Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 04:51 pm BST

PDB ID	:	1PU6
Title	:	Crystal structure of H.pylori 3-methyladenine DNA glycosylase (MagIII)
Authors	:	Eichman, B.F.; O'Rourke, E.J.; Radicella, J.P.; Ellenberger, T.
Deposited on	:	2003-06-24
Resolution	:	1.64 Å(reported)

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

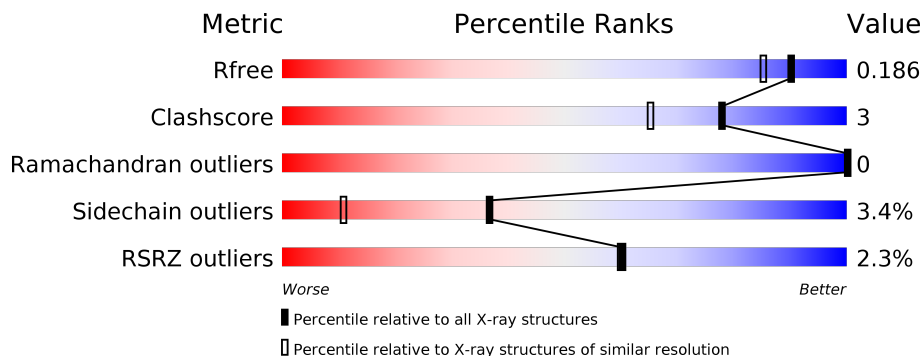|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.64 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	Percentile Ranks	Value
Rfree		0.186
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		3.4%
RSRZ outliers		2.3%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3122 (1.66-1.62)
Clashscore	141614	3268 (1.66-1.62)
Ramachandran outliers	138981	3215 (1.66-1.62)
Sidechain outliers	138945	3215 (1.66-1.62)
RSRZ outliers	127900	3079 (1.66-1.62)

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 >=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	218	3% / 91% / 8%
1	B	218	2% / 91% / 8%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3729 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 3-METHYLADENINE DNA GLYCOSYLASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	217	Total	C	N	O	S	0	0	1
			1731	1131	276	320	4			
1	B	217	Total	C	N	O	S	0	1	0
			1743	1139	280	320	4			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	VAL	MET	SEE REMARK 999	UNP O25323
A	102	GLY	LYS	SEE REMARK 999	UNP O25323
A	120	ARG	LYS	SEE REMARK 999	UNP O25323
A	141	ALA	VAL	SEE REMARK 999	UNP O25323
A	205	KCX	LYS	MODIFIED RESIDUE	UNP O25323
B	1	VAL	MET	SEE REMARK 999	UNP O25323
B	102	GLY	LYS	SEE REMARK 999	UNP O25323
B	120	ARG	LYS	SEE REMARK 999	UNP O25323
B	141	ALA	VAL	SEE REMARK 999	UNP O25323
B	205	KCX	LYS	MODIFIED RESIDUE	UNP O25323

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Cl	0	0
			1	1		

- Molecule 3 is BETA-MERCAPTOETHANOL (three-letter code: BME) (formula: C_2H_6OS).

BME

S2

SH

C1

C2

HO

O1

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total 4	C 2	O 1	S 1	0	0

- Molecule 4 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: $C_6H_{14}O_2$).

MPD

O4

OH

C1

C5

C4(S)

C2

CM

C3

OH

O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total 8	C 6	O 2	0	0
4	B	1	Total 8	C 6	O 2	0	0

- Molecule 5 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	127	Total 127	O 127	0	0
5	B	107	Total 107	O 107	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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: 3-METHYLADENINE DNA GLYCOSYLASE

Chain A:

- Molecule 1: 3-METHYLADENINE DNA GLYCOSYLASE

Chain B:
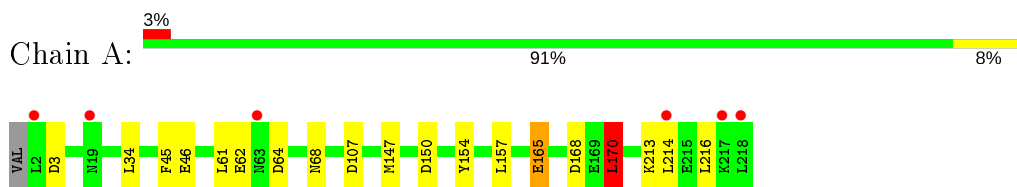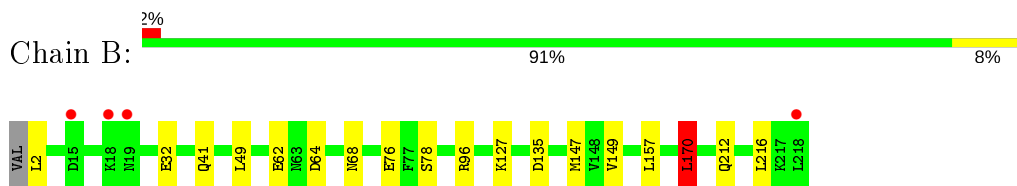
4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	146.60Å 44.40Å 81.52Å 90.00° 106.40° 90.00°	Depositor
Resolution (Å)	24.41 – 1.64 24.41 – 1.64	Depositor EDS
% Data completeness (in resolution range)	99.9 (24.41-1.64) 97.3 (24.41-1.64)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	2.74 (at 1.64Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.151 , 0.185 0.151 , 0.186	Depositor DCC
R_{free} test set	3124 reflections (5.15%)	wwPDB-VP
Wilson B-factor (Å^2)	18.5	Xtriage
Anisotropy	0.218	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 54.4	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3729	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 23.21 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.8751e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: MPD, CL, KCX, BME

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	1.05	1/1752 (0.1%)	0.88	7/2366 (0.3%)
1	B	0.92	0/1767	0.86	6/2382 (0.3%)
All	All	0.98	1/3519 (0.0%)	0.87	13/4748 (0.3%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	147	MET	CG-SD	-5.70	1.66	1.81

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	170	LEU	CB-CG-CD2	8.18	124.91	111.00
1	B	64	ASP	CB-CG-OD2	6.40	124.06	118.30
1	B	170	LEU	CB-CG-CD2	6.33	121.76	111.00
1	A	150	ASP	CB-CG-OD2	6.15	123.84	118.30
1	B	96	ARG	NE-CZ-NH2	-5.97	117.31	120.30
1	A	168	ASP	CB-CG-OD2	5.95	123.65	118.30
1	A	107	ASP	CB-CG-OD2	5.75	123.47	118.30
1	A	64	ASP	CB-CG-OD2	5.74	123.47	118.30
1	B	96	ARG	NE-CZ-NH1	5.67	123.14	120.30
1	B	135	ASP	CB-CG-OD2	5.43	123.19	118.30
1	A	61	LEU	CA-CB-CG	5.34	127.59	115.30
1	B	170	LEU	CA-CB-CG	5.10	127.03	115.30
1	A	3	ASP	CB-CG-OD2	5.01	122.81	118.30

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	1731	0	1706	8	0
1	B	1743	0	1740	11	0
2	A	1	0	0	0	0
3	A	4	0	6	2	0
4	B	16	0	28	4	0
5	A	127	0	0	1	0
5	B	107	0	0	1	0
All	All	3729	0	3480	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 3.

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:B:212:GLN:NE2	4:B:304:MPD:HM1	1.71	1.04
1:B:62:GLU:H	1:B:68:ASN:HD21	1.25	0.82
1:A:62:GLU:H	1:A:68:ASN:HD21	1.32	0.78
1:B:32:GLU:CG	1:B:32:GLU:OE1	2.36	0.73
1:B:212:GLN:HE21	4:B:304:MPD:HM1	1.56	0.64
1:B:32:GLU:CG	1:B:32:GLU:OE2	2.48	0.61
1:B:76:GLU:CD	1:B:78:SER:OG	2.39	0.60
1:B:212:GLN:HE22	4:B:304:MPD:HM1	1.65	0.59
1:B:41:GLN:HG2	5:B:361:HOH:O	2.03	0.58
1:B:62:GLU:H	1:B:68:ASN:ND2	1.98	0.57
3:A:302:BME:H22	5:A:481:HOH:O	2.05	0.56
1:B:212:GLN:NE2	4:B:304:MPD:CM	2.58	0.54
1:A:62:GLU:H	1:A:68:ASN:ND2	2.07	0.47
1:A:46:GLU:CD	1:A:46:GLU:H	2.19	0.46
1:A:154:TYR:CE1	1:A:165:GLU:HG3	2.53	0.44
1:A:46:GLU:CD	1:A:46:GLU:N	2.72	0.43
1:B:149:VAL:HB	1:B:170:LEU:HD13	2.01	0.43
1:A:34:LEU:C	1:A:34:LEU:HD23	2.39	0.42
1:A:157:LEU:HD12	1:A:170:LEU:CD2	2.50	0.42
1:A:45:PHE:CD2	3:A:302:BME:H12	2.55	0.41

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	214/218 (98%)	211 (99%)	3 (1%)	0	100	100
1	B	215/218 (99%)	212 (99%)	3 (1%)	0	100	100
All	All	429/436 (98%)	423 (99%)	6 (1%)	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	176/190 (93%)	171 (97%)	5 (3%)	43	16
1	B	179/190 (94%)	172 (96%)	7 (4%)	32	8
All	All	355/380 (93%)	343 (97%)	12 (3%)	37	11

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	165	GLU
1	A	170	LEU
1	A	213	LYS
1	A	214	LEU
1	A	216	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	2	LEU
1	B	49	LEU
1	B	127	LYS
1	B	147	MET
1	B	157	LEU
1	B	170	LEU
1	B	216	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	27	ASN
1	A	57	ASN
1	A	68	ASN
1	A	179	GLN
1	B	27	ASN
1	B	68	ASN
1	B	179	GLN
1	B	212	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are 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	$\#|Z| > 2$
1	KCX	B	205	1	7,11,12	0.83	0	4,12,14	0.99	0
1	KCX	A	205	1	7,11,12	1.22	0	4,12,14	0.62	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
1	KCX	B	205	1	-	0/7/10/12	-
1	KCX	A	205	1	-	0/7/10/12	-

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.

5.5 Carbohydrates (i)

There are no carbohydrates 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	MPD	B	303	-	7,7,7	0.45	0	9,10,10	0.67	0
4	MPD	B	304	-	7,7,7	0.42	0	9,10,10	0.77	0
3	BME	A	302	-	3,3,3	0.43	0	1,2,2	0.23	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.

Full wwPDB X-ray Structure Validation Report

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	MPD	B	303	-	-	3/5/5/5	-
4	MPD	B	304	-	-	1/5/5/5	-
3	BME	A	302	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	303	MPD	C2-C3-C4-C5
4	B	303	MPD	C2-C3-C4-O4
4	B	303	MPD	O2-C2-C3-C4
4	B	304	MPD	C2-C3-C4-C5

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	304	MPD	4	0
3	A	302	BME	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, 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	216/218 (99%)	-0.30	6 (2%)	53	51	11, 19, 33, 57	13 (6%)
1	B	216/218 (99%)	-0.16	4 (1%)	66	67	13, 22, 38, 43	12 (5%)
All	All	432/436 (99%)	-0.23	10 (2%)	60	60	11, 21, 37, 57	25 (5%)

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	214	LEU	6.1
1	B	19	ASN	4.0
1	A	218	LEU	3.5
1	A	217	LYS	3.5
1	B	218	LEU	3.0
1	A	2	LEU	2.1
1	B	15	ASP	2.1
1	A	19	ASN	2.1
1	B	18	LYS	2.0
1	A	63	ASN	2.0

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, 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(Å^2)	Q<0.9
1	KCX	B	205	12/13	0.94	0.12	16,19,24,24	0
1	KCX	A	205	12/13	0.96	0.11	14,15,20,20	0

6.3 Carbohydrates (i)

There are no carbohydrates 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(Å^2)	Q<0.9
3	BME	A	302	4/4	0.58	0.18	53,54,54,63	0
4	MPD	B	303	8/8	0.82	0.13	58,60,62,63	0
4	MPD	B	304	8/8	0.94	0.13	40,45,46,46	0
2	CL	A	301	1/1	0.99	0.02	31,31,31,31	0

6.5 Other polymers (i)

There are no such residues in this entry.