



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 7, 2020 – 08:11 PM BST

PDB ID : 6IWR
Title : Crystal structure of GalNAc-T7 with UDP, GalNAc and Mn²⁺
Authors : Yu, C.; Yin, Y.X.
Deposited on : 2018-12-06
Resolution : 2.60 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 : 2.13.1
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

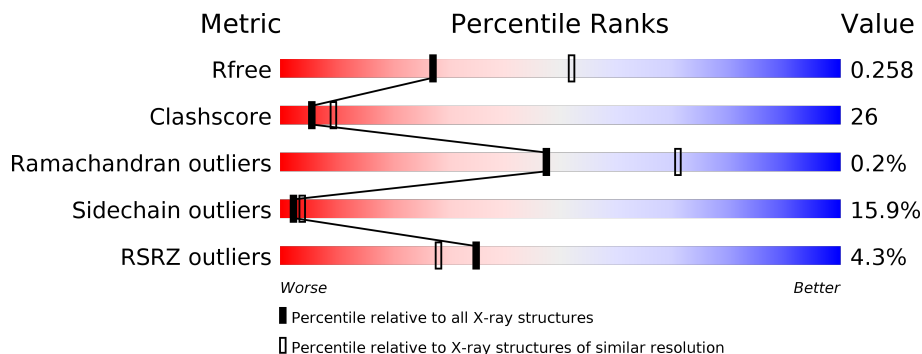
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 $\leq 5\%$. 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	597	
1	B	597	
1	C	597	
1	D	597	
1	E	597	
1	F	597	

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 27098 atoms, of which 26 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 N-acetylgalactosaminyltransferase 7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	546	4440	2855	757	805	23	0	0	0
1	B	546	4440	2855	757	805	23	0	0	0
1	C	546	4440	2855	757	805	23	0	0	0
1	D	538	4380	2814	748	795	23	0	0	0
1	E	546	4440	2855	757	805	23	0	0	0
1	F	546	4440	2855	757	805	23	0	0	0

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

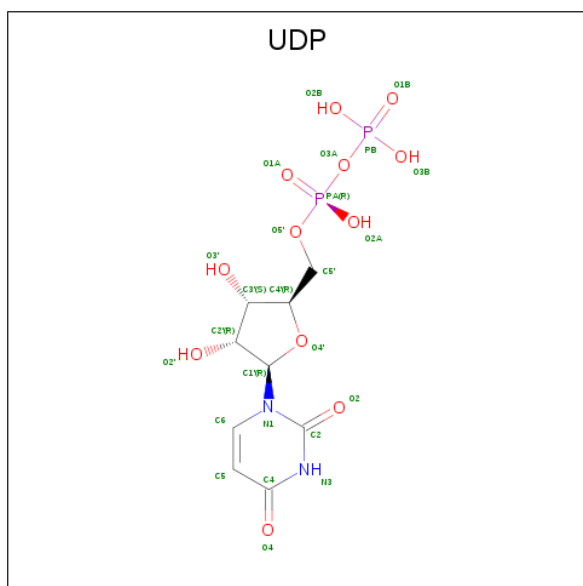
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total 1	Mn 1	0	0
2	E	1	Total 1	Mn 1	0	0
2	B	1	Total 1	Mn 1	0	0
2	C	1	Total 1	Mn 1	0	0
2	A	1	Total 1	Mn 1	0	0
2	F	1	Total 1	Mn 1	0	0

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C₈H₁₅NO₆).



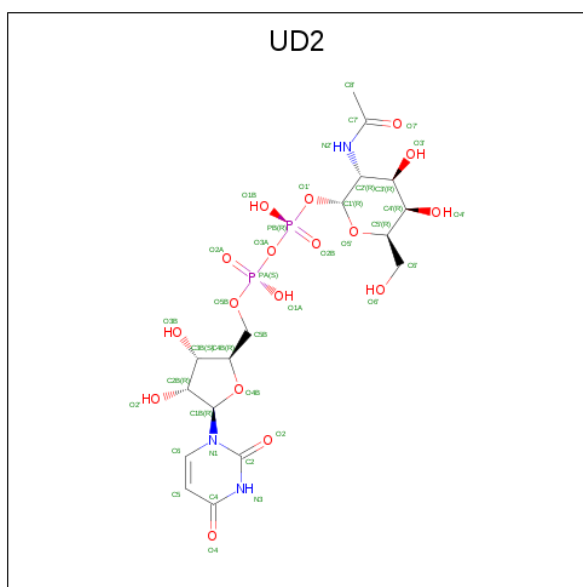
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	N	O		
3	D	1	30	8	15	1	6	0	0

- 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
			Total	C	H	N	O	P		
4	D	1	36	9	11	2	12	2	0	0

- Molecule 5 is URIDINE-DIPHOSPHATE-N-ACETYL GALACTOSAMINE (three-letter code: UD2) (formula: $C_{17}H_{27}N_3O_{17}P_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
5	D	1	39	17	3	17	2	0	0

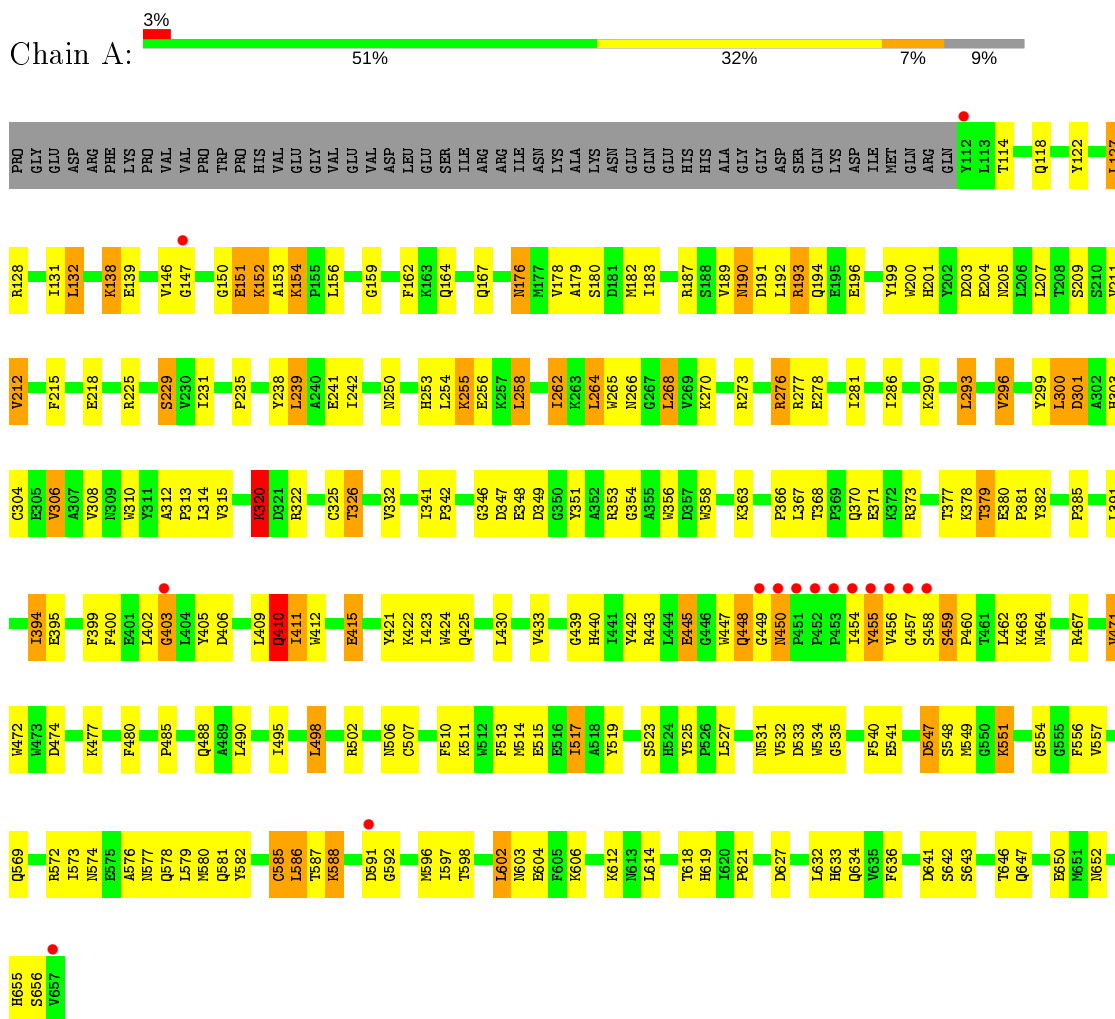
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	100	Total	O	0	0
			100	100		
6	B	77	Total	O	0	0
			77	77		
6	C	60	Total	O	0	0
			60	60		
6	D	55	Total	O	0	0
			55	55		
6	E	68	Total	O	0	0
			68	68		
6	F	47	Total	O	0	0
			47	47		

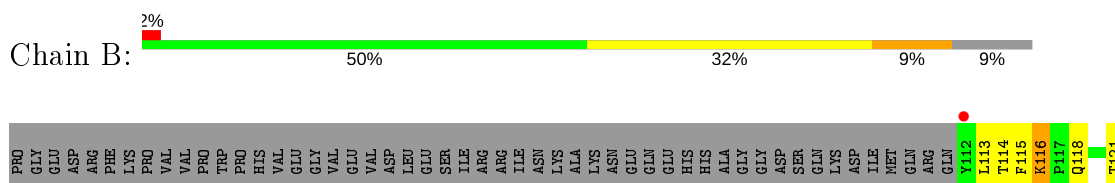
3 Residue-property plots [i](#)

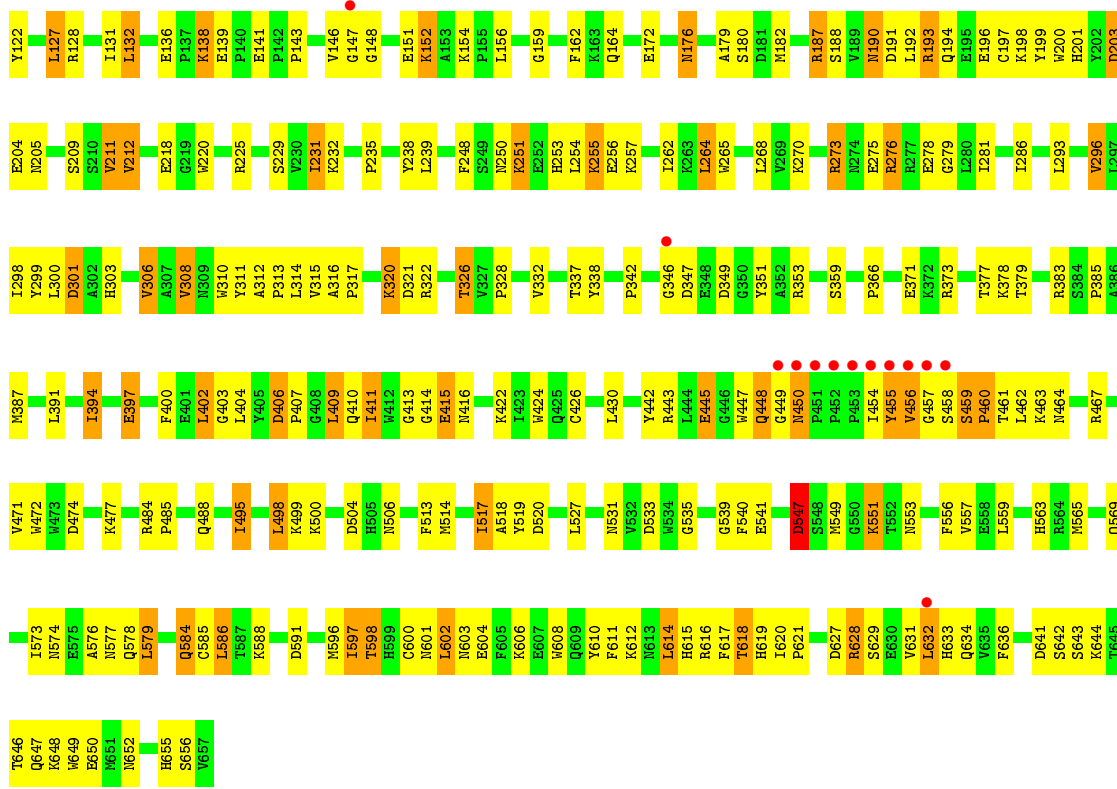
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: N-acetylgalactosaminyltransferase 7

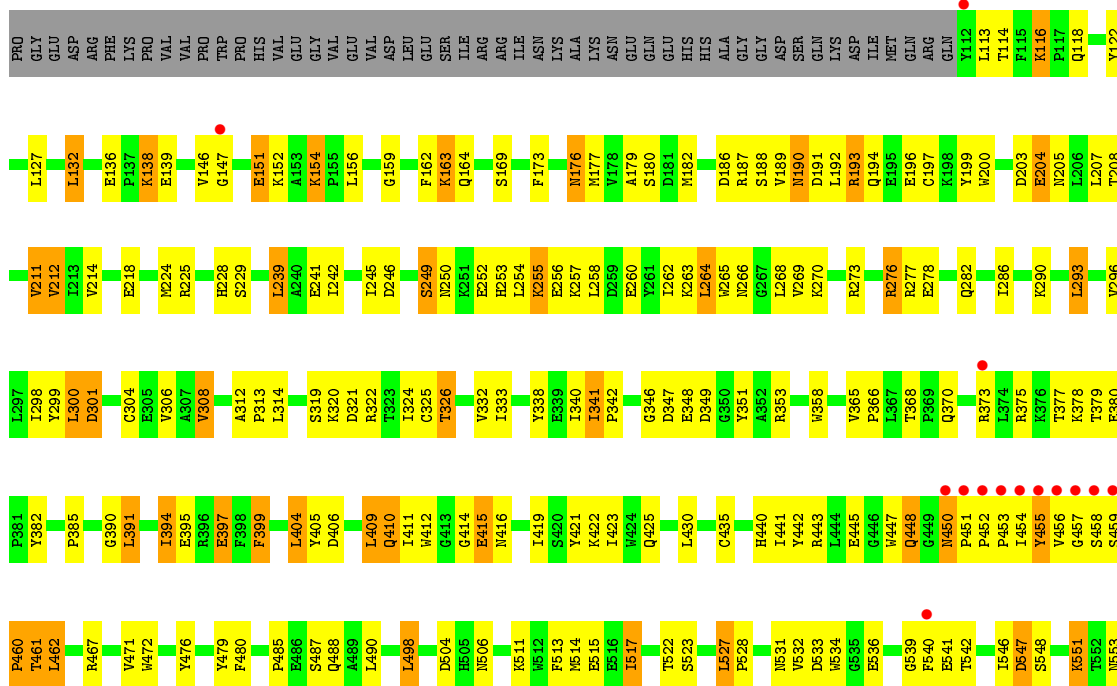
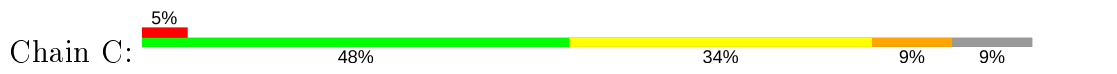


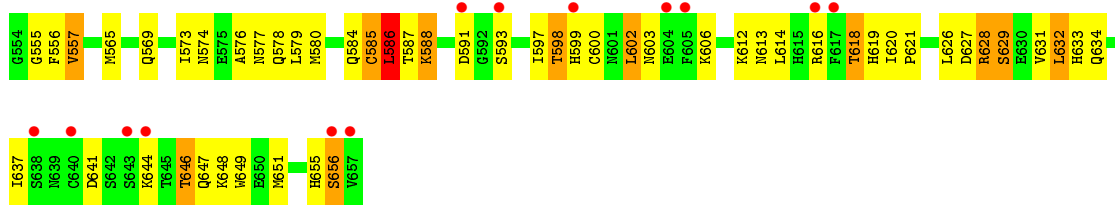
- Molecule 1: N-acetylgalactosaminyltransferase 7



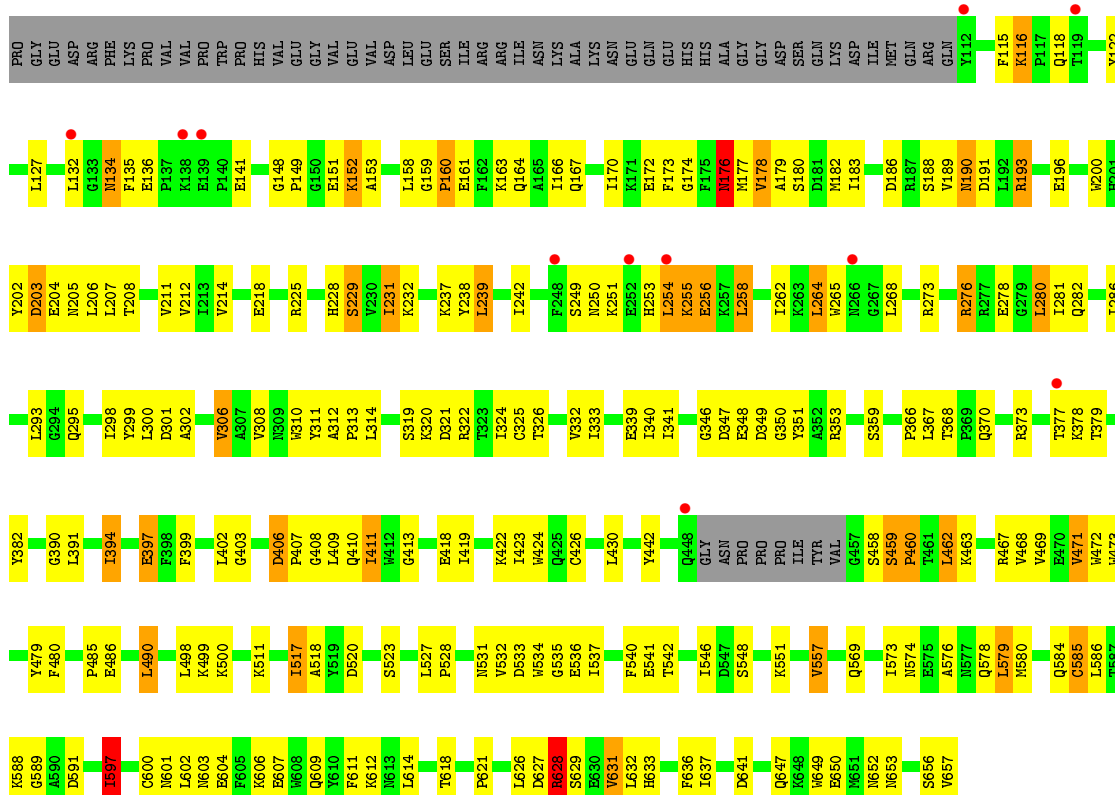


• Molecule 1: N-acetylgalactosaminyltransferase 7

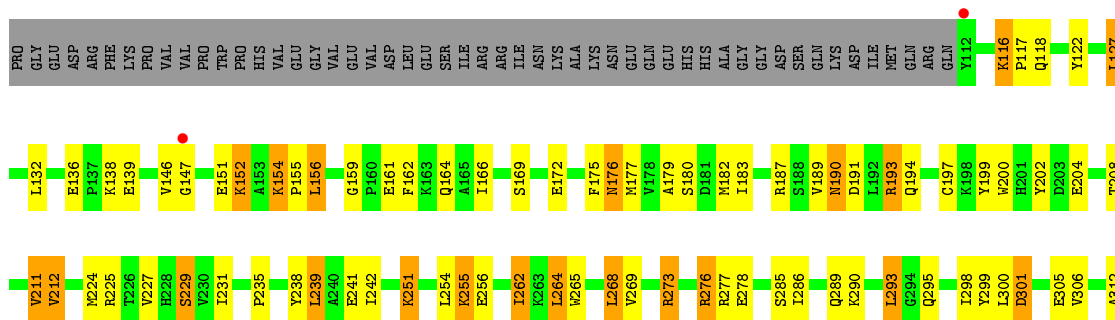


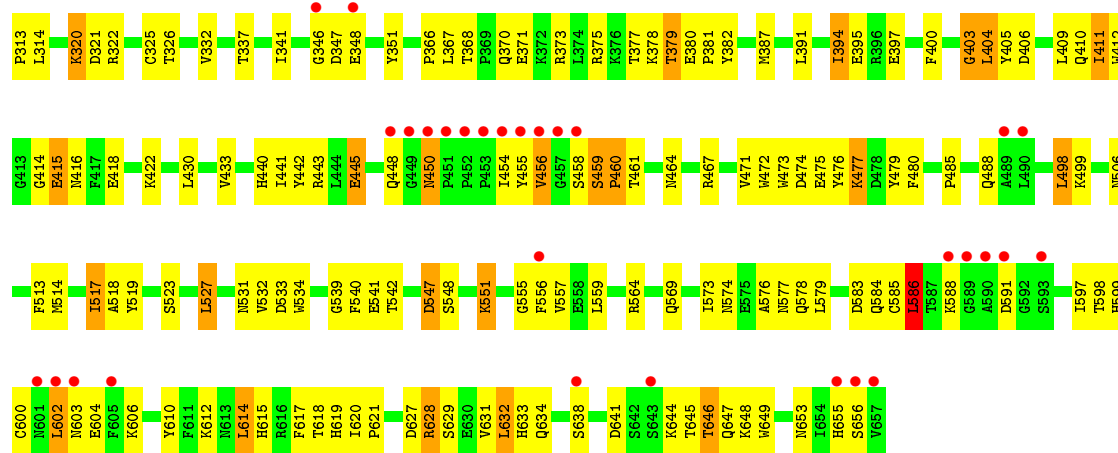


• Molecule 1: N-acetylgalactosaminyltransferase 7

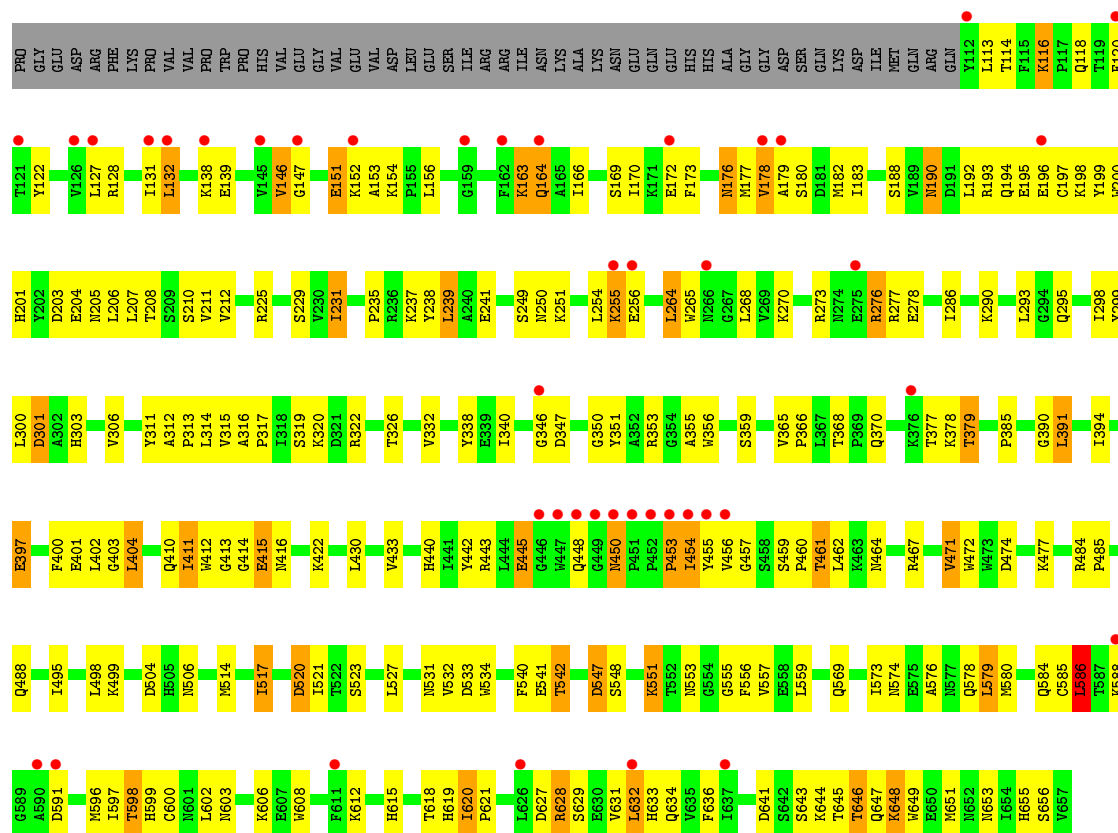


• Molecule 1: N-acetylgalactosaminyltransferase 7





• Molecule 1: N-acetylgalactosaminyltransferase 7



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	137.44Å 158.26Å 251.48Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.25 – 2.60 49.25 – 2.60	Depositor EDS
% Data completeness (in resolution range)	95.7 (49.25-2.60) 95.7 (49.25-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.84 (at 2.61Å)	Xtrriage
Refinement program	PHENIX (dev_2400: ???)	Depositor
R, R_{free}	0.220 , 0.258 0.220 , 0.258	Depositor DCC
R_{free} test set	8025 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	43.2	Xtrriage
Anisotropy	0.130	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 43.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	27098	wwPDB-VP
Average B, all atoms (Å ²)	47.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.65% 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: UD2, UDP, MN, NAG

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	2/4569 (0.0%)	0.71	2/6192 (0.0%)
1	B	0.55	2/4569 (0.0%)	0.70	4/6192 (0.1%)
1	C	0.54	3/4569 (0.1%)	0.68	4/6192 (0.1%)
1	D	0.54	4/4504 (0.1%)	0.70	6/6098 (0.1%)
1	E	0.51	3/4569 (0.1%)	0.66	3/6192 (0.0%)
1	F	0.47	0/4569	0.62	2/6192 (0.0%)
All	All	0.53	14/27349 (0.1%)	0.68	21/37058 (0.1%)

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	B	0	2
1	E	0	1
1	F	0	1
All	All	0	5

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	585	CYS	CB-SG	-7.43	1.69	1.82
1	B	585	CYS	CB-SG	-6.53	1.71	1.82
1	D	585	CYS	CB-SG	-5.51	1.72	1.81
1	E	585	CYS	CB-SG	-5.38	1.73	1.81
1	A	320	LYS	CD-CE	5.38	1.64	1.51

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	403	GLY	N-CA-C	8.30	133.84	113.10
1	B	148	GLY	C-N-CD	-7.82	103.39	120.60
1	A	320	LYS	CD-CE-NZ	6.68	127.06	111.70
1	E	403	GLY	N-CA-C	6.63	129.67	113.10
1	B	547	ASP	CB-CG-OD1	6.62	124.26	118.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	410	GLN	Peptide
1	B	402	LEU	Peptide
1	B	413	GLY	Peptide
1	E	403	GLY	Peptide
1	F	402	LEU	Peptide

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	4440	0	4325	212	0
1	B	4440	0	4325	254	0
1	C	4440	0	4325	244	0
1	D	4380	0	4265	228	0
1	E	4440	0	4325	200	0
1	F	4440	0	4325	227	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	D	15	15	15	1	0
4	D	25	11	11	2	0
5	D	39	0	25	3	0
6	A	100	0	0	8	0
6	B	77	0	0	6	0
6	C	60	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	D	55	0	0	10	0
6	E	68	0	0	1	0
6	F	47	0	0	4	0
All	All	27072	26	25941	1346	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 26.

The worst 5 of 1346 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:456:VAL:HG23	1:B:458:SER:CA	1.41	1.47
1:F:453:PRO:HB2	1:F:455:TYR:CZ	1.56	1.37
1:F:453:PRO:HB2	1:F:455:TYR:CE2	1.62	1.34
1:B:454:ILE:HG22	1:B:457:GLY:CA	1.62	1.27
1:C:454:ILE:O	1:C:458:SER:HA	1.39	1.22

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	544/597 (91%)	516 (95%)	26 (5%)	2 (0%)	34	57
1	B	544/597 (91%)	520 (96%)	24 (4%)	0	100	100
1	C	544/597 (91%)	515 (95%)	27 (5%)	2 (0%)	34	57
1	D	534/597 (89%)	513 (96%)	20 (4%)	1 (0%)	47	71
1	E	544/597 (91%)	518 (95%)	26 (5%)	0	100	100
1	F	544/597 (91%)	509 (94%)	32 (6%)	3 (1%)	25	47
All	All	3254/3582 (91%)	3091 (95%)	155 (5%)	8 (0%)	47	71

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	459	SER
1	F	411	ILE
1	F	413	GLY
1	D	176	ASN
1	C	453	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	Percentiles
1	A	480/525 (91%)	400 (83%)	80 (17%)	2 3
1	B	480/525 (91%)	395 (82%)	85 (18%)	2 3
1	C	480/525 (91%)	397 (83%)	83 (17%)	2 3
1	D	473/525 (90%)	418 (88%)	55 (12%)	5 10
1	E	480/525 (91%)	401 (84%)	79 (16%)	2 3
1	F	480/525 (91%)	406 (85%)	74 (15%)	2 4
All	All	2873/3150 (91%)	2417 (84%)	456 (16%)	2 4

5 of 456 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	395	GLU
1	D	229	SER
1	F	445	GLU
1	C	415	GLU
1	C	586	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 88 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	578	GLN
1	D	425	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	568	ASN
1	C	599	HIS
1	D	118	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](#)

Of 9 ligands modelled in this entry, 6 are monoatomic - leaving 3 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
3	NAG	D	701	-	15,15,15	1.56	2 (13%)	21,21,21	1.04	1 (4%)
4	UDP	D	702	2	20,26,26	1.69	4 (20%)	25,40,40	1.47	2 (8%)
5	UD2	D	704	-	34,41,41	3.85	19 (55%)	45,62,62	2.56	9 (20%)

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	NAG	D	701	-	-	1/6/26/26	0/1/1/1
4	UDP	D	702	2	-	7/14/32/32	0/2/2/2
5	UD2	D	704	-	-	2/24/63/63	0/3/3/3

The worst 5 of 25 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	D	704	UD2	PB-O1'	10.34	1.87	1.60
5	D	704	UD2	C6-N1	8.94	1.46	1.35
5	D	704	UD2	C4-N3	8.18	1.47	1.33
5	D	704	UD2	O4B-C1B	6.92	1.50	1.41
5	D	704	UD2	O4B-C4B	5.90	1.58	1.45

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	D	704	UD2	O3A-PB-O1'	-8.79	84.76	102.48
5	D	704	UD2	O1'-PB-O2B	-7.91	79.79	109.47
5	D	704	UD2	O4B-C1B-C2B	-6.38	97.60	106.93
5	D	704	UD2	PB-O3A-PA	-6.05	112.08	132.83
4	D	702	UDP	PA-O3A-PB	-5.25	114.81	132.83

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	D	704	UD2	C5B-O5B-PA-O3A
4	D	702	UDP	C2'-C1'-N1-C6
4	D	702	UDP	O4'-C1'-N1-C6
4	D	702	UDP	C5'-O5'-PA-O1A
3	D	701	NAG	O5-C5-C6-O6

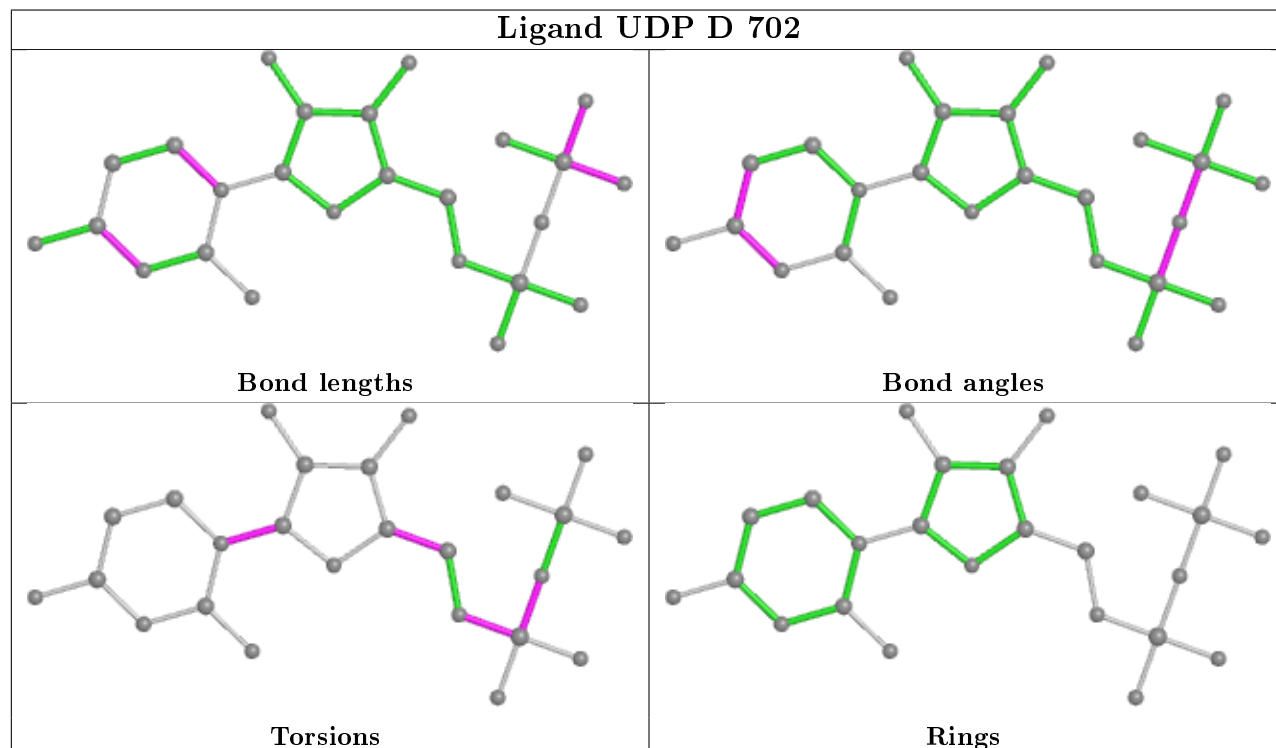
There are no ring outliers.

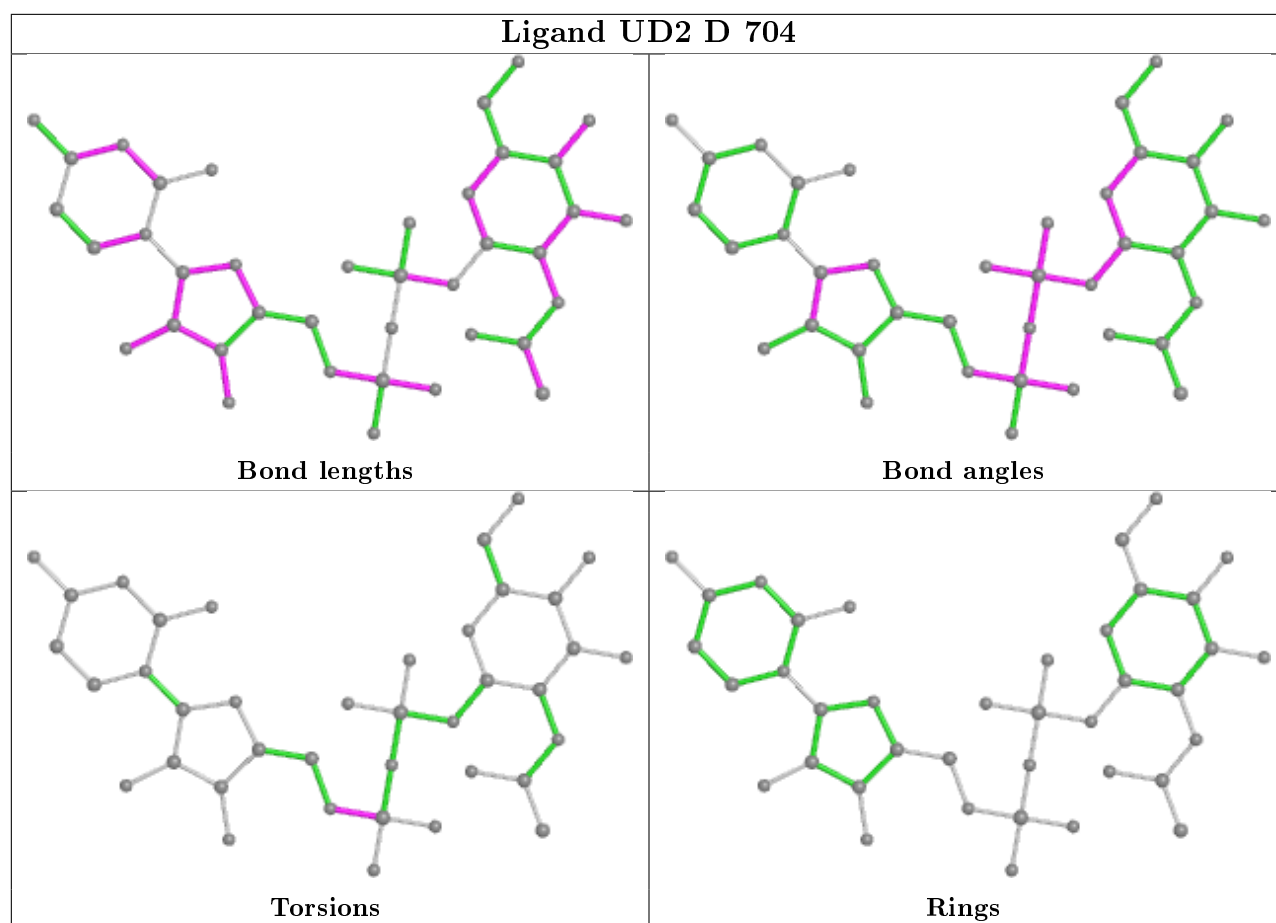
3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	701	NAG	1	0
4	D	702	UDP	2	0
5	D	704	UD2	3	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, 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	546/597 (91%)	-0.10	15 (2%) 54 48	23, 33, 59, 111	0
1	B	546/597 (91%)	-0.00	14 (2%) 56 50	26, 37, 64, 116	0
1	C	546/597 (91%)	0.22	27 (4%) 29 23	24, 42, 82, 138	0
1	D	538/597 (90%)	0.07	11 (2%) 65 60	29, 47, 75, 92	0
1	E	546/597 (91%)	0.19	32 (5%) 22 17	26, 46, 83, 154	0
1	F	546/597 (91%)	0.46	42 (7%) 13 10	34, 59, 86, 136	0
All	All	3268/3582 (91%)	0.14	141 (4%) 35 28	23, 43, 79, 154	0

The worst 5 of 141 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	458	SER	14.3
1	F	455	TYR	11.4
1	C	454	ILE	10.7
1	E	657	VAL	10.0
1	C	456	VAL	9.8

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](#)

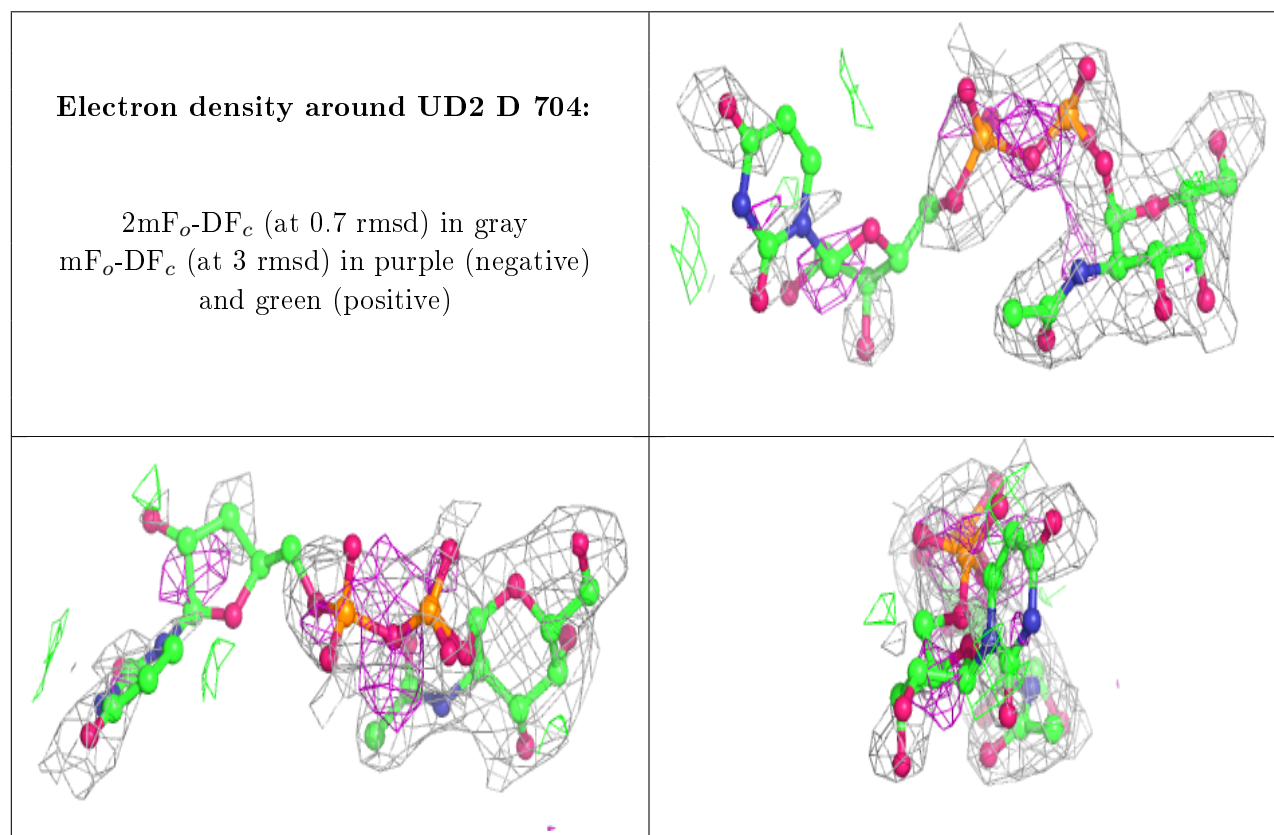
There are no monosaccharides in this entry.

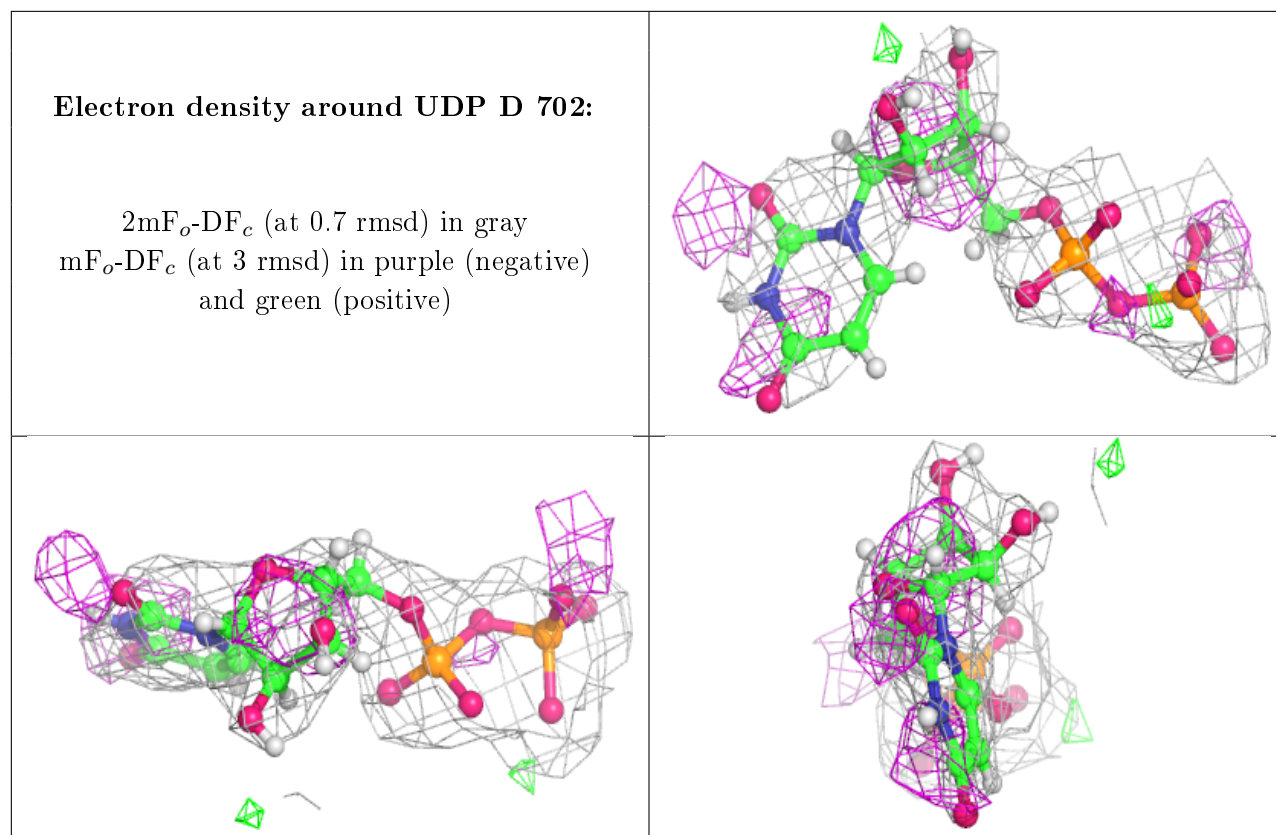
6.4 Ligands [i](#)

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(Å ²)	Q<0.9
2	MN	A	701	1/1	0.22	0.32	100,100,100,100	0
2	MN	F	701	1/1	0.66	0.32	91,91,91,91	0
2	MN	C	701	1/1	0.70	0.27	93,93,93,93	0
2	MN	E	701	1/1	0.78	0.19	82,82,82,82	0
3	NAG	D	701	15/15	0.83	0.33	52,68,79,82	0
5	UD2	D	704	39/39	0.85	0.33	36,81,113,113	0
2	MN	B	701	1/1	0.85	0.22	87,87,87,87	0
4	UDP	D	702	25/25	0.86	0.28	55,69,81,82	0
2	MN	D	703	1/1	0.97	0.14	56,56,56,56	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.