

# wwPDB X-ray Structure Validation Summary Report (i)

#### Aug 10, 2020 – 08:04 AM BST

PDB ID : 4CJC

Title: orthorhombic crystal form of Bogt6a E192Q in complex with UDP-GalNAc,

UDP, GalNAc

Authors: Pham, T.; Stinson, B.; Thiyagarajan, N.; Lizotte-Waniewski, M.; Brew, K.;

Acharya, K.R.

Deposited on : 2013-12-19

Resolution : 3.42 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

The following versions of software and data (see references (1)) were used in the production of this report:

 $Mol Probity \quad : \quad 4.02b\text{--}467$ 

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)

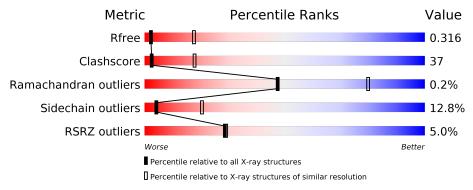
Validation Pipeline (wwPDB-VP) : 2.13.1

## 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: X- $RAY\ DIFFRACTION$ 

The reported resolution of this entry is 3.42 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 $(\# \mathrm{Entries})$	$\begin{array}{c} {\rm Similar \; resolution} \\ (\#{\rm Entries, \; resolution \; range(\AA)}) \end{array}$
$R_{free}$	130704	1486 (3.50-3.34)
Clashscore	141614	1572 (3.50-3.34)
Ramachandran outliers	138981	1534 (3.50-3.34)
Sidechain outliers	138945	1535 (3.50-3.34)
RSRZ outliers	127900	1395 (3.50-3.34)

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 >=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% 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	246	% 	41%	9% •			
1	В	246	51%	39%	7% •			
1	С	246	5%	35%	12% •			
1	D	246	13%	48%	7% •			



# 2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8120 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 GLYCOSYLTRANSFERASE FAMILY 6.

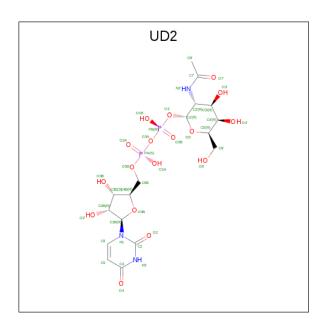
Mol	Chain	Residues	${f Atoms}$					ZeroOcc	AltConf	Trace
1	Λ	241	Total	С	N	О	S	7	0	0
1	A	∠41 	2008	1312	327	362	7	1	U	
1	В	238	Total	С	N	О	S	10	0	0
1	Б	256	1981	1295	322	357	7			
1	С	240	Total C I	N	О	S	10	0	0	
1		240	2000	1306	326	361	7	10	0	
1	1 D	D 236	Total	С	N	О	S	10	0	0
1	ש	∠30	1973	1291	320	355	7	10	U	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	192	GLN	GLU	engineered mutation	UNP A7LVT2
В	192	GLN	GLU	engineered mutation	UNP A7LVT2
С	192	GLN	GLU	engineered mutation	UNP A7LVT2
D	192	GLN	GLU	engineered mutation	UNP A7LVT2

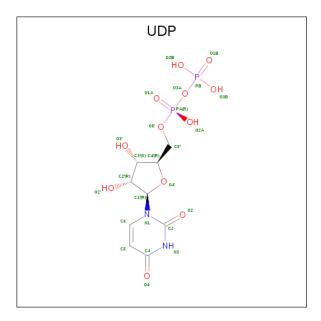
• Molecule 2 is URIDINE-DIPHOSPHATE-N-ACETYLGALACTOSAMINE (three-letter code: UD2) (formula: C<sub>17</sub>H<sub>27</sub>N<sub>3</sub>O<sub>17</sub>P<sub>2</sub>).





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
9	Λ.	1	Total	С	N	О	Р	0	0
2	$\mathcal{L}$ $\mathcal{A}$	1	39	17	3	17	2	0	
9	С	1	Total	С	N	О	Р	0	0
2		1	39	17	3	17	2	0	

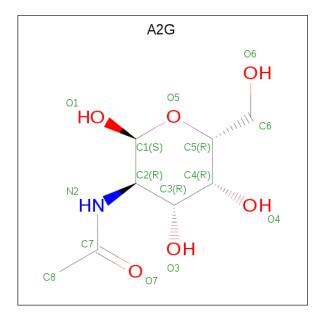
 $\bullet \ \ \mathrm{Molecule\ 3\ is\ URIDINE-5'-DIPHOSPHATE\ (three-letter\ code:\ UDP)\ (formula:\ C_9H_{14}N_2O_{12}P_2)}.$ 



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf		
2	D	1	Total	С	N	О	Р	0	0	
3	3   B	1	25	9	2	12	2	0		
9	D	1	Total	С	N	О	Р	0	0	
)	3 D	1	25	9	2	12	2	U		



• Molecule 4 is 2-acetamido-2-deoxy-alpha-D-galactopyranose (three-letter code: A2G) (formula:  $C_8H_{15}NO_6$ ).



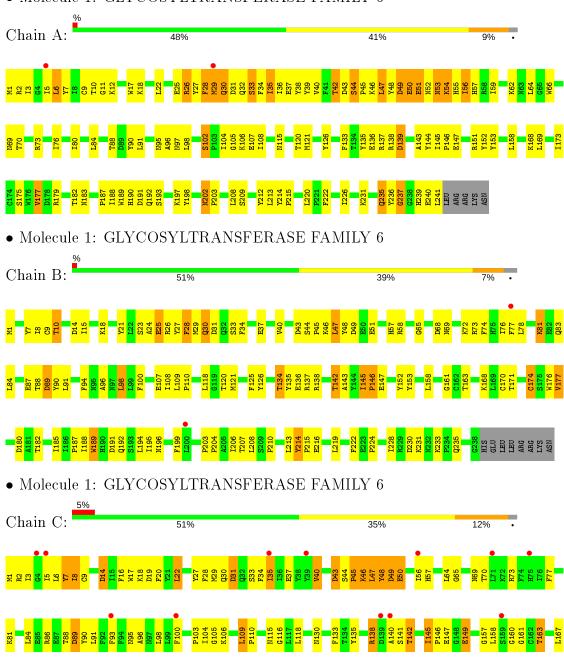
Mol	Chain	Residues	${f Atoms}$			ZeroOcc	AltConf
4	В	1	Total 15			0	0
4	D	1	Total 15		N 1	0	0



# 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.

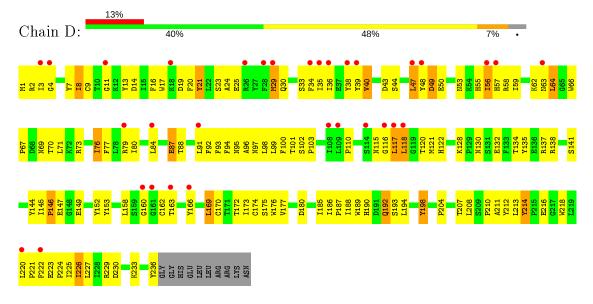
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6







• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	80.12Å 120.15Å 131.83Å	Danagitan
a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	88.80 - 3.42	Depositor
Resolution (A)	88.80 - 3.42	EDS
% Data completeness	97.0 (88.80-3.42)	Depositor
(in resolution range)	90.4 (88.80-3.42)	EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.53 (at 3.41Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
D D.	0.283 , 0.314	Depositor
$R, R_{free}$	0.288 , 0.316	DCC
$R_{free}$ test set	878 reflections $(5.06\%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	84.6	Xtriage
Anisotropy	0.398	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.27, 53.0	EDS
L-test for twinning <sup>2</sup>	$ < L >=0.48, < L^2>=0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	8120	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 4.17% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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: UD2, UDP, A2G

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		
MIOI	Chain	RMSZ	# Z >5	RMSZ	# Z  > 5	
1	A	0.50	0/2074	0.65	$2/2816 \ (0.1\%)$	
1	В	0.47	0/2046	0.65	$4/2778 \ (0.1\%)$	
1	С	0.38	0/2066	0.62	0/2805	
1	D	0.36	0/2038	0.67	$2/2768 \; (0.1\%)$	
All	All	0.43	0/8224	0.65	8/11167 (0.1%)	

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$Observed(^o)$	$\mathbf{Ideal}(^o)$
1	A	49	ASP	N-CA-C	-8.23	88.78	111.00
1	В	47	LEU	CA-CB-CG	-7.24	98.66	115.30
1	D	169	LEU	CA-CB-CG	6.82	130.98	115.30
1	A	237	GLY	N-CA-C	-6.36	97.20	113.10
1	В	33	SER	CB-CA-C	-6.03	98.64	110.10

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	2008	0	1944	201	0
1	В	1981	0	1920	100	0



$\alpha \cdots$	· ·	•	
Continued	trom	mromanne	maaa
-	110116	DICUIUU	$Du_iu_{C}$

Mol	Chain	Non-H	$\mathbf{H}(\mathbf{model})$	$\mathbf{H}(\mathbf{added})$	Clashes	Symm-Clashes
1	С	2000	0	1933	130	0
1	D	1973	0	1914	157	0
2	A	39	0	25	9	0
2	С	39	0	25	5	0
3	В	25	0	11	2	0
3	D	25	0	11	2	0
4	В	15	0	12	0	0
4	D	15	0	12	1	0
All	All	8120	0	7807	586	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 37.

The worst 5 of 586 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{aligned} &  ext{Interatomic} \ &  ext{distance} \ &  ext{(Å)} \end{aligned}$	$egin{array}{c}  ext{Clash} \  ext{overlap } ( ext{Å}) \end{array}$
1:D:29:MET:HG3	1:D:34:PHE:CZ	1.33	1.63
1:C:48:TYR:CD1	1:C:49:ASP:HB3	1.41	1.52
1:A:3:ILE:HD12	1:A:29:MET:CE	1.41	1.46
1:A:188:ILE:HD13	1:A:189:TRP:CD1	1.49	1.44
1:A:3:ILE:CD1	1:A:29:MET:HE3	1.42	1.44

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		
1	A	$239/246 \ (97\%)$	222 (93%)	17 (7%)	0	100	100	
1	В	$236/246 \ (96\%)$	219 (93%)	16 (7%)	1 (0%)	34	69	
1	С	238/246 (97%)	219 (92%)	19 (8%)	0	100	100	



Continued from previous page...

Mol	Chain	Analysed Favoured		d Allowed Outliers		Percentiles	
1	D	234/246 (95%)	218 (93%)	15 (6%)	1 (0%)	34 69	
All	All	947/984 (96%)	878 (93%)	67 (7%)	2 (0%)	47 80	

#### All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	146	PRO
1	В	146	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	Perce	entiles
1	A	$216/221 \; (98\%)$	184 (85%)	32 (15%)	3	15
1	В	$213/221 \ (96\%)$	194 (91%)	19 (9%)	9	37
1	С	215/221 (97%)	177 (82%)	38 (18%)	2	8
1	D	$213/221 \ (96\%)$	192 (90%)	21 (10%)	8	32
All	All	857/884 (97%)	747 (87%)	110 (13%)	4	20

5 of 110 residues with a non-rotameric sidechain are listed below:

Mol	Chain	${f Res}$	Type
1	В	206	ILE
1	С	45	PRO
1	D	117	LEU
1	С	1	MET
1	С	22	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

$\mathbf{Mol}$	Chain	${f Res}$	$\mathbf{Type}$
1	D	53	ASN
1	D	196	ASN



Continued from previous page...

Mol	Chain	Res	Type
1	D	69	ASN
1	A	202	ASN
1	D	130	ASN

#### 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)

6 ligands 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	Mal Thurs Chain I		Dec	Link	Bond lengths			Bond angles		
MIGI	Type	Chain	m Res	Link	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	A2G	D	1238	-	15,15,15	0.42	0	21,21,21	0.39	0
4	A2G	В	1240	_	15,15,15	0.41	0	21,21,21	0.66	0
2	UD2	A	1242	-	34,41,41	0.67	0	45,62,62	1.13	4 (8%)
3	UDP	D	1237	-	20,26,26	0.83	1 (5%)	25,40,40	1.43	2 (8%)
2	UD2	С	1241	-	34,41,41	0.66	0	45,62,62	1.47	7 (15%)
3	UDP	В	1239	-	20,26,26	0.90	1 (5%)	25,40,40	1.26	2 (8%)

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	A2G	D	1238	-	-	4/6/26/26	0/1/1/1
4	A2G	В	1240	-	-	3/6/26/26	0/1/1/1
2	UD2	A	1242	-	-	14/24/63/63	0/3/3/3
3	UDP	D	1237	-	-	2/14/32/32	0/2/2/2
2	UD2	С	1241	-	-	6/24/63/63	0/3/3/3
3	UDP	В	1239	-	-	8/14/32/32	0/2/2/2

#### All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$\operatorname{Observed}( ext{\AA})$	$\operatorname{Ideal}( ext{\AA})$
3	В	1239	UDP	C2-N3	-2.37	1.33	1.38
3	D	1237	UDP	C2-N3	-2.17	1.33	1.38

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$\mathbf{Observed}(^o)$	$\mathbf{Ideal}(^o)$
3	D	1237	UDP	C3'-C2'-C1'	4.45	107.68	100.98
2	С	1241	UD2	O5'-C5'-C4'	4.02	117.00	109.69
2	С	1241	UD2	C3B-C2B-C1B	3.61	106.42	100.98
3	D	1237	UDP	PA-O3A-PB	-3.53	120.73	132.83
3	В	1239	UDP	PA-O3A-PB	-3.27	121.61	132.83

There are no chirality outliers.

5 of 37 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	1238	A2G	O7-C7-N2-C2
4	D	1238	A2G	C8-C7-N2-C2
4	В	1240	A2G	O7-C7-N2-C2
4	В	1240	A2G	C8-C7-N2-C2
2	A	1242	UD2	C3'-C2'-N2'-C7'

There are no ring outliers.

5 monomers are involved in 19 short contacts:

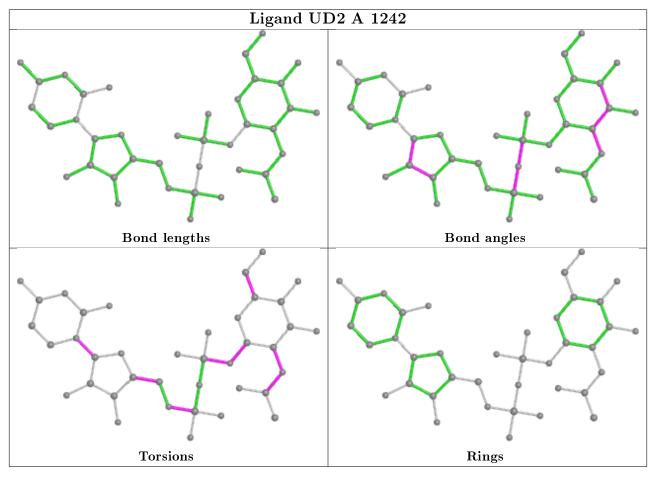
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	1238	A2G	1	0
2	A	1242	UD2	9	0
3	D	1237	UDP	2	0



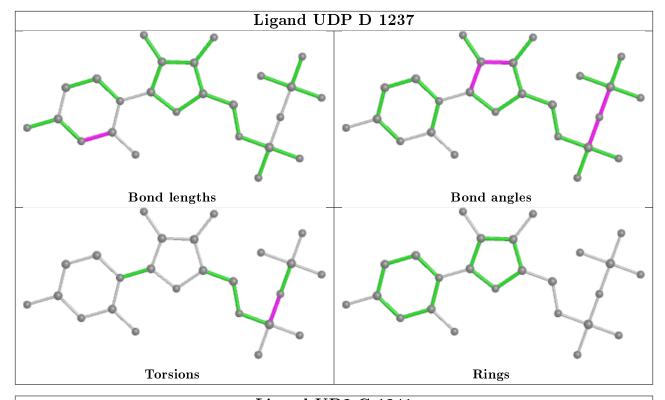
Continued from previous page...

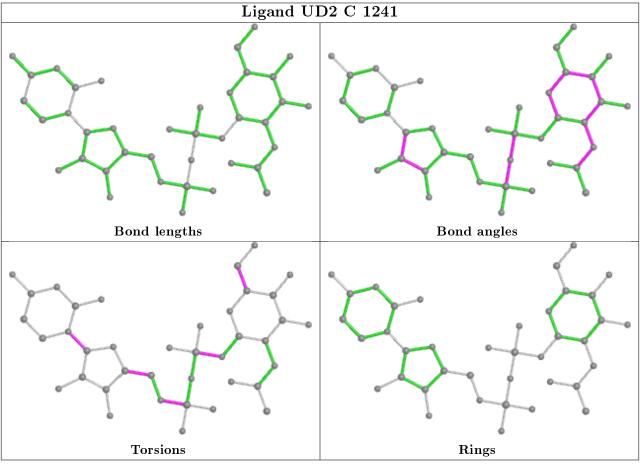
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	С	1241	UD2	5	0
3	В	1239	UDP	2	0

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 then 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.

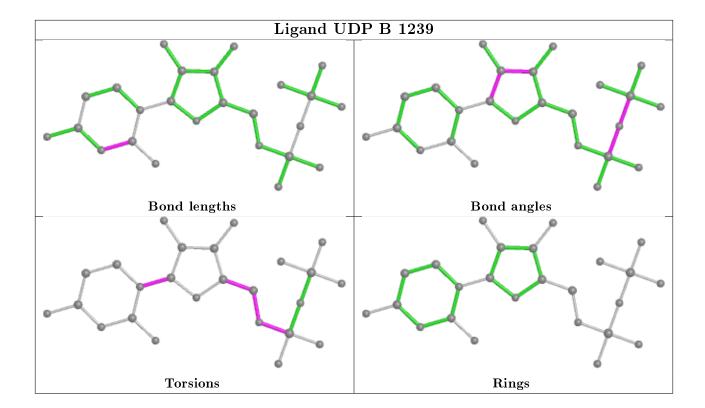












# 5.7 Other polymers (i)

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,  $95^{th}$  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>	$\#\mathrm{RSRZ}{>}2$	$OWAB(A^2)$	Q < 0.9
1	A	241/246 (97%)	-0.12	2 (0%) 86 84	23, 51, 79, 94	2 (0%)
1	В	$238/246 \ (96\%)$	-0.04	2 (0%) 86 84	33, 64, 84, 95	3 (1%)
1	С	240/246 (97%)	0.25	12 (5%) 28 29	25, 97, 148, 184	3 (1%)
1	D	$236/246 \ (95\%)$	0.69	32 (13%) 3 4	46, 119, 199, 358	3 (1%)
All	All	955/984 (97%)	0.19	48 (5%) 28 29	23, 74, 160, 358	11 (1%)

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	116	GLY	7.6
1	D	166	TYR	4.9
1	D	108	ILE	4.6
1	D	161	GLY	4.3
1	D	84	LEU	3.9

### 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.

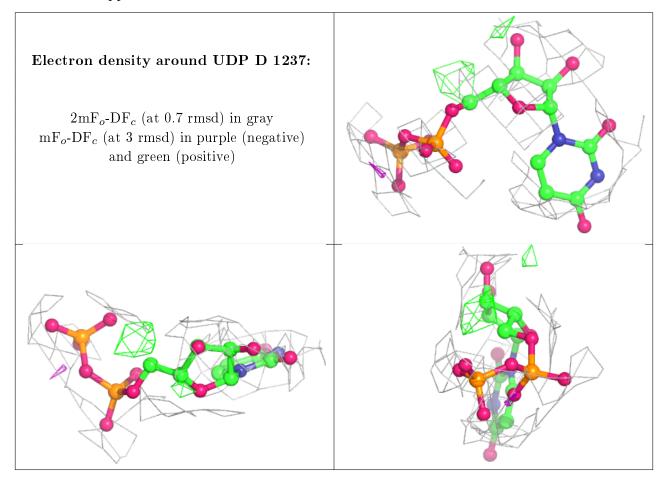
### 6.4 Ligands (i)

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,  $95^{th}$  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.

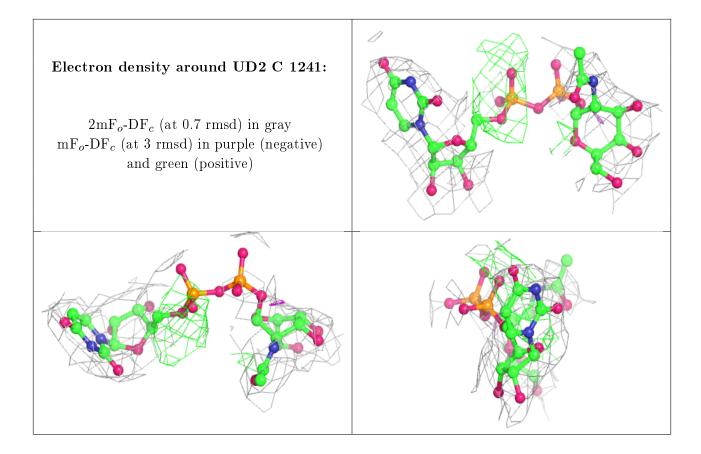


Mol	Type	Chain	${f Res}$	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
3	UDP	D	1237	25/25	0.80	0.20	97,127,136,141	0
4	A2G	В	1240	15/15	0.83	0.23	62,79,86,89	0
2	UD2	С	1241	39/39	0.84	0.21	69,86,97,98	0
3	UDP	В	1239	25/25	0.84	0.25	65,78,105,116	0
4	A2G	D	1238	15/15	0.88	0.19	50,59,71,72	0
2	UD2	A	1242	39/39	0.91	0.19	33,49,62,65	0

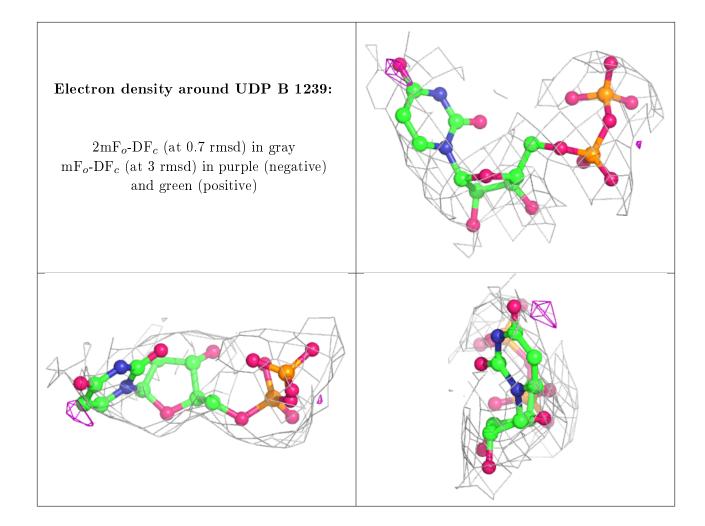
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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



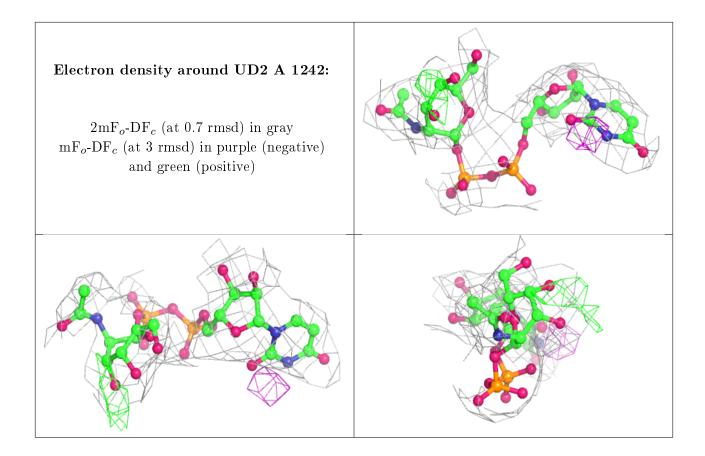












# 6.5 Other polymers (i)

There are no such residues in this entry.

