Full wwPDB X-ray Structure Validation Report ⓘ

Jan 31, 2016 – 07:16 PM GMT

PDB ID	:	1EUG
Title	:	CRYSTAL STRUCTURE OF ESCHERICHIA COLI URACIL DNA GLY-COSYLASE AND ITS COMPLEXES WITH URACIL AND GLYCEROL: STRUCTURE AND GLYCOSYLASE MECHANISM REVISITED
Authors	:	Xiao, G.; Tordova, M.; Jagadeesh, J.; Drohat, A.C.; Stivers, J.T.; Gilliland, G.L.
Deposited on	:	1998-10-12
Resolution	:	1.60 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.60 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	Percentile Ranks	Value
Clashscore		7
Ramachandran outliers		0
Sidechain outliers		3.7%
RSRZ outliers		37.3%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
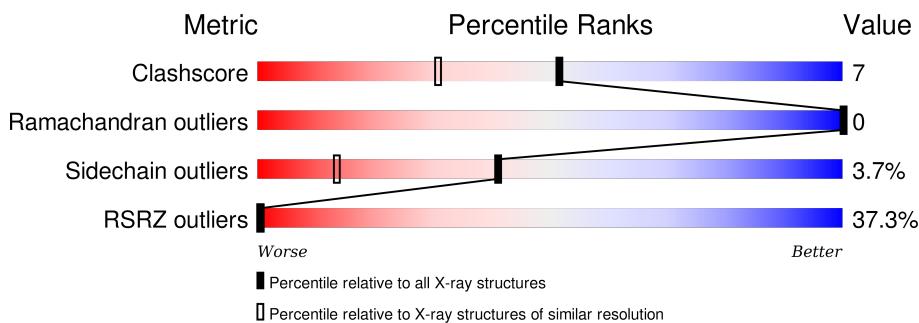
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	102246	2732 (1.60-1.60)
Ramachandran outliers	100387	2654 (1.60-1.60)
Sidechain outliers	100360	2653 (1.60-1.60)
RSRZ outliers	91569	2479 (1.60-1.60)

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 >=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 <=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	37% / 83% / 14%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2090 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 PROTEIN (GLYCOSYLASE).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	225	Total 1790	C 1151	N 318	O 318	S 3	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	213	HIS	ARG	MUTATION	UNP P12295

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	300	Total 300	O 300	0	0

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.

• Molecule 1: PROTEIN (GLYCOSYLASE)
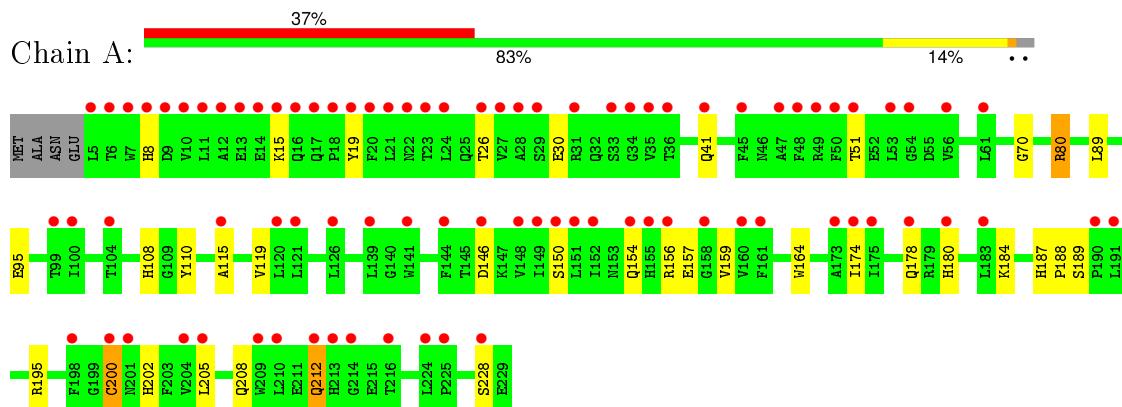
4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	55.13Å 61.32Å 64.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	99.00 – 1.60 44.41 – 1.44	Depositor EDS
% Data completeness (in resolution range)	92.5 (99.00-1.60) 71.6 (44.41-1.44)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.86 (at 1.44Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, R_{free}	0.194 , 0.250 0.322 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	0.1	Xtriage
Anisotropy	7.013	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 65.2	EDS
Estimated twinning fraction	0.025 for -h,l,k	Xtriage
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Outliers	3 of 32709 reflections (0.009%)	Xtriage
F_o,F_c correlation	0.67	EDS
Total number of atoms	2090	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.34% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.42	0/1847	1.10	8/2518 (0.3%)

There are no bond length outliers.

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	80	ARG	NE-CZ-NH2	-12.86	113.87	120.30
1	A	80	ARG	NE-CZ-NH1	12.05	126.32	120.30
1	A	80	ARG	CD-NE-CZ	11.17	139.24	123.60
1	A	200	CYS	C-N-CA	-9.79	97.22	121.70
1	A	195	ARG	NE-CZ-NH2	-7.67	116.47	120.30
1	A	200	CYS	O-C-N	-6.54	112.24	122.70
1	A	146	ASP	CB-CG-OD1	5.30	123.07	118.30
1	A	19	TYR	CB-CG-CD2	5.14	124.09	121.00

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	1790	0	1747	24	0
2	A	300	0	0	3	0
All	All	2090	0	1747	24	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 7.

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:A:164:TRP:HE1	1:A:202:HIS:HD2	1.31	0.78
1:A:212:GLN:N	1:A:212:GLN:HE21	1.82	0.77
1:A:159:VAL:H	1:A:180:HIS:HD2	1.31	0.75
1:A:108:HIS:HD2	1:A:110:TYR:H	1.38	0.71
1:A:212:GLN:HE21	1:A:212:GLN:CA	2.05	0.68
1:A:164:TRP:HE1	1:A:202:HIS:CD2	2.15	0.64
1:A:150:SER:O	1:A:154:GLN:HG3	1.99	0.62
1:A:208:GLN:O	1:A:212:GLN:NE2	2.42	0.52
1:A:70:GLY:O	1:A:80:ARG:HG3	2.12	0.50
1:A:108:HIS:CD2	1:A:110:TYR:H	2.23	0.50
1:A:174:ILE:HG12	2:A:416:HOH:O	2.12	0.48
1:A:184:LYS:HB2	1:A:184:LYS:HE3	1.51	0.48
1:A:159:VAL:H	1:A:180:HIS:CD2	2.20	0.48
1:A:8:HIS:HD2	1:A:15:LYS:NZ	2.13	0.47
1:A:212:GLN:NE2	1:A:212:GLN:CA	2.77	0.46
1:A:156:ARG:NH2	2:A:399:HOH:O	2.48	0.46
1:A:95:GLU:OE2	1:A:200:CYS:O	2.34	0.45
1:A:156:ARG:NH1	2:A:468:HOH:O	2.49	0.45
1:A:187:HIS:CG	1:A:188:PRO:HD2	2.52	0.45
1:A:156:ARG:NH1	1:A:157:GLU:H	2.18	0.42
1:A:26:THR:O	1:A:30:GLU:HG3	2.20	0.41
1:A:159:VAL:N	1:A:180:HIS:HD2	2.07	0.41
1:A:115:ALA:HA	1:A:119:VAL:O	2.21	0.40
1:A:205:LEU:HA	1:A:205:LEU:HD23	1.97	0.40

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	
1	A	223/229 (97%)	215 (96%)	8 (4%)	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	A	191/194 (98%)	184 (96%)	7 (4%)	41	13

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

Mol	Chain	Res	Type
1	A	41	GLN
1	A	51	THR
1	A	89	LEU
1	A	178	GLN
1	A	189	SER
1	A	212	GLN
1	A	228	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	8	HIS
1	A	22	ASN
1	A	108	HIS
1	A	167	HIS
1	A	178	GLN
1	A	180	HIS
1	A	202	HIS
1	A	212	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)

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, 95th 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^2)	Q<0.9
1	A	225/229 (98%)	1.94	84 (37%)	0	0	5, 11, 26, 38	0

All (84) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	5	LEU	7.7
1	A	21	LEU	5.7
1	A	24	LEU	5.7
1	A	10	VAL	5.4
1	A	173	ALA	4.9
1	A	17	GLN	4.8
1	A	20	PHE	4.5
1	A	141	TRP	4.3
1	A	7	TRP	4.2
1	A	11	LEU	4.1
1	A	174	ILE	4.1
1	A	12	ALA	4.0
1	A	35	VAL	4.0
1	A	19	TYR	3.8
1	A	53	LEU	3.8
1	A	151	LEU	3.8
1	A	212	GLN	3.8
1	A	191	LEU	3.7
1	A	36	THR	3.6
1	A	154	GLN	3.6
1	A	6	THR	3.5
1	A	175	ILE	3.5
1	A	45	PHE	3.5
1	A	9	ASP	3.4
1	A	56	VAL	3.4
1	A	15	LYS	3.3
1	A	18	PRO	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	139	LEU	3.2
1	A	144	PHE	3.1
1	A	200	CYS	3.1
1	A	16	GLN	3.1
1	A	51	THR	3.1
1	A	155	HIS	3.1
1	A	161	PHE	3.1
1	A	148	VAL	3.0
1	A	13	GLU	3.0
1	A	33	SER	3.0
1	A	156	ARG	3.0
1	A	160	VAL	3.0
1	A	23	THR	3.0
1	A	183	LEU	2.9
1	A	54	GLY	2.9
1	A	178	GLN	2.9
1	A	8	HIS	2.9
1	A	41	GLN	2.8
1	A	204	VAL	2.8
1	A	28	ALA	2.8
1	A	213	HIS	2.8
1	A	214	GLY	2.8
1	A	126	LEU	2.7
1	A	48	PHE	2.7
1	A	201	ASN	2.7
1	A	225	PRO	2.6
1	A	152	ILE	2.6
1	A	27	VAL	2.6
1	A	149	ILE	2.6
1	A	216	THR	2.6
1	A	50	PHE	2.6
1	A	100	ILE	2.6
1	A	31	ARG	2.5
1	A	228	SER	2.5
1	A	99	THR	2.5
1	A	198	PHE	2.5
1	A	120	LEU	2.4
1	A	26	THR	2.4
1	A	34	GLY	2.4
1	A	190	PRO	2.4
1	A	158	GLY	2.3
1	A	224	LEU	2.3

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	14	GLU	2.3
1	A	104	THR	2.2
1	A	209	TRP	2.2
1	A	180	HIS	2.2
1	A	47	ALA	2.2
1	A	115	ALA	2.2
1	A	150	SER	2.2
1	A	210	LEU	2.1
1	A	146	ASP	2.1
1	A	61	LEU	2.1
1	A	121	LEU	2.1
1	A	29	SER	2.1
1	A	22	ASN	2.0
1	A	49	ARG	2.0
1	A	205	LEU	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.