

# Full wwPDB X-ray Structure Validation Report (i)

#### Oct 8, 2023 – 10:54 PM EDT

PDB ID	:	6W13
Title	:	Human 8-oxoguanine glycosylase interrogating fully intrahelical oxoG lesion
		DNA
Authors	:	Shigdel, U.; Verdine, G.
Deposited on	:	2020-03-03
Resolution	:	2.38  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 (i) 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 (1)) were used in the production of this report:

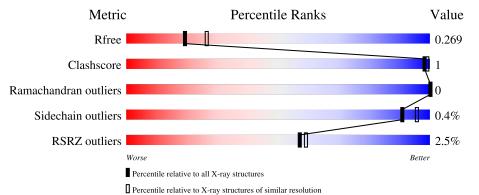
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber $(2001)$
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 2.38 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
$R_{free}$	130704	5509(2.40-2.36)
Clashscore	141614	6082(2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)
RSRZ outliers	127900	5397 (2.40-2.36)

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	А	317	% 97%	
2	В	7	43% 86% 14%	_
3	С	6	50%	-



#### 6W13

# 2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 2870 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 N-glycosylase/DNA lyase.

Mol	Chain	Residues		At	oms			ZeroOcc	AltConf	Trace
1	А	313	Total 2490	C 1579	N 449	0 450	S 12	51	1	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	GLY	-	expression tag	UNP O15527
А	10	SER	-	expression tag	UNP 015527
А	11	GLU	-	expression tag	UNP O15527
А	122	GLN	GLU	engineered mutation	UNP O15527
А	207	CYS	TYR	engineered mutation	UNP 015527
А	253	TRP	CYS	engineered mutation	UNP 015527

• Molecule 2 is a DNA chain called DNA (5'-D(P\*CP\*AP\*(8OG)P\*GP\*TP\*CP\*T)-3').

Mol	Chain	Residues		At	oms			ZeroOcc	AltConf	Trace
2	В	7	Total 144	C 68	N 25	0 44	Р 7	0	0	0

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

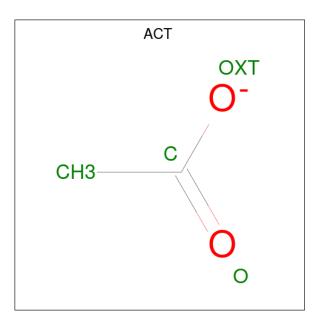
Mol	Chain	Residues		At	$\mathbf{oms}$			ZeroOcc	AltConf	Trace
3	С	6	Total 123	C 58	N 23	O 36	Р 6	0	0	0

• Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	4	Total Mg 4 4	0	0

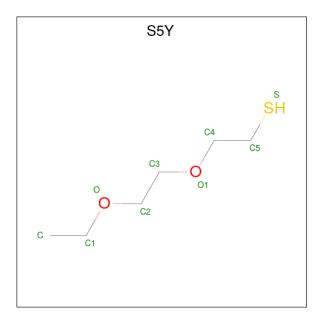
• Molecule 5 is ACETATE ION (three-letter code: ACT) (formula:  $C_2H_3O_2$ ).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
5	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0

• Molecule 6 is 2-(2-ethoxy ethoxy)ethanethiol (three-letter code: S5Y) (formula:  $C_6H_{14}O_2S$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	В	1	Total 16	C 10	0 4	$\frac{S}{2}$	0	1

• Molecule 7 is water.

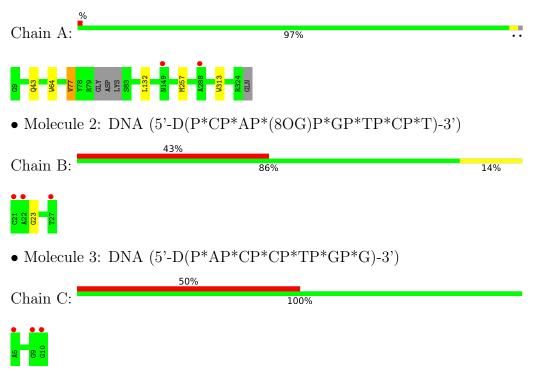


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	А	83	Total         O           83         83	0	0
7	В	2	Total O 2 2	0	0



# 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: N-glycosylase/DNA lyase



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants	90.71Å 90.71Å 210.80Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $120.00^{\circ}$	Depositor
Resolution (Å)	45.35 - 2.38	Depositor
Resolution (A)	45.35 - 2.38	EDS
% Data completeness	99.9 (45.35-2.38)	Depositor
(in resolution range)	99.9 (45.35 - 2.38)	EDS
R <sub>merge</sub>	0.12	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.50 (at 2.37 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
D D.	0.214 , $0.271$	Depositor
$R, R_{free}$	0.218 , $0.269$	DCC
$R_{free}$ test set	1098 reflections $(5.13\%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	42.8	Xtriage
Anisotropy	0.678	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.33, $39.3$	EDS
L-test for twinning <sup>2</sup>	$ \langle L  \rangle = 0.49, \langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2870	wwPDB-VP
Average B, all atoms $(Å^2)$	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.73% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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: ACT, MG, S5Y, 8OG  $\,$ 

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	А	0.46	0/2559	0.66	0/3483	
2	В	0.34	0/133	0.70	0/200	
3	С	0.26	0/137	0.62	0/209	
All	All	0.45	0/2829	0.66	0/3892	

There are no bond length outliers.

There are no bond angle outliers.

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	А	2490	0	2421	3	0
2	В	144	0	79	0	0
3	С	123	0	68	0	0
4	А	4	0	0	0	0
5	А	8	0	6	0	0
6	В	16	0	0	0	0
7	А	83	0	0	0	0
7	В	2	0	0	0	0
All	All	2870	0	2574	3	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 1.

All (3) 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:A:132:LEU:HD12	1:A:313:TRP:CD1	2.37	0.59
1:A:43:GLN:HG2	1:A:257:MET:HE1	1.92	0.51
1:A:64:TRP:CZ3	1:A:77:VAL:HG22	2.50	0.46

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				
1	А	310/317~(98%)	300~(97%)	10 (3%)	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	А	264/266~(99%)	263 (100%)	1 (0%)	91 96	

All (1) residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	А	77	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

#### 5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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).

Т	Mol Type		Chain	Res	Res Link	Bond lengths			Bond angles		
	IVIOI	rybe	Ullalli	nes		Counts	RMSZ	# Z >2	Counts	RMSZ	# Z  > 2
	2	80G	В	23	3,2	22,25,26	2.29	7 (31%)	30,37,40	2.96	11 (36%)

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
2	80G	В	23	3,2	-	0/7/21/22	0/3/3/3

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$\operatorname{Observed}(\operatorname{\AA})$	Ideal(Å)
2	В	23	80G	C5-C4	7.72	1.48	1.37
2	В	23	80G	C4-N9	-3.42	1.32	1.39
2	В	23	80G	C8-N9	-3.22	1.35	1.40
2	В	23	80G	O8-C8	3.17	1.28	1.23
2	В	23	80G	C5-C6	2.27	1.48	1.42
2	В	23	80G	C5-N7	-2.16	1.34	1.37

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	Observed(Å)	Ideal(Å)
2	В	23	80G	C8-N7	-2.03	1.34	1.38

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
2	В	23	80G	N7-C8-N9	7.57	115.56	106.58
2	В	23	80G	C5-C4-N3	-7.01	120.56	127.80
2	В	23	80G	N9-C4-N3	5.69	132.31	125.81
2	В	23	80G	O4'-C1'-N9	4.69	113.00	108.29
2	В	23	80G	C2-N3-C4	4.60	120.50	112.30
2	В	23	80G	C1'-N9-C4	-3.92	120.46	126.54
2	В	23	80G	O8-C8-N7	-3.54	120.23	126.64
2	В	23	80G	C5-N7-C8	-3.41	104.56	109.47
2	В	23	80G	O6-C6-C5	-2.85	120.71	127.24
2	В	23	80G	C6-C5-N7	2.32	136.94	131.96
2	В	23	80G	C2'-C3'-C4'	2.09	107.13	102.76

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 monosaccharides in this entry.

## 5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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	Dec	Dec	Tink	B	ond leng	$\operatorname{gths}$	В	ond ang	gles
				Link	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2		
5	ACT	А	406	-	$3,\!3,\!3$	0.94	0	$3,\!3,\!3$	0.37	0		



Mol	Turne	Chain	Res	Res Link	Bond lengths				Bond angles		
WIOI	Type	Unam			Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z >2	
6	S5Y	В	101[A]	1	8,8,8	0.48	0	7,7,7	0.65	0	
6	S5Y	В	101[B]	1	8,8,8	0.38	0	7,7,7	0.45	0	
5	ACT	А	405	-	$3,\!3,\!3$	0.69	0	3,3,3	1.07	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
6	S5Y	В	101[A]	1	-	4/6/6/6	-
6	S5Y	В	101[B]	1	-	3/6/6/6	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (7) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	В	101[A]	S5Y	O-C2-C3-O1
6	В	101[A]	S5Y	O1-C4-C5-S
6	В	101[A]	S5Y	C-C1-O-C2
6	В	101[B]	S5Y	C5-C4-O1-C3
6	В	101[B]	S5Y	C2-C3-O1-C4
6	В	101[B]	S5Y	O-C2-C3-O1
6	В	101[A]	S5Y	C3-C2-O-C1

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,  $95^{th}$  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	$\mathbf{OWAB}(\mathbf{A}^2)$	Q<0.9
1	А	313/317~(98%)	0.03	2 (0%) 89 89	32, 45, 72, 109	16 (5%)
2	В	6/7~(85%)	2.31	3 (50%) 0 0	81, 97, 129, 134	0
3	С	6/6 (100%)	2.80	3 (50%) 0 0	101, 112, 129, 130	0
All	All	325/330~(98%)	0.12	8 (2%) 57 59	32, 46, 86, 134	16 (4%)

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	С	10	DG	5.7
3	С	9	DG	4.8
2	В	21	DC	4.5
2	В	27	DT	3.5
1	А	149	ASN	3.1
1	А	288	ALA	2.8
2	В	22	DA	2.7
3	С	5	DA	2.3

## 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,  $95^{th}$  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
	2	80G	В	23	23/24	0.93	0.25	66,79,96,97	0



## 6.3 Carbohydrates (i)

There are no monosaccharides 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,  $95^{th}$  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 <sup>2</sup> )	Q<0.9
5	ACT	А	406	4/4	0.70	0.23	$50,\!51,\!63,\!74$	0
6	S5Y	В	101[A]	9/9	0.77	0.21	66,95,103,104	7
6	S5Y	В	101[B]	9/9	0.77	0.21	73,91,104,104	7
5	ACT	А	405	4/4	0.82	0.15	$60,\!63,\!66,\!83$	0
4	MG	А	401	1/1	0.94	0.20	49,49,49,49	0
4	MG	А	403	1/1	0.95	0.30	56, 56, 56, 56	0
4	MG	А	402	1/1	0.98	0.18	$50,\!50,\!50,\!50$	0
4	MG	А	404	1/1	0.99	0.28	$55,\!55,\!55,\!55$	0

## 6.5 Other polymers (i)

There are no such residues in this entry.

