

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

#### Aug 9, 2020 – 02:51 PM BST

PDB ID	:	$2 \mathrm{GMM}$
Title	:	Metal-free (apo) P. angolensis seed lectin in complex with $Man-alpha(1-2)Man$
Authors	:	Garcia-Pino, A.; Buts, L.; Wyns, L.; Loris, R.
Deposited on		
Resolution	:	2.15 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:

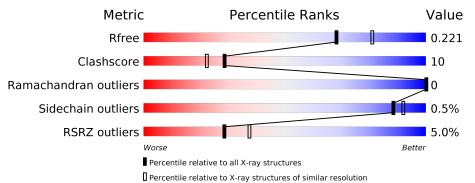
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
$\rm CCP4$	:	$7.0.044 (\mathrm{Gargrove})$
Ideal geometry (proteins)	:	Engh & 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 2.15 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries},{ m resolution\ range}({ m \AA}))$
$R_{free}$	130704	1479 (2.16-2.16)
Clashscore	141614	1585(2.16-2.16)
Ramachandran outliers	138981	$1560 \ (2.16-2.16)$
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	А	252	71%	19%	10%		
1	В	252	4%	17%	8%		
2	С	2	100%				
2	D	2	100%				



#### $2 \mathrm{GMM}$

# 2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3817 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 lectin.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace	
1	Λ	227	Total	С	Ν	Ο	0	6	1
1			1760	1122	293	345	0		
1	В	232	Total	С	Ν	Ο	0	5	1
	D	232	1773	1130	296	347			

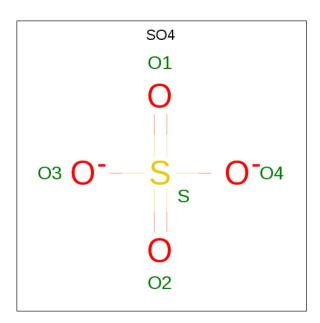
• Molecule 2 is an oligosaccharide called alpha-D-mannopyranose-(1-2)-alpha-D-mannopyran ose.



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	С	2	Total         C         O           23         12         11	0	0	0
2	D	2	Total         C         O           23         12         11	0	0	0

• Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
3	В	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0

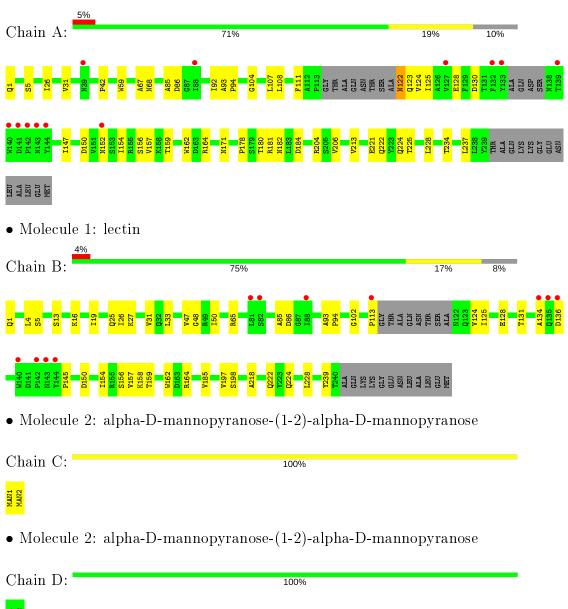
• Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	128	Total O 128 128	0	0
4	В	100	Total O 100 100	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: lectin



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	58.71Å $62.59$ Å $129.96$ Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	25.00 - 2.15	Depositor
Resolution (A)	24.60 - 2.15	EDS
% Data completeness	(Not available) $(25.00-2.15)$	Depositor
(in resolution range)	98.8(24.60-2.15)	EDS
R <sub>merge</sub>	(Not available)	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$8.85 (at 2.15 \text{\AA})$	Xtriage
Refinement program	$CNS \ 1.1$	Depositor
$R, R_{free}$	0.185 , $0.227$	Depositor
It, It free	0.178 , $0.221$	DCC
$R_{free}$ test set	2087 reflections $(7.90%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	27.2	Xtriage
Anisotropy	0.396	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.34 , $60.2$	EDS
L-test for twinning <sup>2</sup>	$ \langle L  \rangle = 0.49, \langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3817	wwPDB-VP
Average B, all atoms $(Å^2)$	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.70% of the height of the origin peak. No significant pseudotranslation is detected.

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



<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: PCA, SO4, MAN  $\,$ 

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		
	Unam	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.34	0/1815	0.66	0/2475	
1	В	0.34	0/1824	0.67	0/2490	
All	All	0.34	0/3639	0.67	0/4965	

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	А	1760	0	1673	39	0
1	В	1773	0	1681	30	0
2	С	23	0	21	3	0
2	D	23	0	21	0	0
3	А	5	0	0	0	0
3	В	5	0	0	0	0
4	А	128	0	0	2	0
4	В	100	0	0	0	0
All	All	3817	0	3396	68	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 10.



Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:152:ASN:ND2	1:A:206:VAL:HG12	1.54	1.21
1:A:152:ASN:HD21	1:A:206:VAL:CG1	1.53	1.20
1:A:152:ASN:ND2	1:A:206:VAL:CG1	2.11	1.09
1:A:152:ASN:HD21	1:A:206:VAL:HG12	0.88	1.04
1:A:164:ARG:HE	1:A:224:GLN:HE22	1.20	0.83

The worst 5 of 68 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

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	Analysed Favoured Allowed		Outliers	Perce	$\mathbf{n}$ tiles
1	А	227/252~(90%)	217~(96%)	10 (4%)	0	100	100
1	В	233/252~(92%)	225~(97%)	8 (3%)	0	100	100
All	All	460/504~(91%)	442 (96%)	18 (4%)	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	А	194/215~(90%)	193~(100%)	1 (0%)	88 92		

Continued on next page...



Continued from previous page...

Mol	Chain	Analysed	Analysed Rotameric Outliers		Percentiles		
1	В	192/215~(89%)	191~(100%)	1 (0%)	88 92		
All	All	386/430~(90%)	384 (100%)	2(0%)	88 92		

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type	
1	А	122	ASN	
1	В	222	GLN	

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	А	224	GLN
1	В	224	GLN
1	В	25	GLN
1	А	122	ASN
1	В	41	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)

2 non-standard protein/DNA/RNA residues 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).

Mal	Mol Type Chain Res Li		Timle	Link Bond lengths			Bond angles			
INIOI	Type	Chain	$\mathbf{Res}$		Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z >2
1	PCA	А	1	1	7,8,9	2.78	2 (28%)	$9,\!10,\!12$	1.92	2 (22%)
1	PCA	В	1	1	7,8,9	2.68	2 (28%)	$9,\!10,\!12$	1.94	3 (33%)

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	А	1	1	-	0/0/11/13	0/1/1/1
1	PCA	В	1	1	-	0/0/11/13	0/1/1/1

Mol	Chain	$\mathbf{Res}$	Type	Atoms		$\operatorname{Observed}(\operatorname{\AA})$	$\operatorname{Ideal}(\operatorname{\AA})$
1	В	1	PCA	CD-N	6.03	1.50	1.34
1	А	1	PCA	CD-N	6.03	1.50	1.34
1	А	1	PCA	CA-N	3.79	1.50	1.46
1	В	1	PCA	CA-N	3.37	1.50	1.46

All (4) bond length outliers are listed below:

All (5) bond angle outliers are listed below:

Mol	Chain	$\mathbf{Res}$	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
1	А	1	PCA	OE-CD-CG	-3.32	120.97	126.76
1	В	1	PCA	OE-CD-CG	-3.28	121.03	126.76
1	В	1	PCA	OE-CD-N	2.83	131.45	124.86
1	А	1	PCA	OE-CD-N	2.75	131.26	124.86
1	В	1	PCA	CB-CA-C	-2.03	109.91	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)

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		Bo	Bond lengths			Bond angles			
	Type	Unam	nes		Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	MAN	С	1	2	12,12,12	0.44	0	$17,\!17,\!17$	0.48	0
2	MAN	С	2	2	11,11,12	0.46	0	$15,\!15,\!17$	0.67	0
2	MAN	D	1	2	12,12,12	0.39	0	17,17,17	0.38	0
2	MAN	D	2	2	11,11,12	0.45	0	$15,\!15,\!17$	0.61	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	$\mathbf{Res}$	Link	Chirals	Torsions	Rings
2	MAN	С	1	2	-	0/2/22/22	0/1/1/1
2	MAN	С	2	2	-	0/2/19/22	0/1/1/1
2	MAN	D	1	2	-	0/2/22/22	0/1/1/1
2	MAN	D	2	2	-	0/2/19/22	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.

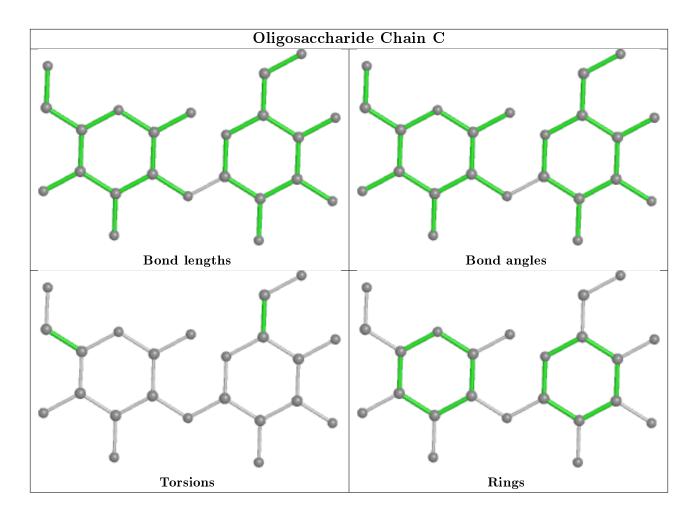
2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	С	1	MAN	1	0
2	С	2	MAN	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 oligosaccharide.

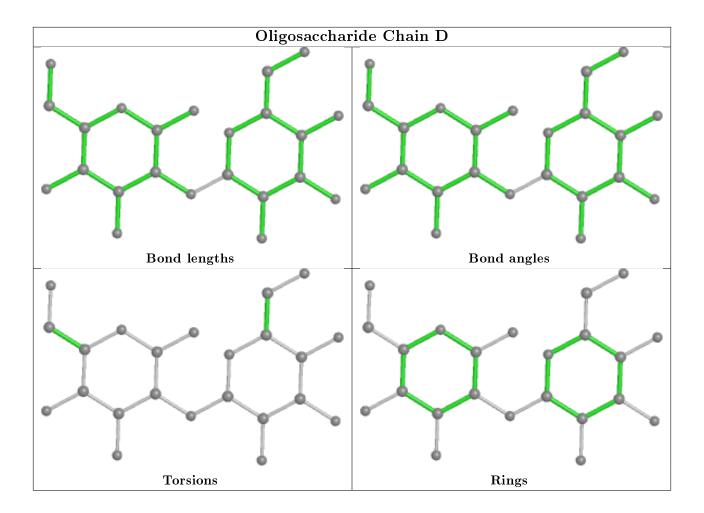












## 5.6 Ligand geometry (i)

2 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	True	Chain	Dec	Res Link	Bond lengths			Bond angles		
INIOI	Type	Chain	nes		Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z >2
3	SO4	В	306	-	4,4,4	0.25	0	$^{6,6,6}$	0.05	0
3	SO4	А	305	-	4,4,4	0.24	0	$^{6,6,6}$	0.05	0

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	${f Analysed}  < {f RSRZ} >  \# {f RSRZ} > 2$		$OWAB(Å^2)$	Q<0.9
1	А	226/252 (89%)	0.07	12 (5%) 26 35	17, 26, 43, 58	0
1	В	231/252 (91%)	-0.01	11 (4%) 30 39	16, 25, 49, 76	0
All	All	457/504~(90%)	0.03	23 (5%) 28 37	16, 25, 49, 76	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	140	TRP	6.2
1	А	132[A]	PHE	5.9
1	В	135	GLN	5.7
1	А	142	PRO	5.4
1	А	143	ASN	4.7

## 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	$\mathbf{B} extsf{-}\mathbf{B} extsf{-}\mathbf{factors}(\mathbf{A}^2)$	Q < 0.9
1	PCA	А	1	8/9	0.95	0.13	$20,\!24,\!24,\!25$	0
1	PCA	В	1	8/9	0.96	0.11	18,22,24,24	0

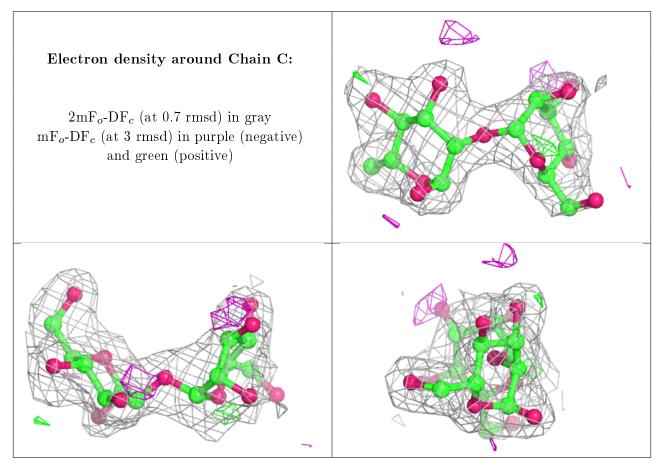
## 6.3 Carbohydrates (i)

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.

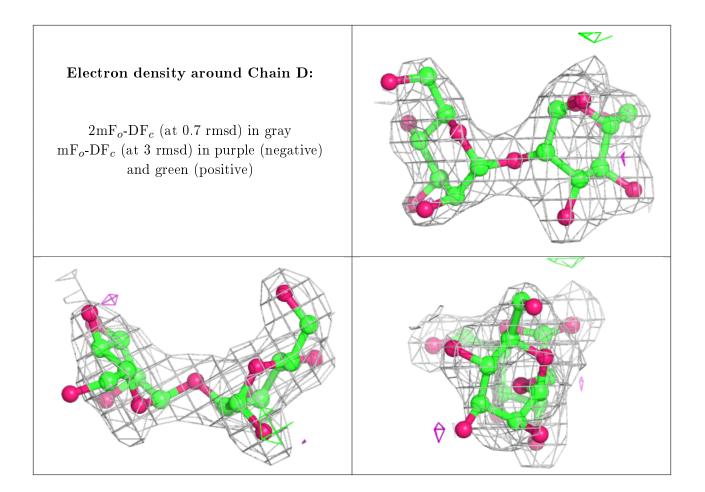


Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-factors}(\mathbf{\AA}^2)$	$Q{<}0.9$
2	MAN	D	2	11/12	0.89	0.27	$50,\!55,\!61,\!63$	0
2	MAN	С	2	11/12	0.89	0.19	$45,\!49,\!55,\!62$	0
2	MAN	С	1	12/12	0.94	0.12	$26,\!32,\!37,\!40$	0
2	MAN	D	1	12/12	0.95	0.10	$29,\!32,\!35,\!41$	0

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	$\mathbf{B} ext{-factors}(\mathbf{\AA}^2)$	$Q{<}0.9$
3	SO4	В	306	5/5	0.94	0.11	$35,\!37,\!38,\!43$	5
3	SO4	А	305	5/5	0.94	0.10	$56,\!58,\!58,\!60$	5

## 6.5 Other polymers (i)

There are no such residues in this entry.

