



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2026 – 07:38 AM UTC

PDB ID : 5ITY / pdb_00005ity
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 ⓘ 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

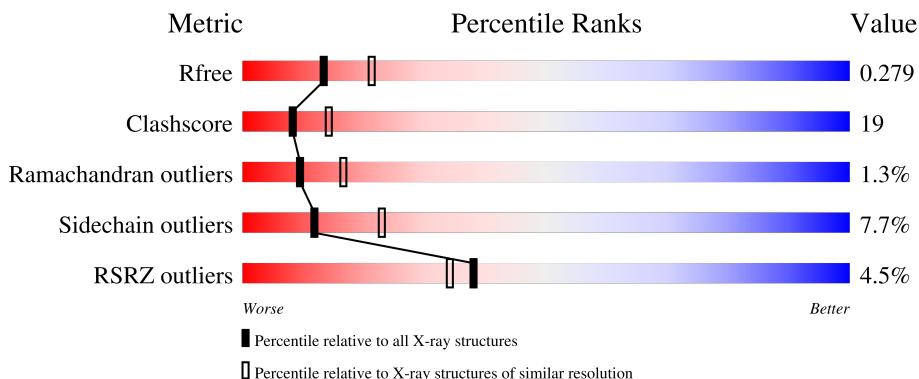
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.48 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}	180053	7589 (2.50-2.46)
Clashscore	190562	8295 (2.50-2.46)
Ramachandran outliers	187476	8164 (2.50-2.46)
Sidechain outliers	187428	8166 (2.50-2.46)
RSRZ outliers	180081	7593 (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 $\leq 5\%$. 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	400	
1	B	400	
1	C	400	
2	D	26	
2	E	26	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	26	 31% 62% 8%

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...

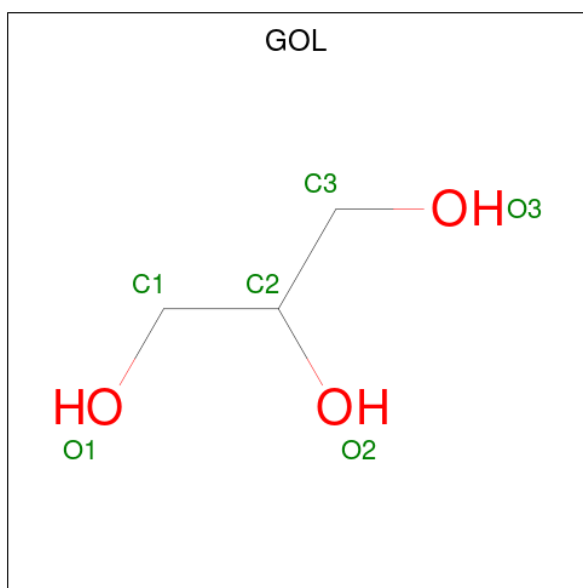
Continued from previous page...

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	Total 527	C 252	N 96	O 155	P 24	0	0	0
2	E	26	Total 527	C 252	N 96	O 155	P 24	0	0	0
2	F	26	Total 527	C 252	N 96	O 155	P 24	0	0	0

- Molecule 3 is GLYCEROL (CCD ID: GOL) (formula: C₃H₈O₃).



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

- Molecule 4 is water.

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

4 Data and refinement statistics

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 91.58 – 2.48	Depositor EDS
% Data completeness (in resolution range)	98.0 (91.58-2.48) 98.4 (91.58-2.48)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.56 (at 2.48Å)	Xtrriage
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 (4.83%)	wwPDB-VP
Wilson B-factor (Å ²)	61.5	Xtrriage
Anisotropy	0.218	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 43.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
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

Xtrriage'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.*

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: GOL, CTG

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	1.13	1/2215 (0.0%)	1.29	16/2992 (0.5%)
1	B	1.09	1/2134 (0.0%)	1.27	16/2886 (0.6%)
1	C	0.70	0/2122	1.07	11/2870 (0.4%)
2	D	0.45	0/564	1.02	3/864 (0.3%)
2	E	0.50	0/564	1.03	4/864 (0.5%)
2	F	0.37	0/564	0.79	1/864 (0.1%)
All	All	0.91	2/8163 (0.0%)	1.16	51/11340 (0.4%)

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

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	188	ILE	C-O	-6.66	1.16	1.24
1	B	117	ILE	C-O	-5.80	1.17	1.24

The worst 5 of 51 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	35	ASN	CA-C-N	-8.82	110.79	120.14
1	A	35	ASN	C-N-CA	-8.82	110.79	120.14
1	B	67	GLN	C-N-CD	8.53	139.36	120.60

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	198	LEU	N-CA-C	8.11	121.04	111.71
1	B	224	ASP	N-CA-C	8.09	120.77	110.33

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	82	1
1	C	2066	0	2042	109	0
2	D	527	0	298	20	0
2	E	527	0	298	14	0
2	F	527	0	298	23	0
3	B	6	0	8	0	0
4	A	186	0	0	10	0
4	B	121	0	0	13	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	289	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 289 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

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
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%)	16	28
1	B	256/400 (64%)	240 (94%)	15 (6%)	1 (0%)	30	46
1	C	255/400 (64%)	217 (85%)	31 (12%)	7 (3%)	4	5
All	All	774/1200 (64%)	709 (92%)	55 (7%)	10 (1%)	9	16

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%)	210 (94%)	14 (6%)	16	31
1	B	216/334 (65%)	202 (94%)	14 (6%)	15	30
1	C	214/334 (64%)	192 (90%)	22 (10%)	7	13
All	All	654/1002 (65%)	604 (92%)	50 (8%)	12	23

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	7	LEU
1	C	76	VAL
1	C	274	ARG
1	C	16	GLU
1	C	69	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	143	GLN
1	C	8	HIS
1	C	139	GLN
1	C	70	GLN
1	C	130	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](#)

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	E	297	2	18,23,24	0.95	0	21,35,38	1.94	8 (38%)
2	CTG	D	297	2	18,23,24	1.02	1 (5%)	21,35,38	2.11	9 (42%)
2	CTG	F	297	2	18,23,24	0.99	1 (5%)	21,35,38	1.80	5 (23%)

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	E	297	2	-	2/7/45/46	0/2/2/2
2	CTG	D	297	2	-	2/7/45/46	0/2/2/2
2	CTG	F	297	2	-	0/7/45/46	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	-2.18	1.34	1.38
2	F	297	CTG	C4-N3	-2.03	1.34	1.37

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	297	CTG	O4-C4-C5	-4.66	118.89	123.13
2	F	297	CTG	C5-C4-N3	4.34	120.93	115.23
2	D	297	CTG	O4'-C1'-N1	-4.14	103.75	108.65
2	D	297	CTG	O4-C4-C5	-3.94	119.55	123.13
2	E	297	CTG	O5-C5-C4	3.49	115.29	109.87

There are no chirality outliers.

All (4) torsion outliers are listed below:

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

There are no ring outliers.

3 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	297	CTG	4	0
2	D	297	CTG	6	0
2	F	297	CTG	2	0

5.5 Carbohydrates [i](#)

There are no oligosaccharides 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	1.10	0	5,5,5	1.10	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](#)

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	265/400 (66%)	-0.23	5 (1%) 66 63	13, 24, 52, 98	5 (1%)
1	B	262/400 (65%)	0.18	11 (4%) 40 36	16, 32, 59, 122	0
1	C	261/400 (65%)	0.79	22 (8%) 17 14	39, 67, 93, 151	0
2	D	25/26 (96%)	-0.13	0 100 100	35, 55, 81, 101	0
2	E	25/26 (96%)	0.18	1 (4%) 42 38	32, 55, 70, 74	0
2	F	25/26 (96%)	0.39	0 100 100	64, 81, 105, 110	0
All	All	863/1278 (67%)	0.24	39 (4%) 38 34	13, 40, 86, 151	5 (0%)

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	106	PRO	6.5
1	C	200	ALA	6.4
1	A	200	ALA	5.5
1	B	240	GLY	5.3
1	A	199	GLU	4.1

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
2	CTG	F	297	22/23	0.90	0.09	63,78,85,87	0
2	CTG	E	297	22/23	0.92	0.09	40,56,61,64	0
2	CTG	D	297	22/23	0.94	0.08	34,47,57,58	0

6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands [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	GOL	B	501	6/6	0.83	0.29	57,80,85,86	0

6.5 Other polymers [i](#)

There are no such residues in this entry.