



wwPDB EM Validation Summary Report ⓘ

Oct 7, 2024 – 02:21 PM JST

PDB ID : 8H7G
EMDB ID : EMD-34520
Title : Cryo-EM structure of the human SAGA complex
Authors : Huang, J.; Zhang, Y.
Deposited on : 2022-10-20
Resolution : 3.70 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:

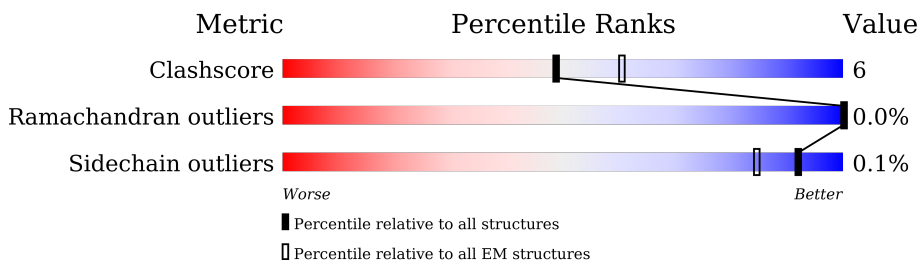
EMDB validation analysis : **FAILED**
MolProbity : 4.02b-467
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ : **FAILED**
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.39

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.70 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.






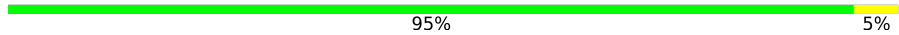
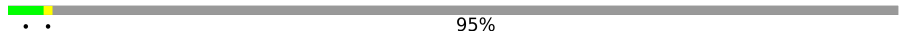
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	C	3859	
2	A	1217	
3	B	86	
4	D	779	
5	E	317	
6	G	374	
7	H	589	
8	I	455	
9	K	622	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	M	264	 37% 8% 55%
11	O	218	 32% 10% 58%
12	R	161	 34% 12% 54%
13	X	19	 95% 5%
14	L	892	 95%

2 Entry composition [i](#)

There are 14 unique types of molecules in this entry. The entry contains 52609 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 Transformation/transcription domain-associated protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	C	3261	26251	16890	4519	4654	188	0	0

- Molecule 2 is a protein called Splicing factor 3B subunit 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	A	1168	9158	5817	1558	1738	45	0	0

- Molecule 3 is a protein called Splicing factor 3B subunit 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	B	45	383	245	67	67	4	0	0

- Molecule 4 is a protein called Transcription factor SPT20 homolog.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	318	2595	1641	448	490	16	0	0

- Molecule 5 is a protein called Transcription initiation protein SPT3 homolog.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	E	225	1770	1116	324	318	12	0	0

- Molecule 6 is a protein called Transcriptional adapter 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	G	188	1505	951	271	277	6	0	0

There are 39 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-38	MET	-	initiating methionine	UNP Q96BN2
G	-37	ASP	-	expression tag	UNP Q96BN2
G	-36	TYR	-	expression tag	UNP Q96BN2
G	-35	LYS	-	expression tag	UNP Q96BN2
G	-34	ASP	-	expression tag	UNP Q96BN2
G	-33	HIS	-	expression tag	UNP Q96BN2
G	-32	ASP	-	expression tag	UNP Q96BN2
G	-31	GLY	-	expression tag	UNP Q96BN2
G	-30	ASP	-	expression tag	UNP Q96BN2
G	-29	TYR	-	expression tag	UNP Q96BN2
G	-28	LYS	-	expression tag	UNP Q96BN2
G	-27	ASP	-	expression tag	UNP Q96BN2
G	-26	HIS	-	expression tag	UNP Q96BN2
G	-25	ASP	-	expression tag	UNP Q96BN2
G	-24	ILE	-	expression tag	UNP Q96BN2
G	-23	ASP	-	expression tag	UNP Q96BN2
G	-22	TYR	-	expression tag	UNP Q96BN2
G	-21	LYS	-	expression tag	UNP Q96BN2
G	-20	ASP	-	expression tag	UNP Q96BN2
G	-19	ASP	-	expression tag	UNP Q96BN2
G	-18	ASP	-	expression tag	UNP Q96BN2
G	-17	ASP	-	expression tag	UNP Q96BN2
G	-16	LYS	-	expression tag	UNP Q96BN2
G	-15	GLY	-	expression tag	UNP Q96BN2
G	-14	GLY	-	expression tag	UNP Q96BN2
G	-13	SER	-	expression tag	UNP Q96BN2
G	-12	GLY	-	expression tag	UNP Q96BN2
G	-11	GLY	-	expression tag	UNP Q96BN2
G	-10	SER	-	expression tag	UNP Q96BN2
G	-9	LEU	-	expression tag	UNP Q96BN2
G	-8	GLU	-	expression tag	UNP Q96BN2
G	-7	VAL	-	expression tag	UNP Q96BN2
G	-6	LEU	-	expression tag	UNP Q96BN2
G	-5	PHE	-	expression tag	UNP Q96BN2
G	-4	GLN	-	expression tag	UNP Q96BN2
G	-3	GLY	-	expression tag	UNP Q96BN2
G	-2	PRO	-	expression tag	UNP Q96BN2
G	-1	LEU	-	expression tag	UNP Q96BN2
G	0	ASP	-	expression tag	UNP Q96BN2

- Molecule 7 is a protein called TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	H	514	4084	2580	706	778	20	0	0

- Molecule 8 is a protein called STAGA complex 65 subunit gamma.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
8	I	175	1419	901	253	260	5	0	0

There are 41 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	415	GLY	-	expression tag	UNP O94864
I	416	SER	-	expression tag	UNP O94864
I	417	GLU	-	expression tag	UNP O94864
I	418	ASN	-	expression tag	UNP O94864
I	419	LEU	-	expression tag	UNP O94864
I	420	TYR	-	expression tag	UNP O94864
I	421	PHE	-	expression tag	UNP O94864
I	422	GLN	-	expression tag	UNP O94864
I	423	GLY	-	expression tag	UNP O94864
I	424	SER	-	expression tag	UNP O94864
I	425	GLY	-	expression tag	UNP O94864
I	426	THR	-	expression tag	UNP O94864
I	427	SER	-	expression tag	UNP O94864
I	428	THR	-	expression tag	UNP O94864
I	429	ALA	-	expression tag	UNP O94864
I	430	TRP	-	expression tag	UNP O94864
I	431	SER	-	expression tag	UNP O94864
I	432	HIS	-	expression tag	UNP O94864
I	433	PRO	-	expression tag	UNP O94864
I	434	GLN	-	expression tag	UNP O94864
I	435	PHE	-	expression tag	UNP O94864
I	436	GLU	-	expression tag	UNP O94864
I	437	LYS	-	expression tag	UNP O94864
I	438	THR	-	expression tag	UNP O94864
I	439	GLY	-	expression tag	UNP O94864
I	440	VAL	-	expression tag	UNP O94864
I	441	SER	-	expression tag	UNP O94864
I	442	ILE	-	expression tag	UNP O94864
I	443	THR	-	expression tag	UNP O94864
I	444	SER	-	expression tag	UNP O94864
I	445	SER	-	expression tag	UNP O94864

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
I	446	GLY	-	expression tag	UNP O94864
I	447	SER	-	expression tag	UNP O94864
I	448	TRP	-	expression tag	UNP O94864
I	449	SER	-	expression tag	UNP O94864
I	450	HIS	-	expression tag	UNP O94864
I	451	PRO	-	expression tag	UNP O94864
I	452	GLN	-	expression tag	UNP O94864
I	453	PHE	-	expression tag	UNP O94864
I	454	GLU	-	expression tag	UNP O94864
I	455	LYS	-	expression tag	UNP O94864

- Molecule 9 is a protein called TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
9	K	352	2703	1720	478	493	12	0	0

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

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
10	M	119	957	607	168	176	6	0	0

- Molecule 11 is a protein called Transcription initiation factor TFIID subunit 10.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
11	O	91	722	466	116	136	4	0	0

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

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
12	R	74	605	379	105	118	3	0	0

- Molecule 13 is a protein called Unassigned sequence.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
13	X	19	95	57	19	19	0	0

- Molecule 14 is a protein called Ataxin-7.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
14	L	44	362	228	70	60	4	0	0

4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	378168	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	50	Depositor
Minimum defocus (nm)	800	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k 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 >5	RMSZ	# Z >5
1	C	0.24	0/26794	0.47	2/36236 (0.0%)
2	A	0.25	0/9345	0.49	1/12680 (0.0%)
3	B	0.23	0/393	0.42	0/528
4	D	0.23	0/2638	0.48	0/3559
5	E	0.23	0/1793	0.49	0/2408
6	G	0.26	0/1538	0.51	2/2089 (0.1%)
7	H	0.24	0/4167	0.49	0/5647
8	I	0.23	0/1449	0.45	0/1959
9	K	0.23	0/2757	0.50	2/3745 (0.1%)
10	M	0.24	0/979	0.50	0/1328
11	O	0.25	0/737	0.46	0/998
12	R	0.24	0/614	0.48	0/829
14	L	0.25	0/374	0.55	0/504
All	All	0.24	0/53578	0.48	7/72510 (0.0%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	G	273	PRO	N-CD-CG	-6.77	93.05	103.20
2	A	504	PRO	CA-N-CD	-6.12	102.94	111.50
9	K	420	PRO	N-CA-CB	5.92	110.41	103.30
6	G	273	PRO	CA-N-CD	-5.83	103.34	111.50
9	K	413	PRO	N-CA-CB	5.78	110.23	103.30

There are no chirality outliers.

There are no planarity outliers.

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	C	26251	0	26857	271	0
2	A	9158	0	9102	124	0
3	B	383	0	364	5	0
4	D	2595	0	2617	34	0
5	E	1770	0	1750	14	0
6	G	1505	0	1482	24	0
7	H	4084	0	4012	68	0
8	I	1419	0	1375	16	0
9	K	2703	0	2670	34	0
10	M	957	0	969	19	0
11	O	722	0	720	17	0
12	R	605	0	598	17	0
13	X	95	0	21	1	0
14	L	362	0	336	5	0
All	All	52609	0	52873	590	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 6.

The worst 5 of 590 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)
11:O:139:LEU:HB3	11:O:144:PHE:HB2	1.61	0.82
1:C:411:LEU:HD12	1:C:412:PRO:HD2	1.61	0.81
2:A:701:LEU:HD11	2:A:712:VAL:HB	1.62	0.81
1:C:3455:LEU:HD12	1:C:3456:PRO:HD2	1.63	0.80
2:A:548:ALA:HB3	2:A:555:VAL:HB	1.68	0.74

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	
1	C	3191/3859 (83%)	3008 (94%)	182 (6%)	1 (0%)	100	100
2	A	1156/1217 (95%)	1117 (97%)	39 (3%)	0	100	100
3	B	43/86 (50%)	43 (100%)	0	0	100	100
4	D	304/779 (39%)	282 (93%)	22 (7%)	0	100	100
5	E	215/317 (68%)	197 (92%)	18 (8%)	0	100	100
6	G	182/374 (49%)	176 (97%)	6 (3%)	0	100	100
7	H	500/589 (85%)	476 (95%)	24 (5%)	0	100	100
8	I	169/455 (37%)	166 (98%)	3 (2%)	0	100	100
9	K	342/622 (55%)	318 (93%)	24 (7%)	0	100	100
10	M	117/264 (44%)	111 (95%)	5 (4%)	1 (1%)	14	47
11	O	87/218 (40%)	83 (95%)	4 (5%)	0	100	100
12	R	72/161 (45%)	63 (88%)	9 (12%)	0	100	100
14	L	42/892 (5%)	35 (83%)	7 (17%)	0	100	100
All	All	6420/9833 (65%)	6075 (95%)	343 (5%)	2 (0%)	100	100

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
10	M	118	ASP
1	C	773	ASP

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	C	2918/3423 (85%)	2911 (100%)	7 (0%)	92	96
2	A	1013/1051 (96%)	1013 (100%)	0	100	100
3	B	41/77 (53%)	41 (100%)	0	100	100
4	D	297/687 (43%)	297 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
5	E	179/273 (66%)	178 (99%)	1 (1%)	84	90
6	G	164/320 (51%)	164 (100%)	0	100	100
7	H	457/521 (88%)	457 (100%)	0	100	100
8	I	152/405 (38%)	152 (100%)	0	100	100
9	K	286/505 (57%)	286 (100%)	0	100	100
10	M	106/235 (45%)	106 (100%)	0	100	100
11	O	78/154 (51%)	78 (100%)	0	100	100
12	R	68/141 (48%)	68 (100%)	0	100	100
14	L	38/779 (5%)	38 (100%)	0	100	100
All	All	5797/8571 (68%)	5789 (100%)	8 (0%)	92	96

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	E	296	ARG
1	C	3125	LYS
1	C	2248	LYS
1	C	2210	ARG
1	C	2357	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	604	GLN
6	G	220	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates [i](#)

There are no oligosaccharides 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.