



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 11:39 AM UTC

PDB ID : 3UFJ / pdb_00003ufj
Title : Human Thymine DNA Glycosylase Bound to Substrate Analog 2'-fluoro-2'-deoxyuridine
Authors : Pozharski, E.; Maiti, A.; Drohat, A.C.
Deposited on : 2011-11-01
Resolution : 2.97 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

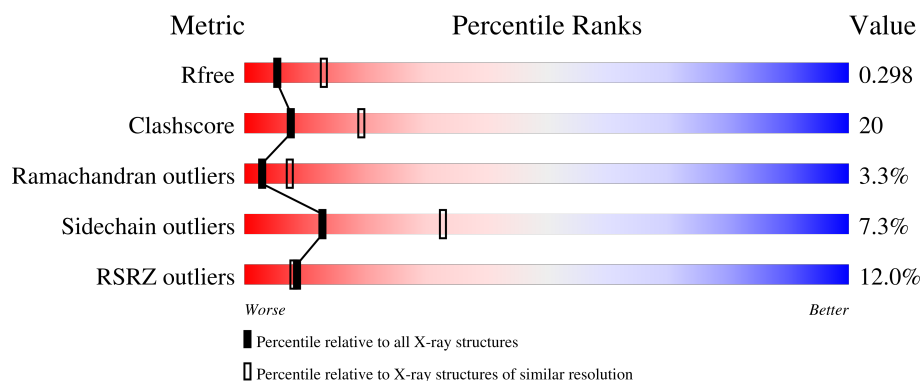
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.97 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))
R_{free}	180053	1130 (2.98-2.94)
Clashscore	190562	1157 (2.98-2.94)
Ramachandran outliers	187476	1101 (2.98-2.94)
Sidechain outliers	187428	1101 (2.98-2.94)
RSRZ outliers	180081	1130 (2.98-2.94)

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 $\leq 5\%$. 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	204	5% 60% 25% 6% 9%
1	B	204	6% 63% 21% 5% 11%
2	C	23	48% 30% 70%
2	E	23	48% 87% 13%
3	D	23	17% 26% 70% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	23	 30% 78% 17% .

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4509 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 G/T mismatch-specific thymine DNA glycosylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	185	Total	C	N	O	S	0	1	0
			1367	875	234	249	9			
1	B	182	Total	C	N	O	S	0	0	0
			1265	809	218	231	7			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	105	GLY	-	expression tag	UNP Q13569
A	106	SER	-	expression tag	UNP Q13569
A	107	HIS	-	expression tag	UNP Q13569
A	108	MET	-	expression tag	UNP Q13569
A	109	ALA	-	expression tag	UNP Q13569
A	110	SER	-	expression tag	UNP Q13569
B	105	GLY	-	expression tag	UNP Q13569
B	106	SER	-	expression tag	UNP Q13569
B	107	HIS	-	expression tag	UNP Q13569
B	108	MET	-	expression tag	UNP Q13569
B	109	ALA	-	expression tag	UNP Q13569
B	110	SER	-	expression tag	UNP Q13569

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	23	Total	C	N	O	P	0	0	0
			473	225	90	136	22			
2	E	23	Total	C	N	O	P	0	0	0
			473	225	90	136	22			

- Molecule 3 is a DNA chain called 5'-D(*CP*CP*AP*CP*TP*GP*CP*TP*CP*AP*(UF2)P*GP*TP*AP*CP*AP*GP*AP*GP*CP*TP*GP*T)-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	D	23	Total 465	C 222	F 1	N 83	O 137	P 22	0	0	0
3	F	23	Total 465	C 222	F 1	N 83	O 137	P 22	0	0	0

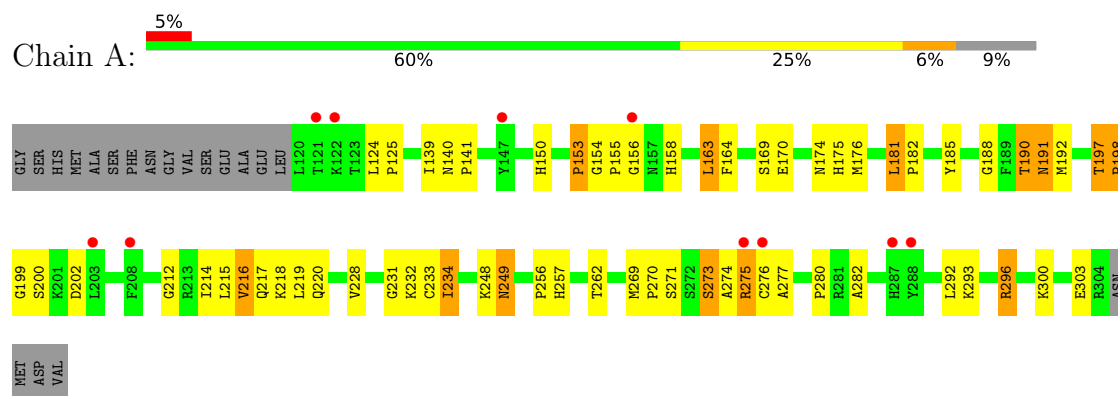
- Molecule 4 is water.

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

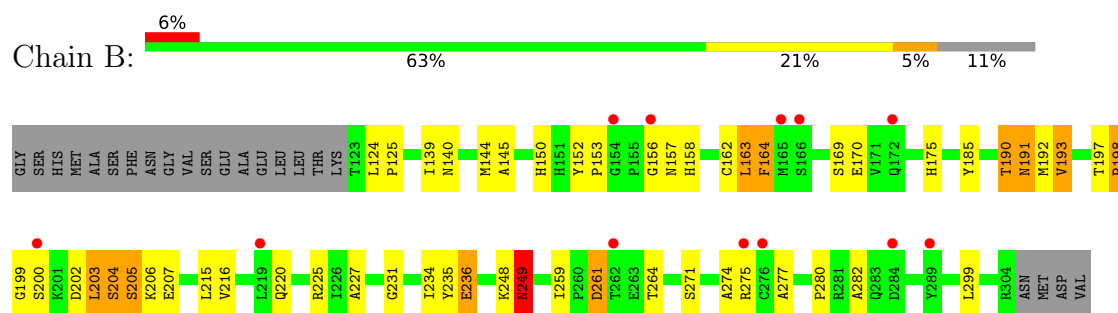
3 Residue-property plots

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.

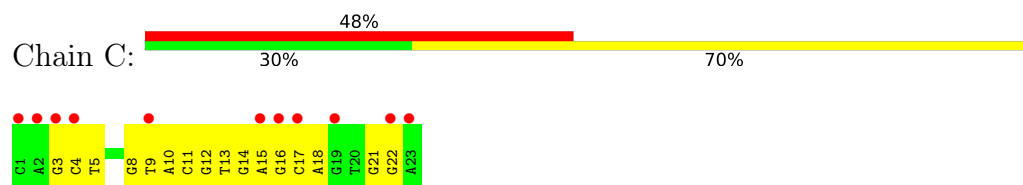
- Molecule 1: G/T mismatch-specific thymine DNA glycosylase



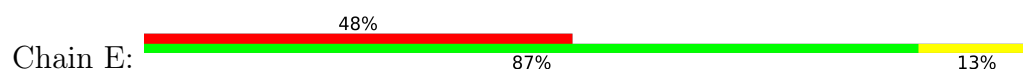
- Molecule 1: G/T mismatch-specific thymine DNA glycosylase

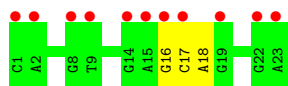


- Molecule 2: 5'-D(*CP*AP*GP*CP*TP*CP*TP*GP*TP*AP*CP*GP*TP*GP*AP*GP*CP*A P*GP*TP*GP*GP*A)-3'

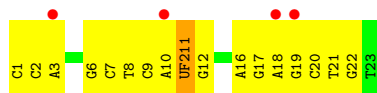


- Molecule 2: 5'-D(*CP*AP*GP*CP*TP*CP*TP*GP*TP*AP*CP*GP*TP*GP*AP*GP*CP*A P*GP*TP*GP*GP*A)-3'

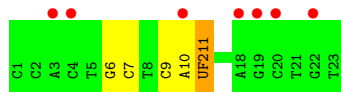
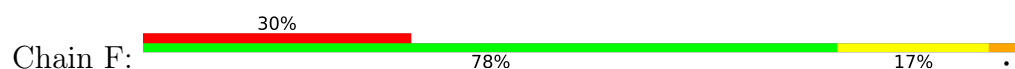




● Molecule 3: 5'-D(*CP*CP*AP*CP*TP*GP*CP*TP*CP*AP*(UF2)P*GP*TP*AP*CP*AP*GP*AP*GP*CP*TP*GP*T)-3'



● Molecule 3: 5'-D(*CP*CP*AP*CP*TP*GP*CP*TP*CP*AP*(UF2)P*GP*TP*AP*CP*AP*GP*AP*GP*CP*TP*GP*T)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α , β , γ	162.54Å 162.54Å 54.96Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.63 – 2.97 40.63 – 2.97	Depositor EDS
% Data completeness (in resolution range)	99.6 (40.63-2.97) 99.8 (40.63-2.97)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.20 (at 2.95Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.263 , 0.304 0.258 , 0.298	Depositor DCC
R_{free} test set	881 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	89.0	Xtriage
Anisotropy	0.541	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 94.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.065 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	4509	wwPDB-VP
Average B, all atoms (Å ²)	111.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.90% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: UF2

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	0.84	1/1406 (0.1%)	0.90	3/1917 (0.2%)
1	B	0.80	0/1299	0.93	0/1778
2	C	0.28	0/531	0.69	0/819
2	E	0.30	0/531	0.65	0/819
3	D	0.29	0/497	0.65	0/762
3	F	0.31	0/497	0.63	0/762
All	All	0.65	1/4761 (0.0%)	0.81	3/6857 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	188	GLY	N-CA	5.31	1.49	1.44

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	197	THR	CA-C-N	5.34	126.52	119.84
1	A	197	THR	C-N-CA	5.34	126.52	119.84
1	A	275	ARG	N-CA-C	-5.23	105.77	113.61

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	1367	0	1235	64	0
1	B	1265	0	1043	70	0
2	C	473	0	245	23	0
2	E	473	0	244	9	0
3	D	465	0	244	22	0
3	F	465	0	245	11	0
4	A	1	0	0	0	0
All	All	4509	0	3256	156	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 20.

The worst 5 of 156 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:156:GLY:N	3:F:9:DC:OP1	1.81	1.14
1:B:277:ALA:HB3	2:C:18:DA:H2	1.12	1.13
1:A:155:PRO:C	3:F:9:DC:OP1	1.98	1.07
1:B:170:GLU:HG2	1:B:185:TYR:OH	1.63	0.98
1:A:275:ARG:HG2	2:E:17:DC:H1'	1.44	0.98

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	184/204 (90%)	158 (86%)	21 (11%)	5 (3%)	4	11
1	B	180/204 (88%)	150 (83%)	23 (13%)	7 (4%)	2	5
All	All	364/408 (89%)	308 (85%)	44 (12%)	12 (3%)	3	7

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	204	SER
1	A	153	PRO
1	A	273	SER
1	B	153	PRO
1	B	205	SER

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	131/178 (74%)	124 (95%)	7 (5%)	20	45
1	B	102/178 (57%)	92 (90%)	10 (10%)	7	21
All	All	233/356 (65%)	216 (93%)	17 (7%)	13	33

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	261	ASP
1	B	299	LEU
1	B	163	LEU
1	B	190	THR
1	B	191	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	175	HIS
1	B	191	ASN
1	B	257	HIS
1	A	191	ASN
1	A	175	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are 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	$\# Z > 2$
3	UF2	F	11	3	18,21,22	0.58	0	25,30,33	1.07	2 (8%)
3	UF2	D	11	3	18,21,22	0.52	0	25,30,33	0.88	2 (8%)

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
3	UF2	F	11	3	-	0/7/25/26	0/2/2/2
3	UF2	D	11	3	-	0/7/25/26	0/2/2/2

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	11	UF2	F2'-C2'-C1'	-3.28	102.30	108.88
3	D	11	UF2	C3'-C2'-C1'	2.82	106.37	103.10
3	F	11	UF2	C3'-C2'-C1'	2.82	106.37	103.10
3	D	11	UF2	C2'-C3'-C4'	2.25	105.40	102.43

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	F	11	UF2	2	0
3	D	11	UF2	2	0

5.5 Carbohydrates [i](#)

There are no oligosaccharides 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	185/204 (90%)	0.41	10 (5%) 31 26	60, 89, 133, 160	1 (0%)
1	B	182/204 (89%)	0.74	12 (6%) 24 21	63, 100, 139, 168	0
2	C	23/23 (100%)	2.02	11 (47%) 0 0	52, 75, 111, 115	23 (100%)
2	E	23/23 (100%)	2.10	11 (47%) 0 0	38, 48, 74, 94	23 (100%)
3	D	22/23 (95%)	1.02	4 (18%) 3 3	49, 75, 111, 124	22 (100%)
3	F	22/23 (95%)	1.73	7 (31%) 1 1	29, 56, 74, 82	22 (100%)
All	All	457/500 (91%)	0.80	55 (12%) 9 8	29, 92, 133, 168	91 (19%)

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	1	DC	7.3
2	E	1	DC	7.2
2	E	16	DG	5.6
3	F	18	DA	4.9
2	E	15	DA	4.9

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, 95th 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
3	UF2	F	11	20/21	0.88	0.39	55,62,71,73	20
3	UF2	D	11	20/21	0.91	0.25	68,77,85,85	20

6.3 Carbohydrates [i](#)

There are no oligosaccharides 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.