Full wwPDB X-ray Structure Validation Report ⓘ

Nov 12, 2026 – 02:12 PM JST

PDB ID	:	24YH / pdb_000024yh
Title	:	3-methyladenine DNA glycosylase mutant D223A bound to DNA/Hx
Authors	:	Xie, W.; Jia, Q.; Jiang, P.
Deposited on	:	2026-03-24
Resolution	:	1.76 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*
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:

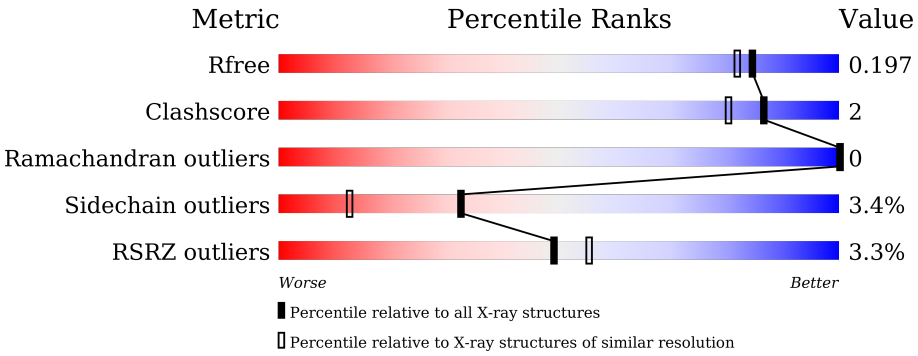
MolProbity	:	4-5-2 with Phenix2.2
Xtriage (Phenix)	:	2.2
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.51

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.76 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	3183 (1.76-1.76)
Clashscore	190562	3299 (1.76-1.76)
Ramachandran outliers	187476	3274 (1.76-1.76)
Sidechain outliers	187428	3274 (1.76-1.76)
RSRZ outliers	180081	3183 (1.76-1.76)

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 <=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	286	3% 88% 7%
1	D	286	2% 88% 7%
2	B	11	100%
3	C	11	18% 82% 9% 9%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5444 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 Base excision DNA repair protein, HhH-GPD superfamily.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	274	Total	C	N	O	S	0	0	0
			2200	1431	364	400	5			
1	D	274	Total	C	N	O	S	0	0	0
			2206	1434	367	400	5			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	223	ALA	ASP	engineered mutation	UNP C5A632
A	279	LEU	-	expression tag	UNP C5A632
A	280	GLU	-	expression tag	UNP C5A632
A	281	HIS	-	expression tag	UNP C5A632
A	282	HIS	-	expression tag	UNP C5A632
A	283	HIS	-	expression tag	UNP C5A632
A	284	HIS	-	expression tag	UNP C5A632
A	285	HIS	-	expression tag	UNP C5A632
A	286	HIS	-	expression tag	UNP C5A632
D	223	ALA	ASP	engineered mutation	UNP C5A632
D	279	LEU	-	expression tag	UNP C5A632
D	280	GLU	-	expression tag	UNP C5A632
D	281	HIS	-	expression tag	UNP C5A632
D	282	HIS	-	expression tag	UNP C5A632
D	283	HIS	-	expression tag	UNP C5A632
D	284	HIS	-	expression tag	UNP C5A632
D	285	HIS	-	expression tag	UNP C5A632
D	286	HIS	-	expression tag	UNP C5A632

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*CP*GP*CP*TP*AP*CP*AP*TP*GP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	P	0	0	0
			220	107	37	66	10			

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*CP*IP*TP*GP*TP*AP*GP*CP*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	11	Total	C	N	O	P	0	0	0
			225	108	44	63	10			

- Molecule 4 is water.

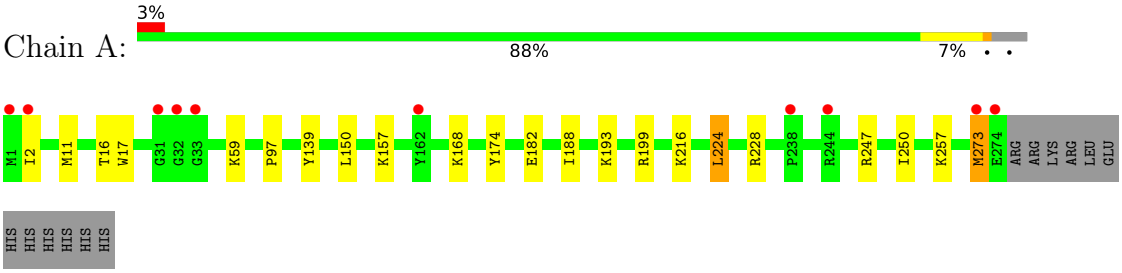
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	265	Total	O	0	0
			265	265		
4	B	20	Total	O	0	0
			20	20		
4	C	16	Total	O	0	0
			16	16		
4	D	292	Total	O	0	0
			292	292		

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.

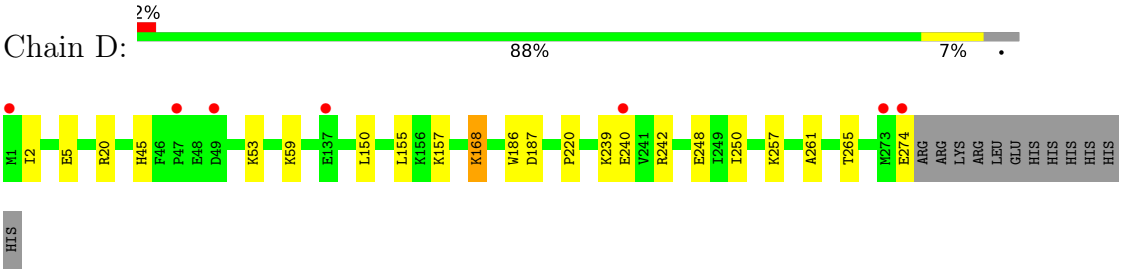
• Molecule 1: Base excision DNA repair protein, HhH-GPD superfamily

Chain A:



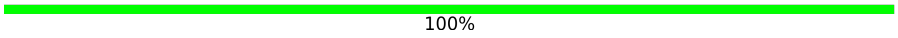
• Molecule 1: Base excision DNA repair protein, HhH-GPD superfamily

Chain D:



• Molecule 2: DNA (5'-D(*TP*CP*GP*CP*TP*AP*CP*AP*TP*GP*T)-3')

Chain B:

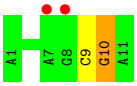


There are no outlier residues recorded for this chain.

• Molecule 3: DNA (5'-D(*AP*CP*IP*TP*GP*TP*AP*GP*CP*GP*A)-3')

Chain C:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	66.10Å 72.19Å 155.33Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	77.67 – 1.76 77.67 – 1.76	Depositor EDS
% Data completeness (in resolution range)	94.5 (77.67-1.76) 94.5 (77.67-1.76)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.25 (at 1.76Å)	Xtriage
Refinement program	PHENIX 1.8	Depositor
R, R_{free}	0.168 , 0.195 0.173 , 0.197	Depositor DCC
R_{free} test set	3451 reflections (4.66%)	wwPDB-VP
Wilson B-factor (Å^2)	22.7	Xtriage
Anisotropy	0.058	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 38.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5444	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.25% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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	1.01	1/2253 (0.0%)	1.20	0/3040
1	D	1.04	1/2259 (0.0%)	1.20	3/3047 (0.1%)
2	B	0.68	0/245	0.95	0/376
3	C	0.71	0/228	0.97	1/348 (0.3%)
All	All	1.00	2/4985 (0.0%)	1.18	4/6811 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	220	PRO	C-O	-5.33	1.19	1.24
1	A	182	GLU	C-O	5.09	1.29	1.23

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	186	TRP	CA-C-N	6.51	129.96	120.71
1	D	186	TRP	C-N-CA	6.51	129.96	120.71
1	D	187	ASP	CA-CB-CG	5.73	118.33	112.60
3	C	10	DG	P-O5'-C5'	-5.19	112.59	120.90

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	2200	0	2190	13	0
1	D	2206	0	2201	7	0
2	B	220	0	127	0	0
3	C	225	0	124	1	0
4	A	265	0	0	4	0
4	B	20	0	0	0	0
4	C	16	0	0	0	0
4	D	292	0	0	2	0
All	All	5444	0	4642	20	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 (20) 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:D:45:HIS:HD2	4:D:502:HOH:O	1.69	0.76
1:D:261:ALA:O	1:D:265:THR:HG23	1.94	0.68
1:A:199:ARG:NH1	4:A:301:HOH:O	2.26	0.68
3:C:9:DC:H2"	3:C:10:DG:H5"	1.76	0.67
1:A:157:LYS:HD2	4:A:444:HOH:O	2.00	0.60
1:D:155:LEU:HD11	1:D:168:LYS:HD3	1.84	0.59
1:A:11:MET:HE1	4:A:430:HOH:O	2.10	0.52
1:A:247:ARG:NH2	4:A:310:HOH:O	2.48	0.47
1:D:150:LEU:O	1:D:168:LYS:NZ	2.46	0.46
1:A:150:LEU:O	1:A:168:LYS:NZ	2.49	0.46
1:A:188:ILE:HG12	1:A:216:LYS:HD3	1.98	0.45
1:D:250:ILE:HG22	1:D:257:LYS:HB2	1.98	0.45
1:A:250:ILE:HG22	1:A:257:LYS:HB2	1.99	0.45
1:D:5:GLU:OE2	4:D:301:HOH:O	2.20	0.44
1:A:174:TYR:CZ	1:D:20:ARG:HD2	2.54	0.41
1:A:11:MET:HG2	1:A:97:PRO:HB3	2.01	0.41
1:A:273:MET:HE2	1:A:273:MET:HB2	1.85	0.40
1:A:16:THR:HA	1:A:139:TYR:O	2.21	0.40
1:A:11:MET:HE2	1:A:17:TRP:NE1	2.37	0.40
1:A:224:LEU:HD11	1:A:228:ARG:NH1	2.37	0.40

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	272/286 (95%)	266 (98%)	6 (2%)	0	100	100
1	D	272/286 (95%)	270 (99%)	2 (1%)	0	100	100
All	All	544/572 (95%)	536 (98%)	8 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	221/237 (93%)	216 (98%)	5 (2%)	44	24
1	D	222/237 (94%)	212 (96%)	10 (4%)	24	7
All	All	443/474 (94%)	428 (97%)	15 (3%)	32	13

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	ILE
1	A	59	LYS
1	A	193	LYS
1	A	224	LEU
1	A	273	MET
1	D	2	ILE
1	D	53	LYS
1	D	59	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	157	LYS
1	D	168	LYS
1	D	239	LYS
1	D	240	GLU
1	D	242	ARG
1	D	248	GLU
1	D	274	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	26	GLN
1	A	114	GLN
1	D	26	GLN
1	D	104	GLN

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)

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.

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^{th} 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	274/286 (95%)	0.18	10 (3%) 46 52	15, 24, 47, 72	0
1	D	274/286 (95%)	0.08	7 (2%) 57 63	15, 21, 43, 72	0
2	B	11/11 (100%)	0.27	0 100 100	24, 30, 54, 56	0
3	C	10/11 (90%)	0.92	2 (20%) 3 3	21, 46, 58, 63	0
All	All	569/594 (95%)	0.14	19 (3%) 49 55	15, 23, 48, 72	0

All (19) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	1	MET	3.9
1	A	32	GLY	3.7
1	D	273	MET	3.6
1	A	1	MET	3.3
1	D	49	ASP	3.3
1	A	31	GLY	2.9
3	C	7	DA	2.8
1	A	273	MET	2.6
1	D	47	PRO	2.6
1	A	33	GLY	2.6
1	A	274	GLU	2.6
1	D	137	GLU	2.6
1	D	274	GLU	2.6
1	A	244	ARG	2.4
1	A	162	TYR	2.3
1	A	238	PRO	2.3
1	D	240	GLU	2.3
3	C	8	DG	2.2
1	A	2	ILE	2.2

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.