



wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 05:28 PM GMT

PDB ID : 3UFJ
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 validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

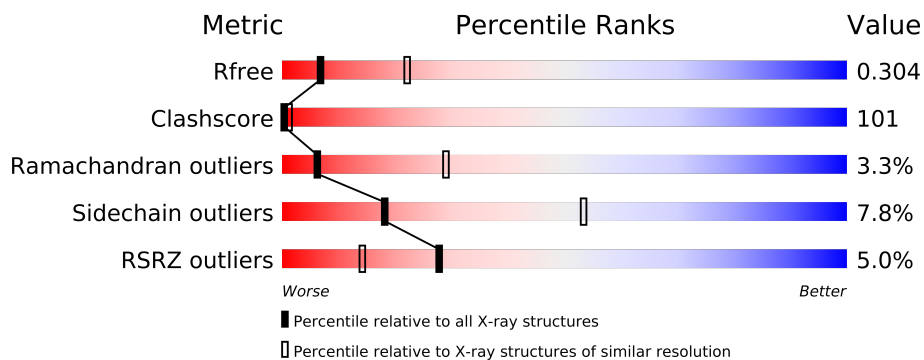
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
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 2.97 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}	66092	1587 (3.00-2.92)
Clashscore	79885	2029 (3.00-2.92)
Ramachandran outliers	78287	1955 (3.00-2.92)
Sidechain outliers	78261	1958 (3.00-2.92)
RSRZ outliers	66119	1588 (3.00-2.92)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	204	
1	B	204	
2	C	23	
2	E	23	
3	D	23	
3	F	23	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4509 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 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.

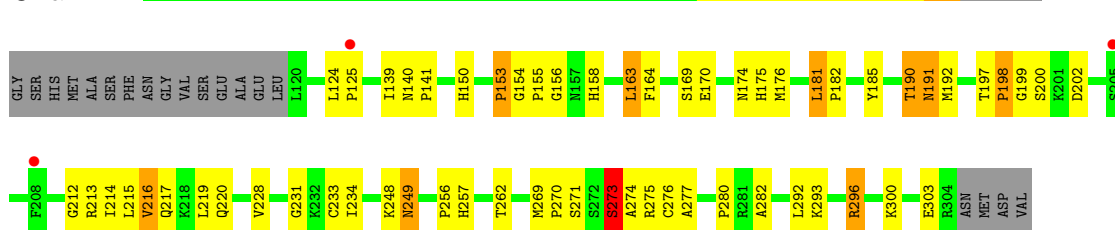
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total 1	O 1	0	0

3 Residue-property plots

These plots are drawn for all protein, RNA 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: G/T mismatch-specific thymine DNA glycosylase

Chain A:



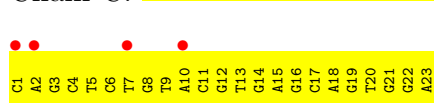
- Molecule 1: G/T mismatch-specific thymine DNA glycosylase

Chain B:



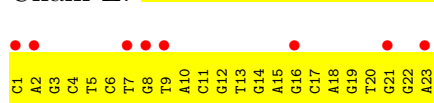
- Molecule 2: 5'-D(*CP*AP*GP*CP*TP*CP*TP*GP*TP*AP*CP*GP*TP*GP*AP*GP*CP*A P*GP*TP*GP*GP*A)-3'

Chain C:



- 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'

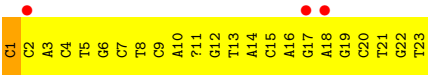
Chain E:



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

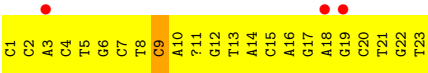
Chain D:





● 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'

Chain F: 



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 0.264 , 0.304	Depositor DCC
R_{free} test set	881 reflections (5.32%)	DCC
Wilson B-factor (Å ²)	89.0	Xtriage
Anisotropy	0.541	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.23 , 75.7	EDS
Estimated twinning fraction	0.065 for h,-h-k,-l	Xtriage
L-test for twinning	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	1 of 17438 reflections (0.006%)	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.

5 Model quality

5.1 Standard geometry

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	0.58	0/1406	0.65	0/1917
1	B	0.57	1/1299 (0.1%)	0.64	1/1778 (0.1%)
2	C	0.25	0/531	0.81	0/819
2	E	0.28	0/531	0.79	0/819
3	D	0.23	0/497	0.78	1/762 (0.1%)
3	F	0.28	0/497	0.75	1/762 (0.1%)
All	All	0.46	1/4761 (0.0%)	0.71	3/6857 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	160	TRP	CD2-CE2	6.41	1.49	1.41

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	9	DC	O5'-P-OP1	-6.04	100.26	105.70
1	B	203	LEU	CA-CB-CG	-5.86	101.82	115.30
3	D	1	DC	P-O3'-C3'	5.06	125.77	119.70

There are no chirality outliers.

There are no planarity outliers.

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	A	1367	0	1235	60	0
1	B	1265	0	1043	70	0
2	C	473	0	245	285	309
2	E	473	0	244	276	335
3	D	465	0	244	353	320
3	F	465	0	245	371	294
4	A	1	0	0	0	0
All	All	4509	0	3256	778	629

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 101.

The worst 5 of 778 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:D:6:DG:C5	2:E:11:DC:H2''	1.19	1.64
3:D:15:DC:H5'	2:E:20:DT:C2	1.31	1.60
2:C:8:DG:C4	3:F:1:DC:H2'	1.35	1.59
3:D:3:DA:C2'	2:E:10:DA:H5'	1.31	1.59
3:D:9:DC:C2	3:F:9:DC:C2	1.91	1.57

The worst 5 of 629 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	Distance(Å)	Clash(Å)
3:D:20:DC:N3	2:E:3:DG:C4[5_554]	0.18	2.02
2:C:6:DC:N3	3:F:23:DT:C4[5_554]	0.31	1.89
3:D:21:DT:N3	2:E:4:DC:N1[5_554]	0.32	1.88
3:D:22:DG:C3'	2:E:5:DT:C2'[5_554]	0.34	1.86
2:C:6:DC:C3'	3:F:23:DT:C4'[5_554]	0.40	1.80

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	184/204 (90%)	158 (86%)	21 (11%)	5 (3%)	8	36
1	B	180/204 (88%)	150 (83%)	23 (13%)	7 (4%)	5	23
All	All	364/408 (89%)	308 (85%)	44 (12%)	12 (3%)	6	29

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%)	123 (94%)	8 (6%)	26	67
1	B	102/178 (57%)	92 (90%)	10 (10%)	12	39
All	All	233/356 (65%)	215 (92%)	18 (8%)	18	54

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	163	LEU
1	B	190	THR
1	B	249	ASN
1	A	273	SER
1	A	296	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	175	HIS
1	B	175	HIS

5.3.3 RNA ⓘ

There are no RNA chains 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	D	11	3	19,21,22	0.82	1 (5%)	23,30,33	1.05	2 (8%)
3	UF2	F	11	3	19,21,22	0.79	0	23,30,33	1.18	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	D	11	3	-	0/6/25/26	0/2/2/2
3	UF2	F	11	3	-	0/6/25/26	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	11	UF2	P-O1P	2.17	1.49	1.46

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	11	UF2	F2'-C2'-C1'	-3.09	102.30	109.61
3	F	11	UF2	C3'-C2'-C1'	2.39	106.37	103.30
3	D	11	UF2	C3'-C2'-C1'	2.39	106.37	103.30
3	D	11	UF2	C2'-C3'-C4'	2.14	105.40	102.10

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 ⓘ

There are no ligands in this entry.

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 ⓘ

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%)	-0.01	3 (1%) 68 32	60, 89, 133, 160	0
1	B	182/204 (89%)	0.02	2 (1%) 77 36	63, 100, 139, 168	0
2	C	23/23 (100%)	1.03	4 (17%) 2 2	94, 134, 199, 206	23 (100%)
2	E	23/23 (100%)	1.66	8 (34%) 1 0	86, 108, 169, 214	23 (100%)
3	D	23/23 (100%)	0.93	3 (13%) 4 2	78, 131, 198, 222	23 (100%)
3	F	23/23 (100%)	1.01	3 (13%) 4 2	64, 116, 168, 187	23 (100%)
All	All	459/500 (91%)	0.24	23 (5%) 28 14	60, 99, 163, 222	92 (20%)

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	1	DC	7.7
3	D	18	DA	3.9
1	B	172	GLN	3.9
2	E	7	DT	3.8
3	F	18	DA	3.6

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(Å ²)	Q<0.9
3	UF2	F	11	20/21	0.47	6.67	55,62,71,73	20
3	UF2	D	11	20/21	0.29	2.51	68,77,85,85	20

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.