wwPDB X-ray Structure Validation Summary Report (i)

Nov 1, 2023 – 02:43 PM JST

PDB ID	:	5ITY
Title	:	Crystal Structure of Human NEIL1(P2G) bound to duplex DNA containing Thymine Glycol
Authors	:	Zhu, C.; Lu, L.; Zhang, J.; Yue, Z.; Song, J.; Zong, S.; Liu, M.; Stovicek, O.; Gao, Y.; Yi, C.
Deposited on	:	2016-03-17
Resolution	:	2.48 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 (1)) were used in the production of this report:

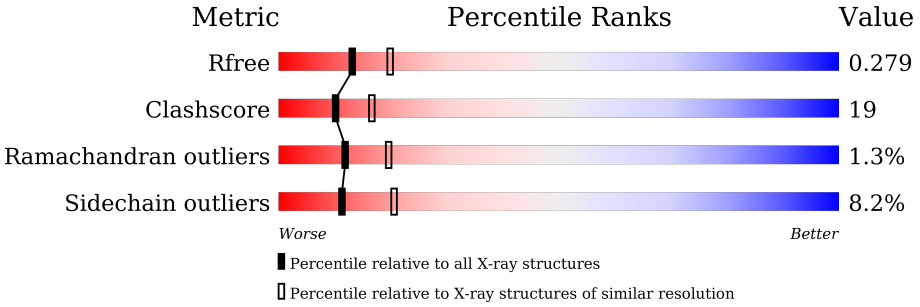
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.48 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	5857 (2.50-2.46)
Clashscore	141614	6594 (2.50-2.46)
Ramachandran outliers	138981	6469 (2.50-2.46)
Sidechain outliers	138945	6471 (2.50-2.46)

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%

Mol	Chain	Length	Quality of chain
1	A	400	53% 12% • 34%
1	B	400	46% 16% • 34%
1	C	400	40% 21% 5% 35%
2	D	26	50% 42% 8%
2	E	26	38% 54% 8%
2	F	26	31% 65% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8239 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 Endonuclease 8-like 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	265	2148	1370	398	370	10	0	5	0
1	B	262	2078	1327	381	360	10	0	0	0
1	C	261	2066	1319	380	357	10	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLY	PRO	engineered mutation	UNP Q96FI4
A	242	ARG	LYS	engineered mutation	UNP Q96FI4
A	391	ALA	-	expression tag	UNP Q96FI4
A	392	ALA	-	expression tag	UNP Q96FI4
A	393	LEU	-	expression tag	UNP Q96FI4
A	394	GLY	-	expression tag	UNP Q96FI4
A	395	HIS	-	expression tag	UNP Q96FI4
A	396	HIS	-	expression tag	UNP Q96FI4
A	397	HIS	-	expression tag	UNP Q96FI4
A	398	HIS	-	expression tag	UNP Q96FI4
A	399	HIS	-	expression tag	UNP Q96FI4
A	400	HIS	-	expression tag	UNP Q96FI4
B	2	GLY	PRO	engineered mutation	UNP Q96FI4
B	242	ARG	LYS	engineered mutation	UNP Q96FI4
B	391	ALA	-	expression tag	UNP Q96FI4
B	392	ALA	-	expression tag	UNP Q96FI4
B	393	LEU	-	expression tag	UNP Q96FI4
B	394	GLY	-	expression tag	UNP Q96FI4
B	395	HIS	-	expression tag	UNP Q96FI4
B	396	HIS	-	expression tag	UNP Q96FI4
B	397	HIS	-	expression tag	UNP Q96FI4
B	398	HIS	-	expression tag	UNP Q96FI4
B	399	HIS	-	expression tag	UNP Q96FI4

Continued on next page...

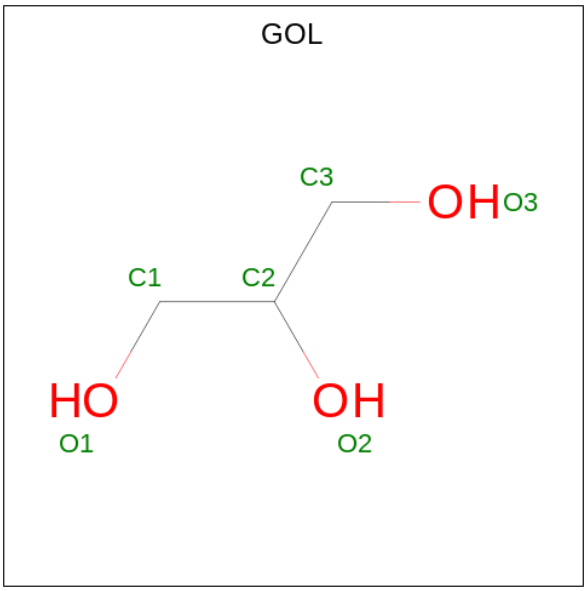
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	400	HIS	-	expression tag	UNP Q96FI4
C	2	GLY	PRO	engineered mutation	UNP Q96FI4
C	242	ARG	LYS	engineered mutation	UNP Q96FI4
C	391	ALA	-	expression tag	UNP Q96FI4
C	392	ALA	-	expression tag	UNP Q96FI4
C	393	LEU	-	expression tag	UNP Q96FI4
C	394	GLY	-	expression tag	UNP Q96FI4
C	395	HIS	-	expression tag	UNP Q96FI4
C	396	HIS	-	expression tag	UNP Q96FI4
C	397	HIS	-	expression tag	UNP Q96FI4
C	398	HIS	-	expression tag	UNP Q96FI4
C	399	HIS	-	expression tag	UNP Q96FI4
C	400	HIS	-	expression tag	UNP Q96FI4

- Molecule 2 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	26	527	252	96	155	24	0	0	0
2	E	26	527	252	96	155	24	0	0	0
2	F	26	527	252	96	155	24	0	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 4 is water.

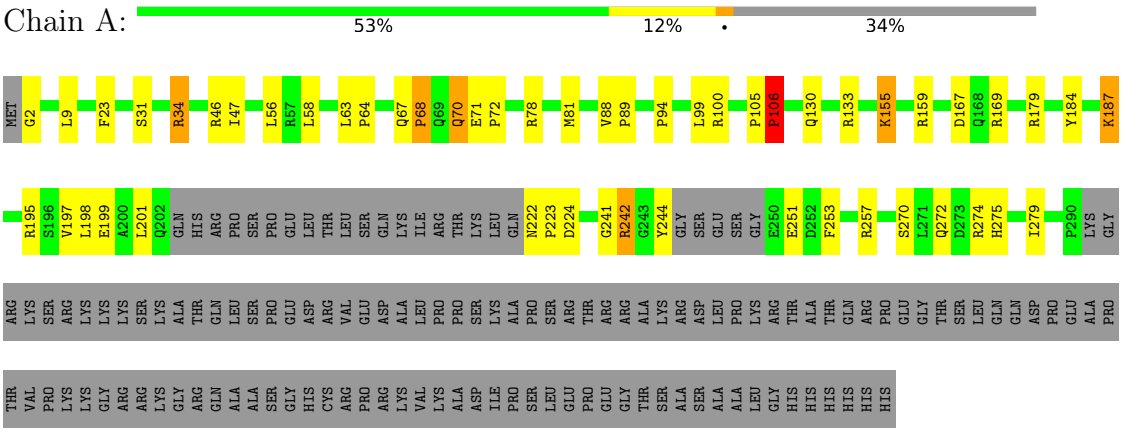
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	186	186	186	0	0
			Total	O		
4	B	121	121	121	0	0
			Total	O		
4	C	8	8	8	0	0
			Total	O		
4	D	29	29	29	0	0
			Total	O		
4	E	13	13	13	0	0
			Total	O		
4	F	3	3	3	0	0

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. 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. 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.

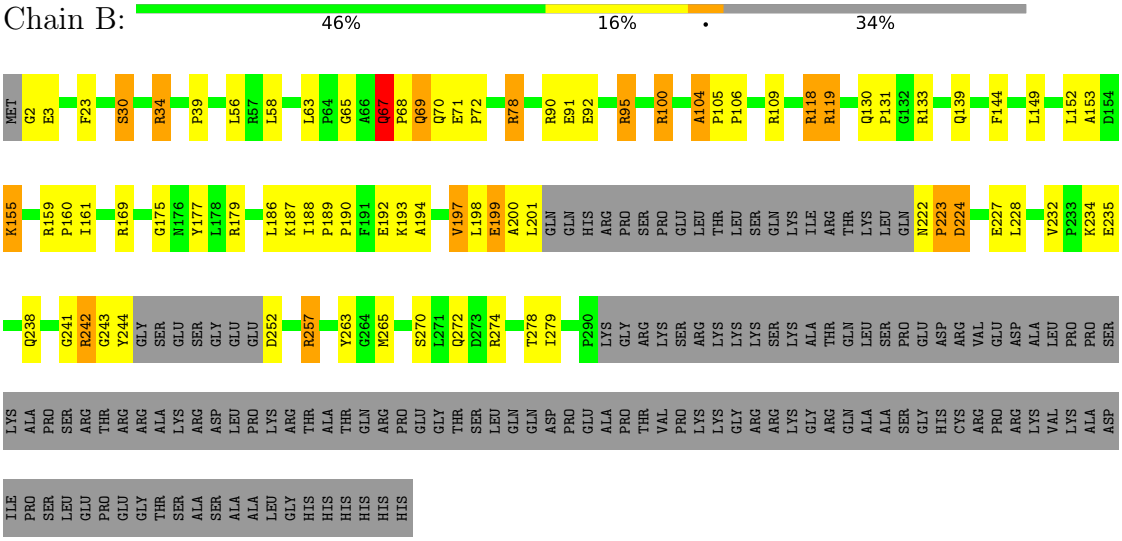
- Molecule 1: Endonuclease 8-like 1

Chain A:



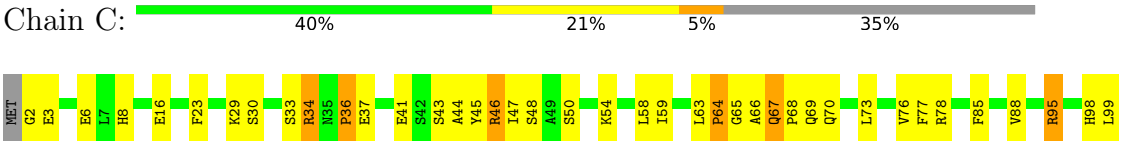
- Molecule 1: Endonuclease 8-like 1

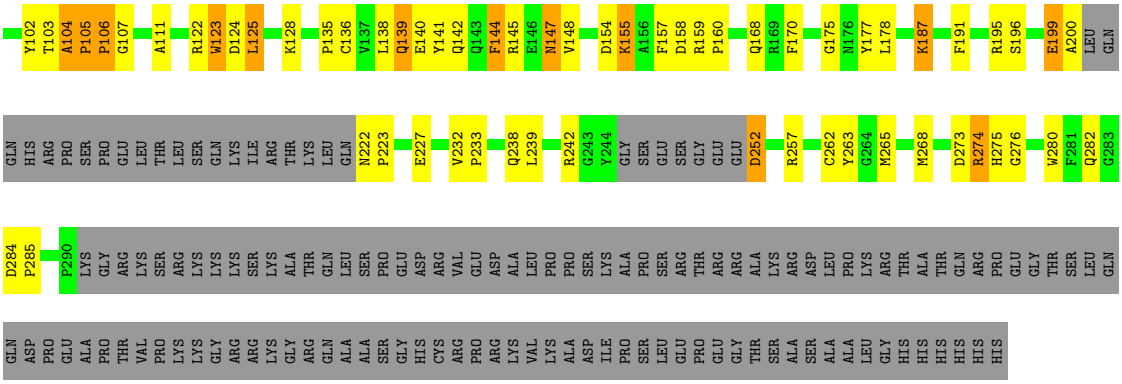
Chain B:



- Molecule 1: Endonuclease 8-like 1

Chain C:



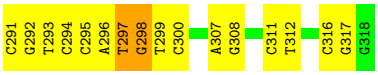


- Molecule 2: DNA (26-MER)

Chain D:

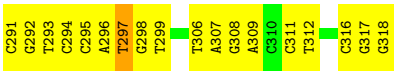
- Molecule 2: DNA (26-MER)

Chain E:



- Molecule 2: DNA (26-MER)

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	73.62Å 108.71Å 169.97Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	91.58 – 2.48 30.48 – 2.48	Depositor EDS
% Data completeness (in resolution range)	98.0 (91.58-2.48) 98.1 (30.48-2.48)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.56 (at 2.48Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.176 , 0.225 0.240 , 0.279	Depositor DCC
R_{free} test set	2387 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å^2)	61.5	Xtriage
Anisotropy	0.218	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 53.0	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	0.94	EDS
Total number of atoms	8239	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.13% 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 (i)

5.1 Standard geometry (i)

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

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	0.90	0/2215	1.02	9/2992 (0.3%)
1	B	0.84	0/2134	1.02	14/2886 (0.5%)
1	C	0.55	0/2122	0.73	0/2870
2	D	0.52	0/564	1.01	3/864 (0.3%)
2	E	0.62	0/564	1.17	3/864 (0.3%)
2	F	0.40	0/564	0.76	0/864
All	All	0.74	0/8163	0.95	29/11340 (0.3%)

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	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	298	DG	O5'-P-OP1	-17.39	89.84	110.70
2	E	298	DG	O5'-P-OP2	12.48	125.68	110.70
1	B	119	ARG	NE-CZ-NH2	-10.05	115.27	120.30
2	D	313	DG	O5'-P-OP1	-8.26	98.26	105.70
1	A	78	ARG	NE-CZ-NH2	-7.20	116.70	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	105	PRO	Peptide
1	B	104	ALA	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	2148	0	2128	66	1
1	B	2078	0	2057	79	1
1	C	2066	0	2042	107	0
2	D	527	0	298	19	0
2	E	527	0	298	12	0
2	F	527	0	298	22	0
3	B	6	0	8	0	0
4	A	186	0	0	10	0
4	B	121	0	0	14	0
4	C	8	0	0	0	0
4	D	29	0	0	2	0
4	E	13	0	0	0	0
4	F	3	0	0	0	0
All	All	8239	0	7129	283	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 19.

The worst 5 of 283 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:43:SER:HB3	1:C:67:GLN:CG	1.41	1.46
1:C:43:SER:CB	1:C:67:GLN:HG2	1.49	1.42
1:C:104:ALA:HB1	1:C:105:PRO:CD	1.69	1.20
1:C:43:SER:OG	1:C:67:GLN:HG2	1.41	1.18
1:B:179:ARG:NH2	4:B:601:HOH:O	1.77	1.15

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:199:GLU:OE1	1:B:34:ARG:NE[3_644]	2.03	0.17

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	263/400 (66%)	252 (96%)	9 (3%)	2 (1%)	19	33
1	B	256/400 (64%)	240 (94%)	15 (6%)	1 (0%)	34	52
1	C	255/400 (64%)	217 (85%)	31 (12%)	7 (3%)	5	6
All	All	774/1200 (64%)	709 (92%)	55 (7%)	10 (1%)	12	19

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	106	PRO
1	C	67	GLN
1	C	105	PRO
1	C	64	PRO
1	C	36	PRO

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	224/334 (67%)	215 (96%)	9 (4%)	31	53
1	B	216/334 (65%)	198 (92%)	18 (8%)	11	20

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	214/334 (64%)	188 (88%)	26 (12%)	5 8
All	All	654/1002 (65%)	601 (92%)	53 (8%)	11 21

5 of 53 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	34	ARG
1	C	85	PHE
1	C	238	GLN
1	C	46	ARG
1	C	70	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	130	GLN
1	C	139	GLN
1	B	70	GLN
1	B	130	GLN
1	C	8	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)

3 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	$\#\|Z\| > 2$
2	CTG	F	297	2	19,23,24	0.98	1 (5%)	21,35,38	1.16	4 (19%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	CTG	D	297	2	19,23,24	1.02	1 (5%)	21,35,38	1.82	8 (38%)
2	CTG	E	297	2	19,23,24	0.94	0	21,35,38	1.78	7 (33%)

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
2	CTG	F	297	2	-	1/7/45/46	0/2/2/2
2	CTG	D	297	2	-	2/7/45/46	0/2/2/2
2	CTG	E	297	2	-	2/7/45/46	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	297	CTG	C2-N3	-2.14	1.34	1.38
2	F	297	CTG	C4-N3	-2.07	1.34	1.37

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	297	CTG	O4'-C1'-N1	-4.14	103.75	108.65
2	E	297	CTG	O5-C5-C4	3.49	115.29	109.87
2	D	297	CTG	C3'-C2'-C1'	-3.05	94.88	102.54
2	E	297	CTG	O5-C5-C5M	-3.01	103.24	109.09
2	D	297	CTG	O3'-C3'-C4'	-2.96	98.78	110.10

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	297	CTG	O4'-C4'-C5'-O5'
2	E	297	CTG	O4'-C4'-C5'-O5'
2	D	297	CTG	C3'-C4'-C5'-O5'
2	E	297	CTG	C3'-C4'-C5'-O5'
2	F	297	CTG	O4'-C4'-C5'-O5'

There are no ring outliers.

3 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	F	297	CTG	2	0
2	D	297	CTG	6	0
2	E	297	CTG	3	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

1 ligand 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	GOL	B	501	-	5,5,5	1.05	0	5,5,5	1.06	0

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	GOL	B	501	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	501	GOL	O1-C1-C2-C3
3	B	501	GOL	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.