



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 5, 2023 – 02:55 AM EDT

PDB ID : 3UBJ  
Title : Influenza hemagglutinin from the 2009 pandemic in complex with ligand LSTa  
Authors : Xu, R.; Wilson, I.A.  
Deposited on : 2011-10-24  
Resolution : 2.25 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](mailto: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 ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) 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 : 2.35  
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) : 2.35

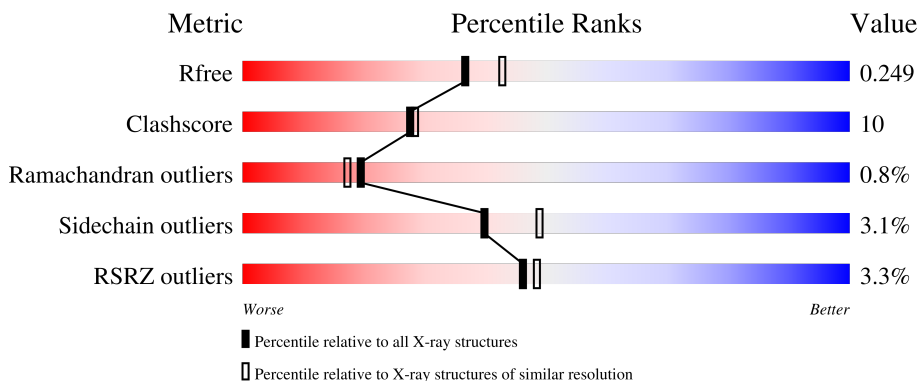
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*


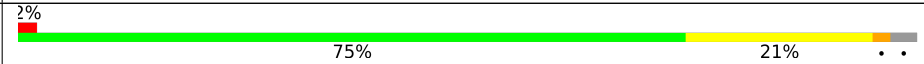
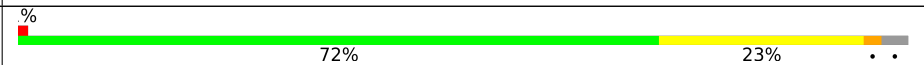

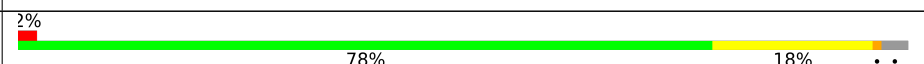
The reported resolution of this entry is 2.25 Å.

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 (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1377 (2.26-2.26)
Clashscore	141614	1487 (2.26-2.26)
Ramachandran outliers	138981	1449 (2.26-2.26)
Sidechain outliers	138945	1450 (2.26-2.26)
RSRZ outliers	127900	1356 (2.26-2.26)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	329	 2% 75% 22% ..
1	C	329	 2% 75% 21% ..
1	E	329	 % 72% 23% ..
1	G	329	 75% 23% ..
1	I	329	 2% 78% 18% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	K	329	 % 71% 25% . .
2	B	177	 12% 71% 23% . . 5%
2	D	177	 3% 68% 27% . .
2	F	177	 2% 75% 19% . .
2	H	177	 % 76% 20% . .
2	J	177	 12% 67% 28% . .
2	L	177	 9% 77% 18% . 5%
3	M	2	 50% 50%
3	O	2	 100%
3	S	2	 50% 50%
4	N	2	 50% 50%
4	P	2	 50% 50%
4	Q	2	 50% 50%
4	R	2	 50% 50%

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 24423 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	323	2522	1595	432	482	13	0	0	0
1	C	320	2501	1582	429	477	13	0	0	0
1	E	319	2492	1578	428	473	13	0	0	0
1	G	323	2522	1595	432	482	13	0	0	0
1	I	320	2501	1582	429	477	13	0	0	0
1	K	320	2504	1585	429	477	13	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	PRO	-	expression tag	UNP C3W5S1
A	10	GLY	-	expression tag	UNP C3W5S1
A	205	CYS	GLY	engineered mutation	UNP C3W5S1
A	220	CYS	ARG	engineered mutation	UNP C3W5S1
C	9	PRO	-	expression tag	UNP C3W5S1
C	10	GLY	-	expression tag	UNP C3W5S1
C	205	CYS	GLY	engineered mutation	UNP C3W5S1
C	220	CYS	ARG	engineered mutation	UNP C3W5S1
E	9	PRO	-	expression tag	UNP C3W5S1
E	10	GLY	-	expression tag	UNP C3W5S1
E	205	CYS	GLY	engineered mutation	UNP C3W5S1
E	220	CYS	ARG	engineered mutation	UNP C3W5S1
G	9	PRO	-	expression tag	UNP C3W5S1
G	10	GLY	-	expression tag	UNP C3W5S1
G	205	CYS	GLY	engineered mutation	UNP C3W5S1
G	220	CYS	ARG	engineered mutation	UNP C3W5S1
I	9	PRO	-	expression tag	UNP C3W5S1

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
I	10	GLY	-	expression tag	UNP C3W5S1
I	205	CYS	GLY	engineered mutation	UNP C3W5S1
I	220	CYS	ARG	engineered mutation	UNP C3W5S1
K	9	PRO	-	expression tag	UNP C3W5S1
K	10	GLY	-	expression tag	UNP C3W5S1
K	205	CYS	GLY	engineered mutation	UNP C3W5S1
K	220	CYS	ARG	engineered mutation	UNP C3W5S1

- Molecule 2 is a protein called Hemagglutinin HA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	169	Total 1360	C 855	N 229	O 270	S 6	0	0	0
2	D	170	Total 1371	C 861	N 233	O 271	S 6	0	0	0
2	F	170	Total 1371	C 861	N 233	O 271	S 6	0	0	0
2	H	172	Total 1389	C 871	N 235	O 277	S 6	0	0	0
2	J	170	Total 1371	C 861	N 233	O 271	S 6	0	0	0
2	L	169	Total 1360	C 855	N 229	O 270	S 6	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	175	SER	-	expression tag	UNP C3W5S1
B	176	GLY	-	expression tag	UNP C3W5S1
B	177	ARG	-	expression tag	UNP C3W5S1
D	175	SER	-	expression tag	UNP C3W5S1
D	176	GLY	-	expression tag	UNP C3W5S1
D	177	ARG	-	expression tag	UNP C3W5S1
F	175	SER	-	expression tag	UNP C3W5S1
F	176	GLY	-	expression tag	UNP C3W5S1
F	177	ARG	-	expression tag	UNP C3W5S1
H	175	SER	-	expression tag	UNP C3W5S1
H	176	GLY	-	expression tag	UNP C3W5S1
H	177	ARG	-	expression tag	UNP C3W5S1
J	175	SER	-	expression tag	UNP C3W5S1
J	176	GLY	-	expression tag	UNP C3W5S1
J	177	ARG	-	expression tag	UNP C3W5S1

*Continued on next page...*

Continued from previous page...

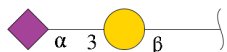
Chain	Residue	Modelled	Actual	Comment	Reference
L	175	SER	-	expression tag	UNP C3W5S1
L	176	GLY	-	expression tag	UNP C3W5S1
L	177	ARG	-	expression tag	UNP C3W5S1

- 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	M	2	28	16	2	10	0	0	0
3	O	2	28	16	2	10	0	0	0
3	S	2	28	16	2	10	0	0	0

- Molecule 4 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
4	N	2	31	17	1	13	0	0	0
4	P	2	31	17	1	13	0	0	0
4	Q	2	31	17	1	13	0	0	0
4	R	2	31	17	1	13	0	0	0

- Molecule 5 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
5	A	1	Total 14	8	1	5	0	0
5	D	1	Total 14	8	1	5	0	0
5	E	1	Total 14	8	1	5	0	0
5	G	1	Total 14	8	1	5	0	0
5	G	1	Total 14	8	1	5	0	0
5	G	1	Total 14	8	1	5	0	0
5	I	1	Total 14	8	1	5	0	0
5	I	1	Total 14	8	1	5	0	0
5	I	1	Total 14	8	1	5	0	0
5	K	1	Total 14	8	1	5	0	0
5	K	1	Total 14	8	1	5	0	0
5	K	1	Total 14	8	1	5	0	0

- Molecule 6 is water.

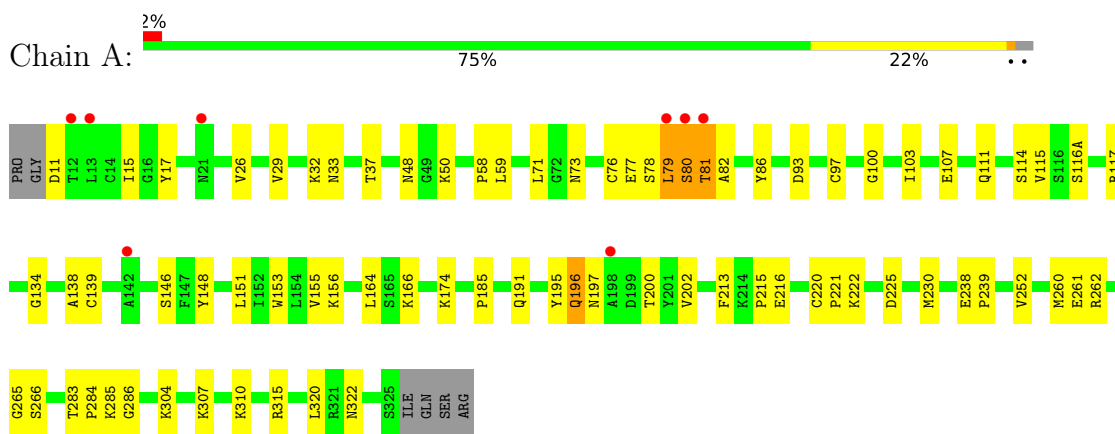
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	98	Total O 98 98	0	0
6	B	47	Total O 47 47	0	0
6	C	71	Total O 71 71	0	0
6	D	52	Total O 52 52	0	0
6	E	91	Total O 91 91	0	0
6	F	40	Total O 40 40	0	0
6	G	84	Total O 84 84	0	0
6	H	54	Total O 54 54	0	0
6	I	104	Total O 104 104	0	0
6	J	43	Total O 43 43	0	0
6	K	65	Total O 65 65	0	0
6	L	34	Total O 34 34	0	0



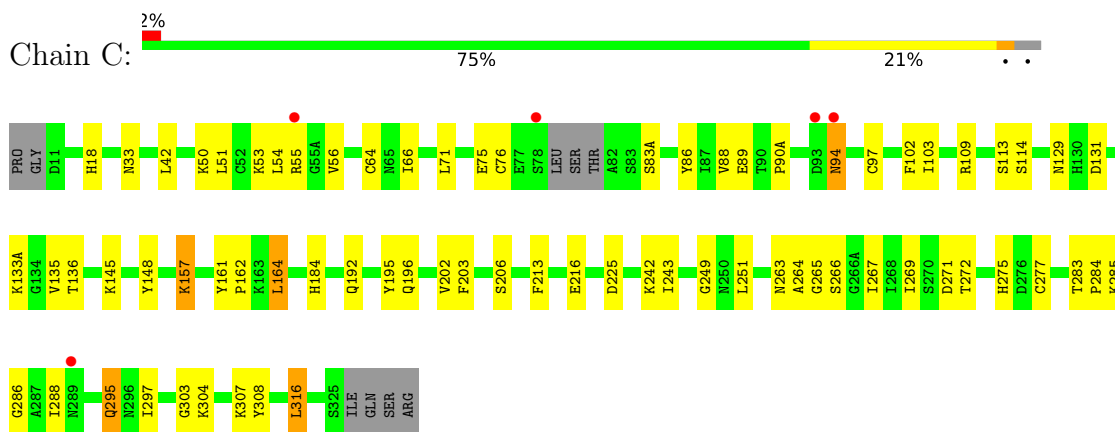
### 3 Residue-property plots [i](#)

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.

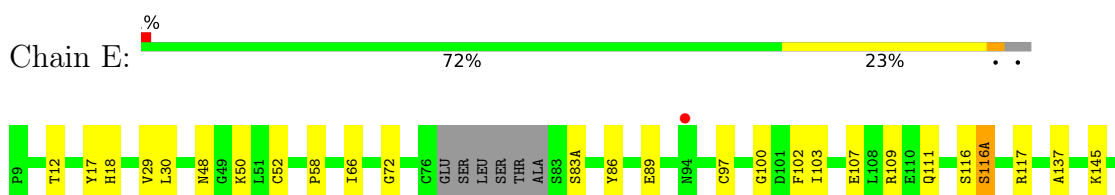
- Molecule 1: Hemagglutinin HA1

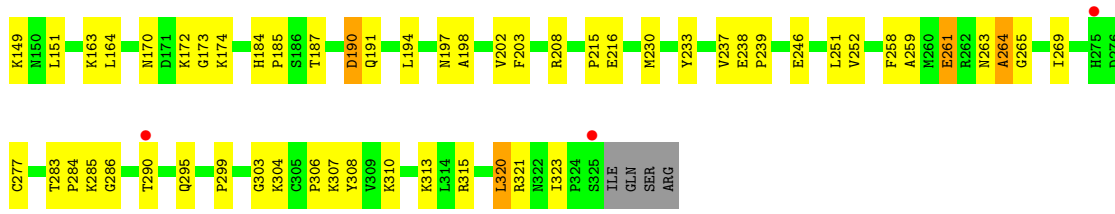


- Molecule 1: Hemagglutinin HA1



- Molecule 1: Hemagglutinin HA1





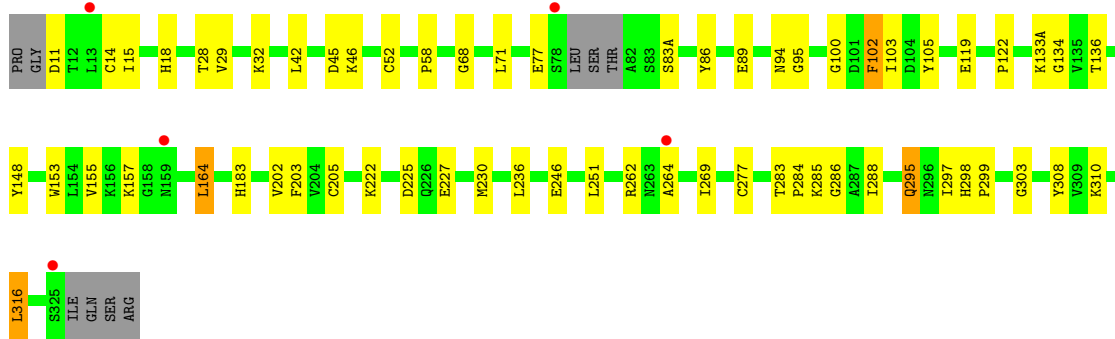
● Molecule 1: Hemagglutinin HA1

Chain G:    ..



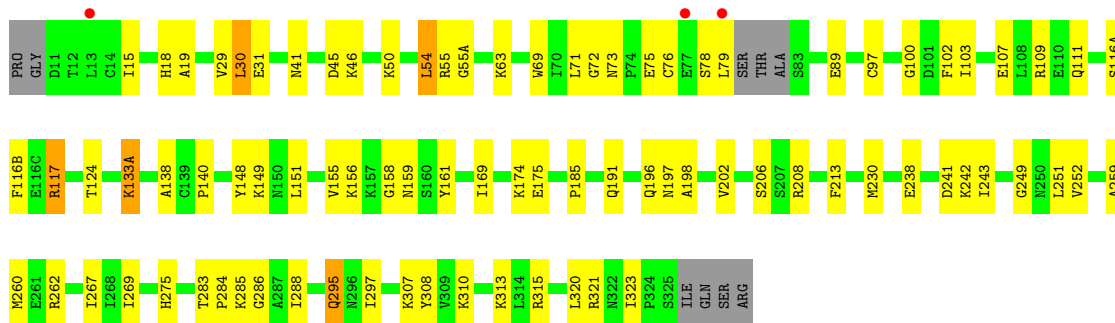
● Molecule 1: Hemagglutinin HA1

Chain I:    ..

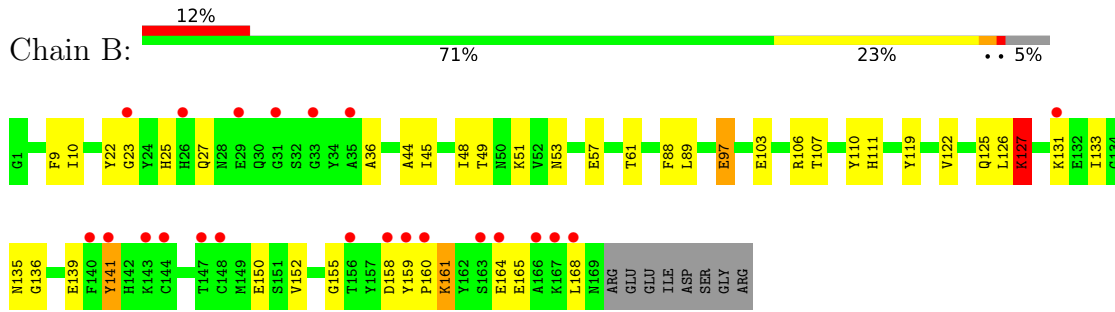


● Molecule 1: Hemagglutinin HA1

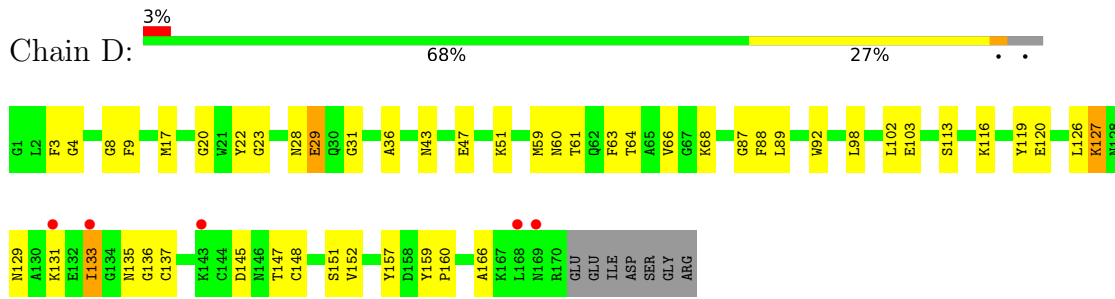
Chain K:    ..



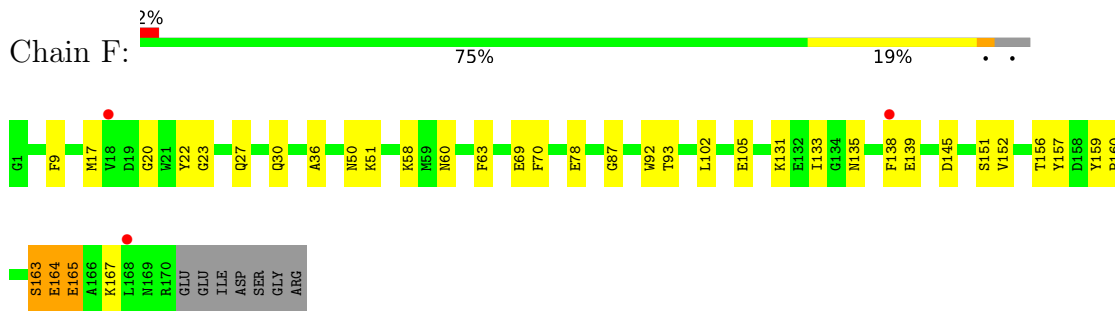
• Molecule 2: Hemagglutinin HA2



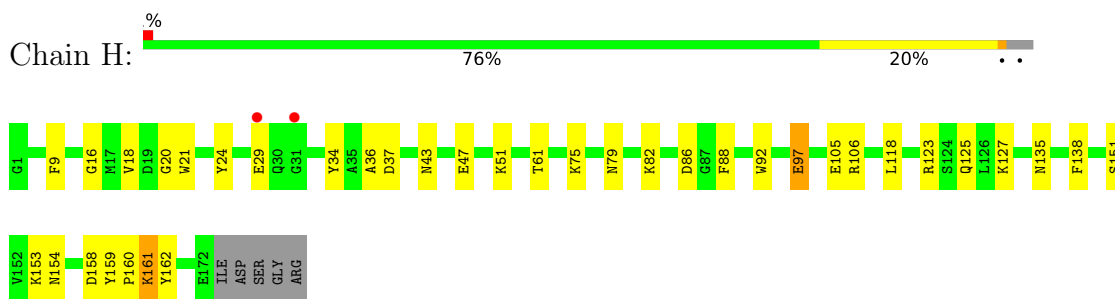
• Molecule 2: Hemagglutinin HA2



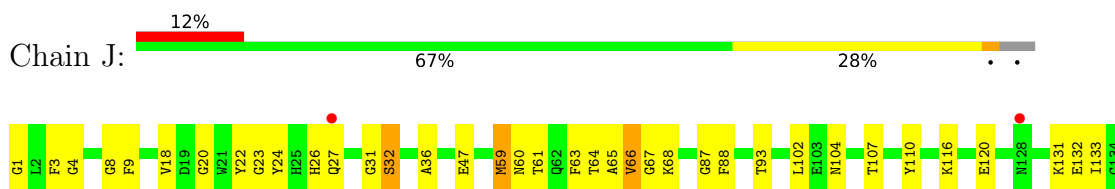
• Molecule 2: Hemagglutinin HA2

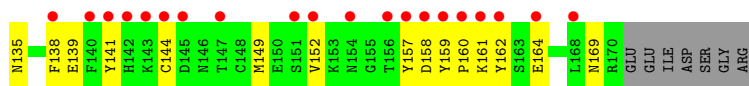


• Molecule 2: Hemagglutinin HA2

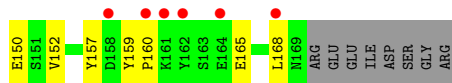
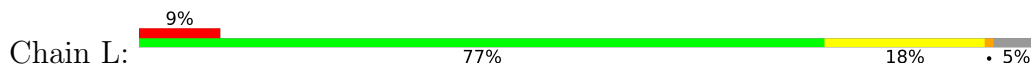


• Molecule 2: Hemagglutinin HA2





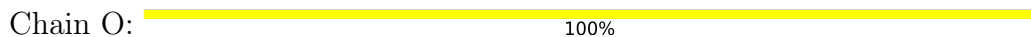
- Molecule 2: Hemagglutinin HA2



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



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



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



- Molecule 4: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose



- Molecule 4: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose



- Molecule 4: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose

Chain Q:  50% 50%

GAL1  
SIA2

- Molecule 4: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose

Chain R:  50% 50%

GAL1  
SIA2

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.85Å 116.09Å 118.40Å 61.05° 77.00° 80.39°	Depositor
Resolution (Å)	49.32 – 2.25 49.31 – 2.24	Depositor EDS
% Data completeness (in resolution range)	91.0 (49.32-2.25) 90.4 (49.31-2.24)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.68 (at 2.25Å)	Xtrriage
Refinement program	PHENIX 1.7_650	Depositor
R, $R_{free}$	0.202 , 0.252 0.198 , 0.249	Depositor DCC
$R_{free}$ test set	6546 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	31.3	Xtrriage
Anisotropy	0.138	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 46.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	24423	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<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.

## 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: SIA, NAG, GAL

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	0.25	0/2586	0.45	0/3516
1	C	0.24	0/2564	0.43	0/3484
1	E	0.25	0/2556	0.46	0/3473
1	G	0.25	0/2586	0.43	0/3516
1	I	0.25	0/2564	0.45	0/3484
1	K	0.24	0/2567	0.43	0/3488
2	B	0.24	0/1388	0.38	0/1871
2	D	0.25	0/1399	0.40	0/1885
2	F	0.25	0/1399	0.39	0/1885
2	H	0.24	0/1417	0.39	0/1909
2	J	0.24	0/1399	0.38	0/1885
2	L	0.24	0/1388	0.37	0/1871
All	All	0.25	0/23813	0.42	0/32267

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	2522	0	2466	57	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2501	0	2444	50	0
1	E	2492	0	2437	64	0
1	G	2522	0	2465	55	0
1	I	2501	0	2441	45	0
1	K	2504	0	2446	62	0
2	B	1360	0	1284	37	0
2	D	1371	0	1296	33	0
2	F	1371	0	1297	31	0
2	H	1389	0	1309	34	0
2	J	1371	0	1297	35	0
2	L	1360	0	1284	26	0
3	M	28	0	25	0	0
3	O	28	0	25	1	0
3	S	28	0	25	1	0
4	N	31	0	26	1	0
4	P	31	0	26	1	0
4	Q	31	0	26	3	0
4	R	31	0	26	2	0
5	A	14	0	13	0	0
5	D	14	0	13	0	0
5	E	14	0	13	0	0
5	G	42	0	39	0	0
5	I	42	0	39	1	0
5	K	42	0	39	1	0
6	A	98	0	0	2	0
6	B	47	0	0	1	0
6	C	71	0	0	0	0
6	D	52	0	0	0	0
6	E	91	0	0	1	0
6	F	40	0	0	2	0
6	G	84	0	0	0	0
6	H	54	0	0	1	0
6	I	104	0	0	1	0
6	J	43	0	0	2	0
6	K	65	0	0	0	0
6	L	34	0	0	1	0
All	All	24423	0	22801	464	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.

The worst 5 of 464 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:C:283:THR:HG22	1:C:285:LYS:H	1.19	1.08
2:H:97:GLU:HG2	2:L:58:LYS:HD2	1.44	0.96
1:A:79:LEU:HA	1:A:81:THR:N	1.84	0.93
1:I:283:THR:HG22	1:I:285:LYS:H	1.35	0.91
1:A:283:THR:HG22	1:A:285:LYS:H	1.39	0.86

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	321/329 (98%)	301 (94%)	15 (5%)	5 (2%)	9 5
1	C	316/329 (96%)	299 (95%)	15 (5%)	2 (1%)	25 25
1	E	315/329 (96%)	299 (95%)	14 (4%)	2 (1%)	25 25
1	G	321/329 (98%)	301 (94%)	19 (6%)	1 (0%)	41 46
1	I	316/329 (96%)	298 (94%)	16 (5%)	2 (1%)	25 25
1	K	316/329 (96%)	298 (94%)	17 (5%)	1 (0%)	41 46
2	B	167/177 (94%)	159 (95%)	7 (4%)	1 (1%)	25 25
2	D	168/177 (95%)	154 (92%)	11 (6%)	3 (2%)	8 4
2	F	168/177 (95%)	156 (93%)	11 (6%)	1 (1%)	25 25
2	H	170/177 (96%)	164 (96%)	6 (4%)	0	100 100
2	J	168/177 (95%)	156 (93%)	7 (4%)	5 (3%)	4 2
2	L	167/177 (94%)	156 (93%)	10 (6%)	1 (1%)	25 25
All	All	2913/3036 (96%)	2741 (94%)	148 (5%)	24 (1%)	19 17

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	127	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	94	ASN
1	A	81	THR
1	A	265	GLY
1	C	264	ALA

### 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	285/290 (98%)	279 (98%)	6 (2%)	53	62
1	C	282/290 (97%)	270 (96%)	12 (4%)	29	33
1	E	281/290 (97%)	274 (98%)	7 (2%)	47	56
1	G	285/290 (98%)	279 (98%)	6 (2%)	53	62
1	I	282/290 (97%)	273 (97%)	9 (3%)	39	47
1	K	283/290 (98%)	272 (96%)	11 (4%)	32	38
2	B	145/152 (95%)	138 (95%)	7 (5%)	25	28
2	D	146/152 (96%)	140 (96%)	6 (4%)	30	36
2	F	146/152 (96%)	141 (97%)	5 (3%)	37	45
2	H	148/152 (97%)	144 (97%)	4 (3%)	44	54
2	J	146/152 (96%)	142 (97%)	4 (3%)	44	54
2	L	145/152 (95%)	141 (97%)	4 (3%)	43	52
All	All	2574/2652 (97%)	2493 (97%)	81 (3%)	40	49

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	155	VAL
1	K	116(A)	SER
1	I	225	ASP
2	J	164	GLU
1	K	159	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
2	H	95	ASN
1	I	159	ASN
2	L	95	ASN
1	K	18	HIS
2	L	60	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](#)

14 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	M	1	1,3	14,14,15	0.56	0	17,19,21	0.98	1 (5%)
3	NAG	M	2	3	14,14,15	0.57	0	17,19,21	0.64	0
4	GAL	N	1	4	11,11,12	0.64	0	15,15,17	0.77	0
4	SIA	N	2	4	20,20,21	0.61	0	24,28,31	1.05	2 (8%)
3	NAG	O	1	1,3	14,14,15	0.55	0	17,19,21	0.83	0
3	NAG	O	2	3	14,14,15	0.53	0	17,19,21	1.08	2 (11%)
4	GAL	P	1	4	11,11,12	0.58	0	15,15,17	0.83	1 (6%)
4	SIA	P	2	4	20,20,21	0.56	0	24,28,31	1.01	2 (8%)
4	GAL	Q	1	4	11,11,12	0.69	0	15,15,17	0.91	0
4	SIA	Q	2	4	20,20,21	0.61	0	24,28,31	0.96	1 (4%)
4	GAL	R	1	4	11,11,12	0.62	0	15,15,17	0.61	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	SIA	R	2	4	20,20,21	0.61	0	24,28,31	0.86	1 (4%)
3	NAG	S	1	1,3	14,14,15	0.51	0	17,19,21	0.85	0
3	NAG	S	2	3	14,14,15	0.49	0	17,19,21	0.75	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
3	NAG	M	1	1,3	-	2/6/23/26	0/1/1/1
3	NAG	M	2	3	-	0/6/23/26	0/1/1/1
4	GAL	N	1	4	-	0/2/19/22	0/1/1/1
4	SIA	N	2	4	-	0/18/34/38	0/1/1/1
3	NAG	O	1	1,3	-	2/6/23/26	0/1/1/1
3	NAG	O	2	3	-	2/6/23/26	0/1/1/1
4	GAL	P	1	4	-	1/2/19/22	0/1/1/1
4	SIA	P	2	4	-	1/18/34/38	0/1/1/1
4	GAL	Q	1	4	-	2/2/19/22	0/1/1/1
4	SIA	Q	2	4	-	0/18/34/38	0/1/1/1
4	GAL	R	1	4	-	0/2/19/22	0/1/1/1
4	SIA	R	2	4	-	0/18/34/38	0/1/1/1
3	NAG	S	1	1,3	-	0/6/23/26	0/1/1/1
3	NAG	S	2	3	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	N	2	SIA	C6-O6-C2	3.01	117.78	111.34
4	P	2	SIA	C6-O6-C2	2.62	116.95	111.34
3	O	2	NAG	C2-N2-C7	-2.38	119.52	122.90
4	P	1	GAL	O5-C5-C6	2.26	110.74	107.20
4	Q	2	SIA	C6-O6-C2	2.24	116.14	111.34

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	O	2	NAG	C8-C7-N2-C2

*Continued on next page...*

*Continued from previous page...*

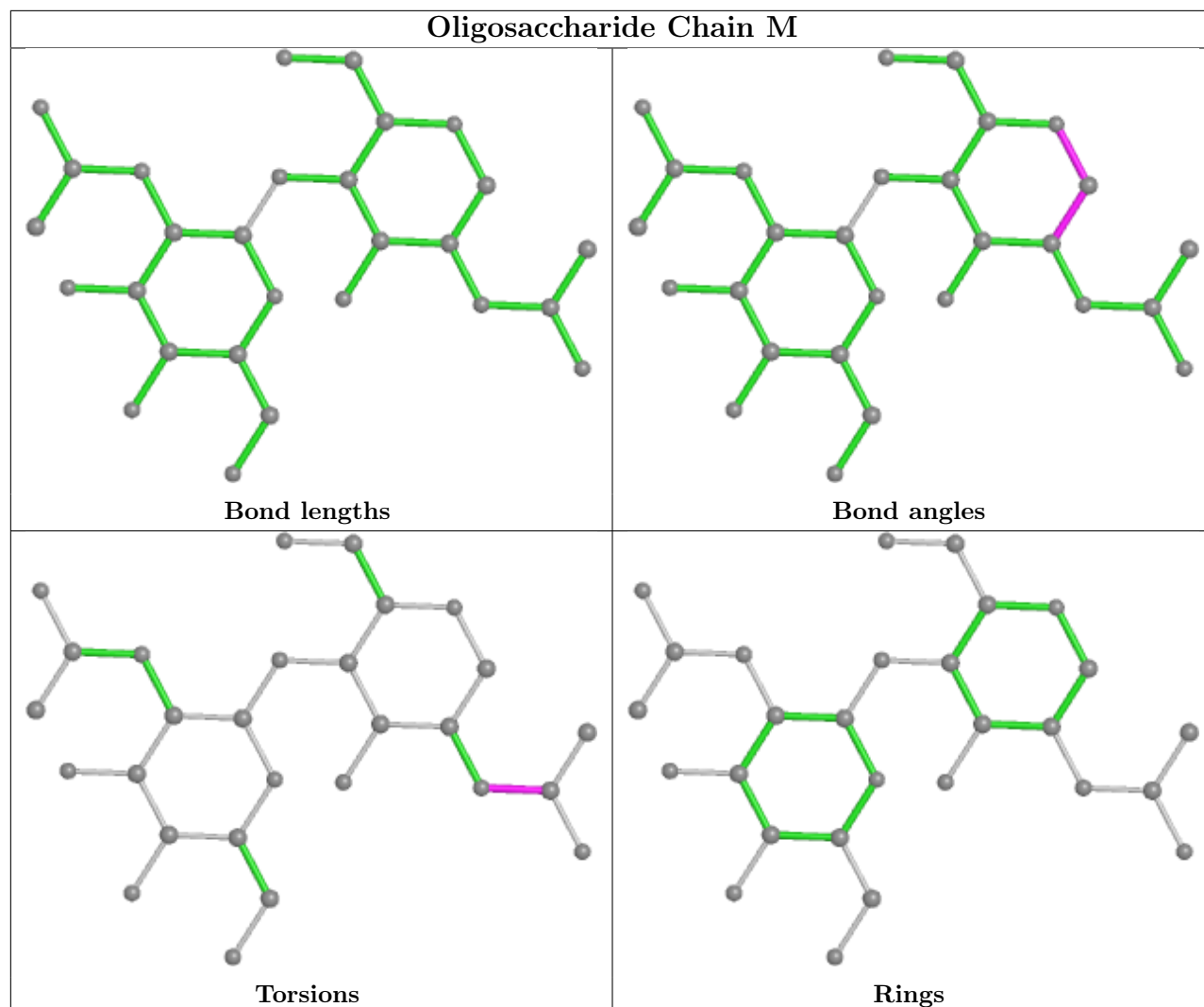
Mol	Chain	Res	Type	Atoms
3	O	2	NAG	O7-C7-N2-C2
3	S	2	NAG	C8-C7-N2-C2
3	S	2	NAG	O7-C7-N2-C2
4	Q	1	GAL	O5-C5-C6-O6

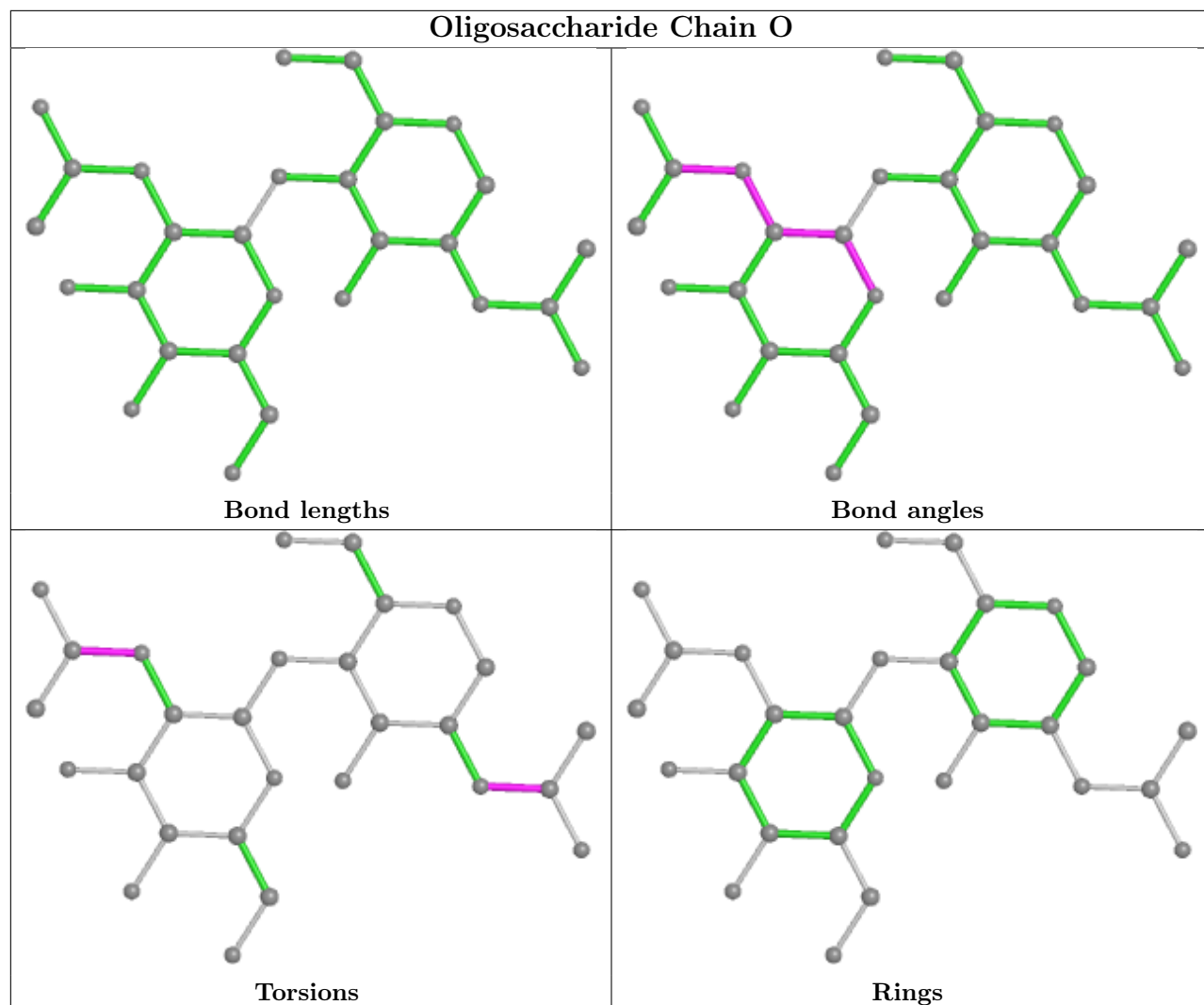
There are no ring outliers.

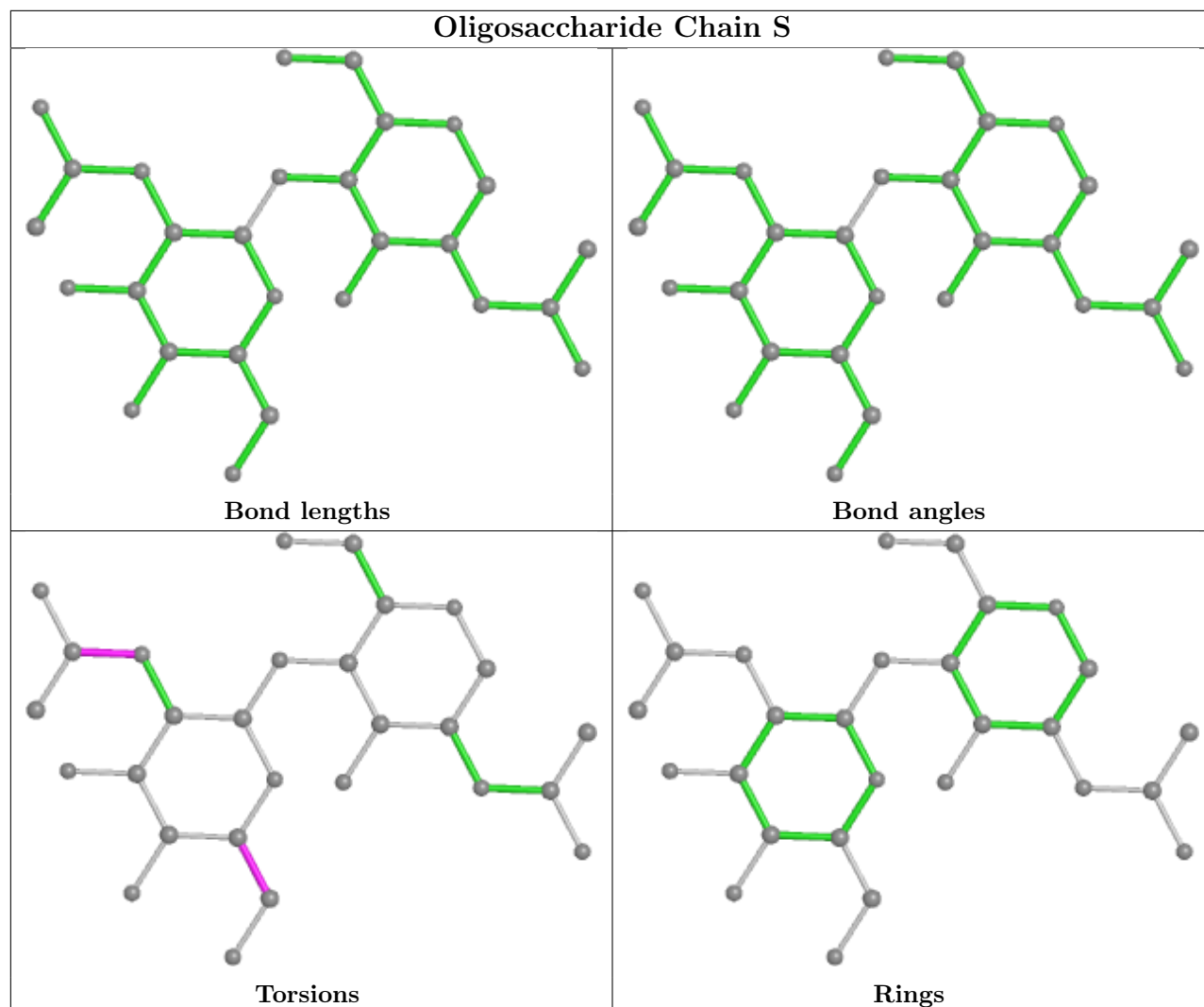
8 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	S	1	NAG	1	0
4	R	2	SIA	2	0
4	Q	2	SIA	3	0
4	Q	1	GAL	1	0
3	O	1	NAG	1	0
4	R	1	GAL	1	0
4	P	2	SIA	1	0
4	N	2	SIA	1	0

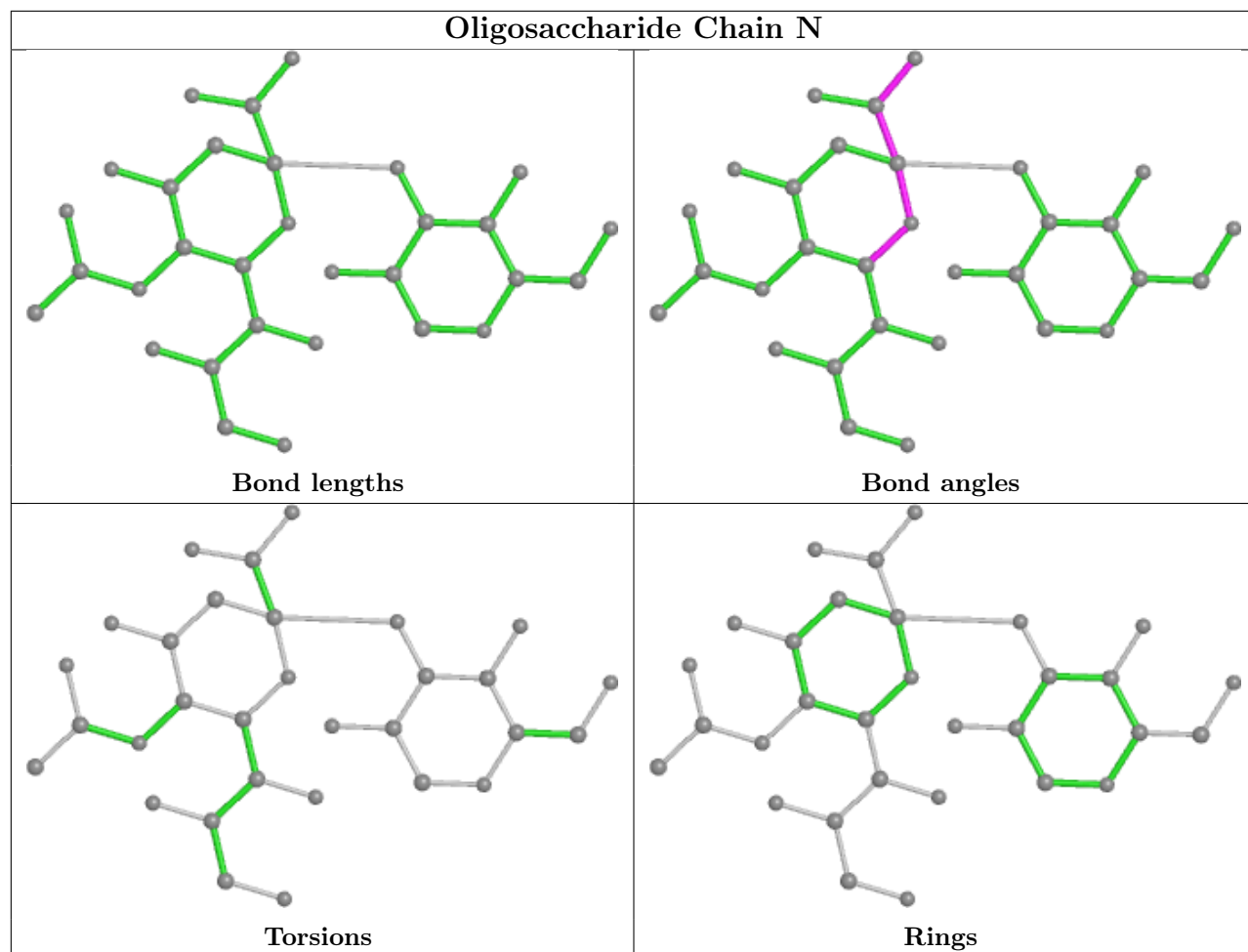
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

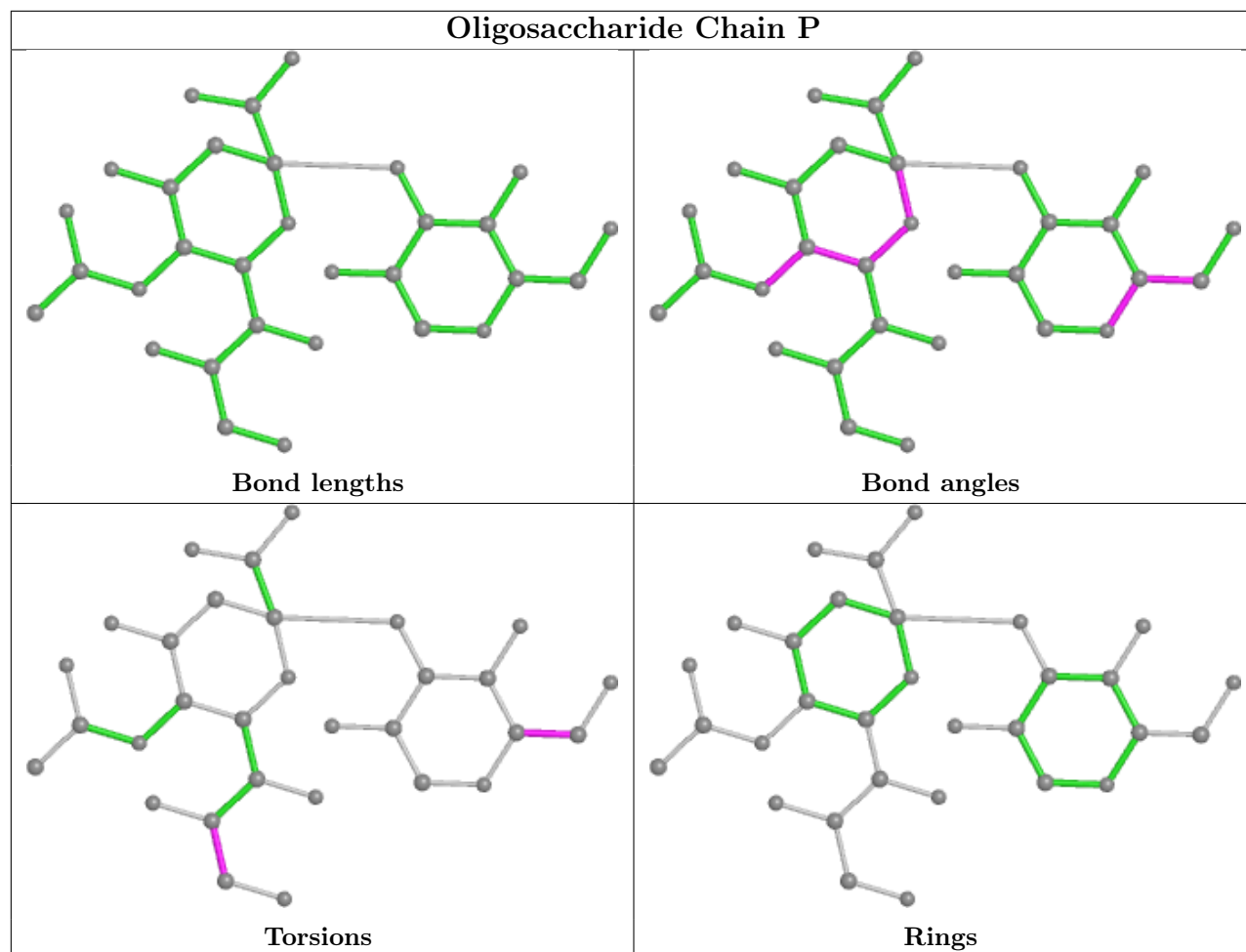


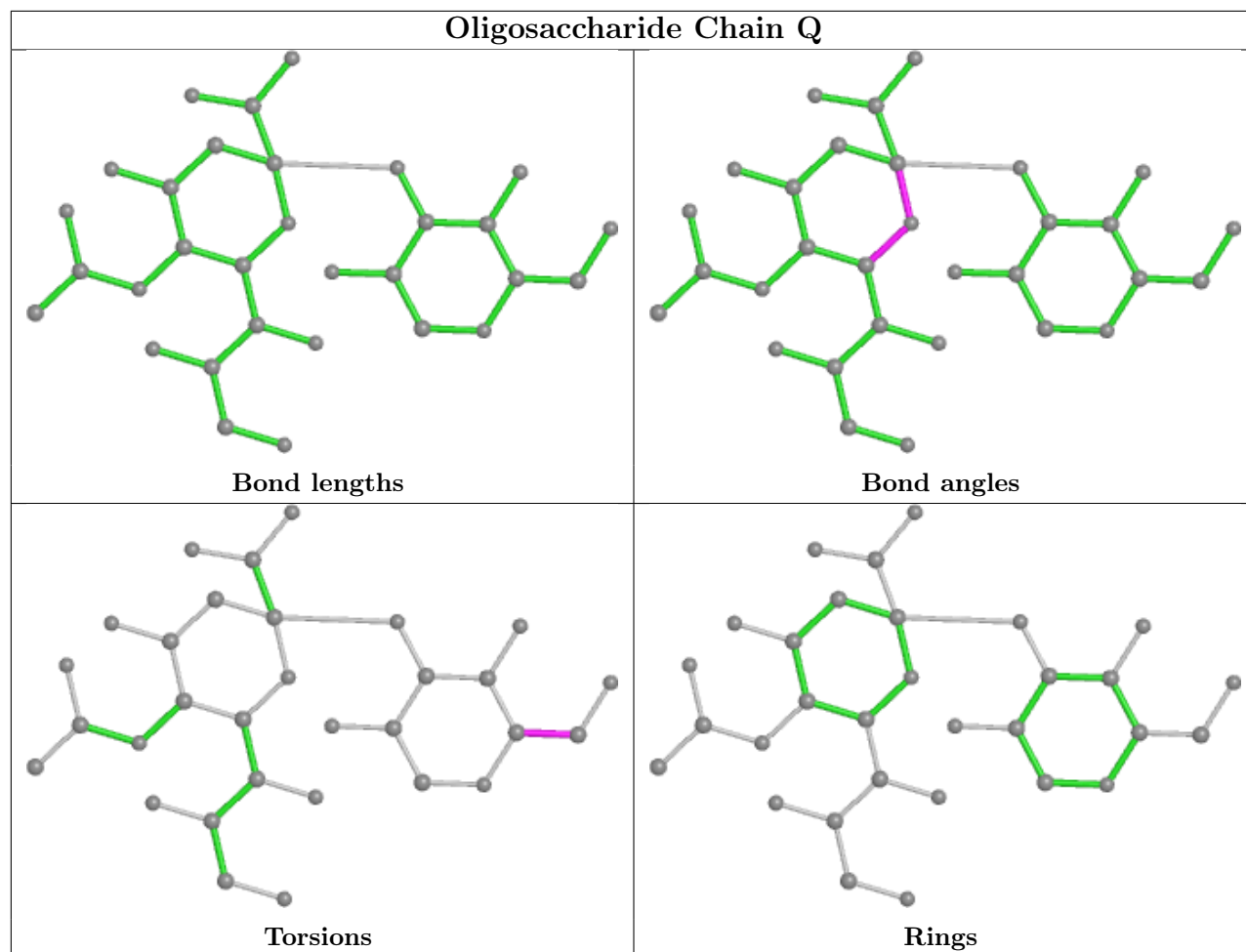


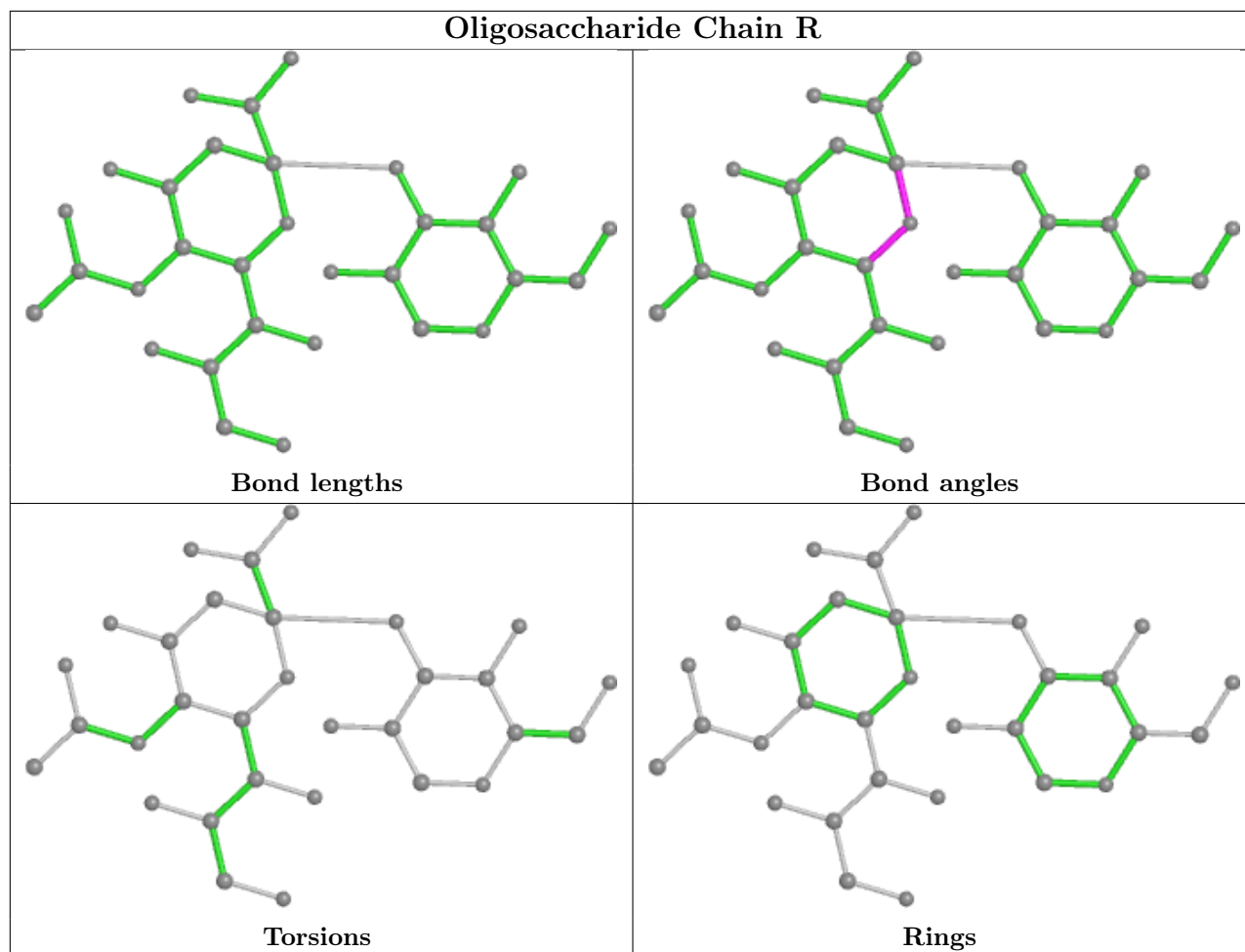












## 5.6 Ligand geometry [i](#)

12 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	E	341	1	14,14,15	0.52	0	17,19,21	0.92	0
5	NAG	K	641	1	14,14,15	0.48	0	17,19,21	0.72	0
5	NAG	A	332	1	14,14,15	0.52	0	17,19,21	0.76	0
5	NAG	G	411	1	14,14,15	0.48	0	17,19,21	0.99	1 (5%)
5	NAG	D	261	2	14,14,15	0.47	0	17,19,21	0.94	0
5	NAG	K	621	1	14,14,15	0.45	0	17,19,21	0.81	1 (5%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
5	NAG	I	521	1	14,14,15	0.46	0	17,19,21	0.79	0
5	NAG	K	651	1	14,14,15	0.52	0	17,19,21	0.77	0
5	NAG	G	451	1	14,14,15	0.50	0	17,19,21	0.98	0
5	NAG	G	431	1	14,14,15	0.53	0	17,19,21	0.81	1 (5%)
5	NAG	I	531	1	14,14,15	0.40	0	17,19,21	1.61	1 (5%)
5	NAG	I	541	1	14,14,15	0.45	0	17,19,21	1.07	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
5	NAG	E	341	1	-	1/6/23/26	0/1/1/1
5	NAG	K	641	1	-	0/6/23/26	0/1/1/1
5	NAG	A	332	1	-	2/6/23/26	0/1/1/1
5	NAG	G	411	1	-	4/6/23/26	0/1/1/1
5	NAG	D	261	2	-	2/6/23/26	0/1/1/1
5	NAG	K	621	1	-	0/6/23/26	0/1/1/1
5	NAG	I	521	1	-	2/6/23/26	0/1/1/1
5	NAG	K	651	1	-	4/6/23/26	0/1/1/1
5	NAG	G	451	1	-	0/6/23/26	0/1/1/1
5	NAG	G	431	1	-	0/6/23/26	0/1/1/1
5	NAG	I	531	1	-	1/6/23/26	0/1/1/1
5	NAG	I	541	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	I	531	NAG	C1-O5-C5	6.00	120.32	112.19
5	G	411	NAG	C1-O5-C5	2.72	115.88	112.19
5	I	541	NAG	C1-O5-C5	2.37	115.40	112.19
5	K	621	NAG	C1-O5-C5	2.14	115.09	112.19
5	G	431	NAG	C2-N2-C7	-2.03	120.01	122.90

There are no chirality outliers.

5 of 18 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	I	541	NAG	O5-C5-C6-O6
5	I	521	NAG	C8-C7-N2-C2
5	I	541	NAG	C4-C5-C6-O6
5	G	411	NAG	O5-C5-C6-O6
5	I	521	NAG	O7-C7-N2-C2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	K	641	NAG	1	0
5	I	531	NAG	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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	323/329 (98%)	-0.23	8 (2%) 57 60	20, 34, 57, 75	0
1	C	320/329 (97%)	-0.23	5 (1%) 72 74	27, 41, 59, 81	0
1	E	319/329 (96%)	-0.26	4 (1%) 77 79	22, 35, 50, 74	0
1	G	323/329 (98%)	-0.26	1 (0%) 94 94	21, 35, 51, 64	0
1	I	320/329 (97%)	-0.39	5 (1%) 72 74	22, 34, 54, 76	0
1	K	320/329 (97%)	-0.17	3 (0%) 84 85	25, 40, 58, 84	0
2	B	169/177 (95%)	0.68	22 (13%) 3 3	22, 46, 95, 111	0
2	D	170/177 (96%)	-0.02	5 (2%) 51 55	22, 47, 65, 81	0
2	F	170/177 (96%)	-0.08	3 (1%) 68 71	22, 42, 61, 71	0
2	H	172/177 (97%)	-0.17	2 (1%) 79 81	23, 42, 62, 74	0
2	J	170/177 (96%)	0.46	22 (12%) 3 3	21, 47, 81, 87	0
2	L	169/177 (95%)	0.32	16 (9%) 8 8	23, 47, 83, 99	0
All	All	2945/3036 (97%)	-0.10	96 (3%) 46 48	20, 38, 70, 111	0

The worst 5 of 96 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	166	ALA	7.6
2	B	160	PRO	7.2
2	B	159	TYR	5.8
1	E	290	THR	5.8
2	B	163	SER	5.7

### 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<sup>th</sup> 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.

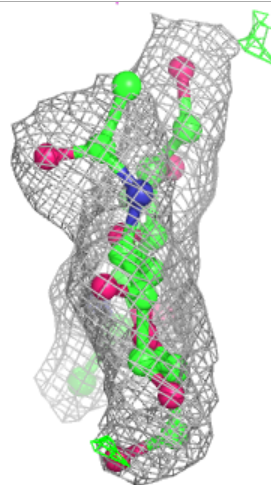
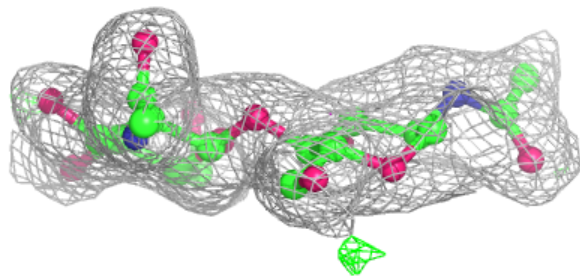
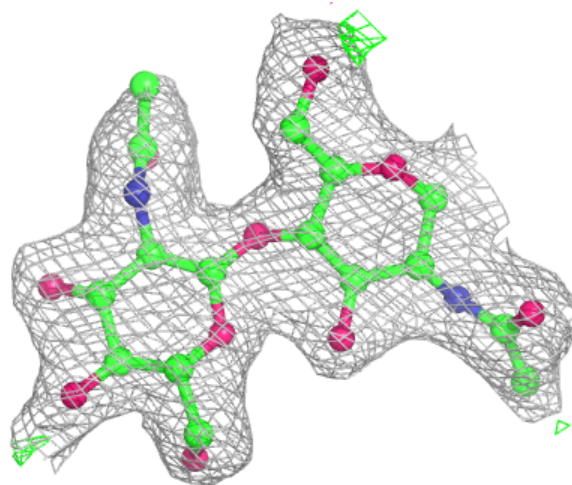
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
4	GAL	P	1	11/12	0.68	0.40	69,74,80,80	0
3	NAG	O	2	14/15	0.74	0.31	62,76,83,86	0
4	GAL	R	1	11/12	0.79	0.22	70,81,89,90	0
3	NAG	O	1	14/15	0.82	0.23	58,64,68,70	0
3	NAG	S	2	14/15	0.84	0.18	69,73,82,83	0
4	SIA	P	2	20/21	0.86	0.27	42,54,60,65	0
4	GAL	N	1	11/12	0.86	0.13	57,66,70,70	0
4	GAL	Q	1	11/12	0.88	0.14	50,60,65,65	0
4	SIA	R	2	20/21	0.88	0.14	42,56,66,68	0
3	NAG	M	2	14/15	0.90	0.13	46,53,62,66	0
4	SIA	N	2	20/21	0.91	0.11	36,46,54,55	0
3	NAG	S	1	14/15	0.93	0.14	49,58,66,69	0
4	SIA	Q	2	20/21	0.93	0.10	35,40,46,48	0
3	NAG	M	1	14/15	0.94	0.09	34,43,47,49	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.



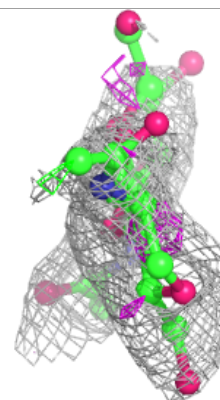
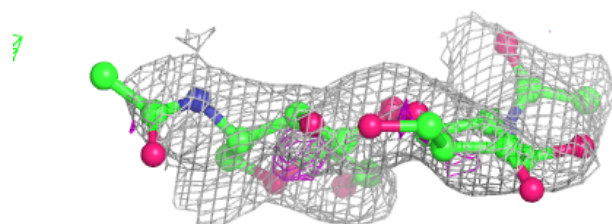
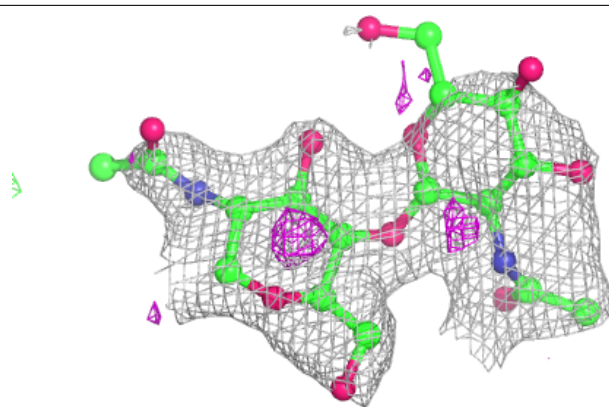
**Electron density around Chain M:**

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

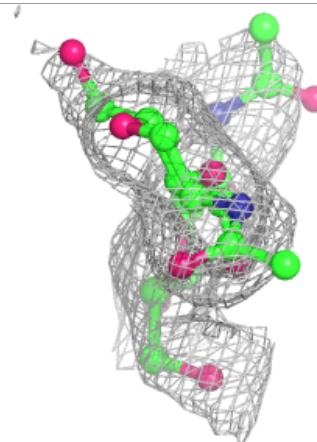
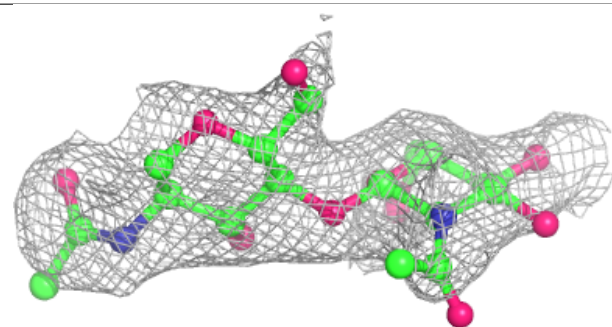
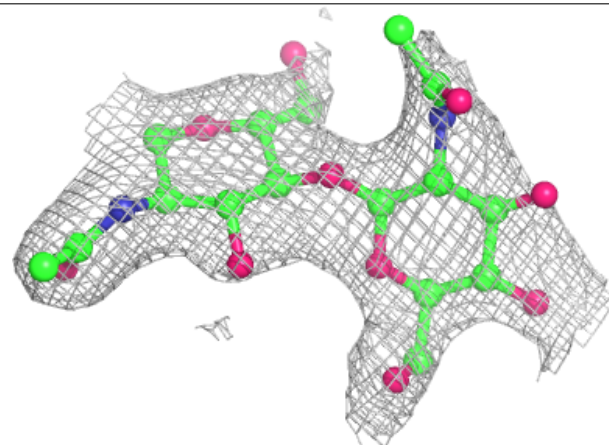


**Electron density around Chain O:**

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

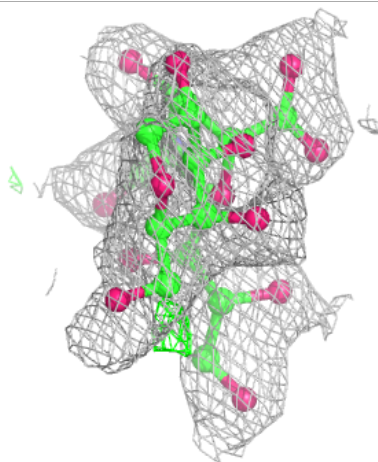
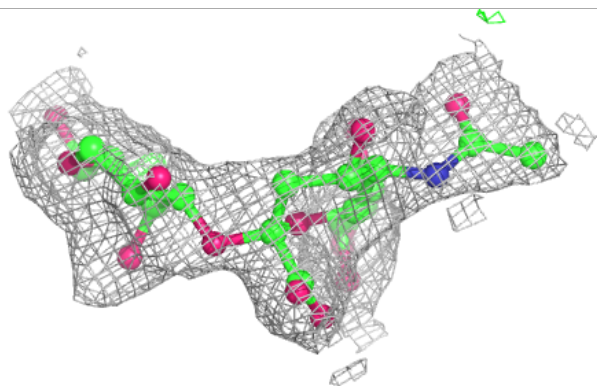
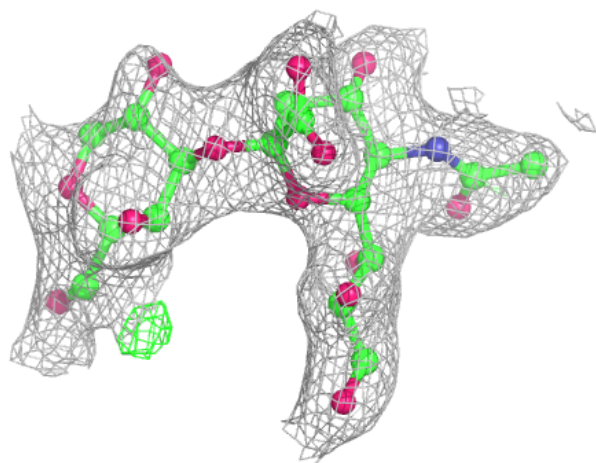
**Electron density around Chain S:**

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



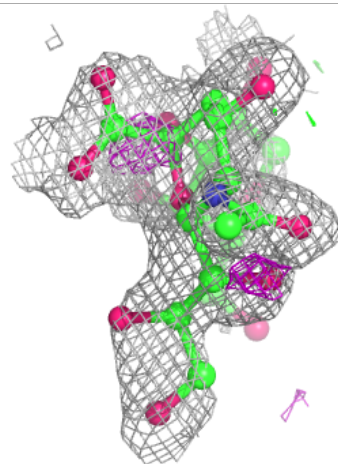
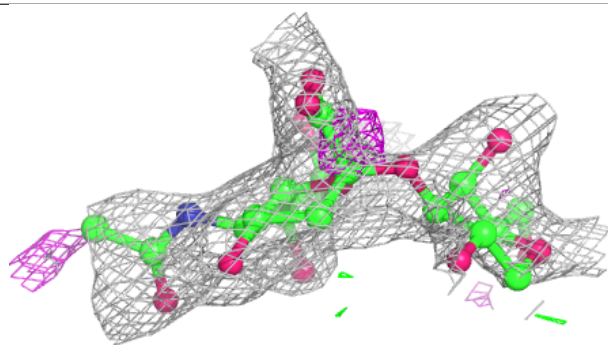
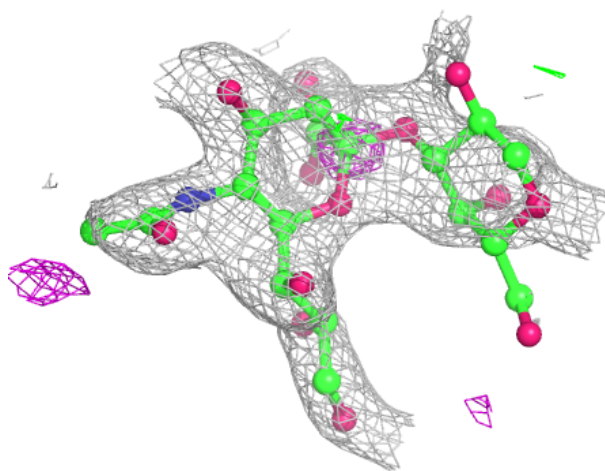
**Electron density around Chain N:**

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



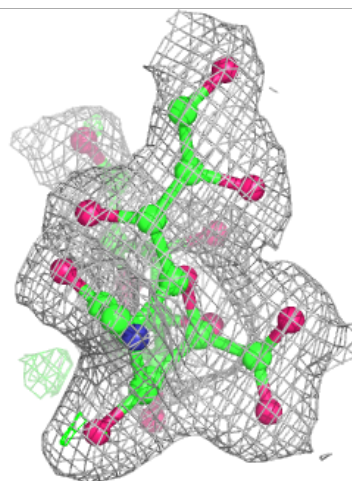
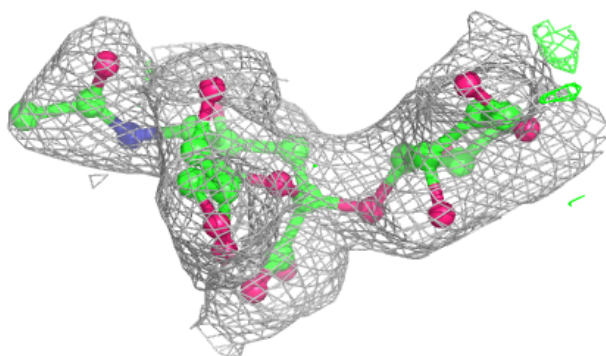
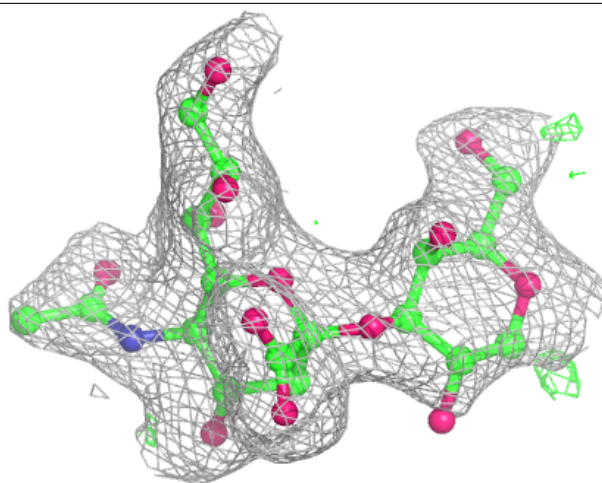
**Electron density around Chain P:**

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

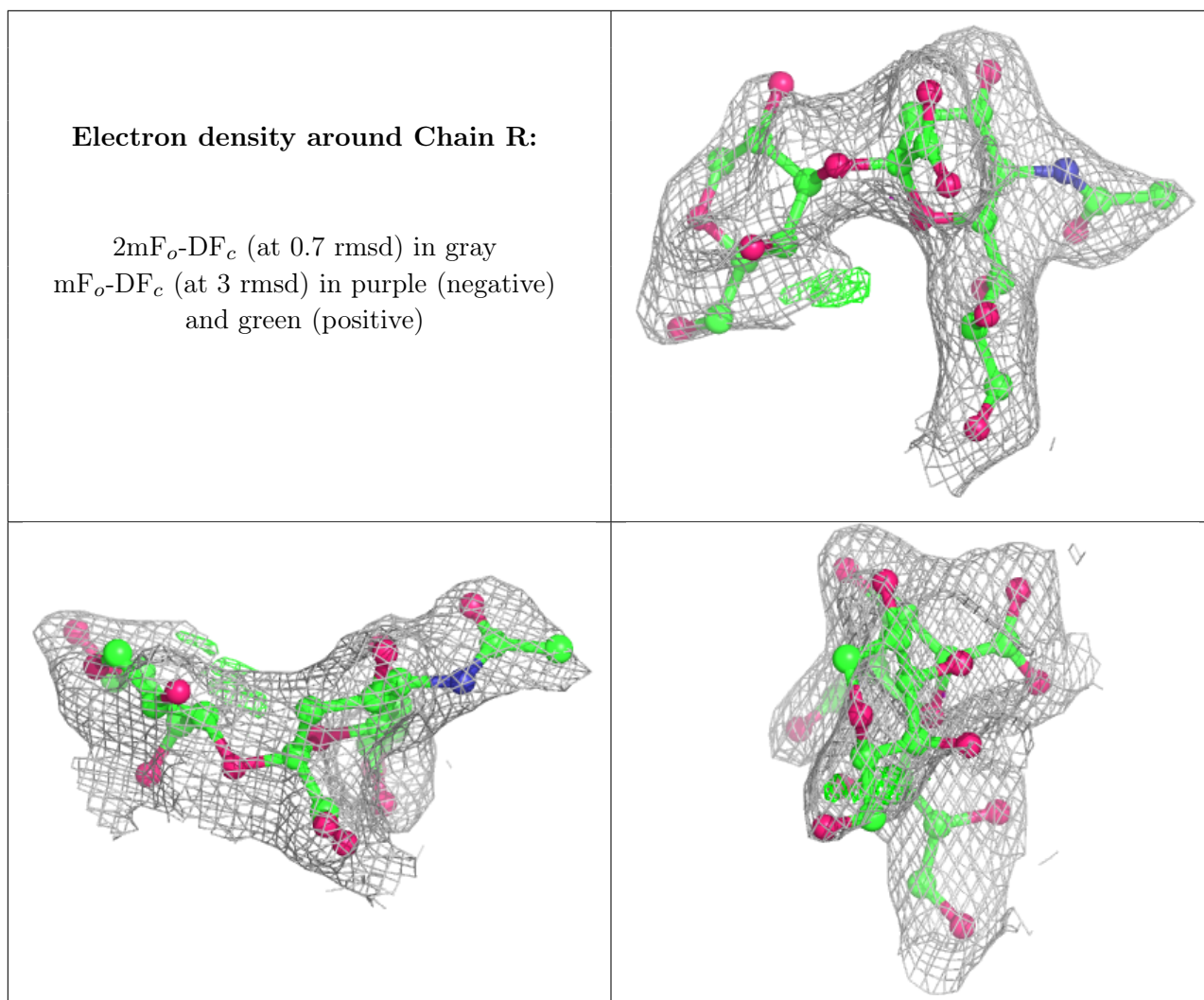


**Electron density around Chain Q:**

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







## 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
5	NAG	G	451	14/15	0.70	0.34	59,86,95,99	0
5	NAG	I	521	14/15	0.70	0.20	73,87,91,93	0
5	NAG	G	411	14/15	0.73	0.29	64,73,79,80	0
5	NAG	A	332	14/15	0.75	0.35	55,63,71,72	0
5	NAG	E	341	14/15	0.77	0.25	68,75,83,85	0
5	NAG	K	621	14/15	0.80	0.18	58,69,78,82	0
5	NAG	K	651	14/15	0.83	0.16	60,77,87,88	0
5	NAG	I	531	14/15	0.87	0.14	40,54,58,59	0
5	NAG	D	261	14/15	0.88	0.15	57,65,72,73	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
5	NAG	K	641	14/15	0.88	0.18	55,61,68,74	0
5	NAG	I	541	14/15	0.88	0.14	54,66,70,76	0
5	NAG	G	431	14/15	0.91	0.09	38,52,61,67	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.