



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 13, 2020 – 06:31 PM BST

PDB ID : 4W6Q
Title : Glycosyltransferase C from *Streptococcus agalactiae*
Authors : Zhu, F.; Zhang, H.; Wu, H.
Deposited on : 2014-08-20
Resolution : 2.70 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 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 : 2.14.4.dev1
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) : 2.14.4.dev1

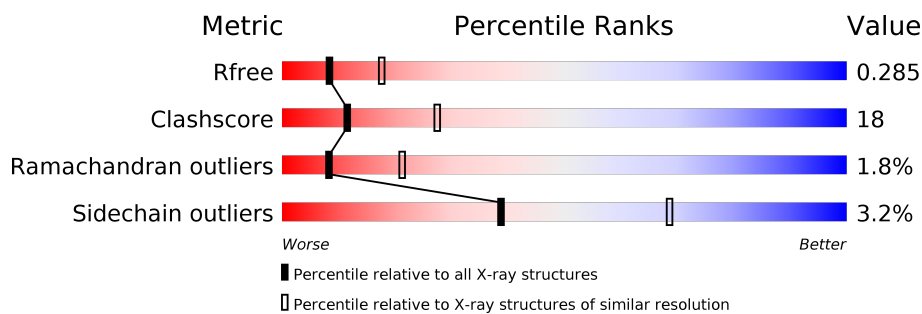
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	333	
1	B	333	
1	C	333	
1	D	333	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11158 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.

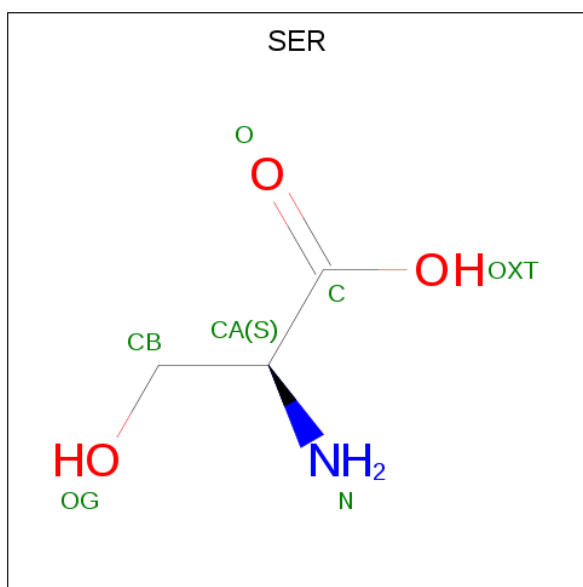
- Molecule 1 is a protein called glucosyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	331	Total	C	N	O	S	0	0	0
			2668	1713	443	497	15			
1	B	332	Total	C	N	O	S	0	0	0
			2674	1716	444	499	15			
1	C	332	Total	C	N	O	S	0	0	0
			2674	1716	444	499	15			
1	D	332	Total	C	N	O	S	0	0	0
			2674	1716	444	499	15			

There are 8 discrepancies between the modelled and reference sequences:

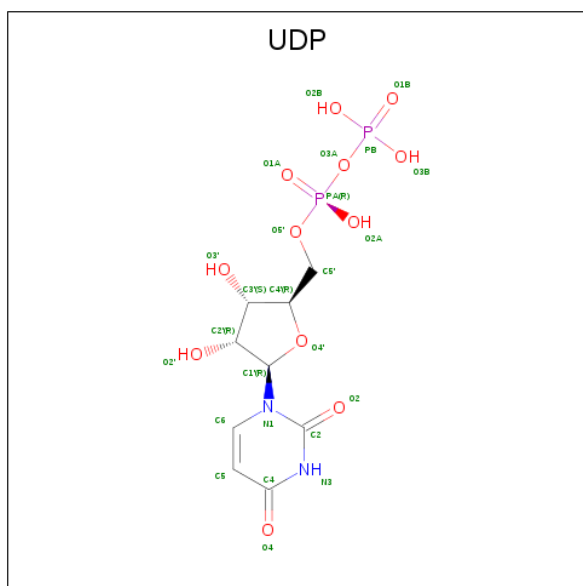
Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ALA	-	expression tag	UNP Q3D9X5
A	0	CYS	-	expression tag	UNP Q3D9X5
B	-1	ALA	-	expression tag	UNP Q3D9X5
B	0	CYS	-	expression tag	UNP Q3D9X5
C	-1	ALA	-	expression tag	UNP Q3D9X5
C	0	CYS	-	expression tag	UNP Q3D9X5
D	-1	ALA	-	expression tag	UNP Q3D9X5
D	0	CYS	-	expression tag	UNP Q3D9X5

- Molecule 2 is SERINE (three-letter code: SER) (formula: C₃H₇NO₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			6	3	1	2		

- 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
3	A	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	B	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	C	1	Total	C	N	O	P	0	0
			25	9	2	12	2		

Continued on next page...

Continued from previous page...

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

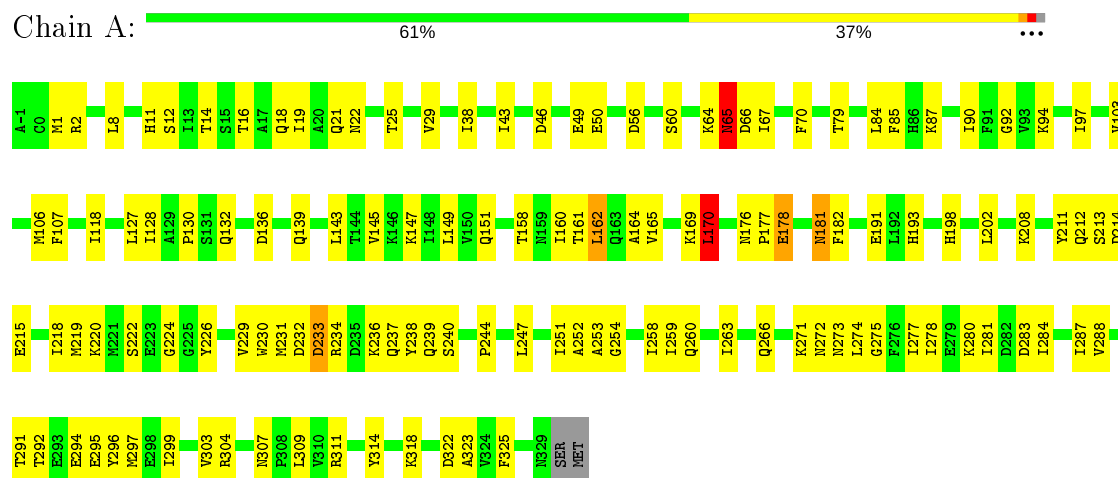
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	96	Total	O	0	0
			96	96		
4	B	94	Total	O	0	0
			94	94		
4	C	102	Total	O	0	0
			102	102		
4	D	70	Total	O	0	0
			70	70		

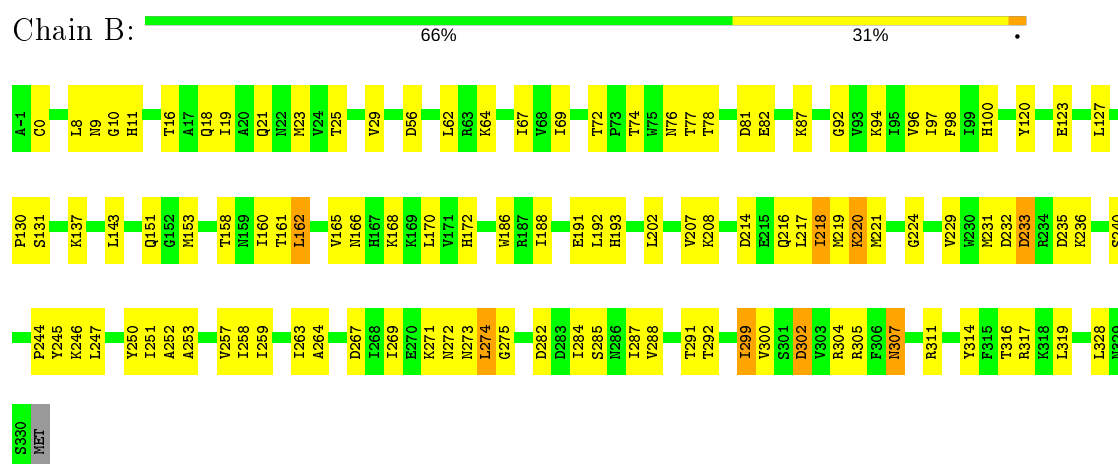
3 Residue-property plots

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.

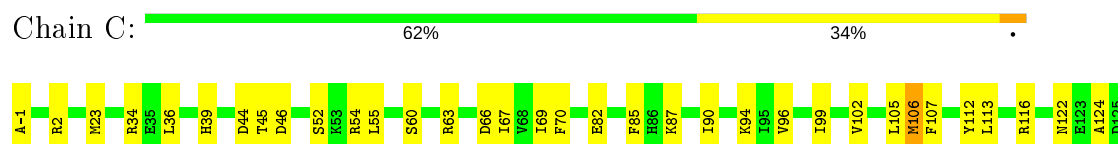
- Molecule 1: glucosyltransferase

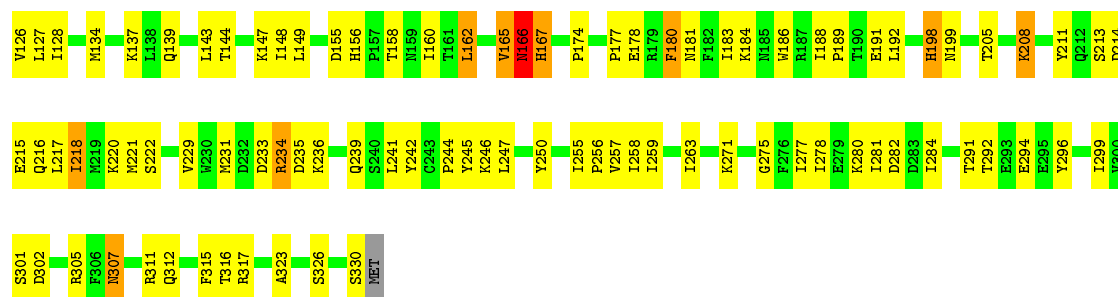


- Molecule 1: glucosyltransferase

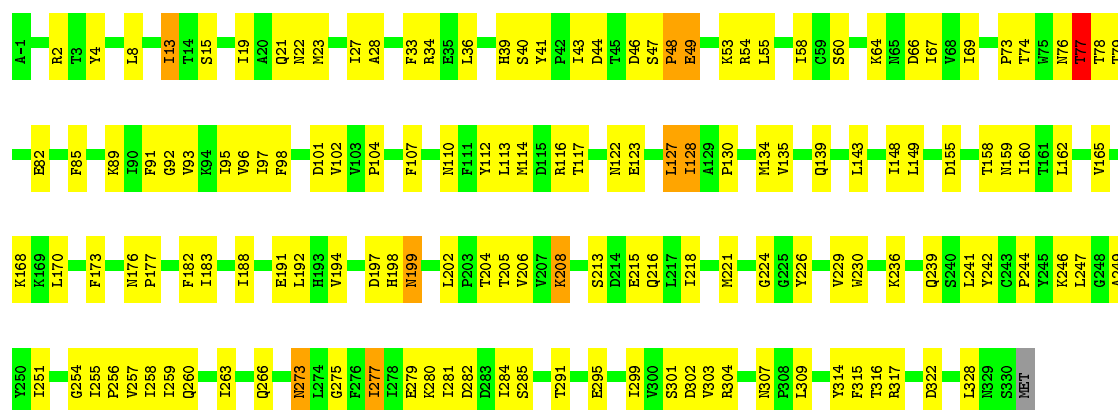


- Molecule 1: glucosyltransferase





• Molecule 1: glucosyltransferase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	77.20Å 99.27Å 188.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.91 – 2.70 29.91 – 2.09	Depositor EDS
% Data completeness (in resolution range)	99.4 (29.91-2.70) 98.0 (29.91-2.09)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.74 (at 2.10Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.184 , 0.284 0.186 , 0.285	Depositor DCC
R_{free} test set	2000 reflections (2.38%)	wwPDB-VP
Wilson B-factor (Å ²)	33.9	Xtriage
Anisotropy	0.331	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 67.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	11158	wwPDB-VP
Average B, all atoms (Å ²)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.77% 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: UDP

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	0.49	0/2728	0.70	2/3701 (0.1%)
1	B	0.52	0/2734	0.65	3/3709 (0.1%)
1	C	0.48	0/2734	0.66	1/3709 (0.0%)
1	D	0.46	0/2734	0.67	1/3709 (0.0%)
All	All	0.48	0/10930	0.67	7/14828 (0.0%)

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
1	C	0	2
All	All	0	3

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	162	LEU	CA-CB-CG	7.21	131.87	115.30
1	A	170	LEU	CA-CB-CG	6.78	130.89	115.30
1	B	274	LEU	CA-CB-CG	-6.07	101.35	115.30
1	D	162	LEU	CA-CB-CG	5.58	128.12	115.30
1	B	218	ILE	CG1-CB-CG2	-5.37	99.58	111.40

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	65	ASN	Peptide
1	C	165	VAL	Peptide
1	C	166	ASN	Peptide

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	2668	0	2662	103	0
1	B	2674	0	2667	86	0
1	C	2674	0	2667	99	0
1	D	2674	0	2667	104	0
2	A	6	0	4	1	0
3	A	25	0	11	0	0
3	B	25	0	11	0	0
3	C	25	0	11	1	0
3	D	25	0	11	0	0
4	A	96	0	0	9	0
4	B	94	0	0	5	0
4	C	102	0	0	9	0
4	D	70	0	0	3	0
All	All	11158	0	10711	379	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 18.

The worst 5 of 379 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:C:233:ASP:HA	1:C:236:LYS:HD3	1.25	1.11
1:B:56:ASP:OD1	1:B:87:LYS:NZ	1.96	0.97
1:B:165:VAL:HG11	1:B:300:VAL:HG11	1.50	0.92
1:C:292:THR:HG22	1:C:294:GLU:H	1.38	0.89
1:C:166:ASN:HA	1:C:296:TYR:HE1	1.40	0.85

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	329/333 (99%)	304 (92%)	17 (5%)	8 (2%)	6	15
1	B	330/333 (99%)	300 (91%)	28 (8%)	2 (1%)	25	50
1	C	330/333 (99%)	303 (92%)	21 (6%)	6 (2%)	8	21
1	D	330/333 (99%)	302 (92%)	20 (6%)	8 (2%)	6	15
All	All	1319/1332 (99%)	1209 (92%)	86 (6%)	24 (2%)	8	21

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	65	ASN
1	B	233	ASP
1	C	166	ASN
1	D	48	PRO
1	D	107	PHE

5.3.2 Protein sidechains

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	298/300 (99%)	292 (98%)	6 (2%)	55	81
1	B	299/300 (100%)	292 (98%)	7 (2%)	50	78
1	C	299/300 (100%)	289 (97%)	10 (3%)	38	67
1	D	299/300 (100%)	284 (95%)	15 (5%)	24	51
All	All	1195/1200 (100%)	1157 (97%)	38 (3%)	39	68

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	198	HIS
1	C	234	ARG
1	D	281	ILE
1	C	208	LYS
1	C	242	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	181	ASN
1	D	239	GLN
1	D	176	ASN
1	B	273	ASN
1	D	163	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](#)

5 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
3	UDP	C	401	-	20,26,26	1.29	1 (5%)	25,40,40	1.07	3 (12%)
2	SER	A	401	-	4,5,6	0.60	0	0,5,7	0.00	-
3	UDP	D	401	-	20,26,26	1.17	1 (5%)	25,40,40	1.02	2 (8%)
3	UDP	B	401	-	20,26,26	1.12	1 (5%)	25,40,40	0.98	1 (4%)
3	UDP	A	402	-	20,26,26	1.25	1 (5%)	25,40,40	1.12	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
3	UDP	C	401	-	-	3/14/32/32	0/2/2/2
2	SER	A	401	-	-	0/2/4/6	-
3	UDP	D	401	-	-	4/14/32/32	0/2/2/2
3	UDP	B	401	-	-	0/14/32/32	0/2/2/2
3	UDP	A	402	-	-	6/14/32/32	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	401	UDP	C4-N3	3.76	1.39	1.33
3	A	402	UDP	C4-N3	3.32	1.38	1.33
3	B	401	UDP	C4-N3	2.95	1.38	1.33
3	D	401	UDP	C4-N3	2.88	1.38	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	401	UDP	C3'-C2'-C1'	2.78	105.17	100.98
3	D	401	UDP	PA-O3A-PB	-2.74	123.42	132.83
3	A	402	UDP	PA-O3A-PB	-2.71	123.53	132.83
3	B	401	UDP	PA-O3A-PB	-2.65	123.73	132.83
3	C	401	UDP	PA-O3A-PB	-2.56	124.04	132.83

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	401	UDP	C5'-O5'-PA-O3A

Continued on next page...

Continued from previous page...

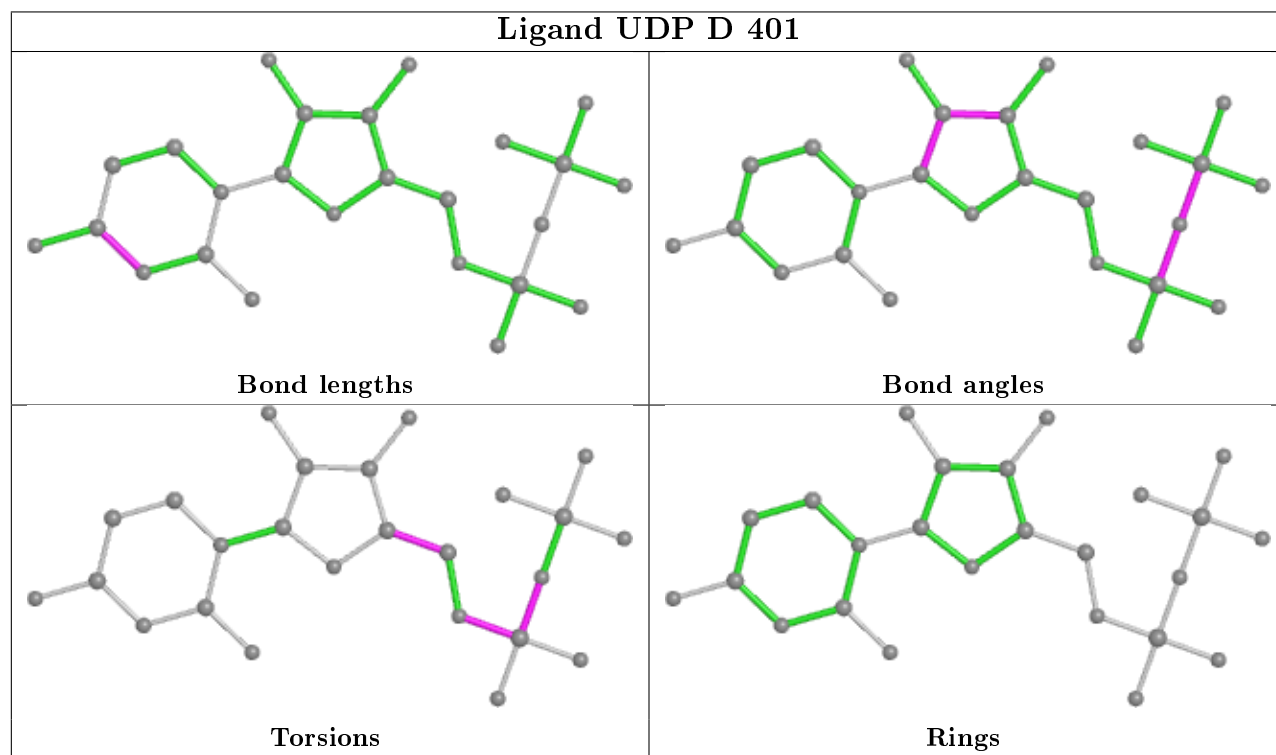
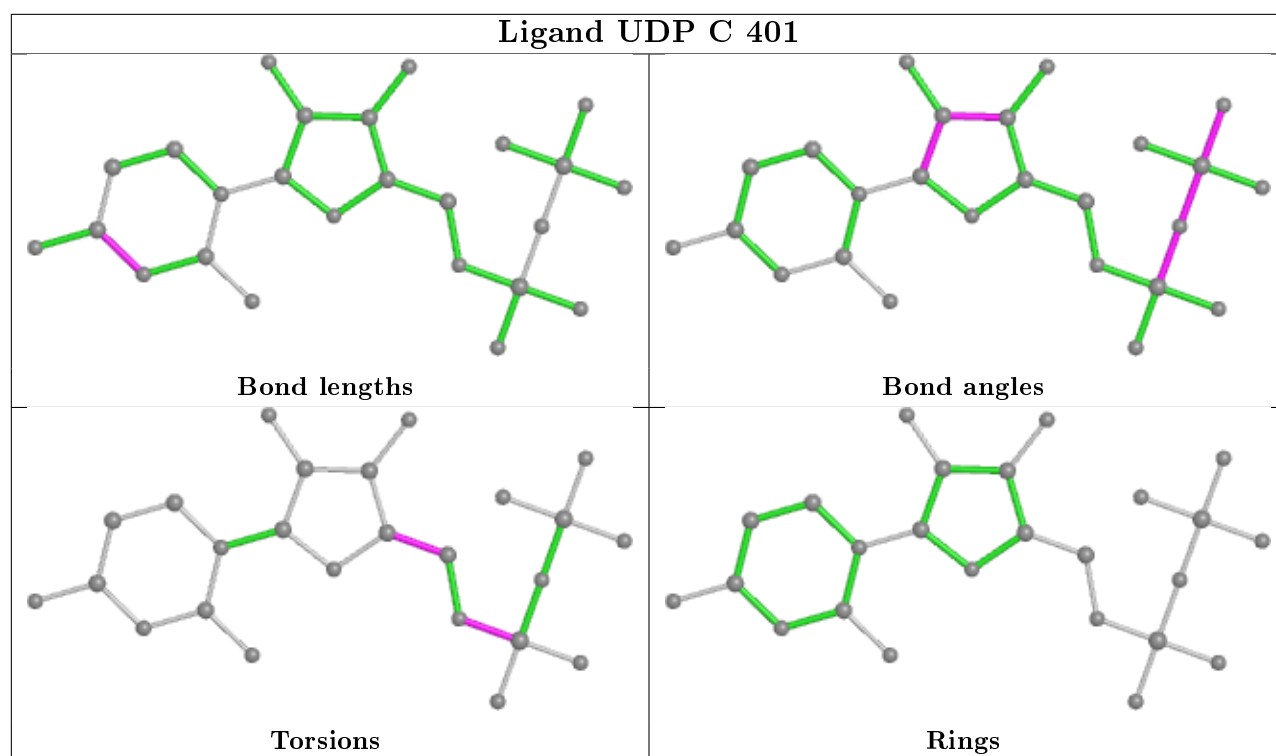
Mol	Chain	Res	Type	Atoms
3	D	401	UDP	C5'-O5'-PA-O3A
3	A	402	UDP	C5'-O5'-PA-O1A
3	A	402	UDP	C5'-O5'-PA-O2A
3	A	402	UDP	C5'-O5'-PA-O3A

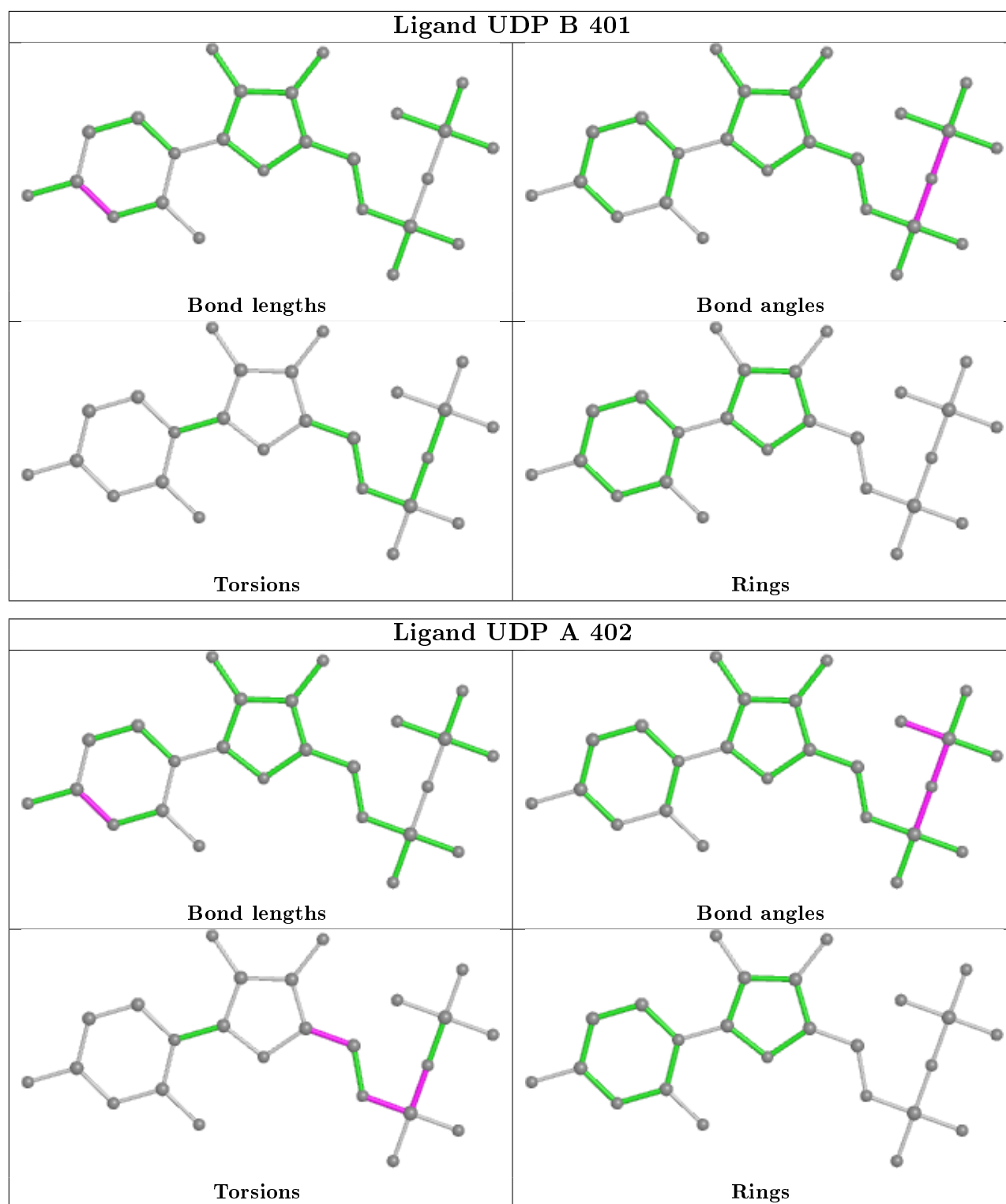
There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	401	UDP	1	0
2	A	401	SER	1	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 ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

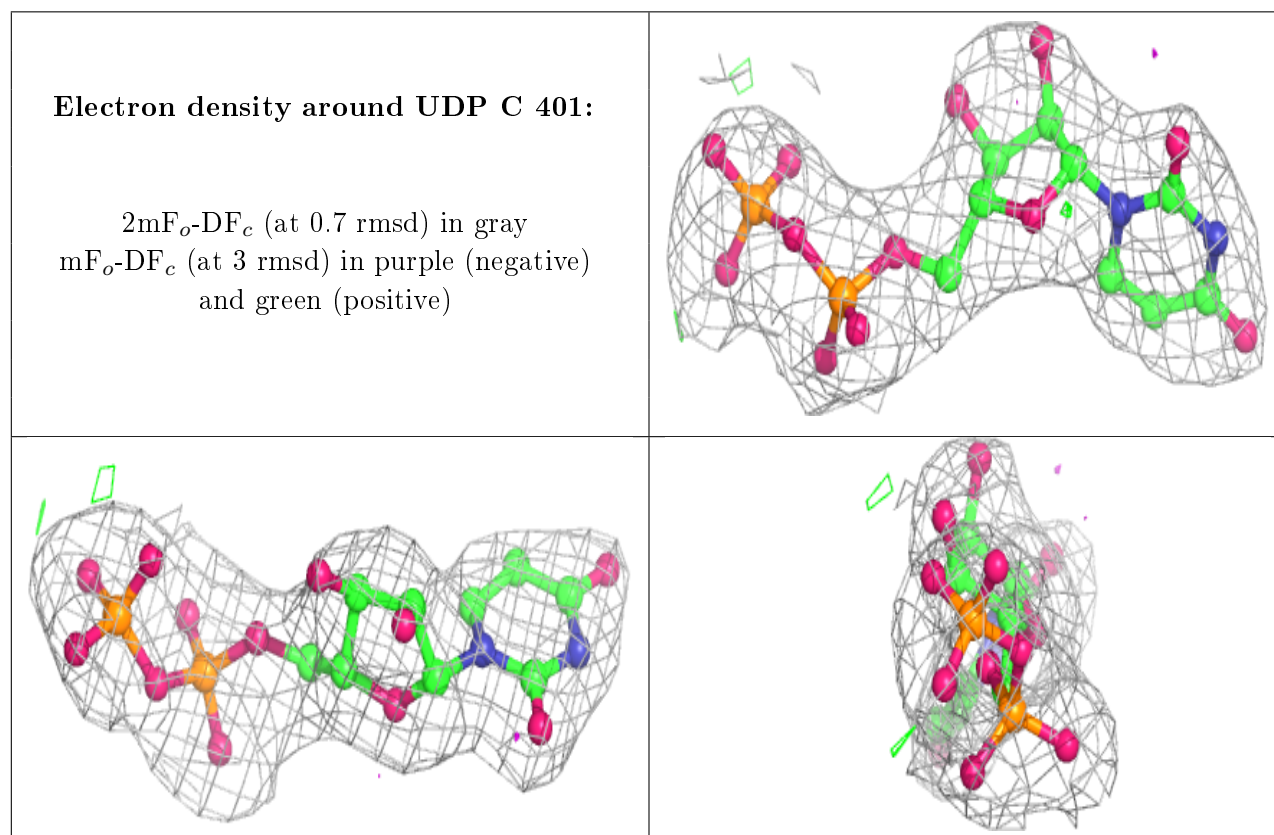
6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

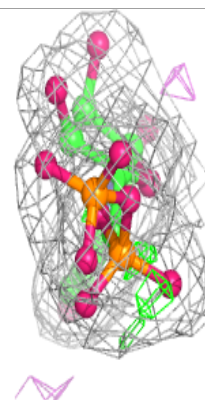
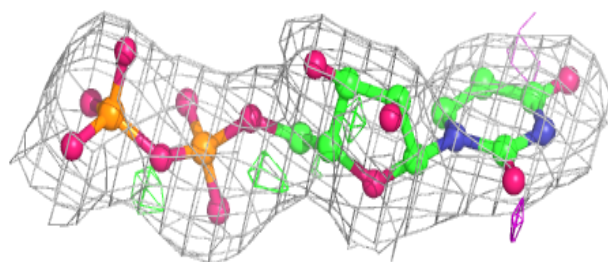
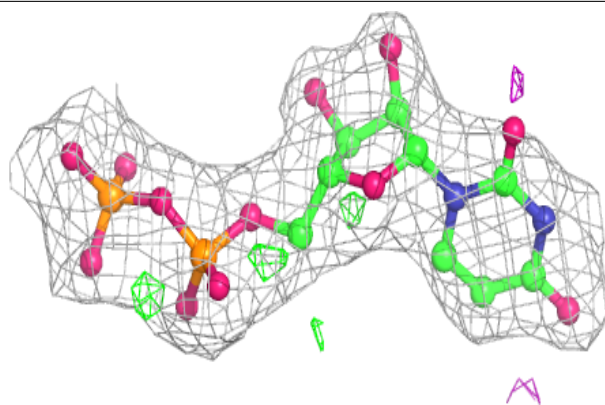
Unable to reproduce the depositors R factor - this section is therefore empty.

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.

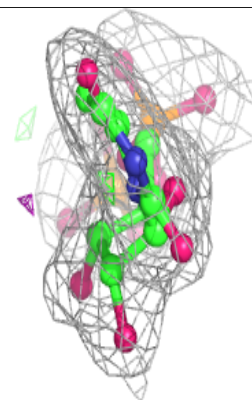
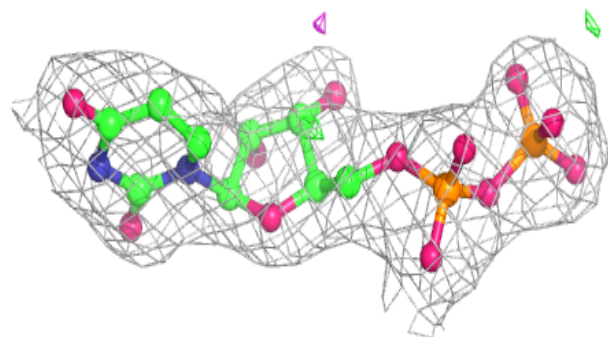
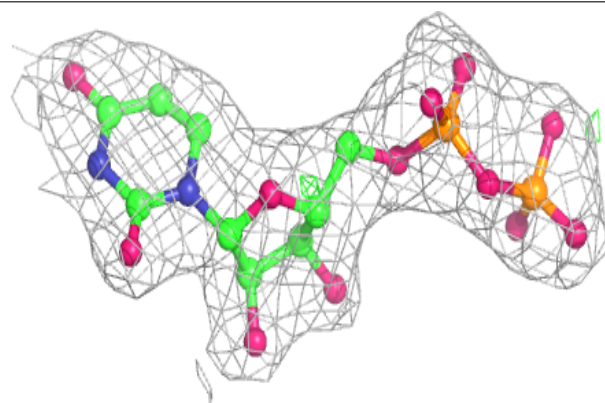


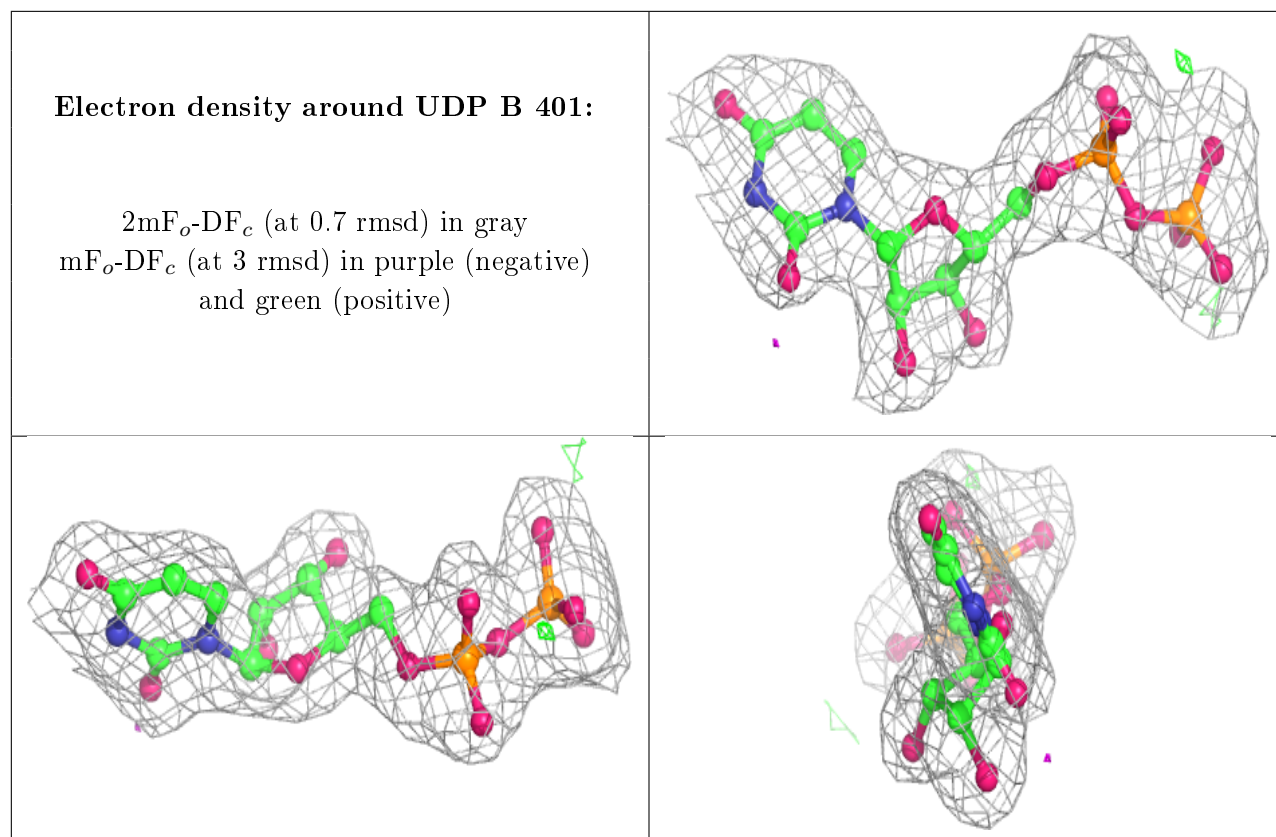
Electron density around UDP D 401:

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

**Electron density around UDP A 402:**

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





6.5 Other polymers [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.