



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 21, 2020 – 07:16 AM BST

PDB ID : 5A57  
Title : The structure of GH101 from Streptococcus pneumoniae TIGR4 in complex with PUGT  
Authors : Gregg, K.J.; Suits, M.D.L.; Deng, L.; Vocadlo, D.J.; Boraston, A.B.  
Deposited on : 2015-06-16  
Resolution : 1.46 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

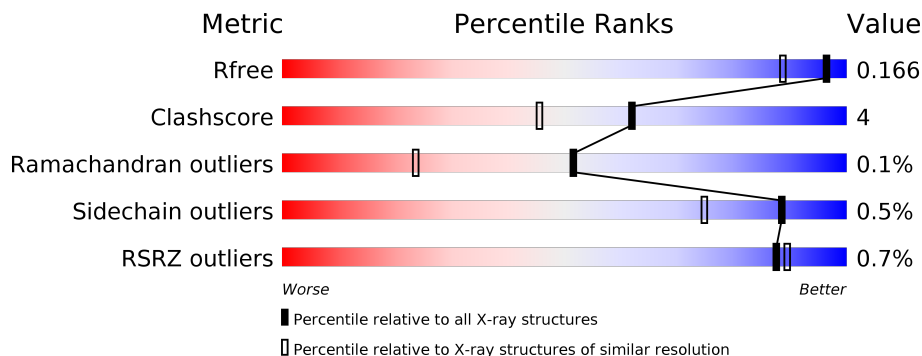
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.46 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	1156 (1.46-1.46)
Clashscore	141614	1202 (1.46-1.46)
Ramachandran outliers	138981	1178 (1.46-1.46)
Sidechain outliers	138945	1178 (1.46-1.46)
RSRZ outliers	127900	1139 (1.46-1.46)

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 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	A	1117	

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
3	CIT	A	2431	-	-	X	-
5	EDO	A	2440	-	-	X	-

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Type</b>	<b>Chain</b>	<b>Res</b>	<b>Chirality</b>	<b>Geometry</b>	<b>Clashes</b>	<b>Electron density</b>
5	EDO	A	2441	-	X	X	-
5	EDO	A	2448	-	-	X	-

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 11169 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 ENDO-ALPHA-N-ACETYLGALACTOSAMINIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1107	8983	5665	1535	1756	27	0	42	0

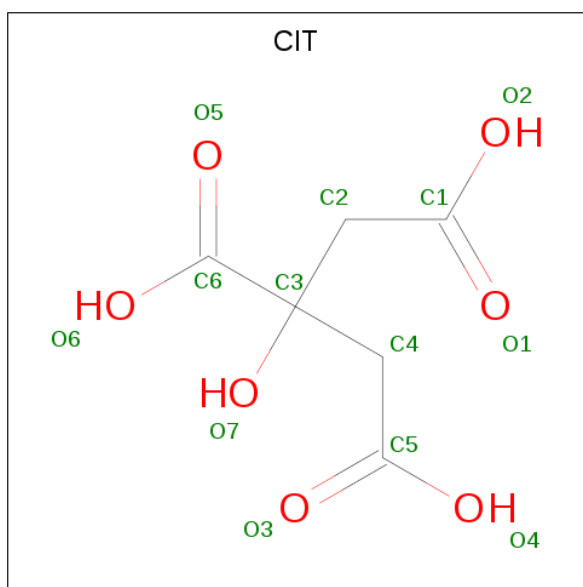
There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	316	MET	-	expression tag	UNP Q2MGH6
A	1427	GLU	-	expression tag	UNP Q2MGH6
A	1428	HIS	-	expression tag	UNP Q2MGH6
A	1429	HIS	-	expression tag	UNP Q2MGH6
A	1430	HIS	-	expression tag	UNP Q2MGH6
A	1431	HIS	-	expression tag	UNP Q2MGH6
A	1432	HIS	-	expression tag	UNP Q2MGH6
A	461	ASP	ASN	conflict	UNP Q2MGH6
A	1165	ASN	ASP	conflict	UNP Q2MGH6
A	1202	GLU	GLN	conflict	UNP Q2MGH6
A	1264	ASP	ASN	conflict	UNP Q2MGH6

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

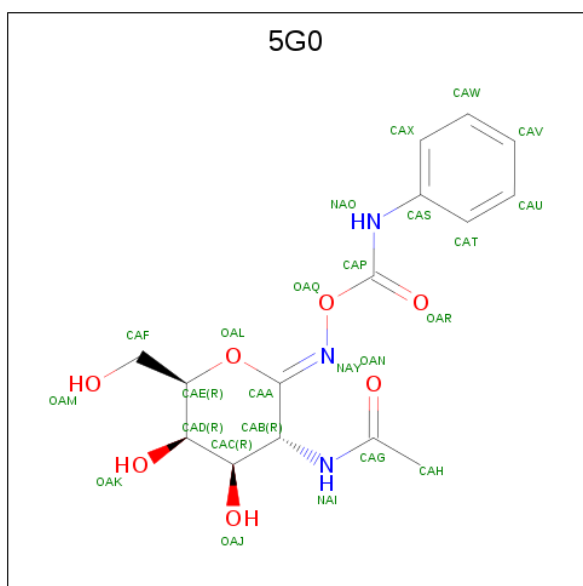
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Ca	0	0
			3	3		

- Molecule 3 is CITRIC ACID (three-letter code: CIT) (formula: C<sub>6</sub>H<sub>8</sub>O<sub>7</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			13	6	7		
3	A	1	Total	C	O	0	0
			13	6	7		

- Molecule 4 is (Z)-[(3R,4R,5R,6R)-3-acetamido-6-(hydroxymethyl)-4,5-bis(oxidanyl)oxan-2-ylidene]amino] N-phenylcarbamate (three-letter code: 5G0) (formula: C<sub>15</sub>H<sub>19</sub>N<sub>3</sub>O<sub>7</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	
4	A	1	Total	C	N	O	0	0
			25	15	3	7		

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



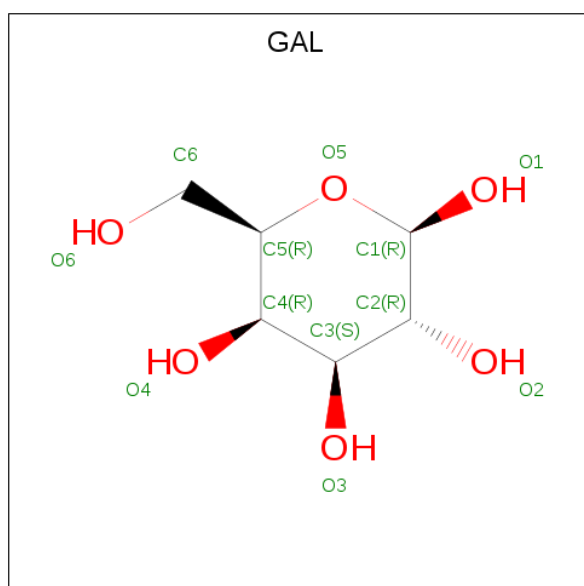
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0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

*Continued on next page...*

Continued from previous page...

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

- Molecule 6 is beta-D-galactopyranose (three-letter code: GAL) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total C O 11 6 5	0	0

- Molecule 7 is water.

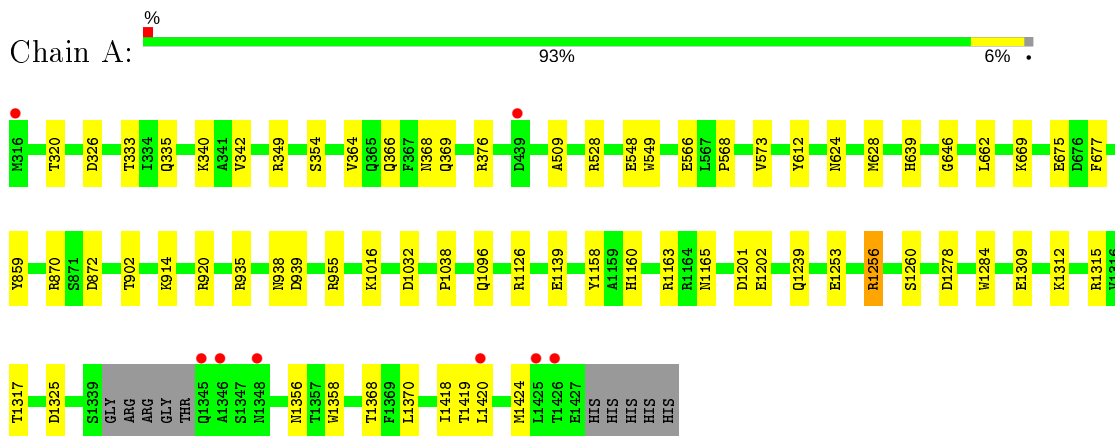
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
7	A	2029	Total 2029	O 2029	0	0



### 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: ENDO-ALPHA-N-ACETYL GALACTOSAMINIDASE



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	87.15Å 122.19Å 139.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.11 – 1.46 39.11 – 1.46	Depositor EDS
% Data completeness (in resolution range)	99.5 (39.11-1.46) 99.5 (39.11-1.46)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.99 (at 1.46Å)	Xtrriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.141 , 0.160 0.149 , 0.166	Depositor DCC
$R_{free}$ test set	12815 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	10.7	Xtrriage
Anisotropy	0.216	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 43.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	11169	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% 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: CA, GAL, EDO, 5G0, CIT

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	0.84	2/9314 (0.0%)	0.92	17/12609 (0.1%)

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

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	566	GLU	CD-OE2	-5.33	1.19	1.25
1	A	1358	TRP	CD2-CE2	5.09	1.47	1.41

All (17) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	349	ARG	NE-CZ-NH1	7.71	124.15	120.30
1	A	1325	ASP	CB-CG-OD1	7.66	125.19	118.30
1	A	349	ARG	NE-CZ-NH2	-7.06	116.77	120.30
1	A	1126	ARG	NE-CZ-NH1	6.84	123.72	120.30
1	A	939	ASP	CB-CG-OD1	6.66	124.30	118.30
1	A	528	ARG	NE-CZ-NH1	6.65	123.62	120.30
1	A	935	ARG	NE-CZ-NH1	6.02	123.31	120.30
1	A	1032	ASP	CB-CG-OD1	5.86	123.57	118.30
1	A	955	ARG	NE-CZ-NH2	-5.51	117.55	120.30
1	A	935	ARG	NE-CZ-NH2	-5.36	117.62	120.30
1	A	920[A]	ARG	NE-CZ-NH1	-5.32	117.64	120.30

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	920[B]	ARG	NE-CZ-NH1	-5.32	117.64	120.30
1	A	326	ASP	CB-CG-OD2	-5.30	113.53	118.30
1	A	1253	GLU	OE1-CD-OE2	-5.20	117.06	123.30
1	A	1278	ASP	CB-CG-OD1	5.08	122.87	118.30
1	A	1256[A]	ARG	NE-CZ-NH2	-5.02	117.79	120.30
1	A	1256[B]	ARG	NE-CZ-NH2	-5.02	117.79	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	859	TYR	Sidechain

## 5.2 Too-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 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	8983	0	8701	61	0
2	A	3	0	0	0	0
3	A	26	0	10	8	0
4	A	25	0	18	5	0
5	A	92	0	137	27	0
6	A	11	0	10	0	0
7	A	2029	0	0	25	3
All	All	11169	0	8876	79	3

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 4.

All (79) 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:A:1317[B]:THR:CG2	1:A:1368[B]:THR:HG22	1.69	1.22
1:A:1317[B]:THR:HG22	1:A:1368[B]:THR:CG2	1.71	1.20
1:A:1139[A]:GLU:OE1	7:A:4619:HOH:O	1.69	1.09

Continued on next page...

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:2431:CIT:O4	3:A:2431:CIT:O2	1.71	1.08
1:A:1418:ILE:HD12	1:A:1424[B]:MET:SD	1.94	1.05
1:A:1368[B]:THR:HG21	7:A:4938:HOH:O	1.62	0.98
4:A:2433:5G0:HAH2	4:A:2433:5G0:HAO	1.29	0.96
5:A:2440:EDO:H22	7:A:4672:HOH:O	1.68	0.92
5:A:2440:EDO:H12	7:A:4176:HOH:O	1.83	0.79
1:A:1418:ILE:CD1	1:A:1424[B]:MET:SD	2.70	0.78
1:A:1256[B]:ARG:NH2	7:A:3963:HOH:O	1.99	0.77
1:A:1201:ASP:H	3:A:2431:CIT:H41	1.49	0.76
5:A:2440:EDO:H21	7:A:4675:HOH:O	1.84	0.76
1:A:548[B]:GLU:OE1	7:A:3434:HOH:O	2.03	0.75
1:A:902:THR:H	1:A:938:ASN:HD21	1.34	0.74
1:A:335:GLN:NE2	7:A:3039:HOH:O	2.18	0.73
5:A:2436:EDO:O1	7:A:4402:HOH:O	2.04	0.73
1:A:675:GLU:HG3	5:A:2448:EDO:H12	1.73	0.69
5:A:2450:EDO:O1	7:A:5023:HOH:O	2.11	0.69
1:A:870:ARG:HB3	5:A:2441:EDO:H22	1.76	0.67
1:A:1096:GLN:HE21	5:A:2445:EDO:H22	1.62	0.65
4:A:2433:5G0:HAC	4:A:2433:5G0:HAH3	1.77	0.64
1:A:1038:PRO:HD2	5:A:2450:EDO:H11	1.79	0.63
1:A:1163[B]:ARG:NE	7:A:4671:HOH:O	2.28	0.63
1:A:639:HIS:NE2	5:A:2450:EDO:H12	2.14	0.62
1:A:1163[A]:ARG:NH1	5:A:2440:EDO:O1	2.32	0.61
1:A:1368[A]:THR:HG21	7:A:4938:HOH:O	2.01	0.59
1:A:914:LYS:NZ	7:A:4253:HOH:O	2.34	0.59
5:A:2440:EDO:H22	5:A:2441:EDO:H11	1.85	0.59
5:A:2447:EDO:O2	7:A:3963:HOH:O	1.81	0.57
1:A:1201:ASP:H	3:A:2431:CIT:C4	2.16	0.57
1:A:1418:ILE:HD12	1:A:1424[A]:MET:HE1	1.87	0.56
1:A:333[B]:THR:HG22	7:A:3039:HOH:O	2.05	0.56
1:A:1202:GLU:OE1	7:A:4714:HOH:O	2.17	0.55
1:A:1096:GLN:HE21	5:A:2445:EDO:C2	2.19	0.55
1:A:1315[B]:ARG:NH2	1:A:1424[B]:MET:SD	2.80	0.55
1:A:1317[B]:THR:HG22	1:A:1368[B]:THR:HG22	0.76	0.54
1:A:669:LYS:NZ	7:A:3774:HOH:O	2.40	0.53
4:A:2433:5G0:HAH1	7:A:4611:HOH:O	2.08	0.53
1:A:1418:ILE:HD12	1:A:1424[A]:MET:CE	2.39	0.53
1:A:675:GLU:CG	5:A:2448:EDO:H12	2.40	0.52
1:A:872:ASP:H	5:A:2441:EDO:H21	1.74	0.52
1:A:872:ASP:HB3	5:A:2441:EDO:H21	1.92	0.51
1:A:1201:ASP:N	3:A:2431:CIT:H41	2.21	0.50

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:624:ASN:HA	1:A:628[A]:MET:HE1	1.93	0.50
5:A:2440:EDO:C1	7:A:4176:HOH:O	2.50	0.49
1:A:366[A]:GLN:HG2	1:A:368:ASN:OD1	2.12	0.49
5:A:2440:EDO:C2	7:A:4672:HOH:O	2.43	0.49
1:A:870:ARG:CB	5:A:2441:EDO:H22	2.43	0.48
1:A:369:GLN:HE21	1:A:376:ARG:NE	2.10	0.48
1:A:366[A]:GLN:HG3	7:A:3055:HOH:O	2.14	0.48
1:A:1165[A]:ASN:ND2	7:A:4680:HOH:O	2.47	0.47
4:A:2433:5G0:CAC	4:A:2433:5G0:HAH3	2.45	0.46
1:A:364[B]:VAL:HG22	5:A:2438:EDO:H11	1.98	0.46
3:A:2431:CIT:C1	3:A:2431:CIT:O4	2.58	0.46
1:A:902:THR:N	1:A:938:ASN:HD21	2.06	0.46
1:A:369:GLN:NE2	1:A:376:ARG:CD	2.78	0.45
5:A:2454:EDO:H12	7:A:4359:HOH:O	2.15	0.45
1:A:1309:GLU:HB2	1:A:1312:LYS:HD2	1.99	0.44
1:A:1201:ASP:HB2	3:A:2431:CIT:H42	2.00	0.43
1:A:675:GLU:HA	5:A:2448:EDO:H12	2.00	0.43
1:A:675:GLU:CB	5:A:2448:EDO:H12	2.48	0.43
1:A:364[B]:VAL:HG22	5:A:2438:EDO:C1	2.48	0.43
1:A:333[A]:THR:HG22	1:A:342:VAL:HG22	2.02	0.42
1:A:612:TYR:HA	1:A:646:GLY:O	2.20	0.41
5:A:2440:EDO:H22	7:A:5011:HOH:O	2.20	0.41
1:A:1158:TYR:CE2	1:A:1160:HIS:HB2	2.56	0.41
1:A:1260:SER:O	1:A:1284:TRP:HA	2.20	0.41
1:A:1202:GLU:HG2	3:A:2431:CIT:O7	2.20	0.41
1:A:1016:LYS:NZ	7:A:4402:HOH:O	2.42	0.41
1:A:509:ALA:HA	1:A:573:VAL:O	2.21	0.41
1:A:872:ASP:H	5:A:2441:EDO:C2	2.33	0.41
1:A:1201:ASP:H	3:A:2431:CIT:C5	2.33	0.41
1:A:1370:LEU:HD22	1:A:1418:ILE:HD11	2.03	0.41
1:A:549:TRP:CD2	1:A:568:PRO:HB3	2.56	0.41
1:A:1420:LEU:CD2	1:A:1424[B]:MET:SD	3.09	0.40
4:A:2433:5G0:HAH2	4:A:2433:5G0:NAO	2.12	0.40
1:A:902:THR:H	1:A:938:ASN:ND2	2.10	0.40
1:A:340[B]:LYS:HE3	1:A:354[B]:SER:OG	2.21	0.40

All (3) 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)
7:A:3820:HOH:O	7:A:3820:HOH:O[2_555]	1.85	0.35

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
7:A:3017:HOH:O	7:A:4884:HOH:O[3_544]	2.01	0.19
7:A:3027:HOH:O	7:A:4976:HOH:O[3_544]	2.16	0.04

### 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	1146/1117 (103%)	1114 (97%)	31 (3%)	1 (0%)	51 24

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	662	LEU

#### 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	973/942 (103%)	967 (99%)	6 (1%)	86 69

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	320[A]	THR
1	A	320[B]	THR
1	A	677	PHE

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	1239	GLN
1	A	1356	ASN
1	A	1419	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	369	GLN
1	A	938	ASN
1	A	1096	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](#)

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 [i](#)

Of 30 ligands modelled in this entry, 3 are monoatomic - leaving 27 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	$\# Z  > 2$
5	EDO	A	2456	-	3,3,3	0.67	0	2,2,2	0.63	0
5	EDO	A	2440	-	3,3,3	0.73	0	2,2,2	0.85	0
6	GAL	A	2457	4	11,11,12	0.90	0	15,15,17	0.84	0



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
5	EDO	A	2442	-	3,3,3	0.49	0	2,2,2	0.34	0
5	EDO	A	2454	-	3,3,3	0.39	0	2,2,2	0.47	0
5	EDO	A	2441	-	3,3,3	1.85	1 (33%)	2,2,2	2.51	1 (50%)
5	EDO	A	2437	-	3,3,3	0.80	0	2,2,2	0.11	0
5	EDO	A	2450	-	3,3,3	0.39	0	2,2,2	0.67	0
5	EDO	A	2443	-	3,3,3	0.83	0	2,2,2	0.44	0
5	EDO	A	2447	-	3,3,3	0.36	0	2,2,2	0.46	0
5	EDO	A	2435	-	3,3,3	0.61	0	2,2,2	0.28	0
4	5G0	A	2433	6	24,26,26	2.79	3 (12%)	24,35,35	1.22	2 (8%)
5	EDO	A	2451	-	3,3,3	0.62	0	2,2,2	0.37	0
5	EDO	A	2452	-	3,3,3	0.48	0	2,2,2	0.34	0
5	EDO	A	2453	-	3,3,3	0.97	0	2,2,2	0.34	0
5	EDO	A	2445	-	3,3,3	0.85	0	2,2,2	0.55	0
5	EDO	A	2449	-	3,3,3	0.39	0	2,2,2	0.26	0
3	CIT	A	2431	-	3,12,12	2.47	2 (66%)	3,17,17	4.45	3 (100%)
5	EDO	A	2439	-	3,3,3	0.51	0	2,2,2	0.20	0
5	EDO	A	2434	-	3,3,3	0.38	0	2,2,2	0.87	0
5	EDO	A	2444	-	3,3,3	0.50	0	2,2,2	0.32	0
3	CIT	A	2432	-	3,12,12	4.39	3 (100%)	3,17,17	2.00	1 (33%)
5	EDO	A	2448	-	3,3,3	0.51	0	2,2,2	0.52	0
5	EDO	A	2455	-	3,3,3	0.33	0	2,2,2	0.48	0
5	EDO	A	2436	-	3,3,3	0.33	0	2,2,2	0.43	0
5	EDO	A	2438	-	3,3,3	0.40	0	2,2,2	0.38	0
5	EDO	A	2446	-	3,3,3	0.35	0	2,2,2	0.16	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
5	EDO	A	2456	-	-	0/1/1/1	-
5	EDO	A	2440	-	-	1/1/1/1	-
6	GAL	A	2457	4	-	0/2/19/22	0/1/1/1
5	EDO	A	2442	-	-	1/1/1/1	-
5	EDO	A	2454	-	-	0/1/1/1	-
5	EDO	A	2441	-	-	1/1/1/1	-
5	EDO	A	2437	-	-	0/1/1/1	-
5	EDO	A	2450	-	-	1/1/1/1	-
5	EDO	A	2443	-	-	0/1/1/1	-
5	EDO	A	2447	-	-	0/1/1/1	-
5	EDO	A	2435	-	-	0/1/1/1	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	5G0	A	2433	6	-	5/12/35/35	0/2/2/2
5	EDO	A	2451	-	-	0/1/1/1	-
5	EDO	A	2452	-	-	0/1/1/1	-
5	EDO	A	2453	-	-	0/1/1/1	-
5	EDO	A	2445	-	-	1/1/1/1	-
5	EDO	A	2449	-	-	0/1/1/1	-
3	CIT	A	2431	-	-	5/6/16/16	-
5	EDO	A	2439	-	-	0/1/1/1	-
5	EDO	A	2434	-	-	0/1/1/1	-
5	EDO	A	2444	-	-	0/1/1/1	-
3	CIT	A	2432	-	-	0/6/16/16	-
5	EDO	A	2448	-	-	1/1/1/1	-
5	EDO	A	2455	-	-	1/1/1/1	-
5	EDO	A	2436	-	-	1/1/1/1	-
5	EDO	A	2438	-	-	0/1/1/1	-
5	EDO	A	2446	-	-	0/1/1/1	-

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	2433	5G0	CAH-CAG	-10.35	1.29	1.50
4	A	2433	5G0	OAN-CAG	7.22	1.39	1.23
3	A	2432	CIT	C4-C3	-6.13	1.46	1.54
4	A	2433	5G0	OAQ-NAY	-3.75	1.35	1.44
3	A	2432	CIT	C2-C3	-3.75	1.49	1.54
3	A	2431	CIT	C4-C3	-3.43	1.50	1.54
5	A	2441	EDO	O1-C1	-2.84	1.27	1.42
3	A	2432	CIT	O7-C3	-2.50	1.39	1.43
3	A	2431	CIT	C2-C3	-2.20	1.51	1.54

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2431	CIT	C3-C2-C1	5.49	123.78	114.98
3	A	2431	CIT	C3-C4-C5	4.29	121.85	114.98
4	A	2433	5G0	CAB-NAI-CAG	3.58	127.95	121.84
5	A	2441	EDO	O2-C2-C1	-3.50	86.75	111.91
3	A	2431	CIT	C4-C3-C2	3.29	118.13	109.33
3	A	2432	CIT	C4-C3-C2	3.03	117.42	109.33
4	A	2433	5G0	OAL-CAE-CAD	2.26	113.75	109.73

There are no chirality outliers.

All (18) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2431	CIT	C1-C2-C3-O7
4	A	2433	5G0	OAN-CAG-NAI-CAB
4	A	2433	5G0	CAH-CAG-NAI-CAB
3	A	2431	CIT	C2-C3-C4-C5
5	A	2440	EDO	O1-C1-C2-O2
5	A	2436	EDO	O1-C1-C2-O2
5	A	2455	EDO	O1-C1-C2-O2
5	A	2442	EDO	O1-C1-C2-O2
4	A	2433	5G0	CAA-CAB-NAI-CAG
3	A	2431	CIT	C1-C2-C3-C6
3	A	2431	CIT	C6-C3-C4-C5
5	A	2445	EDO	O1-C1-C2-O2
3	A	2431	CIT	O7-C3-C4-C5
5	A	2450	EDO	O1-C1-C2-O2
5	A	2441	EDO	O1-C1-C2-O2
4	A	2433	5G0	CAT-CAS-NAO-CAP
4	A	2433	5G0	CAX-CAS-NAO-CAP
5	A	2448	EDO	O1-C1-C2-O2

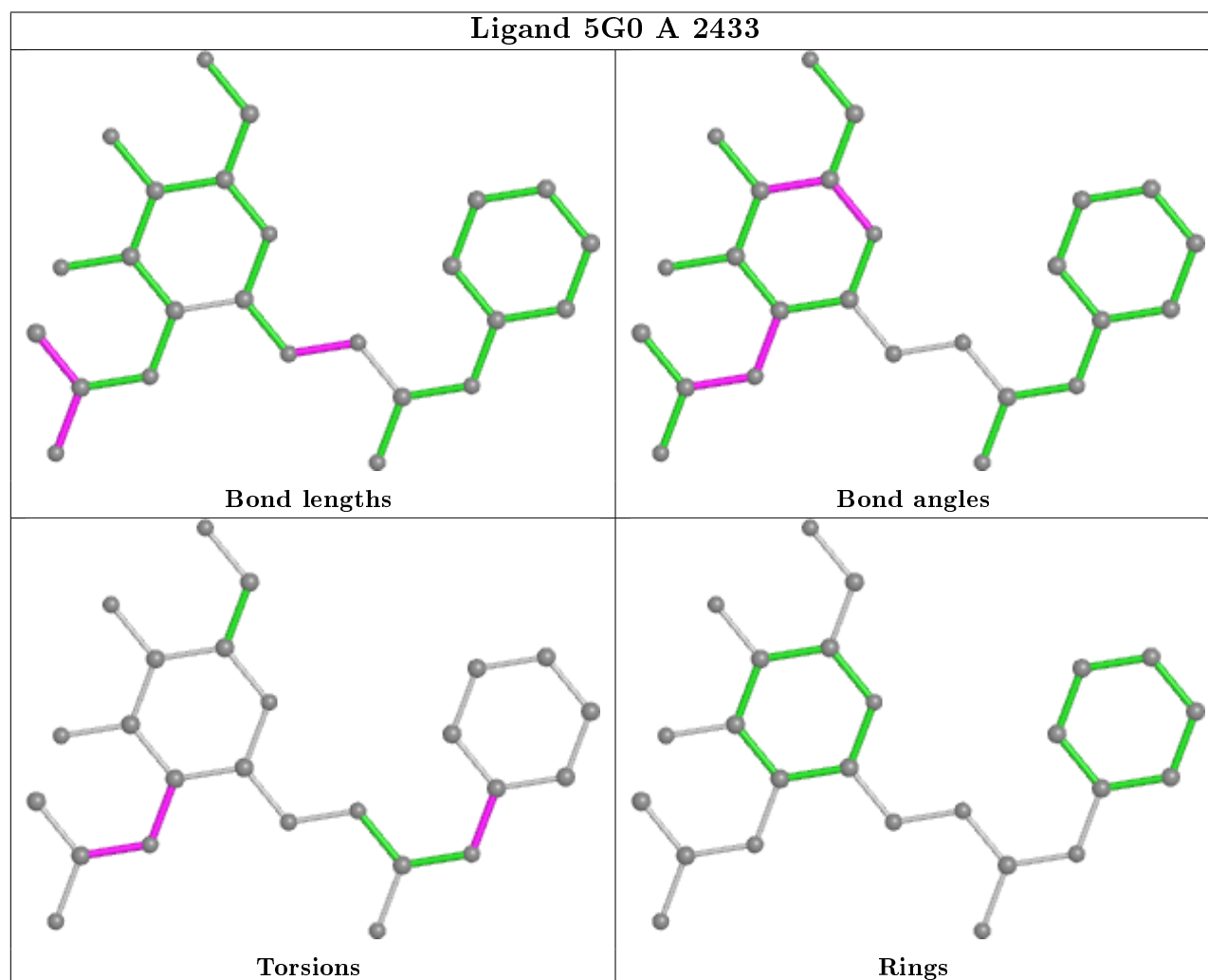
There are no ring outliers.

11 monomers are involved in 40 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	2440	EDO	8	0
5	A	2454	EDO	1	0
5	A	2441	EDO	6	0
5	A	2450	EDO	3	0
5	A	2447	EDO	1	0
4	A	2433	5G0	5	0
5	A	2445	EDO	2	0
3	A	2431	CIT	8	0
5	A	2448	EDO	4	0
5	A	2436	EDO	1	0
5	A	2438	EDO	2	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 [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, 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	A	1107/1117 (99%)	-0.39	8 (0%) <b>87</b> <b>89</b>	5, 10, 21, 50	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1425	LEU	4.0
1	A	1345	GLN	3.6
1	A	1346	ALA	3.5
1	A	316	MET	3.3
1	A	1420	LEU	2.7
1	A	1348	ASN	2.4
1	A	439	ASP	2.1
1	A	1426	THR	2.1

### 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](#)

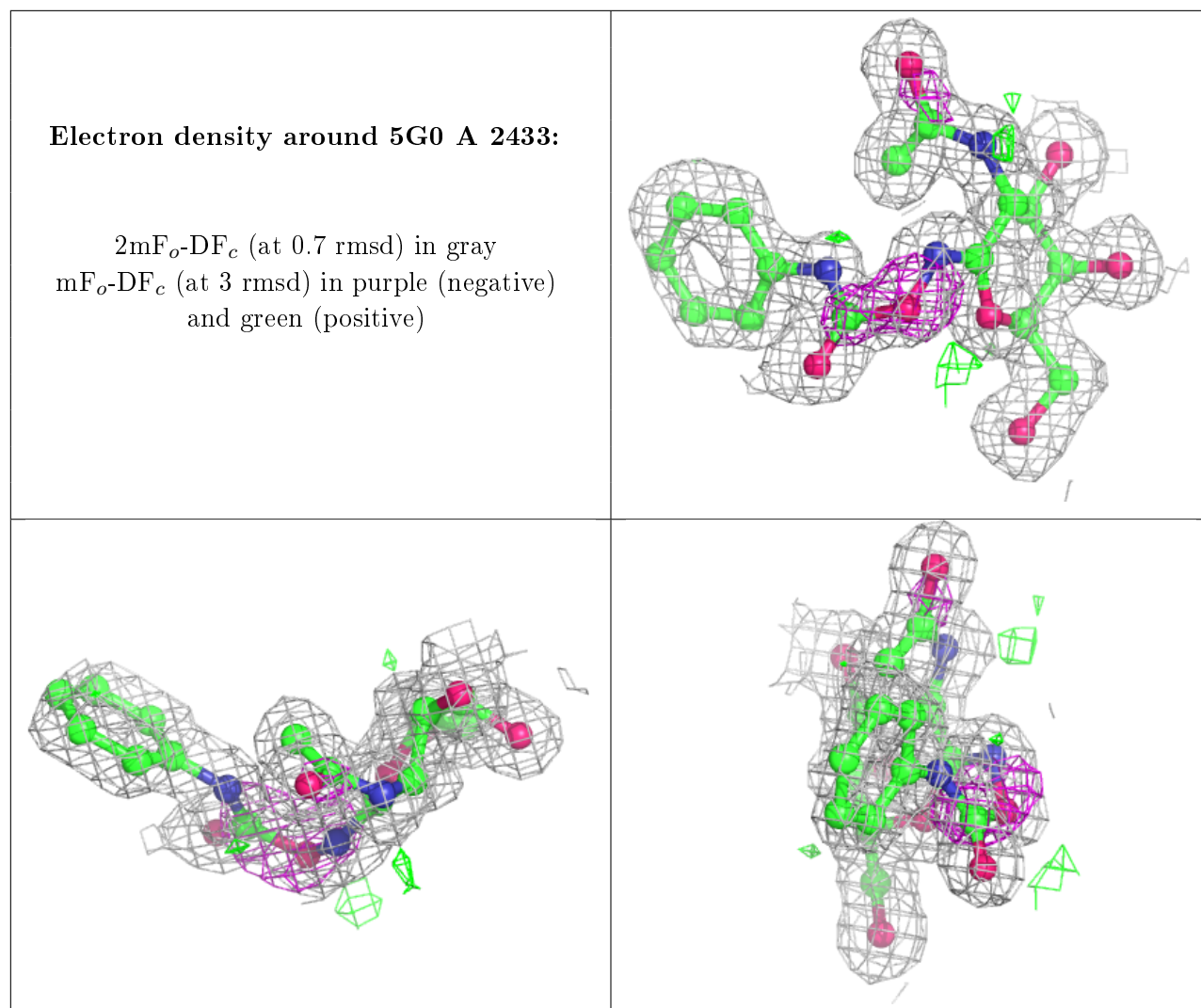
There are no monosaccharides in this entry.

### 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
5	EDO	A	2440	4/4	0.73	0.17	22,24,25,31	0
5	EDO	A	2455	4/4	0.78	0.22	39,41,42,45	0
5	EDO	A	2436	4/4	0.78	0.18	28,31,31,39	0
5	EDO	A	2445	4/4	0.80	0.37	26,30,31,44	0
5	EDO	A	2452	4/4	0.82	0.15	35,36,39,42	0
5	EDO	A	2450	4/4	0.83	0.16	18,25,29,30	0
5	EDO	A	2441	4/4	0.86	0.18	14,15,17,24	0
5	EDO	A	2442	4/4	0.86	0.20	25,28,29,34	0
5	EDO	A	2449	4/4	0.87	0.27	32,33,37,38	0
5	EDO	A	2435	4/4	0.88	0.13	22,25,26,30	0
5	EDO	A	2439	4/4	0.88	0.13	23,23,26,29	0
3	CIT	A	2431	13/13	0.90	0.23	13,29,41,42	0
5	EDO	A	2443	4/4	0.90	0.11	16,17,18,24	0
5	EDO	A	2434	4/4	0.91	0.13	21,23,23,25	0
5	EDO	A	2447	4/4	0.92	0.14	17,22,22,27	0
5	EDO	A	2448	4/4	0.93	0.21	28,28,29,30	0
5	EDO	A	2453	4/4	0.93	0.12	10,13,15,17	0
3	CIT	A	2432	13/13	0.93	0.15	14,19,24,26	0
5	EDO	A	2451	4/4	0.94	0.10	16,17,17,18	0
4	5G0	A	2433	25/25	0.94	0.12	9,12,31,33	0
5	EDO	A	2438	4/4	0.94	0.09	12,12,14,14	0
5	EDO	A	2456	4/4	0.95	0.13	10,11,14,15	0
5	EDO	A	2454	4/4	0.96	0.10	19,24,27,30	0
5	EDO	A	2444	4/4	0.97	0.07	17,18,19,20	0
5	EDO	A	2437	4/4	0.97	0.12	13,13,14,15	0
6	GAL	A	2457	11/12	0.98	0.05	8,9,11,11	0
5	EDO	A	2446	4/4	0.98	0.08	15,15,15,16	0
2	CA	A	2428	1/1	0.99	0.05	9,9,9,9	0
2	CA	A	2430	1/1	1.00	0.04	7,7,7,7	0
2	CA	A	2429	1/1	1.00	0.03	7,7,7,7	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.