



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 06:30 pm GMT

PDB ID : 1UUG  
Title : ESCHERICHIA COLI URACIL-DNA GLYCOSYLASE:INHIBITOR COM-  
PLEX WITH WILD-TYPE UDG AND WILD-TYPE UGI  
Authors : Mol, C.D.; Arvai, A.S.; Putnam, C.D.; Tainer, J.A.  
Deposited on : 1998-10-31  
Resolution : 2.40 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

<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
Xtrriage (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

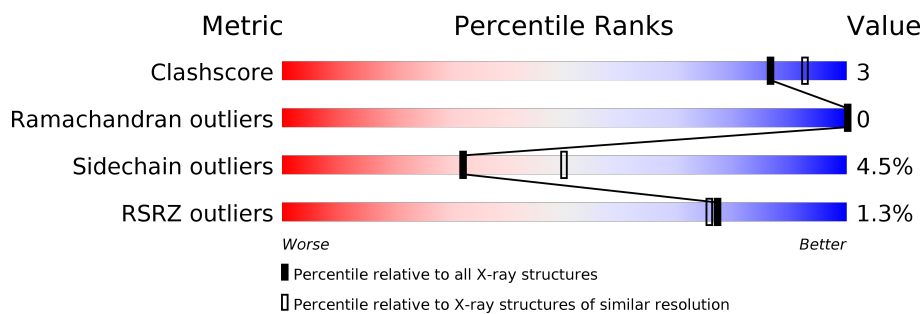
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 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	112137	3674 (2.40-2.40)
Ramachandran outliers	110173	3616 (2.40-2.40)
Sidechain outliers	110143	3617 (2.40-2.40)
RSRZ outliers	101464	3195 (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. 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	A	229	<div> <div>%</div> <div> <div></div> <div>88%</div> <div>9%</div> <div>.</div> </div> </div>
1	C	229	<div> <div>%</div> <div> <div></div> <div>88%</div> <div>7%</div> <div>..</div> </div> </div>
2	B	84	<div> <div>%</div> <div> <div></div> <div>89%</div> <div>8%</div> <div>.</div> </div> </div>
2	D	84	<div> <div>%</div> <div> <div></div> <div>89%</div> <div>8%</div> <div>.</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6412 atoms, of which 1403 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 URACIL-DNA GLYCOSYLASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	222	Total	C	H	N	O	S	0	0	0
			2154	1138	388	316	309	3			
1	C	221	Total	C	H	N	O	S	0	0	0
			2148	1135	387	315	308	3			

- Molecule 2 is a protein called URACIL-DNA GLYCOSYLASE INHIBITOR.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	82	Total	C	H	N	O	S	0	0	0
			773	406	126	99	139	3			
2	D	82	Total	C	H	N	O	S	0	0	0
			773	406	126	99	139	3			

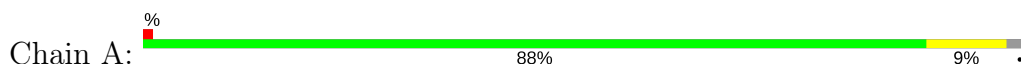
- Molecule 3 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	72	Total	H	O	0	0
			216	144	72		
3	B	25	Total	H	O	0	0
			75	50	25		
3	C	75	Total	H	O	0	0
			225	150	75		
3	D	16	Total	H	O	0	0
			48	32	16		

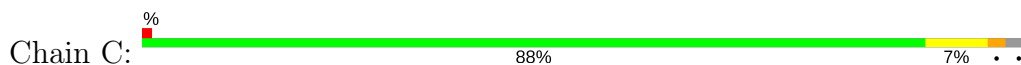
### 3 Residue-property plots [i](#)

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.

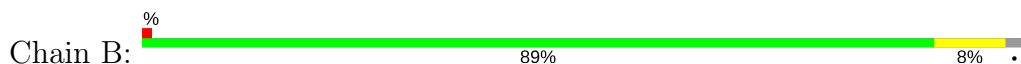
- Molecule 1: URACIL-DNA GLYCOSYLASE



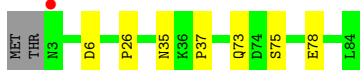
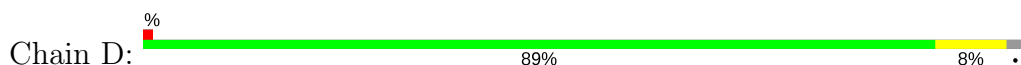
- Molecule 1: URACIL-DNA GLYCOSYLASE



- Molecule 2: URACIL-DNA GLYCOSYLASE INHIBITOR



- Molecule 2: URACIL-DNA GLYCOSYLASE INHIBITOR



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.99Å 86.71Å 108.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.40 19.95 – 2.42	Depositor EDS
% Data completeness (in resolution range)	97.7 (20.00-2.40) 98.2 (19.95-2.42)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.78 (at 2.41Å)	Xtriage
Refinement program	X-PLOR 3.851	Depositor
R, $R_{free}$	0.190 , 0.232 0.189 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	32.4	Xtriage
Anisotropy	0.062	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 55.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6412	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.58% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.39	0/1822	0.61	0/2485
1	C	0.38	0/1817	0.61	0/2478
2	B	0.37	0/655	0.61	0/887
2	D	0.36	0/655	0.64	0/887
All	All	0.38	0/4949	0.61	0/6737

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	A	1766	388	1736	9	1
1	C	1761	387	1731	11	0
2	B	647	126	640	2	0
2	D	647	126	640	2	0
3	A	72	144	0	0	1
3	B	25	50	0	1	0
3	C	75	150	0	0	0
3	D	16	32	0	0	0
All	All	5009	1403	4747	24	1

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 (24) 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:C:117:GLN:HE22	1:C:219:ASP:H	1.38	0.72
1:C:107:ASN:HD22	1:C:107:ASN:H	1.51	0.56
1:A:105:ARG:HG3	1:A:105:ARG:HH11	1.71	0.55
1:A:14:GLU:HG3	1:A:147:LYS:HG2	1.89	0.54
1:C:97:GLU:OE2	1:C:105:ARG:HD3	2.09	0.53
1:C:73:HIS:HE1	1:C:78:SER:OG	1.92	0.53
2:B:39:SER:HB2	3:B:897:HOH:O	2.09	0.52
1:A:30:GLU:OE1	1:A:129:ARG:HD3	2.12	0.49
1:C:107:ASN:HD22	1:C:107:ASN:N	2.12	0.48
1:C:23:THR:O	1:C:26:THR:HG22	2.13	0.48
2:B:29:VAL:HG21	2:B:41:ILE:HG12	1.96	0.46
2:D:26:PRO:HA	2:D:37:PRO:HG2	1.97	0.45
1:C:107:ASN:ND2	1:C:107:ASN:H	2.15	0.45
1:C:154:GLN:HA	1:C:179:ARG:HH12	1.82	0.45
2:D:73:GLN:HA	2:D:78:GLU:O	2.18	0.44
1:A:105:ARG:HA	1:A:106:PRO:HD3	1.68	0.44
1:A:186:PRO:HG2	1:A:195:ARG:O	2.18	0.43
1:A:5:LEU:HD23	1:A:5:LEU:HA	1.76	0.42
1:C:224:LEU:HA	1:C:224:LEU:HD12	1.92	0.42
1:A:100:ILE:HA	1:A:101:PRO:HD3	1.81	0.42
1:A:224:LEU:HD12	1:A:224:LEU:HA	1.93	0.41
1:C:32:GLN:HE21	1:C:32:GLN:HB3	1.70	0.41
1:A:216:THR:HA	1:A:217:PRO:HD3	1.94	0.40
1:C:153:ASN:O	1:C:179:ARG:NH1	2.55	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:42:LYS:HZ1	3:A:891:HOH:H1[4_456]	1.20	0.40

## 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	
1	A	220/229 (96%)	214 (97%)	6 (3%)	0	100	100
1	C	219/229 (96%)	213 (97%)	6 (3%)	0	100	100
2	B	80/84 (95%)	80 (100%)	0	0	100	100
2	D	80/84 (95%)	80 (100%)	0	0	100	100
All	All	599/626 (96%)	587 (98%)	12 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	
1	A	188/194 (97%)	181 (96%)	7 (4%)	39	59
1	C	188/194 (97%)	178 (95%)	10 (5%)	26	42
2	B	76/78 (97%)	72 (95%)	4 (5%)	26	42
2	D	76/78 (97%)	73 (96%)	3 (4%)	37	56
All	All	528/544 (97%)	504 (96%)	24 (4%)	32	50

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	13	GLU
1	A	16	GLN
1	A	53	LEU
1	A	107	ASN
1	A	157	GLU
1	A	195	ARG
1	A	205	LEU
2	B	4	LEU
2	B	5	SER
2	B	75	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	79	ASN
1	C	5	LEU
1	C	19	TYR
1	C	21	LEU
1	C	32	GLN
1	C	53	LEU
1	C	105	ARG
1	C	107	ASN
1	C	120	LEU
1	C	205	LEU
1	C	224	LEU
2	D	6	ASP
2	D	35	ASN
2	D	75	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (14) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	22	ASN
1	A	169	GLN
1	A	208	GLN
2	B	3	ASN
2	B	76	ASN
1	C	22	ASN
1	C	32	GLN
1	C	41	GLN
1	C	73	HIS
1	C	107	ASN
1	C	117	GLN
1	C	169	GLN
1	C	181	HIS
1	C	194	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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	A	222/229 (96%)	-0.51	3 (1%) 75 74	12, 21, 41, 67	0
1	C	221/229 (96%)	-0.37	3 (1%) 75 74	16, 27, 50, 60	0
2	B	82/84 (97%)	-0.33	1 (1%) 79 77	13, 28, 47, 62	0
2	D	82/84 (97%)	-0.23	1 (1%) 79 77	16, 31, 53, 66	0
All	All	607/626 (96%)	-0.40	8 (1%) 77 75	12, 25, 50, 67	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	226	ALA	4.2
1	C	225	PRO	2.7
1	A	8	HIS	2.3
1	A	16	GLN	2.3
2	D	3	ASN	2.2
1	C	5	LEU	2.2
1	C	178	GLN	2.1
2	B	15	GLN	2.0

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.