W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 12, 2024 – 06:46 PM EST

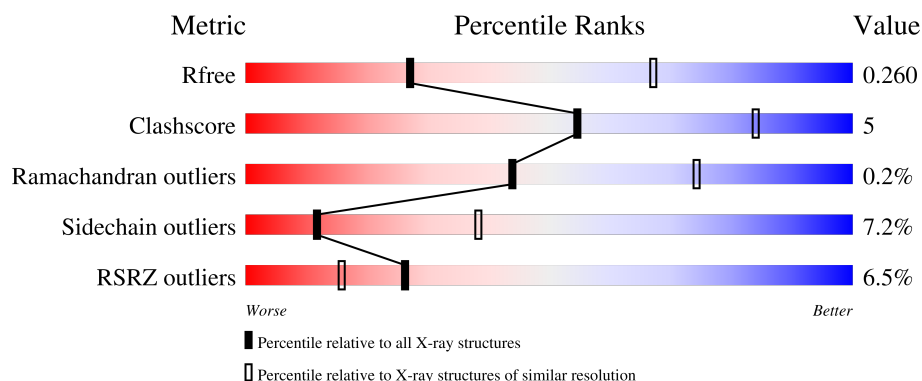|		
PDB ID	:	3S12
Title	:	Crystal structure of H5N1 influenza virus hemagglutinin, strain YU562 crystal form 1
Authors	:	DuBois, R.M.; Zaraket, H.; Reddivari, M.; Heath, R.J.; White, S.W.; Russell, C.J.
Deposited on	:	2011-05-14
Resolution	:	3.10 Å(reported)

---

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
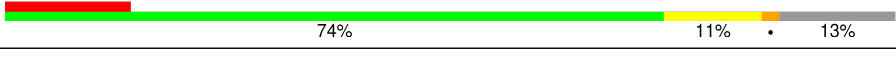*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	Percentile Ranks	Value
Rfree		0.260
Clashscore		5
Ramachandran outliers		0.2%
Sidechain outliers		7.2%
RSRZ outliers		6.5%

Worse — *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1351 (3.10-3.10)
Clashscore	180529	1454 (3.10-3.10)
Ramachandran outliers	177936	1391 (3.10-3.10)
Sidechain outliers	177891	1391 (3.10-3.10)
RSRZ outliers	164620	1351 (3.10-3.10)

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	336	2% ... 77% ... 16% ... • 5%
2	B	182	14% ... 74% ... 11% • 13%
3	C	2	50% ... 50%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 3860 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 HA1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	318	Total 2506	C 1581	N 433	O 477	S 15	0	0	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	ASP	-	expression tag	UNP Q80A30
A	-1	LEU	-	expression tag	UNP Q80A30
A	0	GLY	-	expression tag	UNP Q80A30
A	1	SER	-	expression tag	UNP Q80A30
A	2	MET	-	expression tag	UNP Q80A30
A	3	ALA	-	expression tag	UNP Q80A30
A	4	ASP	-	expression tag	UNP Q80A30
A	5	PRO	-	expression tag	UNP Q80A30
A	6	GLY	-	expression tag	UNP Q80A30
A	7	TYR	-	expression tag	UNP Q80A30
A	8	LEU	-	expression tag	UNP Q80A30
A	9	LEU	-	expression tag	UNP Q80A30
A	10	GLU	-	expression tag	UNP Q80A30

- Molecule 2 is a protein called Hemagglutinin HA2 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	158	Total 1286	C 797	N 222	O 259	S 8	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	177	ARG	-	expression tag	UNP Q80A30
B	178	SER	-	expression tag	UNP Q80A30
B	179	LEU	-	expression tag	UNP Q80A30

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	180	VAL	-	expression tag	UNP Q80A30
B	181	PRO	-	expression tag	UNP Q80A30
B	182	ARG	-	expression tag	UNP Q80A30

- Molecule 3 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.
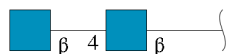
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	2	28	16	2	10	0	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
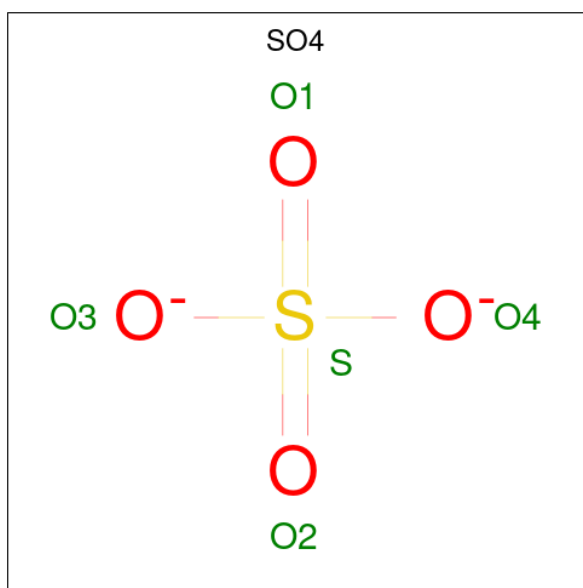
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
4	A	1	5	4	1	0	0
4	A	1	5	4	1	0	0

- Molecule 5 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

NAG

O6
OH

O5
O1
HO O
C6

C1(R) C5(R)

C2(R) C4(S)
C3(R)
N2
HN OH
O4

C7
OH
C8 O O3
O7
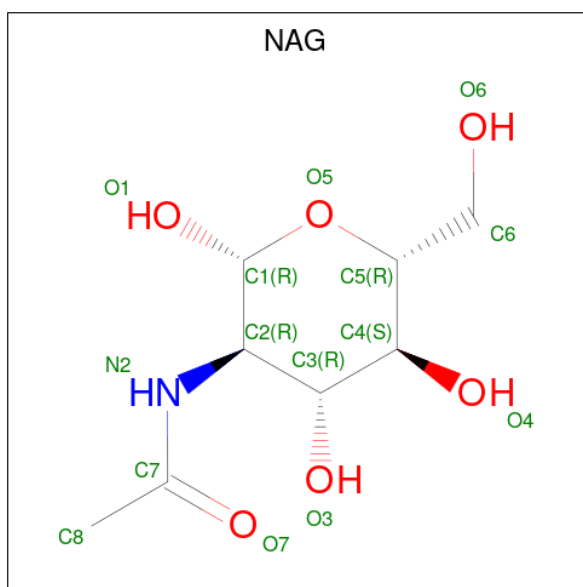
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	12	Total	O	0	0
			12	12		
6	B	4	Total	O	0	0
			4	4		

3 Residue-property plots ⓘ

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 o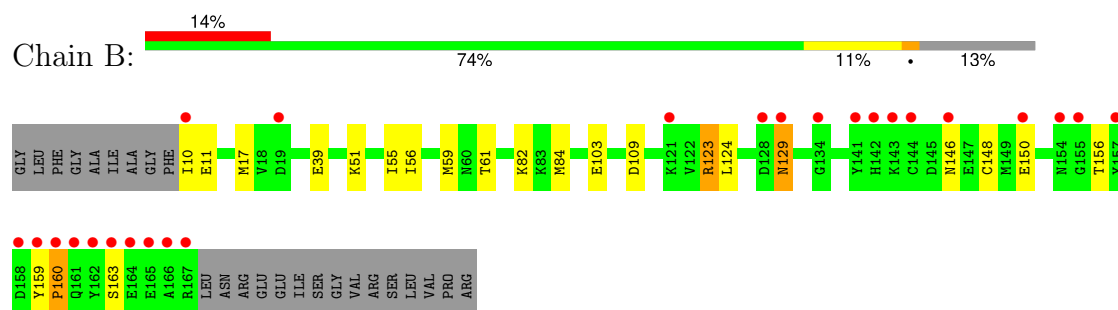r more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Hemagglutinin HA1 chain
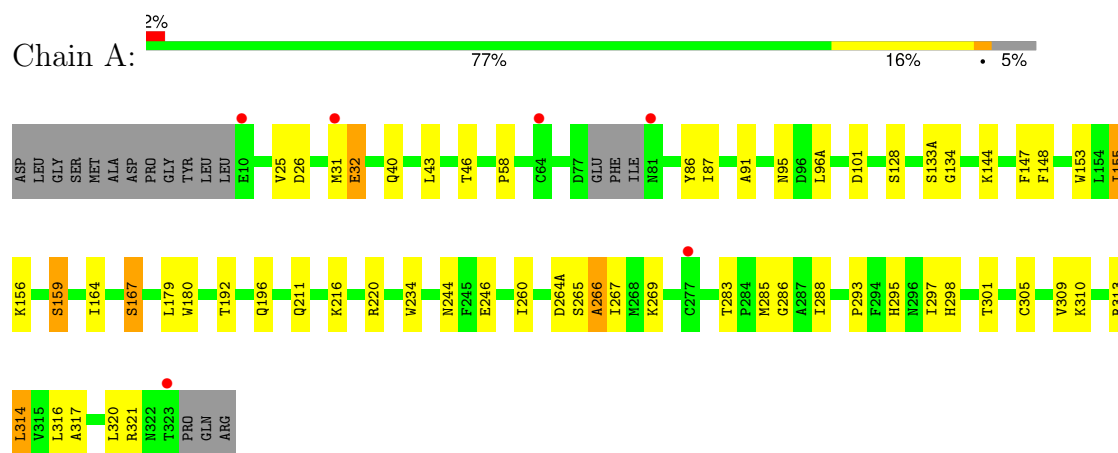
• Molecule 2: Hemagglutinin HA2 chain

• Molecule 3: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 3 2 1	Depositor
Cell constants a, b, c, α, β, γ	112.55Å 112.55Å 134.69Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	134.69 – 3.10 134.69 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.7 (134.69-3.10) 99.7 (134.69-3.10)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.72 (at 3.12Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.217 , 0.264 0.215 , 0.260	Depositor DCC
R_{free} test set	940 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å^2)	40.6	Xtriage
Anisotropy	0.134	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 32.3	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.019 for -h,-k,l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	3860	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.14% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.
[2] Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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: SO4, 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	A	0.39	0/2565	0.57	1/3482 (0.0%)
2	B	0.41	0/1311	0.54	0/1764
All	All	0.39	0/3876	0.56	1/5246 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	266	ALA	N-CA-CB	-5.19	102.84	110.10

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	2506	0	2445	27	0
2	B	1286	0	1186	11	0
3	C	28	0	25	0	0
4	A	10	0	0	0	0
5	A	14	0	13	0	0
6	A	12	0	0	0	0
6	B	4	0	0	0	0
All	All	3860	0	3669	35	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 5.

The worst 5 of 35 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:A:283:THR:HG22	1:A:285:MET:H	1.30	0.92
1:A:283:THR:HB	1:A:286:GLY:O	1.81	0.80
1:A:265:SER:OG	1:A:266:ALA:N	2.22	0.71
1:A:32:GLU:OE1	1:A:321:ARG:NH2	2.24	0.70
2:B:10:ILE:HG22	2:B:11:GLU:H	1.57	0.69

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	314/336 (94%)	292 (93%)	22 (7%)	0	100	100
2	B	156/182 (86%)	146 (94%)	9 (6%)	1 (1%)	22	53
All	All	470/518 (91%)	438 (93%)	31 (7%)	1 (0%)	44	74

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	160	PRO

5.3.2 Protein sidechains ⓘ

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	282/298 (95%)	262 (93%)	20 (7%)	12	39
2	B	137/156 (88%)	127 (93%)	10 (7%)	11	37
All	All	419/454 (92%)	389 (93%)	30 (7%)	12	38

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	264(A)	ASP
2	B	148	CYS
1	A	313	ARG
2	B	163	SER
2	B	109	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	95	ASN
2	B	117	ASN
2	B	129	ASN
2	B	125	GLN
1	A	224	ASN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5   Carbohydrates ⓘ

2 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		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
3	NAG	C	1	1,3	14,14,15	0.56	0	17,19,21	0.84	0
3	NAG	C	2	3	14,14,15	0.47	0	17,19,21	1.23	2 (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
3	NAG	C	1	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	2	3	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

| Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	2	NAG	C3-C4-C5	2.38	114.55	110.23
3	C	2	NAG	C4-C3-C2	2.33	114.43	111.02

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.

Oligosaccharide Chain C

Bond lengths   Bond angles

Torsions   Rings
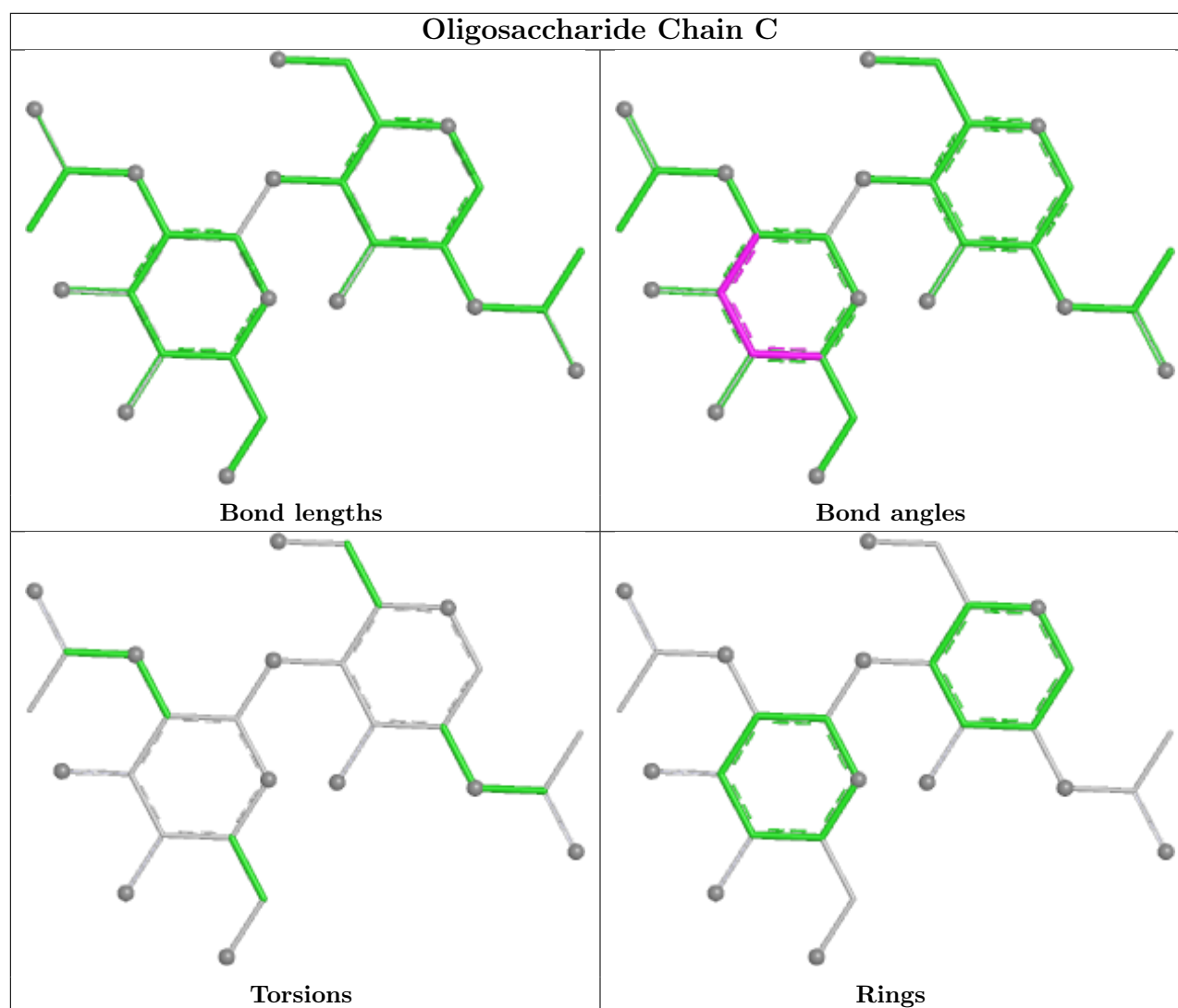
5.6 Ligand geometry ⓘ

3 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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
5	NAG	A	329	1	14,14,15	0.60	0	17,19,21	0.88	1 (5%)
4	SO4	A	327	-	4,4,4	0.22	0	6,6,6	0.33	0
4	SO4	A	328	-	4,4,4	0.24	0	6,6,6	0.16	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	NAG	A	329	1	-	1/6/23/26	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})
5	A	329	NAG	C4-C3-C2	2.05	114.02	111.02

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	329	NAG	C4-C5-C6-O6

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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(Å²)	Q<0.9
1	A	318/336 (94%)	-0.24	6 (1%)	66	47	13, 28, 47, 53	0
2	B	158/182 (86%)	0.78	25 (15%)	6	3	23, 48, 92, 100	0
All	All	476/518 (91%)	0.10	31 (6%)	26	16	13, 34, 72, 100	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	164	GLU	6.4
2	B	167	ARG	5.0
2	B	128	ASP	4.9
2	B	10	ILE	4.7
2	B	163	SER	4.5

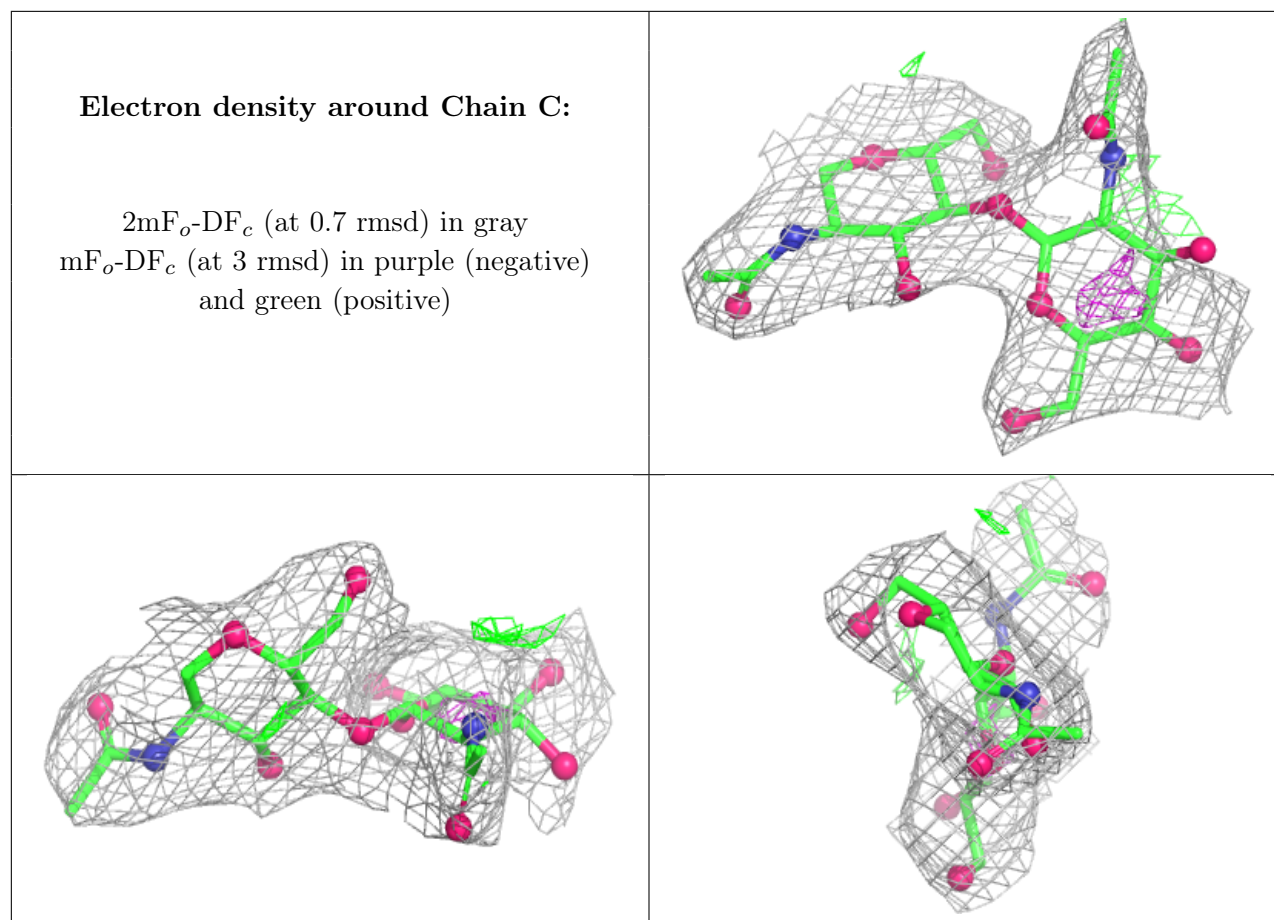## 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

6.3 Carbohydrates ⓘ

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, 95th 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	B-factors(Å²)	Q<0.9
3	NAG	C	2	14/15	0.74	0.17	58,61,62,62	0
3	NAG	C	1	14/15	0.92	0.11	44,46,50,54	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.

Electron density around Chain C:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

6.4 Ligands ⓘ

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, 95th 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	B-factors(Å^2)	Q<0.9
5	NAG	A	329	14/15	0.82	0.15	61,64,65,65	0
4	SO4	A	327	5/5	0.96	0.09	42,42,42,42	0
4	SO4	A	328	5/5	0.99	0.05	37,37,38,38	0

6.5 Other polymers ⓘ

There are no such residues in this entry.