

wwPDB X-ray Structure Validation Summary Report (i)

Aug 8, 2020 – 11:54 PM BST

PDB ID : 4BGY

Title : H5 (VN1194) Influenza Virus Haemagglutinin in Complex with Avian Receptor

Analogue 3'-SLN

Authors: Xiong, X.; Coombs, P.; Martin, S.R.; Liu, J.; Xiao, H.; McCauley, J.W.;

Locher, K.; Walker, P.A.; Collins, P.J.; Kawaoka, Y.; Skehel, J.J.; Gamblin,

S.J.

Deposited on : 2013-03-29

Resolution : 2.68 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)

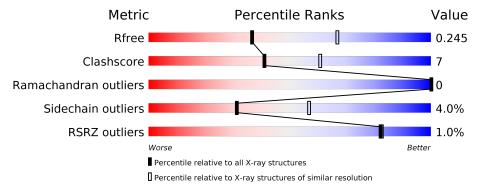
Validation Pipeline (wwPDB-VP) : 2.13.1

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 2.68 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	3863 (2.70-2.66)
Clashscore	141614	4210 (2.70-2.66)
Ramachandran outliers	138981	4141 (2.70-2.66)
Sidechain outliers	138945	4141 (2.70-2.66)
RSRZ outliers	127900	3780 (2.70-2.66)

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	326	82%	15%	
2	В	166	85%	12%	
3	С	2	100%		
3	D	2	100%		
3	F	2	100%		
4	Е	3	67%	33%	



The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mo	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	NAG	F	1	X	-	_	-



2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 4137 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 HEMAGGLUTININ.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	Λ	321	Total	С	N	О	S	0	0	0
1	A	321	2547	1609	439	484	15	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Α	325	THR	ARG	conflict	UNP Q6DQ34

• Molecule 2 is a protein called HEMAGGLUTININ.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
2	В	162	Total 1312	C 819	N 226	O 259	S 8	0	0	0

• Molecule 3 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
3	С	2	Total C 28 10	N 3 2	O 10	0	0	0
3	D	2	Total C	N 3 2		0	0	0
3	F	2	Total C	N 3 2	O 10	0	0	0

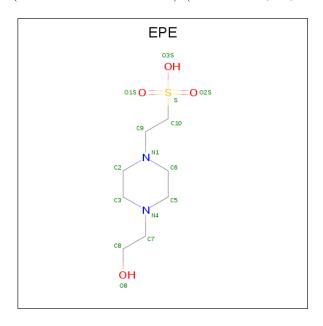
• Molecule 4 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galacto pyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.





Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace
4	Е	3	Total C 46 25	N O 2 19	0	0	0

• Molecule 5 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	D	1	Total	С	N	О	S	0	0
9	D	1	15	8	2	4	1	0	0

• Molecule 6 is water.

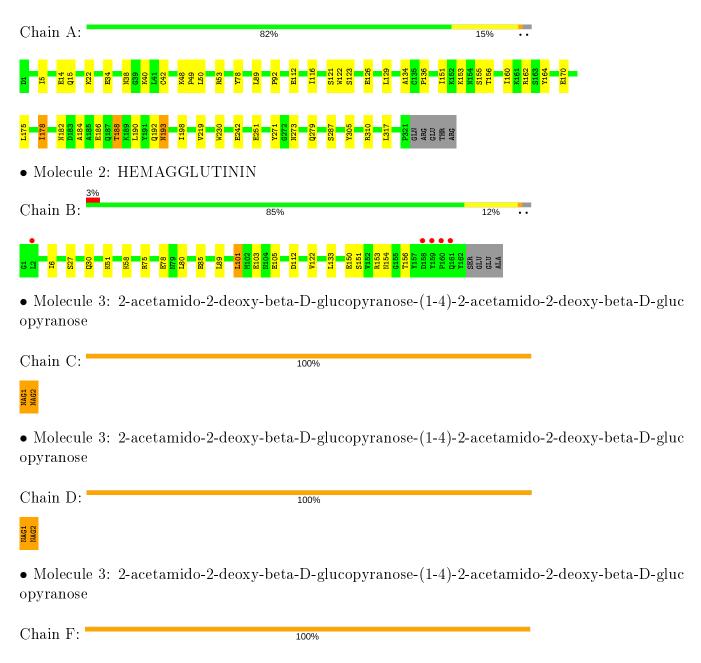
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	73	Total O 73 73	0	0
6	В	60	Total O 60 60	0	0



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.









 \bullet Molecule 4: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain E: 67% 33%





4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants	101.40Å 101.40Å 450.00Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.94 - 2.68	Depositor
resolution (A)	40.90 - 2.68	EDS
% Data completeness	99.9 (40.94-2.68)	Depositor
(in resolution range)	$100.0 \ (40.90 - 2.68)$	EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$2.90 \; ({\rm at} \; 2.69 {\rm \AA})$	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
P. P.	0.202 , 0.248	Depositor
R, R_{free}	0.204 , 0.245	DCC
R_{free} test set	1307 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å ²)	52.6	Xtriage
Anisotropy	0.712	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.30 , 44.5	EDS
L-test for twinning ²	$ < L >=0.49, < L^2>=0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o , F_c correlation	0.94	EDS
Total number of atoms	4137	wwPDB-VP
Average B, all atoms $(Å^2)$	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.00% 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: EPE, SIA, GAL, 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		
IVIOI		RMSZ	# Z >5	RMSZ	# Z > 5	
1	A	0.35	0/2609	0.56	0/3545	
2	В	0.40	0/1339	0.57	0/1802	
All	All	0.37	0/3948	0.56	0/5347	

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	$\mathbf{H}(\mathbf{model})$	$\mathbf{H}(\mathbf{added})$	Clashes	Symm-Clashes
1	A	2547	0	2489	39	1
2	В	1312	0	1218	13	0
3	С	28	0	25	2	0
3	D	28	0	25	1	0
3	F	28	0	25	1	0
4	E	46	0	40	3	0
5	В	15	0	18	1	0
6	A	73	0	0	1	0
6	В	60	0	0	2	0
All	All	4137	0	3840	56	1

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 7.

The worst 5 of 56 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{aligned} ext{Interatomic} \ ext{distance} & (ext{Å}) \end{aligned}$	$egin{aligned} ext{Clash} \ ext{overlap } (ext{Å}) \end{aligned}$
2:B:150:GLU:OE1	2:B:153:ARG:NH1	2.08	0.86
1:A:126:GLU:HG2	1:A:129:LEU:CD1	2.11	0.81
3:C:1:NAG:O4	3:C:2:NAG:H83	1.87	0.75
1:A:129:LEU:HD12	1:A:151:ILE:HD11	1.70	0.71
1:A:126:GLU:HG2	1:A:129:LEU:HD12	1.71	0.70

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	$egin{array}{l} ext{Interatomic} \ ext{distance} \ (ext{Å}) \end{array}$	$egin{aligned} ext{Clash} \ ext{overlap } (ext{Å}) \end{aligned}$
1:A:170:GLU:OE2	1:A:170:GLU:OE2[11_565]	2.07	0.13

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	$_{ m ntiles}$
1	A	319/326 (98%)	304 (95%)	15 (5%)	0	100	100
2	В	160/166~(96%)	153 (96%)	7 (4%)	0	100	100
All	All	479/492 (97%)	457 (95%)	22 (5%)	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	289/294~(98%)	278 (96%)	11 (4%)	33 59
2	В	138/141 (98%)	132 (96%)	6 (4%)	29 54
All	All	427/435 (98%)	410 (96%)	17 (4%)	31 57

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	192	GLN
1	A	193	ASN
2	В	58	LYS
1	A	188	THR
2	В	85	GLU

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

Mol	Chain	Res	Type
1	A	15	GLN
1	A	193	ASN
2	В	79	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)

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

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	ain Res Link		Вс	ond leng	ths	В	ond ang	les	
WIOI	Type	Chain	res	1169	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	NAG	С	1	1,3	14,14,15	0.48	0	17,19,21	1.19	1 (5%)	
3	NAG	С	2	3	14,14,15	0.63	0	17,19,21	1.55	4 (23%)	
3	NAG	D	1	1,3	14,14,15	0.51	0	17,19,21	1.60	3 (17%)	
3	NAG	D	2	3	14,14,15	0.67	0	17,19,21	1.41	2 (11%)	
4	NAG	Е	1	4	15,15,15	0.41	0	21,21,21	1.42	1 (4%)	
4	GAL	Е	2	4	11,11,12	0.28	0	15,15,17	0.89	1 (6%)	
4	SIA	Е	3	4	17,20,21	0.55	0	21,28,31	1.20	4 (19%)	
3	NAG	F	1	3,2	14,14,15	0.66	0	17,19,21	1.26	2 (11%)	
3	NAG	F	2	3	14,14,15	0.55	0	17,19,21	1.24	1 (5%)	

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	${f Torsions}$	Rings
3	NAG	С	1	1,3	-	2/6/23/26	0/1/1/1
3	NAG	С	2	3	-	4/6/23/26	0/1/1/1
3	NAG	D	1	1,3	-	2/6/23/26	0/1/1/1
3	NAG	D	2	3	-	3/6/23/26	0/1/1/1
4	NAG	Ε	1	4	-	0/6/26/26	0/1/1/1
4	GAL	Ε	2	4	-	2/2/19/22	0/1/1/1
4	SIA	E	3	4	-	1/14/34/38	0/1/1/1
3	NAG	F	1	3,2	1/1/5/7	2/6/23/26	0/1/1/1
3	NAG	F	2	3	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\mathbf{Observed}(^o)$	$Ideal(^{o})$
4	E	1	NAG	C1-C2-N2	-5.30	104.58	110.73
3	D	1	NAG	C1-O5-C5	4.63	118.46	112.19
3	С	2	NAG	O5-C1-C2	-3.79	105.30	111.29
3	С	1	NAG	O5-C5-C6	3.71	113.02	107.20
3	F	2	NAG	C1-C2-N2	-3.55	104.42	110.49



All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	F	1	NAG	C1

5 of 16 torsion outliers are listed below:

Mol	Chain	${f Res}$	Type	${f Atoms}$
3	D	1	NAG	O5-C5-C6-O6
3	D	2	NAG	C4-C5-C6-O6
3	F	1	NAG	O5-C5-C6-O6
4	E	2	GAL	O5-C5-C6-O6
3	D	2	NAG	O5-C5-C6-O6

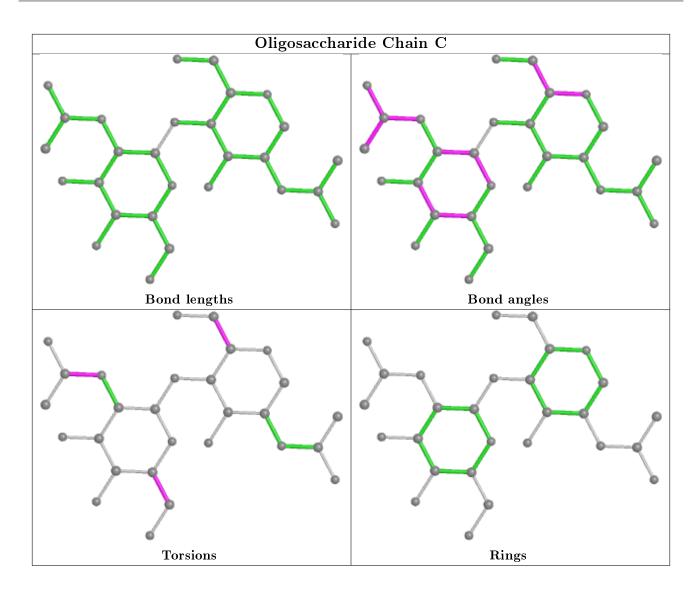
There are no ring outliers.

7 monomers are involved in 7 short contacts:

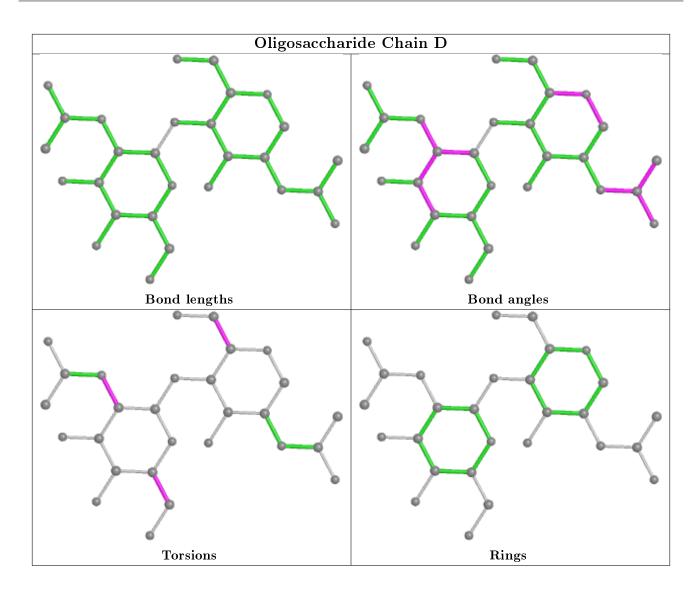
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	2	NAG	1	0
3	F	1	NAG	1	0
3	D	1	NAG	1	0
3	F	2	NAG	1	0
3	С	1	NAG	2	0
4	E	3	SIA	3	0
3	С	2	NAG	1	0

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

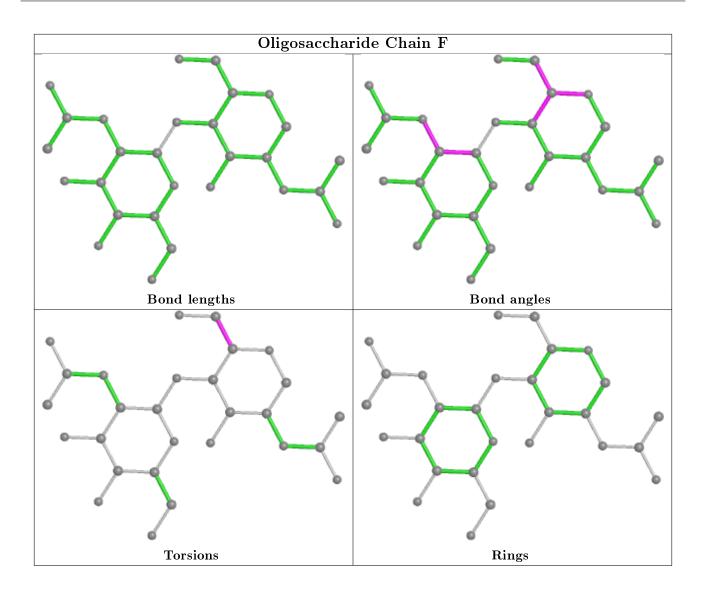




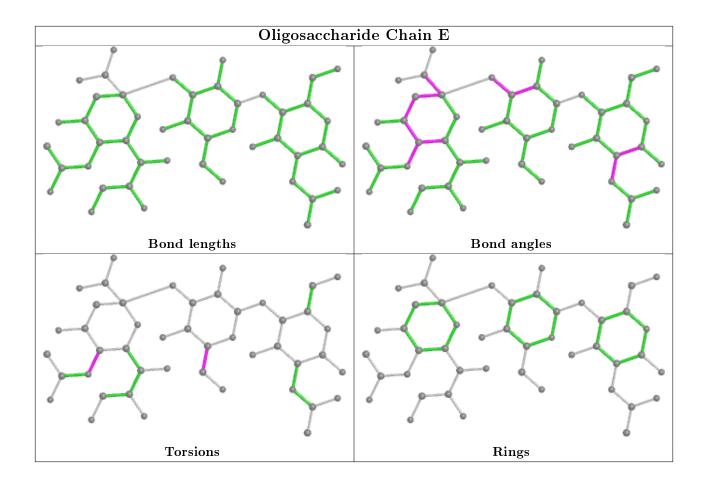












5.6 Ligand geometry (i)

1 ligand 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	Bo	ond leng	$ ag{ths}$	Bond angles		
MIOI	туре	Chain	nes	Lilik	Counts	RMSZ	# Z > 2	Counts	$\mid RMSZ \mid \# Z > 2$	# Z > 2
5	EPE	В	1165	-	15,15,15	2.03	1 (6%)	18,20,20	2.86	9 (50%)

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	EPE	В	1165	-	-	3/9/19/19	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	${f Observed(\AA)}$	$\operatorname{Ideal}(ext{\AA})$
5	В	1165	EPE	C10-S	-7.40	1.67	1.77

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$Observed(^o)$	$Ideal(^{o})$
5	В	1165	EPE	C6-N1-C2	5.10	120.30	108.83
5	В	1165	EPE	C5-N4-C3	4.54	119.06	108.83
5	В	1165	EPE	O2S-S-C10	4.33	112.13	106.92
5	В	1165	EPE	C9-N1-C6	3.66	120.60	111.23
5	В	1165	EPE	C7-N4-C5	3.62	120.48	111.23

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	В	1165	EPE	C10-C9-N1-C6
5	В	1165	EPE	N4-C7-C8-O8
5	В	1165	EPE	C8-C7-N4-C5

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	В	1165	EPE	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, 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 $>$	$\#\mathrm{RSRZ}{>}2$	$\mathbf{OWAB}(\mathrm{\AA}^2)$	Q < 0.9
1	A	321/326 (98%)	-0.26	0 100 100	44, 83, 107, 133	0
2	В	162/166~(97%)	0.02	5 (3%) 49 48	36, 60, 104, 134	0
All	All	483/492 (98%)	-0.16	5 (1%) 82 82	36, 77, 107, 134	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	В	160	PRO	3.6
2	В	159	TYR	2.4
2	В	161	GLN	2.4
2	В	2	LEU	2.3
2	В	158	ASP	2.2

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
3	NAG	С	2	14/15	0.83	0.40	100,121,127,127	0
3	NAG	D	2	14/15	0.87	0.32	112,118,123,126	0
3	NAG	F	2	14/15	0.88	0.35	108,122,130,132	0
3	NAG	F	1	14/15	0.89	0.26	83,94,109,119	0

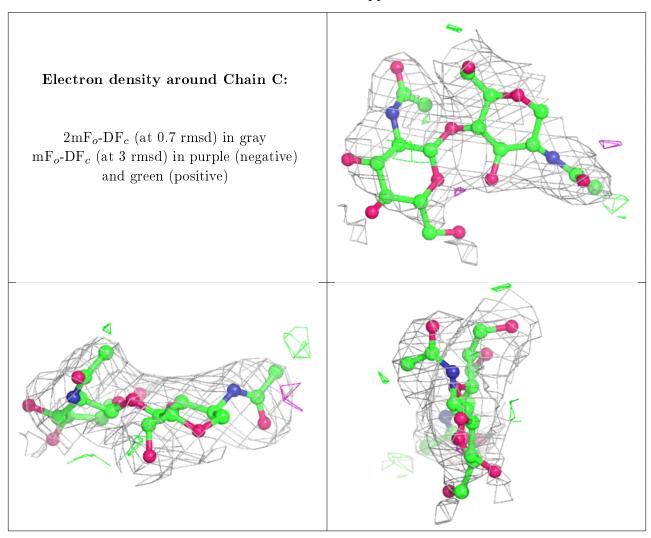
Continued on next page...



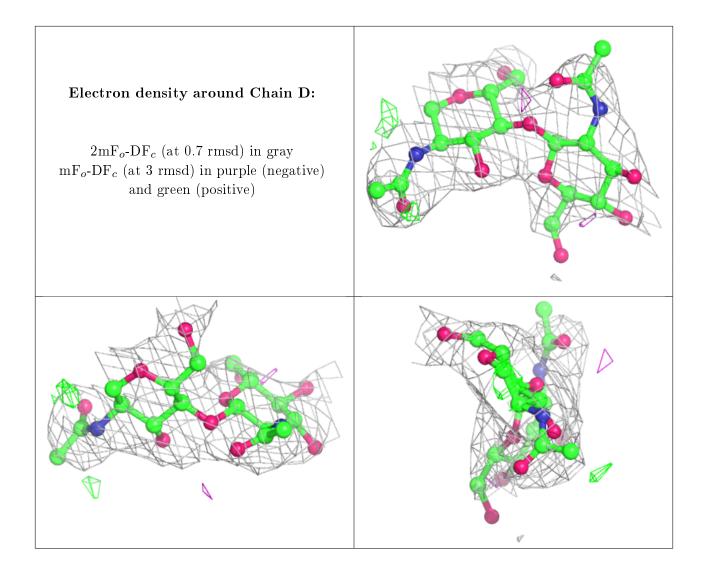
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	${f B\text{-factors}}({f \AA}^2)$	Q < 0.9
4	GAL	E	2	11/12	0.90	0.19	89,97,106,108	0
3	NAG	D	1	14/15	0.90	0.19	76,85,94,101	0
4	NAG	E	1	15/15	0.91	0.28	112,118,121,121	0
3	NAG	С	1	14/15	0.91	0.21	75,85,93,109	0
4	SIA	Е	3	20/21	0.95	0.16	70,80,87,89	0

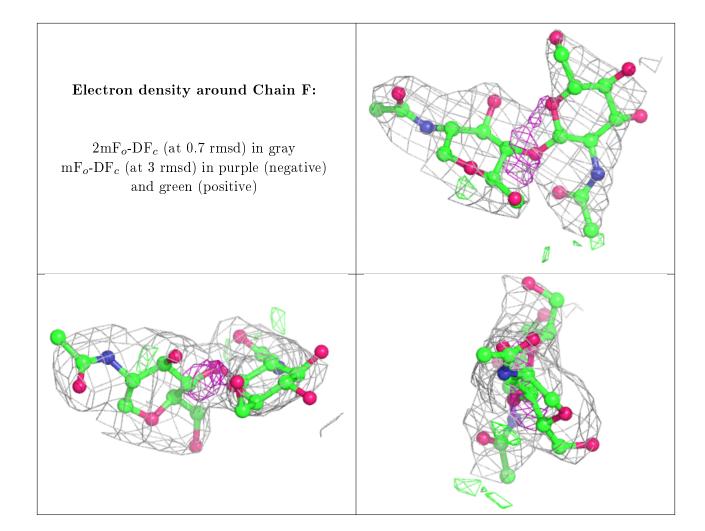
The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.



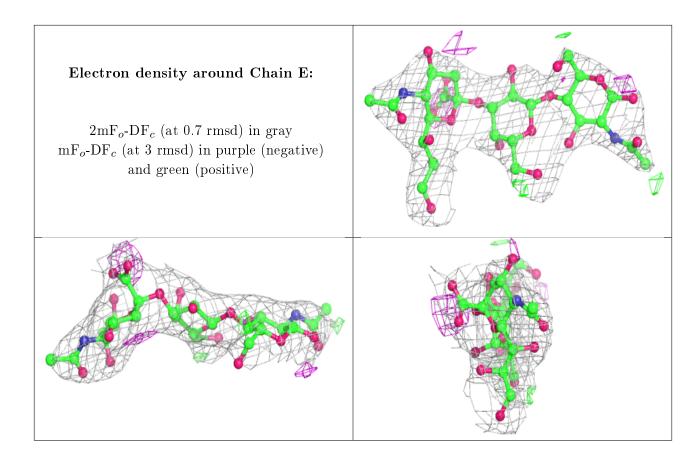












6.4 Ligands (i)

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	${f B-factors}({f \AA}^2)$	Q < 0.9
5	EPE	В	1165	15/15	0.88	0.31	79,82,114,117	0

6.5 Other polymers (i)

There are no such residues in this entry.

