



wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Sep 23, 2017 – 08:10 PM EDT

PDB ID : 5OF4
EMDB ID: : EMD-3802
Title : The cryo-EM structure of human TFIIH
Authors : Greber, B.J.; Nguyen, T.H.D.; Fang, J.; Afonine, P.V.; Adams, P.D.; Nogales, E.
Deposited on : unknown
Resolution : 4.40 Å(reported)

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

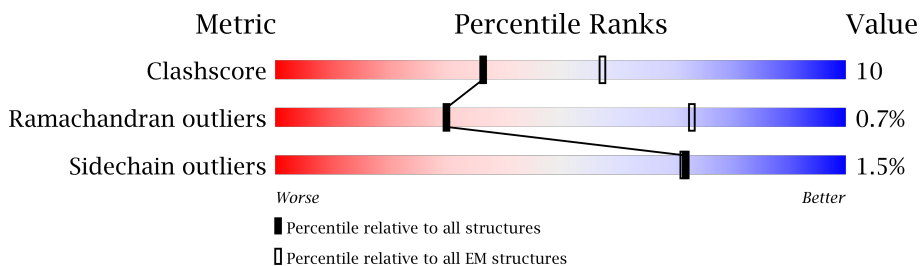
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 4.40 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.



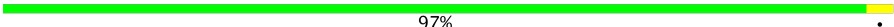
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	553	
2	B	760	
3	D	85	
4	E	395	
5	F	308	
6	G	71	
7	H	124	
8	Z	270	
9	Y	232	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	X	78	 97%

2 Entry composition [i](#)

There are 11 unique types of molecules in this entry. The entry contains 17200 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 TFIID basal transcription factor complex helicase XPB subunit,XPB,TFIID basal transcription factor complex helicase XPB subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	499	3887	2469	676	718	24	0	0

- Molecule 2 is a protein called TFIID basal transcription factor complex helicase XPD subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	724	5671	3626	992	1025	28	0	0

- Molecule 3 is a protein called General transcription factor IID subunit 4,p52,General transcription factor IID subunit 4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	D	81	643	412	118	112	1	0	0

- Molecule 4 is a protein called General transcription factor IID subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	E	184	1451	918	251	272	10	0	0

- Molecule 5 is a protein called General transcription factor IID subunit 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	F	205	1498	954	257	276	11	0	0

- Molecule 6 is a protein called General transcription factor IID subunit 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	G	66	522	336	83	100	3	0	0

- Molecule 7 is a protein called MAT1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	H	124	620	372	124	124		0	0

- Molecule 8 is a protein called Unassigned secondary structure elements..

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
8	Z	270	1350	810	270	270		0	0

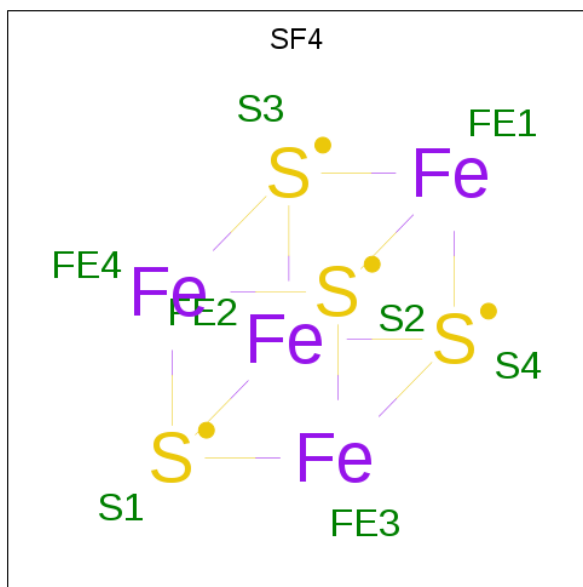
- Molecule 9 is a protein called Unassigned secondary structure elements (p52 region).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
9	Y	232	1160	696	232	232		0	0

- Molecule 10 is a protein called Unassigned secondary structure elements (XPB NTE region).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
10	X	78	390	234	78	78		0	0

- Molecule 11 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe₄S₄).

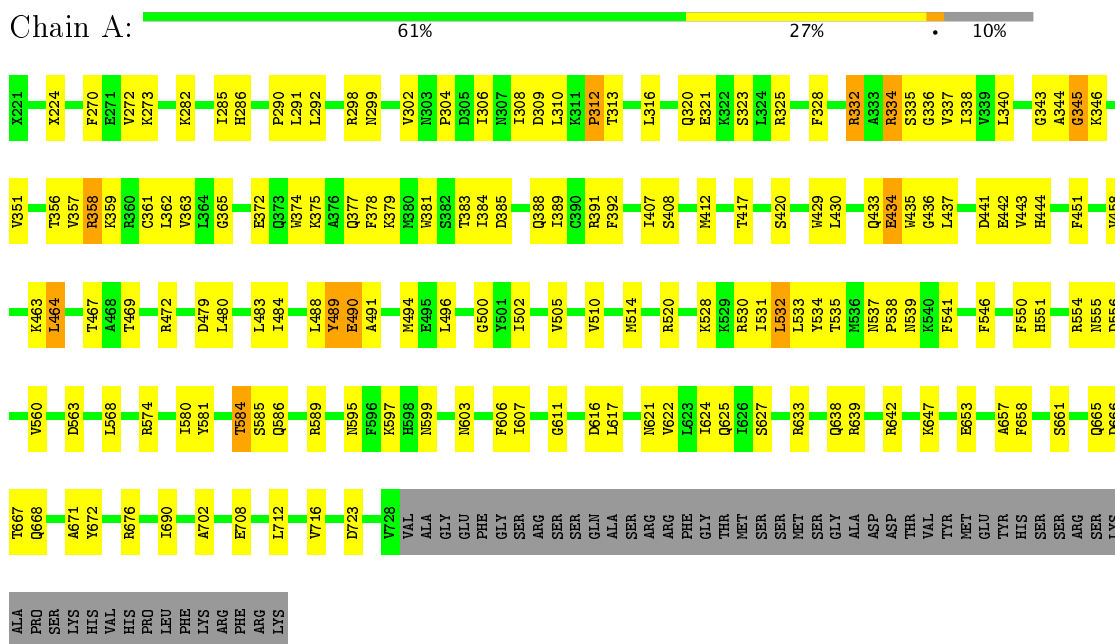


Mol	Chain	Residues	Atoms			AltConf
			Total	Fe	S	
11	B	1	8	4	4	0

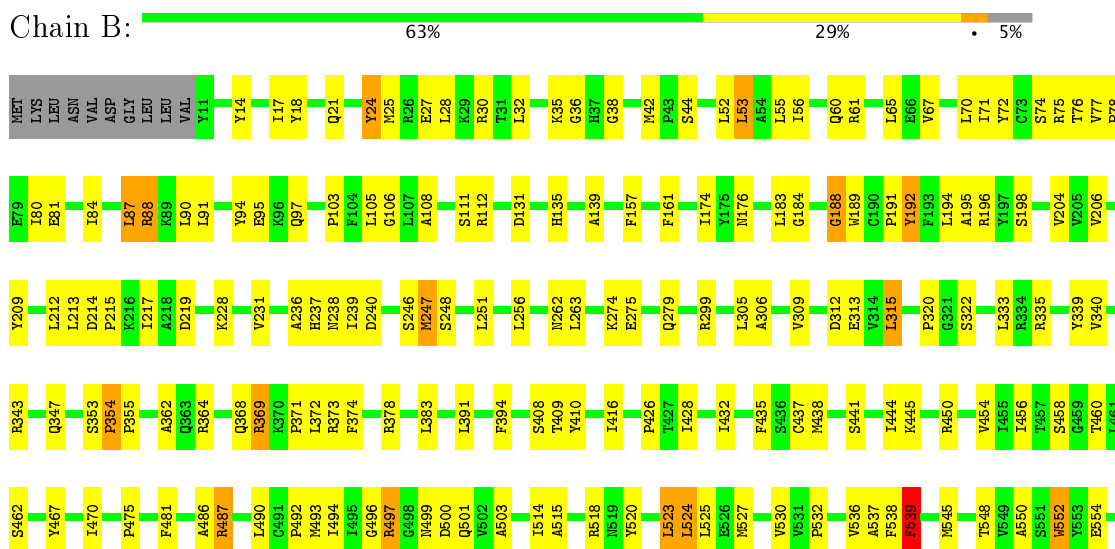
3 Residue-property plots

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: TFIIF basal transcription factor complex helicase XPB subunit,XPB,TFIIF basal transcription factor complex helicase XPB subunit

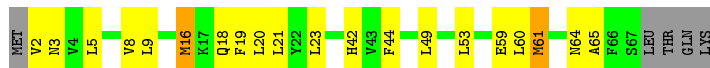


- Molecule 2: TFIIF basal transcription factor complex helicase XPD subunit

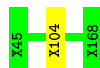


SER
ALA

- Molecule 6: General transcription factor IIH subunit 5

Chain G:  65% 25% 7%

- Molecule 7: MAT1

Chain H:  99%

- Molecule 8: Unassigned secondary structure elements.

Chain Z:  97%

- Molecule 9: Unassigned secondary structure elements (p52 region)

Chain Y:  93% 6%

- Molecule 10: Unassigned secondary structure elements (XPB NTE region)

Chain X:  97%

4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	122900	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION; Correction in RELION based on values determined in CTFFIND4.	Depositor
Microscope	FEI TITAN	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	40	Depositor
Minimum defocus (nm)	2000	Depositor
Maximum defocus (nm)	4500	Depositor
Magnification	37879	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality i

5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: SF4

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	1.00	7/3779 (0.2%)	1.07	17/5102 (0.3%)
2	B	1.08	11/5790 (0.2%)	1.22	44/7853 (0.6%)
3	D	0.92	1/631 (0.2%)	0.96	2/855 (0.2%)
4	E	1.13	3/1478 (0.2%)	1.24	10/2001 (0.5%)
5	F	1.02	1/1518 (0.1%)	1.11	12/2061 (0.6%)
6	G	0.72	0/528	0.86	2/713 (0.3%)
All	All	1.04	23/13724 (0.2%)	1.15	87/18585 (0.5%)

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	6
2	B	0	9
3	D	0	1
6	G	0	1
7	H	0	1
8	Z	0	2
9	Y	0	3
All	All	0	23

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	489	TYR	C-O	-9.39	1.05	1.23
3	D	394	TRP	CE3-CZ3	9.33	1.54	1.38
4	E	91	PHE	CG-CD1	7.93	1.50	1.38
2	B	591	GLY	C-O	7.24	1.35	1.23
2	B	539	PHE	N-CA	-7.20	1.31	1.46

The worst 5 of 87 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	E	207	VAL	CG1-CB-CG2	10.98	128.47	110.90
2	B	196	ARG	NE-CZ-NH1	10.96	125.78	120.30
1	A	639	ARG	NE-CZ-NH2	-9.64	115.48	120.30
2	B	684	PHE	CB-CG-CD1	9.34	127.34	120.80
1	A	332	ARG	NE-CZ-NH1	8.44	124.52	120.30

There are no chirality outliers.

5 of 23 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	272	VAL	Peptide
1	A	312	PRO	Peptide
1	A	328	PHE	Peptide
1	A	412	MET	Peptide
1	A	434	GLU	Peptide

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3887	0	3750	94	0
2	B	5671	0	5575	140	0
3	D	643	0	622	14	0
4	E	1451	0	1492	32	0
5	F	1498	0	1414	18	0
6	G	522	0	528	14	0
7	H	620	0	131	0	0
8	Z	1350	0	315	4	0
9	Y	1160	0	279	10	0
10	X	390	0	96	1	0
11	B	8	0	0	0	0
All	All	17200	0	14202	301	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 10.

The worst 5 of 301 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)
2:B:106:GLY:HA2	2:B:204:VAL:O	1.71	0.91
1:A:671:ALA:HB2	6:G:65:ALA:HB1	1.56	0.88
1:A:224:UNK:O	1:A:292:LEU:HA	1.74	0.87
1:A:429:TRP:O	1:A:433:GLN:HB2	1.75	0.85
3:D:408:LEU:HB3	6:G:5:LEU:HD12	1.59	0.84

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	A	461/553 (83%)	402 (87%)	57 (12%)	2 (0%)	38	77
2	B	722/760 (95%)	614 (85%)	101 (14%)	7 (1%)	18	61
3	D	75/85 (88%)	71 (95%)	3 (4%)	1 (1%)	14	56
4	E	182/395 (46%)	171 (94%)	11 (6%)	0	100	100
5	F	199/308 (65%)	182 (92%)	15 (8%)	2 (1%)	18	61
6	G	64/71 (90%)	56 (88%)	8 (12%)	0	100	100
All	All	1703/2172 (78%)	1496 (88%)	195 (12%)	12 (1%)	30	68

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	313	THR
2	B	428	ILE
2	B	592	ARG
2	B	14	TYR
2	B	685	ALA

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	400/454 (88%)	395 (99%)	5 (1%)	73	87
2	B	586/664 (88%)	576 (98%)	10 (2%)	66	85
3	D	64/73 (88%)	64 (100%)	0	100	100
4	E	169/352 (48%)	165 (98%)	4 (2%)	54	78
5	F	144/272 (53%)	142 (99%)	2 (1%)	71	86
6	G	59/64 (92%)	59 (100%)	0	100	100
All	All	1422/1879 (76%)	1401 (98%)	21 (2%)	72	86

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	487	ARG
2	B	539	PHE
4	E	174	LEU
2	B	437	CYS
4	E	194	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	638	GLN
1	A	665	GLN
4	E	60	HIS
1	A	621	ASN
4	E	235	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
11	SF4	B	1000	2	0,12,12	0.00	-	0,24,24	0.00	-

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
11	SF4	B	1000	2	-	0/0/48/48	0/6/5/5

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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
9	Y	18
8	Z	15
10	X	6
1	A	1

The worst 5 of 40 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	Z	181:UNK	C	201:UNK	N	103.83
1	Z	165:UNK	C	171:UNK	N	91.40
1	Z	20:UNK	C	82:UNK	N	60.92
1	Y	320:UNK	C	373:UNK	N	57.49
1	Z	487:UNK	C	501:UNK	N	56.59