



wwPDB EM Validation Summary Report ⓘ

May 21, 2024 – 10:09 AM JST

PDB ID : 8HMY
EMDB ID : EMD-34904
Title : Cryo-EM structure of the human pre-catalytic TSEN/pre-tRNA complex
Authors : Zhang, X.; Yang, F.; Zhan, X.; Shi, Y.
Deposited on : 2022-12-06
Resolution : 2.94 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 : 20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ : **FAILED**
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.36.2

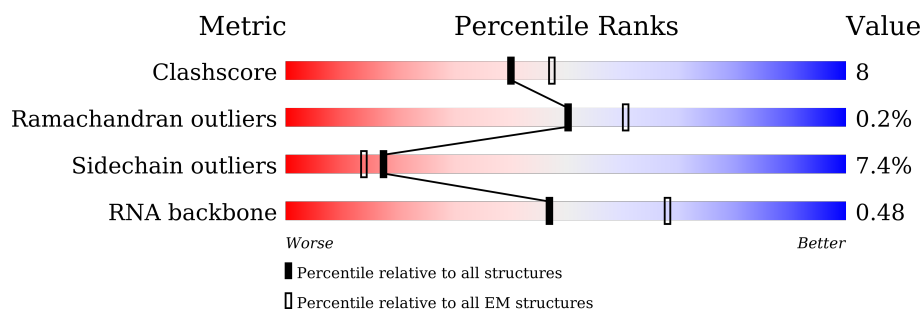
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 2.94 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	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826
RNA backbone	4643	859

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	A	485	43% 20% . 35%
2	B	330	61% 18% . 20%
3	T	114	45% 28% 6% 21%
4	C	546	39% 12% . 48%
5	D	213	40% 18% . 41%
6	E	445	85% 15%

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 11675 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 tRNA-splicing endonuclease subunit Sen2.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	314	Total	C	N	O	S	0	0
			2608	1683	445	467	13		

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	initiating methionine	UNP Q8NCE0
A	-18	ALA	-	expression tag	UNP Q8NCE0
A	-17	SER	-	expression tag	UNP Q8NCE0
A	-16	ASP	-	expression tag	UNP Q8NCE0
A	-15	TYR	-	expression tag	UNP Q8NCE0
A	-14	LYS	-	expression tag	UNP Q8NCE0
A	-13	ASP	-	expression tag	UNP Q8NCE0
A	-12	ASP	-	expression tag	UNP Q8NCE0
A	-11	ASP	-	expression tag	UNP Q8NCE0
A	-10	ASP	-	expression tag	UNP Q8NCE0
A	-9	LYS	-	expression tag	UNP Q8NCE0
A	-8	ALA	-	expression tag	UNP Q8NCE0
A	-7	SER	-	expression tag	UNP Q8NCE0
A	-6	ASP	-	expression tag	UNP Q8NCE0
A	-5	GLU	-	expression tag	UNP Q8NCE0
A	-4	VAL	-	expression tag	UNP Q8NCE0
A	-3	ASP	-	expression tag	UNP Q8NCE0
A	-2	ALA	-	expression tag	UNP Q8NCE0
A	-1	GLY	-	expression tag	UNP Q8NCE0
A	0	THR	-	expression tag	UNP Q8NCE0
A	377	ALA	HIS	engineered mutation	UNP Q8NCE0

- Molecule 2 is a protein called tRNA-splicing endonuclease subunit Sen34.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	264	Total	C	N	O	S	0	0
			2057	1298	386	369	4		

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-19	MET	-	initiating methionine	UNP Q9BSV6
B	-18	ALA	-	expression tag	UNP Q9BSV6
B	-17	SER	-	expression tag	UNP Q9BSV6
B	-16	ASP	-	expression tag	UNP Q9BSV6
B	-15	TYR	-	expression tag	UNP Q9BSV6
B	-14	LYS	-	expression tag	UNP Q9BSV6
B	-13	ASP	-	expression tag	UNP Q9BSV6
B	-12	ASP	-	expression tag	UNP Q9BSV6
B	-11	ASP	-	expression tag	UNP Q9BSV6
B	-10	ASP	-	expression tag	UNP Q9BSV6
B	-9	LYS	-	expression tag	UNP Q9BSV6
B	-8	ALA	-	expression tag	UNP Q9BSV6
B	-7	SER	-	expression tag	UNP Q9BSV6
B	-6	ASP	-	expression tag	UNP Q9BSV6
B	-5	GLU	-	expression tag	UNP Q9BSV6
B	-4	VAL	-	expression tag	UNP Q9BSV6
B	-3	ASP	-	expression tag	UNP Q9BSV6
B	-2	ALA	-	expression tag	UNP Q9BSV6
B	-1	GLY	-	expression tag	UNP Q9BSV6
B	0	THR	-	expression tag	UNP Q9BSV6
B	255	ALA	HIS	engineered mutation	UNP Q9BSV6

- Molecule 3 is a RNA chain called Pre-tRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	T	90	Total	C	N	O	P	0	0
			1888	838	336	624	90		

- Molecule 4 is a protein called tRNA-splicing endonuclease subunit Sen54.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	C	285	Total	C	N	O	S	0	0
			2263	1440	405	412	6		

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-19	MET	-	initiating methionine	UNP Q7Z6J9
C	-18	ALA	-	expression tag	UNP Q7Z6J9
C	-17	SER	-	expression tag	UNP Q7Z6J9
C	-16	ASP	-	expression tag	UNP Q7Z6J9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-15	TYR	-	expression tag	UNP Q7Z6J9
C	-14	LYS	-	expression tag	UNP Q7Z6J9
C	-13	ASP	-	expression tag	UNP Q7Z6J9
C	-12	ASP	-	expression tag	UNP Q7Z6J9
C	-11	ASP	-	expression tag	UNP Q7Z6J9
C	-10	ASP	-	expression tag	UNP Q7Z6J9
C	-9	LYS	-	expression tag	UNP Q7Z6J9
C	-8	ALA	-	expression tag	UNP Q7Z6J9
C	-7	SER	-	expression tag	UNP Q7Z6J9
C	-6	ASP	-	expression tag	UNP Q7Z6J9
C	-5	GLU	-	expression tag	UNP Q7Z6J9
C	-4	VAL	-	expression tag	UNP Q7Z6J9
C	-3	ASP	-	expression tag	UNP Q7Z6J9
C	-2	ALA	-	expression tag	UNP Q7Z6J9
C	-1	GLY	-	expression tag	UNP Q7Z6J9
C	0	THR	-	expression tag	UNP Q7Z6J9

- Molecule 5 is a protein called Chromosome 1 open reading frame 19, isoform CRA_a.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	D	126	Total	C	N	O	S	0	0
			984	631	154	191	8		

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-37	MET	-	initiating methionine	UNP A0A2U3TZM3
D	-36	ALA	-	expression tag	UNP A0A2U3TZM3
D	-35	SER	-	expression tag	UNP A0A2U3TZM3
D	-34	SER	-	expression tag	UNP A0A2U3TZM3
D	-33	ALA	-	expression tag	UNP A0A2U3TZM3
D	-32	TRP	-	expression tag	UNP A0A2U3TZM3
D	-31	SER	-	expression tag	UNP A0A2U3TZM3
D	-30	HIS	-	expression tag	UNP A0A2U3TZM3
D	-29	PRO	-	expression tag	UNP A0A2U3TZM3
D	-28	GLN	-	expression tag	UNP A0A2U3TZM3
D	-27	PHE	-	expression tag	UNP A0A2U3TZM3
D	-26	GLU	-	expression tag	UNP A0A2U3TZM3
D	-25	LYS	-	expression tag	UNP A0A2U3TZM3
D	-24	GLY	-	expression tag	UNP A0A2U3TZM3
D	-23	GLY	-	expression tag	UNP A0A2U3TZM3
D	-22	GLY	-	expression tag	UNP A0A2U3TZM3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-21	SER	-	expression tag	UNP A0A2U3TZM3
D	-20	GLY	-	expression tag	UNP A0A2U3TZM3
D	-19	GLY	-	expression tag	UNP A0A2U3TZM3
D	-18	GLY	-	expression tag	UNP A0A2U3TZM3
D	-17	SER	-	expression tag	UNP A0A2U3TZM3
D	-16	GLY	-	expression tag	UNP A0A2U3TZM3
D	-15	GLY	-	expression tag	UNP A0A2U3TZM3
D	-14	SER	-	expression tag	UNP A0A2U3TZM3
D	-13	ALA	-	expression tag	UNP A0A2U3TZM3
D	-12	TRP	-	expression tag	UNP A0A2U3TZM3
D	-11	SER	-	expression tag	UNP A0A2U3TZM3
D	-10	HIS	-	expression tag	UNP A0A2U3TZM3
D	-9	PRO	-	expression tag	UNP A0A2U3TZM3
D	-8	GLN	-	expression tag	UNP A0A2U3TZM3
D	-7	PHE	-	expression tag	UNP A0A2U3TZM3
D	-6	GLU	-	expression tag	UNP A0A2U3TZM3
D	-5	LYS	-	expression tag	UNP A0A2U3TZM3
D	-4	GLY	-	expression tag	UNP A0A2U3TZM3
D	-3	SER	-	expression tag	UNP A0A2U3TZM3
D	-2	ALA	-	expression tag	UNP A0A2U3TZM3
D	-1	ALA	-	expression tag	UNP A0A2U3TZM3
D	0	ALA	-	expression tag	UNP A0A2U3TZM3

- Molecule 6 is a protein called Polyribonucleotide 5'-hydroxyl-kinase Clp1.

Mol	Chain	Residues	Atoms				AltConf	Trace
6	E	379	Total	C	N	O	0	0
			1869	1111	379	379		

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	-19	MET	-	initiating methionine	UNP Q92989
E	-18	ALA	-	expression tag	UNP Q92989
E	-17	SER	-	expression tag	UNP Q92989
E	-16	ASP	-	expression tag	UNP Q92989
E	-15	TYR	-	expression tag	UNP Q92989
E	-14	LYS	-	expression tag	UNP Q92989
E	-13	ASP	-	expression tag	UNP Q92989
E	-12	ASP	-	expression tag	UNP Q92989
E	-11	ASP	-	expression tag	UNP Q92989
E	-10	ASP	-	expression tag	UNP Q92989

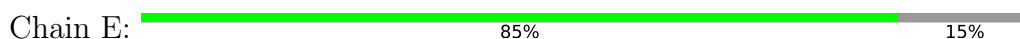
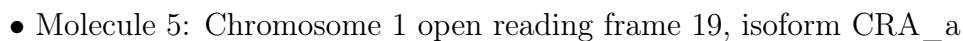
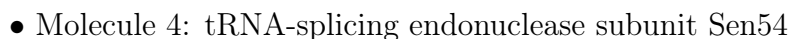
Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	-9	LYS	-	expression tag	UNP Q92989
E	-8	ALA	-	expression tag	UNP Q92989
E	-7	SER	-	expression tag	UNP Q92989
E	-6	ASP	-	expression tag	UNP Q92989
E	-5	GLU	-	expression tag	UNP Q92989
E	-4	VAL	-	expression tag	UNP Q92989
E	-3	ASP	-	expression tag	UNP Q92989
E	-2	ALA	-	expression tag	UNP Q92989
E	-1	GLY	-	expression tag	UNP Q92989
E	0	THR	-	expression tag	UNP Q92989

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	AltConf
7	T	6	Total Mg 6 6	0



MET	ALA	SER	ASP	TYR	LYS	ASP	ASP	ASP	ASP	LYS	ALA	SER	ASP	VAL	ALA	GLY	THR	MET	GLY	GLU	GLU	ALA	ASN	ASP	ASP	LYS	LYS	P11	V329	GLY	ALA	PRO	THR	ILE	PRO	ASP	SER	CYS	LEU	PRO	LEU	GLY	MET	SER	GLN	GLU	ASP	ASN	GLN	LEU	LYS	L352	L370	ALA	GLU
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	-----	-----

GLY	THR	GLU	GLU	ASN	LEU	SER	GLU	THR	S382	M422	ASP	LEU	LYS
-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	-----	-----	-----

4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	199001	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)	1500	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

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

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	0.29	0/2670	0.50	0/3602
2	B	0.29	0/2102	0.45	0/2848
3	T	0.39	0/2110	0.78	0/3290
4	C	0.33	0/2319	0.49	2/3140 (0.1%)
5	D	0.29	0/1004	0.46	0/1366
6	E	0.34	0/1866	0.67	0/2592
All	All	0.33	0/12071	0.58	2/16838 (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
4	C	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	C	411	PRO	CA-N-CD	-8.51	99.59	111.50
4	C	524	VAL	C-N-CA	-5.06	111.67	122.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	C	521	PRO	Peptide

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	A	2608	0	2637	57	0
2	B	2057	0	2089	40	0
3	T	1888	0	951	20	0
4	C	2263	0	2234	49	0
5	D	984	0	987	24	0
6	E	1869	0	827	0	0
7	T	6	0	0	0	0
All	All	11675	0	9725	162	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 8.

The worst 5 of 162 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)
3:T:19:A:N6	3:T:71:C:H42	1.43	1.14
3:T:19:A:H61	3:T:71:C:N4	1.50	1.07
3:T:19:A:H61	3:T:71:C:H42	0.88	0.86
3:T:19:A:N1	3:T:71:C:N3	2.30	0.80
2:B:248:PRO:HB2	4:C:45:PRO:HG3	1.69	0.73

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	A	308/485 (64%)	280 (91%)	28 (9%)	0	100	100
2	B	260/330 (79%)	248 (95%)	12 (5%)	0	100	100
4	C	281/546 (52%)	260 (92%)	20 (7%)	1 (0%)	34	64
5	D	124/213 (58%)	112 (90%)	10 (8%)	2 (2%)	9	30
6	E	373/445 (84%)	369 (99%)	4 (1%)	0	100	100
All	All	1346/2019 (67%)	1269 (94%)	74 (6%)	3 (0%)	50	76

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	C	522	GLN
5	D	162	PRO
5	D	164	PRO

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	A	288/427 (67%)	257 (89%)	31 (11%)	6	19
2	B	211/260 (81%)	196 (93%)	15 (7%)	14	38
4	C	242/464 (52%)	230 (95%)	12 (5%)	24	54
5	D	113/171 (66%)	108 (96%)	5 (4%)	28	59
All	All	854/1322 (65%)	791 (93%)	63 (7%)	17	36

5 of 63 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	435	GLU
4	C	516	ARG
2	B	77	ARG
4	C	486	CYS
5	D	80	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
4	C	69	GLN
5	D	117	ASN
1	A	446	GLN
2	B	89	GLN
2	B	110	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	T	89/114 (78%)	26 (29%)	0

5 of 26 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	T	13	C
3	T	15	A
3	T	16	U
3	T	19	A
3	T	20	U

There are no RNA pucker outliers to report.

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 ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.