



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 10, 2020 – 05:58 AM BST

PDB ID : 1TAU  
Title : TAQ POLYMERASE (E.C.2.7.7.7)/DNA/B-OCTYLGLUCOSIDE COMPLEX  
Authors : Eom, S.H.; Wang, J.; Steitz, T.A.  
Deposited on : 1996-06-17  
Resolution : 3.00 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](mailto: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 following versions of software and data (see [references ⓘ](#)) 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 : 2.13.1  
buster-report : 1.1.7 (2018)  
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) : 2.13.1

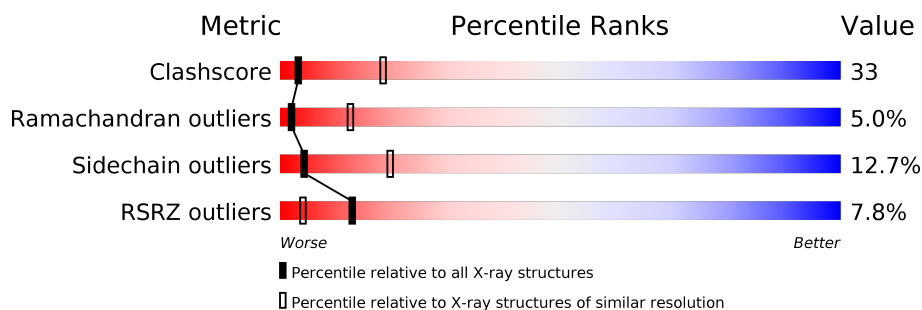
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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  $\geq 3$ , 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	T	8	<div> <div>75%</div> <div>100%</div> </div>
2	P	8	<div> <div>75%</div> <div>100%</div> </div>
3	A	832	<div> <div>6%</div> <div>41%</div> <div>45%</div> <div>9%</div> <div>5%</div> </div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 8122 atoms, of which 1448 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 DNA chain called DNA (5'-D(\*GP\*CP\*GP\*AP\*TP\*CP\*CP\*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	T	8	Total	C	H	N	O	P	0	0	0
			181	77	20	31	46	7			

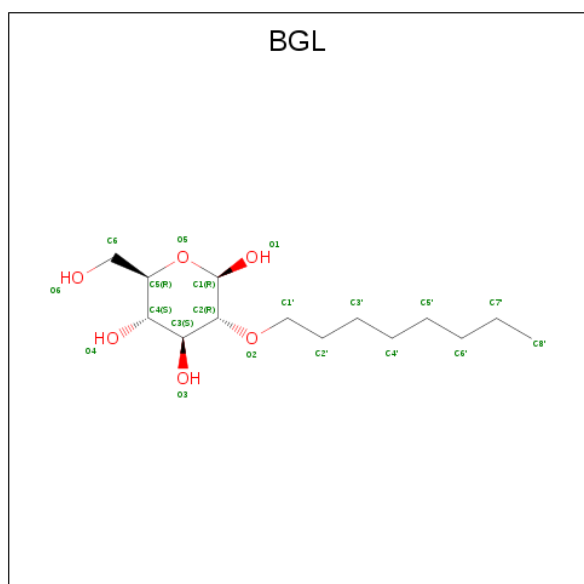
- Molecule 2 is a DNA chain called DNA (5'-D(\*CP\*GP\*GP\*AP\*TP\*CP\*GP\*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	P	8	Total	C	H	N	O	P	0	0	0
			181	77	20	31	46	7			

- Molecule 3 is a protein called PROTEIN (TAQ POLYMERASE).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	A	793	Total	C	H	N	O	S	0	0	0
			7739	4039	1408	1125	1153	14			

- Molecule 4 is 2-O-octyl-beta-D-glucopyranose (three-letter code: BGL) (formula: C<sub>14</sub>H<sub>28</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			20	14	6		

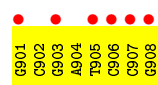
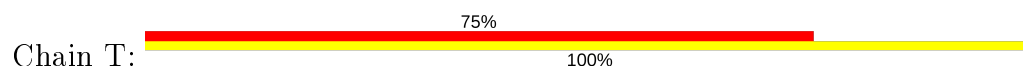
- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Zn	0	0
			1	1		

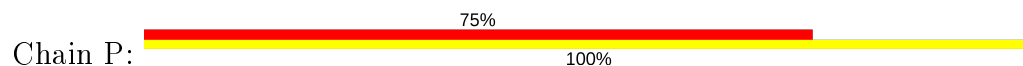
### 3 Residue-property plots

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.

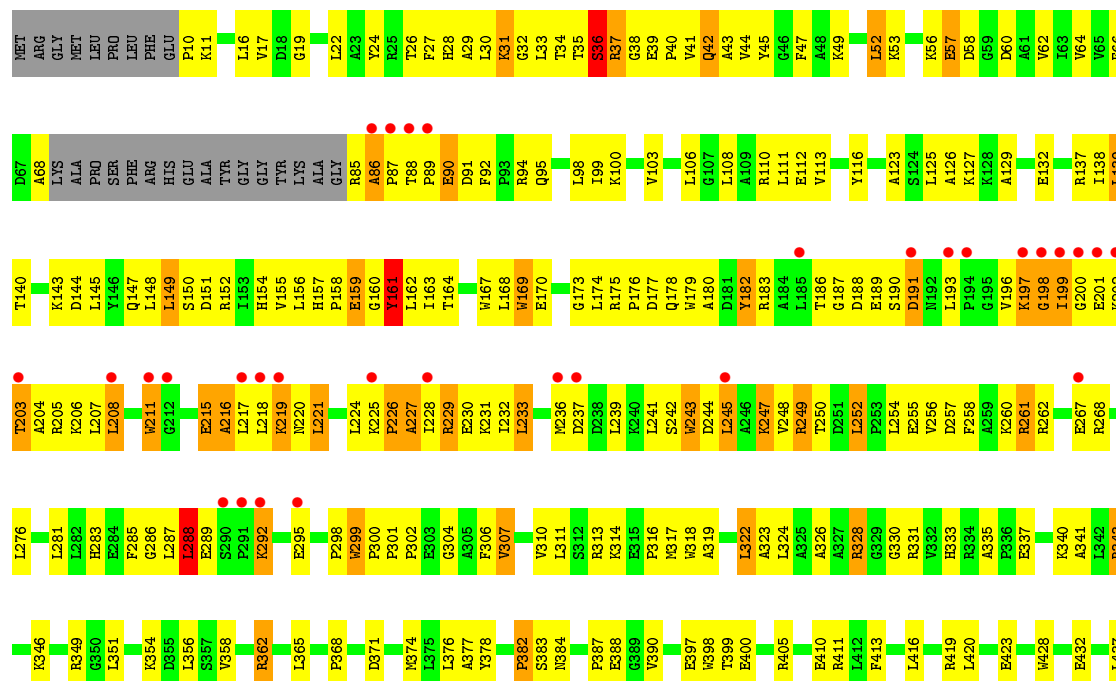
- Molecule 1: DNA (5'-D(\*GP\*CP\*GP\*AP\*TP\*CP\*CP\*G)-3')



- Molecule 2: DNA (5'-D(\*CP\*GP\*GP\*AP\*TP\*CP\*GP\*C)-3')



- Molecule 3: PROTEIN (TAQ POLYMERASE)



V806	E734	K663	Q582	LYS	S438
M607	A735	T664	M583	ARG	
E808	R736	I665	I584	SS13	A442
G809	V737	N666	P585	TS14	T447
V810	K738	F667	V586	SS15	G448
Y811	S739	G668	R587	AS16	V449
P812	V740	V669	T588	AS17	R450
	R741	L670	P589	V518	
V815	E742	G671	L590	L519	V453
P816	A743	G672	G591	E520	
L817	A744	M673	Q592	AS21	
	E745	S674	R593	L522	L456
E820	R746	A675	I594	R523	
V821	M747	H676	AS95	E524	A464
G822	A748	R677		AS25	E466
I823	F749		A600	H526	I467
			E601	P527	A468
D826	M750	Q680	E602	I528	R469
H827	M751	E681	G603	V529	L470
L828	P752	L682	M604	E530	E471
	V753	A683	L605	K531	A472
K831	Q754	I684	L606	I532	E473
E832	G755	P685	V607	L533	V474
	T756	Y686	A608	D534	F475
	A757	E687	L609	Y535	R476
	A758	E688	H610	R536	L477
	K762		Y611	E537	A478
		A691		L538	G479
	K767	F692	I614	TS39	H480
	L768	I693		K540	F481
	F769	E694	R617	L541	F482
	P770	R695	V618		M483
	R771		L619	TS44	L484
	E774	S699	G624	Y545	N485
		F700	D625	D547	S486
	R778	P701	E626	P548	R487
M779	K702	K702	I627	L549	D488
L780			L628	P550	Q489
L781	W706	W706	I629	D551	L490
Q782	I707	I707	R630	L552	E491
Q783	E708	E708	V631	I553	R492
V783	K709	K709	F632	H554	V493
H784			Q633	P555	L494
D785	E712	E712	E634	AS56	F495
	E713	E713	G635		D496
E790	G714	G714	R636	L560	E497
	R715	R715	D637	H561	L498
K793	R716	R716	I638	TS62	GLY
E794	R717	R717	H639		LEU
R795	G718	G718	T640	Q566	PRO
A796	Y719	Y719	E641	TS67	ALA
E797			T642	AS68	ILE
A798	I722	I722			GLY
V799	L723	L723	R651	H573	LYS
A800				L574	THR
R801	R727	R727	V654		GLU
L802	R728	R728	D655	P579	LYS
A803	Y729	Y729		M580	THR
K804			L733	L581	GLY
E805	L733	L733	M658		

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.60Å 107.60Å 170.01Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	8.00 – 3.00 45.46 – 3.00	Depositor EDS
% Data completeness (in resolution range)	(Not available) (8.00-3.00) 88.0 (45.46-3.00)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.11 (at 3.01Å)	Xtriage
Refinement program	X-PLOR	Depositor
R, $R_{free}$	0.244 , 0.305 0.267 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.6	Xtriage
Anisotropy	0.310	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 92.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.033 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	8122	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.56% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: ZN, BGL

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	T	0.74	0/180	1.08	0/276
2	P	0.87	0/180	0.87	0/276
3	A	0.46	0/6464	0.76	5/8749 (0.1%)
All	All	0.49	0/6824	0.78	5/9301 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	781	LEU	CA-CB-CG	6.48	130.21	115.30
3	A	151	ASP	N-CA-C	-5.34	96.58	111.00
3	A	568	ALA	N-CA-C	5.13	124.85	111.00
3	A	281	LEU	CA-CB-CG	5.11	127.04	115.30
3	A	191	ASP	N-CA-C	-5.09	97.27	111.00

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	T	161	20	91	36	0
2	P	161	20	91	57	0
3	A	6331	1408	6407	381	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	20	0	26	1	0
5	A	1	0	0	0	0
All	All	6674	1448	6615	441	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 33.

The worst 5 of 441 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)
2:P:954:DT:H2'	2:P:953:DC:C6	1.74	1.23
2:P:954:DT:H2'	2:P:953:DC:C5	1.77	1.20
1:T:905:DT:H4'	3:A:580:ASN:HB2	1.29	1.13
2:P:955:DA:H2''	2:P:954:DT:C7	1.78	1.12
1:T:906:DC:H4'	3:A:544:THR:HG23	1.19	1.10

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
3	A	785/832 (94%)	653 (83%)	93 (12%)	39 (5%)	<b>2</b> <b>12</b>

5 of 39 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	31	LYS
3	A	37	ARG
3	A	86	ALA
3	A	161	TYR
3	A	220	ASN

### 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
3	A	655/683 (96%)	572 (87%)	83 (13%)	<b>4</b> <b>19</b>

5 of 83 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	337	GLU
3	A	519	LEU
3	A	782	GLN
3	A	343	ARG
3	A	390	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	583	ASN
3	A	592	GLN
3	A	750	ASN
3	A	566	GLN
3	A	639	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](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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	BGL	A	1001	-	20,20,20	1.17	3 (15%)	24,25,25	2.59	12 (50%)

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	BGL	A	1001	-	-	3/11/31/31	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1001	BGL	O3-C3	-3.12	1.35	1.43
4	A	1001	BGL	O4-C4	-2.48	1.37	1.43
4	A	1001	BGL	C4-C3	2.32	1.58	1.52

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1001	BGL	O2-C2-C3	5.54	121.41	109.83
4	A	1001	BGL	C1-O5-C5	-5.36	103.55	113.66
4	A	1001	BGL	C6-C5-C4	4.20	122.83	113.00
4	A	1001	BGL	O3-C3-C4	-4.02	101.05	110.35
4	A	1001	BGL	C4-C3-C2	3.73	118.20	109.68

There are no chirality outliers.

All (3) torsion outliers are listed below:

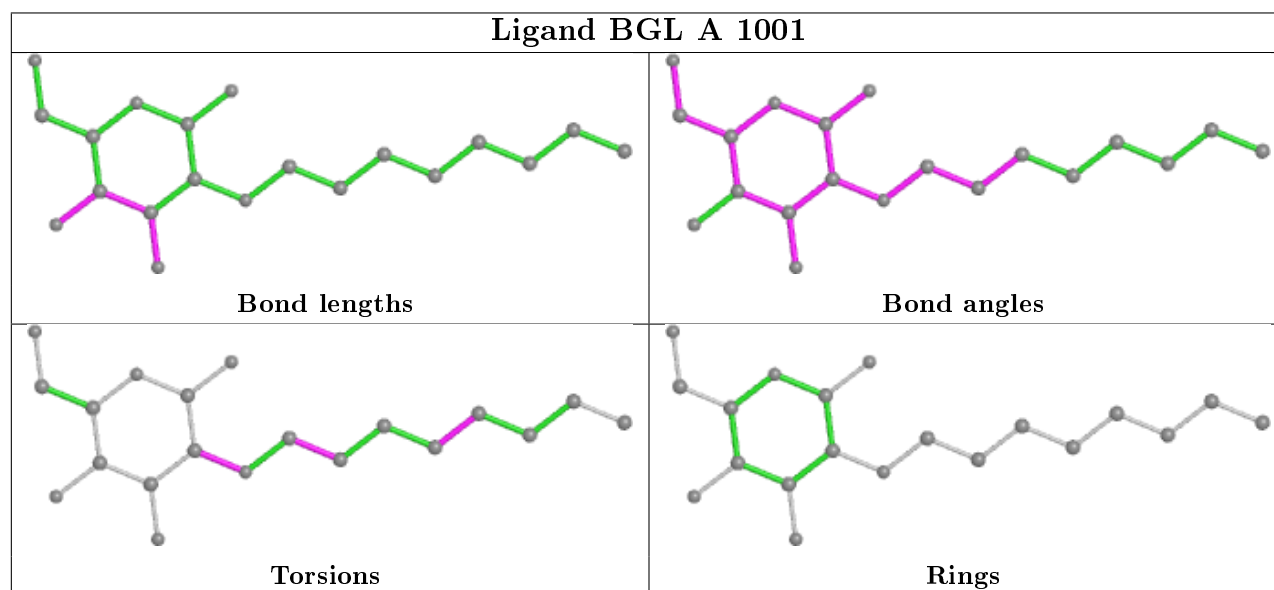
Mol	Chain	Res	Type	Atoms
4	A	1001	BGL	C3-C2-O2-C1'
4	A	1001	BGL	O2-C1'-C2'-C3'
4	A	1001	BGL	C3'-C4'-C5'-C6'

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1001	BGL	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	300:PRO	C	301:PRO	N	3.05

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	T	8/8 (100%)	3.00	6 (75%) 0 0	57, 59, 61, 61	0
2	P	8/8 (100%)	3.50	6 (75%) 0 0	57, 60, 62, 62	0
3	A	793/832 (95%)	0.14	51 (6%) 19 6	10, 37, 95, 100	0
All	All	809/848 (95%)	0.20	63 (7%) 13 4	10, 38, 95, 100	0

The worst 5 of 63 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	517	ALA	10.6
3	A	211	TRP	7.7
3	A	194	PRO	7.6
3	A	518	VAL	7.4
3	A	516	ALA	6.8

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

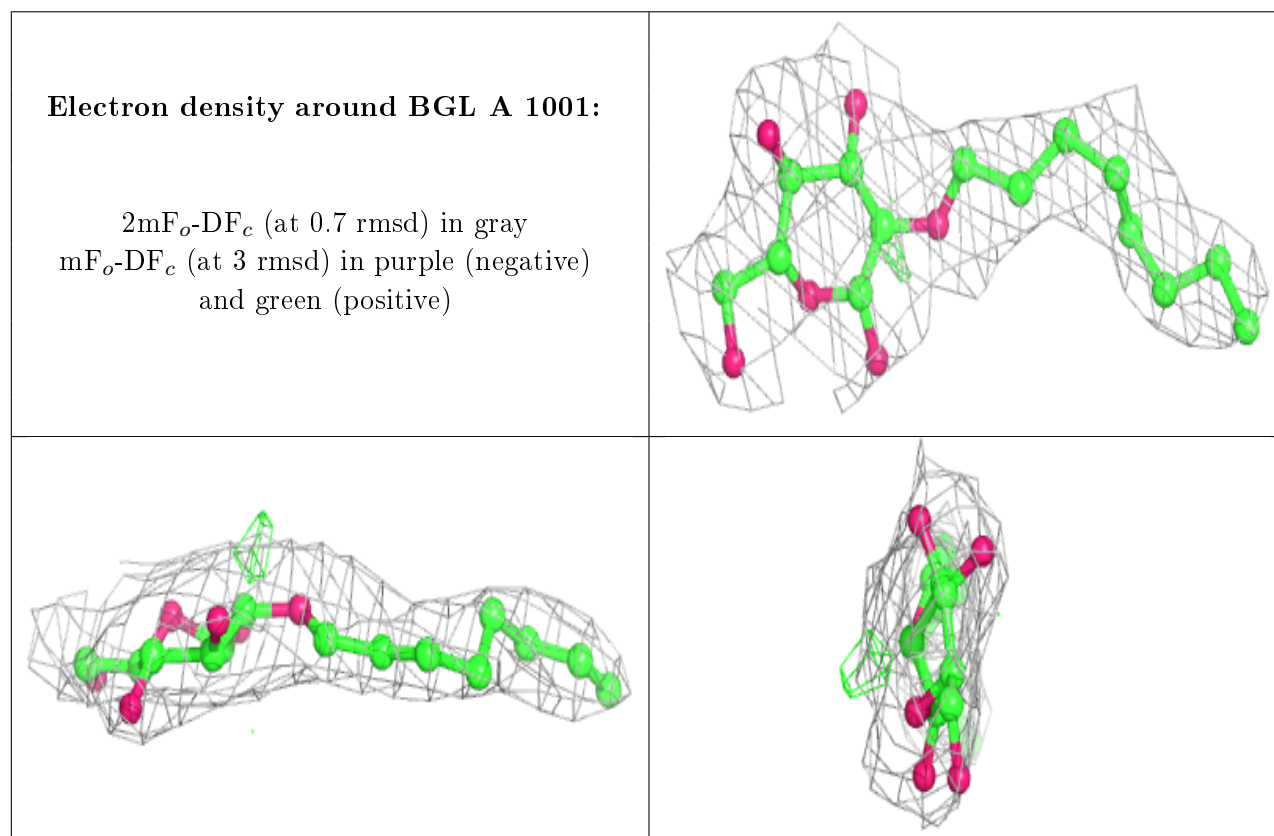
There are no monosaccharides in this entry.

### 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
4	BGL	A	1001	20/20	0.94	0.21	10,22,41,46	0
5	ZN	A	900	1/1	0.97	0.08	58,58,58,58	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



## 6.5 Other polymers [i](#)

There are no such residues in this entry.