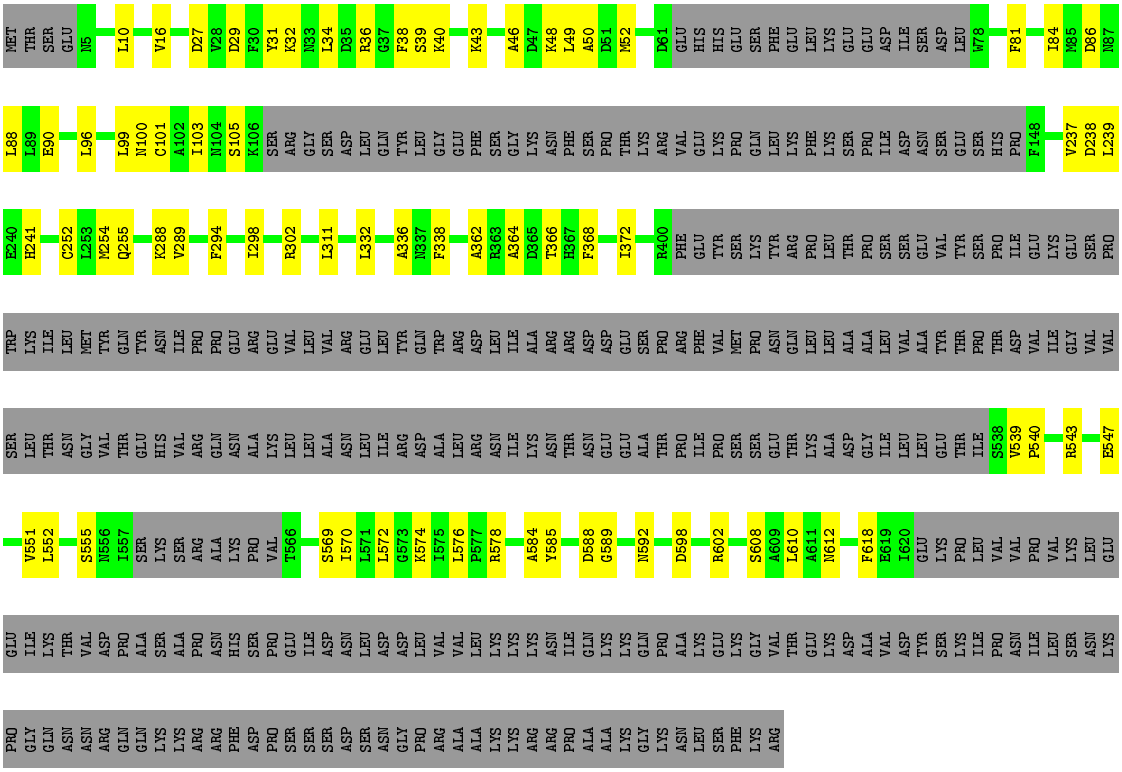




• Molecule 12: Exosome complex exonuclease RRP6

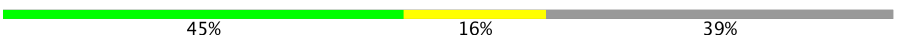
Chain KK:

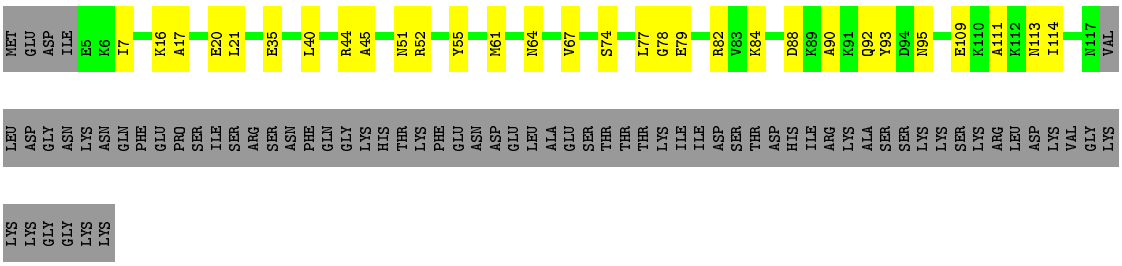


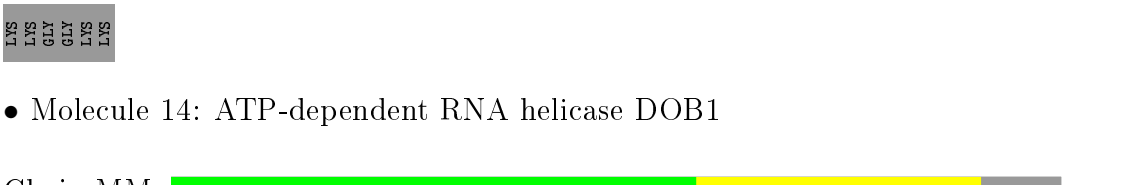


• Molecule 13: Exosome complex protein LRP1

Chain LL:

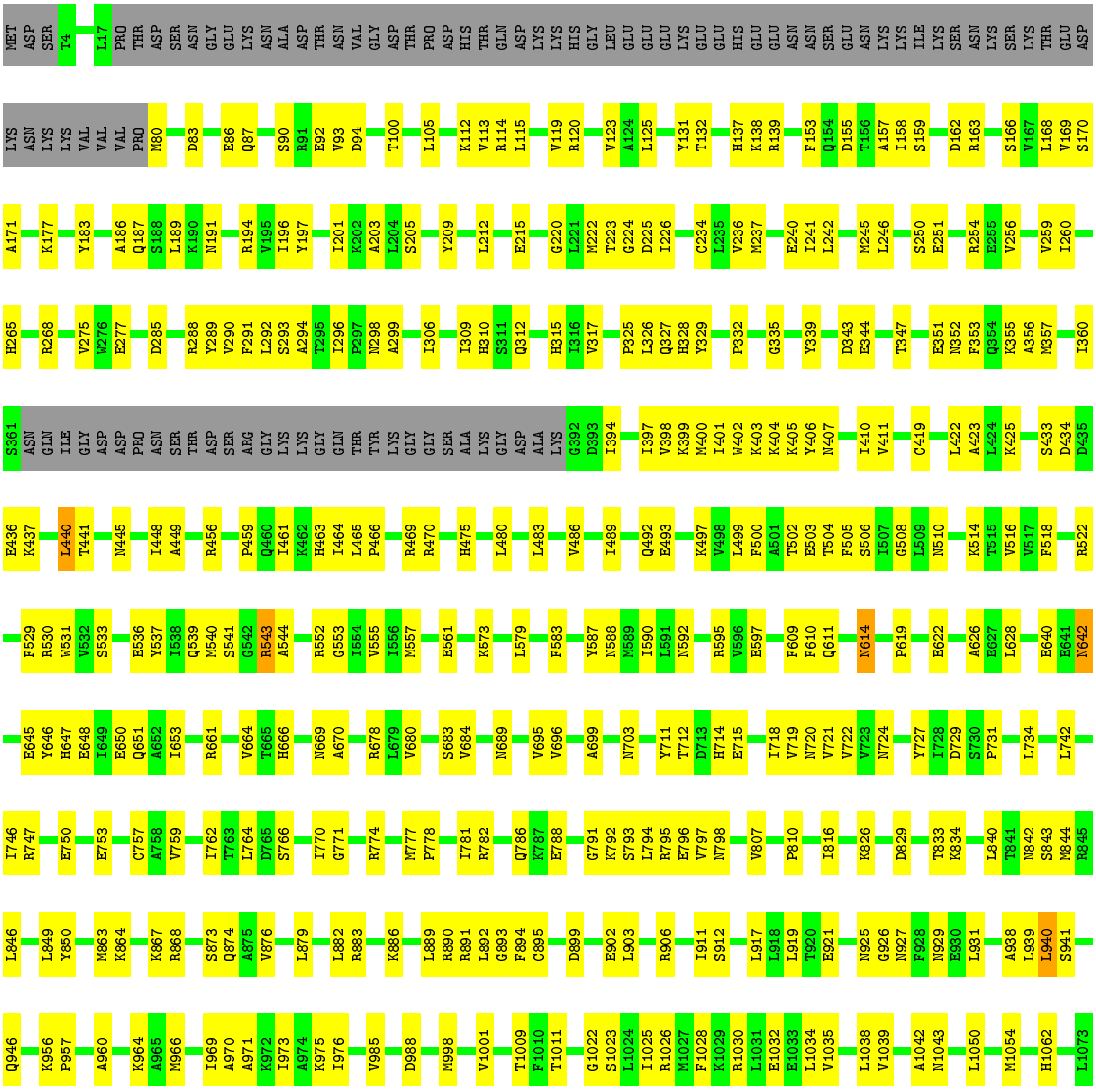






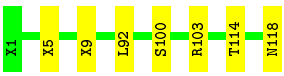
• Molecule 14: ATP-dependent RNA helicase DOB1

Chain MM:



• Molecule 15: M-phase phosphoprotein 6 homolog,M-phase phosphoprotein 6 homolog,Nuclear exosome-associated RNA binding protein,M-phase phosphoprotein 6 homolog

Chain NN:


X1
X5
X9
L92
S100
R103
T114
N118

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	22439	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 ($e^-/Å^2$)	38.4	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

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	2	0.76	0/418	1.51	6/645 (0.9%)
10	II	0.48	0/1676	0.66	0/2277
11	JJ	0.42	0/7575	0.60	1/10290 (0.0%)
12	KK	0.35	0/2540	0.51	0/3497
13	LL	0.39	0/903	0.58	0/1210
14	MM	0.49	0/7773	0.66	3/10521 (0.0%)
15	NN	0.44	0/225	0.68	0/301
2	AA	0.57	0/2340	0.64	1/3161 (0.0%)
3	BB	0.53	0/1910	0.68	0/2579
4	CC	0.50	0/2629	0.69	2/3569 (0.1%)
5	DD	0.54	0/1722	0.70	0/2339
6	EE	0.51	0/2093	0.69	0/2849
7	FF	0.53	0/1660	0.65	0/2241
8	GG	0.55	0/1828	0.68	0/2486
9	HH	0.52	0/2269	0.66	0/3066
All	All	0.49	0/37561	0.66	13/51031 (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
10	II	0	1
11	JJ	0	1
14	MM	0	3
3	BB	0	1
4	CC	0	3
5	DD	0	1
6	EE	0	1
7	FF	0	2
9	HH	0	1
All	All	0	14

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	2	167	U	N1-C2-O2	13.94	132.56	122.80
1	2	167	U	N3-C2-O2	-11.19	114.36	122.20
1	2	167	U	C2-N1-C1'	11.17	131.10	117.70
1	2	167	U	C6-N1-C1'	-8.17	109.77	121.20
14	MM	940	LEU	CA-CB-CG	-6.03	101.43	115.30

There are no chirality outliers.

5 of 14 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	BB	166	TYR	Peptide
4	CC	335	ASP	Peptide
4	CC	352	ALA	Peptide
4	CC	54	ILE	Peptide
5	DD	188	GLU	Peptide

5.2 Too-close contacts (i)

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	2	378	0	190	3	0
2	AA	2304	0	2265	65	0
3	BB	1886	0	1904	55	0
4	CC	2589	0	2607	73	0
5	DD	1701	0	1755	49	0
6	EE	2050	0	2063	57	0
7	FF	1638	0	1590	39	0
8	GG	1792	0	1747	42	0
9	HH	2236	0	2215	54	0
10	II	1653	0	1616	41	0
11	JJ	7430	0	7350	215	0
12	KK	2517	0	1819	57	0
13	LL	894	0	917	22	0
14	MM	7627	0	7521	227	0
15	NN	277	0	238	6	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	36972	0	35797	932	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 13.

The worst 5 of 932 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)
5:DD:25:LYS:H	5:DD:96:GLU:HB3	1.50	0.76
9:HH:356:GLY:HA2	13:LL:35:GLU:HB2	1.70	0.73
5:DD:138:ILE:HG12	5:DD:145:ILE:HG12	1.68	0.73
11:JJ:732:LEU:HD23	11:JJ:735:ASN:HD22	1.53	0.73
14:MM:939:LEU:HD11	14:MM:966:MET:HG2	1.70	0.73

There are no symmetry-related clashes.

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	
2	AA	297/303 (98%)	276 (93%)	21 (7%)	0	100	100
3	BB	242/248 (98%)	224 (93%)	18 (7%)	0	100	100
4	CC	332/393 (84%)	304 (92%)	27 (8%)	1 (0%)	43	80
5	DD	222/245 (91%)	209 (94%)	13 (6%)	0	100	100
6	EE	266/267 (100%)	252 (95%)	14 (5%)	0	100	100
7	FF	209/250 (84%)	191 (91%)	18 (9%)	0	100	100
8	GG	235/242 (97%)	216 (92%)	19 (8%)	0	100	100
9	HH	287/361 (80%)	262 (91%)	25 (9%)	0	100	100
10	II	214/301 (71%)	198 (92%)	16 (8%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
11	JJ	942/1003 (94%)	875 (93%)	67 (7%)	0	100	100
12	KK	404/733 (55%)	366 (91%)	38 (9%)	0	100	100
13	LL	111/184 (60%)	108 (97%)	3 (3%)	0	100	100
14	MM	972/1073 (91%)	909 (94%)	63 (6%)	0	100	100
15	NN	27/40 (68%)	23 (85%)	4 (15%)	0	100	100
All	All	4760/5643 (84%)	4413 (93%)	346 (7%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	CC	337	GLU

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	
2	AA	255/265 (96%)	255 (100%)	0	100	100
3	BB	210/219 (96%)	207 (99%)	3 (1%)	69	85
4	CC	282/349 (81%)	281 (100%)	1 (0%)	92	95
5	DD	196/216 (91%)	196 (100%)	0	100	100
6	EE	238/241 (99%)	238 (100%)	0	100	100
7	FF	181/219 (83%)	180 (99%)	1 (1%)	87	93
8	GG	194/210 (92%)	194 (100%)	0	100	100
9	HH	243/313 (78%)	242 (100%)	1 (0%)	92	95
10	II	174/249 (70%)	172 (99%)	2 (1%)	76	87
11	JJ	812/901 (90%)	808 (100%)	4 (0%)	90	94
12	KK	150/671 (22%)	150 (100%)	0	100	100
13	LL	99/168 (59%)	98 (99%)	1 (1%)	78	89
14	MM	810/953 (85%)	804 (99%)	6 (1%)	85	92

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
15	NN	25/25 (100%)	25 (100%)	0	100 100
All	All	3869/4999 (77%)	3850 (100%)	19 (0%)	90 94

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
11	JJ	29	ASN
11	JJ	169	ARG
14	MM	543	ARG
10	II	130	ASN
14	MM	614	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 43 such sidechains are listed below:

Mol	Chain	Res	Type
11	JJ	555	HIS
12	KK	604	GLN
14	MM	874	GLN
11	JJ	775	ASN
11	JJ	855	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	2	22/23 (95%)	15 (68%)	0

5 of 15 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	2	161	U
1	2	162	U
1	2	163	U
1	2	164	A
1	2	165	A

There are no RNA pucker outliers to report.

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
15	NN	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	NN	11:UNK	C	90:PRO	N	57.41