

wwPDB EM Validation Summary Report (i)

Dec 16, 2024 – 07:17 PM EST

PDB ID	:	7ML3
EMDB ID	:	EMD-23907
Title	:	General transcription factor TFIIH (weak binding)
Authors	:	Yang, C.; Fujiwara, R.; Kim, H.J.; Gorbea Colon, J.J.; Steimle, S.; Garcia,
		B.A.; Murakami, K.
Deposited on	:	2021-04-27
Resolution	:	7.60 Å(reported)
Based on initial model	:	5OQJ

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 (i) 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 (1)) were used in the production of this report:

EMDB validation analysis	:	0.0.1.dev113
Mogul	:	2022.3.0, CSD as543be (2022)
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 7.60 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)	$\mathop{{\rm EM}}\limits_{{\rm (\#Entries)}}$		
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 <=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		Qualit	y of chain	
1	3	321	23%	16% •	57%	6
2	2	513	13%		44%	•• 10%
3	0	778	23%		69%	5% •
4	1	537	15%		26% •	32%
5	4	338	8%		49%	• 16%
6	6	461	- 30%		42%	• 24%
7	5	72	22% 19%		64%	8% 8%
8	7	843	23%	38%	•	35%



Mol	Chain	Length		Quality of chain
9	Ν	30	10% 13%	87%
10	Т	30	20%	13% 73%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
12	SF4	0	801	-	-	Х	-



2 Entry composition (i)

There are 12 unique types of molecules in this entry. The entry contains 22648 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 BJ4_G0050160.mRNA.1.CDS.1.

Mol	Chain	Residues		At	oms	AltConf	Trace		
1	3	138	Total 860	C 533	N 160	O 160	${ m S} 7$	0	0

• Molecule 2 is a protein called RNA polymerase II transcription factor B subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	2	460	Total 3011	C 1856	N 562	0 584	S 9	0	0

• Molecule 3 is a protein called General transcription and DNA repair factor IIH helicase subunit XPD.

Mol	Chain	Residues		Α	AltConf	Trace			
3	0	754	Total 6108	C 3891	N 1032	0 1147	S 38	0	0

• Molecule 4 is a protein called Tfb1.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	1	367	Total 2411	C 1536	N 438	0 430	${f S}7$	0	0

• Molecule 5 is a protein called General transcription and DNA repair factor IIH subunit TFB4.

Mol	Chain	Residues		At	AltConf	Trace			
5	4	284	Total 2041	C 1310	N 343	O 376	S 12	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
4	113	UNK	ASP	conflict	UNP A0A7I9C5C2
4	114	UNK	MET	conflict	UNP A0A7I9C5C2



• Molecule 6 is a protein called General transcription and DNA repair factor IIH.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	6	351	Total 2527	C 1590	N 454	0 456	S 27	0	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
6	412	UNK	ILE	conflict	UNP A0A7I9FQL5
6	413	UNK	LEU	conflict	UNP A0A7I9FQL5
6	414	UNK	LYS	conflict	UNP A0A7I9FQL5
6	415	UNK	ASN	conflict	UNP A0A7I9FQL5
6	416	UNK	HIS	conflict	UNP A0A7I9FQL5
6	417	UNK	LYS	conflict	UNP A0A7I9FQL5
6	418	UNK	ASN	conflict	UNP A0A7I9FQL5
6	419	UNK	ASP	conflict	UNP A0A7I9FQL5
6	420	UNK	LYS	conflict	UNP A0A7I9FQL5
6	421	UNK	LEU	conflict	UNP A0A7I9FQL5
6	422	UNK	LEU	conflict	UNP A0A7I9FQL5
6	423	UNK	THR	conflict	UNP A0A7I9FQL5
6	424	UNK	SER	conflict	UNP A0A7I9FQL5

• Molecule 7 is a protein called General transcription and DNA repair factor IIH subunit TFB5.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	5	66	Total	С	Ν	Ο	\mathbf{S}	0	0
1	5	00	498	314	89	93	2	0	0

• Molecule 8 is a protein called General transcription and DNA repair factor IIH helicase subunit XPB.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	7	548	Total 3947	C 2438	N 729	O 757	S 23	0	0

• Molecule 9 is a DNA chain called non-template strand DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	N	30	Total 624	C 298	N 113	0 183	Р 30	0	0

• Molecule 10 is a DNA chain called template strand DNA.



Mol	Chain	Residues	Atoms					AltConf	Trace
10	Т	30	Total 606	C 291	N 108	O 177	Р 30	0	0

• Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	AltConf
11	3	2	Total Zn 2 2	0
11	4	1	Total Zn 1 1	0
11	6	4	Total Zn 4 4	0

• Molecule 12 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).



Mol	Chain	Residues	Atoms	AltConf
12	0	1	Total Fe S 8 4 4	0



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: BJ4_G0050160.mRNA.1.CDS.1























4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	101497	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE	Depositor
	CORRECTION	
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose $(e^-/\text{\AA}^2)$	45	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor
Maximum map value	0.072	Depositor
Minimum map value	0.000	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.018	Depositor
Map size (Å)	273.47998, 273.47998, 272.41998	wwPDB
Map dimensions	258, 258, 257	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.06, 1.06, 1.06	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: ZN, 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).

Mal	Chain	B	ond lengths	E	Bond angles
INIOI	Unam	RMSZ	# Z > 5	RMSZ	# Z > 5
1	3	1.27	6/870~(0.7%)	0.84	7/1190~(0.6%)
2	2	1.03	19/3057~(0.6%)	1.20	16/4071~(0.4%)
3	0	0.81	27/6226~(0.4%)	0.79	23/8407~(0.3%)
4	1	0.62	7/1896~(0.4%)	0.60	2/2543~(0.1%)
5	4	0.84	8/2062~(0.4%)	0.77	8/2805~(0.3%)
6	6	1.04	14/2506~(0.6%)	0.84	6/3402~(0.2%)
7	5	0.73	3/502~(0.6%)	1.00	5/677~(0.7%)
8	7	1.06	29/4017~(0.7%)	1.13	32/5374~(0.6%)
9	Ν	2.51	44/700~(6.3%)	2.68	55/1081~(5.1%)
10	Т	2.75	36/678~(5.3%)	2.77	48/1041~(4.6%)
All	All	1.12	193/22514~(0.9%)	1.15	202/30591~(0.7%)

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	3	0	2
2	2	0	4
3	0	0	2
5	4	0	2
6	6	0	2
8	7	0	8
All	All	0	20

The worst 5 of 193 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	3	34	CYS	CB-SG	-30.99	1.29	1.82
6	6	366	CYS	CB-SG	-24.74	1.40	1.82



Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	2	504	PHE	CE1-CZ	-22.91	0.93	1.37
6	6	426	ARG	CZ-NH2	21.06	1.60	1.33
8	7	572	GLU	CG-CD	19.68	1.81	1.51

The worst 5 of 202 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
2	2	286	ARG	NE-CZ-NH2	-49.68	95.46	120.30
8	7	757	ARG	NE-CZ-NH1	27.93	134.26	120.30
10	Т	102	DT	O4'-C1'-N1	26.90	126.83	108.00
10	Т	92	DA	O4'-C1'-N9	26.29	126.41	108.00
10	Т	106	DA	O4'-C1'-N9	26.22	126.35	108.00

There are no chirality outliers.

5 of 20 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	2	286	ARG	Sidechain
2	2	468	TYR	Sidechain
2	2	68	SER	Peptide
1	3	71	GLN	Peptide
1	3	75	ASP	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	3	860	0	621	127	0
2	2	3011	0	2600	427	0
3	0	6108	0	6165	1150	0
4	1	2411	0	1882	220	0
5	4	2041	0	1954	302	0
6	6	2527	0	2321	404	0
7	5	498	0	506	129	0
8	7	3947	0	3575	814	0
9	N	624	0	342	87	0
10	Т	606	0	339	44	0



Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
11	3	2	0	0	0	0
11	4	1	0	0	0	0
11	6	4	0	0	0	0
12	0	8	0	0	4	0
All	All	22648	0	20305	3535	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 82.

The worst 5 of 3535 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)
2:2:501:VAL:HA	2:2:504:PHE:CE2	1.33	1.63
3:0:614:HIS:CE1	3:0:675:ASP:HA	1.28	1.61
2:2:380:ARG:HH11	2:2:444:TRP:CB	1.16	1.57
8:7:757:ARG:CD	8:7:757:ARG:NE	1.68	1.51
8:7:572:GLU:CD	8:7:572:GLU:CG	1.81	1.47

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	3	136/321~(42%)	110 (81%)	24 (18%)	2(2%)	8 40
2	2	456/513~(89%)	378~(83%)	78 (17%)	0	100 100
3	0	752/778~(97%)	657~(87%)	94 (12%)	1 (0%)	48 83
4	1	256/537~(48%)	223~(87%)	30 (12%)	3 (1%)	11 44
5	4	279/338~(82%)	218 (78%)	61~(22%)	0	100 100
6	6	336/461~(73%)	278 (83%)	55 (16%)	3 (1%)	14 52



	J	1 1 3						
Mol	Chain	Analysed	Favoured	Allowed	Outliers	Perce	entiles	
7	5	64/72~(89%)	50~(78%)	14(22%)	0	100	100	
8	7	544/843~(64%)	469~(86%)	75 (14%)	0	100	100	
All	All	2823/3863~(73%)	2383~(84%)	431 (15%)	9~(0%)	38	73	

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	1	230	PRO
6	6	116	THR
6	6	411	PRO
4	1	389	LEU
4	1	239	PRO

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent side chain 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	3	53/303~(18%)	52 (98%)	1 (2%)	52 69
2	2	258/468~(55%)	253~(98%)	5(2%)	52 69
3	0	686/707~(97%)	685 (100%)	1 (0%)	92 95
4	1	169/395~(43%)	167~(99%)	2(1%)	67 78
5	4	198/298~(66%)	198 (100%)	0	100 100
6	6	247/406~(61%)	241 (98%)	6(2%)	44 62
7	5	53/66~(80%)	53~(100%)	0	100 100
8	7	379/737~(51%)	371 (98%)	8 (2%)	48 66
All	All	2043/3380~(60%)	2020 (99%)	23 (1%)	69 80

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	6	366	CYS
8	7	541	MET



Continued from previous page...

Mol	Chain	Res	Type
8	7	534	LYS
8	7	546	LYS
3	0	649	ARG

Sometimes side chains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such side chains are listed below:

Mol	Chain	Res	Type
8	7	442	ASN
8	7	644	GLN
8	7	447	GLN
8	7	528	ASN
5	4	64	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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 7 are monoatomic - leaving 1 for Mogul analysis.

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	
12	SF4	0	801	3	0,12,12	-	-	-	



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
12	SF4	0	801	3	-	-	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.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
12	0	801	SF4	4	0

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
4	1	4

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	1	393:UNK	С	465:UNK	Ν	85.89
1	1	519:UNK	С	537:GLU	Ν	12.88
1	1	355:UNK	С	368:UNK	Ν	12.28
1	1	567:HIS	С	573:GLN	Ν	4.45



6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-23907. 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 (i)

6.1.1 Primary map



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

6.2 Central slices (i)

6.2.1 Primary map



X Index: 129



Y Index: 129



Z Index: 128

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: 147

Y Index: 129

Z Index: 133

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

6.4 Orthogonal standard-deviation projections (False-color) (i)

6.4.1 Primary map



The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.



6.5 Orthogonal surface views (i)

6.5.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.018. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.6 Mask visualisation (i)

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



7 Map analysis (i)

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

7.1 Map-value distribution (i)



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 279 nm^3 ; this corresponds to an approximate mass of 252 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)

This section was not generated. The rotationally averaged power spectrum is only generated for cubic maps.



8 Fourier-Shell correlation (i)

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



9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-23907 and PDB model 7ML3. Per-residue inclusion information can be found in section 3 on page 7.

9.1 Map-model overlay (i)



The images above show the 3D surface view of the map at the recommended contour level 0.018 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)



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)



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 (0.018).



9.4 Atom inclusion (i)



At the recommended contour level, 84% of all backbone atoms, 73% 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 (0.018) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score	
All	0.7300	0.2150	1.0
0	0.7490	0.2480	
1	0.6630	0.2230	
2	0.7400	0.1980	
3	0.6570	0.2150	
4	0.7360	0.2550	
5	0.6950	0.1500	
6	0.8230	0.2810	
7	0.7280	0.1390	0.0
N	0.6550	0.1560	<0.0
Т	0.5880	0.1270	

