

wwPDB X-ray Structure Validation Summary Report (i)

Aug 9, 2020 – 12:04 PM BST

PDB ID : 4UFK

Title: Mouse Galactocerebrosidase complexed with dideoxy-imino-lyxitol DIL

Authors: Hill, C.H.; Viuff, A.H.; Spratley, S.J.; Salamone, S.; Christensen, S.H.; Read,

R.J.; Moriarty, N.W.; Jensen, H.H.; Deane, J.E.

Deposited on : 2015-03-17

Resolution : 2.40 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 (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 : 2.13.1

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)

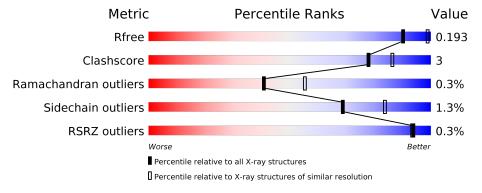
Validation Pipeline (wwPDB-VP) : 2.13.1

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 2.40 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	$\begin{array}{c} \text{Whole archive} \\ (\#\text{Entries}) \end{array}$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries,\ resolution\ range(\AA)}) \end{array}$
R_{free}	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	654	89%	9%	•
2	В	2	100%		
2	С	2	100%		



2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5429 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 GALACTOCEREBROSIDASE.

\mathbf{Mol}	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace			
1	A	640	Total 5140	C 3330	N 853	O 942	S 15	0	3	0	

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	15	HIS	=	expression tag	UNP P54818
A	16	HIS	-	expression tag	UNP P54818
A	17	HIS	-	expression tag	UNP P54818
A	18	HIS	-	expression tag	UNP P54818
A	19	HIS	-	expression tag	UNP P54818
A	20	HIS	-	expression tag	UNP P54818
A	21	ILE	_	expression tag	UNP P54818
A	22	GLU	-	expression tag	UNP P54818
A	23	GLY	-	expression tag	UNP P54818
A	24	ARG	_	expression tag	UNP P54818
A	25	GLY	_	expression tag	UNP P54818
A	26	ALA	-	expression tag	UNP P54818

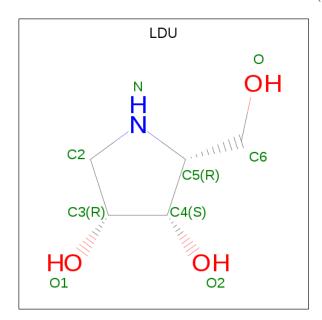
• Molecule 2 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-a cetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	В	2	Total C N O 28 16 2 10	0	0	0
2	С	2	Total C N O 28 16 2 10	0	0	0

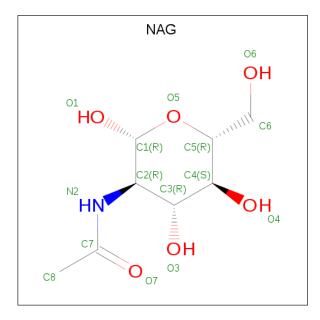


• Molecule 3 is DIDEOXY-IMINO-LYXITOL (three-letter code: LDU) (formula: C₅H₁₁NO₃).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf		
3	A	1	Total 9	C 5	N 1	O 3	0	0

• Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf		
4	A	1	Total	C 8	N 1	O 5	0	0

Continued on next page...



Continued from previous page...

	Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
	4	A	1	Total 14			O 5	0	0
-									
	4	A	1	Total 14	8		5	0	0

• Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Ca 1 1	0	0

• Molecule 6 is water.

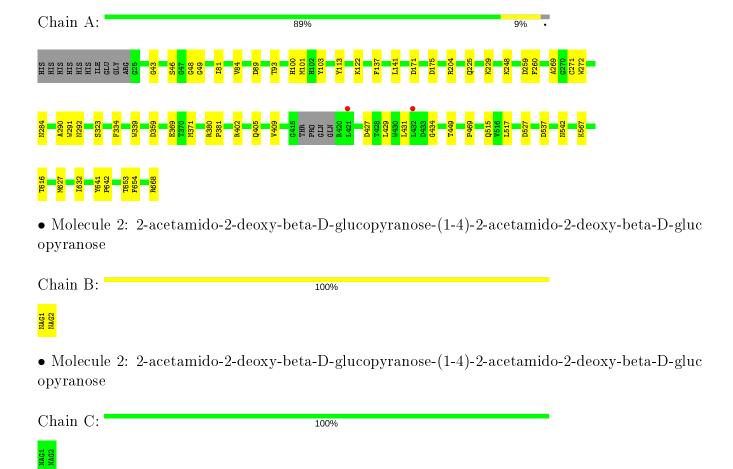
Mol	Chain	Residues	${f Atoms}$	ZeroOcc	AltConf
6	A	181	Total O 181 181	0	0



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: GALACTOCEREBROSIDASE





4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants	248.82Å 248.82Å 77.85Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 120.00°	Depositor
Resolution (Å)	73.22 - 2.40	Depositor
resolution (A)	73.22 - 2.40	EDS
% Data completeness	99.7 (73.22-2.40)	Depositor
(in resolution range)	95.0 (73.22-2.40)	EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	1.56 (at 2.40Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
P. P.	0.153 , 0.194	Depositor
R, R_{free}	0.156 , 0.193	DCC
R_{free} test set	1833 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å ²)	40.4	Xtriage
Anisotropy	0.227	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.32 , 40.0	EDS
L-test for twinning ²	$ < L > = 0.48, < L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	5429	wwPDB-VP
Average B, all atoms $(Å^2)$	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.12% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities 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: CA, NAG, LDU

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	
Ī	1	A	0.36	0/5312	0.49	0/7240	

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	$\mathbf{H}(\mathbf{model})$	H(added)	Clashes	Symm-Clashes
1	A	5140	0	4902	30	0
2	В	28	0	25	0	0
2	С	28	0	25	0	0
3	A	9	0	11	0	0
4	A	42	0	39	2	0
5	A	1	0	0	0	0
6	A	181	0	0	1	0
All	All	5429	0	5002	30	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 30 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.



Atom-1	Atom-2	$egin{aligned} ext{Interatomic} \ ext{distance} \ (ext{Å}) \end{aligned}$	$egin{aligned} ext{Clash} \ ext{overlap} & (ext{Å}) \end{aligned}$	
1:A:175:ASP:OD1	1:A:204:ARG:NH2	2.13	0.80	
1:A:402:ARG:NH1	1:A:434:GLY:O	2.23	0.72	
1:A:409:VAL:HG21	1:A:429:LEU:HD13	1.77	0.66	
1:A:627:MET:HG2	1:A:632:ILE:HD13	1.87	0.55	
1:A:269:ALA:HB2	1:A:323:SER:HB2	1.90	0.54	

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	639/654 (98%)	617 (97%)	20 (3%)	2 (0%)	41 55	

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	93	THR
1	A	260	PHE

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	536/555~(97%)	529 (99%)	7 (1%)	69 84		

5 of 7 residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	A	271	CYS
1	A	527	ASP
1	A	427	ASP
1	A	113	TYR
1	A	515	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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		
MIOI	Type				Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	NAG	В	1	1,2	14,14,15	1.63	1 (7%)	17,19,21	1.00	2 (11%)
2	NAG	В	2	2	14,14,15	1.10	2 (14%)	17,19,21	0.51	0
2	NAG	С	1	1,2	14,14,15	0.46	0	17,19,21	0.68	0
2	NAG	С	2	2	14,14,15	0.35	0	17,19,21	0.39	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	NAG	В	1	1,2	-	2/6/23/26	0/1/1/1
2	NAG	В	2	2	_	4/6/23/26	0/1/1/1
2	NAG	С	1	1,2	_	2/6/23/26	0/1/1/1
2	NAG	С	2	2	-	1/6/23/26	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(ext{\AA})$	$\operatorname{Ideal}(ext{\AA})$
2	В	1	NAG	O5-C1	-5.98	1.34	1.43
2	В	2	NAG	O5-C1	-2.98	1.39	1.43
2	В	2	NAG	C1-C2	2.71	1.56	1.52

All (2) bond angle outliers are listed below:

	Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^{o})$
Ī	2	В	1	NAG	C4-C3-C2	2.20	114.24	111.02
	2	В	1	NAG	O4-C4-C3	-2.19	105.28	110.35

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

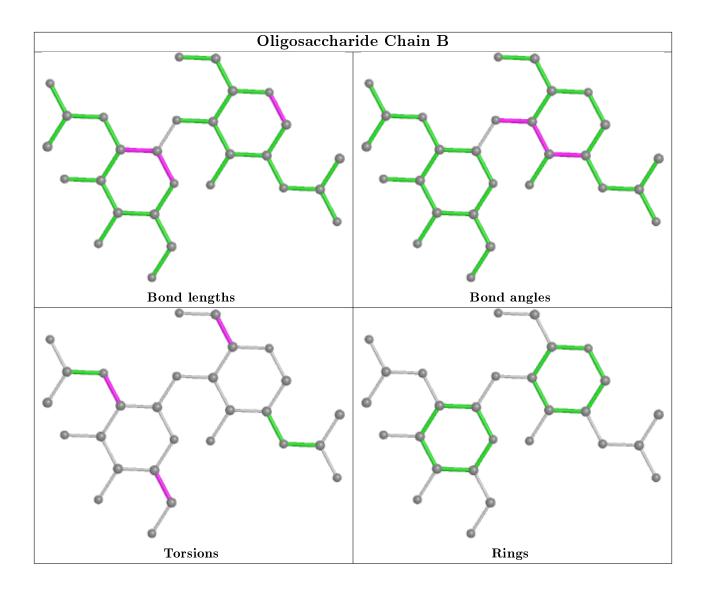
Mol	Chain	Res	Type	Atoms
2	В	2	NAG	O5-C5-C6-O6
2	В	1	NAG	C4-C5-C6-O6
2	В	2	NAG	C4-C5-C6-O6
2	В	1	NAG	O5-C5-C6-O6
2	В	2	NAG	C1-C2-N2-C7

There are no ring outliers.

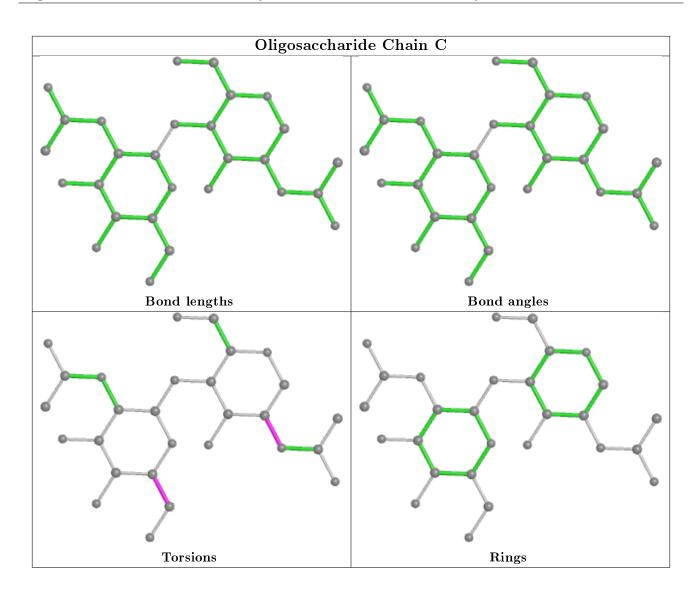
No monomer is involved in short contacts.

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









5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 1 is monoatomic - leaving 4 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	Tuno	Chain	Res	Link	Bo	Bond lengths			Bond angles		
10101	Type			LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2	
4	NAG	A	1387	1	14,14,15	0.45	0	17,19,21	0.59	1 (5%)	
4	NAG	A	2542	-	14,14,15	0.73	1 (7%)	17,19,21	0.53	0	
3	LDU	A	1001	_	9,9,9	0.41	0	6,12,12	0.45	0	



Mol	Type	Chain	Chain	Res	Link	Bond lengths			Bond angles		
			nes	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	$\mid \# Z > 2 \mid$	
4	NAG	A	1542	-	14,14,15	0.71	1 (7%)	17,19,21	0.52	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
4	NAG	A	1387	1	-	0/6/23/26	0/1/1/1
4	NAG	A	2542	-	-	2/6/23/26	0/1/1/1
3	LDU	A	1001	-	-	0/2/15/15	0/1/1/1
4	NAG	A	1542	-	-	2/6/23/26	0/1/1/1

All (2) bond length outliers are listed below:

	Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\mathbf{Observed}(\mathbf{\mathring{A}})$	$\operatorname{Ideal}(ext{\AA})$
Ī	4	A	2542	NAG	O5-C1	-2.40	1.39	1.43
	4	A	1542	NAG	O5-C1	-2.07	1.40	1.43

All (1) bond angle outliers are listed below:

\mathbf{Mol}	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^{o})$
4	A	1387	NAG	C1-O5-C5	2.09	115.02	112.19

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	2542	NAG	O5-C5-C6-O6
4	A	2542	NAG	C4-C5-C6-O6
4	A	1542	NAG	C4-C5-C6-O6
4	A	1542	NAG	O5-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1542	NAG	2	0



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	$\langle { m RSRZ} \rangle$	$\#\mathrm{RSRZ}{>}2$	$OWAB(A^2)$	Q < 0.9
1	A	640/654 (97%)	-0.50	2 (0%) 94 93	26, 45, 80, 105	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	421	LEU	2.4
1	A	432	LEU	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

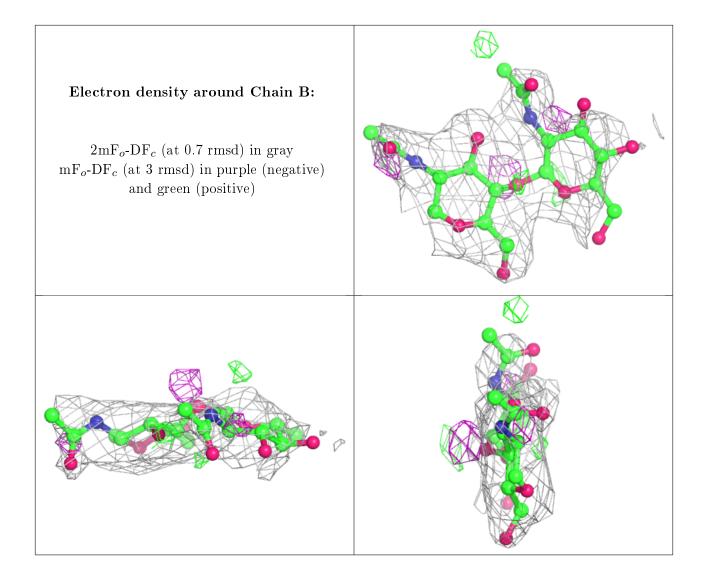
6.3 Carbohydrates (i)

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.

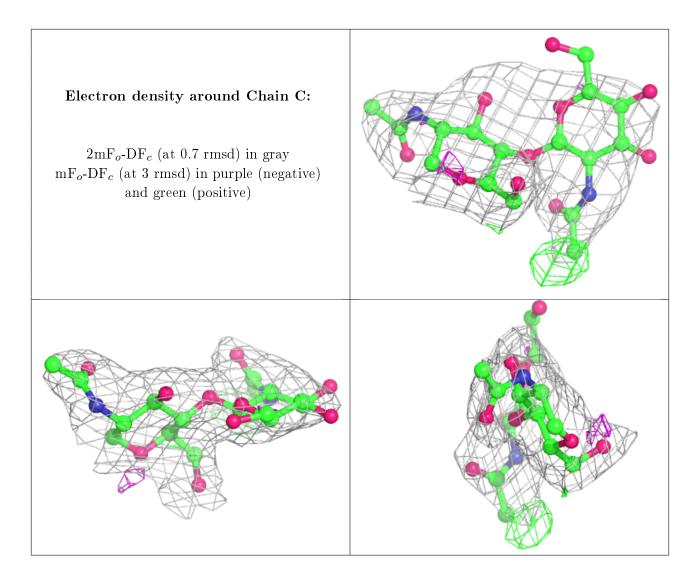
Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
2	NAG	В	2	14/15	0.82	0.27	103,113,116,117	0
2	NAG	С	2	14/15	0.87	0.17	100,107,113,113	0
2	NAG	В	1	14/15	0.90	0.14	57,71,78,92	0
2	NAG	С	1	14/15	0.94	0.09	54,69,79,91	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.









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
4	NAG	A	1387	14/15	0.76	0.24	81,95,103,103	0
4	NAG	A	1542	14/15	0.85	0.15	53,67,74,76	0
4	NAG	A	2542	14/15	0.89	0.20	80,88,93,93	0
3	LDU	A	1001	9/9	0.98	0.12	32,34,38,41	0
5	CA	A	3669	1/1	0.99	0.07	55,55,55,55	0



6.5 Other polymers (i)

There are no such residues in this entry.

