



## wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Jan 21, 2020 – 12:39 PM EST

PDB ID : 6TB4  
EMDB ID: : EMD-10438  
Title : Structure of SAGA bound to TBP  
Authors : Papai, G.; Frechard, A.; Kolesnikova, O.; Crucifix, C.; Schultz, P.; Ben-Shem, A.  
Deposited on : 2019-10-31  
Resolution : 3.80 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

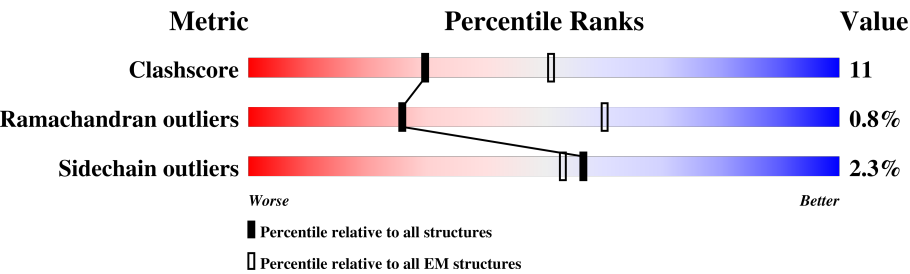
MolProbity : 4.02b-467  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.4

# 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 3.80 Å.

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



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

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  $\geq 3$ , 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	M	243	37% 31% 6% 26%
2	A	448	31% 7% 61%
3	C	698	8% . 91%
4	F	517	33% 7% 59%
5	D	341	37% 21% . 39%
6	E	1191	10% . . 87%
7	J	217	36% 7% 56%
8	K	609	17% 7% . 75%
9	G	722	52% 20% . 28%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	H	485	<div><div></div><div>69%17%13%</div></div>
11	I	153	<div><div></div><div>65%16%20%</div></div>
12	L	3825	<div><div></div><div>66%11%22%</div></div>
13	B	76	<div><div></div><div>95%5%</div></div>

## 2 Entry composition

There are 13 unique types of molecules in this entry. The entry contains 40740 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 TATA-box Binding Protein (TBP).

Mol	Chain	Residues	Atoms					AltConf	Trace
1	M	180	Total	C	N	O	S	0	0
			1415	921	242	246	6		

- Molecule 2 is a protein called Transcriptional coactivator HFI1/ADA1.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	A	173	Total	C	N	O	S	0	0
			1300	816	228	250	6		

- Molecule 3 is a protein called SAGA-associated factor 73 (Sgf73).

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	65	Total	C	N	O	S	0	0
			518	331	94	90	3		

- Molecule 4 is a protein called Spt20.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	F	210	Total	C	N	O	S	0	0
			1682	1071	292	315	4		

- Molecule 5 is a protein called Subunit of the SAGA and SAGA-like transcriptional regulatory complexes, interacts with Spt15p to act.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	D	209	Total	C	N	O	S	0	0
			1616	1016	298	295	7		

- Molecule 6 is a protein called Subunit of the SAGA transcriptional regulatory complex, involved in proper assembly of the complex.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	E	154	Total	C	N	O	S	0	0
			1232	784	208	233	7		

- Molecule 7 is a protein called Transcription initiation factor TFIID subunit 10.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	J	96	Total	C	N	O	S	0	0
			768	489	120	156	3		

- Molecule 8 is a protein called Subunit (61/68 kDa) of TFIID and SAGA complexes.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	K	154	Total	C	N	O	S	0	0
			1192	747	216	226	3		

- Molecule 9 is a protein called Subunit (90 kDa) of TFIID and SAGA complexes.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	G	522	Total	C	N	O	S	0	0
			4075	2581	719	756	19		

- Molecule 10 is a protein called Subunit (60 kDa) of TFIID and SAGA complexes.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	H	421	Total	C	N	O	S	0	0
			3263	2084	556	617	6		

- Molecule 11 is a protein called Subunit (17 kDa) of TFIID and SAGA complexes, involved in RNA polymerase II transcription initiation.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	I	123	Total	C	N	O	S	0	0
			981	632	169	178	2		

- Molecule 12 is a protein called Transcription-associated protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	L	2968	Total	C	N	O	S	0	0
			22318	14296	3864	4071	87		

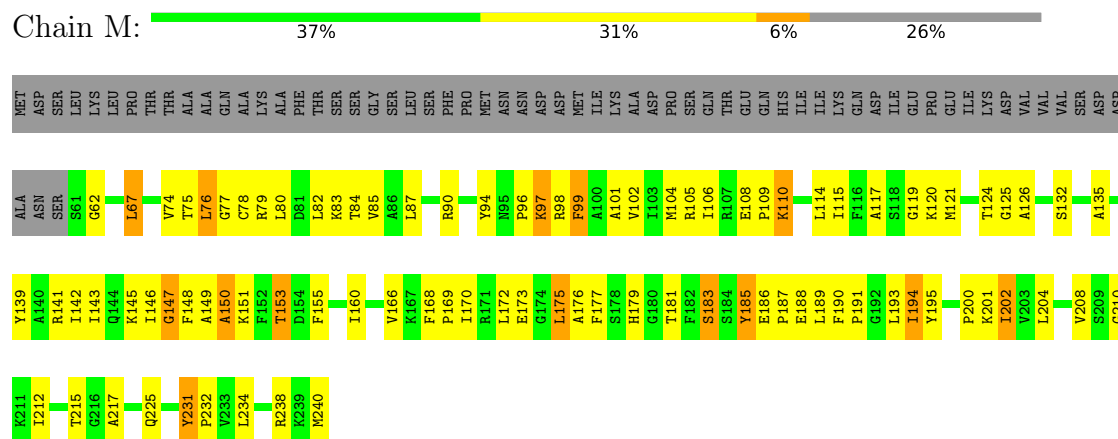
- Molecule 13 is a protein called Transcriptional adapter 3 (Ada3).

Mol	Chain	Residues	Atoms				AltConf	Trace
13	B	76	Total	C	N	O	0	0
			380	228	76	76		

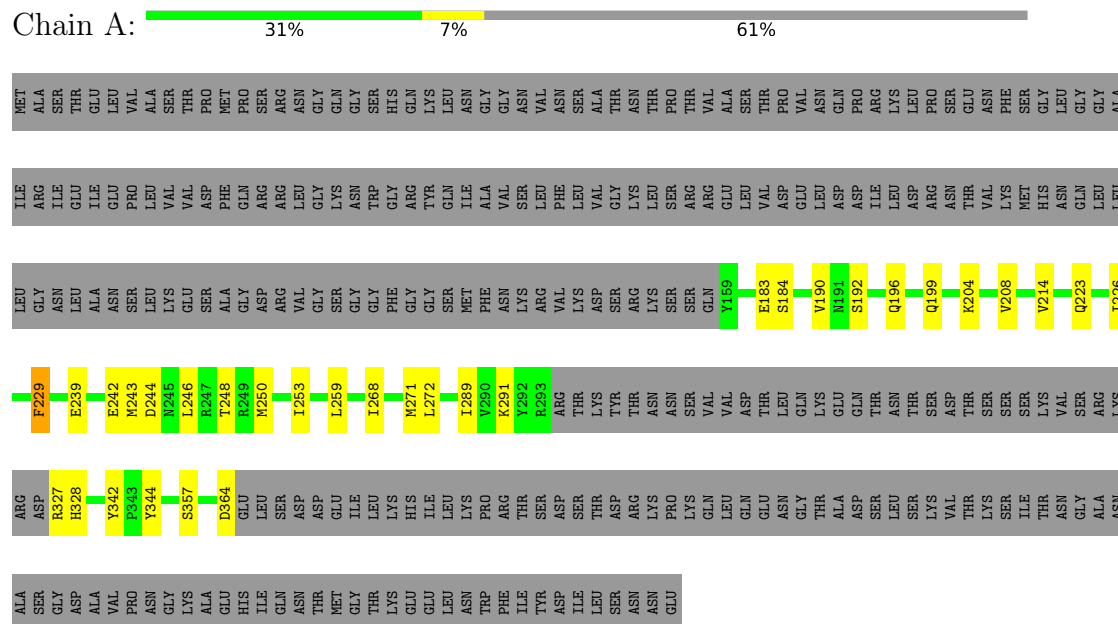
### 3 Residue-property plots

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: TATA-box Binding Protein (TBP)



#### • Molecule 2: Transcriptional coactivator HFI1/ADA1



#### • Molecule 3: SAGA-associated factor 73 (Sgf73)



[illegible]

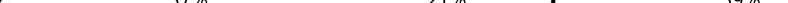
- Molecule 4: Spt20

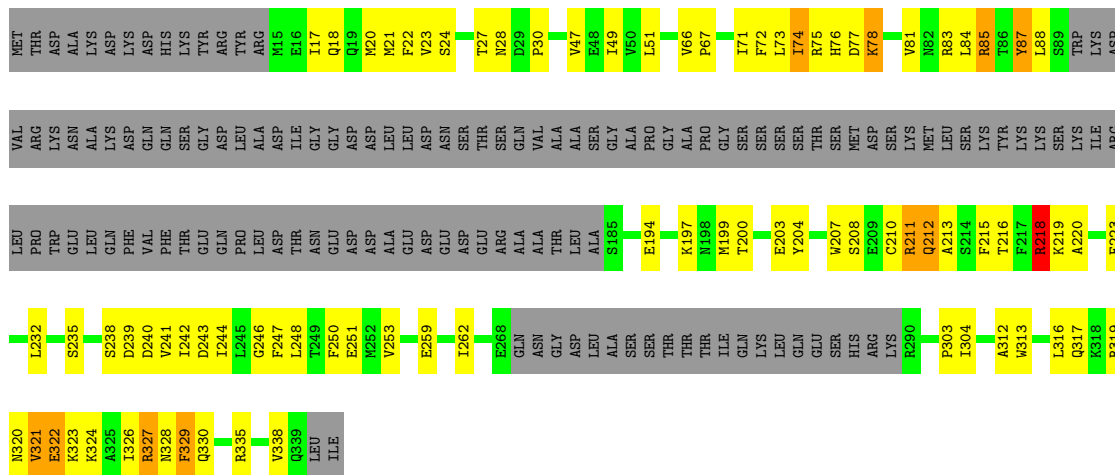
Chain F:  33% 7% 59%

GLN	ASN	ASP	THR	THR	LYS	MET
ASN	ASP	ASP	ASN	ASN	ASN	SER
HIS	THR	SER	ASN	GLU	TYR	GLN
GLN	GLN	GLN	LYS	LYS	X64	ILE
GLN	GLN	GLN	LYS	LYS	P80	GLN
GLN	GLY	GLY	GLU	GLU	L83	SER
ALA	ALA	SER	PRO	SER	T84	GLN
ASN	ALA	SER	SER	LYS	F85	VAL
GLY	ILE	VAL	LYS	VAL	E89	GLY
ASP	GLN	THR	VAL	VAL		ALA
GLU	GLN	SER	PRO	PRO	B89	LYS
LEU	ARG	SER	K184	K184		ASN
LEU	SER	GLN			R93	ASN
SER	MET	GLN	I189	I189	F94	MET
SER	ILE	F324	M325	L190	G95	GLN
LEU	GLN	R326	R191	R191		THR
ASP	GLN				F111	ALA
ASP	GLY					GLN
LYS	ARG	K396	L195	L195		PRO
LYS	ALA	GLU			V115	GLN
ILE	GLU	GLU	D200	D200		ALA
PRO	GLN	ARG	L201	L201	E118	GLN
LYS	GLN	LEU	L202	L202	E119	GLN
PRO	GLN	GLN	Y203	Y203		ARG
PRO	GLN	GLU	Q204	Q204	P122	PRO
ALA	GLN	SER	T205	T205		ILE
LYS	PRO	GLY			K129	ASN
LYS	ASP	GLY	F212	F212		GLY
GLN	GLN	SER			V133	SER
GLY	GLU	SER	L218	L218	Q134	VAL
LYS	GLN	SER	N219	N219		THR
MET	LEU	ARG			E137	LEU
THR	GLN	GLN	T222	T222		SER
LYS	GLU	ARG	E223	E223	L142	ASN
LYS	GLN	PRO			R143	GLY
GLY	GLN	LEU	N277	N277		GLN
MET	GLN	GLN	A278	A278	R149	ARG
THR	GLN	PRO	D279	D279	HIS	ILE
THR	GLN	GLY			LEU	ASN
LEU	PRO	GLN	D282	D282	VAL	PRO
GLN	GLN	ILE	TYR	TYR	ALA	GLN
HIS	GLN	PRO	ARG	ARG	HIS	ASN
GLN	ALA	GLY	LYS	LYS	ILE	LEU
ASN	MET	MET	LYS	LYS	ASN	THR
VAL	MET	MET	LYS	LYS	ASP	PRO
ALA	SER	LEU	HIS	HIS	ALA	GLN
GLY	GLN	SER	GLU	GLU	LYS	GLN
LYS	GLN	SER	ASP	ASP	ASP	GLN
THR	PHE	MET	MET	MET	ASP	ARG
ILE	GLN	SER	ALA	ALA	SER	GLN
PRO	GLN	SER	GLN	GLN	GLN	LEU
SER	GLN	ASN	HIS	HIS	VAL	LEU
ALA	LEU	THR	GLY	GLY	SER	GLN
GLY	GLY	GLY	S295	S295	SER	GLN
SER	ASN	ASN			SER	LYS
THR	LEU	MET	L304	L304	GLU	VAL
PRO	TYR	ASN	S305	S305	SER	LEU
PRO	GLN	THR	D306	D306	SER	HIS
THR	ARG	MET	R307	R307	ALA	GLN
THR	GLN	LEU	PHE	PHE	ASN	LYS
ALA	GLN	ASP	ASP	ASP	ASN	LEU



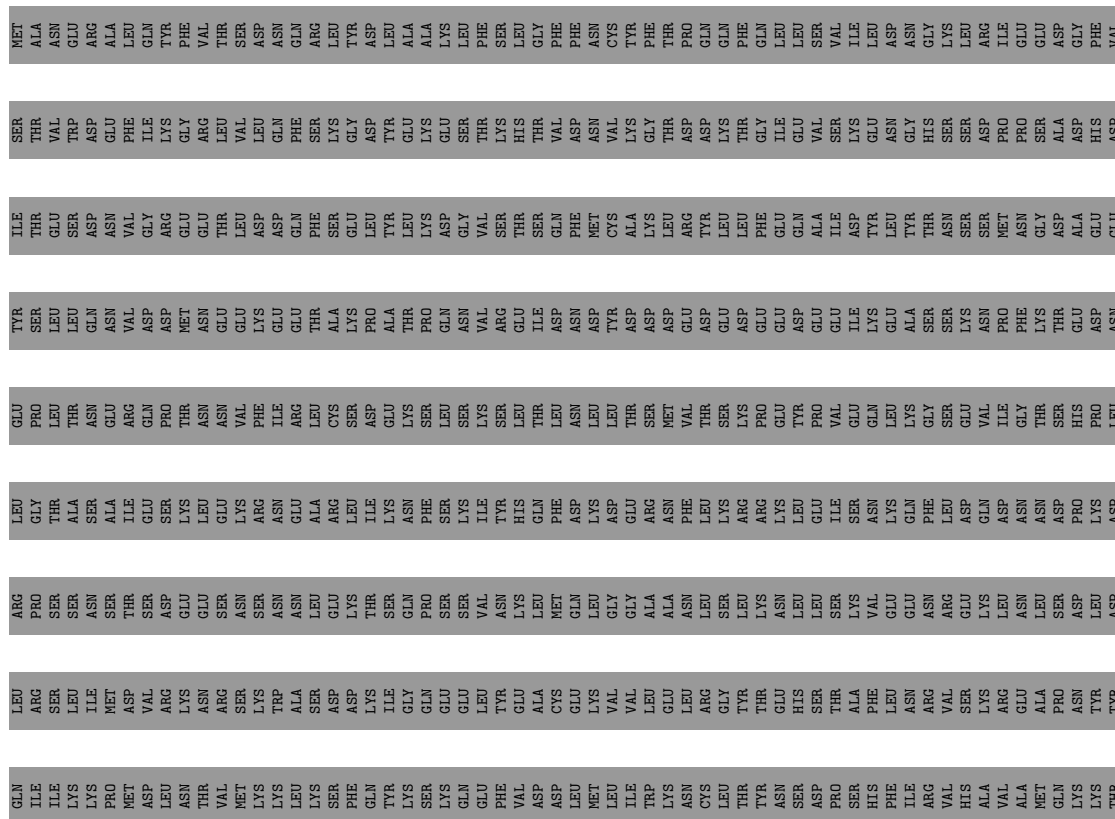
- Molecule 5: Subunit of the SAGA and SAGA-like transcriptional regulatory complexes, interacts with Spt15p to act

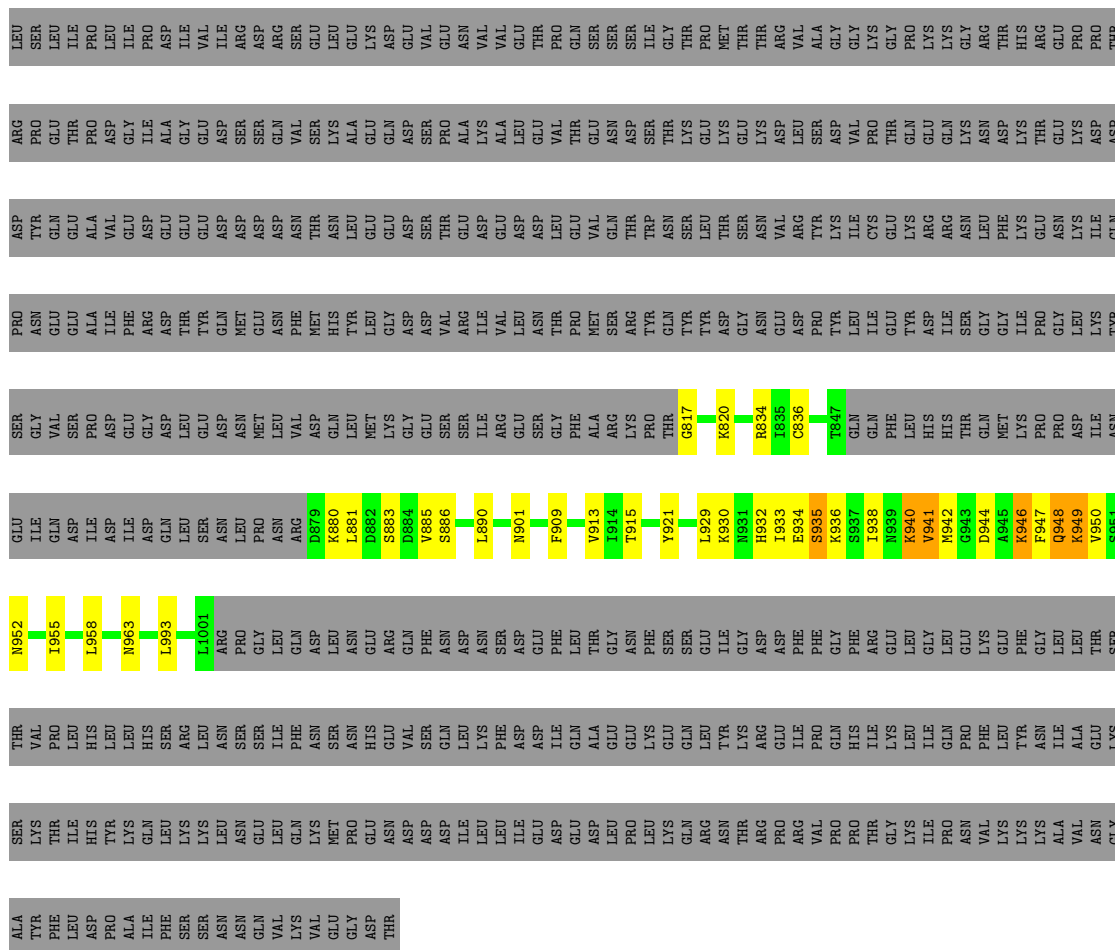
Chain D:  37% 21% . 39%



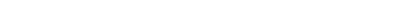
- Molecule 6: Subunit of the SAGA transcriptional regulatory complex, involved in proper assembly of the complex

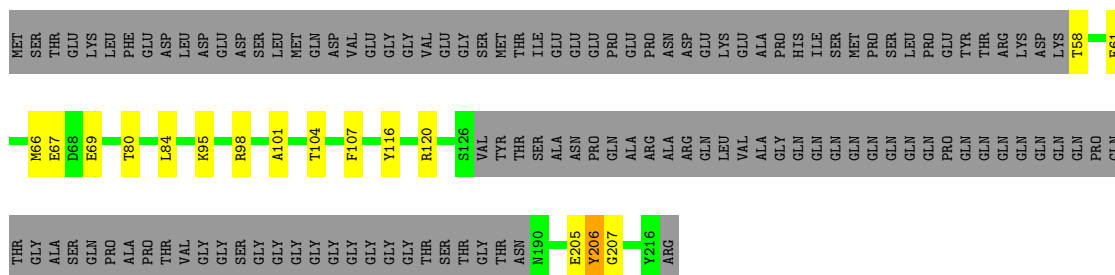
Chain E:  10% .. 87%





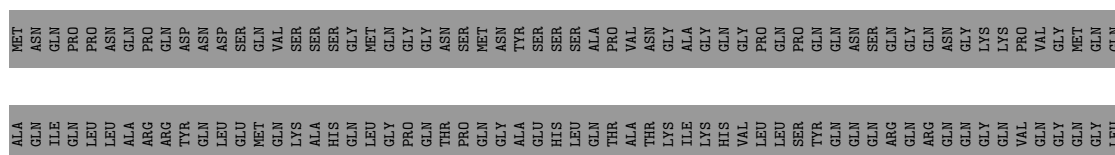
- Molecule 7: Transcription initiation factor TFIID subunit 10

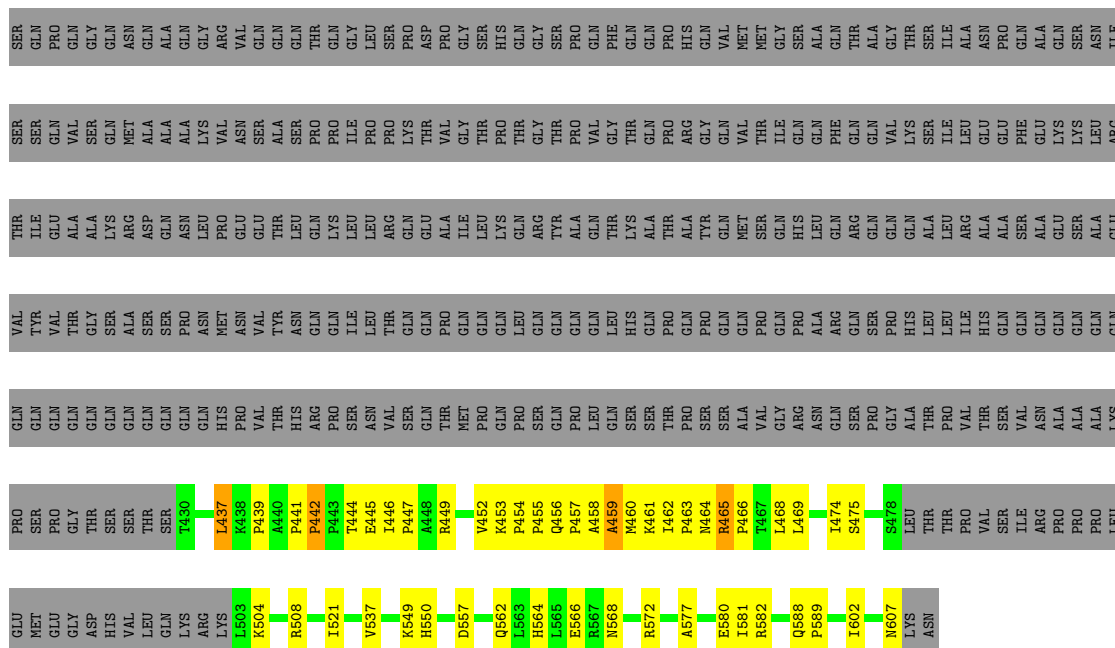
Chain J:  36% 7% 56%



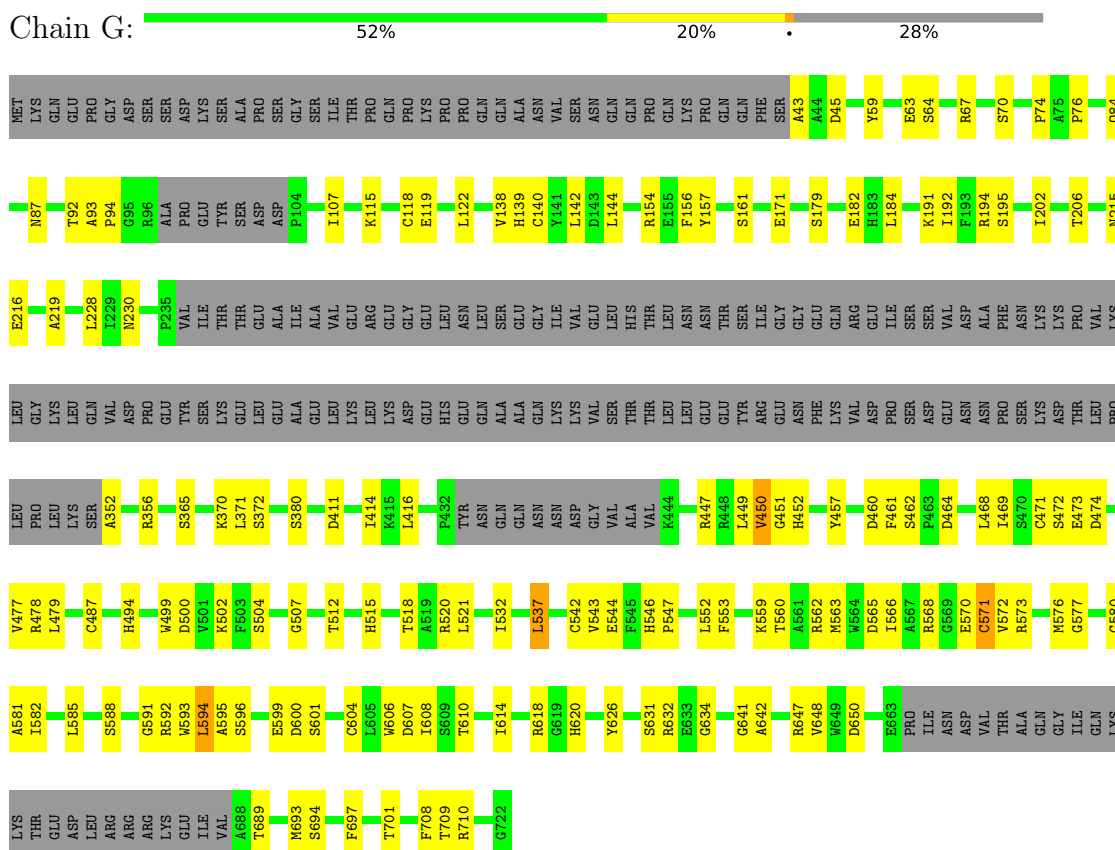
- Molecule 8: Subunit (61/68 kDa) of TFIID and SAGA complexes

Chain K:  17% 7% . 75%



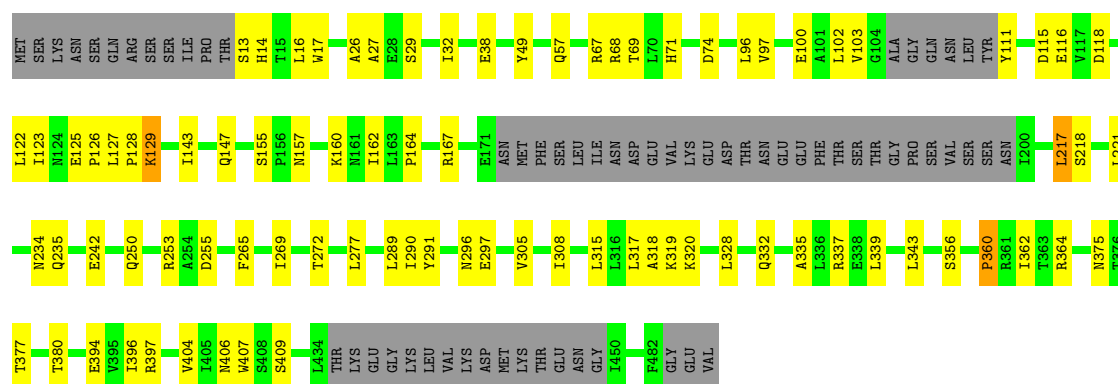


- Molecule 9: Subunit (90 kDa) of TFIID and SAGA complexes

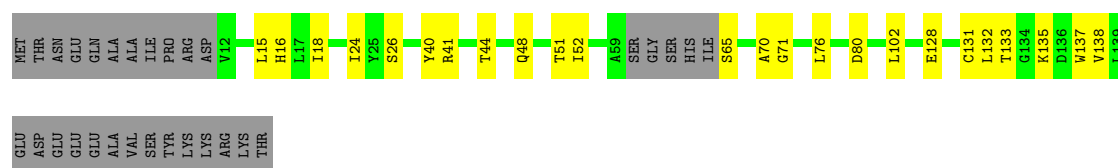


- Molecule 10: Subunit (60 kDa) of TFIID and SAGA complexes

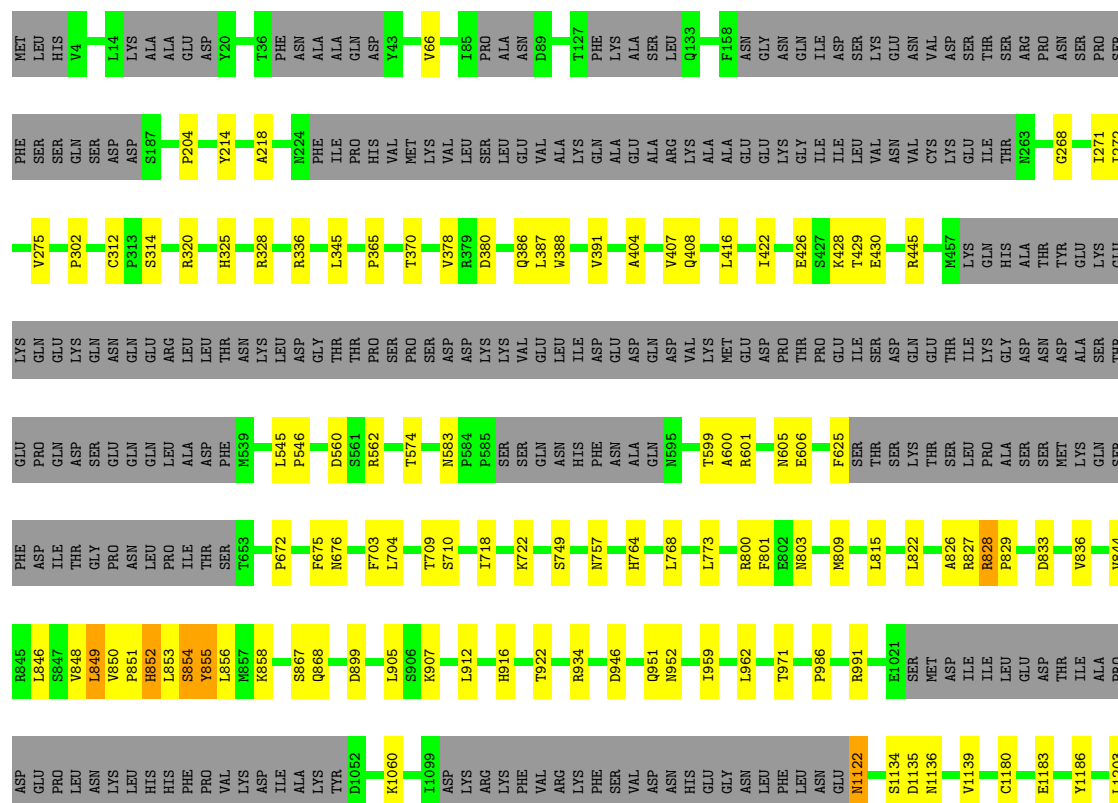




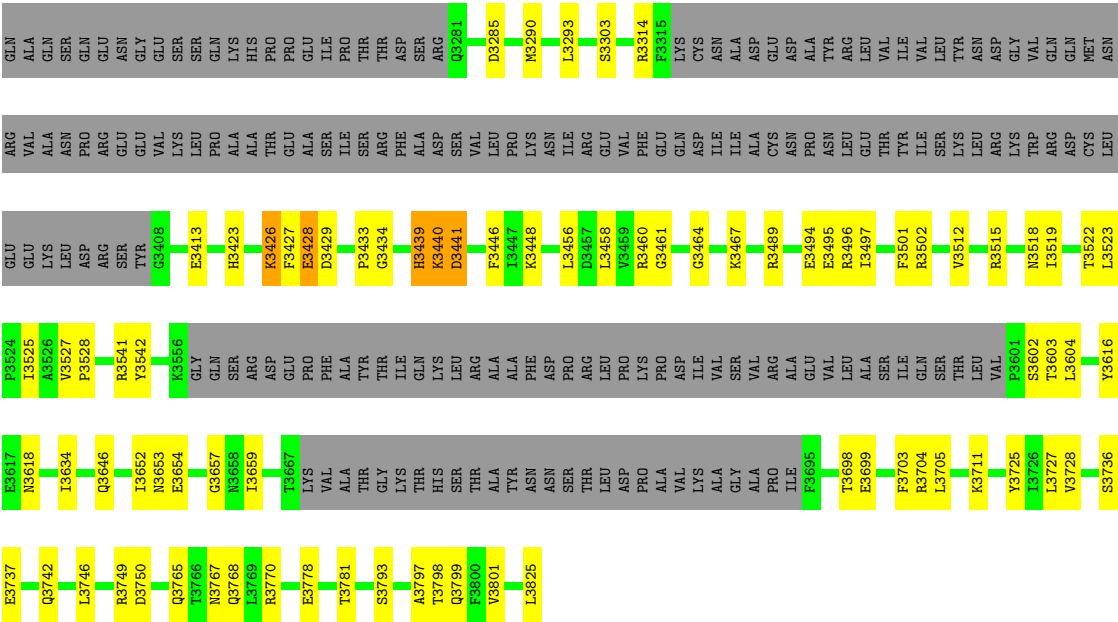
- Molecule 11: Subunit (17 kDa) of TFIID and SAGA complexes, involved in RNA polymerase II transcription initiation



- Molecule 12: Transcription-associated protein





● Molecule 13: Transcriptional adapter 3 (Ada3)

Chain B: 

95%

5%



## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	354104	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 ( $e^-/\text{\AA}^2$ )	52.8	Depositor
Minimum defocus (nm)	0.8	Depositor
Maximum defocus (nm)	4.5	Depositor
Magnification	105000	Depositor
Image detector	GATAN K2 QUANTUM (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	M	0.88	2/1442 (0.1%)	0.78	2/1942 (0.1%)
10	H	0.38	0/3315	0.60	0/4500
11	I	0.44	0/1006	0.63	0/1374
12	L	0.33	0/22712	0.54	0/30825
2	A	0.44	0/1319	0.60	0/1794
3	C	0.35	0/528	0.61	0/710
4	F	0.34	0/1718	0.58	0/2335
5	D	0.51	0/1641	0.65	0/2213
6	E	0.45	0/1246	0.62	0/1667
7	J	0.47	0/779	0.60	0/1051
8	K	0.42	0/1213	0.66	0/1647
9	G	0.52	0/4177	0.60	0/5661
All	All	0.41	2/41096 (0.0%)	0.58	2/55719 (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	M	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	M	194	ILE	C-N	27.38	1.97	1.34
1	M	183	SER	C-N	6.38	1.48	1.34

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	M	62	GLY	N-CA-C	6.94	130.46	113.10
1	M	194	ILE	CA-C-N	-6.13	103.71	117.20



There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	M	183	SER	Mainchain

## 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	M	1415	0	1492	88	0
2	A	1300	0	1254	30	0
3	C	518	0	528	9	0
4	F	1682	0	1622	30	0
5	D	1616	0	1558	122	0
6	E	1232	0	1276	40	0
7	J	768	0	754	11	0
8	K	1192	0	1214	89	0
9	G	4075	0	3934	102	0
10	H	3263	0	3258	69	0
11	I	981	0	982	18	0
12	L	22318	0	20960	376	0
13	B	380	0	85	2	0
All	All	40740	0	38917	865	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 865 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)
8:K:469:LEU:HD13	12:L:2788:ILE:CG2	1.31	1.51
8:K:469:LEU:CD1	12:L:2788:ILE:HG21	1.57	1.34
12:L:828:ARG:H	12:L:829:PRO:HD3	1.06	1.20
1:M:194:ILE:HG22	1:M:195:TYR:N	1.57	1.19
1:M:194:ILE:C	1:M:195:TYR:N	1.97	1.17

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	M	178/243 (73%)	147 (83%)	24 (14%)	7 (4%)	3	33
2	A	169/448 (38%)	151 (89%)	18 (11%)	0	100	100
3	C	61/698 (9%)	56 (92%)	5 (8%)	0	100	100
4	F	202/517 (39%)	177 (88%)	24 (12%)	1 (0%)	31	71
5	D	203/341 (60%)	177 (87%)	19 (9%)	7 (3%)	4	37
6	E	150/1191 (13%)	135 (90%)	13 (9%)	2 (1%)	13	54
7	J	92/217 (42%)	82 (89%)	9 (10%)	1 (1%)	16	57
8	K	150/609 (25%)	125 (83%)	23 (15%)	2 (1%)	13	54
9	G	512/722 (71%)	448 (88%)	63 (12%)	1 (0%)	49	83
10	H	413/485 (85%)	365 (88%)	48 (12%)	0	100	100
11	I	119/153 (78%)	98 (82%)	21 (18%)	0	100	100
12	L	2874/3825 (75%)	2623 (91%)	233 (8%)	18 (1%)	27	67
All	All	5123/9449 (54%)	4584 (90%)	500 (10%)	39 (1%)	26	62

5 of 39 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	M	77	GLY
1	M	99	PHE
1	M	110	LYS
5	D	200	THR
5	D	327	ARG

### 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	M	152/208 (73%)	138 (91%)	14 (9%)	10	41
2	A	134/394 (34%)	132 (98%)	2 (2%)	67	85
3	C	55/627 (9%)	55 (100%)	0	100	100
4	F	179/471 (38%)	179 (100%)	0	100	100
5	D	166/306 (54%)	154 (93%)	12 (7%)	16	52
6	E	142/1101 (13%)	136 (96%)	6 (4%)	32	66
7	J	85/183 (46%)	84 (99%)	1 (1%)	74	88
8	K	133/524 (25%)	130 (98%)	3 (2%)	53	78
9	G	439/635 (69%)	430 (98%)	9 (2%)	56	80
10	H	352/438 (80%)	348 (99%)	4 (1%)	76	88
11	I	104/130 (80%)	103 (99%)	1 (1%)	78	89
12	L	2156/3450 (62%)	2114 (98%)	42 (2%)	60	82
All	All	4097/8467 (48%)	4003 (98%)	94 (2%)	57	78

5 of 94 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
9	G	537	LEU
12	L	429	THR
12	L	3314	ARG
9	G	542	CYS
10	H	129	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
9	G	391	ASN
10	H	64	HIS
12	L	2633	ASN
5	D	317	GLN
12	L	1632	HIS

### 5.3.3 RNA ⓘ

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 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
13	B	2
1	M	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	414:UNK	C	462:UNK	N	28.86
1	B	471:UNK	C	509:UNK	N	23.98
1	M	194:ILE	C	195:TYR	N	1.97