



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 2, 2024 – 04:12 PM JST

PDB ID : 8XKD
Title : beta-1,4-galactosyltransferase
Authors : Luo, G.; Huang, Z.; Chen, J.; Hou, X.; Zhu, Y.; Ni, D.; Xu, W.; Zhang, W.;
Rao, Y.; Mu, W.
Deposited on : 2023-12-23
Resolution : 1.73 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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) : 1.21
EDS : 3.0
buster-report : 1.1.7 (2018)
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.004 (Gargrove)
Density-Fitness : 1.0.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.40

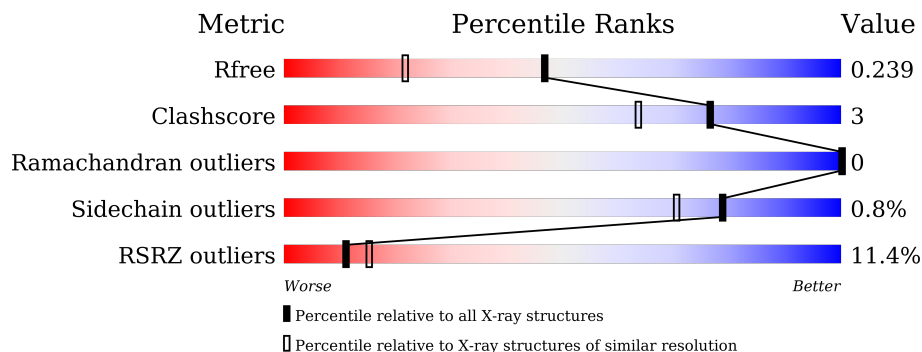
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.73 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}	164625	1043 (1.74-1.74)
Clashscore	180529	1119 (1.74-1.74)
Ramachandran outliers	177936	1112 (1.74-1.74)
Sidechain outliers	177891	1112 (1.74-1.74)
RSRZ outliers	164620	1043 (1.74-1.74)

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 of 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 $\leq 5\%$. 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	251	<div> <div>8%</div> <div>79%</div> <div>5%</div> <div>16%</div> </div>
1	B	251	<div> <div>11%</div> <div>77%</div> <div>6%</div> <div>17%</div> </div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3745 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 25 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	212	Total	C	N	O	S	0	2	0
			1757	1133	288	328	8			
1	B	209	Total	C	N	O	S	0	3	0
			1737	1121	284	324	8			

There are 36 discrepancies between the modelled and reference sequences:

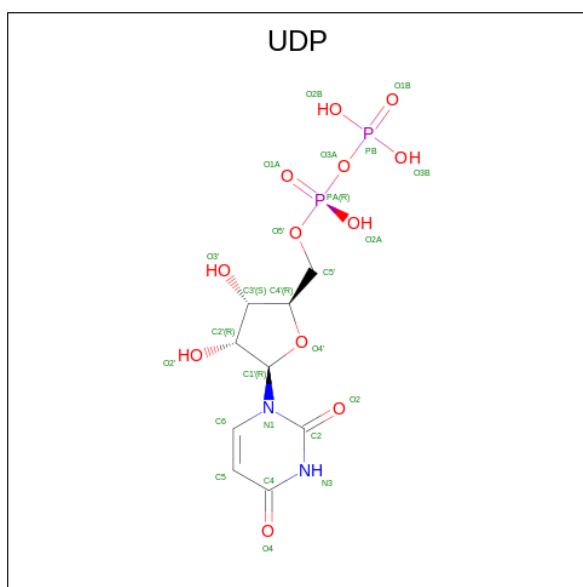
Chain	Residue	Modelled	Actual	Comment	Reference
A	234	ILE	-	expression tag	UNP A0A5D0ENI3
A	235	ASN	-	expression tag	UNP A0A5D0ENI3
A	236	LEU	-	expression tag	UNP A0A5D0ENI3
A	237	LYS	-	expression tag	UNP A0A5D0ENI3
A	238	TYR	-	expression tag	UNP A0A5D0ENI3
A	239	GLU	-	expression tag	UNP A0A5D0ENI3
A	240	LYS	-	expression tag	UNP A0A5D0ENI3
A	241	ARG	-	expression tag	UNP A0A5D0ENI3
A	242	LYS	-	expression tag	UNP A0A5D0ENI3
A	243	HIS	-	expression tag	UNP A0A5D0ENI3
A	244	LEU	-	expression tag	UNP A0A5D0ENI3
A	245	GLU	-	expression tag	UNP A0A5D0ENI3
A	246	HIS	-	expression tag	UNP A0A5D0ENI3
A	247	HIS	-	expression tag	UNP A0A5D0ENI3
A	248	HIS	-	expression tag	UNP A0A5D0ENI3
A	249	HIS	-	expression tag	UNP A0A5D0ENI3
A	250	HIS	-	expression tag	UNP A0A5D0ENI3
A	251	HIS	-	expression tag	UNP A0A5D0ENI3
B	234	ILE	-	expression tag	UNP A0A5D0ENI3
B	235	ASN	-	expression tag	UNP A0A5D0ENI3
B	236	LEU	-	expression tag	UNP A0A5D0ENI3
B	237	LYS	-	expression tag	UNP A0A5D0ENI3
B	238	TYR	-	expression tag	UNP A0A5D0ENI3
B	239	GLU	-	expression tag	UNP A0A5D0ENI3
B	240	LYS	-	expression tag	UNP A0A5D0ENI3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	241	ARG	-	expression tag	UNP A0A5D0ENI3
B	242	LYS	-	expression tag	UNP A0A5D0ENI3
B	243	HIS	-	expression tag	UNP A0A5D0ENI3
B	244	LEU	-	expression tag	UNP A0A5D0ENI3
B	245	GLU	-	expression tag	UNP A0A5D0ENI3
B	246	HIS	-	expression tag	UNP A0A5D0ENI3
B	247	HIS	-	expression tag	UNP A0A5D0ENI3
B	248	HIS	-	expression tag	UNP A0A5D0ENI3
B	249	HIS	-	expression tag	UNP A0A5D0ENI3
B	250	HIS	-	expression tag	UNP A0A5D0ENI3
B	251	HIS	-	expression tag	UNP A0A5D0ENI3

- Molecule 2 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula: $C_9H_{14}N_2O_{12}P_2$) (labeled as "Ligand of Interest" by depositor).



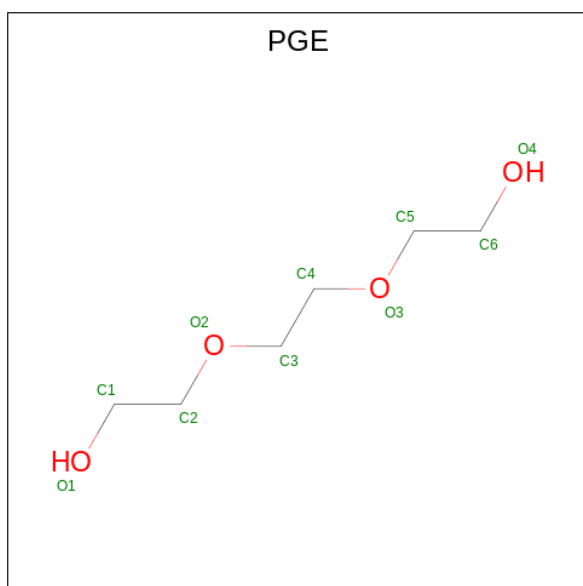
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
2	B	1	Total	C	N	O	P	0	0
			25	9	2	12	2		

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			7	4	3		

- Molecule 4 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			10	6	4		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	95	Total 95	O 95	0	0
5	B	89	Total 89	O 89	0	0

i

- Molecule 1: Glycosyltransferase family 25 protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	82.60Å 92.81Å 71.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.71 – 1.73 46.71 – 1.73	Depositor EDS
% Data completeness (in resolution range)	100.0 (46.71-1.73) 100.0 (46.71-1.73)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.47 (at 1.73Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.200 , 0.234 0.208 , 0.239	Depositor DCC
R_{free} test set	2986 reflections (5.16%)	wwPDB-VP
Wilson B-factor (Å ²)	29.3	Xtriage
Anisotropy	0.042	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 28.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	3745	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.66% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: PGE, UDP, PEG

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	0.56	0/1802	0.70	1/2439 (0.0%)
1	B	0.54	0/1783	0.69	0/2412
All	All	0.55	0/3585	0.70	1/4851 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	72	MET	CG-SD-CE	-6.71	89.46	100.20

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	1757	0	1730	12	0
1	B	1737	0	1714	14	0
2	A	25	0	11	1	0
2	B	25	0	11	2	0
3	B	7	0	10	0	0
4	B	10	0	14	1	0
5	A	95	0	0	0	0
5	B	89	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	3745	0	3490	20	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 3.

The worst 5 of 20 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:B:132:PRO:HB3	1:B:157:PHE:CD2	2.29	0.68
1:B:22:HIS:ND1	5:B:401:HOH:O	2.29	0.65
1:A:145:GLN:HE21	1:B:101:ASN:HD21	1.53	0.56
1:A:147:ARG:HD2	1:B:105:PHE:CZ	2.44	0.53
1:A:40:PHE:O	2:A:301:UDP:N3	2.41	0.53

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	212/251 (84%)	209 (99%)	3 (1%)	0	100	100
1	B	208/251 (83%)	204 (98%)	4 (2%)	0	100	100
All	All	420/502 (84%)	413 (98%)	7 (2%)	0	100	100

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	194/230 (84%)	193 (100%)	1 (0%)	86	81
1	B	192/230 (84%)	190 (99%)	2 (1%)	73	61
All	All	386/460 (84%)	383 (99%)	3 (1%)	79	70

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	146	GLN
1	B	157	PHE
1	B	215	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	195	GLN
1	A	213	GLN
1	B	213	GLN
1	A	156	HIS
1	A	145	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

4 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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	UDP	A	301	-	24,26,26	1.27	3 (12%)	37,40,40	1.50	8 (21%)
4	PGE	B	303	-	9,9,9	0.58	0	8,8,8	0.31	0
3	PEG	B	302	-	6,6,6	0.61	0	5,5,5	0.32	0
2	UDP	B	301	-	24,26,26	1.19	4 (16%)	37,40,40	1.46	6 (16%)

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	UDP	A	301	-	-	4/16/32/32	0/2/2/2
4	PGE	B	303	-	-	5/7/7/7	-
3	PEG	B	302	-	-	2/4/4/4	-
2	UDP	B	301	-	-	0/16/32/32	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	301	UDP	C2-N1	3.58	1.44	1.38
2	B	301	UDP	C2-N1	2.98	1.43	1.38
2	B	301	UDP	O2-C2	2.22	1.27	1.23
2	B	301	UDP	C2-N3	-2.19	1.34	1.38
2	B	301	UDP	C5-C4	-2.16	1.38	1.43

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	301	UDP	C5-C4-N3	3.89	120.67	114.84
2	A	301	UDP	C5-C4-N3	3.66	120.32	114.84
2	B	301	UDP	C4-N3-C2	-3.47	122.01	126.58
2	A	301	UDP	C4-N3-C2	-3.34	122.18	126.58
2	B	301	UDP	O4-C4-C5	-2.99	119.91	125.16

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

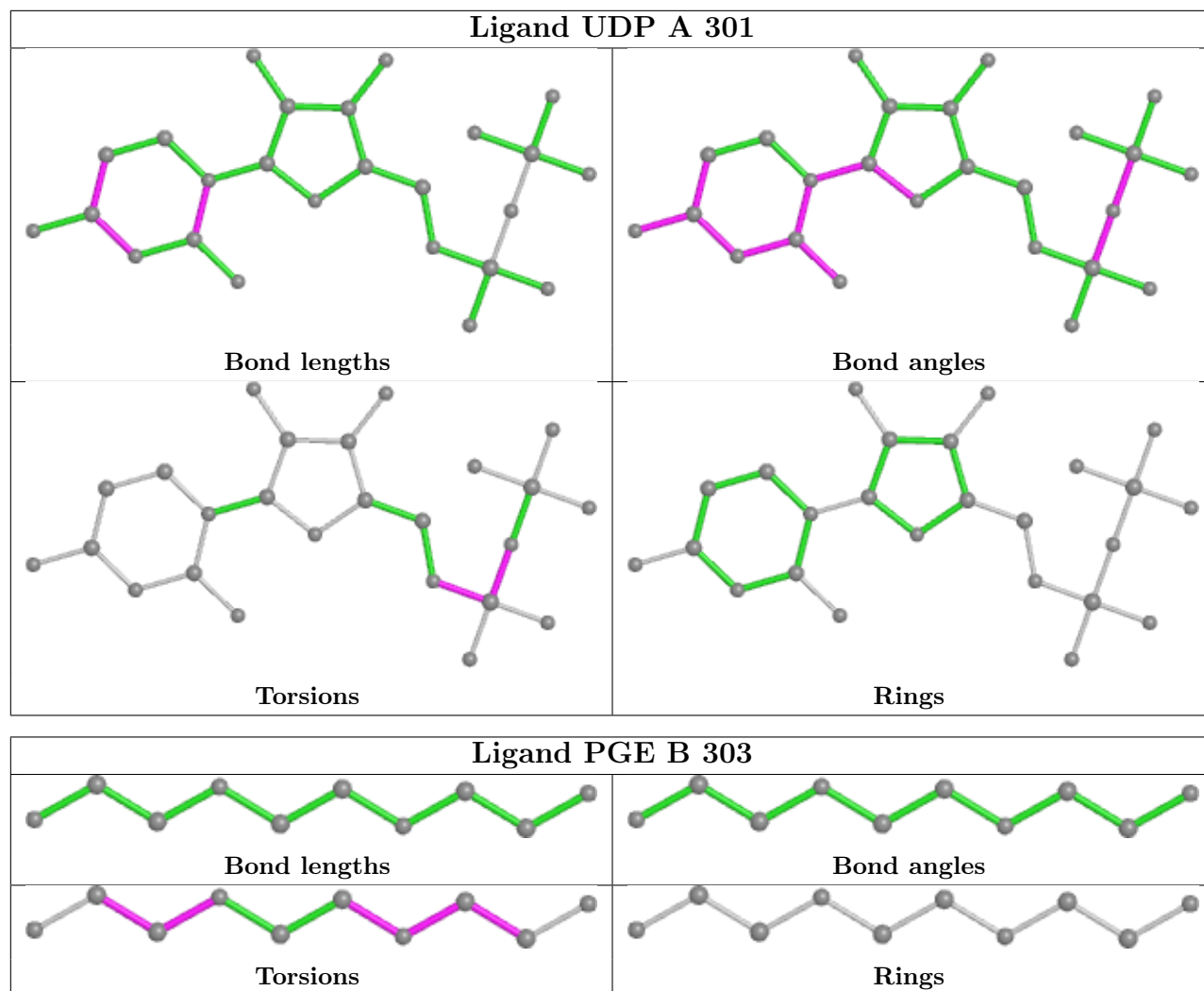
Mol	Chain	Res	Type	Atoms
2	A	301	UDP	C5'-O5'-PA-O1A
3	B	302	PEG	O2-C3-C4-O4
4	B	303	PGE	O3-C5-C6-O4
2	A	301	UDP	PB-O3A-PA-O5'
4	B	303	PGE	C1-C2-O2-C3

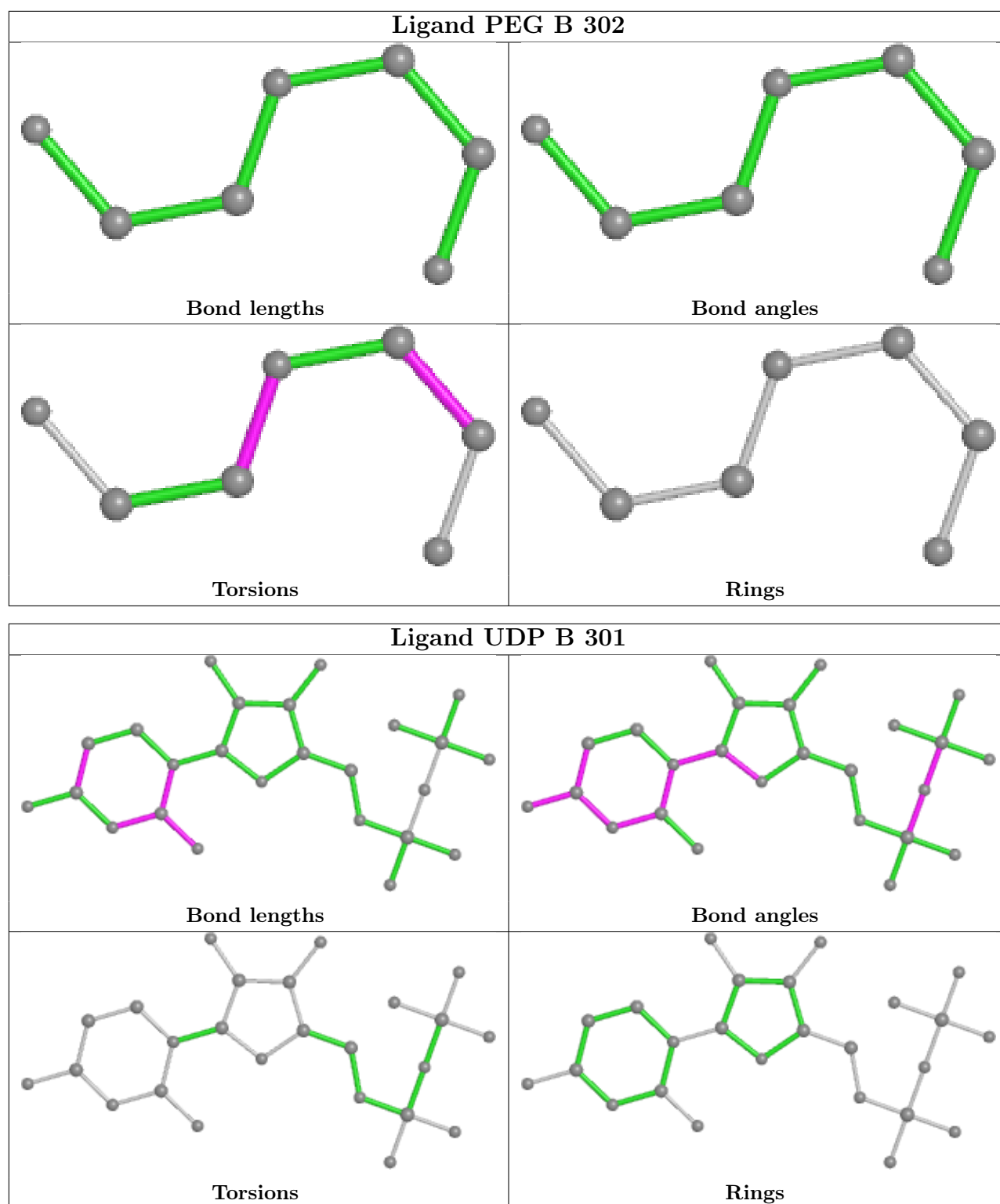
There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	301	UDP	1	0
4	B	303	PGE	1	0
2	B	301	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 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 ⓘ

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, 95th 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 ²)	Q<0.9
1	A	212/251 (84%)	0.67	20 (9%)	15 21	19, 30, 55, 103	2 (0%)
1	B	209/251 (83%)	0.75	28 (13%)	8 11	16, 32, 60, 75	3 (1%)
All	All	421/502 (83%)	0.71	48 (11%)	11 15	16, 31, 58, 103	5 (1%)

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	196	ILE	7.5
1	B	215	LEU	7.4
1	A	215	LEU	7.0
1	A	130	LEU	6.8
1	B	157	PHE	6.2

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](#)

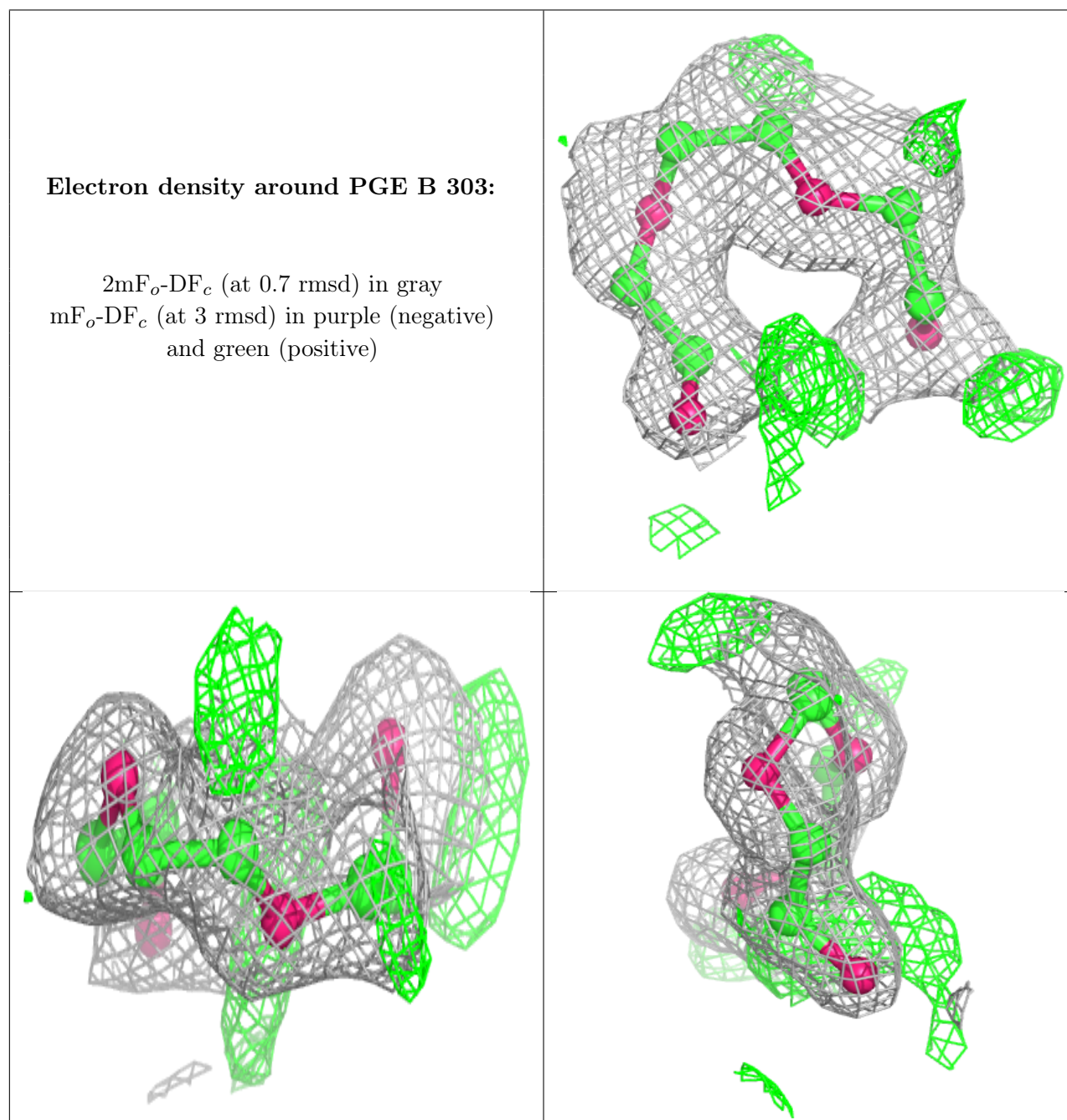
There are no monosaccharides in this entry.

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, 95th 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.

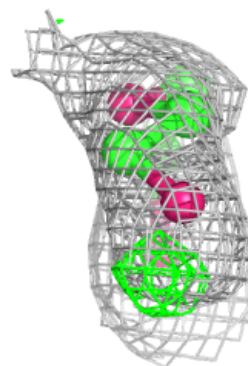
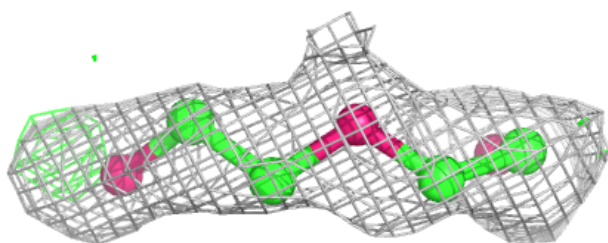
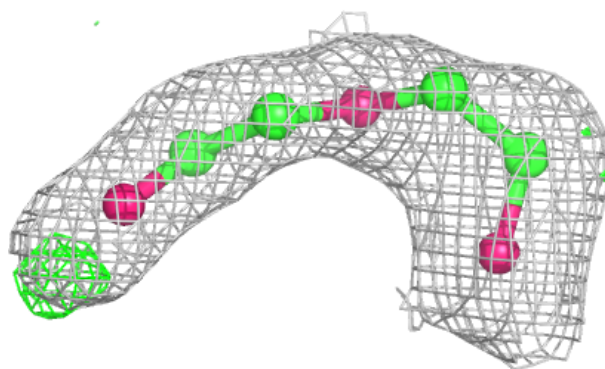
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	PGE	B	303	10/10	0.71	0.20	53,56,58,60	0
3	PEG	B	302	7/7	0.77	0.20	54,55,57,60	0
2	UDP	B	301	25/25	0.82	0.13	35,46,91,94	0
2	UDP	A	301	25/25	0.86	0.12	29,39,93,97	0

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.



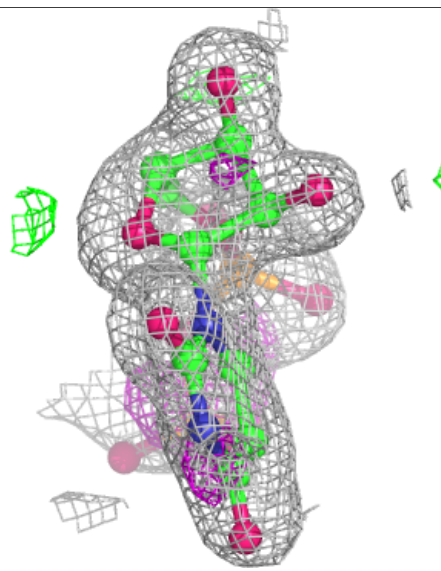
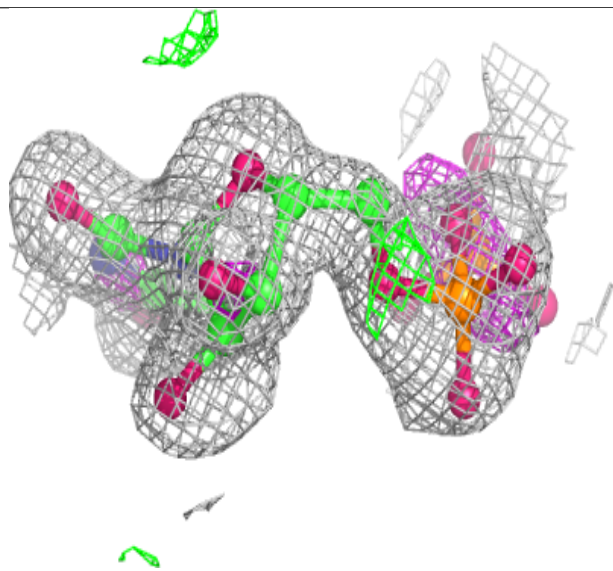
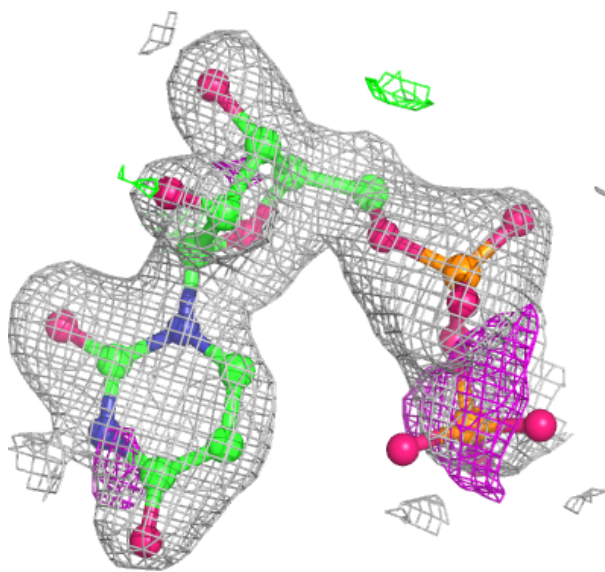
Electron density around PEG B 302:

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



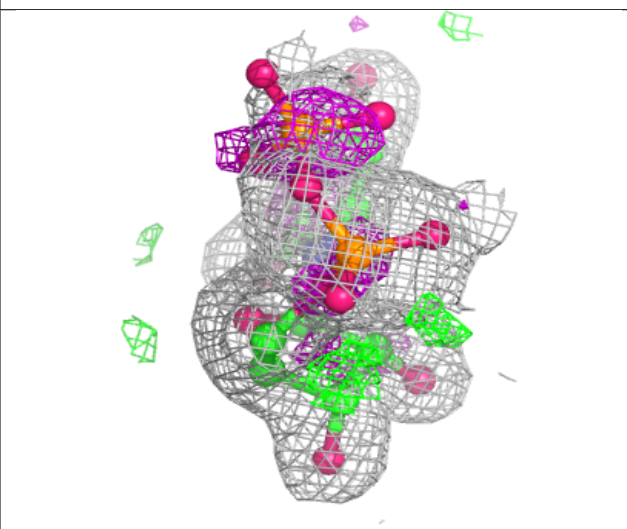
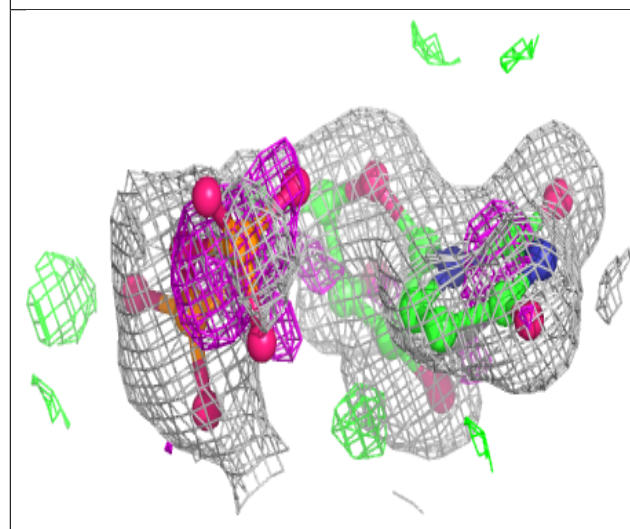
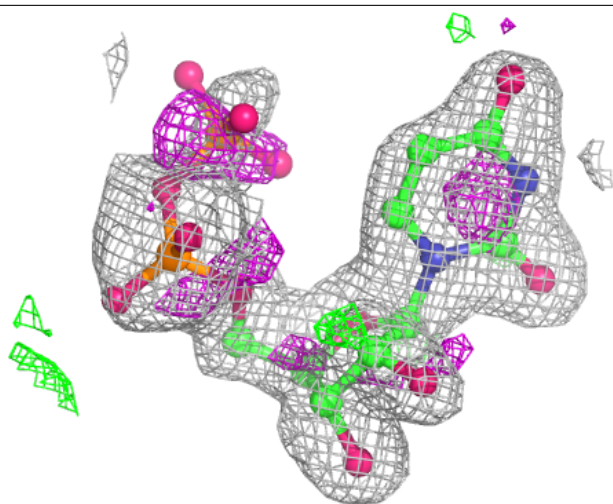
Electron density around UDP B 301:

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



Electron density around UDP A 301:

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



6.5 Other polymers [i](#)

There are no such residues in this entry.