



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 12:05 pm BST

PDB ID : 5NPF  
Title : Crystal structure of txGH116 (beta-glucosidase from *Thermoanaerobacterium xylolyticum*) in complex with beta Cyclophellitol Cyclosulfate probe ME594  
Authors : Wu, L.; Offen, W.A.; Breen, I.Z.; Davies, G.J.  
Deposited on : 2017-04-16  
Resolution : 1.38 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 : 2.11  
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) : 2.11

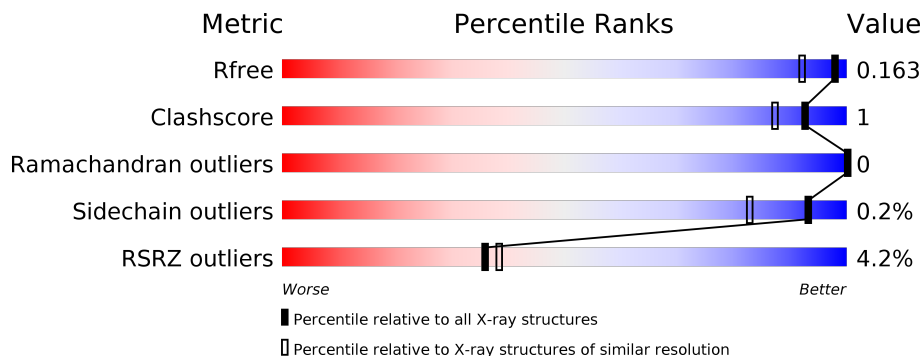
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.38 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	2907 (1.40-1.36)
Clashscore	141614	3037 (1.40-1.36)
Ramachandran outliers	138981	2970 (1.40-1.36)
Sidechain outliers	138945	2969 (1.40-1.36)
RSRZ outliers	127900	2846 (1.40-1.36)

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	798	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7059 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 Glucosylceramidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	763	6423	4143	1039	1214	27	0	29	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	17	MET	-	initiating methionine	UNP F6BL85
A	18	ALA	-	expression tag	UNP F6BL85
A	807	LEU	-	expression tag	UNP F6BL85
A	808	GLU	-	expression tag	UNP F6BL85
A	809	HIS	-	expression tag	UNP F6BL85
A	810	HIS	-	expression tag	UNP F6BL85
A	811	HIS	-	expression tag	UNP F6BL85
A	812	HIS	-	expression tag	UNP F6BL85
A	813	HIS	-	expression tag	UNP F6BL85
A	814	HIS	-	expression tag	UNP F6BL85

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

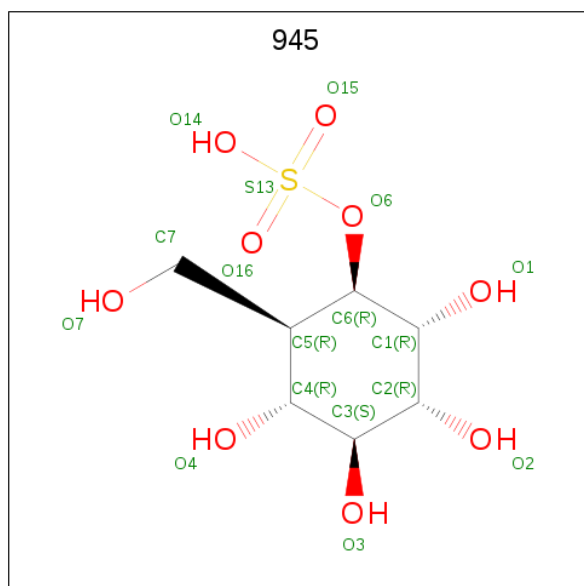


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0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

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Ca 1 1	0	0

- Molecule 4 is [(1 {R},2 {R},3 {R},4 {S},5 {R})-2-(hydroxymethyl)-3,4,5,6-tetrakis(oxidany l)cyclohexyl] hydrogen sulfate (three-letter code: 945) (formula: C<sub>7</sub>H<sub>14</sub>O<sub>9</sub>S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O S 16 7 8 1	0	0

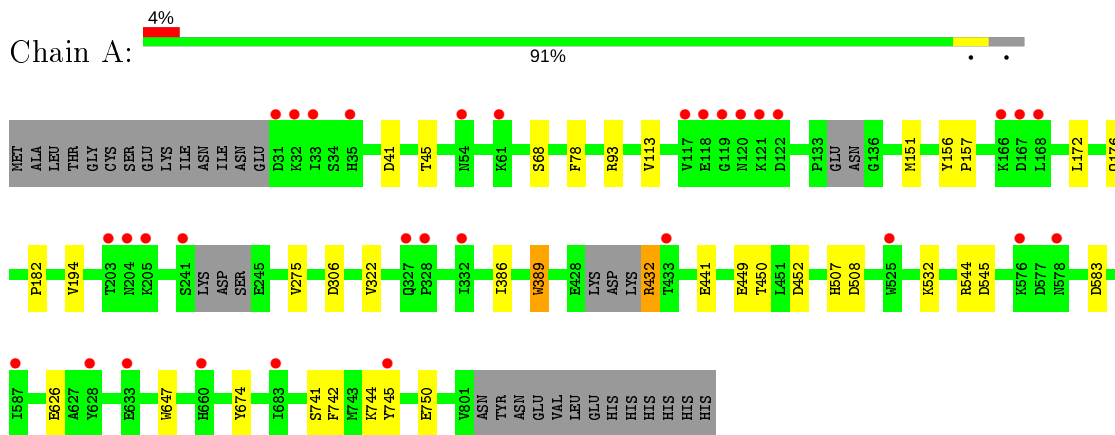
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	567	Total O 567 567	0	0

### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Glucosylceramidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	178.18Å 53.81Å 83.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	51.57 – 1.38 51.51 – 1.38	Depositor EDS
% Data completeness (in resolution range)	99.2 (51.57-1.38) 99.2 (51.51-1.38)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.48 (at 1.38Å)	Xtrriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.125 , 0.161 0.134 , 0.163	Depositor DCC
$R_{free}$ test set	8245 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	15.4	Xtrriage
Anisotropy	0.450	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 48.9	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	0.98	EDS
Total number of atoms	7059	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	19.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.65% 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, 945, EDO

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	0.82	4/6609 (0.1%)	0.91	13/8941 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	441	GLU	CD-OE2	9.63	1.36	1.25
1	A	68	SER	CB-OG	5.99	1.50	1.42
1	A	68	SER	CA-CB	5.44	1.61	1.52
1	A	626	GLU	CG-CD	-5.04	1.44	1.51

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	432	ARG	NE-CZ-NH1	11.84	126.22	120.30
1	A	432	ARG	NE-CZ-NH2	-9.80	115.40	120.30
1	A	742	PHE	CB-CG-CD2	-6.46	116.28	120.80
1	A	306	ASP	CB-CG-OD1	6.01	123.71	118.30
1	A	545	ASP	CB-CG-OD1	5.67	123.41	118.30

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	A	6423	0	6167	19	1
2	A	52	0	77	3	0
3	A	1	0	0	0	0
4	A	16	0	0	0	0
5	A	567	0	0	1	1
All	All	7059	0	6244	19	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 1.

The worst 5 of 19 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:432:ARG:HD2	2:A:909:EDO:H12	1.80	0.63
1:A:45[B]:THR:HG22	1:A:151:MET:HE3	1.85	0.58
1:A:744:LYS:NZ	5:A:1001:HOH:O	2.20	0.52
1:A:386:ILE:HA	1:A:389[A]:TRP:CD1	2.46	0.51
1:A:45[B]:THR:HG22	1:A:151:MET:CE	2.41	0.49

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:750:GLU:OE2	5:A:1359:HOH:O[2_675]	2.11	0.09

## 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	784/798 (98%)	757 (97%)	27 (3%)	0	<b>100</b> <b>100</b>

There are no Ramachandran outliers to report.

### 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	681/685 (99%)	680 (100%)	1 (0%)	<a href="#">93</a> <a href="#">83</a>

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	674	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 15 ligands modelled in this entry, 1 is monoatomic - leaving 14 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
2	EDO	A	911	-	3,3,3	0.21	0	2,2,2	0.42	0
2	EDO	A	910	-	3,3,3	1.18	0	2,2,2	0.64	0
2	EDO	A	909	-	3,3,3	0.65	0	2,2,2	0.25	0
2	EDO	A	904	-	3,3,3	0.51	0	2,2,2	0.47	0
4	945	A	915	-	16,16,17	1.84	3 (18%)	18,24,26	1.88	6 (33%)
2	EDO	A	901	-	3,3,3	0.17	0	2,2,2	0.87	0
2	EDO	A	913	-	3,3,3	0.83	0	2,2,2	0.29	0
2	EDO	A	905	-	3,3,3	0.40	0	2,2,2	0.70	0
2	EDO	A	902	-	3,3,3	1.03	0	2,2,2	1.71	1 (50%)
2	EDO	A	903	-	3,3,3	0.30	0	2,2,2	0.23	0
2	EDO	A	906	-	3,3,3	0.43	0	2,2,2	0.86	0
2	EDO	A	912	-	3,3,3	0.51	0	2,2,2	0.30	0
2	EDO	A	908	-	3,3,3	0.89	0	2,2,2	0.22	0
2	EDO	A	907	-	3,3,3	0.84	0	2,2,2	0.33	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
2	EDO	A	911	-	-	1/1/1/1	-
2	EDO	A	910	-	-	0/1/1/1	-
2	EDO	A	909	-	-	0/1/1/1	-
2	EDO	A	904	-	-	1/1/1/1	-
4	945	A	915	-	-	0/7/27/31	0/1/1/1
2	EDO	A	901	-	-	0/1/1/1	-
2	EDO	A	913	-	-	0/1/1/1	-
2	EDO	A	905	-	-	0/1/1/1	-
2	EDO	A	902	-	-	1/1/1/1	-
2	EDO	A	903	-	-	0/1/1/1	-
2	EDO	A	906	-	-	1/1/1/1	-
2	EDO	A	912	-	-	1/1/1/1	-
2	EDO	A	908	-	-	0/1/1/1	-
2	EDO	A	907	-	-	0/1/1/1	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	915	945	C1-C6	4.47	1.60	1.52
4	A	915	945	O6-C6	-4.17	1.40	1.47
4	A	915	945	C2-C3	2.12	1.55	1.52

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	915	945	C2-C1-C6	-3.63	102.78	111.21
4	A	915	945	C7-C5-C6	3.29	116.48	111.46
4	A	915	945	O14-S13-O16	3.17	119.52	108.49
4	A	915	945	O6-C6-C1	3.00	114.91	108.54
4	A	915	945	C1-C2-C3	-2.60	106.90	110.69

There are no chirality outliers.

All (5) torsion outliers are listed below:

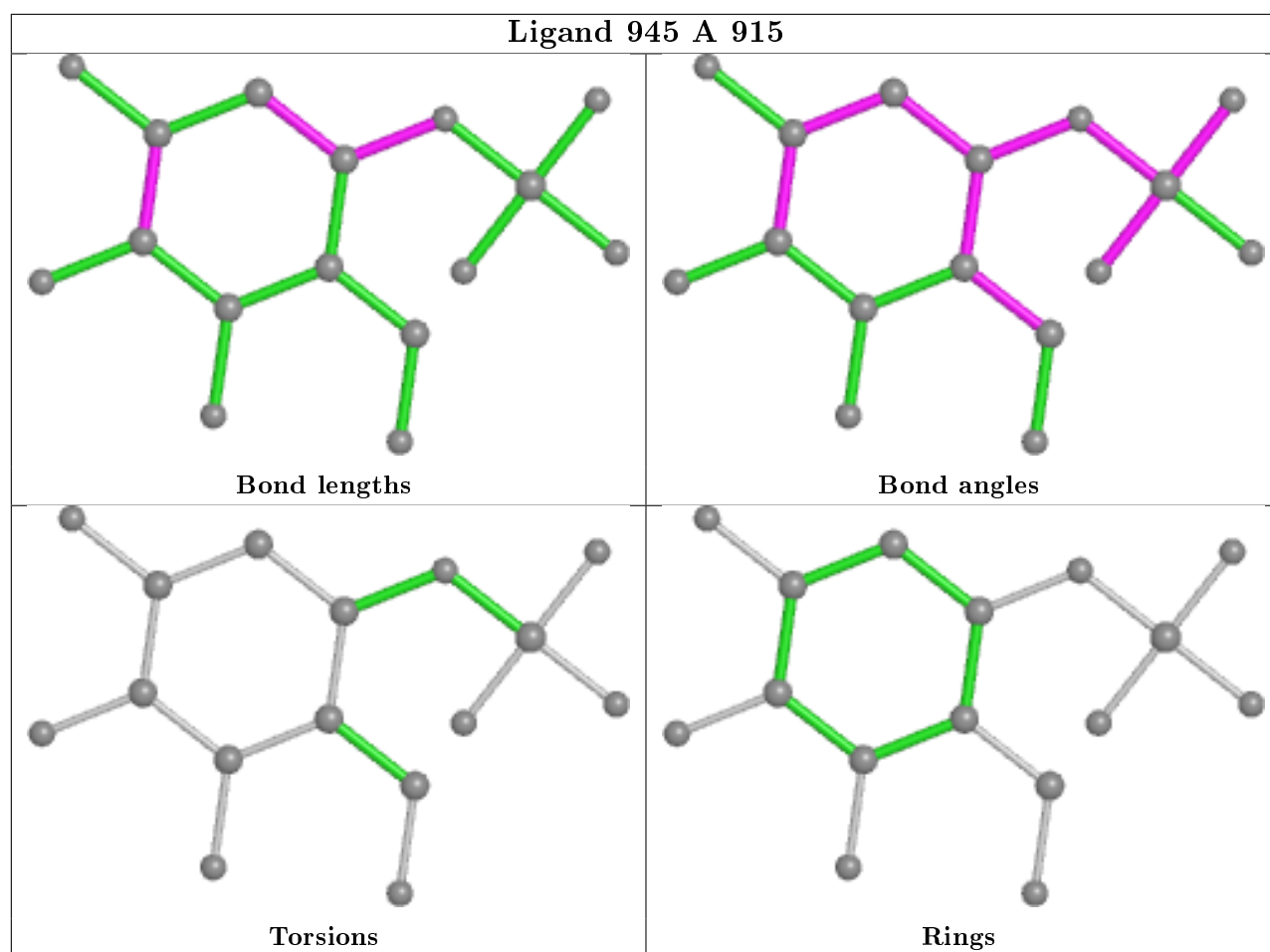
Mol	Chain	Res	Type	Atoms
2	A	906	EDO	O1-C1-C2-O2
2	A	911	EDO	O1-C1-C2-O2
2	A	904	EDO	O1-C1-C2-O2
2	A	912	EDO	O1-C1-C2-O2
2	A	902	EDO	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	911	EDO	1	0
2	A	909	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	763/798 (95%)	0.48	32 (4%) 36 38	9, 16, 36, 76	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	35	HIS	5.5
1	A	119	GLY	5.5
1	A	121	LYS	4.4
1	A	204[A]	ASN	4.2
1	A	433	THR	4.1

### 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 carbohydrates 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.

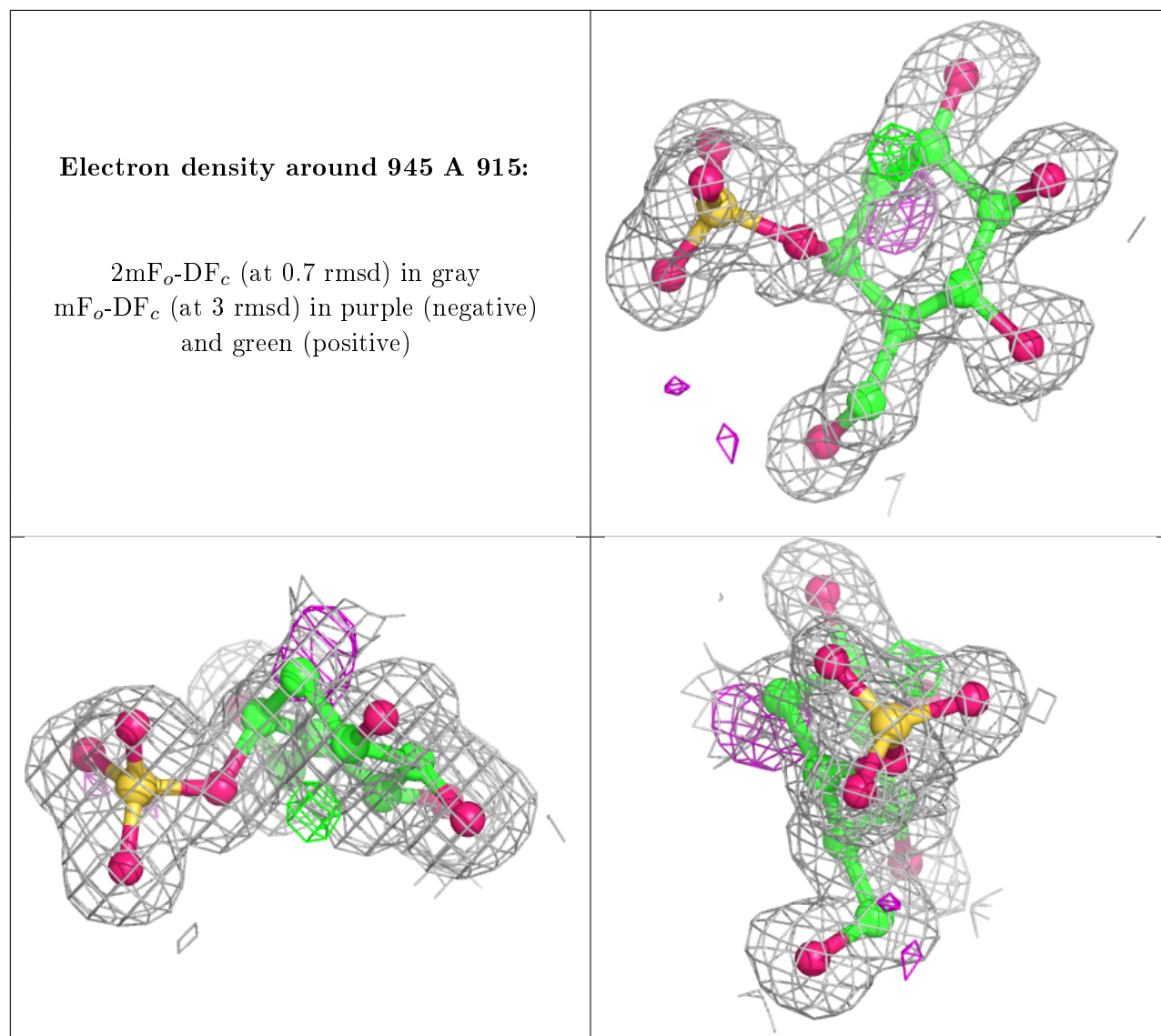
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	EDO	A	911	4/4	0.77	0.17	39,39,45,50	0
2	EDO	A	904	4/4	0.79	0.16	34,35,42,44	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	EDO	A	909	4/4	0.80	0.11	36,38,38,44	0
2	EDO	A	906	4/4	0.81	0.26	29,35,45,53	0
2	EDO	A	905	4/4	0.85	0.23	38,40,49,52	0
2	EDO	A	912	4/4	0.85	0.35	40,40,44,66	0
2	EDO	A	913	4/4	0.86	0.17	35,40,41,45	0
2	EDO	A	901	4/4	0.87	0.11	26,28,31,36	0
2	EDO	A	902	4/4	0.89	0.14	20,21,28,28	0
2	EDO	A	907	4/4	0.94	0.09	22,28,30,32	0
2	EDO	A	908	4/4	0.95	0.12	15,16,18,19	0
4	945	A	915	16/17	0.95	0.09	14,16,24,26	0
2	EDO	A	903	4/4	0.96	0.21	28,32,33,35	0
2	EDO	A	910	4/4	0.97	0.12	13,19,19,21	0
3	CA	A	914	1/1	1.00	0.05	25,25,25,25	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.