



# wwPDB X-ray Structure Validation Summary Report

Mar 1, 2014 – 01:27 AM GMT

PDB ID : 1F4R  
Title : CRYSTAL STRUCTURE OF THE HUMAN AAG DNA REPAIR GLYCO-  
SYLASE COMPLEXED WITH 1,N6-ETHENOADENINE-DNA  
Authors : Lau, A.Y.; Wyatt, M.D.; Glassner, B.J.; Samson, L.D.; Ellenberger, T.  
Deposited on : 2000-06-08  
Resolution : 2.40 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>

---

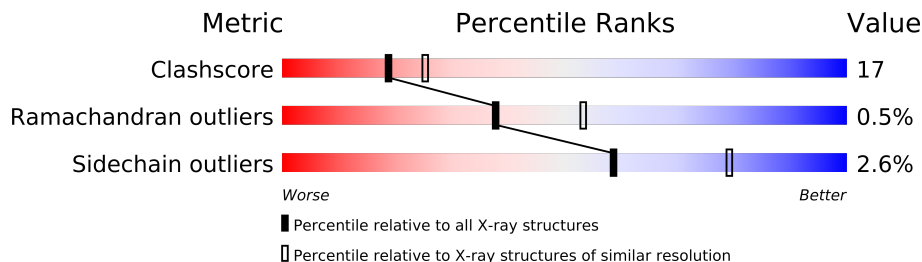
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 21963  
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 2.40 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))
Clashscore	79885	2789 (2.40-2.40)
Ramachandran outliers	78287	2736 (2.40-2.40)
Sidechain outliers	78261	2737 (2.40-2.40)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	D	13	
2	E	13	
3	A	219	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 2064 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 DNA chain called DNA (5'-D(\*GP\*AP\*CP\*AP\*TP\*GP\*(EDA)P\*TP\*TP\*GP\*CP\*CP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	12	Total	C	N	O	P	0	0	0
			245	119	45	70	11			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	12	Total	C	N	O	P	0	0	0
			245	117	45	71	12			

- Molecule 3 is a protein called 3-METHYL-ADENINE DNA GLYCOSYLASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	199	Total	C	N	O	S	0	0	0
			1525	963	284	269	9			

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Na	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	39	Total	O	0	0
			39	39		
5	D	5	Total	O	0	0
			5	5		
5	E	4	Total	O	0	0
			4	4		

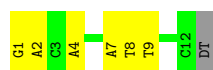
### 3 Residue-property plots

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.

Note EDS was not executed.

- Molecule 1: DNA (5'-D(\*GP\*AP\*CP\*AP\*TP\*GP\*(EDA)P\*TP\*TP\*GP\*CP\*CP\*T)-3')

Chain D: 



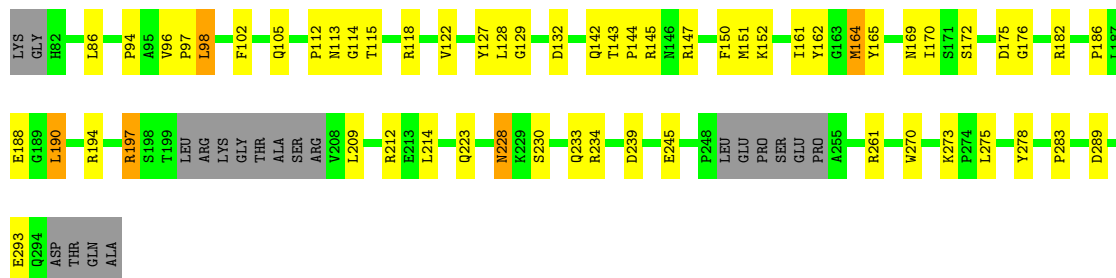
- Molecule 2: DNA (5'-D(\*GP\*GP\*CP\*AP\*AP\*TP\*CP\*AP\*TP\*GP\*TP\*CP\*A)-3')

Chain E: 



- Molecule 3: 3-METHYL-ADENINE DNA GLYCOSYLASE

Chain A: 



## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	42.10Å 57.30Å 125.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	500.00 – 2.40	Depositor
% Data completeness (in resolution range)	99.4 (500.00-2.40)	Depositor
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS	Depositor
R, $R_{free}$	0.239 , 0.276	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2064	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

## 5 Model quality

### 5.1 Standard geometry

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

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	0.32	0/247	0.71	0/377
2	E	0.26	0/274	0.70	0/420
3	A	0.36	0/1554	0.64	0/2101
All	All	0.34	0/2075	0.66	0/2898

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
3	A	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
3	A	127	TYR	Sidechain

### 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	D	245	0	137	7	0
2	E	245	0	136	12	0
3	A	1525	0	1511	56	0
4	A	1	0	0	0	0
5	A	39	0	0	0	0
5	D	5	0	0	0	0
5	E	4	0	0	0	0
All	All	2064	0	1784	63	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 17.

The worst 5 of 63 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:A:143:THR:HB	3:A:144:PRO:HD2	1.53	0.89
3:A:86:LEU:HD11	3:A:105:GLN:HE21	1.43	0.83
3:A:197:ARG:HH11	3:A:197:ARG:HG2	1.45	0.80
3:A:212:ARG:HD2	3:A:283:PRO:O	1.86	0.74
1:D:1:DG:H2"	1:D:2:DA:C8	2.23	0.73

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
3	A	193/219 (88%)	186 (96%)	6 (3%)	1 (0%)	38 53

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	164	MET

### 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
3	A	152/177 (86%)	148 (97%)	4 (3%)	59 79

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	98	LEU
3	A	190	LEU
3	A	197	ARG
3	A	228	ASN

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

Mol	Chain	Res	Type
3	A	105	GLN
3	A	195	GLN
3	A	228	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

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).



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
1	EDA	D	7	1	24,26,27	0.93	2 (8%)	32,38,41	1.49	7 (21%)

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	EDA	D	7	1	-	0/6/21/22	0/1/4/4

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	7	EDA	C10-N1	2.48	1.41	1.37
1	D	7	EDA	C6-N1	-2.11	1.37	1.42

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	7	EDA	C10-N1-C2	4.24	137.02	130.65
1	D	7	EDA	N6-C6-N1	3.37	117.66	112.70
1	D	7	EDA	C11-N6-C6	-3.05	100.11	103.95
1	D	7	EDA	C8-N9-C4	-2.89	104.69	106.90
1	D	7	EDA	C2-N1-C6	-2.44	120.69	122.22

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is 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.

## 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 ⓘ

EDS was not executed - this section will therefore be empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

EDS was not executed - this section will therefore be empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section will therefore be empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section will therefore be empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section will therefore be empty.