



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 10:52 AM UTC

PDB ID : 4YOC / pdb_00004yoc
Title : Crystal Structure of human DNMT1 and USP7/HAUSP complex
Authors : Cheng, J.; Yang, H.; Fang, J.; Gong, R.; Wang, P.; Li, Z.; Xu, Y.
Deposited on : 2015-03-11
Resolution : 2.92 Å(reported)

This is a wwPDB X-ray Structure 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/XrayValidationReportHelp>

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:

MolProbity : 4-5-2 with Phenix2.0
Xtrriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

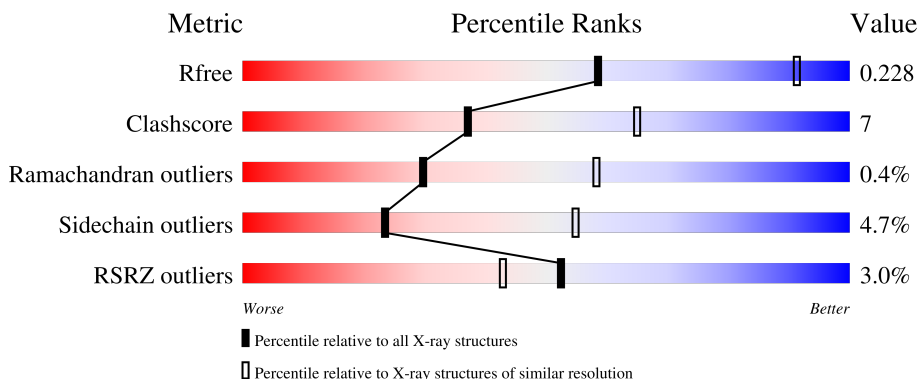
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.92 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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	180053	2995 (2.94-2.90)
Clashscore	190562	3213 (2.94-2.90)
Ramachandran outliers	187476	3128 (2.94-2.90)
Sidechain outliers	187428	3130 (2.94-2.90)
RSRZ outliers	180081	2995 (2.94-2.90)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower 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 $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1004	
2	C	548	

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 11386 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, 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 DNA (cytosine-5)-methyltransferase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	882	7033	4453	1240	1296	44	0	1	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	597	SER	-	expression tag	UNP P26358
A	598	GLU	-	expression tag	UNP P26358
A	599	PHE	-	expression tag	UNP P26358

- Molecule 2 is a protein called Ubiquitin carboxyl-terminal hydrolase 7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	529	4340	2762	733	818	27	0	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	555	GLY	-	expression tag	UNP Q93009
C	556	PRO	-	expression tag	UNP Q93009
C	557	LEU	-	expression tag	UNP Q93009
C	558	GLY	-	expression tag	UNP Q93009
C	559	SER	-	expression tag	UNP Q93009

- Molecule 3 is ZINC ION (CCD ID: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	Zn	0	0
			2	2		

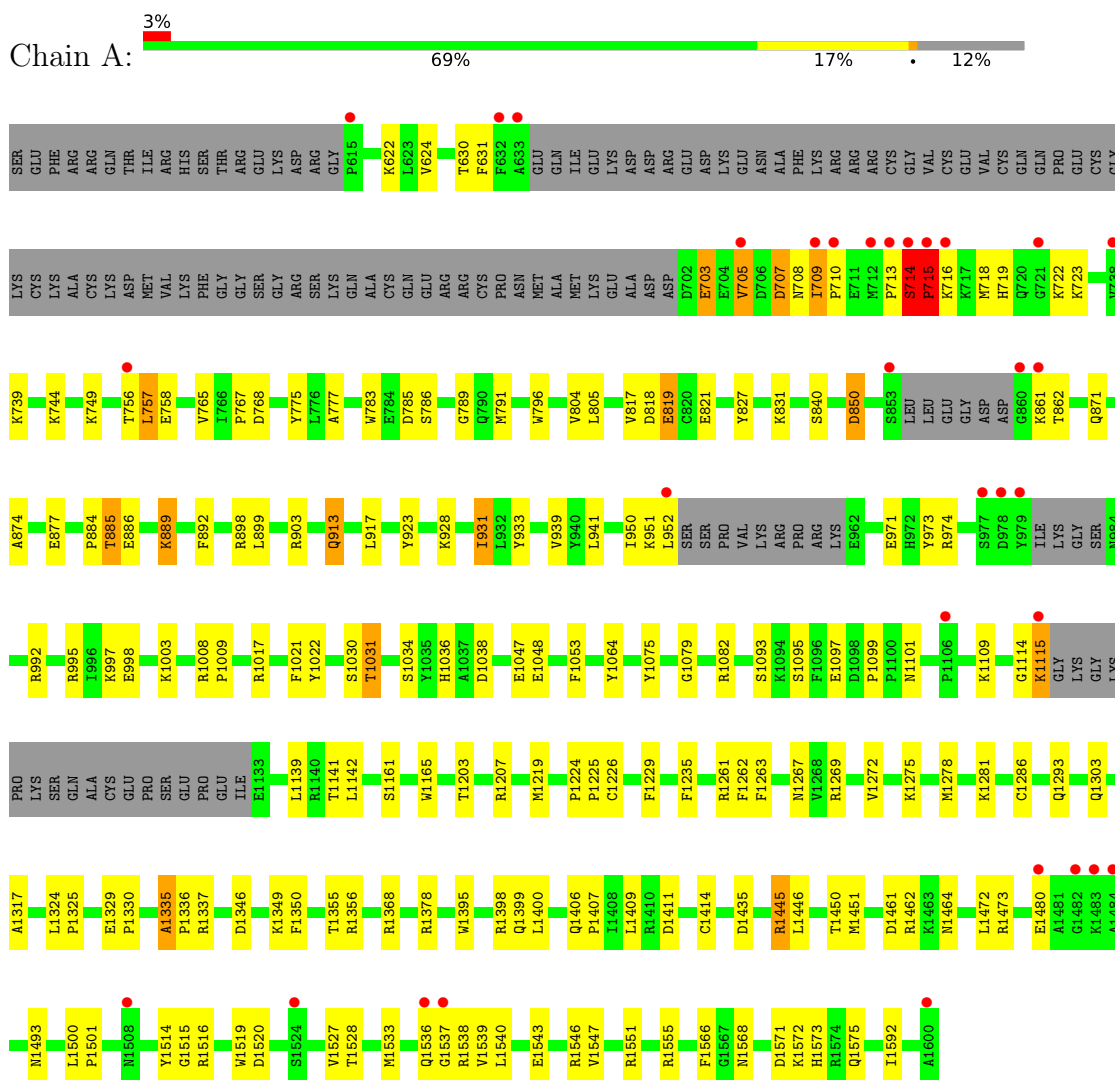
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	9	Total O 9 9	0	0
4	C	2	Total O 2 2	0	0

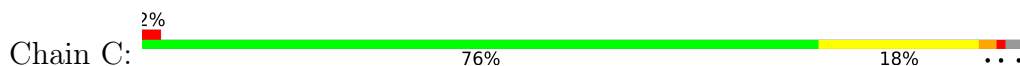
3 Residue-property plots [i](#)

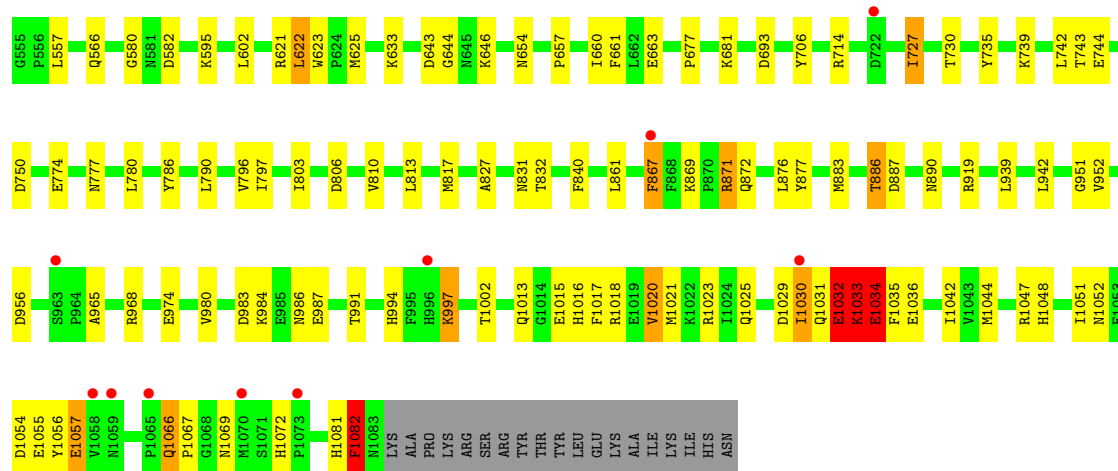
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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 dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). 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: DNA (cytosine-5)-methyltransferase 1



- Molecule 2: Ubiquitin carboxyl-terminal hydrolase 7





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	110.23Å 111.62Å 163.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.42 – 2.92 49.42 – 2.92	Depositor EDS
% Data completeness (in resolution range)	99.2 (49.42-2.92) 99.1 (49.42-2.92)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.96 (at 2.91Å)	Xtrriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.208 , 0.258 0.213 , 0.228	Depositor DCC
R_{free} test set	2248 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	61.7	Xtrriage
Anisotropy	0.220	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 51.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.007 for k,h,-l	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11386	wwPDB-VP
Average B, all atoms (Å ²)	66.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.36% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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

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	0.30	0/7214	0.83	11/9755 (0.1%)
2	C	0.28	0/4440	0.78	9/5995 (0.2%)
All	All	0.30	0/11654	0.81	20/15750 (0.1%)

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	5
2	C	0	2
All	All	0	7

There are no bond length outliers.

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	709	ILE	N-CA-C	9.14	117.86	107.89
2	C	951	GLY	N-CA-C	6.94	121.00	110.18
2	C	1082	PHE	CB-CA-C	-6.78	106.85	116.34
1	A	1079	GLY	CA-C-N	6.50	125.72	118.97
1	A	1079	GLY	C-N-CA	6.50	125.72	118.97

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1536	GLN	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	707	ASP	Peptide
1	A	708	ASN	Peptide
1	A	714	SER	Peptide
1	A	715	PRO	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	7033	0	6901	98	0
2	C	4340	0	4252	62	0
3	A	2	0	0	0	0
4	A	9	0	0	0	0
4	C	2	0	0	0	0
All	All	11386	0	11153	159	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 7.

The worst 5 of 159 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)
1:A:1036:HIS:O	1:A:1398:ARG:NH2	2.20	0.74
2:C:1033:LYS:HE3	2:C:1036:GLU:HB2	1.76	0.67
2:C:867:PHE:H	2:C:867:PHE:HD1	1.42	0.67
2:C:625:MET:HG3	2:C:633:LYS:HG2	1.79	0.65
2:C:919:ARG:NH2	2:C:956:ASP:OD1	2.31	0.64

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 X-ray entries followed by that with respect to entries

of similar resolution.

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	871/1004 (87%)	830 (95%)	39 (4%)	2 (0%)	43 71
2	C	527/548 (96%)	499 (95%)	25 (5%)	3 (1%)	21 50
All	All	1398/1552 (90%)	1329 (95%)	64 (5%)	5 (0%)	30 58

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	1033	LYS
2	C	1081	HIS
1	A	819	GLU
2	C	980	VAL
1	A	705	VAL

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 X-ray entries followed by that with respect to entries of similar resolution.

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	764/869 (88%)	730 (96%)	34 (4%)	24 56
2	C	484/501 (97%)	460 (95%)	24 (5%)	22 52
All	All	1248/1370 (91%)	1190 (95%)	58 (5%)	23 56

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1207	ARG
2	C	1057	GLU
2	C	622	LEU
2	C	1054	ASP
2	C	1020	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such

sidechains are listed below:

Mol	Chain	Res	Type
2	C	1081	HIS
2	C	1013	GLN
2	C	700	ASN
2	C	581	ASN
2	C	851	ASN

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 2 ligands modelled in this entry, 2 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 [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	882/1004 (87%)	0.01	32 (3%) 46 37	23, 54, 116, 158	1 (0%)
2	C	529/548 (96%)	0.09	10 (1%) 66 58	31, 61, 136, 173	0
All	All	1411/1552 (90%)	0.04	42 (2%) 52 43	23, 58, 127, 173	1 (0%)

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1483	LYS	5.4
1	A	633	ALA	5.3
1	A	709	ILE	5.0
2	C	996	HIS	4.6
1	A	715	PRO	4.2

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q < 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	ZN	A	1701	1/1	0.99	0.03	76,76,76,76	0
3	ZN	A	1702	1/1	0.99	0.03	59,59,59,59	0

6.5 Other polymers [i](#)

There are no such residues in this entry.