



# Full wwPDB X-ray Structure Validation Report ⓘ

May 18, 2020 – 03:43 pm BST

PDB ID : 1EMJ  
Title : URACIL-DNA GLYCOSYLASE BOUND TO DNA CONTAINING A 4'-TH IO-2'DEOXYURIDINE ANALOG PRODUCT  
Authors : Parikh, S.S.; Walcher, G.; Jones, G.D.; Slupphaug, G.; Krokan, H.E.; Blackburn, G.M.; Tainer, J.A.  
Deposited on : 2000-03-16  
Resolution : 2.00 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](mailto: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 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)  
Xtrriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

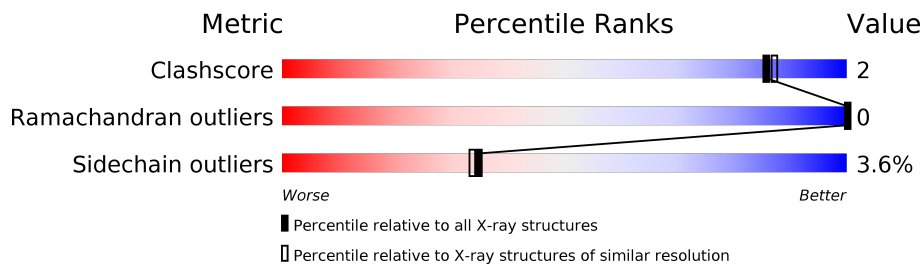
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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 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  $\leq 5\%$ .

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	B	9	
2	C	10	
3	A	223	

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 2353 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 DNA (5'-D(\*TP\*GP\*TP\*(ASU)P\*AP\*TP\*CP\*TP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	N	O	P				S
1	B	9	171	84	23	55	8	1	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	10	205	99	45	52	9	0	0	0

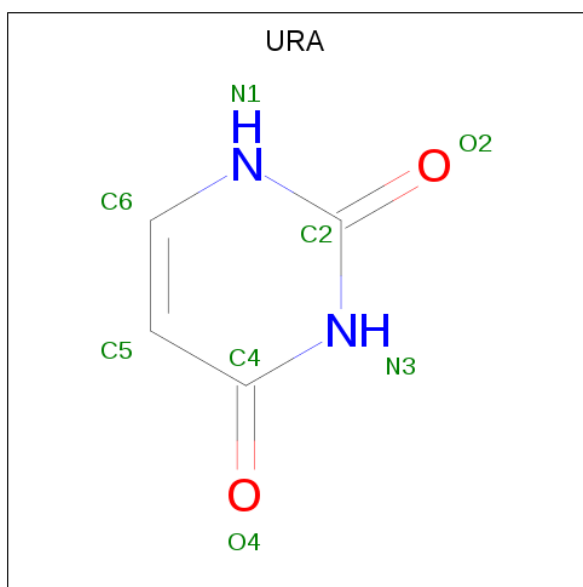
- Molecule 3 is a protein called URACIL-DNA GLYCOSYLASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	223	1808	1168	318	316	6	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	82	MET	PRO	CONFLICT	UNP P13051
A	83	GLU	VAL	CONFLICT	UNP P13051
A	84	PHE	GLY	CONFLICT	UNP P13051

- Molecule 4 is URACIL (three-letter code: URA) (formula: C<sub>4</sub>H<sub>4</sub>N<sub>2</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
4	A	1	8	4	2	2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	12	Total	O	0	0
			12	12		
5	C	5	Total	O	0	0
			5	5		
5	A	144	Total	O	0	0
			144	144		

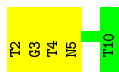
### 3 Residue-property plots [i](#)

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. 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. 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(\*TP\*GP\*TP\*(ASU)P\*AP\*TP\*CP\*TP\*T)-3')

Chain B: 



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

Chain C: 

There are no outlier residues recorded for this chain.

- Molecule 3: URACIL-DNA GLYCOSYLASE

Chain A: 



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	48.55Å 64.85Å 96.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00	Depositor
% Data completeness (in resolution range)	99.9 (20.00-2.00)	Depositor
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS	Depositor
R, $R_{free}$	0.228 , 0.257	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2353	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	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: URA, ASU

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	B	0.30	0/175	0.65	0/266
2	C	0.26	0/232	0.63	0/356
3	A	0.63	0/1869	0.69	0/2533
All	All	0.58	0/2276	0.68	0/3155

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

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	B	171	0	103	3	0
2	C	205	0	113	0	0
3	A	1808	0	1762	6	0
4	A	8	0	3	1	0
5	A	144	0	0	1	0
5	B	12	0	0	0	0
5	C	5	0	0	0	0
All	All	2353	0	1981	9	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 (9) 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)
3:A:83:GLU:HG3	3:A:84:PHE:H	1.59	0.67
3:A:250:GLN:HG2	3:A:265:GLN:HE21	1.60	0.67
1:B:5:ASU:H2'	4:A:55:URA:N1	2.15	0.62
3:A:245:TRP:HE1	3:A:283:HIS:HD2	1.48	0.61
1:B:2:DT:H2''	1:B:3:DG:O5'	2.03	0.58
3:A:234:ASN:O	3:A:260:ARG:NH1	2.47	0.48
3:A:83:GLU:HA	3:A:83:GLU:OE2	2.15	0.46
1:B:3:DG:H2''	1:B:4:DT:H71	1.99	0.44
3:A:276:ARG:NH1	5:A:570:HOH:O	2.50	0.44

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	221/223 (99%)	216 (98%)	5 (2%)	0	<b>100</b>   <b>100</b>

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
3	A	195/195 (100%)	188 (96%)	7 (4%)	35   34



All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	83	GLU
3	A	94	SER
3	A	168	PRO
3	A	215	ASN
3	A	235	GLN
3	A	283	HIS
3	A	291	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	92	HIS
3	A	124	GLN
3	A	161	GLN
3	A	265	GLN
3	A	283	HIS

### 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).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
1	ASU	B	5	1	7,12,13	0.91	0	9,16,19	1.26	2 (22%)

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	ASU	B	5	1	-	0/3/17/18	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	5	ASU	C1'-C2'-C3'	2.35	108.60	102.61
1	B	5	ASU	C2'-C1'-S4'	-2.27	103.22	106.23

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	5	ASU	1	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

1 ligand 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
4	URA	A	55	-	6,8,8	1.73	2 (33%)	4,10,10	10.92	3 (75%)

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
4	URA	A	55	-	-	-	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	55	URA	C4-N3	3.37	1.38	1.33
4	A	55	URA	C5-C6	-2.16	1.34	1.38

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
4	A	55	URA	N1-C2-N3	-16.85	115.04	128.43
4	A	55	URA	C6-N1-C2	13.27	120.97	114.42
4	A	55	URA	C5-C4-N3	-4.02	114.47	123.31

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	55	URA	1	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

### 6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.