

# Full wwPDB X-ray Structure Validation Report (i)

Oct 23, 2021 – 11:32 AM EDT

PDB ID : 6CEL

Title : CBH1 (E212Q) CELLOPENTAOSE COMPLEX

Authors: Divne, C.; Stahlberg, J.; Jones, T.A.

Deposited on : 1997-09-24

Resolution : 1.70 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 : 2.23.2

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)

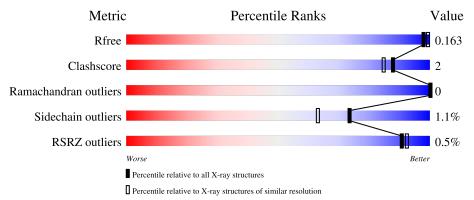
 $Validation\ Pipeline\ (wwPDB-VP) \quad : \quad 2.23.2$ 

## 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 1.70 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	Similar resolution		
Metric	$(\# \mathrm{Entries})$	$(\#  ext{Entries},  ext{ resolution range}( ext{Å}))$		
$R_{free}$	130704	4298 (1.70-1.70)		
Clashscore	141614	4695 (1.70-1.70)		
Ramachandran outliers	138981	4610 (1.70-1.70)		
Sidechain outliers	138945	4610 (1.70-1.70)		
RSRZ outliers	127900	4222 (1.70-1.70)		

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 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 <=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	434	94%	6%				
2	В	4	100%					
3	С	5	80%	20%				



# 2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 3840 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.

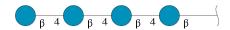
• Molecule 1 is a protein called 1,4-BETA-D-GLUCAN CELLOBIOHYDROLASE I.

$\mathbf{Mol}$	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	A	434	Total 3224	C 1977	N 534	O 687	S 26	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Α	94	ASP	GLY	cloning artifact	UNP P00725
A	212	GLN	GLU	engineered mutation	UNP P00725

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



Mo	l Chai	in	Residues	Atoms		ZeroOcc	AltConf	Trace	
2	В		4	Total 45	C 24		0	0	0

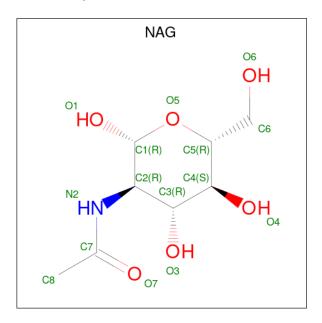
• Molecule 3 is an oligosaccharide called beta-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-beta-D-glucopyranose.



Mol	Chain	Residues	At	oms		ZeroOcc	AltConf	Trace
3	С	5	Total 56	C 30	O 26	0	0	0



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



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

• Molecule 5 is COBALT (II) ION (three-letter code: CO) (formula: Co).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	2	Total Co 2 2	0	0

• Molecule 6 is water.

$\mathbf{Mol}$	Chain	Residues	${f Atoms}$	ZeroOcc	AltConf
6	A	485	Total O 485 485	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: 1,4-BETA-D-GLUCAN CELLOBIOHYDROLASE I
Chain A: 94% 6%
Molecule 2: beta-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-beta-D-glucopyranose
Chain B: 100%
Molecule 3: beta-D-glucopyranose-(1-4)-beta-D-gluco



# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants	83.05Å 82.95Å 110.85Å	Donositon
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	7.50 - 1.70	Depositor
Resolution (A)	66.41 - 1.65	EDS
% Data completeness	(Not available) (7.50-1.70)	Depositor
(in resolution range)	92.8 (66.41-1.65)	EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	4.31  (at  1.65Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
D.D.	0.190 , $0.235$	Depositor
$R, R_{free}$	0.166 , $0.163$	DCC
$R_{free}$ test set	2000 reflections $(4.66\%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	9.1	Xtriage
Anisotropy	0.840	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	$0.35\;,52.2$	EDS
L-test for twinning <sup>2</sup>	$< L > = 0.49, < L^2> = 0.33$	Xtriage
Estimated twinning fraction	0.016 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3840	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	11.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 6.15% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



<sup>&</sup>lt;sup>1</sup>Intensities 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: NAG, PCA, BGC, CO

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).

ſ	Mol	Chain	Bond	lengths	Bond angles		
			RMSZ	# Z  > 5	RMSZ	# Z  > 5	
Ī	1	A	0.35	0/3292	0.67	0/4488	

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	3224	0	2924	13	0
2	В	45	0	39	0	0
3	С	56	0	48	0	0
4	A	28	0	26	0	0
5	A	2	0	0	0	0
6	A	485	0	0	1	0
All	All	3840	0	3037	13	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 2.

All (13) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.



Atom-1	Atom-2	$egin{array}{ll}  ext{Interatomic} \  ext{distance } ( ext{Å}) \end{array}$	Clash overlap (Å)
1:A:54:ASN:OD1	1:A:55:THR:HG22	2.11	0.50
1:A:198:ASN:HB3	1:A:201:THR:OG1	2.11	0.49
1:A:381:TYR:HA	1:A:382:PRO:C	2.33	0.49
1:A:213:MET:HB2	1:A:352:PHE:HZ	1.79	0.47
1:A:227:PRO:HD2	1:A:261:CYS:O	2.18	0.44
1:A:313:GLN:HE21	1:A:313:GLN:HA	1.82	0.43
1:A:123:LEU:HD21	1:A:353:LYS:HE3	2.00	0.43
1:A:209:CYS:O	1:A:210:CYS:HB3	2.20	0.41
1:A:173:ASP:OD1	1:A:177:PRO:HD3	2.20	0.41
1:A:137:PRO:HD2	1:A:140:LEU:HD22	2.02	0.41
1:A:186:GLN:HG2	6:A:805:HOH:O	2.21	0.41
1:A:54:ASN:ND2	1:A:194:PRO:HB2	2.36	0.41
1:A:257:ASP:OD2	1:A:260:GLY:HA2	2.21	0.40

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	432/434 (100%)	427 (99%)	5 (1%)	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	356/356 (100%)	352 (99%)	4 (1%)	73 63		

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	55	THR
1	A	223	GLU
1	A	249	ASP
1	A	313	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	98	GLN
1	A	313	GLN
1	A	373	ASN

#### 5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	B	ond leng	$\operatorname{gths}$	В	ond ang	gles
				Lilik	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
1	PCA	A	1	1	7,8,9	0.91	0	9,10,12	1.42	1 (11%)

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
1	PCA	A	1	1	-	0/0/11/13	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$\operatorname{Observed}(^{o})$	$\operatorname{Ideal}({}^{o})$
1	A	1	PCA	CB-CA-C	-2.69	109.00	112.70

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

#### 5.5 Carbohydrates (i)

9 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	Во	ond leng	ths	В	ond ang	cles
MIOI	Type	Chain	nes	Les   Lilik	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	BGC	В	1	2	12,12,12	0.37	0	17,17,17	0.36	0
2	BGC	В	2	2	11,11,12	0.38	0	15,15,17	0.39	0
2	BGC	В	3	2	11,11,12	0.34	0	15,15,17	0.43	0
2	BGC	В	4	2	11,11,12	0.36	0	15,15,17	0.44	0
3	BGC	С	1	3	12,12,12	0.28	0	17,17,17	0.56	0
3	BGC	С	2	3	11,11,12	0.39	0	15,15,17	0.84	1 (6%)
3	BGC	С	3	3	11,11,12	0.46	0	15,15,17	0.54	0
3	BGC	С	4	3	11,11,12	0.35	0	15,15,17	0.47	0
3	BGC	С	5	3	11,11,12	0.39	0	15,15,17	0.36	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	BGC	В	1	2	-	0/2/22/22	0/1/1/1
2	BGC	В	2	2	-	0/2/19/22	0/1/1/1
2	BGC	В	3	2	-	0/2/19/22	0/1/1/1
2	BGC	В	4	2	-	0/2/19/22	0/1/1/1
3	BGC	С	1	3	-	0/2/22/22	0/1/1/1
3	BGC	С	2	3	-	0/2/19/22	0/1/1/1
3	BGC	С	3	3	-	0/2/19/22	0/1/1/1
3	BGC	С	4	3	-	0/2/19/22	0/1/1/1
3	BGC	С	5	3	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}({}^o)$	
3	С	2	BGC	C1-C2-C3	2.73	113.03	109.67	

There are no chirality outliers.

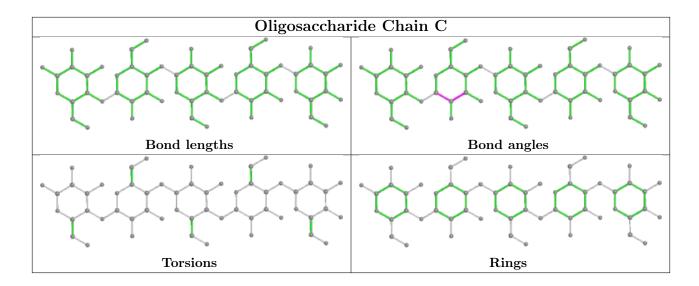
There are no torsion outliers.

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 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	Chain	Res	Link	Bo	ond leng	ths	В	ond ang	les
MIOI	Type		nes	LillK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2	
4	NAG	A	435	1	14,14,15	0.59	0	17,19,21	0.59	0	
4	NAG	A	436	1	14,14,15	0.40	0	17,19,21	0.63	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
ĺ	4	NAG	A	435	1	-	0/6/23/26	0/1/1/1
	4	NAG	A	436	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.



There are no ring outliers.

No monomer is involved in short contacts.

# 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	<rsrz></rsrz>	> #RSRZ $>$ 2		$OWAB(Å^2)$	Q < 0.9	
1	A	433/434 (99%)	-0.35	2 (0%)	91	92	4, 8, 22, 47	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	253	GLY	2.3
1	A	434	GLY	2.2

#### 6.2 Non-standard residues in protein, DNA, RNA chains (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
1	PCA	A	1	8/9	0.95	0.07	7,7,9,10	0

#### 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	BGC	В	1	12/12	0.54	0.20	37,40,41,42	0
2	BGC	В	2	11/12	0.85	0.14	26,31,33,33	0
2	BGC	В	4	11/12	0.90	0.11	15,16,17,17	0
3	BGC	С	5	11/12	0.90	0.12	14,16,17,17	0

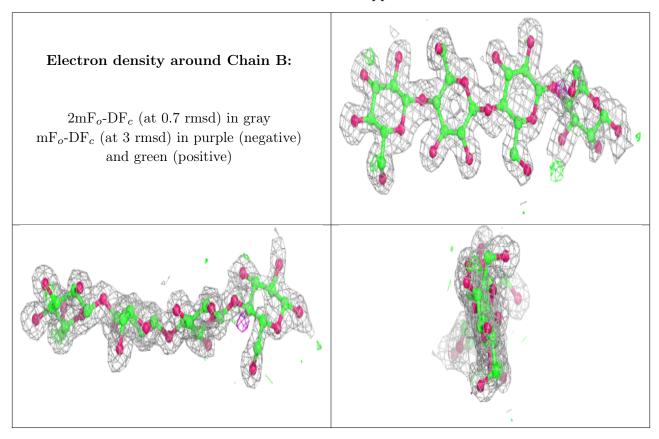
Continued on next page...



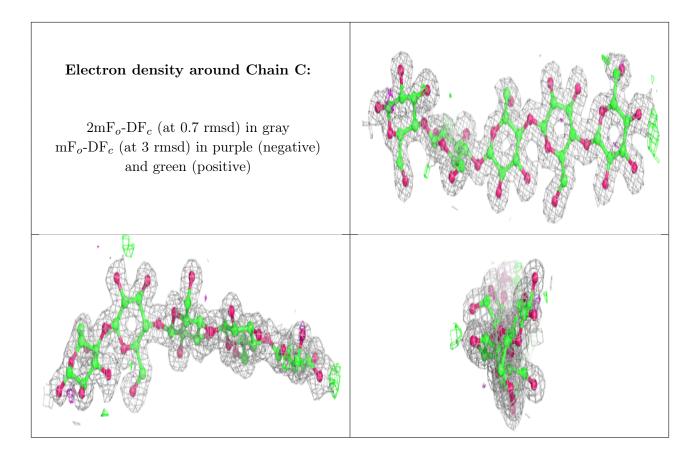
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
3	BGC	С	1	12/12	0.92	0.13	13,14,17,18	0
2	BGC	В	3	11/12	0.92	0.10	17,19,21,22	0
3	BGC	С	2	11/12	0.95	0.09	10,12,13,13	0
3	BGC	С	4	11/12	0.97	0.07	8,9,11,12	0
3	BGC	С	3	11/12	0.97	0.07	8,10,10,11	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	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
4	NAG	A	436	14/15	0.80	0.17	51,51,53,53	0
4	NAG	A	435	14/15	0.92	0.10	9,12,15,15	0
5	CO	A	461	1/1	0.94	0.08	33,33,33,33	0
5	CO	A	1000	1/1	1.00	0.06	7,7,7,7	1

### 6.5 Other polymers (i)

There are no such residues in this entry.

