wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 02:07 PM BST

PDB ID	:	4ADV
EMDB ID:	:	EMD-2017
Title	:	Structure of the E. coli methyltransferase KsgA bound to the E. coli 30S ribosomal subunit
Authors	:	Boehringer, D.; O'Farrell, H.C.; Rife, J.P.; Ban, N.
Deposited on	:	2012-01-03
Resolution	:	13.50 Å(reported)
Based on PDB ID	:	2AVY, 1QYR

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

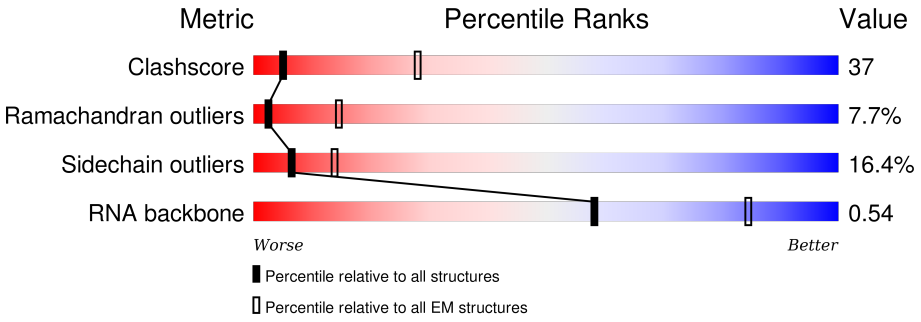
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

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 13.50 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686
RNA backbone	3027	244

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	A	1542	23% 55% 13% 9%
2	B	240	19% 52% 19% • 9%
3	C	232	27% 48% 13% • 11%
4	D	205	27% 53% 19% •
5	E	166	28% 50% 11% • 10%
6	F	135	24% 37% 11% • 26%
7	G	178	26% 46% 11% • 16%
8	H	129	40% 50% 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	I	129	22% 53% 22%
10	J	103	25% 50% 17% 5%
11	K	128	24% 49% 16% 9%
12	L	123	24% 59% 18%
13	M	117	26% 54% 18%
14	N	100	26% 53% 16%
15	O	89	45% 47% 6%
16	P	82	35% 54% 11%
17	Q	83	19% 59% 18%
18	R	74	24% 43% 7% 26%
19	S	91	25% 51% 9% 13%
20	T	86	41% 49% 7%
21	U	71	14% 27% 28% 28%
22	V	252	83% 16%

2 Entry composition

There are 23 unique types of molecules in this entry. The entry contains 50888 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 16S RIBOSOMAL RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1410	Total	C	N	O	P	0	0
			30262	13496	5559	9797	1410		

- Molecule 2 is a protein called 30S RIBOSOMAL PROTEIN S2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	218	Total	C	N	O	S	0	0
			1704	1081	305	311	7		

- Molecule 3 is a protein called 30S RIBOSOMAL PROTEIN S3.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	206	Total	C	N	O	S	0	0
			1624	1028	305	288	3		

- Molecule 4 is a protein called 30S RIBOSOMAL PROTEIN S4.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	205	Total	C	N	O	S	0	0
			1643	1026	315	298	4		

- Molecule 5 is a protein called 30S RIBOSOMAL PROTEIN S5.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	150	Total	C	N	O	S	0	0
			1105	687	211	201	6		

- Molecule 6 is a protein called 30S RIBOSOMAL PROTEIN S6.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	100	Total	C	N	O	S	0	0
			817	515	148	148	6		

- Molecule 7 is a protein called 30S RIBOSOMAL PROTEIN S7.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	150	Total	C	N	O	S	0	0
			1174	730	226	214	4		

- Molecule 8 is a protein called 30S RIBOSOMAL PROTEIN S8.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	129	Total	C	N	O	S	0	0
			979	616	173	184	6		

- Molecule 9 is a protein called 30S RIBOSOMAL PROTEIN S9.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	127	Total	C	N	O	S	0	0
			1022	634	206	179	3		

- Molecule 10 is a protein called 30S RIBOSOMAL PROTEIN S10.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	J	98	Total	C	N	O	S	0	0
			786	493	150	142	1		

- Molecule 11 is a protein called 30S RIBOSOMAL PROTEIN S11.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	K	117	Total	C	N	O	S	0	0
			877	540	174	160	3		

- Molecule 12 is a protein called 30S RIBOSOMAL PROTEIN S12.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	L	123	Total	C	N	O	S	0	0
			955	590	196	165	4		

- Molecule 13 is a protein called 30S RIBOSOMAL PROTEIN S13.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	M	114	Total	C	N	O	S	0	0
			883	546	178	156	3		

- Molecule 14 is a protein called 30S RIBOSOMAL PROTEIN S14.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	N	96	Total	C	N	O	S	0	0
			774	483	160	128	3		

- Molecule 15 is a protein called 30S RIBOSOMAL PROTEIN S15.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	O	88	Total	C	N	O	S	0	0
			716	440	146	129	1		

- Molecule 16 is a protein called 30S RIBOSOMAL PROTEIN S16.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	P	82	Total	C	N	O	S	0	0
			649	406	128	114	1		

- Molecule 17 is a protein called 30S RIBOSOMAL PROTEIN S17.

Mol	Chain	Residues	Atoms					AltConf	Trace
17	Q	80	Total	C	N	O	S	0	0
			648	411	121	113	3		

- Molecule 18 is a protein called 30S RIBOSOMAL PROTEIN S18.

Mol	Chain	Residues	Atoms				AltConf	Trace
18	R	55	Total	C	N	O	0	0
			455	288	86	81		

- Molecule 19 is a protein called 30S RIBOSOMAL PROTEIN S19.

Mol	Chain	Residues	Atoms					AltConf	Trace
19	S	79	Total	C	N	O	S	0	0
			637	408	120	107	2		

- Molecule 20 is a protein called 30S RIBOSOMAL PROTEIN S20.

Mol	Chain	Residues	Atoms					AltConf	Trace
20	T	85	Total	C	N	O	S	0	0
			665	411	137	114	3		

- Molecule 21 is a protein called 30S RIBOSOMAL PROTEIN S21.

Mol	Chain	Residues	Atoms					AltConf	Trace
21	U	51	Total	C	N	O	S	0	0
			425	265	86	73	1		

- Molecule 22 is a protein called RIBOSOMAL RNA SMALL SUBUNIT METHYLTRANSFERASE A.

Mol	Chain	Residues	Atoms					AltConf	Trace
22	V	252	Total	C	N	O	S	0	0
			1949	1241	336	359	13		

- Molecule 23 is water.

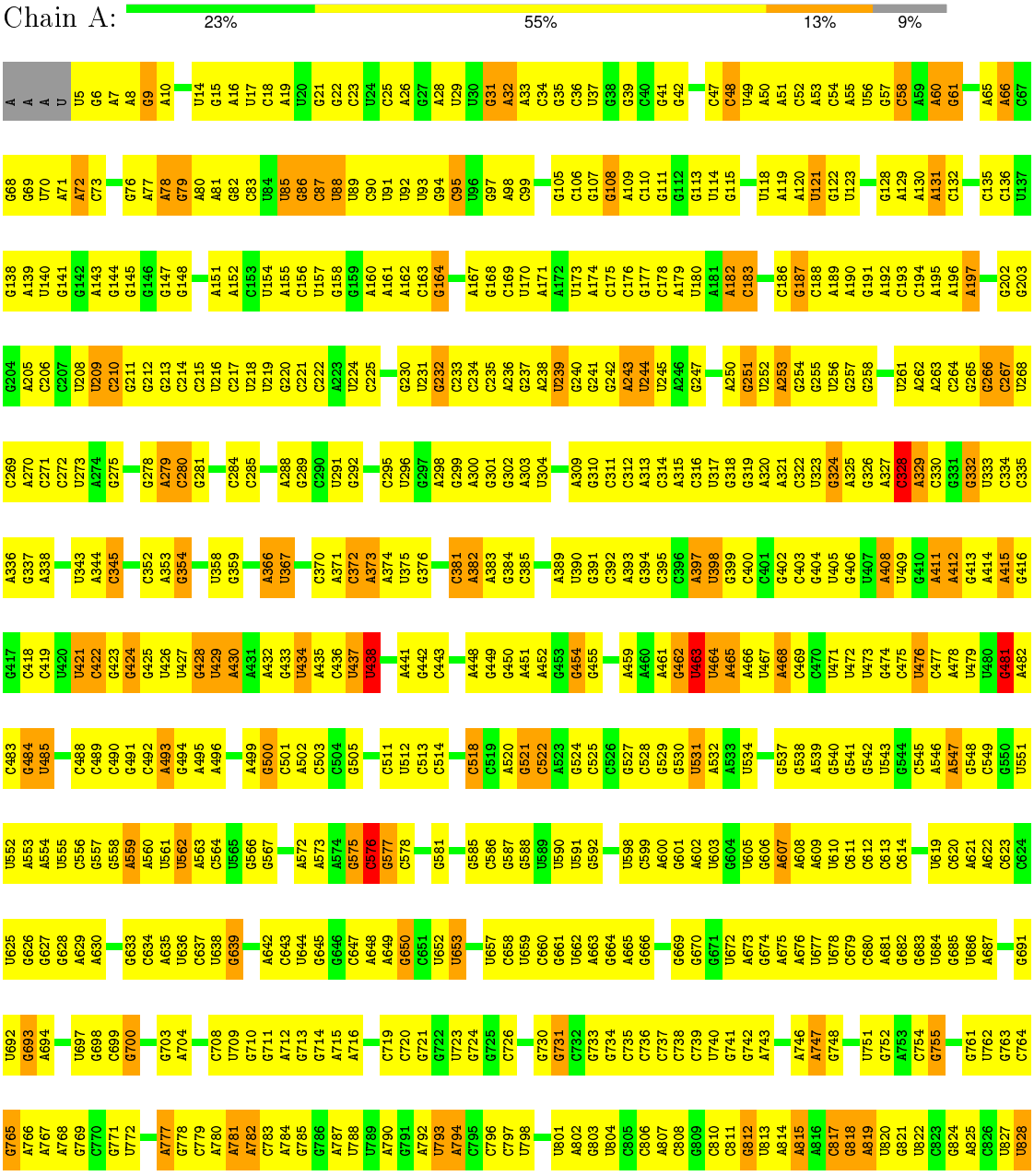
Mol	Chain	Residues	Atoms		AltConf
23	V	139	Total	O	0
			139	139	

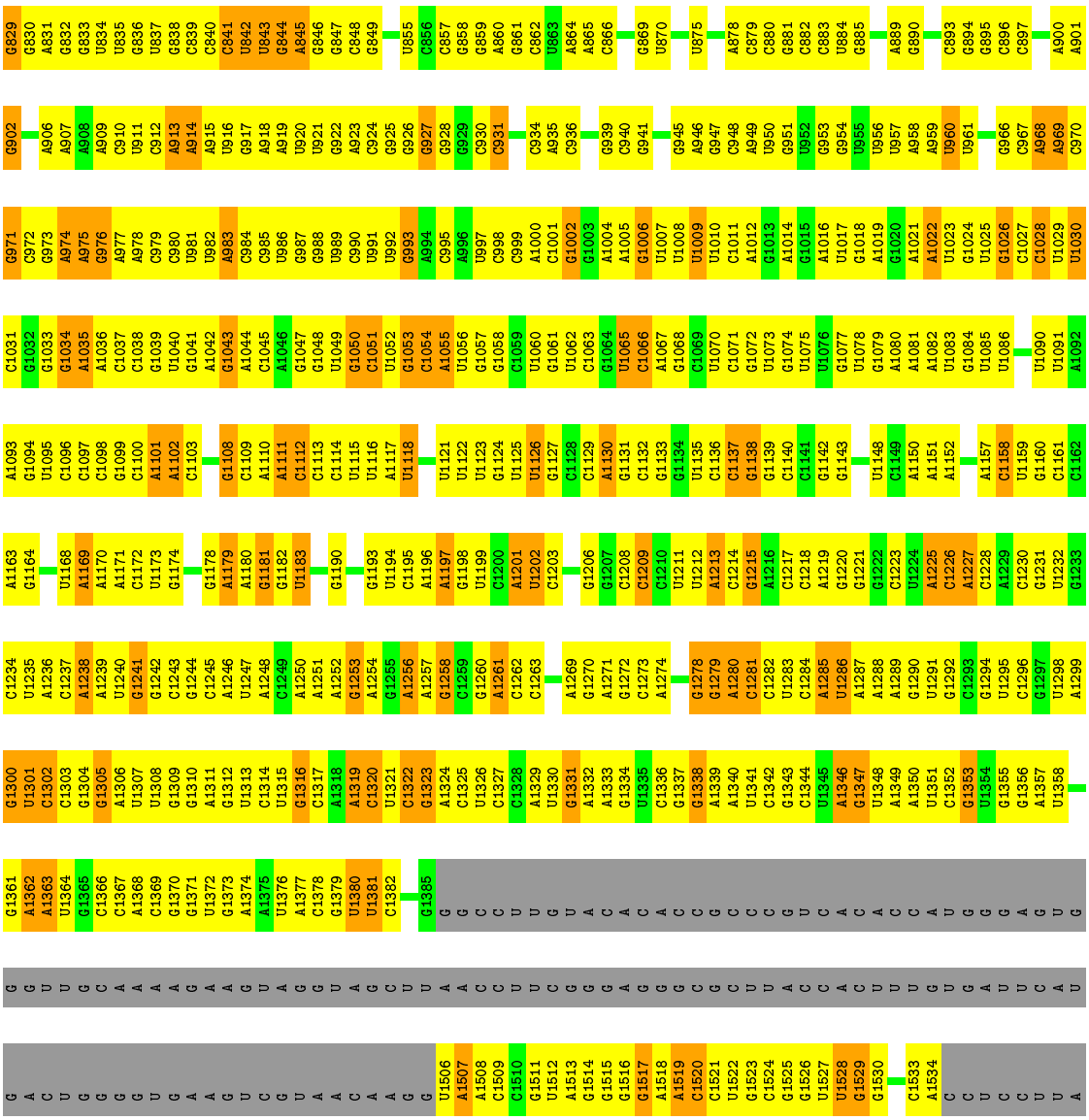
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: 16S RIBOSOMAL RNA

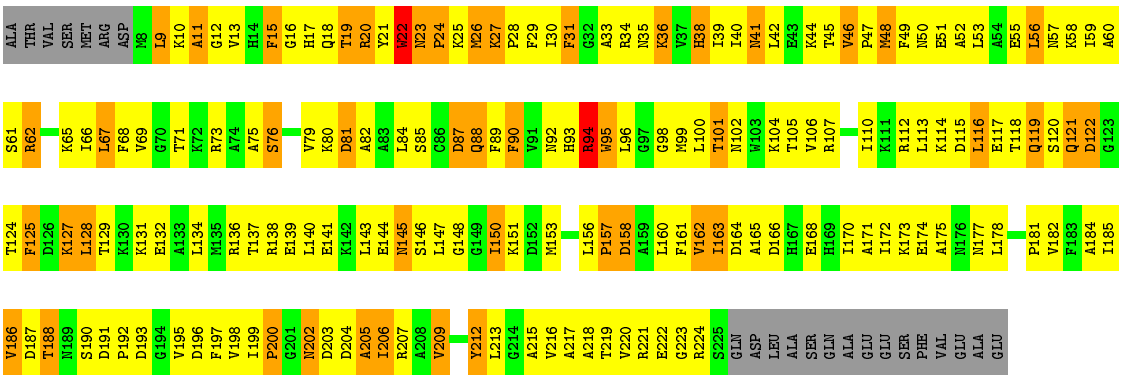
Chain A:





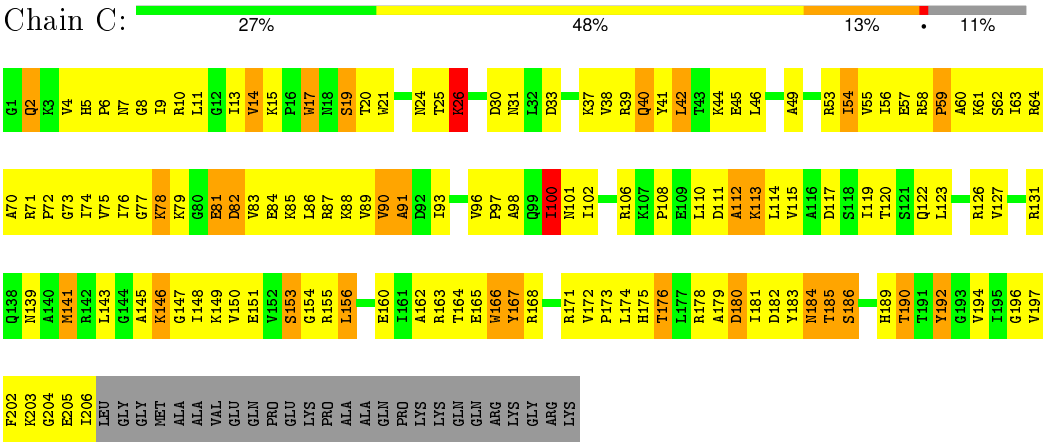
• Molecule 2: 30S RIBOSOMAL PROTEIN S2

Chain B:



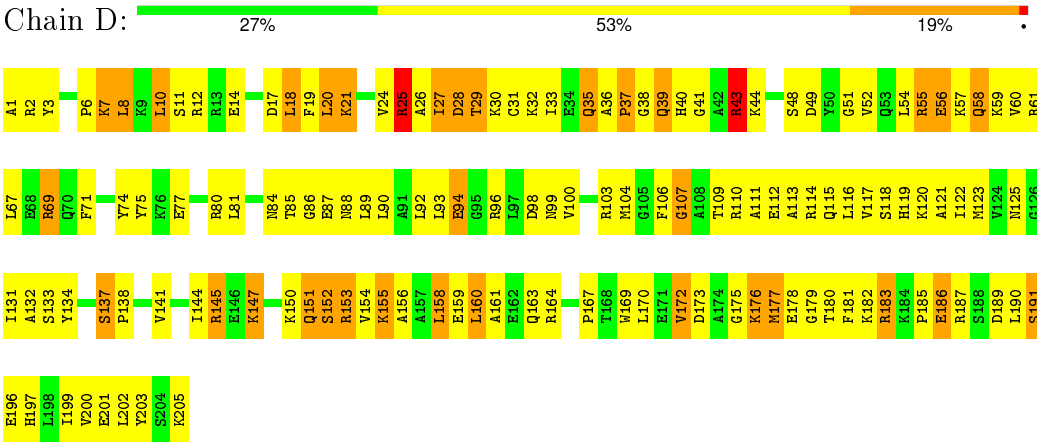
• Molecule 3: 30S RIBOSOMAL PROTEIN S3

Chain C:



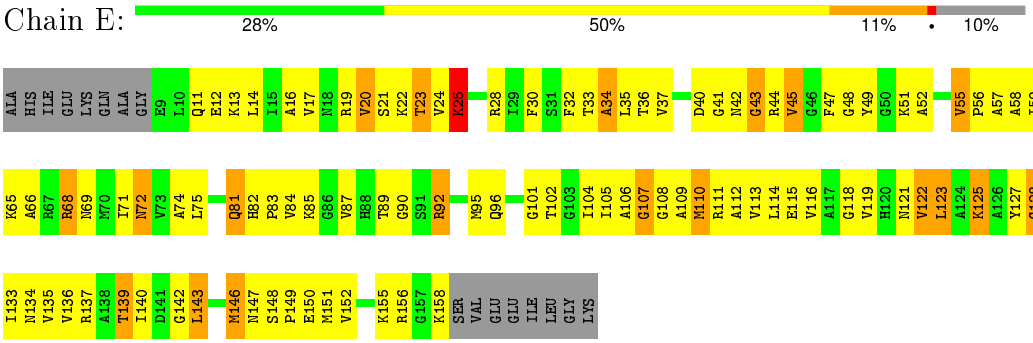
- Molecule 4: 30S RIBOSOMAL PROTEIN S4

Chain D:



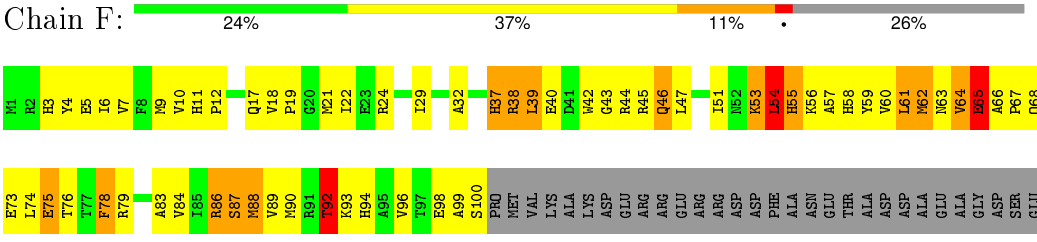
- Molecule 5: 30S RIBOSOMAL PROTEIN S5

Chain E:



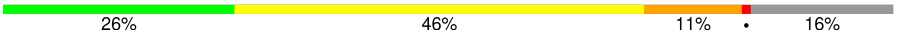
- Molecule 6: 30S RIBOSOMAL PROTEIN S6

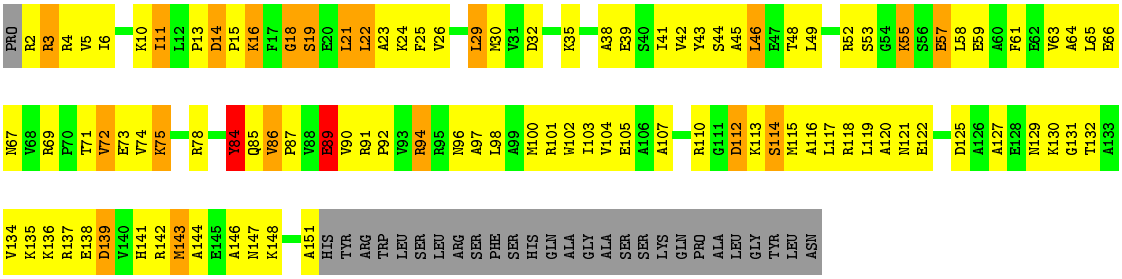
Chain F:



- Molecule 7: 30S RIBOSOMAL PROTEIN S7

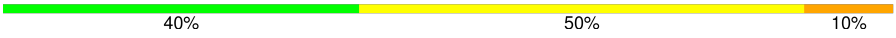
Chain G:

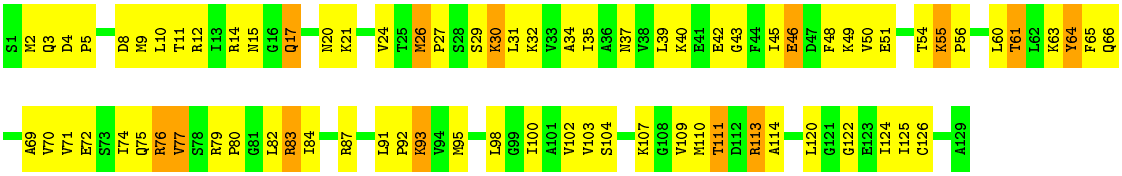




- Molecule 8: 30S RIBOSOMAL PROTEIN S8

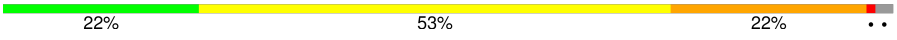
Chain H:

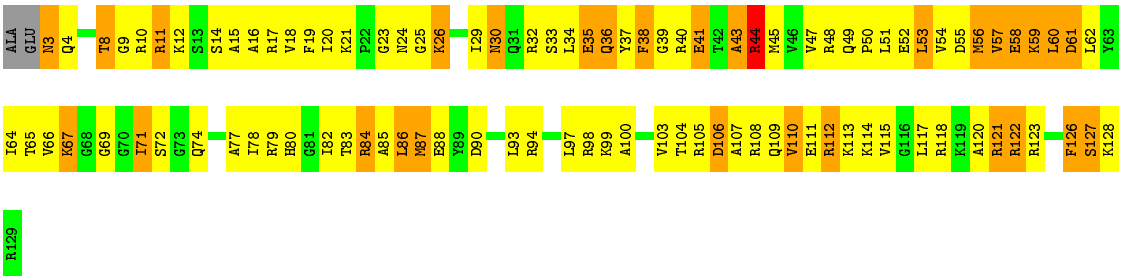




- Molecule 9: 30S RIBOSOMAL PROTEIN S9

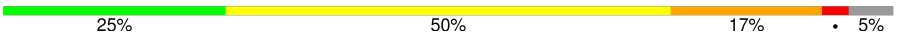
Chain I:

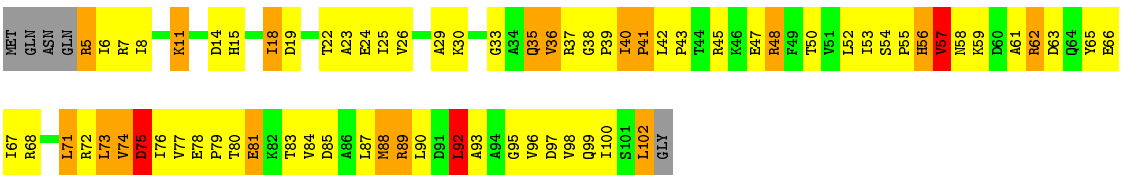




- Molecule 10: 30S RIBOSOMAL PROTEIN S10

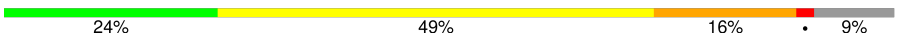
Chain J:

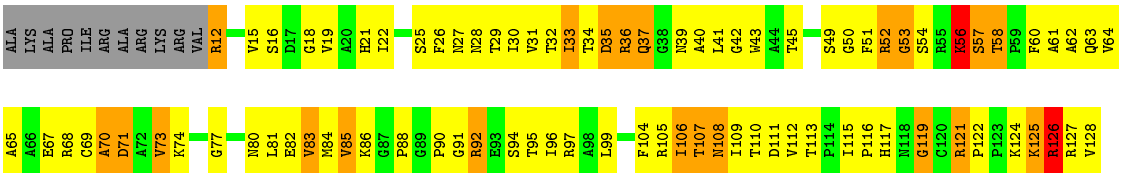




- Molecule 11: 30S RIBOSOMAL PROTEIN S11

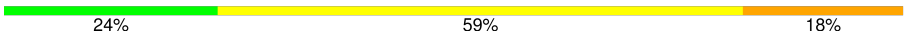
Chain K:

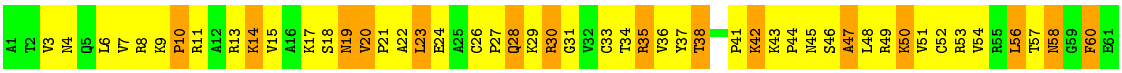


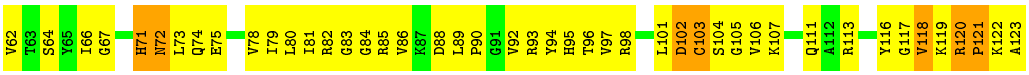


- Molecule 12: 30S RIBOSOMAL PROTEIN S12

Chain L:

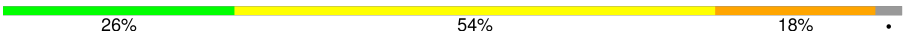


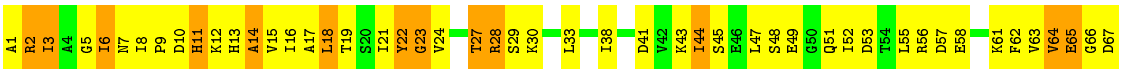


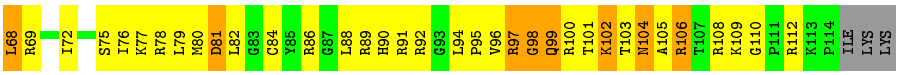


- Molecule 13: 30S RIBOSOMAL PROTEIN S13

Chain M:

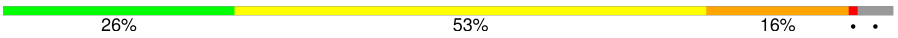


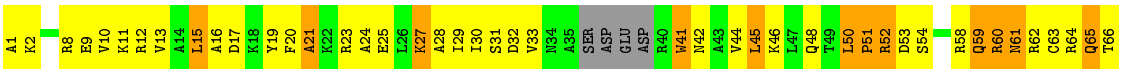




- Molecule 14: 30S RIBOSOMAL PROTEIN S14

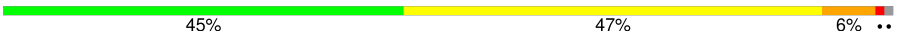
Chain N:

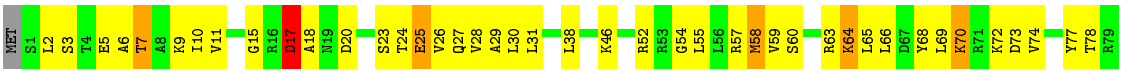


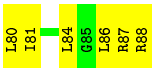


- Molecule 15: 30S RIBOSOMAL PROTEIN S15

Chain O:

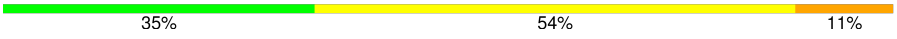






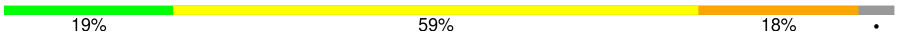
- Molecule 16: 30S RIBOSOMAL PROTEIN S16

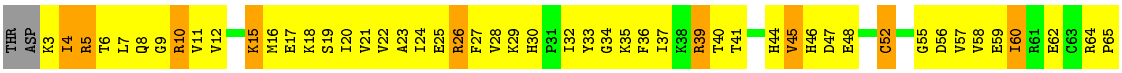
Chain P:



- Molecule 17: 30S RIBOSOMAL PROTEIN S17

Chain Q:



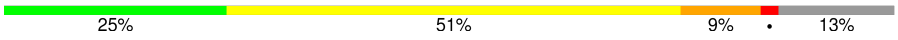


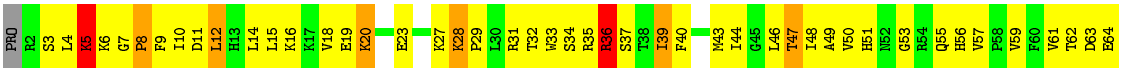
- Molecule 18: 30S RIBOSOMAL PROTEIN S18

Chain R:

- Molecule 19: 30S RIBOSOMAL PROTEIN S19

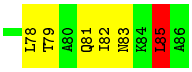
Chain S:





- Molecule 20: 30S RIBOSOMAL PROTEIN S20

Chain T:

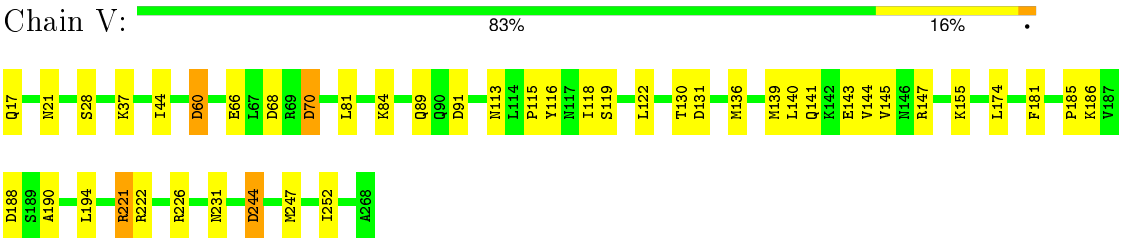


- Molecule 21: 30S RIBOSOMAL PROTEIN S21

Chain U:

• Molecule 22: RIBOSOMAL RNA SMALL SUBUNIT METHYLTRANSFERASE A

Chain V:



4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	Not provided	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	PER IMAGE, CTFFIND3	Depositor
Microscope	FEI TECNAI 20	Depositor
Voltage (kV)	200	Depositor
Electron dose ($e^-/Å^2$)	20	Depositor
Minimum defocus (nm)	2000	Depositor
Maximum defocus (nm)	4000	Depositor
Magnification	62000	Depositor
Image detector	GATAN US4000	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	A	0.27	2/33885 (0.0%)	0.75	7/52858 (0.0%)
10	J	0.23	0/796	0.49	0/1077
11	K	0.24	0/893	0.46	0/1205
12	L	0.22	0/969	0.47	0/1300
13	M	0.21	0/892	0.48	0/1193
14	N	0.25	0/785	0.46	0/1043
15	O	0.23	0/724	0.45	0/966
16	P	0.26	0/659	0.44	0/884
17	Q	0.24	0/657	0.46	0/881
18	R	0.23	0/462	0.46	0/621
19	S	0.25	0/652	0.46	0/877
2	B	0.25	0/1735	0.47	0/2338
20	T	0.24	0/671	0.41	0/888
21	U	1.01	4/430 (0.9%)	0.74	2/570 (0.4%)
22	V	0.62	0/1991	0.80	7/2710 (0.3%)
3	C	0.23	0/1651	0.45	0/2225
4	D	0.23	0/1665	0.46	0/2227
5	E	0.23	0/1118	0.45	0/1504
6	F	0.25	0/835	0.48	0/1128
7	G	0.23	0/1187	0.44	0/1591
8	H	0.23	0/989	0.45	0/1326
9	I	0.24	0/1034	0.45	0/1375
All	All	0.29	6/54680 (0.0%)	0.68	16/80787 (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	15
21	U	0	1
All	All	0	16

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
21	U	15	LEU	C-N	-15.01	0.99	1.34
21	U	25	ALA	C-N	-9.28	1.16	1.33
1	A	463	U	O3'-P	-6.70	1.53	1.61
21	U	29	ALA	C-N	6.63	1.49	1.34
21	U	11	PHE	C-N	-5.36	1.21	1.34

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	765	G	N9-C1'-C2'	-8.16	103.02	112.00
22	V	221	ARG	NE-CZ-NH2	-6.96	116.82	120.30
1	A	438	U	N1-C1'-C2'	-6.29	105.08	112.00
22	V	221	ARG	NE-CZ-NH1	6.22	123.41	120.30
1	A	232	G	C5'-C4'-C3'	-6.21	106.07	116.00

There are no chirality outliers.

5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	187	G	Sidechain
1	A	281	G	Sidechain
1	A	437	U	Sidechain
1	A	438	U	Sidechain
1	A	58	C	Sidechain

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	A	30262	0	15198	1460	0
2	B	1704	0	1720	428	0
3	C	1624	0	1699	158	0
4	D	1643	0	1710	176	0
5	E	1105	0	1145	170	0
6	F	817	0	808	76	0
7	G	1174	0	1230	100	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	H	979	0	1032	81	0
9	I	1022	0	1070	141	0
10	J	786	0	828	96	0
11	K	877	0	887	107	0
12	L	955	0	1019	103	0
13	M	883	0	944	84	0
14	N	774	0	827	92	0
15	O	716	0	742	50	0
16	P	649	0	666	78	0
17	Q	648	0	691	81	0
18	R	455	0	478	39	0
19	S	637	0	665	78	0
20	T	665	0	714	53	0
21	U	425	0	447	102	0
22	V	1949	0	1949	46	0
23	V	139	0	0	11	0
All	All	50888	0	36469	3247	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 37.

The worst 5 of 3247 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:870:U:C6	2:B:27:LYS:HD3	1.28	1.64
1:A:829:G:C3'	2:B:30:ILE:HG12	1.18	1.62
1:A:830:G:C4	2:B:22:TRP:HB3	1.40	1.53
1:A:17:U:C5'	1:A:1079:G:H5'	1.39	1.49
1:A:830:G:C5	2:B:22:TRP:HB3	1.47	1.49

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	B	216/240 (90%)	145 (67%)	53 (24%)	18 (8%)	1 18
3	C	204/232 (88%)	135 (66%)	49 (24%)	20 (10%)	1 14
4	D	203/205 (99%)	130 (64%)	55 (27%)	18 (9%)	1 17
5	E	148/166 (89%)	108 (73%)	33 (22%)	7 (5%)	3 32
6	F	98/135 (73%)	69 (70%)	23 (24%)	6 (6%)	2 26
7	G	148/178 (83%)	103 (70%)	37 (25%)	8 (5%)	2 29
8	H	127/129 (98%)	105 (83%)	19 (15%)	3 (2%)	7 47
9	I	125/129 (97%)	86 (69%)	28 (22%)	11 (9%)	1 17
10	J	96/103 (93%)	63 (66%)	21 (22%)	12 (12%)	0 8
11	K	115/128 (90%)	75 (65%)	27 (24%)	13 (11%)	0 10
12	L	121/123 (98%)	74 (61%)	30 (25%)	17 (14%)	0 6
13	M	112/117 (96%)	87 (78%)	14 (12%)	11 (10%)	1 14
14	N	92/100 (92%)	59 (64%)	24 (26%)	9 (10%)	1 14
15	O	86/89 (97%)	66 (77%)	18 (21%)	2 (2%)	8 48
16	P	80/82 (98%)	59 (74%)	17 (21%)	4 (5%)	3 31
17	Q	78/83 (94%)	48 (62%)	26 (33%)	4 (5%)	2 30
18	R	53/74 (72%)	27 (51%)	19 (36%)	7 (13%)	0 7
19	S	77/91 (85%)	57 (74%)	14 (18%)	6 (8%)	1 20
20	T	83/86 (96%)	69 (83%)	8 (10%)	6 (7%)	1 22
21	U	49/71 (69%)	22 (45%)	12 (24%)	15 (31%)	0 0
22	V	250/252 (99%)	247 (99%)	2 (1%)	1 (0%)	39 80
All	All	2561/2813 (91%)	1834 (72%)	529 (21%)	198 (8%)	2 20

5 of 198 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	19	THR
2	B	22	TRP
2	B	163	ILE
2	B	186	VAL
3	C	2	GLN

5.3.2 Protein sidechains (i)

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

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

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	B	180/198 (91%)	142 (79%)	38 (21%)	1 9
3	C	170/189 (90%)	142 (84%)	28 (16%)	3 19
4	D	172/172 (100%)	140 (81%)	32 (19%)	2 14
5	E	113/125 (90%)	92 (81%)	21 (19%)	2 14
6	F	87/116 (75%)	68 (78%)	19 (22%)	1 9
7	G	123/146 (84%)	102 (83%)	21 (17%)	2 17
8	H	104/104 (100%)	87 (84%)	17 (16%)	3 20
9	I	105/106 (99%)	83 (79%)	22 (21%)	1 9
10	J	86/90 (96%)	66 (77%)	20 (23%)	1 7
11	K	90/98 (92%)	70 (78%)	20 (22%)	1 9
12	L	103/103 (100%)	88 (85%)	15 (15%)	4 24
13	M	92/95 (97%)	70 (76%)	22 (24%)	1 7
14	N	79/83 (95%)	67 (85%)	12 (15%)	3 22
15	O	76/77 (99%)	69 (91%)	7 (9%)	11 43
16	P	65/65 (100%)	56 (86%)	9 (14%)	4 27
17	Q	74/77 (96%)	60 (81%)	14 (19%)	2 13
18	R	48/64 (75%)	45 (94%)	3 (6%)	22 59
19	S	70/78 (90%)	60 (86%)	10 (14%)	4 25
20	T	65/65 (100%)	56 (86%)	9 (14%)	4 27
21	U	44/61 (72%)	36 (82%)	8 (18%)	2 15
22	V	213/216 (99%)	206 (97%)	7 (3%)	45 76
All	All	2159/2328 (93%)	1805 (84%)	354 (16%)	6 19

5 of 354 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
8	H	17	GLN
10	J	5	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
19	S	47	THR
8	H	48	PHE
9	I	30	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 69 such sidechains are listed below:

Mol	Chain	Res	Type
7	G	121	ASN
9	I	30	ASN
19	S	51	HIS
7	G	147	ASN
8	H	20	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	1407/1542 (91%)	261 (18%)	24 (1%)

5 of 261 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	7	A
1	A	9	G
1	A	14	U
1	A	31	G
1	A	32	A

5 of 24 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	576	C
1	A	975	A
1	A	1302	C
1	A	913	A
1	A	960	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

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.