Full wwPDB X-ray Structure Validation Report (i)

Sep 21, 2020 – 04:51 PM BST

PDB ID	:	4C2S
Title	:	Crystal structure of the fucosylgalactoside alpha N- acetylgalactosaminyltr ansferase (GTA P156L mutant) in complex with UDP and deoxy-H-antigen acceptor
Authors	:	Weadge, J.T.; Palcic, M.; Henriksen, A.
Deposited on	:	2013-08-19
Resolution	:	2.48 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

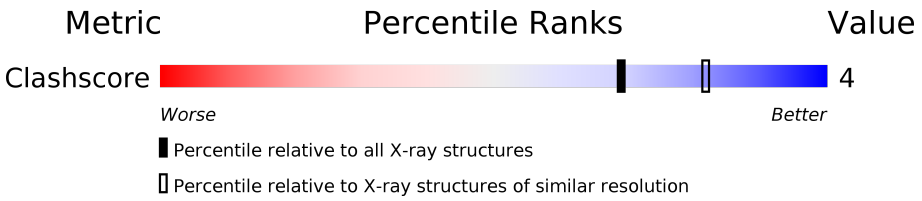
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.6

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 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))
Clashscore	141614	6594 (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 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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	292	89% 11%
1	B	292	92% 7% •
2	C	2	50% 50%
2	D	2	50% 50%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4959 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 HISTO-BLOOD GROUP ABO SYSTEM TRANSFERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	291	2370	1529	415	414	12	0	0	0
1	B	290	2356	1523	408	413	12	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	63	MET	-	expression tag	UNP P16442
A	156	LEU	PRO	engineered mutation	UNP P16442
B	63	MET	-	expression tag	UNP P16442
B	156	LEU	PRO	engineered mutation	UNP P16442

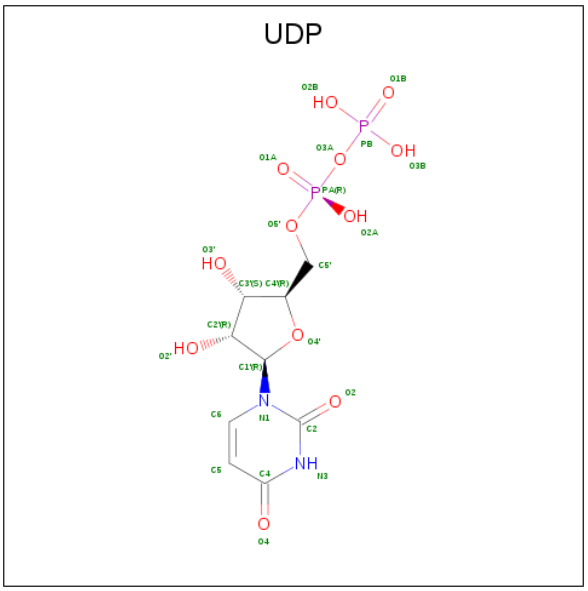
- Molecule 2 is an oligosaccharide called alpha-L-fucopyranose-(1-2)-hexyl 3-deoxy-beta-D-galactopyranoside.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
			Total	C	O			
2	C	2	27	18	9	0	0	0
2	D	2	27	18	9	0	0	0

- Molecule 3 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mn		
3	B	1	1	1	0	0
3	A	1	1	1	0	0

- Molecule 4 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula: $C_9H_{14}N_2O_{12}P_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
4	B	1	Total	C	N	O	P	0	0
			25	9	2	12	2		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	69	Total	O	0	0
			69	69		
5	B	58	Total	O	0	0
			58	58		

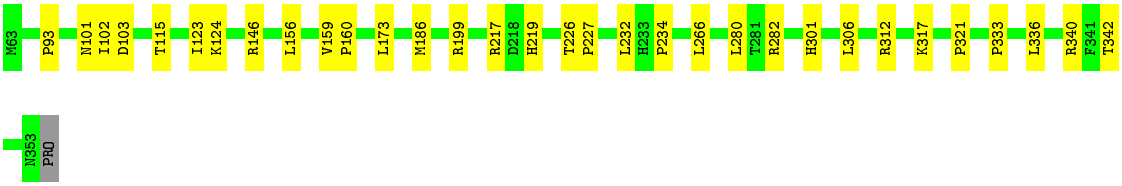
3 Residue-property plots (i)

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. 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. 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.

Note EDS failed to run properly.

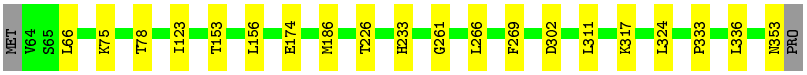
- Molecule 1: HISTO-BLOOD GROUP ABO SYSTEM TRANSFERASE

Chain A:



- Molecule 1: HISTO-BLOOD GROUP ABO SYSTEM TRANSFERASE

Chain B:



- Molecule 2: alpha-L-fucopyranose-(1-2)-hexyl 3-deoxy-beta-D-galactopyranoside

Chain C:

- Molecule 2: alpha-L-fucopyranose-(1-2)-hexyl 3-deoxy-beta-D-galactopyranoside

Chain D:

4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	147.51Å 52.61Å 80.27Å 90.00° 89.97° 90.00°	Depositor
Resolution (Å)	36.88 – 2.48	Depositor
% Data completeness (in resolution range)	99.5 (36.88-2.48)	Depositor
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.84 (at 2.48Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.145 , 0.182	Depositor
Wilson B-factor (Å^2)	16.4	Xtriage
Anisotropy	0.181	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.286 for -h,-k,l	Xtriage
Reported twinning fraction	0.938 for H, K, L 0.062 for h,-k,-l	Depositor
Outliers	1 of 22069 reflections (0.005%)	Xtriage
Total number of atoms	4959	wwPDB-VP
Average B, all atoms (Å^2)	12.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.29% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] 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.

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: UDP, MN, DLG, FUC

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	0.78	0/2435	0.70	0/3305
1	B	0.76	0/2421	0.68	0/3289
All	All	0.77	0/4856	0.69	0/6594

There are no bond length outliers.

There are no bond angle outliers.

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	2370	0	2336	23	0
1	B	2356	0	2320	13	0
2	C	27	0	33	0	0
2	D	27	0	33	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	25	0	11	0	0
4	B	25	0	11	0	0
5	A	69	0	0	1	0
5	B	58	0	0	1	0
All	All	4959	0	4744	35	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 4.

All (35) 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:282:ARG:NH1	5:A:2039:HOH:O	2.10	0.83
1:B:226:THR:HG21	1:B:317:LYS:HB2	1.66	0.78
1:A:226:THR:HG21	1:A:317:LYS:HB2	1.79	0.65
1:B:75:LYS:HG3	1:B:78:THR:HB	1.82	0.61
1:A:199:ARG:HG3	1:A:199:ARG:O	2.02	0.60
1:B:353:ASN:HB3	5:B:2059:HOH:O	2.03	0.59
1:A:173:LEU:HD21	1:A:199:ARG:HH22	1.68	0.58
1:A:173:LEU:HD21	1:A:199:ARG:NH2	2.18	0.57
1:A:93:PRO:HG2	1:A:317:LYS:HG2	1.87	0.56
1:A:186:MET:HG3	1:A:301:HIS:HB3	1.91	0.53
1:A:280:LEU:HD21	1:A:306:LEU:HD13	1.93	0.51
1:A:333:PRO:HD2	1:A:336:LEU:HD12	1.92	0.51
1:B:156:LEU:HD21	1:B:174:GLU:HB2	1.93	0.51
1:A:219:HIS:O	1:A:340:ARG:HD3	2.11	0.51
1:A:186:MET:CG	1:A:301:HIS:HB3	2.41	0.51
1:B:333:PRO:HD2	1:B:336:LEU:HD12	1.94	0.50
1:B:123:ILE:HD13	1:B:153:THR:HA	1.94	0.49
1:A:312:ARG:HD2	1:B:66:LEU:O	2.13	0.49
1:A:115:THR:OG1	1:A:146:ARG:NH1	2.46	0.48
1:B:269:PHE:HB2	1:B:324:LEU:HD12	1.94	0.48
1:A:101:ASN:OD1	1:A:103:ASP:HB2	2.15	0.46
1:B:261:GLY:HA2	1:B:311:LEU:HD13	1.96	0.46
1:A:156:LEU:H	1:A:156:LEU:HD12	1.79	0.46
1:B:233:HIS:HA	1:B:266:LEU:HG	1.98	0.46
1:A:226:THR:OG1	1:A:227:PRO:HD2	2.17	0.45
1:A:159:VAL:HA	1:A:160:PRO:HD3	1.89	0.44
1:A:217:ARG:HD3	1:A:342:THR:OG1	2.18	0.44
1:B:153:THR:O	1:B:174:GLU:HA	2.18	0.43
1:A:232:LEU:HD12	1:A:321:PRO:HD3	2.01	0.42
1:A:234:PRO:HD3	1:A:266:LEU:HD11	2.00	0.42
1:B:186:MET:HA	1:B:302:ASP:HB3	2.01	0.42
1:B:75:LYS:HG2	1:B:75:LYS:H	1.59	0.41
1:A:102:ILE:HD12	1:A:102:ILE:HA	1.90	0.41
1:A:232:LEU:HB2	1:A:321:PRO:HD3	2.02	0.41
1:A:123:ILE:HG22	1:A:124:LYS:HG3	2.02	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

There are no protein backbone outliers to report in this entry.

5.3.2 Protein sidechains (i)

There are no protein residues with a non-rotameric sidechain to report in this entry.

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)

4 monosaccharides 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	DLG	C	1	2	16,17,17	1.25	4 (25%)	18,21,21	0.84	0
2	FUC	C	2	2	10,10,11	0.82	0	14,14,16	0.89	0
2	DLG	D	1	2	16,17,17	1.29	2 (12%)	18,21,21	0.99	1 (5%)
2	FUC	D	2	2	10,10,11	0.75	0	14,14,16	0.79	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	DLG	C	1	2	-	4/9/25/25	0/1/1/1
2	FUC	C	2	2	-	-	0/1/1/1
2	DLG	D	1	2	-	4/9/25/25	0/1/1/1
2	FUC	D	2	2	-	-	0/1/1/1

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	1	DLG	O1-C1	3.22	1.45	1.40
2	C	1	DLG	O1-C1	2.60	1.44	1.40
2	D	1	DLG	C15-C14	-2.52	1.33	1.51
2	C	1	DLG	C15-C14	-2.47	1.34	1.51
2	C	1	DLG	C3-C2	2.32	1.56	1.52
2	C	1	DLG	O5-C5	-2.00	1.39	1.44

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	1	DLG	O5-C1-C2	-2.04	107.66	111.31

There are no chirality outliers.

All (8) torsion outliers are listed below:

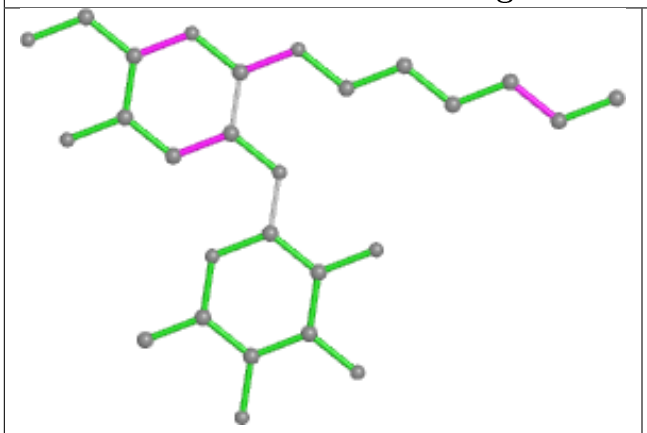
Mol	Chain	Res	Type	Atoms
2	C	1	DLG	O1-C11-C12-C13
2	C	1	DLG	C12-C13-C14-C15
2	D	1	DLG	C4-C5-C6-O6
2	C	1	DLG	O5-C5-C6-O6
2	D	1	DLG	O1-C11-C12-C13
2	C	1	DLG	O5-C1-O1-C11
2	D	1	DLG	O5-C5-C6-O6
2	D	1	DLG	C11-C12-C13-C14

There are no ring outliers.

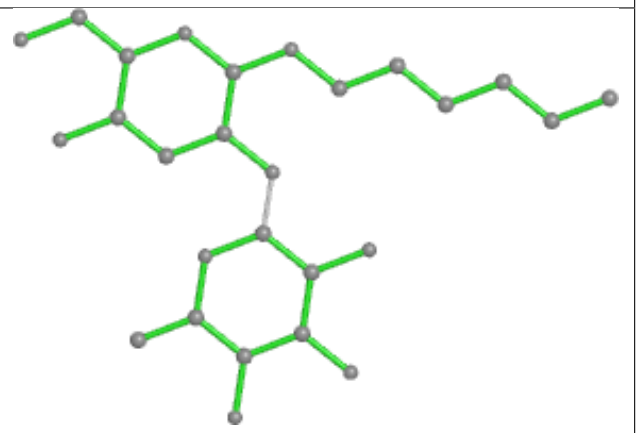
No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

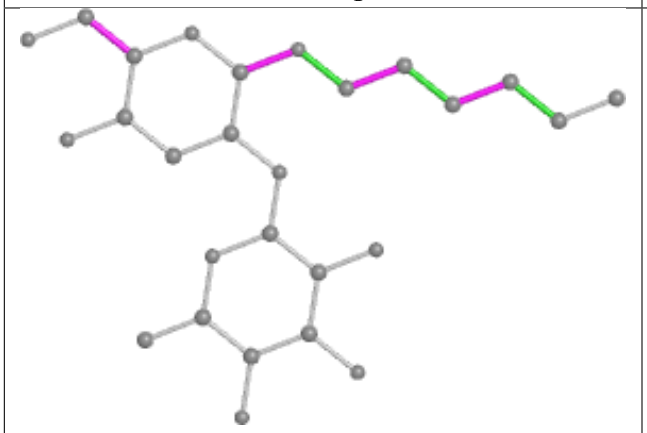
Oligosaccharide Chain C



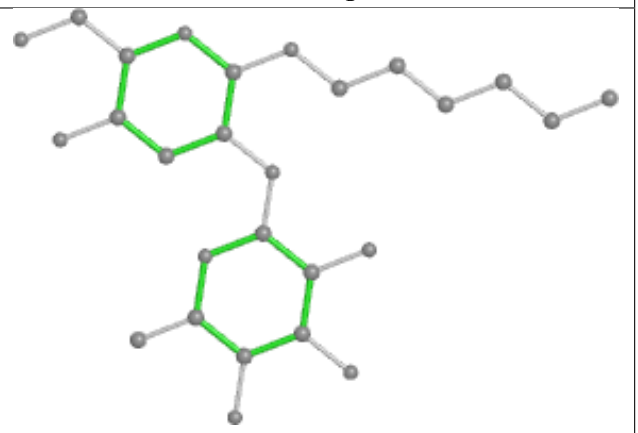
Bond lengths



Bond angles

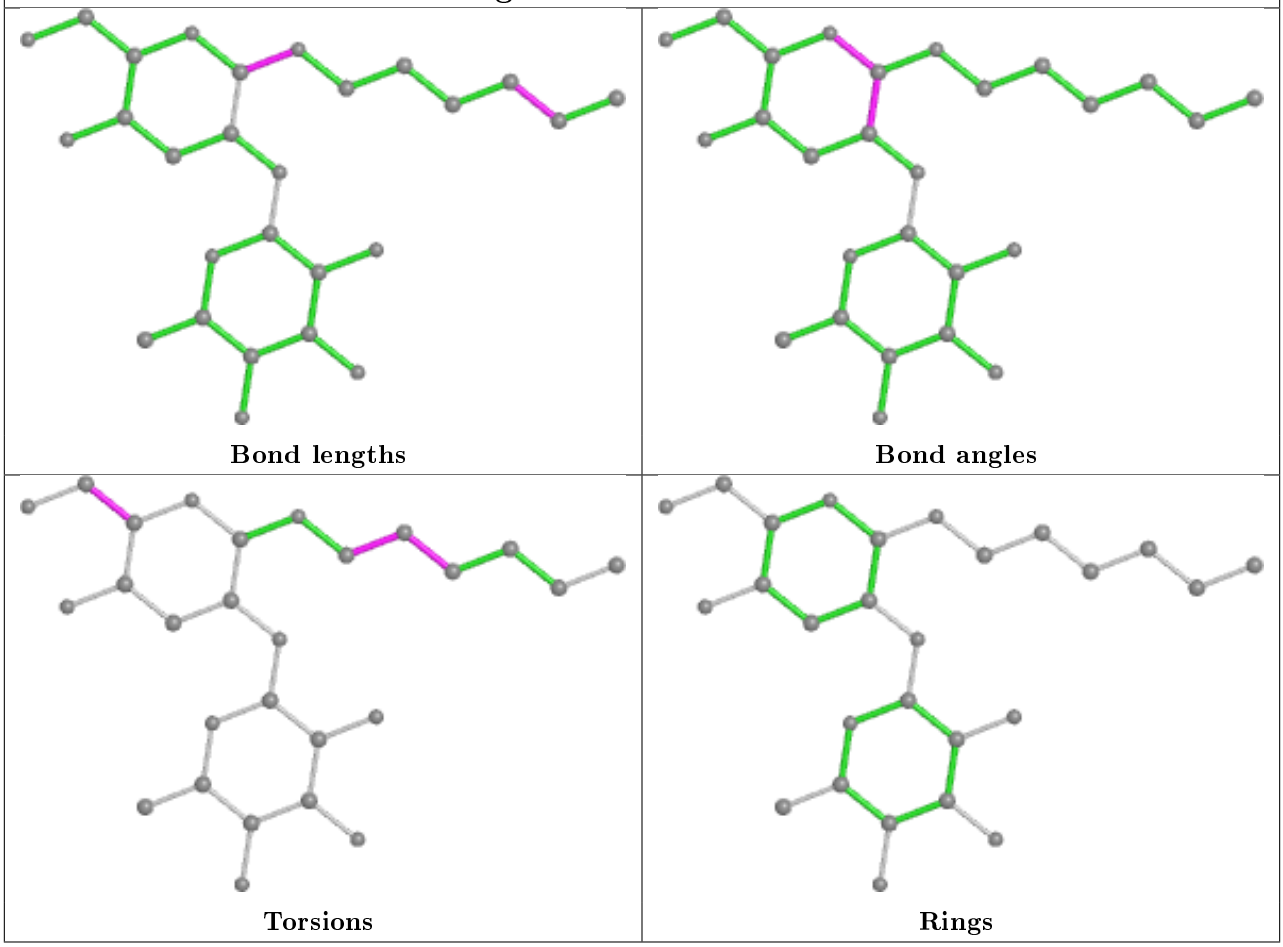


Torsions



Rings

Oligosaccharide Chain D



5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	UDP	B	475	3	20,26,26	0.81	0	25,40,40	1.04	3 (12%)
4	UDP	A	475	3	20,26,26	1.17	3 (15%)	25,40,40	1.04	1 (4%)

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	UDP	B	475	3	-	1/14/32/32	0/2/2/2
4	UDP	A	475	3	-	1/14/32/32	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	475	UDP	C2'-C1'	-2.32	1.50	1.53
4	A	475	UDP	C6-N1	-2.26	1.33	1.35
4	A	475	UDP	C4-N3	2.18	1.36	1.33

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	475	UDP	PA-O3A-PB	-3.23	121.75	132.83
4	B	475	UDP	PA-O3A-PB	-3.15	122.02	132.83
4	B	475	UDP	C6-N1-C2	-2.44	117.32	121.20
4	B	475	UDP	C3'-C2'-C1'	2.02	104.03	100.98

There are no chirality outliers.

All (2) torsion outliers are listed below:

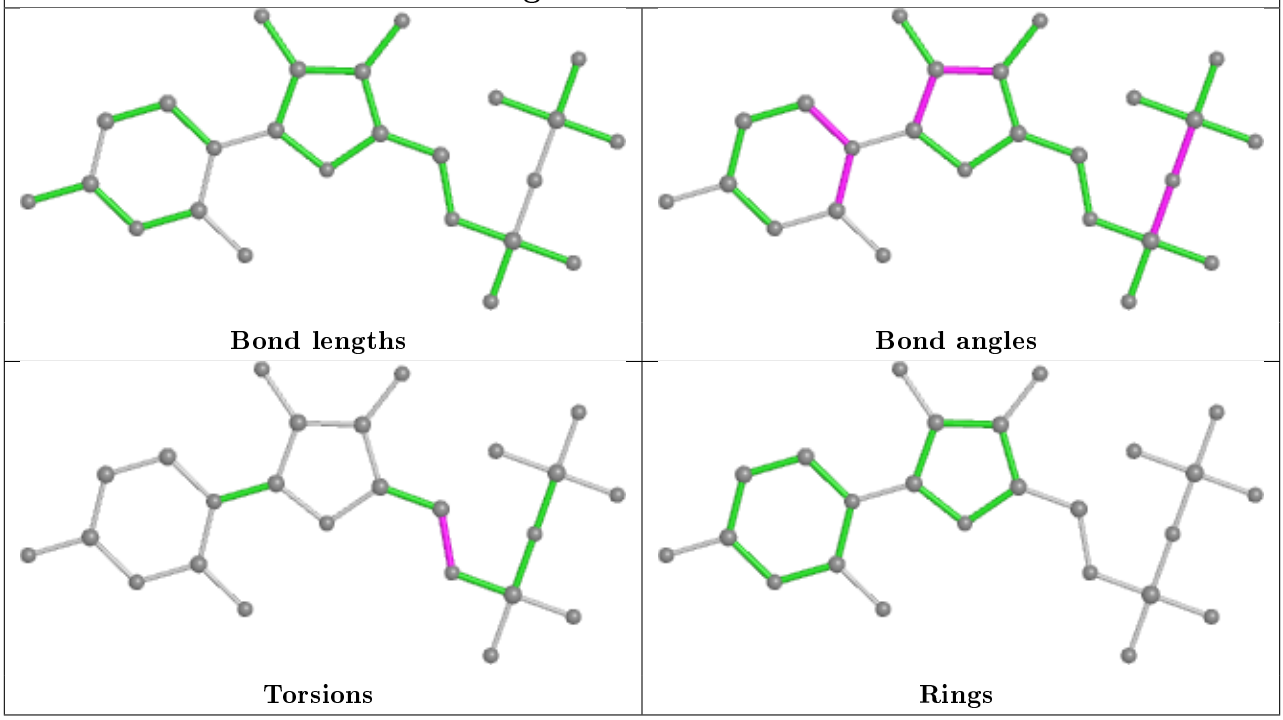
Mol	Chain	Res	Type	Atoms
4	B	475	UDP	C4'-C5'-O5'-PA
4	A	475	UDP	C4'-C5'-O5'-PA

There are no ring outliers.

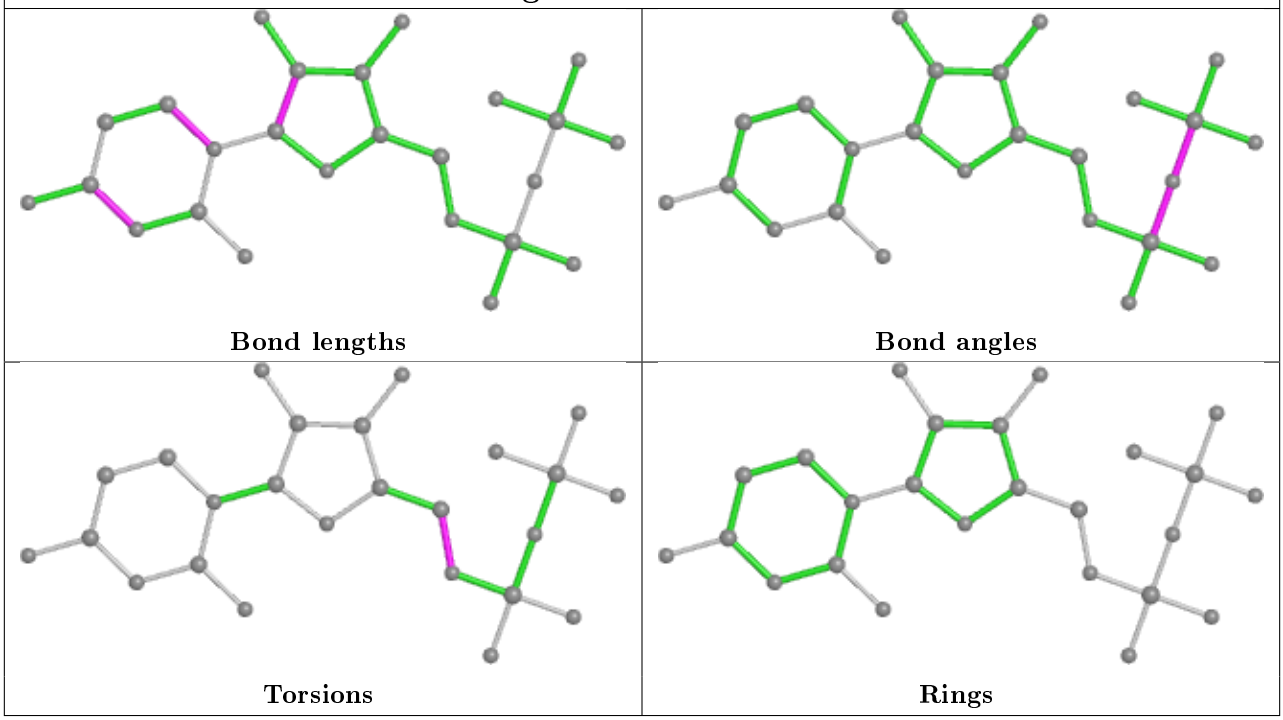
No monomer is involved in short contacts.

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.

Ligand UDP B 475



Ligand UDP A 475



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)

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates (i)

EDS failed to run properly - this section is therefore empty.

6.4 Ligands (i)

EDS failed to run properly - this section is therefore empty.

6.5 Other polymers (i)

EDS failed to run properly - this section is therefore empty.