



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 2, 2021 – 04:44 PM EDT

PDB ID : 7KZ0  
Title : Human MBD4 glycosylase domain bound to DNA containing substrate analog  
2'-deoxy-pseudouridine  
Authors : Pidugu, L.S.; Pozharski, E.; Drohat, A.C.  
Deposited on : 2020-12-09  
Resolution : 1.57 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.23.2  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.23.2

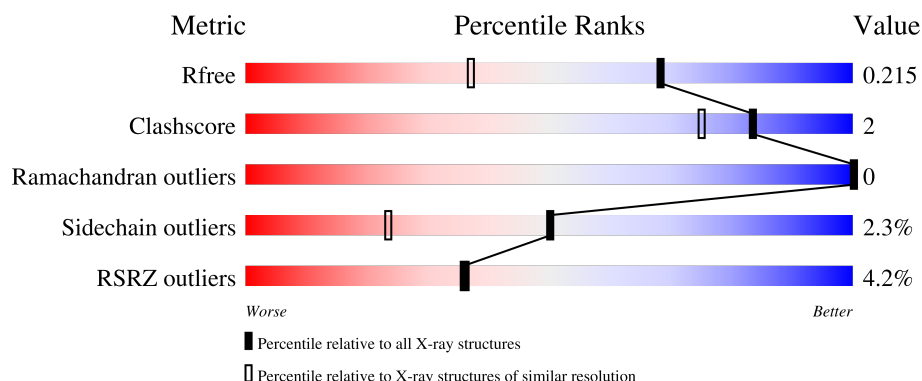
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*




The reported resolution of this entry is 1.57 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))
$R_{free}$	130704	5534 (1.60-1.56)
Clashscore	141614	5861 (1.60-1.56)
Ramachandran outliers	138981	5708 (1.60-1.56)
Sidechain outliers	138945	5703 (1.60-1.56)
RSRZ outliers	127900	5431 (1.60-1.56)

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  $\geq 3$ , 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	174	 4% 79% 17%
2	C	12	 58% 42%
3	D	12	 50% 50%

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 2022 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 Methyl-CpG-binding domain protein 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	144	Total	C	N	O	S	0	3	0
			1247	821	212	212	2			

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	407	MET	-	expression tag	UNP O95243
A	408	GLY	-	expression tag	UNP O95243
A	409	SER	-	expression tag	UNP O95243
A	410	SER	-	expression tag	UNP O95243
A	411	HIS	-	expression tag	UNP O95243
A	412	HIS	-	expression tag	UNP O95243
A	413	HIS	-	expression tag	UNP O95243
A	414	HIS	-	expression tag	UNP O95243
A	415	HIS	-	expression tag	UNP O95243
A	416	HIS	-	expression tag	UNP O95243
A	417	SER	-	expression tag	UNP O95243
A	418	SER	-	expression tag	UNP O95243
A	419	GLY	-	expression tag	UNP O95243
A	420	LEU	-	expression tag	UNP O95243
A	421	VAL	-	expression tag	UNP O95243
A	422	PRO	-	expression tag	UNP O95243
A	423	ARG	-	expression tag	UNP O95243
A	424	GLY	-	expression tag	UNP O95243
A	425	SER	-	expression tag	UNP O95243

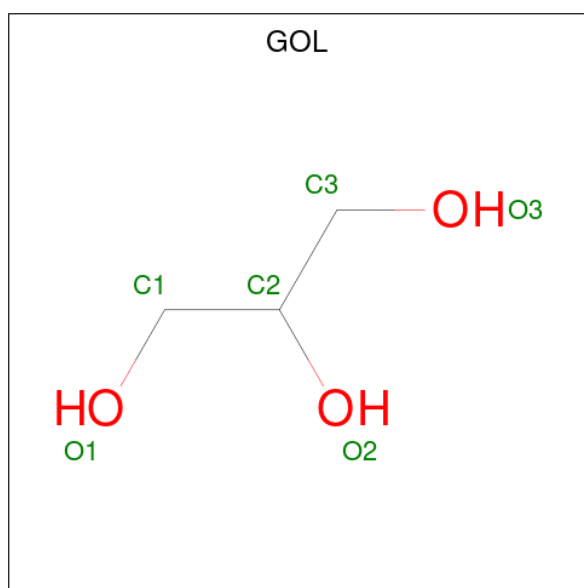
- Molecule 2 is a DNA chain called DNA (5'-D(\*CP\*CP\*AP\*GP\*CP\*GP\*(P2U)P\*GP\*CP\*AP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	12	Total	C	N	O	P	0	0	0
			241	114	47	69	11			

- Molecule 3 is a DNA chain called DNA (5'-D(\*GP\*CP\*TP\*GP\*CP\*GP\*CP\*GP\*CP\*TP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	12	Total	C	N	O	P	0	0	0
			245	116	46	72	11			

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			6	3	3		
4	A	1	Total	C	O	0	0
			6	3	3		
4	D	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total K	0	0
			1 1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	174	Total O	0	3
			177 177		

*Continued on next page...*


*Continued from previous page...*

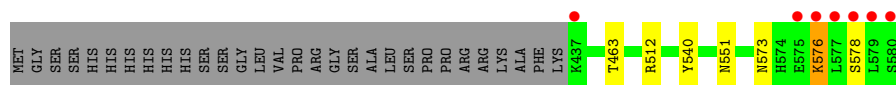
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	49	Total	O	0	3
			52	52		
6	D	40	Total	O	0	1
			41	41		

### 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.

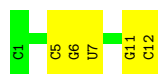
- Molecule 1: Methyl-CpG-binding domain protein 4

Chain A: 



- Molecule 2: DNA (5'-D(\*CP\*CP\*AP\*GP\*CP\*GP\*(P2U)P\*GP\*CP\*AP\*GP\*C)-3')

Chain C: 



- Molecule 3: DNA (5'-D(\*GP\*CP\*TP\*GP\*CP\*GP\*CP\*GP\*CP\*TP\*GP\*G)-3')

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	41.36Å 56.86Å 105.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.65 – 1.57 38.65 – 1.57	Depositor EDS
% Data completeness (in resolution range)	99.3 (38.65-1.57) 99.6 (38.65-1.57)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.29 (at 1.57Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.188 , 0.215 0.189 , 0.215	Depositor DCC
$R_{free}$ test set	1771 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.9	Xtriage
Anisotropy	0.297	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 51.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2022	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.12% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: P2U, GOL, K

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	0.34	0/1302	0.50	1/1769 (0.1%)
2	C	0.73	0/248	0.84	0/378
3	D	0.69	0/274	0.88	0/422
All	All	0.48	0/1824	0.64	1/2569 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	512	ARG	NE-CZ-NH1	-5.01	117.80	120.30

There are no chirality outliers.

There are no planarity outliers.

### 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	1247	0	1233	2	0
2	C	241	0	133	2	0
3	D	245	0	136	4	0
4	A	12	0	16	0	0
4	D	6	0	8	2	0
5	A	1	0	0	0	0
6	A	177	0	0	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	C	52	0	0	0	0
6	D	41	0	0	0	0
All	All	2022	0	1526	8	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 2.

All (8) 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:573:ASN:HA	1:A:576:LYS:HD2	1.81	0.64
2:C:5:DC:H2''	2:C:6:DG:C8	2.33	0.64
2:C:11:DG:H2''	2:C:12:DC:H5''	1.82	0.62
3:D:7:DC:H2'	3:D:8:DG:C8	2.45	0.52
3:D:4:DG:O6	4:D:101:GOL:H32	2.11	0.51
3:D:1:DG:H2''	3:D:2:DC:H5'	2.02	0.42
3:D:3:DT:H73	4:D:101:GOL:H11	2.02	0.42
1:A:463:THR:HB	1:A:540:TYR:CE2	2.56	0.41

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 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	145/174 (83%)	144 (99%)	1 (1%)	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 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	136/158 (86%)	133 (98%)	3 (2%)	52 25

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	551	ASN
1	A	576	LYS
1	A	578	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	551	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

1 non-standard protein/DNA/RNA residue 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	$\# Z  > 2$
2	P2U	C	7	2	16,20,21	1.27	3 (18%)	17,28,31	6.13	5 (29%)

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
2	P2U	C	7	2	-	0/7/21/22	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	7	P2U	C4-N3	3.10	1.38	1.33
2	C	7	P2U	C6-C5	-2.37	1.35	1.38
2	C	7	P2U	C4-C5	2.27	1.46	1.41

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	7	P2U	N1-C2-N3	-18.00	114.12	128.43
2	C	7	P2U	C4-N3-C2	14.48	127.36	115.14
2	C	7	P2U	C5-C4-N3	-8.25	114.73	125.36
2	C	7	P2U	C6-N1-C2	4.55	122.87	115.36
2	C	7	P2U	O4'-C1'-C2'	3.10	108.57	103.57

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 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	$\# Z  > 2$
4	GOL	A	603	-	5,5,5	0.07	0	5,5,5	0.20	0
4	GOL	D	101	-	5,5,5	0.03	0	5,5,5	0.24	0
4	GOL	A	601	-	5,5,5	0.08	0	5,5,5	0.18	0

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
4	GOL	A	603	-	-	0/4/4/4	-
4	GOL	D	101	-	-	0/4/4/4	-
4	GOL	A	601	-	-	0/4/4/4	-

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 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	101	GOL	2	0

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	144/174 (82%)	0.27	7 (4%) 29 29	15, 22, 43, 71	0
2	C	11/12 (91%)	-0.06	0 100 100	17, 32, 42, 46	0
3	D	12/12 (100%)	-0.02	0 100 100	18, 34, 38, 39	0
All	All	167/198 (84%)	0.22	7 (4%) 36 36	15, 22, 43, 71	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	579	LEU	5.6
1	A	578	SER	5.0
1	A	580	SER	4.2
1	A	577	LEU	3.1
1	A	437	LYS	3.1
1	A	576	LYS	2.9
1	A	575	GLU	2.5

### 6.2 Non-standard residues in protein, DNA, RNA chains [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, 95<sup>th</sup> 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(Å <sup>2</sup> )	Q<0.9
2	P2U	C	7	19/20	0.96	0.07	15,17,26,27	0

### 6.3 Carbohydrates [i](#)

There are no monosaccharides 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
4	GOL	D	101	6/6	0.72	0.20	41,47,52,58	0
4	GOL	A	601	6/6	0.81	0.12	29,34,37,38	0
4	GOL	A	603	6/6	0.94	0.09	25,26,27,28	0
5	K	A	602	1/1	0.99	0.04	22,22,22,22	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.