wwPDB EM Validation Summary Report

Nov 22, 2022 – 01:33 PM JST

PDB ID	:	7EGI
EMDB ID	:	EMD-31118
Title	:	TFIID in rearranged conformation
Authors	:	Chen, X.; Wu, Z.; Li, J.; Zhao, D.; Xu, Y.
Deposited on	:	2021-03-24
Resolution	:	9.82 Å(reported)

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

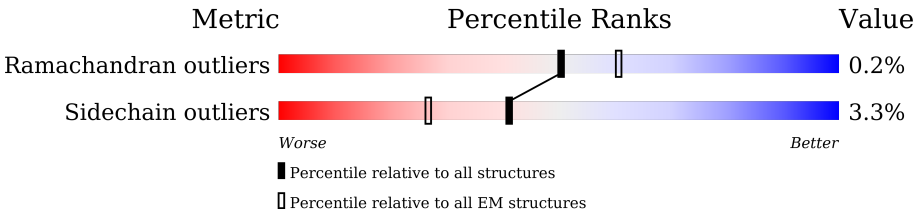
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

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 9.82 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.



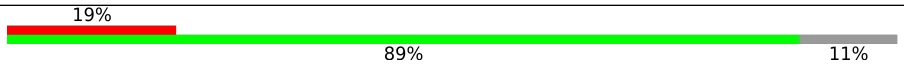
Metric	Whole archive (#Entries)	EM structures (#Entries)
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1872	27% 71%
2	B	1199	79% 20%
3	D	1085	17% 82%
3	d	1085	14% 85%
4	E	800	66% 32%
4	e	800	9% 66% 33%
5	F	677	58% 39%
5	f	677	58% 40%
6	G	349	7% 39% 59%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	H	310	
8	I	264	
8	i	264	
9	J	218	
9	j	218	
10	L	161	
10	l	161	
11	O	109	
12	P	339	
13	Q	376	
14	c	929	
15	k	211	
16	m	124	

2 Entry composition (i)

There are 16 unique types of molecules in this entry. The entry contains 43358 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 Transcription initiation factor TFIID subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	549	Total	C	N	O	S	0	0
			4505	2879	781	818	27		

- Molecule 2 is a protein called Transcription initiation factor TFIID subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	963	Total	C	N	O	S	0	0
			7796	5011	1315	1412	58		

- Molecule 3 is a protein called Transcription initiation factor TFIID subunit 4.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	197	Total	C	N	O	S	0	0
			1614	997	302	310	5		
3	d	158	Total	C	N	O	S	0	0
			1307	814	238	252	3		

- Molecule 4 is a protein called Transcription initiation factor TFIID subunit 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	545	Total	C	N	O	S	0	0
			4359	2765	756	817	21		
4	e	539	Total	C	N	O	S	0	0
			4327	2746	748	814	19		

- Molecule 5 is a protein called Transcription initiation factor TFIID subunit 6.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	F	412	Total	C	N	O	S	0	0
			3143	1994	548	583	18		
5	f	403	Total	C	N	O	S	0	0
			3081	1954	533	576	18		

- Molecule 6 is a protein called Transcription initiation factor TFIID subunit 7.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	G	144	Total	C	N	O	S	0	0
			1171	742	215	210	4		

- Molecule 7 is a protein called Transcription initiation factor TFIID subunit 8.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	H	209	Total	C	N	O	S	0	0
			1622	1026	281	310	5		

- Molecule 8 is a protein called Transcription initiation factor TFIID subunit 9.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	I	120	Total	C	N	O	S	0	0
			959	610	166	177	6		
8	i	121	Total	C	N	O	S	0	0
			967	615	167	178	7		

- Molecule 9 is a protein called Transcription initiation factor TFIID subunit 10.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	J	90	Total	C	N	O	S	0	0
			720	466	115	135	4		
9	j	95	Total	C	N	O	S	0	0
			759	488	124	143	4		

- Molecule 10 is a protein called Transcription initiation factor TFIID subunit 12.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	L	76	Total	C	N	O	S	0	0
			622	388	109	122	3		
10	l	107	Total	C	N	O	S	0	0
			876	547	158	166	5		

- Molecule 11 is a protein called Transcription initiation factor IIA subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	O	97	Total	C	N	O	S	0	0
			771	491	133	145	2		

- Molecule 12 is a protein called TATA-box-binding protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	P	177	Total	C	N	O	S	0	0
			1412	918	249	238	7		

- Molecule 13 is a protein called Transcription initiation factor IIA subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	Q	101	Total	C	N	O	S	0	0
			827	524	139	160	4		

- Molecule 14 is a protein called Transcription initiation factor TFIID subunit 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	c	127	Total	C	N	O	S	0	0
			1011	638	174	193	6		

- Molecule 15 is a protein called Transcription initiation factor TFIID subunit 11.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	k	98	Total	C	N	O	S	0	0
			785	499	142	139	5		

- Molecule 16 is a protein called Transcription initiation factor TFIID subunit 13.

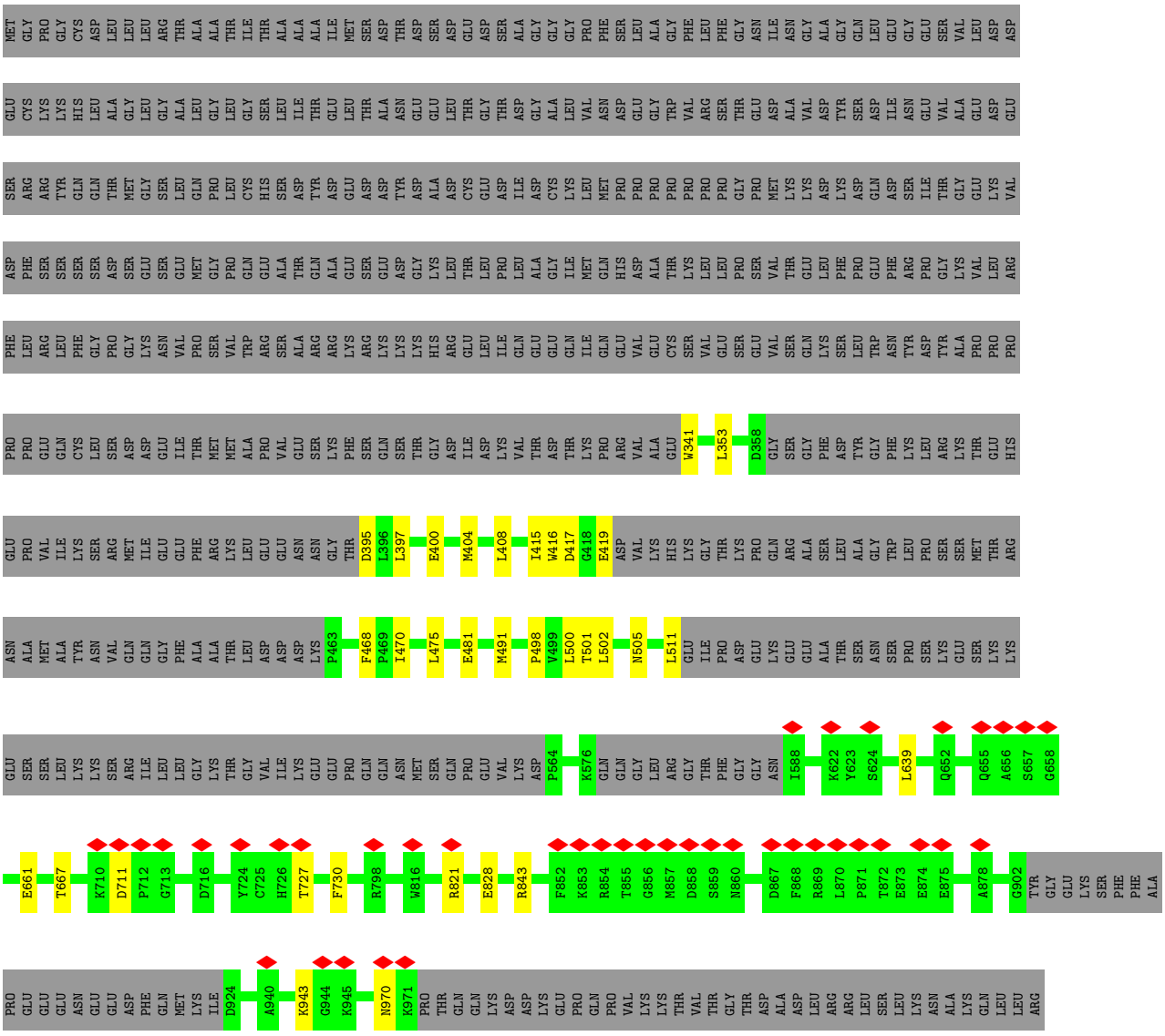
Mol	Chain	Residues	Atoms					AltConf	Trace
16	m	87	Total	C	N	O	S	0	0
			724	456	131	131	6		

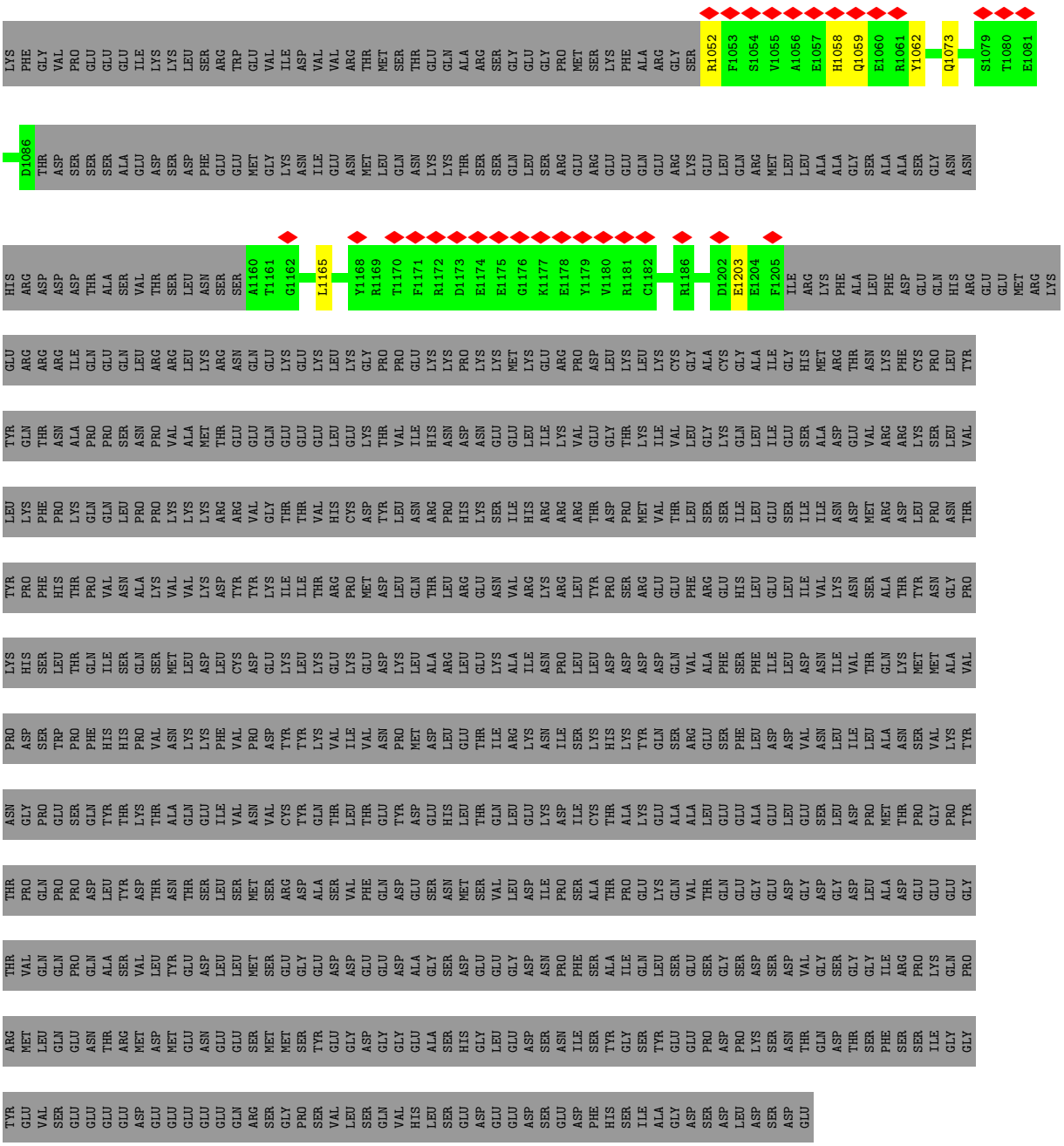
3 Residue-property plots (i)

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 and atom inclusion in map density. 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. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). 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: Transcription initiation factor TFIID subunit 1

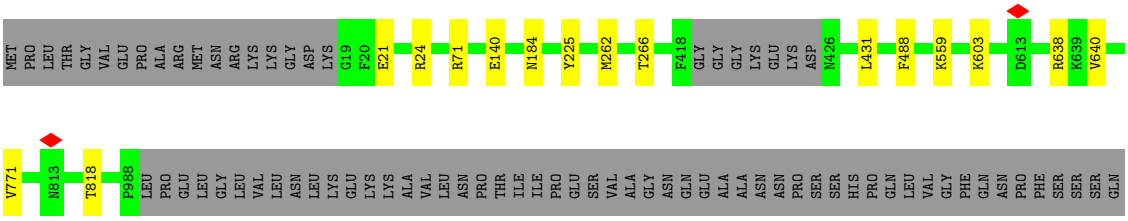
Chain A:

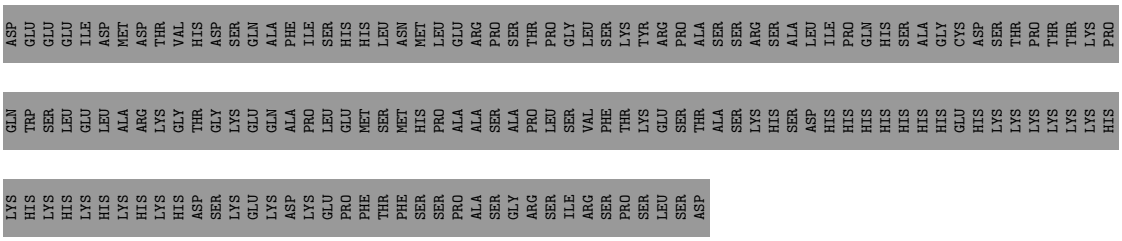




• Molecule 2: Transcription initiation factor TFIID subunit 2

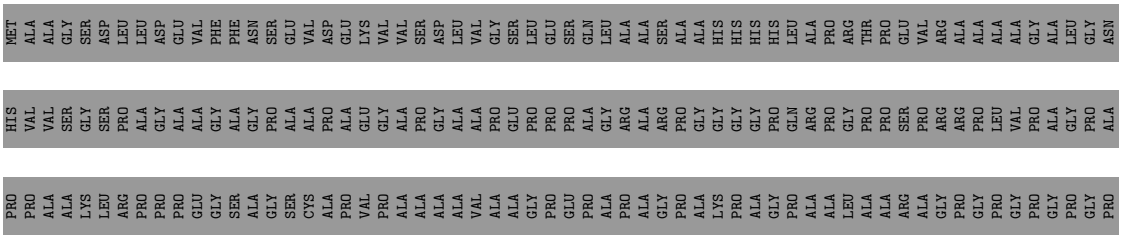
Chain B:

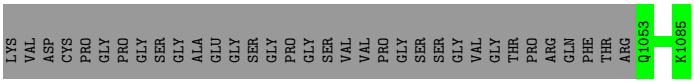




• Molecule 3: Transcription initiation factor TFIID subunit 4

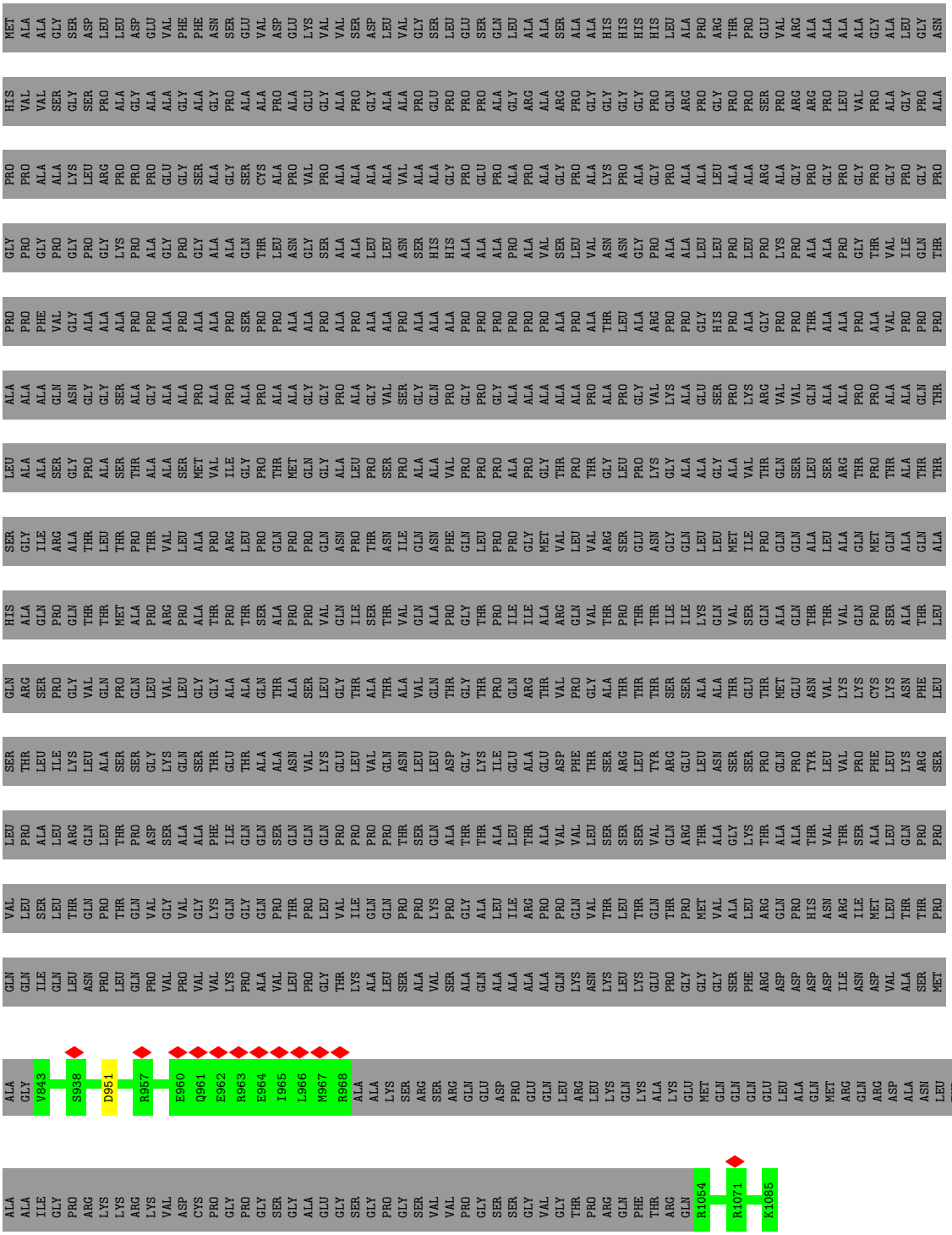
Chain D:





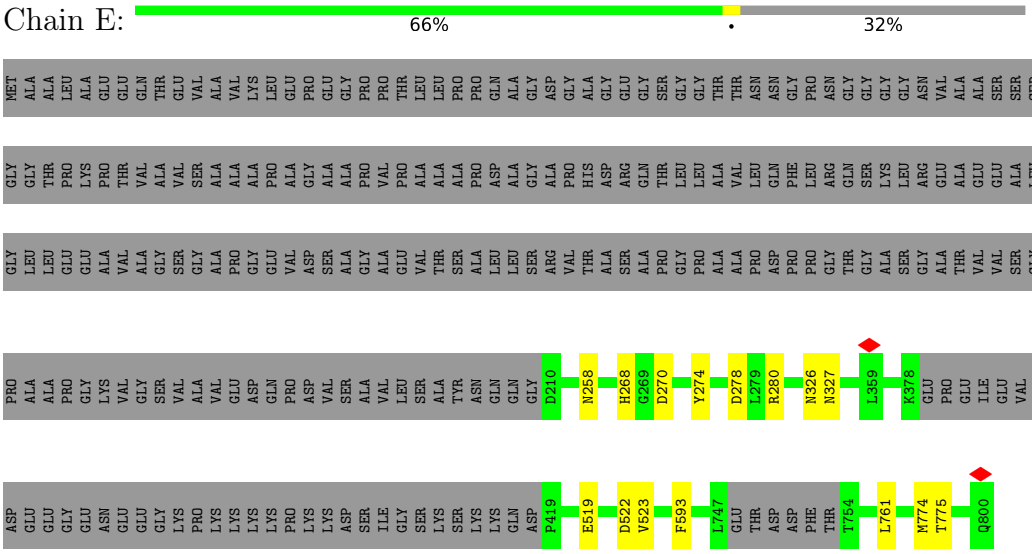
• Molecule 3: Transcription initiation factor TFIID subunit 4

Chain d:



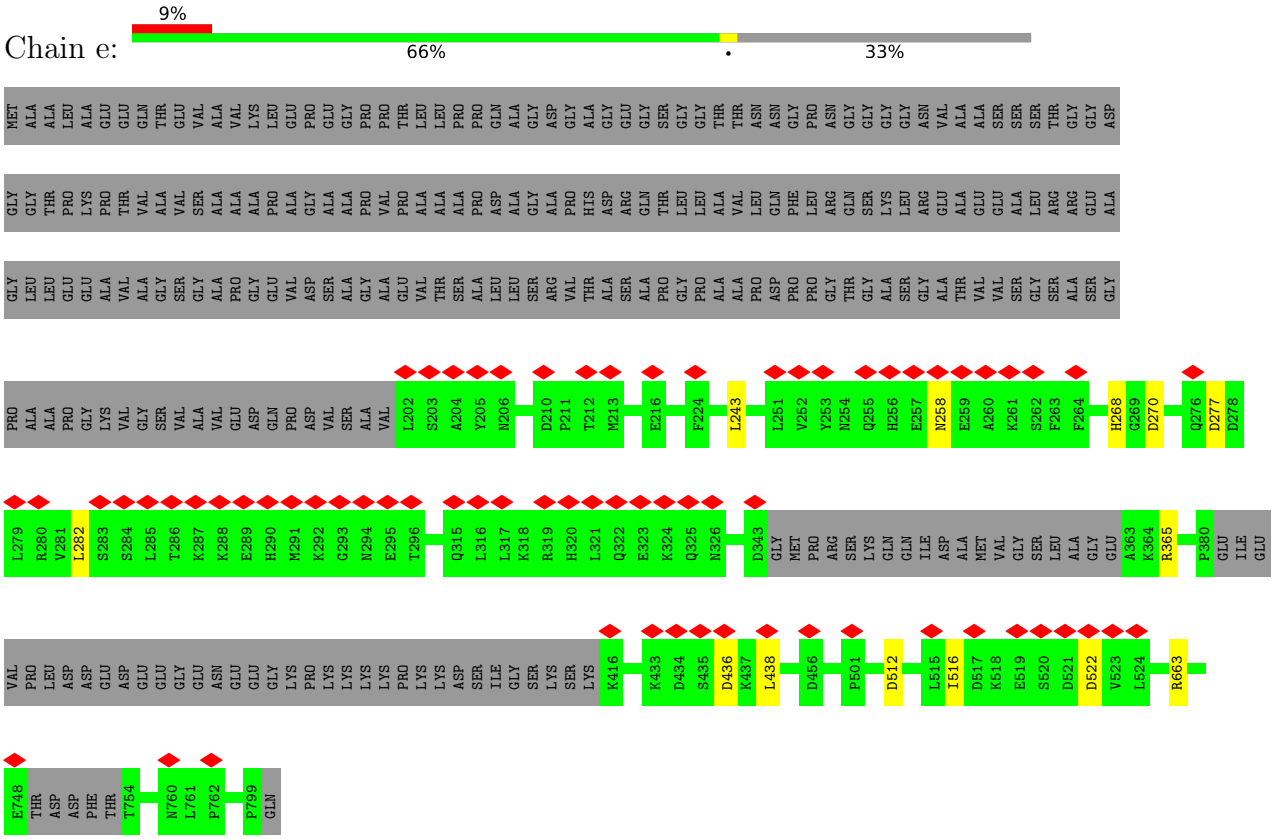
• Molecule 4: Transcription initiation factor TFIID subunit 5

Chain E:



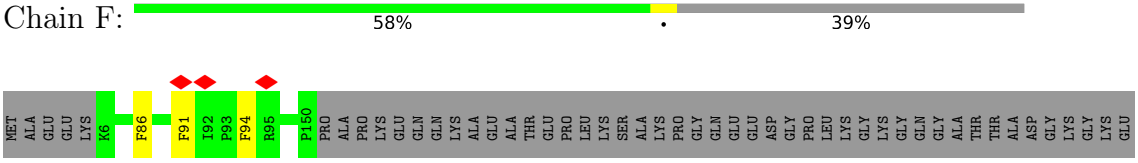
• Molecule 4: Transcription initiation factor TFIID subunit 5

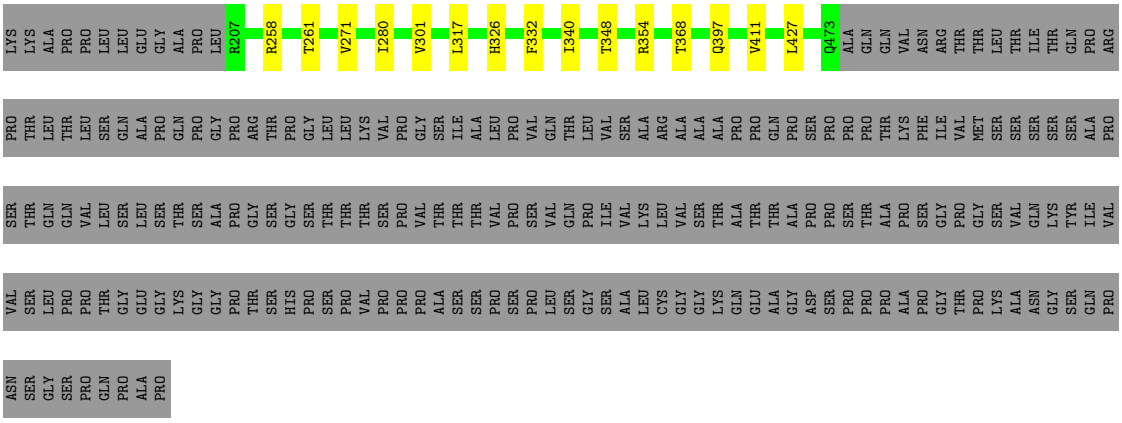
Chain e:



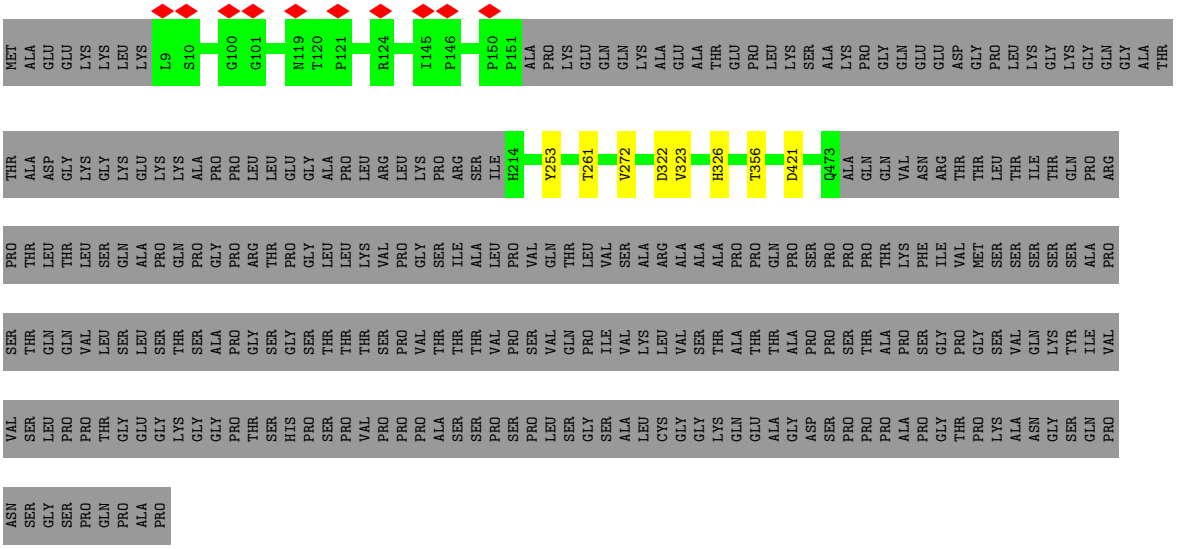
• Molecule 5: Transcription initiation factor TFIID subunit 6

Chain F:

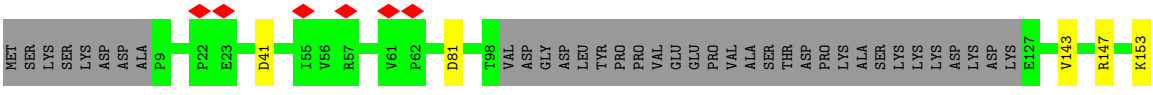


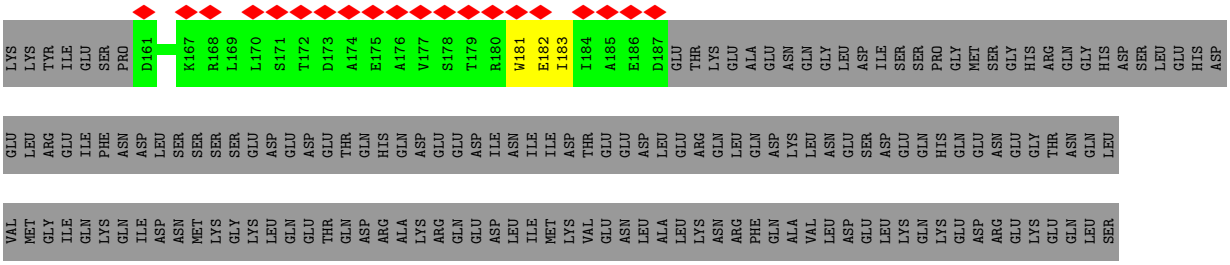


• Molecule 5: Transcription initiation factor TFIID subunit 6



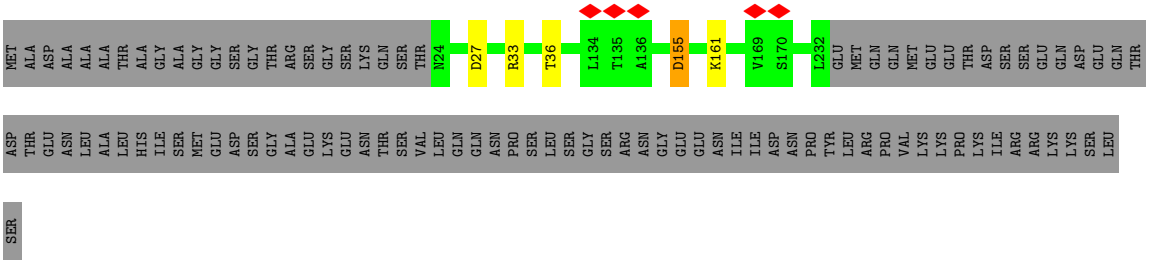
• Molecule 6: Transcription initiation factor TFIID subunit 7





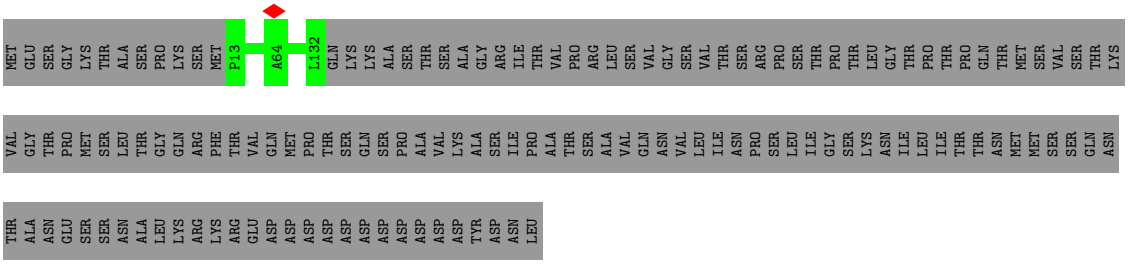
• Molecule 7: Transcription initiation factor TFIID subunit 8

Chain H:



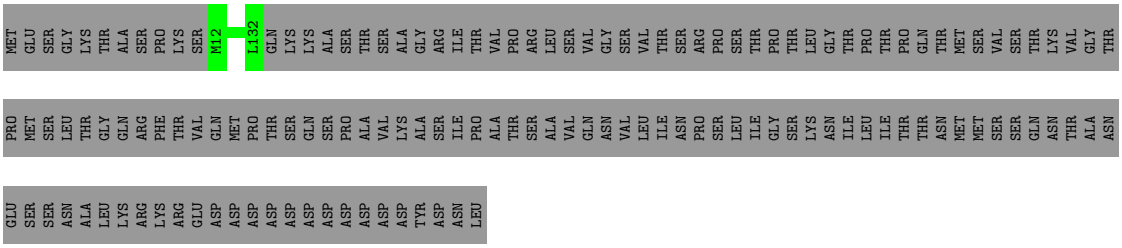
• Molecule 8: Transcription initiation factor TFIID subunit 9

Chain I:



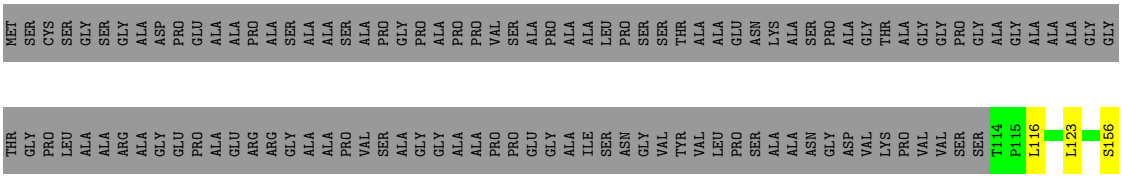
• Molecule 8: Transcription initiation factor TFIID subunit 9

Chain i:

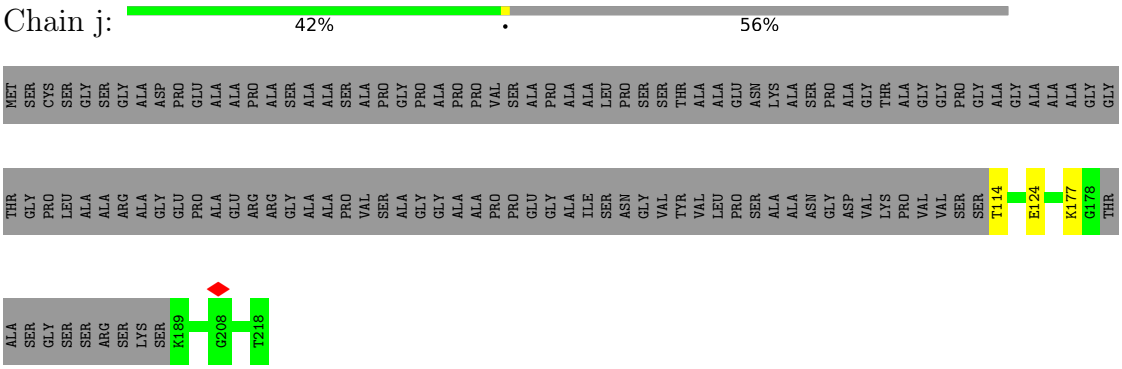


• Molecule 9: Transcription initiation factor TFIID subunit 10

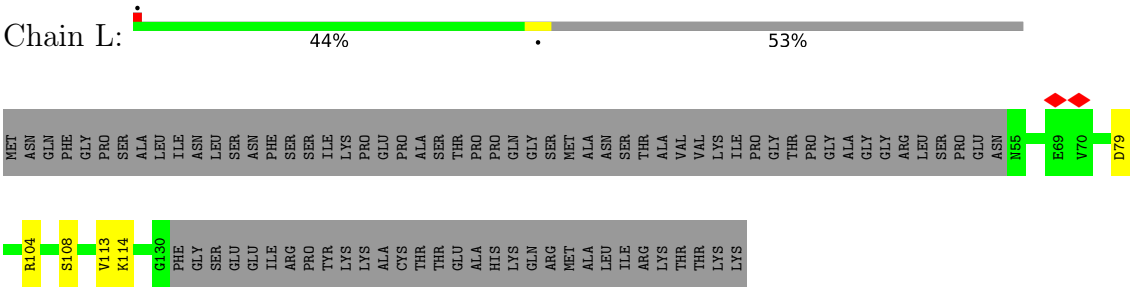
Chain J:



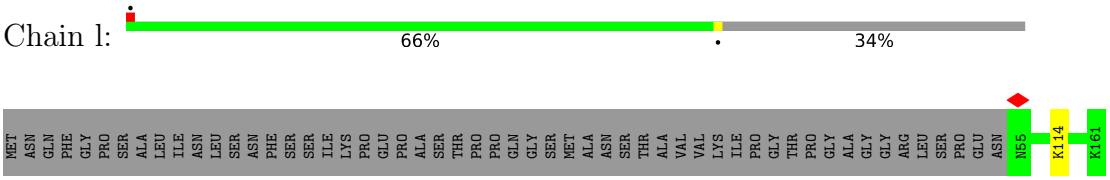
• Molecule 9: Transcription initiation factor TFIID subunit 10



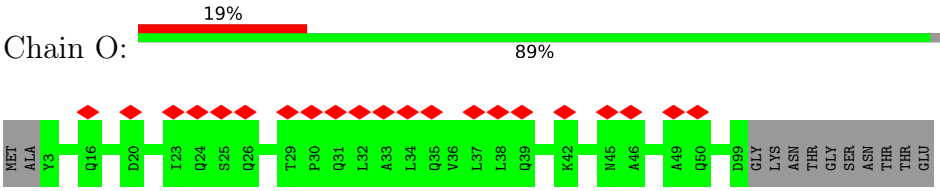
• Molecule 10: Transcription initiation factor TFIID subunit 12



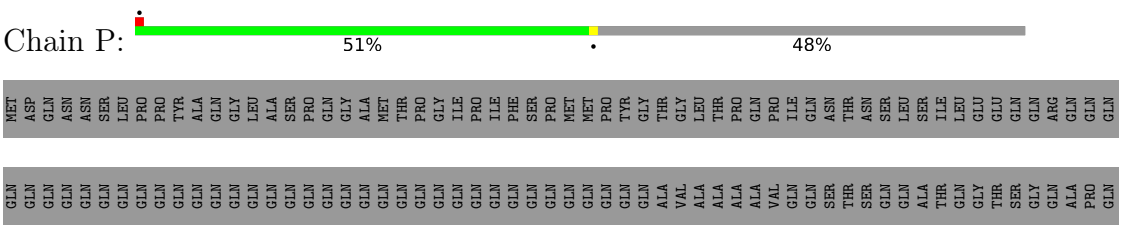
• Molecule 10: Transcription initiation factor TFIID subunit 12

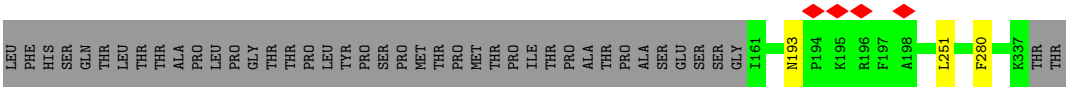


• Molecule 11: Transcription initiation factor IIA subunit 2



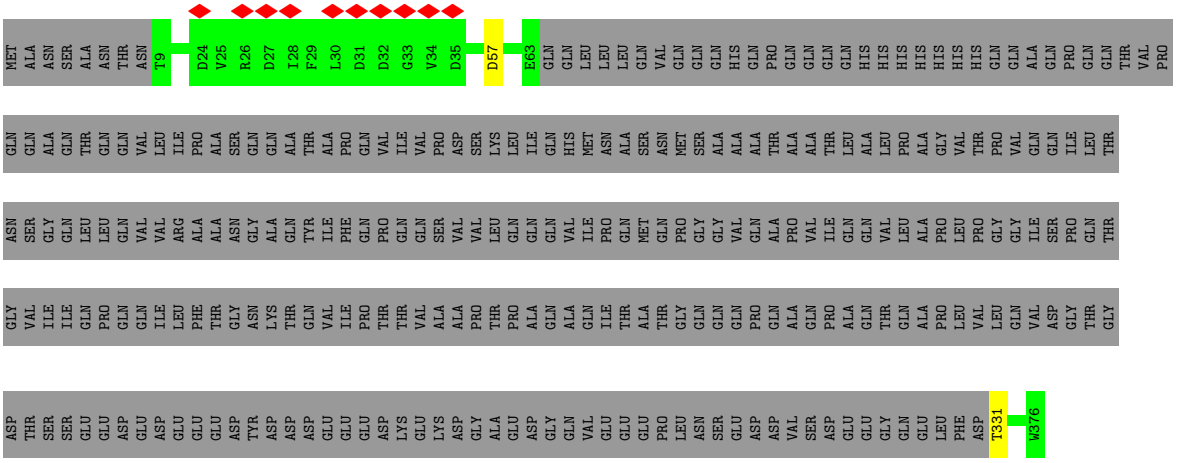
• Molecule 12: TATA-box-binding protein





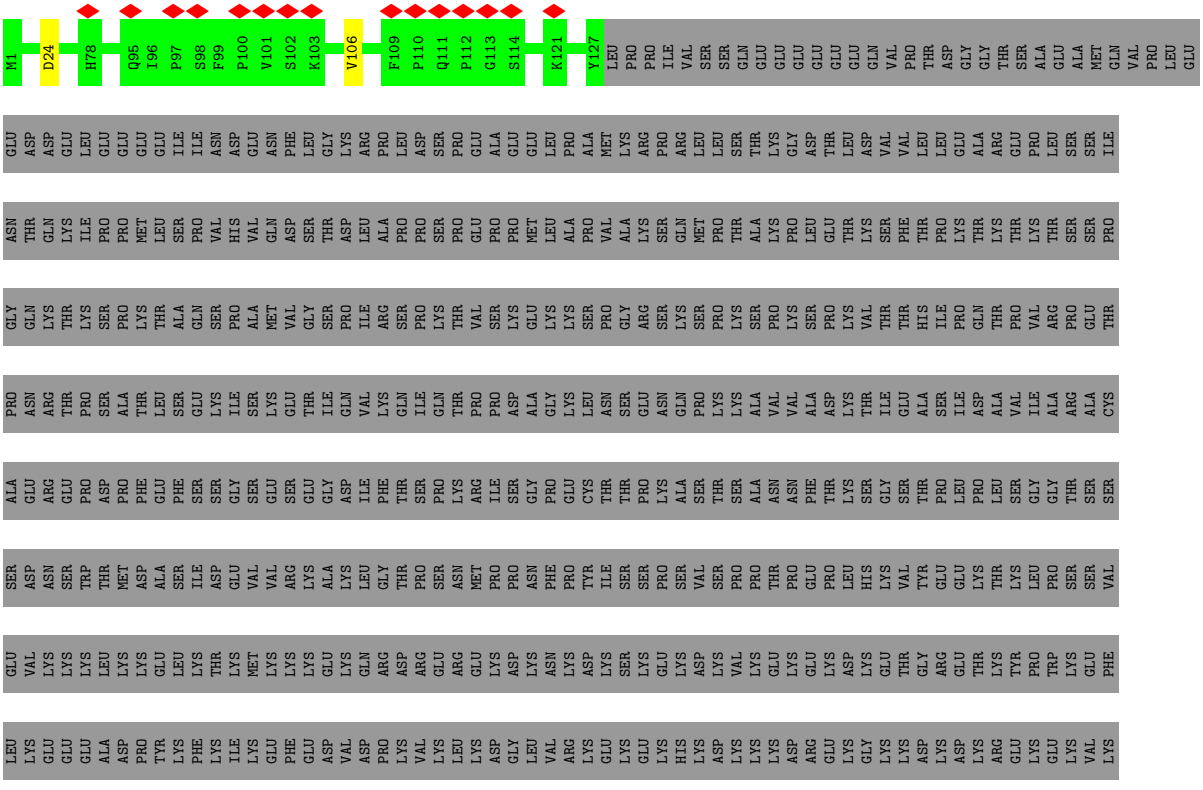
• Molecule 13: Transcription initiation factor IIA subunit 1

Chain Q:



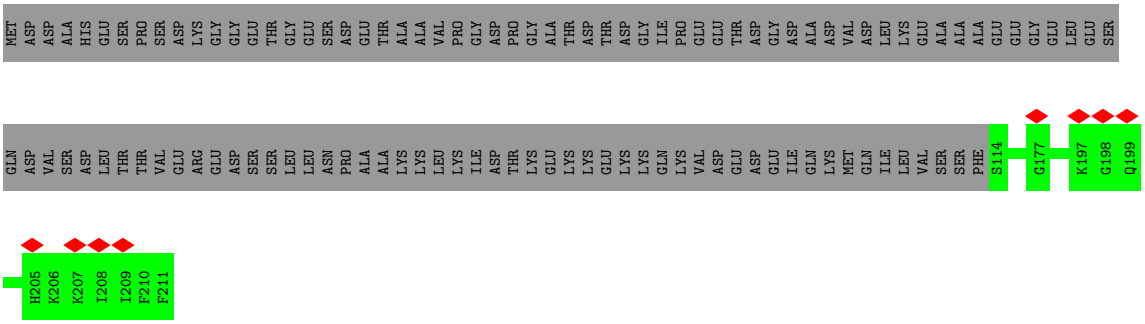
• Molecule 14: Transcription initiation factor TFIID subunit 3

Chain c:



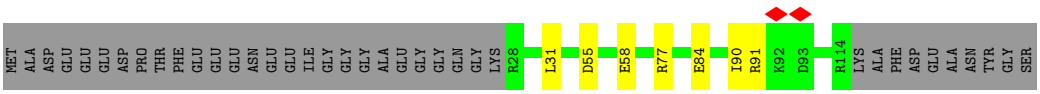
• Molecule 15: Transcription initiation factor TFIID subunit 11

Chain k:



• Molecule 16: Transcription initiation factor TFIID subunit 13

Chain m:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	64496	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 ($e^-/Å^2$)	50	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.040	Depositor
Minimum map value	-0.012	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.002	Depositor
Recommended contour level	0.01	Depositor
Map size (Å)	432.0, 432.0, 432.0	wwPDB
Map dimensions	320, 320, 320	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.35, 1.35, 1.35	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	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.50	0/4621	0.67	0/6240
2	B	0.46	0/7993	0.60	0/10836
3	D	0.43	0/1628	0.53	0/2178
3	d	0.39	0/1321	0.53	0/1772
4	E	0.39	0/4465	0.59	0/6045
4	e	0.43	0/4433	0.59	0/6004
5	F	0.48	0/3201	0.69	0/4347
5	f	0.40	0/3140	0.63	0/4268
6	G	0.51	0/1190	0.62	0/1601
7	H	0.49	0/1662	0.67	0/2272
8	I	0.27	0/981	0.47	0/1332
8	i	0.29	0/989	0.46	0/1343
9	J	0.57	0/736	0.69	0/998
9	j	0.53	0/775	0.63	0/1049
10	L	0.48	0/630	0.71	0/852
10	l	0.44	0/888	0.55	0/1194
11	O	0.34	0/781	0.62	0/1061
12	P	0.46	0/1438	0.57	0/1935
13	Q	0.36	0/842	0.63	0/1134
14	c	0.39	0/1035	0.54	0/1406
15	k	0.30	0/799	0.47	0/1070
16	m	0.59	0/733	0.64	0/977
All	All	0.44	0/44281	0.61	0/59914

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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
1	A	533/1872 (28%)	514 (96%)	18 (3%)	1 (0%)	47 81
2	B	959/1199 (80%)	912 (95%)	47 (5%)	0	100 100
3	D	191/1085 (18%)	182 (95%)	9 (5%)	0	100 100
3	d	154/1085 (14%)	150 (97%)	4 (3%)	0	100 100
4	E	539/800 (67%)	515 (96%)	23 (4%)	1 (0%)	47 81
4	e	531/800 (66%)	482 (91%)	48 (9%)	1 (0%)	47 81
5	F	408/677 (60%)	390 (96%)	17 (4%)	1 (0%)	47 81
5	f	399/677 (59%)	383 (96%)	16 (4%)	0	100 100
6	G	138/349 (40%)	135 (98%)	3 (2%)	0	100 100
7	H	207/310 (67%)	193 (93%)	13 (6%)	1 (0%)	29 69
8	I	118/264 (45%)	115 (98%)	3 (2%)	0	100 100
8	i	119/264 (45%)	115 (97%)	4 (3%)	0	100 100
9	J	86/218 (39%)	82 (95%)	4 (5%)	0	100 100
9	j	91/218 (42%)	85 (93%)	4 (4%)	2 (2%)	6 35
10	L	74/161 (46%)	72 (97%)	2 (3%)	0	100 100
10	l	105/161 (65%)	100 (95%)	5 (5%)	0	100 100
11	O	95/109 (87%)	86 (90%)	9 (10%)	0	100 100
12	P	175/339 (52%)	165 (94%)	9 (5%)	1 (1%)	25 66
13	Q	97/376 (26%)	89 (92%)	8 (8%)	0	100 100
14	c	125/929 (14%)	116 (93%)	9 (7%)	0	100 100
15	k	96/211 (46%)	91 (95%)	5 (5%)	0	100 100
16	m	85/124 (68%)	82 (96%)	3 (4%)	0	100 100
All	All	5325/12228 (44%)	5054 (95%)	263 (5%)	8 (0%)	50 81

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	E	523	VAL
4	e	522	ASP
9	j	124	GLU
5	F	411	VAL
9	j	177	LYS

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
1	A	487/1665 (29%)	448 (92%)	39 (8%)	12 35
2	B	876/1083 (81%)	860 (98%)	16 (2%)	59 77
3	D	175/815 (22%)	167 (95%)	8 (5%)	27 52
3	d	146/815 (18%)	145 (99%)	1 (1%)	84 90
4	E	478/657 (73%)	464 (97%)	14 (3%)	42 64
4	e	475/657 (72%)	463 (98%)	12 (2%)	47 68
5	F	328/574 (57%)	311 (95%)	17 (5%)	23 48
5	f	322/574 (56%)	314 (98%)	8 (2%)	47 68
6	G	132/322 (41%)	124 (94%)	8 (6%)	18 44
7	H	178/270 (66%)	173 (97%)	5 (3%)	43 65
8	I	106/235 (45%)	106 (100%)	0	100 100
8	i	107/235 (46%)	107 (100%)	0	100 100
9	J	79/154 (51%)	72 (91%)	7 (9%)	9 30
9	j	83/154 (54%)	82 (99%)	1 (1%)	71 83
10	L	71/141 (50%)	66 (93%)	5 (7%)	15 40
10	l	98/141 (70%)	97 (99%)	1 (1%)	76 86
11	O	84/98 (86%)	84 (100%)	0	100 100
12	P	153/293 (52%)	151 (99%)	2 (1%)	69 81
13	Q	91/324 (28%)	89 (98%)	2 (2%)	52 71
14	c	113/833 (14%)	111 (98%)	2 (2%)	59 77

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
15	k	87/182 (48%)	87 (100%)	0	100 100
16	m	80/106 (76%)	73 (91%)	7 (9%)	10 31
All	All	4749/10328 (46%)	4594 (97%)	155 (3%)	41 61

5 of 155 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
10	L	104	ARG
5	f	326	HIS
12	P	193	ASN
4	e	277	ASP
16	m	58	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 74 such sidechains are listed below:

Mol	Chain	Res	Type
13	Q	37	GLN
10	l	119	HIS
3	d	943	GLN
5	f	325	ASN
3	D	936	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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 monosaccharides 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)

There are no chain breaks in this entry.

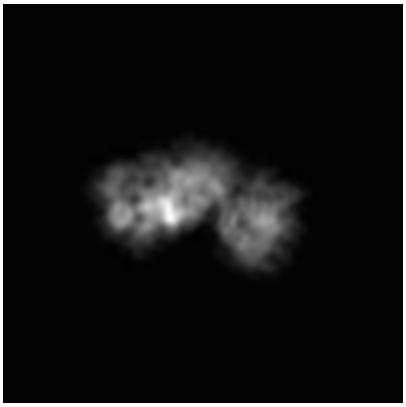
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-31118. These allow visual inspection of the internal detail of the map and identification of artifacts.

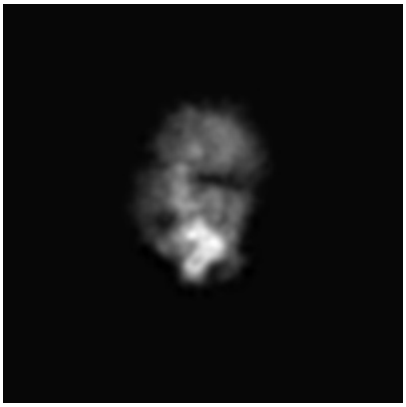
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

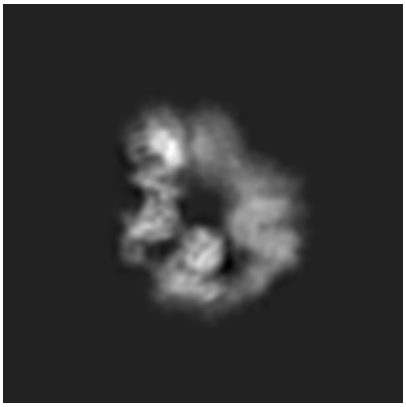
6.1.1 Primary map



X



Y

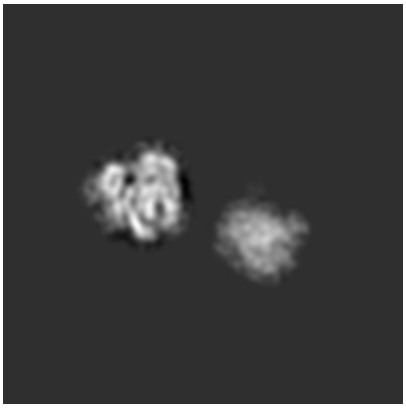


Z

The images above show the map projected in three orthogonal directions.

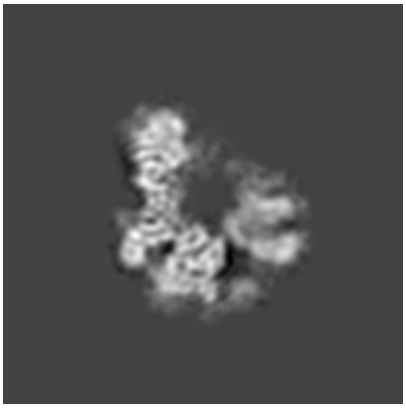
6.2 Central slices

6.2.1 Primary map



X Index: 160

Y Index: 160



Z Index: 160

The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

6.3.1 Primary map

X Index: 132

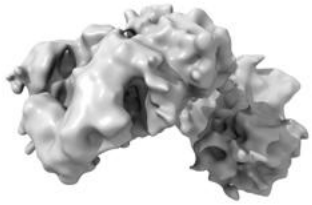
Y Index: 133

Z Index: 155

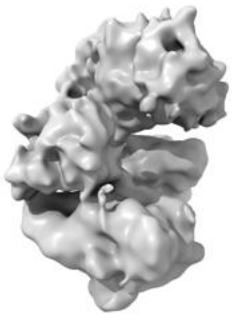
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal surface views (i)

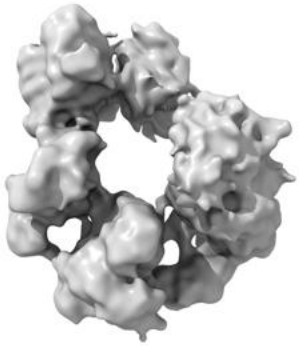
6.4.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.01. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

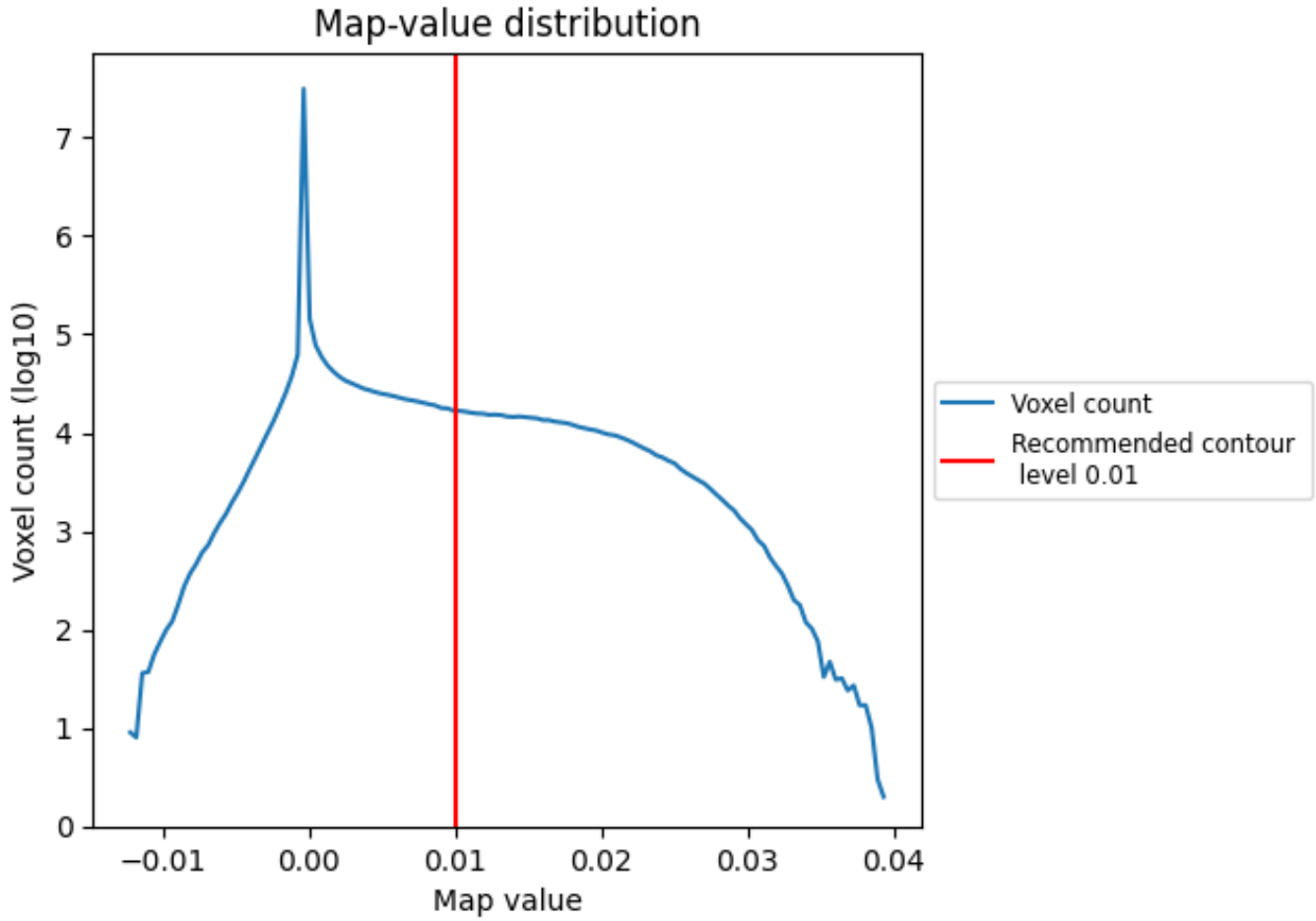
6.5 Mask visualisation (i)

This section was not generated. No masks/segmentation were deposited.

7 Map analysis

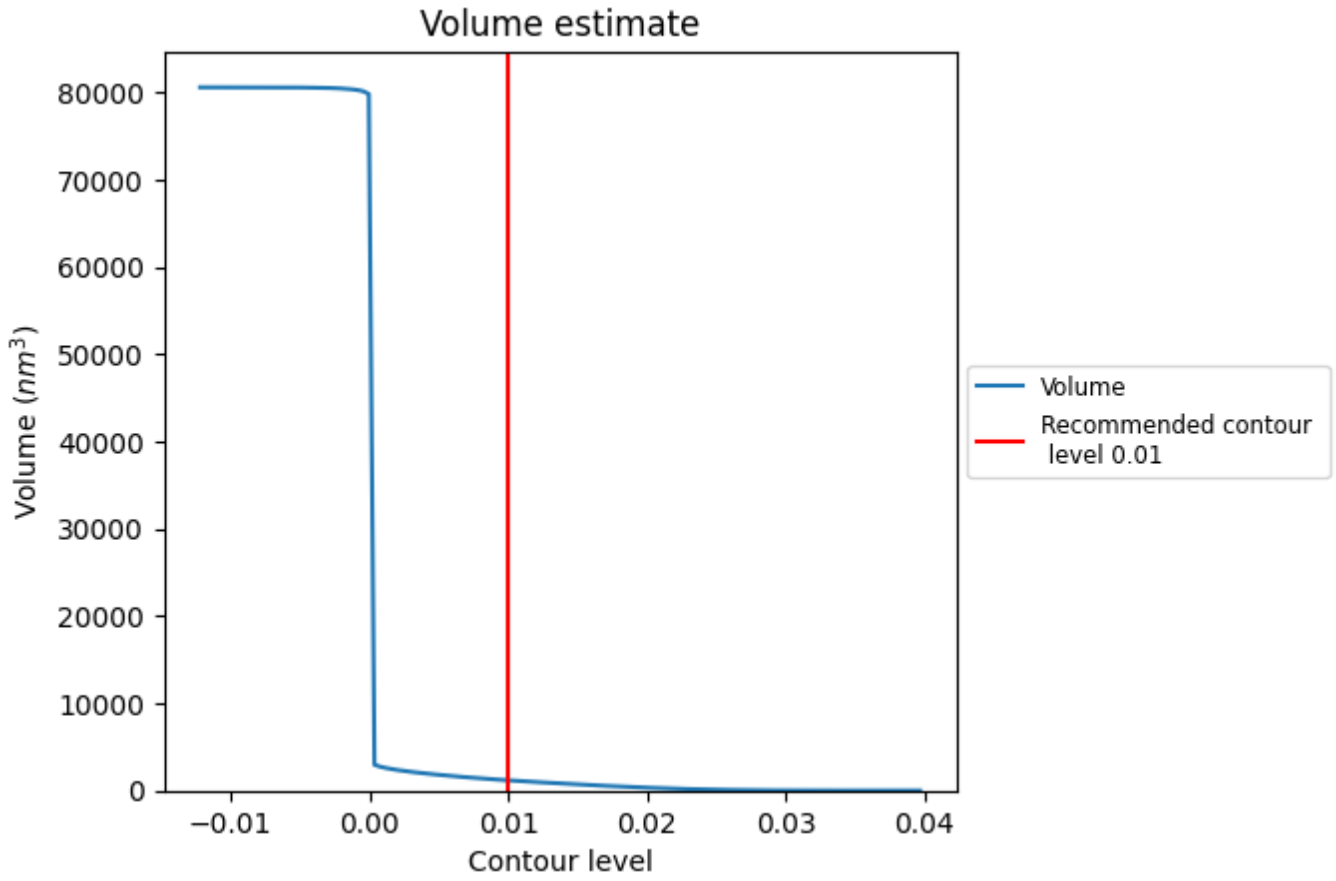
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

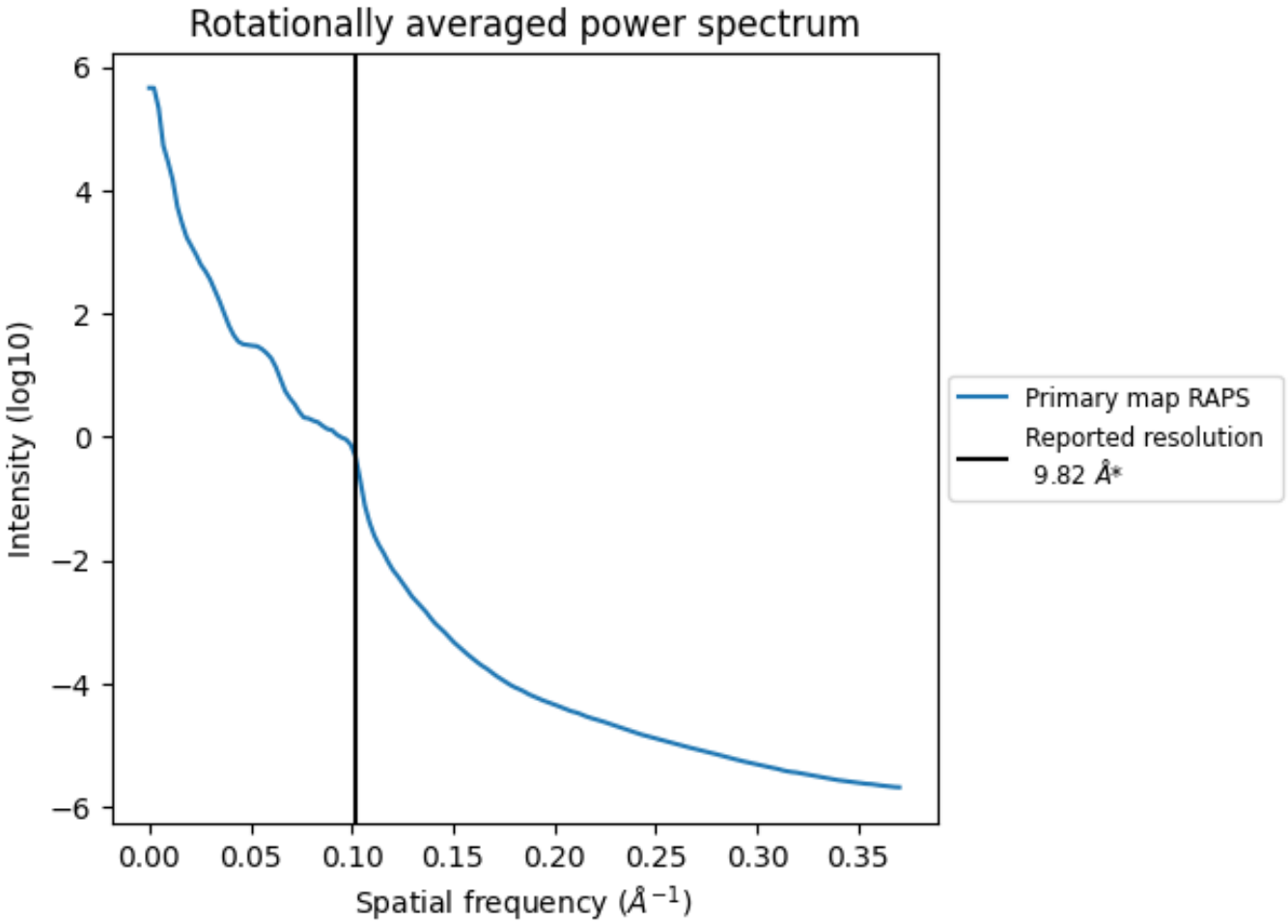
7.2 Volume estimate (i)



The volume at the recommended contour level is 1175 nm^3; this corresponds to an approximate mass of 1062 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum (i)



*Reported resolution corresponds to spatial frequency of 0.102 Å^{-1}

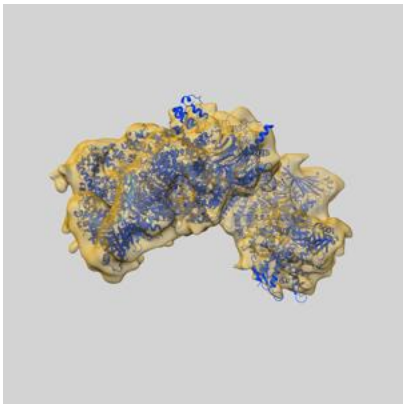
8 Fourier-Shell correlation

This section was not generated. No FSC curve or half-maps provided.

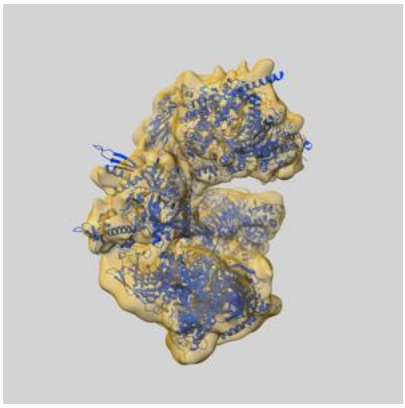
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-31118 and PDB model 7EGI. Per-residue inclusion information can be found in section 3 on page 7.

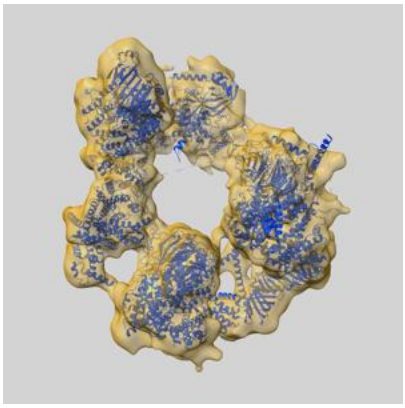
9.1 Map-model overlay



X



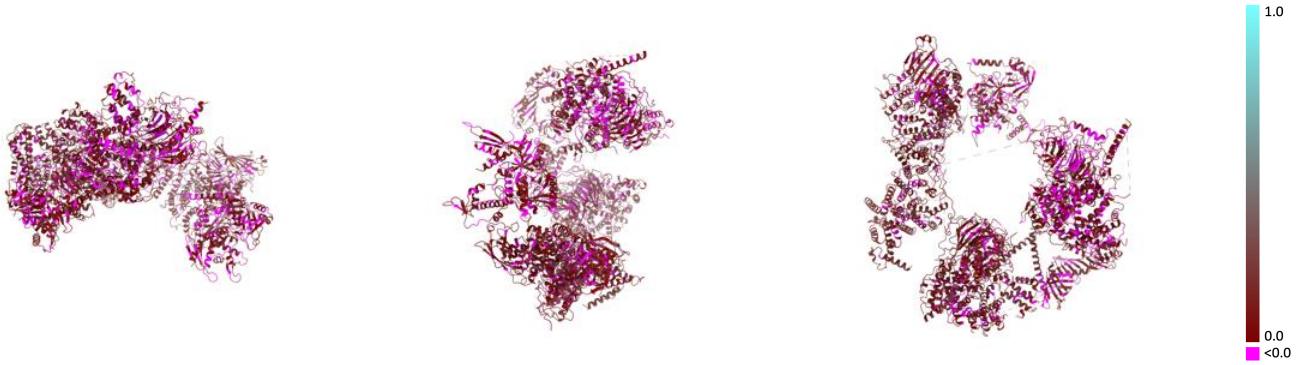
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.01 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

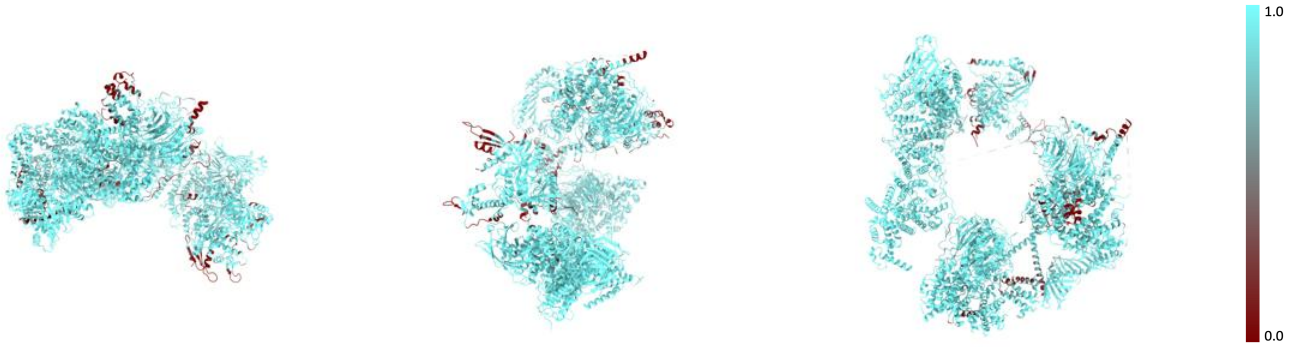
9.2 Q-score mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

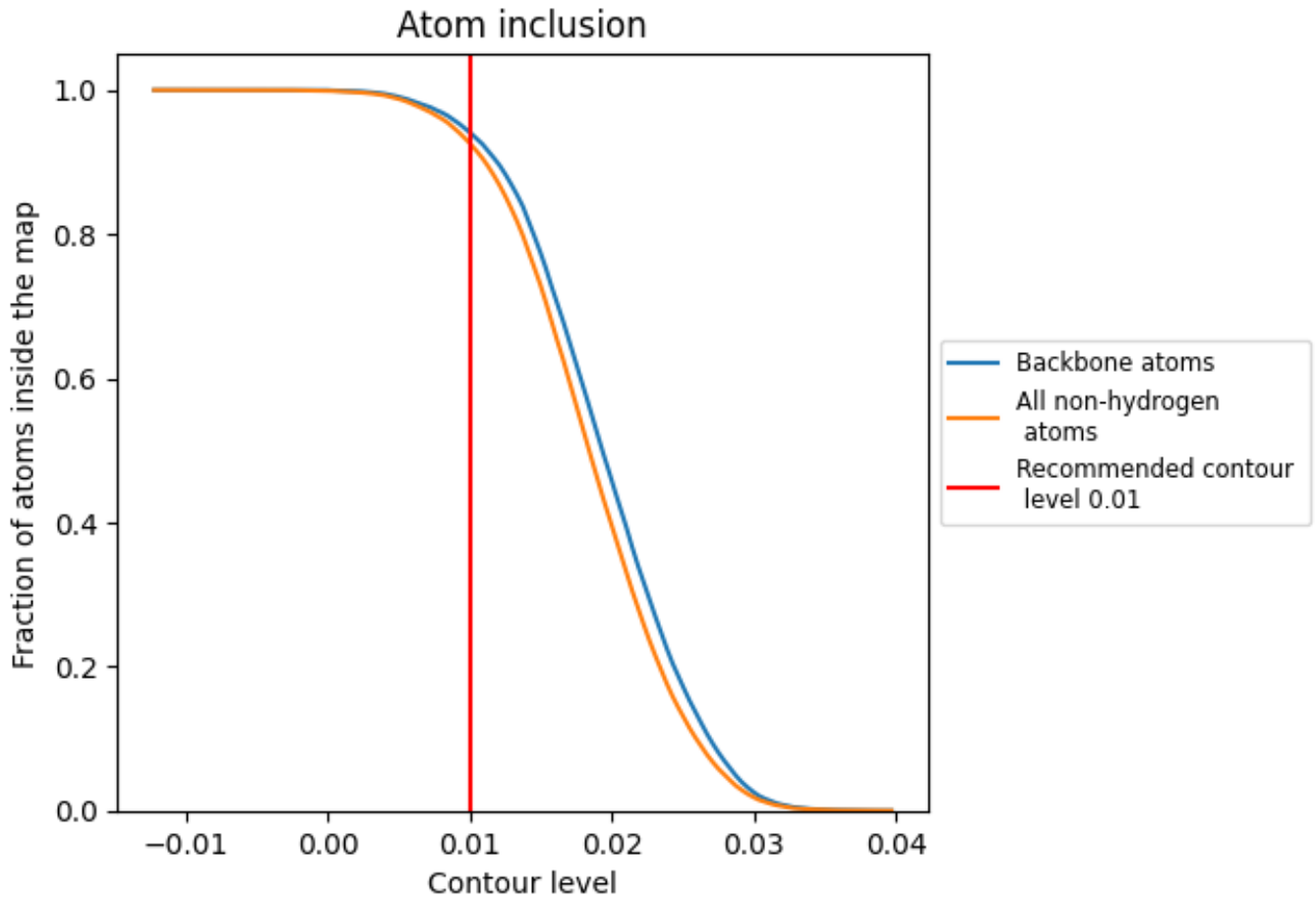
9.3 Atom inclusion mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.01).

9.4 Atom inclusion (i)



At the recommended contour level, 94% of all backbone atoms, 93% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (0.01) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.9267	0.0790
A	0.8375	0.0620
B	0.9799	0.0950
D	0.8532	0.0940
E	0.9728	0.0850
F	0.9702	0.1150
G	0.8273	0.0320
H	0.9716	0.1260
I	0.9711	0.1130
J	0.9958	0.1020
L	0.9755	0.1130
O	0.7741	0.0740
P	0.9652	0.0700
Q	0.8868	0.0520
c	0.8687	0.0420
d	0.9027	0.0490
e	0.8480	0.0310
f	0.9590	0.0980
i	0.9777	0.0700
j	0.9664	0.0590
k	0.8958	0.0510
l	0.9697	0.0740
m	0.9501	0.0700