



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 05:19 pm GMT

PDB ID : 1YFH  
Title : wt Human O6-Alkylguanine-DNA Alkyltransferase Bound To DNA Contain-  
ing an Alkylated Cytosine  
Authors : Duguid, E.M.; Rice, P.A.; He, C.  
Deposited on : 2004-12-31  
Resolution : 3.01 Å(reported)

This is a wwPDB X-ray Structure 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/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

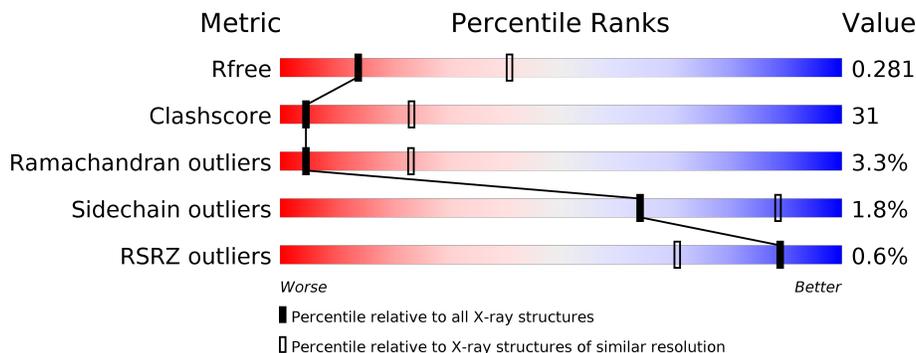
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.01 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))
$R_{free}$	100719	1924 (3.04-3.00)
Clashscore	112137	2279 (3.04-3.00)
Ramachandran outliers	110173	2207 (3.04-3.00)
Sidechain outliers	110143	2210 (3.04-3.00)
RSRZ outliers	101464	1948 (3.04-3.00)

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. 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	D	16	
1	F	16	
2	E	16	
2	G	16	
3	A	179	
3	B	179	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	C	179	 46% 35% 17%

## 2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 4946 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 DNA chain called 5'-D(\*GP\*TP\*GP\*GP\*AP\*TP\*GP\*(XCY)P\*GP\*TP\*GP\*TP\*AP\*GP\*GP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	D	16	Total 343	C 167	N 64	O 97	P 15	0	0	0
1	F	16	Total 343	C 167	N 64	O 97	P 15	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	E	16	Total 315	C 152	N 58	O 90	P 15	0	0	0
2	G	16	Total 315	C 152	N 58	O 90	P 15	0	0	0

- Molecule 3 is a protein called Methylated-DNA--protein-cysteine methyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	165	Total 1252	C 805	N 221	O 218	S 8	0	0	0
3	B	161	Total 1223	C 786	N 216	O 213	S 8	0	0	0
3	C	149	Total 1146	C 737	N 203	O 198	S 8	0	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total 1	Zn 1	0	0
4	A	1	Total 1	Zn 1	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total	Zn	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	O	0	0
			1	1		
5	B	1	Total	O	0	0
			1	1		
5	C	2	Total	O	0	0
			2	2		
5	F	2	Total	O	0	0
			2	2		

### 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.

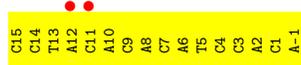
- Molecule 1: 5'-D(\*GP\*TP\*GP\*GP\*AP\*TP\*GP\*(XCY)P\*GP\*TP\*GP\*TP\*AP\*GP\*GP\*T)-3',



- Molecule 1: 5'-D(\*GP\*TP\*GP\*GP\*AP\*TP\*GP\*(XCY)P\*GP\*TP\*GP\*TP\*AP\*GP\*GP\*T)-3',



- Molecule 2: 5'-D(\*CP\*CP\*TP\*AP\*CP\*AP\*CP\*AP\*CP\*AP\*TP\*CP\*CP\*AP\*CP\*A)-3'



L176  
G177  
K178  
P179

• Molecule 3: Methylated-DNA--protein-cysteine methyltransferase

Chain B:  47% 40% 10%

MET ASP LYS ASP C5 E6 W7 K6 R9 T10 T11 L12 D13 S14 L21 L22 G23 C24 G27 L28 H29 G35 LYS GLY THR SER ALA ALA ASP ALA VAL VAL V46 P47 A48 P49 A50 A51 V52 L53 P56 E57 P58 L59 Y69 F70 H71 Q72 P73 E74 A75

L84 H85 H86 P87 V88 F89 E92 S93 F94 T95 R96 Q97 W100 K101 K104 V106 K107 F108 G109 E110 V111 I112 S113 Q116 L117 A118 A119 P124 K125 A126 V130 M134 R135 P144 R147 V148 V149 C150 S151 S152 G153 A154 M157 Y158 S159 G160 G161 P80

E172 R175 LEU GLY LYS PRO

• Molecule 3: Methylated-DNA--protein-cysteine methyltransferase

Chain C:  46% 35% 17%

MET ASP LYS ASP C5 E6 W7 K6 R9 T10 T11 L12 D13 S14 P15 L16 G17 K18 L19 E20 L21 S22 G23 C24 L28 H29 L33 L34 G35 LYS GLY THR SER ALA ALA ASP ALA VAL VAL VAL PRO PRO PRD ALA ALA VAL LEU GLY G55 P56 L59 F70 H71 Q72 P73 E74

A75 I76 E77 A83 L84 H85 H86 P87 V88 F89 Q90 S93 F94 T95 R96 L99 W100 K101 L102 L103 K104 K107 F108 G109 E110 I111 I112 S113 Y114 Q115 Q116 K125 A126 V130 M134 I143 P144 R147 V148 V149 C150 S151 S152 V155 GLY ASN TYR SER GLY G161

L168 E172 R175 L176 G177 LYS PRO

## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	58.15Å 102.71Å 87.94Å 90.00° 106.77° 90.00°	Depositor
Resolution (Å)	29.17 – 3.01 29.16 – 3.01	Depositor EDS
% Data completeness (in resolution range)	88.0 (29.17-3.01) 88.0 (29.16-3.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.98 (at 3.00Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.246 , 0.286 0.247 , 0.281	Depositor DCC
$R_{free}$ test set	834 reflections (4.80%)	DCC
Wilson B-factor (Å <sup>2</sup> )	73.5	Xtrriage
Anisotropy	0.278	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 46.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.031 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	4946	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	76.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.07% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, XCY

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	D	0.57	0/353	0.78	0/544
1	F	0.63	0/353	0.80	0/544
2	E	0.61	0/352	0.74	0/538
2	G	0.63	0/352	0.76	0/538
3	A	0.48	0/1285	0.66	0/1744
3	B	0.47	0/1255	0.66	0/1705
3	C	0.47	0/1174	0.69	0/1590
All	All	0.51	0/5124	0.70	0/7203

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	G	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	G	15	DC	Sidechain

## 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	D	343	0	192	27	0
1	F	343	0	192	27	0
2	E	315	0	180	28	0
2	G	315	0	180	22	0
3	A	1252	0	1268	75	0
3	B	1223	0	1234	79	0
3	C	1146	0	1159	56	0
4	A	1	0	0	0	0
4	B	1	0	0	1	0
4	C	1	0	0	0	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
5	C	2	0	0	1	0
5	F	2	0	0	1	0
All	All	4946	0	4405	291	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 31.

The worst 5 of 291 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)
2:E:14:DC:H2''	2:E:13:DT:H71	1.41	1.03
3:B:101:LYS:HA	3:B:101:LYS:HE3	1.44	0.98
1:D:1:DG:H2''	1:D:2:DT:H5'	1.45	0.96
3:B:7:MET:CE	3:B:24:CYS:HB3	1.97	0.94
3:B:7:MET:HB3	3:B:49:PRO:HB3	1.47	0.93

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	
3	A	161/179 (90%)	138 (86%)	18 (11%)	5 (3%)	5	25
3	B	157/179 (88%)	130 (83%)	21 (13%)	6 (4%)	4	20
3	C	143/179 (80%)	121 (85%)	18 (13%)	4 (3%)	6	27
All	All	461/537 (86%)	389 (84%)	57 (12%)	15 (3%)	4	23

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	72	GLN
3	C	72	GLN
3	A	55	GLY
3	B	153	GLY
3	B	160	GLY

### 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	
3	A	131/141 (93%)	129 (98%)	2 (2%)	70	91
3	B	128/141 (91%)	124 (97%)	4 (3%)	45	80
3	C	121/141 (86%)	120 (99%)	1 (1%)	85	95
All	All	380/423 (90%)	373 (98%)	7 (2%)	64	88

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

Mol	Chain	Res	Type
3	B	79	PHE
3	C	96	ARG
3	B	101	LYS
3	A	179	PRO
3	B	151	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	61	GLN
3	B	90	GLN
3	C	72	GLN
3	A	157	ASN
3	C	61	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](#)

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	XCY	D	8	1	24,30,31	1.46	5 (20%)	30,41,44	1.09	1 (3%)
1	XCY	F	8	1	24,30,31	1.39	5 (20%)	30,41,44	1.07	1 (3%)

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	XCY	D	8	1	-	0/10/28/29	0/3/3/3
1	XCY	F	8	1	-	0/10/28/29	0/3/3/3

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	8	XCY	CD2-CG	2.01	1.43	1.38
1	F	8	XCY	CD2-CG	2.05	1.43	1.38
1	F	8	XCY	CE1-CZ	2.07	1.43	1.38
1	D	8	XCY	CE1-CZ	2.31	1.43	1.38
1	F	8	XCY	CD1-CG	2.33	1.43	1.38

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	8	XCY	CH-N4-C4	5.34	134.31	123.35
1	D	8	XCY	CH-N4-C4	5.49	134.62	123.35

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	F	8	XCY	5	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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.

## 5.7 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	D	15/16 (93%)	0.68	1 (6%) 19 7	65, 89, 101, 105	0
1	F	15/16 (93%)	-0.34	0 100 100	47, 73, 91, 105	0
2	E	16/16 (100%)	0.69	2 (12%) 4 1	66, 94, 101, 102	0
2	G	16/16 (100%)	-0.06	0 100 100	55, 75, 88, 98	0
3	A	165/179 (92%)	-0.05	0 100 100	59, 74, 82, 88	0
3	B	161/179 (89%)	-0.07	0 100 100	66, 76, 85, 93	0
3	C	149/179 (83%)	-0.13	0 100 100	58, 74, 83, 92	0
All	All	537/601 (89%)	-0.04	3 (0%) 89 71	47, 75, 89, 105	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	11	DC	2.9
1	D	14	DG	2.2
2	E	12	DA	2.1

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
1	XCY	F	8	28/29	0.94	0.20	-	77,93,93,93	0
1	XCY	D	8	28/29	0.95	0.16	-	73,73,77,77	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q < 0.9
4	ZN	C	302	1/1	0.97	0.10	-1.69	77,77,77,77	0
4	ZN	B	303	1/1	0.99	0.06	-2.00	77,77,77,77	0
4	ZN	A	301	1/1	0.99	0.06	-2.12	77,77,77,77	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.