



# wwPDB X-ray Structure Validation Summary Report

Feb 26, 2014 – 11:02 PM GMT

PDB ID : 2H56  
Title : Crystal structure of DNA-3-methyladenine glycosidase (10174367) from *Bacillus halodurans* at 2.55 Å resolution  
Authors : Joint Center for Structural Genomics (JCSG)  
Deposited on : 2006-05-25  
Resolution : 2.55 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>

---

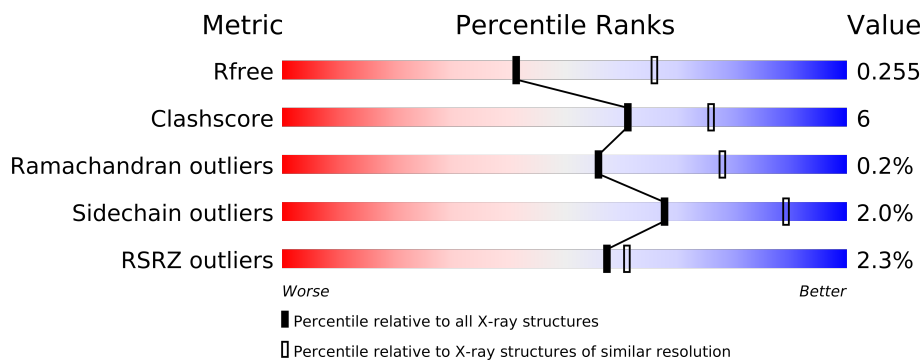
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.55 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	3413 (2.58-2.50)
Clashscore	79885	4284 (2.58-2.50)
Ramachandran outliers	78287	4193 (2.58-2.50)
Sidechain outliers	78261	4195 (2.58-2.50)
RSRZ outliers	66119	3414 (2.58-2.50)

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	233	
1	B	233	
1	C	233	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4855 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 DNA-3-methyladenine glycosidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	218	Total	C	N	O	S	Se	0	0	0
			1650	1062	281	301	2	4			
1	B	217	Total	C	N	O	S	Se	0	0	0
			1653	1067	282	298	2	4			
1	C	204	Total	C	N	O	S	Se	0	0	0
			1512	968	262	276	2	4			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MSE	-	LEADER SEQUENCE	UNP Q9KC25
A	-10	GLY	-	LEADER SEQUENCE	UNP Q9KC25
A	-9	SER	-	LEADER SEQUENCE	UNP Q9KC25
A	-8	ASP	-	LEADER SEQUENCE	UNP Q9KC25
A	-7	LYS	-	LEADER SEQUENCE	UNP Q9KC25
A	-6	ILE	-	LEADER SEQUENCE	UNP Q9KC25
A	-5	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	-4	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	-3	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	-2	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	-1	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	0	HIS	-	LEADER SEQUENCE	UNP Q9KC25
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
A	141	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
A	143	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
A	144	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
B	-11	MSE	-	LEADER SEQUENCE	UNP Q9KC25
B	-10	GLY	-	LEADER SEQUENCE	UNP Q9KC25
B	-9	SER	-	LEADER SEQUENCE	UNP Q9KC25
B	-8	ASP	-	LEADER SEQUENCE	UNP Q9KC25
B	-7	LYS	-	LEADER SEQUENCE	UNP Q9KC25
B	-6	ILE	-	LEADER SEQUENCE	UNP Q9KC25
B	-5	HIS	-	LEADER SEQUENCE	UNP Q9KC25

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	-3	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	-2	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	-1	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	0	HIS	-	LEADER SEQUENCE	UNP Q9KC25
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
B	141	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
B	143	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
B	144	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
C	-11	MSE	-	LEADER SEQUENCE	UNP Q9KC25
C	-10	GLY	-	LEADER SEQUENCE	UNP Q9KC25
C	-9	SER	-	LEADER SEQUENCE	UNP Q9KC25
C	-8	ASP	-	LEADER SEQUENCE	UNP Q9KC25
C	-7	LYS	-	LEADER SEQUENCE	UNP Q9KC25
C	-6	ILE	-	LEADER SEQUENCE	UNP Q9KC25
C	-5	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	-4	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	-3	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	-2	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	-1	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	0	HIS	-	LEADER SEQUENCE	UNP Q9KC25
C	1	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
C	141	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
C	143	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25
C	144	MSE	MET	MODIFIED RESIDUE	UNP Q9KC25

- Molecule 2 is water.

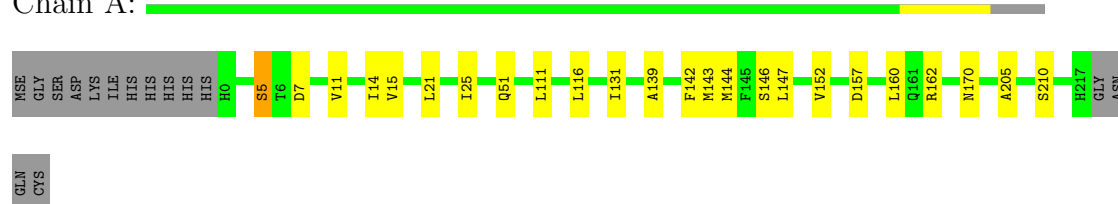
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	19	Total O 19 19	0	0
2	B	13	Total O 13 13	0	0
2	C	8	Total O 8 8	0	0

### 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.

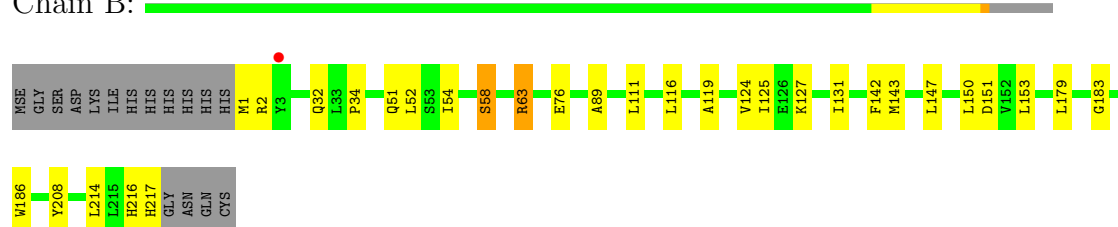
- Molecule 1: DNA-3-methyladenine glycosidase

Chain A:



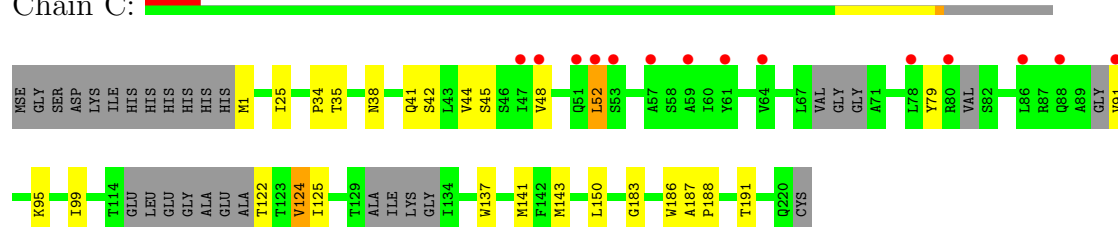
- Molecule 1: DNA-3-methyladenine glycosidase

Chain B:



- Molecule 1: DNA-3-methyladenine glycosidase

Chain C:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	141.75Å 141.75Å 85.95Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.40 – 2.55 46.40 – 2.55	Depositor EDS
% Data completeness (in resolution range)	98.0 (46.40-2.55) 98.0 (46.40-2.55)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.92 (at 2.54Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.196 , 0.246 0.209 , 0.255	Depositor DCC
$R_{free}$ test set	1606 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.1	Xtriage
Anisotropy	0.289	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 81.0	EDS
Estimated twinning fraction	0.044 for h,-h-k,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 31729 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4855	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.49% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 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	0.71	0/1679	0.72	0/2268
1	B	0.68	0/1685	0.69	1/2277 (0.0%)
1	C	0.66	0/1537	0.65	0/2073
All	All	0.69	0/4901	0.69	1/6618 (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	B	151	ASP	CB-CG-OD1	5.29	123.06	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 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	1650	0	1612	17	0
1	B	1653	0	1609	23	0
1	C	1512	0	1394	16	0
2	A	19	0	0	0	0
2	B	13	0	0	0	0
2	C	8	0	0	0	0
All	All	4855	0	4615	55	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 6.

The worst 5 of 55 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:125:ILE:CG1	1:B:143:MSE:HE1	1.49	1.40
1:A:142:PHE:CE2	1:A:147:LEU:HD11	2.00	0.96
1:B:125:ILE:CG1	1:B:143:MSE:CE	2.42	0.96
1:B:1:MSE:HE3	1:B:34:PRO:N	2.07	0.69
1:C:52:LEU:HD22	1:C:95:LYS:NZ	2.08	0.69

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	216/233 (93%)	208 (96%)	8 (4%)	0	100	100
1	B	215/233 (92%)	209 (97%)	6 (3%)	0	100	100
1	C	192/233 (82%)	180 (94%)	11 (6%)	1 (0%)	38	60
All	All	623/699 (89%)	597 (96%)	25 (4%)	1 (0%)	56	78

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	79	TYR

### 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	159/188 (85%)	157 (99%)	2 (1%)	80	94
1	B	158/188 (84%)	156 (99%)	2 (1%)	80	94
1	C	138/188 (73%)	133 (96%)	5 (4%)	47	72
All	All	455/564 (81%)	446 (98%)	9 (2%)	68	89

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	35	THR
1	C	124	VAL
1	C	45	SER
1	B	58	SER
1	C	42	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	51	GLN
1	B	216	HIS
1	B	217	HIS

### 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	218/233 (93%)	-0.20	0 100 100	61, 69, 83, 91	0
1	B	217/233 (93%)	-0.13	1 (0%) 88 90	54, 70, 82, 91	0
1	C	204/233 (87%)	0.29	14 (6%) 17 17	53, 71, 84, 91	0
All	All	639/699 (91%)	-0.02	15 (2%) 57 61	53, 70, 83, 91	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	57	ALA	6.0
1	C	53	SER	5.1
1	C	86	LEU	4.4
1	C	61	TYR	4.2
1	C	64	VAL	4.1

### 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.