



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 04:14 PM GMT

PDB ID : 4E9G  
Title : structure of the glycosylase domain of MBD4 bound to thymine containing DNA  
Authors : Morera, S.; Vigouroux, A.  
Deposited on : 2012-03-21  
Resolution : 2.35 Å(reported)

This is a wwPDB validation summary 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

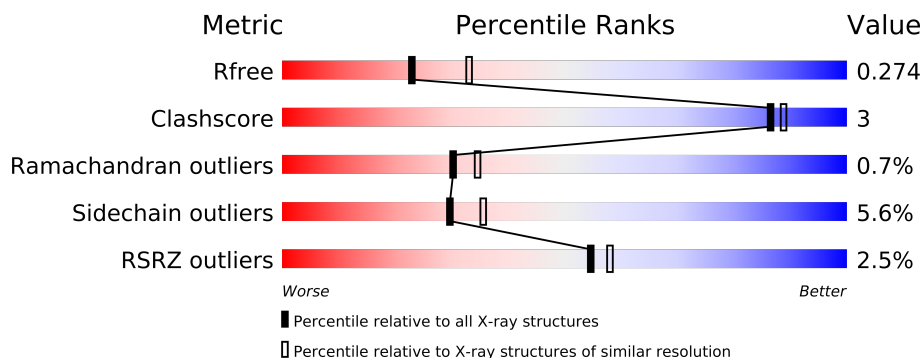
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.35 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))
$R_{free}$	66092	3327 (2.40-2.32)
Clashscore	79885	1064 (2.38-2.34)
Ramachandran outliers	78287	1048 (2.38-2.34)
Sidechain outliers	78261	1049 (2.38-2.34)
RSRZ outliers	66119	3330 (2.40-2.32)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	161	
2	C	12	
3	D	12	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 1695 atoms, of which 0 are hydrogen and 0 are deuterium.

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	138	Total	C	N	O	S	0	0	0
			1176	775	198	201	2			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	426	MET	-	EXPRESSION TAG	UNP O95243
A	560	ALA	ASP	ENGINEERED MUTATION	UNP O95243
A	581	HIS	-	EXPRESSION TAG	UNP O95243
A	582	HIS	-	EXPRESSION TAG	UNP O95243
A	583	HIS	-	EXPRESSION TAG	UNP O95243
A	584	HIS	-	EXPRESSION TAG	UNP O95243
A	585	HIS	-	EXPRESSION TAG	UNP O95243
A	586	HIS	-	EXPRESSION TAG	UNP O95243

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	12	Total	C	N	O	P	0	0	0
			242	115	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 water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	20	Total 20	O 20	0	0
4	C	6	Total 6	O 6	0	0
4	D	6	Total 6	O 6	0	0

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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\*TP\*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, $\alpha$ , $\beta$ , $\gamma$	40.08Å 61.47Å 96.56Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.97 – 2.35 37.97 – 2.35	Depositor EDS
% Data completeness (in resolution range)	97.8 (37.97-2.35) 98.1 (37.97-2.35)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.32 (at 2.34Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, $R_{free}$	0.205 , 0.258 0.213 , 0.274	Depositor DCC
$R_{free}$ test set	1012 reflections (11.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	42.7	Xtriage
Anisotropy	0.341	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 34.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 10209 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	1695	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.60% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 5 Model quality i

### 5.1 Standard geometry i

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	0.51	0/1219	0.67	0/1660
2	C	1.12	0/271	1.85	7/416 (1.7%)
3	D	1.11	0/274	1.90	9/422 (2.1%)
All	All	0.75	0/1764	1.21	16/2498 (0.6%)

There are no bond length outliers.

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	D	1	DG	P-O3'-C3'	10.55	132.36	119.70
2	C	2	DC	O4'-C1'-N1	9.59	114.71	108.00
2	C	7	DT	P-O3'-C3'	8.07	129.38	119.70
3	D	8	DG	P-O3'-C3'	7.49	128.68	119.70
3	D	10	DT	O4'-C1'-N1	7.30	113.11	108.00

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1176	0	1155	6	0
2	C	242	0	135	2	0
3	D	245	0	136	3	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	20	0	0	0	0
4	C	6	0	0	1	0
4	D	6	0	0	0	0
All	All	1695	0	1426	8	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 3.

The worst 5 of 8 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:D:7:DC:H2''	3:D:8:DG:C8	2.30	0.65
2:C:6:DG:H3'	4:C:102:HOH:O	2.09	0.52
1:A:466:LEU:O	2:C:7:DT:H3'	2.16	0.45
1:A:556:VAL:HG12	1:A:558:PRO:HD3	1.99	0.45
1:A:507:GLY:O	1:A:508:LEU:C	2.55	0.44

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	136/161 (84%)	129 (95%)	6 (4%)	1 (1%)	30 34

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	574	HIS

### 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	126/148 (85%)	119 (94%)	7 (6%)	30	35

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	493	THR
1	A	562	LYS
1	A	501	GLU
1	A	480	LYS
1	A	502	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	467	ASN
1	A	564	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 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.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	138/161 (85%)	0.04	2 (1%) 72 76	25, 41, 64, 96	0
2	C	12/12 (100%)	0.19	1 (8%) 11 13	36, 57, 67, 78	0
3	D	12/12 (100%)	0.41	1 (8%) 11 13	46, 65, 79, 81	0
All	All	162/185 (87%)	0.08	4 (2%) 54 58	25, 43, 68, 96	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	12	DC	2.7
1	A	561	HIS	2.5
3	D	1	DG	2.5
1	A	527	GLN	2.1

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

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

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

There are no ligands in this entry.

### 6.5 Other polymers ⓘ

There are no such residues in this entry.