

Full wwPDB X-ray Structure Validation Report (i)

Aug 8, 2020 – 12:34 PM BST

PDB ID	:	4RUQ
Title	:	Carp Fishelectin, apo form
Authors	:	Capaldi, S.; Faggion, B.; Carrizo, M.E.; Destefanis, L.; Gonzalez, M.C.; Perd-
		uca, M.; Bovi, M.; Galliano, M.; Monaco, H.L.
Deposited on		
$\operatorname{Resolution}$:	1.35 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:

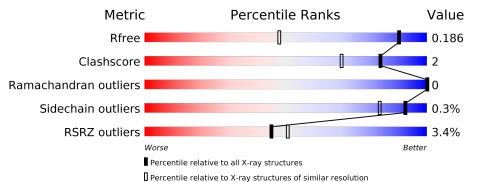
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
$\operatorname{CCP4}$:	$7.0.044 (\mathrm{Gargrove})$
Ideal geometry (proteins)	:	Engh & Huber (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 1.35 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	1509(1.38-1.34)
Clashscore	141614	1551(1.38-1.34)
Ramachandran outliers	138981	1530(1.38-1.34)
Sidechain outliers	138945	1530 (1.38-1.34)
RSRZ outliers	127900	1487 (1.38-1.34)

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	238	94%	5% •
1	В	238	^{2%} 95%	•
2	С	2	50% 50%	
2	D	2	100%	



2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 4029 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 Fish-egg lectin.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	А	236		С 1117		0 342	S 13	0	0	0
1	В	237	Total 1778	C 1121		0 344	S 13	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	119	LEU	ILE	$\operatorname{conflict}$	UNP P68512
В	119	LEU	ILE	$\operatorname{conflict}$	UNP P68512

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



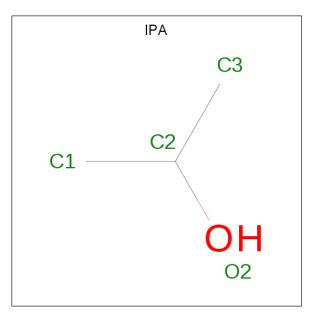
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	С	2	Total C N O 28 16 2 10	0	0	0
2	D	2	Total C N O 28 16 2 10	0	0	0

• Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	В	1	Total Ca 1 1	0	0
3	А	1	Total Ca 1 1	0	0



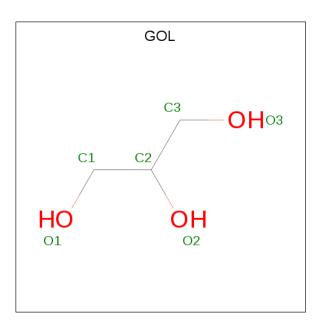
• Molecule 4 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C₃H₈O).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 3 & 1 \end{array}$	0	0
4	A	1	Total C O	0	0
4	В	1	Total C O	0	0
4	B	1	$\begin{array}{ccc} 4 & 3 & 1 \\ \hline \text{Total} & \text{C} & \text{O} \end{array}$	0	0
4			$\begin{array}{c ccc} 4 & 3 & 1 \\ \hline \text{Total} & \text{C} & \text{O} \end{array}$	0	0
4	4 B	1	4 3 1	0	0

• Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	1	$\begin{array}{ccc} {\rm Total} & {\rm C} & {\rm O} \\ 6 & 3 & 3 \end{array}$	0	0
5	В	1	$\begin{array}{ccc} {\rm Total} & {\rm C} & {\rm O} \\ 6 & 3 & 3 \end{array}$	0	0
5	В	1	$\begin{array}{ccc} \text{Total} \text{C} \text{O} \\ 6 3 3 \end{array}$	0	0

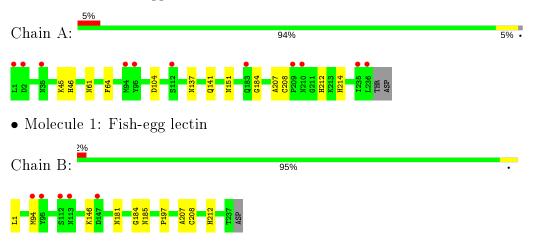
• Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	А	176	Total O 176 176	0	0
6	В	208	Total O 208 208	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: Fish-egg lectin

• Molecule 2: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain C:	50%	50%	
NAG2 NAG2			
• Molecule 2 opyranose	: 2-acetamido-2-deoxy-beta-E	m 0-glucopyranose-(1-4)-2-acetamid	o-2-deoxy-beta-D-gluc

Chain D:

100%



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	44.31Å 70.99Å 163.03Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 - 1.35	Depositor
Resolution (A)	38.93 - 1.35	EDS
% Data completeness	99.5 (30.00-1.35)	Depositor
(in resolution range)	99.5(38.93 - 1.35)	EDS
R _{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$3.81 (at 1.35 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
D D.	0.172 , 0.185	Depositor
R, R_{free}	0.173 , 0.186	DCC
R_{free} test set	5667 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	8.5	Xtriage
Anisotropy	0.367	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.37, 40.9	EDS
L-test for twinning ²	$ \langle L \rangle = 0.47, \langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4029	wwPDB-VP
Average B, all atoms $(Å^2)$	10.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.24% of the height of the origin peak. No significant pseudotranslation is detected.

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.



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: GOL, CA, IPA, 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	А	0.30	0/1806	0.53	0/2452	
1	В	0.31	0/1813	0.56	0/2462	
All	All	0.30	0/3619	0.55	0/4914	

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	1771	0	1727	10	0
1	В	1778	0	1734	7	0
2	С	28	0	25	0	0
2	D	28	0	25	0	0
3	А	1	0	0	0	0
3	В	1	0	0	0	0
4	А	8	0	16	0	0
4	В	12	0	24	1	0
5	А	6	0	8	1	0
5	В	12	0	16	0	0
6	A	176	0	0	2	0

Continued on next page...



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	В	208	0	0	0	0
All	All	4029	0	3575	17	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 (17) 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:207:ALA:H	1:B:212:HIS:HE1	1.23	0.87
1:A:207:ALA:H	1:A:212:HIS:HE1	1.21	0.85
1:A:137:ASN:HD21	1:A:141:GLN:HE21	1.34	0.74
1:B:181:ASN:HD21	1:B:185:ASN:HD22	1.42	0.68
1:A:207:ALA:H	1:A:212:HIS:CE1	2.11	0.64
1:A:104:ASP:H	1:A:151:ASN:HD21	1.49	0.60
1:B:207:ALA:H	1:B:212:HIS:CE1	2.13	0.60
1:B:208:CYS:H	1:B:212:HIS:CE1	2.27	0.51
1:A:208:CYS:H	1:A:212:HIS:CE1	2.29	0.51
1:A:214:HIS:HE1	6:A:407:HOH:O	1.95	0.49
1:A:45:LYS:NZ	1:A:61:ASN:HD21	2.12	0.47
1:B:184:GLY:O	1:B:212:HIS:HD2	1.99	0.46
1:B:1:LEU:HG	4:B:306:IPA:H13	1.99	0.44
1:A:64:PHE:CZ	5:A:306:GOL:H32	2.55	0.42
1:A:184:GLY:O	1:A:212:HIS:HD2	2.03	0.41
1:A:46:HIS:HD2	6:A:568:HOH:O	2.02	0.41
1:B:146:LYS:HD3	1:B:197:PRO:HG2	2.04	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	Perce	ntiles
1	А	234/238~(98%)	231~(99%)	3 (1%)	0	100	100
1	В	235/238~(99%)	233~(99%)	2(1%)	0	100	100
All	All	469/476~(98%)	464 (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	Analysed Rotameric Outliers		Percentiles		
1	А	199/201~(99%)	199~(100%)	0	100 100		
1	В	200/201~(100%)	199~(100%)	1 (0%)	88 74		
All	All	399/402~(99%)	398~(100%)	1 (0%)	92 83		

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	В	94	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (13) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	46	HIS
1	А	61	ASN
1	А	67	GLN
1	А	141	GLN
1	А	151	ASN
1	А	185	ASN
1	А	212	HIS
1	А	214	HIS
1	В	118	GLN
1	В	139	ASN
1	В	185	ASN
1	В	188	GLN

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type
1	В	212	HIS

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 Lin		Link	Bond lengths			Bond angles			
	Type	Cham	nes		Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	NAG	С	1	1,2	14, 14, 15	0.54	0	$17,\!19,\!21$	0.74	0
2	NAG	С	2	2	14,14,15	0.50	0	$17,\!19,\!21$	0.98	1(5%)
2	NAG	D	1	1,2	14,14,15	0.54	0	17,19,21	0.74	0
2	NAG	D	2	2	14,14,15	0.48	0	17,19,21	0.74	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	С	1	1,2	-	0/6/23/26	0/1/1/1
2	NAG	С	2	2	-	0/6/23/26	0/1/1/1
2	NAG	D	1	1,2	-	0/6/23/26	0/1/1/1
2	NAG	D	2	2	-	0/6/23/26	0/1/1/1

There are no bond length outliers.



All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
2	С	2	NAG	C1-O5-C5	2.03	114.95	112.19

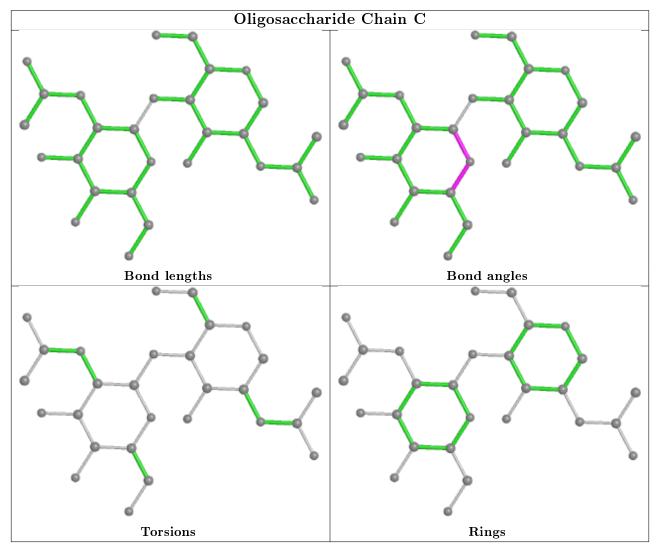
There are no chirality outliers.

There are no torsion outliers.

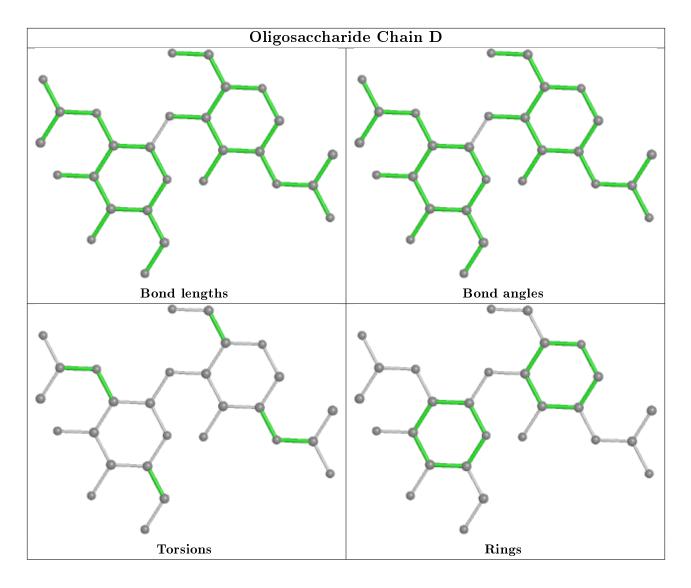
There are no ring outliers.

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 10 ligands modelled in this entry, 2 are monoatomic - leaving 8 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).

Mal	Mol Type		Res	5 Link	Bond lengths			Bond angles		
	туре	Chain	nes		Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z >2
4	IPA	В	304	-	3,3,3	0.52	0	3,3,3	0.21	0
4	IPA	А	305	-	3,3,3	0.53	0	3,3,3	0.19	0
5	GOL	В	307	-	5, 5, 5	0.23	0	$5,\!5,\!5$	0.36	0
4	IPA	В	305	-	3,3,3	0.52	0	$3,\!3,\!3$	0.14	0



Mal	Mol Type	e Chain	Res	Link	Bond lengths			Bond angles		
	туре	Cham	nes		Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
4	IPA	А	304	-	3,3,3	0.50	0	3,3,3	0.21	0
4	IPA	В	306	-	3,3,3	0.48	0	3,3,3	0.25	0
5	GOL	А	306	-	5, 5, 5	0.24	0	$5,\!5,\!5$	0.35	0
5	GOL	В	308	-	5, 5, 5	0.40	0	$5,\!5,\!5$	0.59	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
5	GOL	В	308	-	-	4/4/4/4	-
5	GOL	В	307	-	-	0/4/4/4	-
5	GOL	А	306	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	В	308	GOL	O1-C1-C2-C3
5	В	308	GOL	C1-C2-C3-O3
5	В	308	GOL	O1-C1-C2-O2
5	В	308	GOL	O2-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	В	306	IPA	1	0
5	А	306	GOL	1	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, 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	$OWAB(Å^2)$	$\mathbf{Q}{<}0.9$
1	А	236/238~(99%)	0.18	11 (4%) 31 36	6, 10, 17, 23	0
1	В	237/238~(99%)	0.11	5 (2%) 63 69	5, 8, 13, 21	0
All	All	473/476~(99%)	0.15	16 (3%) 45 51	5, 9, 15, 23	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	112	SER	5.4
1	А	236	LEU	5.4
1	А	1	LEU	4.9
1	В	112	SER	4.5
1	А	2	ASP	4.4
1	В	113	ASN	4.1
1	А	95	TYR	4.1
1	В	95	TYR	4.0
1	А	210	ASN	3.2
1	В	147	ASP	3.0
1	А	235	ILE	2.4
1	А	94	MET	2.3
1	В	94	MET	2.3
1	А	209	PRO	2.2
1	А	35	ASN	2.1
1	А	183	GLN	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.

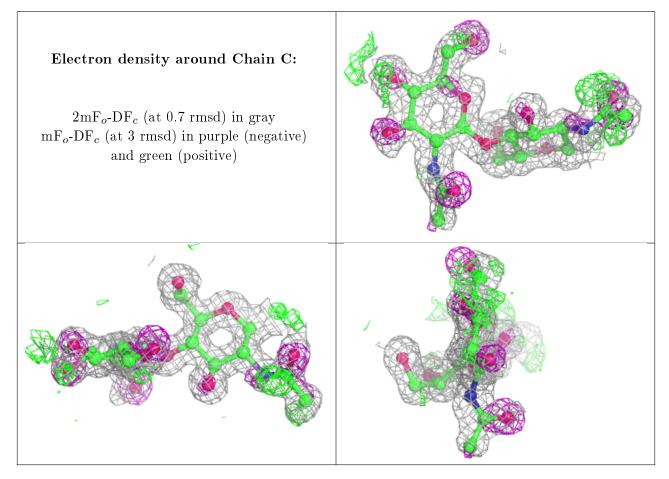


6.3 Carbohydrates (i)

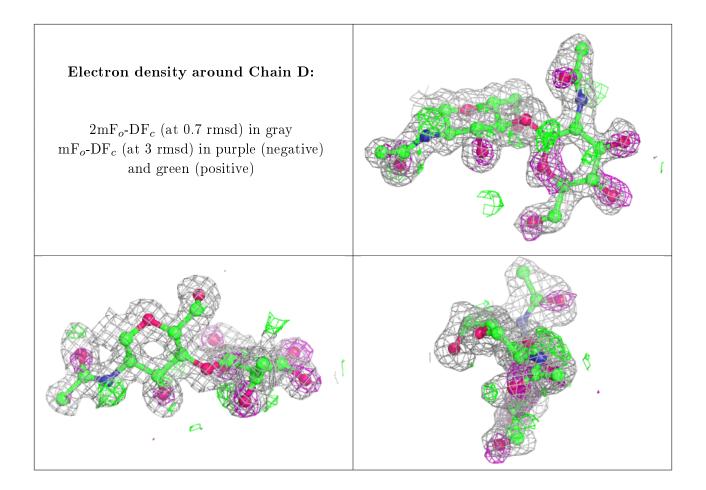
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-factors}(\mathbf{\AA}^2)$	Q<0.9
2	NAG	С	2	14/15	0.75	0.21	$11,\!13,\!15,\!16$	0
2	NAG	D	2	14/15	0.77	0.33	$14,\!17,\!18,\!18$	0
2	NAG	С	1	14/15	0.89	0.10	$7,\!8,\!10,\!10$	0
2	NAG	D	1	14/15	0.92	0.10	$7,\!9,\!10,\!12$	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	$\mathbf{B} ext{-factors}(\mathbf{A}^2)$	$Q{<}0.9$
5	GOL	В	308	6/6	0.83	0.14	$13,\!13,\!14,\!15$	0
4	IPA	В	306	4/4	0.89	0.19	$16,\!16,\!16,\!17$	0
4	IPA	А	305	4/4	0.92	0.13	$18,\!19,\!19,\!19$	0
4	IPA	А	304	4/4	0.95	0.08	9,9,9,9	0
5	GOL	В	307	6/6	0.95	0.08	$10,\!10,\!11,\!11$	0
4	IPA	В	305	4/4	0.95	0.11	$12,\!13,\!13,\!13$	0
5	GOL	А	306	6/6	0.96	0.12	$12,\!12,\!13,\!13$	0
4	IPA	В	304	4/4	0.98	0.05	5, 5, 5, 5	0
3	CA	А	303	1/1	1.00	0.05	5, 5, 5, 5	0
3	CA	В	303	1/1	1.00	0.04	3, 3, 3, 3	0



6.5 Other polymers (i)

There are no such residues in this entry.

