wwPDB X-ray Structure Validation Summary Report (i)

Apr 21, 2024 – 03:00 am BST

PDB ID	:	1OE6
Title	:	Xenopus SMUG1, an anti-mutator uracil-DNA Glycosylase
Authors	:	Wibley, J.E.A.; Pearl, L.H.
Deposited on	:	2003-03-19
Resolution	:	2.65 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 (i) 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 (i)) were used in the production of this report:

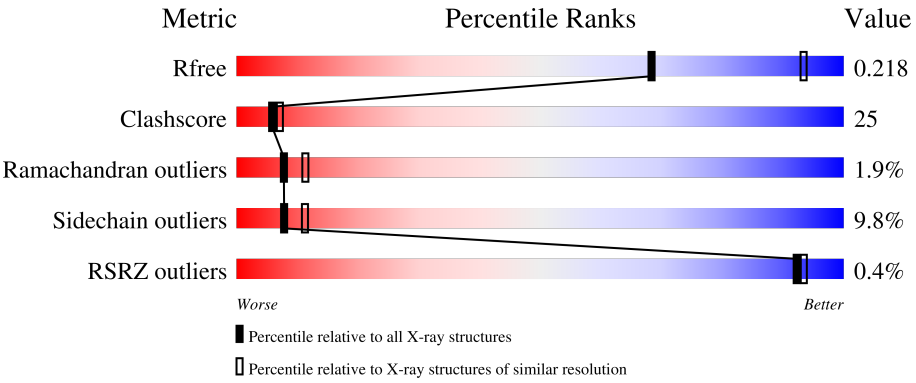
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.65 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}	130704	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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 <=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	247	46% 39% 11%
1	B	247	% 49% 34% 11% 5%
2	E	12	92% 8%
3	F	12	58% 42%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	HMU	A	1282	-	X	-	-
4	HMU	A	1283	-	X	-	-
5	GOL	A	1284	-	X	-	-
6	IPA	A	1285[A]	-	X	X	-
6	IPA	A	1285[B]	-	X	X	-

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 4499 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 SINGLE-STRAND SELECTIVE MONOFUNCTIONAL URACIL DNA GLYCOSYLASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	244	Total	C	N	O	S	0	2	0
			1938	1243	339	343	13			
1	B	245	Total	C	N	O	S	0	0	0
			1939	1244	339	344	12			

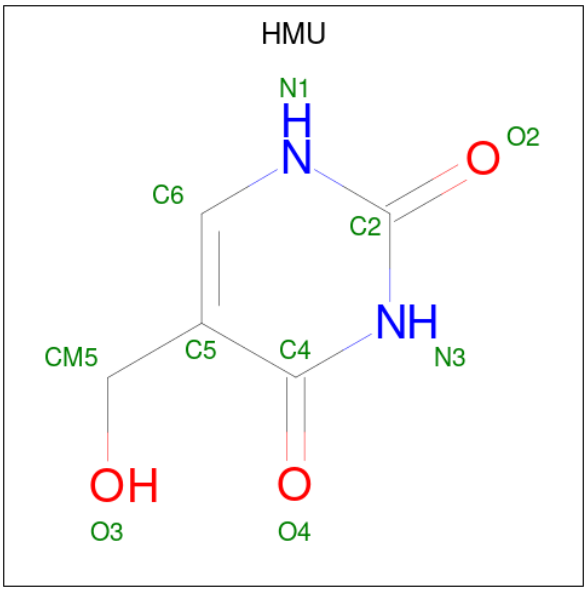
- Molecule 2 is a DNA chain called 5'-D(*CP*CP*CP*GP*TP*GP*AP*GP*TP*CP*CP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	12	Total	C	N	O	P	0	0	0
			241	115	44	71	11			

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

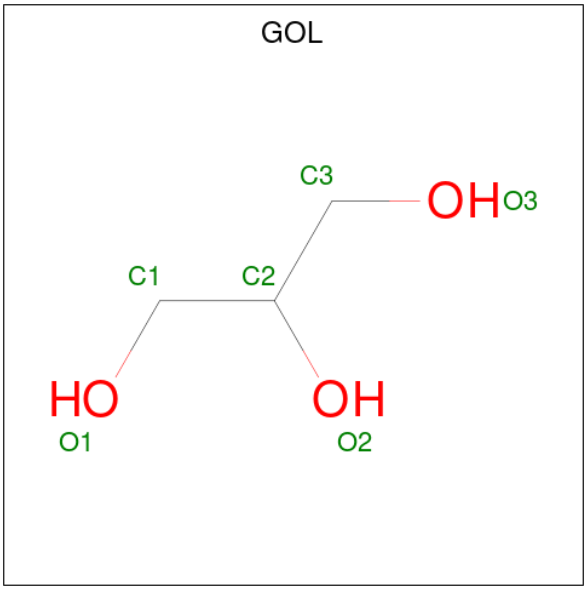
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	12	Total	C	N	O	P	0	0	0
			237	112	46	68	11			

- Molecule 4 is 5-HYDROXYMETHYL URACIL (three-letter code: HMU) (formula: $C_5H_6N_2O_3$).



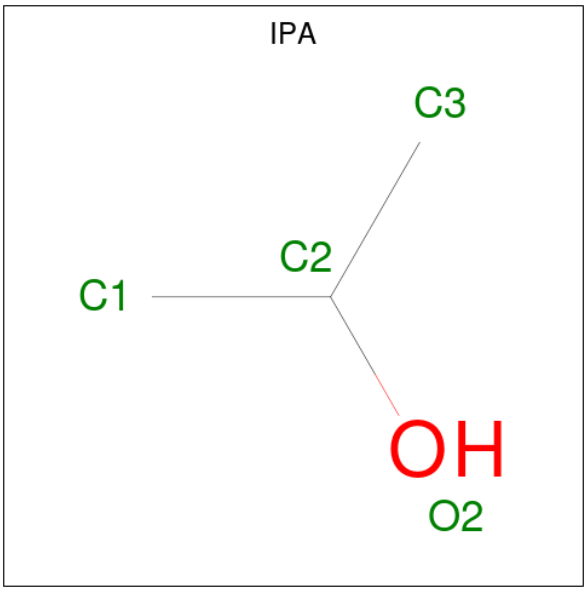
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C N O 10 5 2 3	0	0
4	A	1	Total C N O 10 5 2 3	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 6 3 3	0	0

- Molecule 6 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C_3H_8O).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total C O 8 6 2	0	1

- Molecule 7 is water.

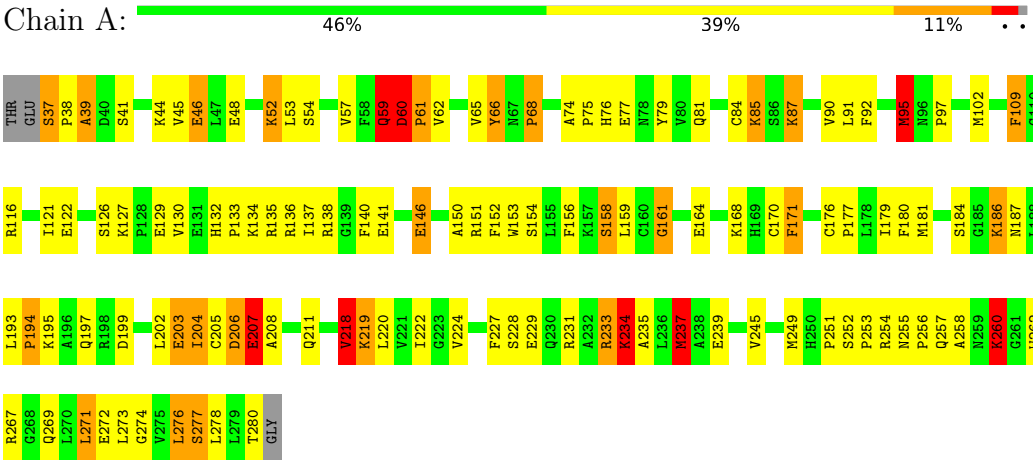
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	53	Total O 53 53	0	0
7	B	45	Total O 45 45	0	0
7	E	6	Total O 6 6	0	0
7	F	6	Total O 6 6	0	0

3 Residue-property plots (i)

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.

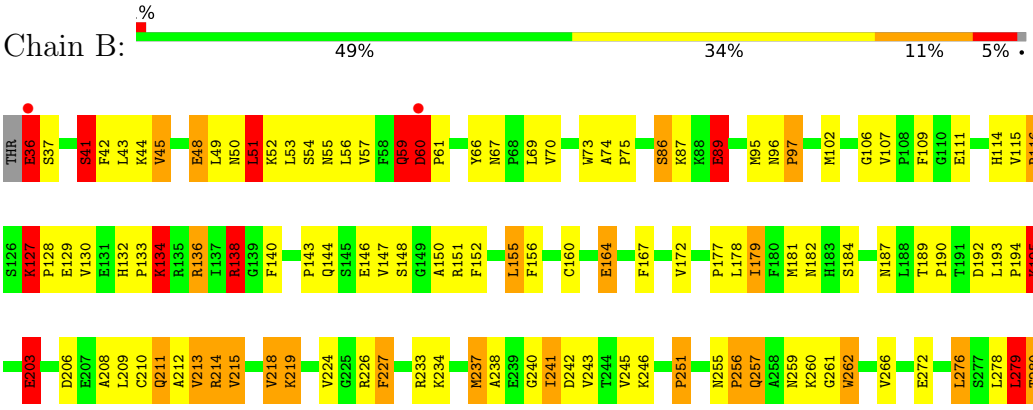
• Molecule 1: SINGLE-STRAND SELECTIVE MONOFUNCTIONAL URACIL DNA GLYCOSYLASE

Chain A:



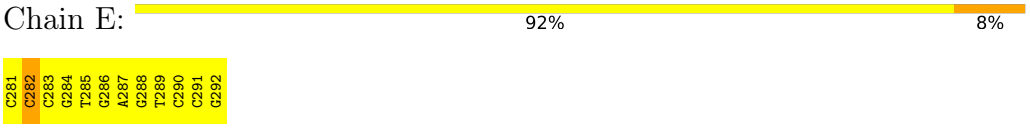
• Molecule 1: SINGLE-STRAND SELECTIVE MONOFUNCTIONAL URACIL DNA GLYCOSYLASE

Chain B:



• Molecule 2: 5'-D(*CP*CP*CP*GP*TP*GP*AP*GP*TP*CP*CP*G)-3'

Chain E:



• Molecule 3: 5'-D(*CP*GP*GP*AP*CP*TP*3DRP*AP*CP*GP*GP*G)-3'

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	120.97Å 86.52Å 78.46Å 90.00° 118.53° 90.00°	Depositor
Resolution (Å)	69.01 – 2.65 43.26 – 2.65	Depositor EDS
% Data completeness (in resolution range)	99.4 (69.01-2.65) 99.4 (43.26-2.65)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.24 (at 2.65Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.154 , 0.218 0.166 , 0.218	Depositor DCC
R_{free} test set	1063 reflections (5.15%)	wwPDB-VP
Wilson B-factor (Å^2)	47.3	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 60.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4499	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.55% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 3DR, IPA, GOL, HMU

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	2.48	103/1997 (5.2%)	1.81	48/2707 (1.8%)
1	B	2.33	82/1990 (4.1%)	1.69	31/2698 (1.1%)
2	E	4.32	58/269 (21.6%)	5.34	124/413 (30.0%)
3	F	5.25	81/253 (32.0%)	5.51	119/387 (30.7%)
All	All	2.79	324/4509 (7.2%)	2.54	322/6205 (5.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	3
All	All	0	4

The worst 5 of 324 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	298	DT	C1'-N1	18.26	1.73	1.49
3	F	296	DA	C8-N7	17.16	1.43	1.31
2	E	292	DG	C8-N7	16.76	1.41	1.30
3	F	296	DA	C6-N1	16.14	1.46	1.35
1	A	46	GLU	CD-OE2	15.73	1.43	1.25

The worst 5 of 322 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	F	298	DT	O4'-C1'-N1	22.18	123.53	108.00
3	F	297	DC	C5-C4-N4	18.73	133.31	120.20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	F	293	DC	N1-C2-O2	18.57	130.04	118.90
2	E	291	DC	O5'-P-OP1	-18.05	89.04	110.70
2	E	291	DC	O4'-C1'-N1	16.89	119.82	108.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	59	GLN	Peptide
1	B	196	ALA	Peptide
1	B	279	LEU	Peptide
1	B	59	GLN	Peptide

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	1938	0	1943	125	0
1	B	1939	0	1942	65	0
2	E	241	0	136	8	0
3	F	237	0	132	12	0
4	A	20	0	12	4	0
5	A	6	0	5	0	0
6	A	8	0	13	64	0
7	A	53	0	0	12	0
7	B	45	0	0	11	0
7	E	6	0	0	2	0
7	F	6	0	0	2	0
All	All	4499	0	4183	212	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 25.

The worst 5 of 212 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:B:127:LYS:CD	1:B:127:LYS:CG	1.74	1.59
1:B:134:LYS:CD	1:B:134:LYS:CE	1.77	1.57
1:B:134:LYS:CD	1:B:134:LYS:CG	1.76	1.56
3:F:298:DT:C1'	3:F:298:DT:N1	1.73	1.48
1:A:237[B]:MET:SD	1:A:237[B]:MET:CG	2.06	1.44

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
1	A	244/247 (99%)	223 (91%)	18 (7%)	3 (1%)	13 19
1	B	243/247 (98%)	218 (90%)	19 (8%)	6 (2%)	5 7
All	All	487/494 (99%)	441 (91%)	37 (8%)	9 (2%)	8 12

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	60	ASP
1	B	60	ASP
1	B	196	ALA
1	B	195	LYS
1	B	240	GLY

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	215/215 (100%)	198 (92%)	17 (8%)	12 19
1	B	214/215 (100%)	188 (88%)	26 (12%)	5 6
All	All	429/430 (100%)	386 (90%)	43 (10%)	8 10

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	138	ARG
1	B	214	ARG
1	B	155	LEU
1	B	197	GLN
1	B	241	ILE

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	50	ASN
1	B	78	ASN
1	B	230	GLN
1	A	175	HIS
1	A	76	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	$\#\|Z\| > 2$
3	3DR	F	299	3	8,11,12	2.87	3 (37%)	9,14,17	2.20	5 (55%)

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	3DR	F	299	3	-	3/3/15/16	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	299	3DR	O3'-C3'	6.02	1.56	1.43
3	F	299	3DR	O5'-C5'	3.94	1.54	1.44
3	F	299	3DR	O4'-C4'	2.85	1.49	1.44

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	F	299	3DR	C2'-C3'-C4'	-3.31	95.91	102.75
3	F	299	3DR	O4'-C1'-C2'	-2.83	100.65	106.37
3	F	299	3DR	O3'-C3'-C4'	2.82	120.88	110.10
3	F	299	3DR	O4'-C4'-C3'	2.79	107.84	103.73
3	F	299	3DR	C1'-C2'-C3'	-2.03	100.92	103.20

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	F	299	3DR	C4'-C5'-O5'-P
3	F	299	3DR	C3'-C4'-C5'-O5'
3	F	299	3DR	O4'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	F	299	3DR	1	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

5 ligands 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	$\#\|Z\| > 2$
4	HMU	A	1282	-	10,10,10	3.93	8 (80%)	11,13,13	3.39	7 (63%)
6	IPA	A	1285[B]	-	3,3,3	1.31	0	3,3,3	3.12	3 (100%)
6	IPA	A	1285[A]	-	3,3,3	1.37	0	3,3,3	3.46	3 (100%)
5	GOL	A	1284	-	5,5,5	19.37	3 (60%)	5,5,5	10.93	4 (80%)
4	HMU	A	1283	-	10,10,10	2.34	4 (40%)	11,13,13	3.73	6 (54%)

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	HMU	A	1282	-	-	2/2/2/2	0/1/1/1
5	GOL	A	1284	-	-	2/4/4/4	-
4	HMU	A	1283	-	-	2/2/2/2	0/1/1/1

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1284	GOL	O2-C2	34.90	2.47	1.43
5	A	1284	GOL	O3-C3	25.29	2.49	1.42
4	A	1282	HMU	O2-C2	6.78	1.37	1.23
4	A	1282	HMU	C2-N1	5.47	1.44	1.36
4	A	1282	HMU	C6-C5	5.34	1.43	1.36

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	1284	GOL	O2-C2-C3	-17.37	32.62	109.12

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	1284	GOL	O3-C3-C2	-15.69	34.97	110.20
4	A	1283	HMU	O2-C2-N1	-8.05	113.93	122.79
4	A	1282	HMU	O4-C4-C5	-7.08	114.42	124.96
4	A	1283	HMU	N1-C2-N3	6.62	122.63	115.13

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1282	HMU	C4-C5-CM5-O3
4	A	1282	HMU	C6-C5-CM5-O3
5	A	1284	GOL	C1-C2-C3-O3
5	A	1284	GOL	O2-C2-C3-O3
4	A	1283	HMU	C6-C5-CM5-O3

There are no ring outliers.

4 monomers are involved in 68 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1282	HMU	3	0
6	A	1285[B]	IPA	33	0
6	A	1285[A]	IPA	31	0
4	A	1283	HMU	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

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^{th} 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	244/247 (98%)	-0.52	0 100 100	20, 40, 61, 72	0
1	B	245/247 (99%)	-0.45	2 (0%) 86 85	22, 48, 68, 76	0
2	E	12/12 (100%)	-0.21	0 100 100	40, 61, 75, 82	0
3	F	11/12 (91%)	0.16	0 100 100	38, 60, 85, 87	0
All	All	512/518 (98%)	-0.47	2 (0%) 92 93	20, 45, 68, 87	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	60	ASP	3.9
1	B	36	GLU	2.6

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. The B-factors column lists the minimum, median, 95^{th} 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	3DR	F	299	11/12	0.97	0.14	67,75,80,83	0

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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, 95^{th} 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
5	GOL	A	1284	6/6	0.52	0.36	73,84,87,90	0
4	HMU	A	1283	10/10	0.81	0.29	80,97,101,103	0
4	HMU	A	1282	10/10	0.87	0.23	46,62,75,80	0
6	IPA	A	1285[A]	4/4	0.97	0.24	124,125,126,126	4
6	IPA	A	1285[B]	4/4	0.97	0.24	125,125,126,129	4

6.5 Other polymers

There are no such residues in this entry.