WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB EM Validation Summary Report (i)

Oct 26, 2025 – 02:21 PM JST

PDB ID	:	8XYL / pdb_00008xyl
EMDB ID	:	EMD-38777
Title	:	Cryo-EM structure of Tetrahymena DNA methyltransferase complex MTA1c
Authors	:	Xu, Q.; Shi, Z.B.
Deposited on	:	2024-01-19
Resolution	:	2.79 Å(reported)
Based on initial model	:	.

The following versions of software and data (see references (i)) were used in the production of this report:

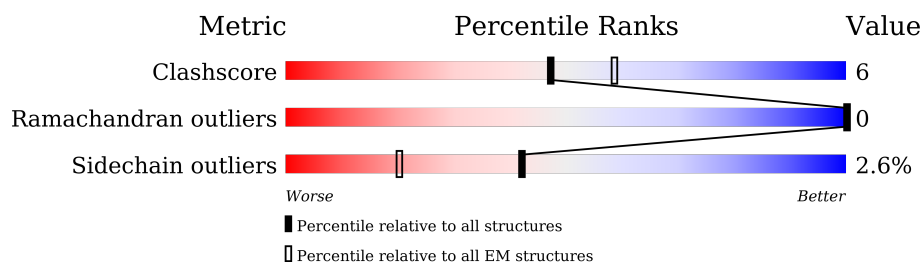|		
EMDB validation analysis	:	FAILED
MolProbity	:	4-5-2 with Phenix2.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
EM percentile statistics	:	NOT EXECUTED
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.46

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 2.79 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	Percentile Ranks	Value
Clashscore		6
Ramachandran outliers		0
Sidechain outliers		2.6%

Worse   *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all EM structures

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	378	48% 11% 41%
2	B	357	61% 13% 27%
3	C	364	7% • 91%
4	D	146	73% 14% 12%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 5312 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 MT-a70 family protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	224	Total 1826	C 1164	N 320	O 332	S 10	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	GLY	-	expression tag	UNP Q22GC0
A	-4	PRO	-	expression tag	UNP Q22GC0
A	-3	GLU	-	expression tag	UNP Q22GC0
A	-2	PHE	-	expression tag	UNP Q22GC0
A	-1	LYS	-	expression tag	UNP Q22GC0
A	0	LEU	-	expression tag	UNP Q22GC0

- Molecule 2 is a protein called Methyltransferase MT, putative.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	262	Total 2166	C 1387	N 356	O 414	S 9	0	0

There are 33 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	GLY	-	expression tag	UNP Q22XT1
B	-3	PRO	-	expression tag	UNP Q22XT1
B	-2	GLY	-	expression tag	UNP Q22XT1
B	-1	ARG	-	expression tag	UNP Q22XT1
B	0	PRO	-	expression tag	UNP Q22XT1
B	325	TRP	-	expression tag	UNP Q22XT1
B	326	SER	-	expression tag	UNP Q22XT1
B	327	HIS	-	expression tag	UNP Q22XT1
B	328	PRO	-	expression tag	UNP Q22XT1
B	329	GLN	-	expression tag	UNP Q22XT1
B	330	PHE	-	expression tag	UNP Q22XT1

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	331	GLU	-	expression tag	UNP Q22XT1
B	332	LYS	-	expression tag	UNP Q22XT1
B	333	GLY	-	expression tag	UNP Q22XT1
B	334	GLY	-	expression tag	UNP Q22XT1
B	335	GLY	-	expression tag	UNP Q22XT1
B	336	SER	-	expression tag	UNP Q22XT1
B	337	GLY	-	expression tag	UNP Q22XT1
B	338	GLY	-	expression tag	UNP Q22XT1
B	339	GLY	-	expression tag	UNP Q22XT1
B	340	SER	-	expression tag	UNP Q22XT1
B	341	GLY	-	expression tag	UNP Q22XT1
B	342	GLY	-	expression tag	UNP Q22XT1
B	343	GLY	-	expression tag	UNP Q22XT1
B	344	SER	-	expression tag	UNP Q22XT1
B	345	TRP	-	expression tag	UNP Q22XT1
B	346	SER	-	expression tag	UNP Q22XT1
B	347	HIS	-	expression tag	UNP Q22XT1
B	348	PRO	-	expression tag	UNP Q22XT1
B	349	GLN	-	expression tag	UNP Q22XT1
B	350	PHE	-	expression tag	UNP Q22XT1
B	351	GLU	-	expression tag	UNP Q22XT1
B	352	LYS	-	expression tag	UNP Q22XT1

- Molecule 3 is a protein called Myb-like DNA-binding domain protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
3	C	32	269	171	45	53	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-3	GLY	-	expression tag	UNP Q22VV9
C	-2	PRO	-	expression tag	UNP Q22VV9
C	-1	GLY	-	expression tag	UNP Q22VV9
C	0	ARG	-	expression tag	UNP Q22VV9
C	1	PRO	-	expression tag	UNP Q22VV9

- Molecule 4 is a protein called Transmembrane protein, putative.

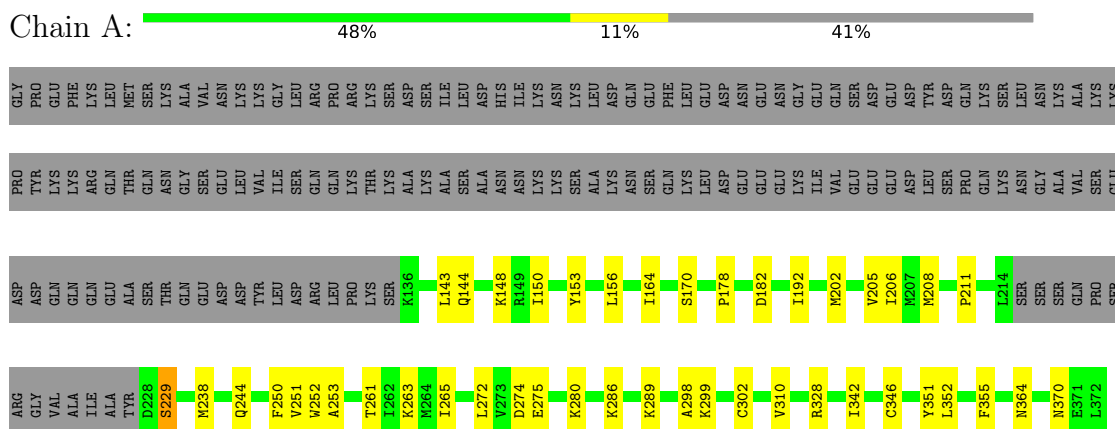| Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	128	1051	666	177	204	4	0	0

There are 4 discrepancies between the modelled and reference sequences:

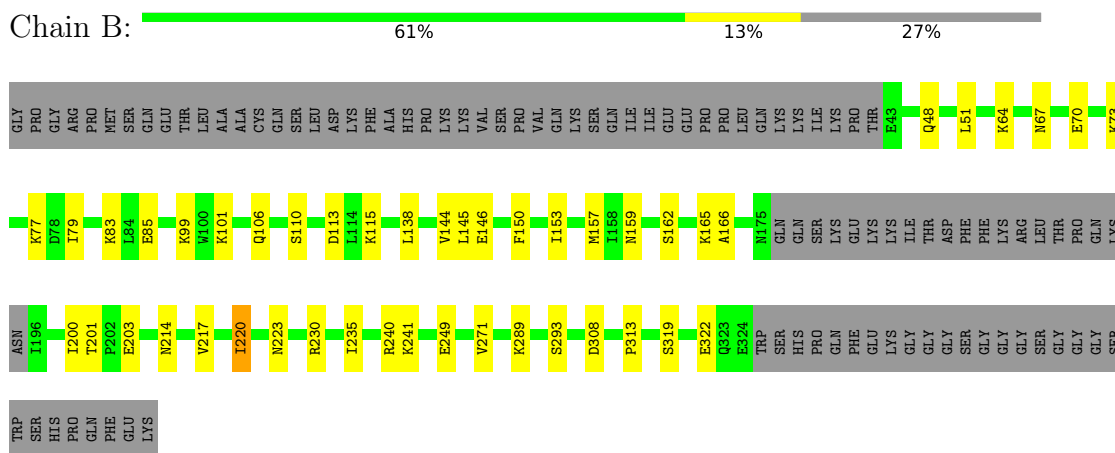| Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	GLY	-	expression tag	UNP I7M8B9
D	-2	PRO	-	expression tag	UNP I7M8B9
D	-1	GLU	-	expression tag	UNP I7M8B9
D	0	PHE	-	expression tag	UNP I7M8B9

3 Residue-property plots ⓘ

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. 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.

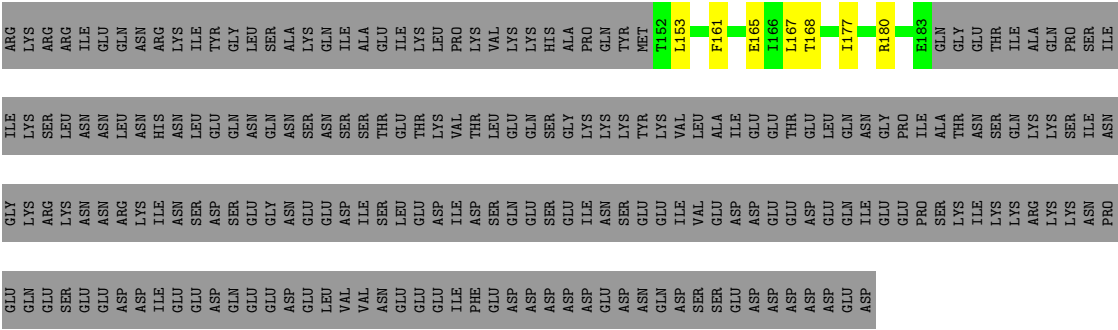• Molecule 1: MT-a70 family protein

Chain A:

• Molecule 2: Methyltransferase MT, putative

Chain B:

• Molecule 3: Myb-like DNA-binding domain protein

Chain C:

ARG LYS ARG ARG ILE GLU GLN ASN ARG LYS ILE TYR GLY LEU SER ALA LYS GLN ILE ALA GLU ILE LYS LEU PRO LYS VAL LYS LYS HIS ALA PRO GLN TYR MET T152 L153 F161 E165 I166 L167 T168 I177 R180 E183 GLN GLY GLU THR ILE ALA GLN PRO SER ILE

ILE LYS SER LEU ASN ASN LEU ASN HIS ASN LEU GLU GLN GLN ASN GLN ASN SER ASN SER THR GLU THR LYS VAL THR LEU GLU GLN GLN SER GLY LYS LYS LYS TYR LYS VAL LEU LEU ALA ILE GLU GLU THR GLU LEU GLN ASN GLY PRO ILE ALA THR ASN SER GLN LYS LYS SER ILE ASN

GLY LYS ARG ASN ASN ALA ARG LYS ILE LEU ASN SER ASP GLU SER GLU GLY ASN GLU GLU ASP ILE SER LEU GLU ASP ILE ASP SER GLN GLU SER GLU ILE ASN SER GLU GLU VAL GLU ASP ASP GLU THR GLU ASP GLU GLN ILE GLU GLU GLU PRO SER LYS ILE LYS LYS ARG LYS LYS ASN PRO

GLU GLN GLU SER GLU GLU ASP ASP ILE GLU GLU GLU ASP GLU GLU GLN GLU ASP LEU VAL VAL ASN GLU GLU GLU ILE PHE GLU ASP ASP ASP ASP GLU ASP ASN GLN ASP LEU ASP SER SER GLU ASP ASP ASP ASP GLU GLU ASP

• Molecule 4: Transmembrane protein, putative

Chain D: 73% 14% 12%

GLY PRO GLU PHE MET LYS LYS ASN GLY LYS SER GLN ASN ASN Q10 P11 L12 M21 R22 L25 S26 C31 L32 A36 L43 I62 F65 G66 V67 I73 E77 I97 S109 Q113 K118 K123 L130 K137 LEU GLN GLN MET GLN

4 Experimental information ⓘ

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	333246	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	TFS KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose $(e^-/Å^2)$	50	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	81000	Depositor
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor

5   Model quality ⓘ

5.1   Standard geometry ⓘ

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	0.17	0/1863	0.40	0/2504
2	B	0.15	0/2209	0.38	0/2965
3	C	0.13	0/271	0.38	0/365
4	D	0.13	0/1068	0.35	0/1430
All	All	0.15	0/5411	0.38	0/7264

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	1826	0	1826	26	0
2	B	2166	0	2160	25	0
3	C	269	0	273	5	0
4	D	1051	0	1046	12	0
All	All	5312	0	5305	62	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 62 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)
4:D:73:ILE:O	4:D:77:GLU:HB3	1.93	0.68
3:C:161:PHE:HB3	3:C:165:GLU:HG3	1.79	0.64
2:B:201:THR:HG22	2:B:203:GLU:H	1.63	0.62
1:A:286:LYS:HG2	2:B:51:LEU:HD21	1.81	0.62
2:B:64:LYS:HD2	2:B:313:PRO:HB2	1.84	0.60

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	220/378 (58%)	203 (92%)	17 (8%)	0	100	100
2	B	258/357 (72%)	246 (95%)	12 (5%)	0	100	100
3	C	30/364 (8%)	29 (97%)	1 (3%)	0	100	100
4	D	126/146 (86%)	120 (95%)	6 (5%)	0	100	100
All	All	634/1245 (51%)	598 (94%)	36 (6%)	0	100	100

There are no Ramachandran outliers to report.

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	199/337 (59%)	195 (98%)	4 (2%)	50	81
2	B	243/324 (75%)	236 (97%)	7 (3%)	37	71

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	C	30/335 (9%)	29 (97%)	1 (3%)	33	67
4	D	114/130 (88%)	111 (97%)	3 (3%)	41	75
All	All	586/1126 (52%)	571 (97%)	15 (3%)	42	75

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	220	ILE
4	D	123	LYS
2	B	271	VAL
4	D	130	LEU
3	C	167	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	158	ASN
3	C	172	ASN
4	D	133	ASN
4	D	10	GLN
1	A	312	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 ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

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.