



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 7, 2024 – 04:40 am GMT

PDB ID : 6HKU  
Title : Trichodysplasia spinulosa-associated polyomavirus (TSPyV) VP1 in complex with sialylated precision glycooligomers  
Authors : Rustmeier, N.H.; Stehle, T.  
Deposited on : 2018-09-08  
Resolution : 1.98 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 : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
buster-report : 1.1.7 (2018)  
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) : 2.36

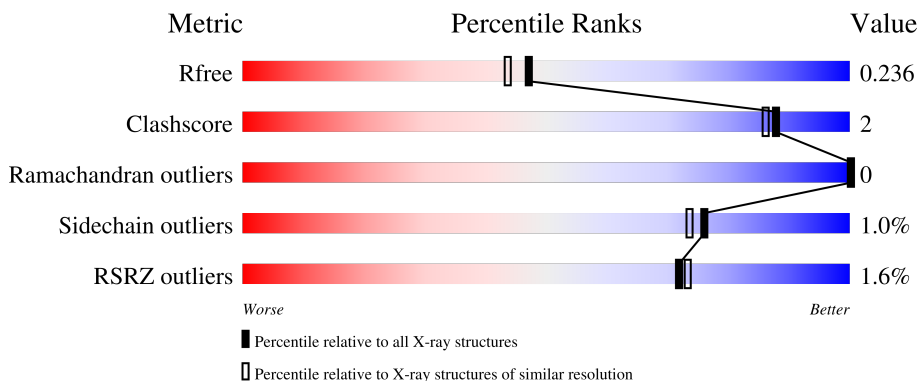
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.98 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.



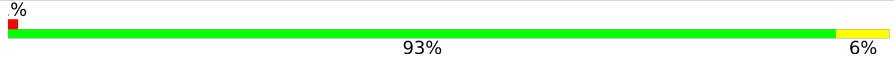
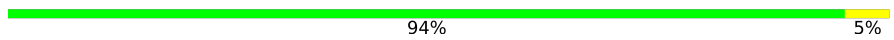
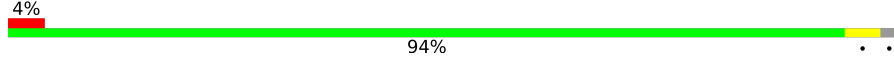
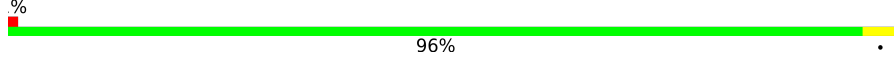
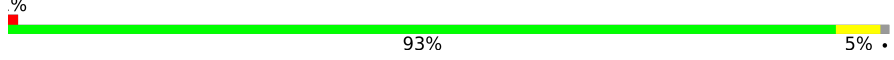

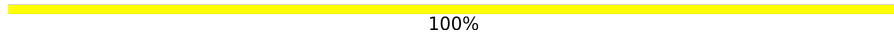
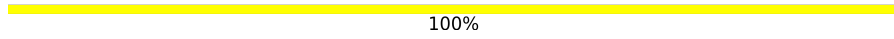
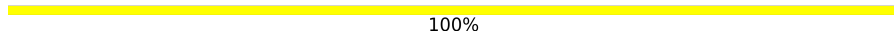
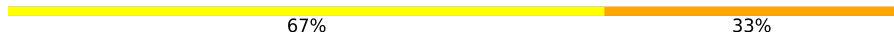
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	11647 (2.00-1.96)
Clashscore	141614	1014 (1.98-1.98)
Ramachandran outliers	138981	1006 (1.98-1.98)
Sidechain outliers	138945	1006 (1.98-1.98)
RSRZ outliers	127900	11410 (2.00-1.96)

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	272	 94% 5%
1	B	272	 92% 6% ..
1	C	272	 92% 8%
1	D	272	 94% ..
1	E	272	 92% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	272	 % 93% 6%
1	G	272	 % 94% 5%
1	H	272	 % 94% . .
1	I	272	 % 96% .
1	J	272	 % 93% 5% .
2	K	2	 50% 50%
2	L	2	 100%
2	M	2	 100%
3	N	3	 100%
3	O	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:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	GAL	N	2	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 23268 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 Capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	272	Total 2094	C 1318	N 349	O 414	S 13	0	0	0
1	B	267	Total 2057	C 1296	N 342	O 407	S 12	0	0	0
1	C	272	Total 2102	C 1320	N 348	O 421	S 13	0	2	0
1	D	265	Total 2023	C 1281	N 335	O 395	S 12	0	0	1
1	E	271	Total 2088	C 1315	N 348	O 413	S 12	0	1	0
1	F	271	Total 2076	C 1307	N 344	O 413	S 12	0	0	0
1	G	271	Total 2088	C 1315	N 346	O 415	S 12	0	1	0
1	H	267	Total 2028	C 1281	N 337	O 398	S 12	0	0	0
1	I	272	Total 2083	C 1312	N 349	O 410	S 12	0	0	0
1	J	268	Total 2054	C 1293	N 341	O 408	S 12	0	1	0

There are 10 discrepancies between the modelled and reference sequences:

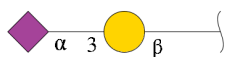
Chain	Residue	Modelled	Actual	Comment	Reference
A	107	SER	CYS	conflict	UNP E2ESL7
B	107	SER	CYS	conflict	UNP E2ESL7
C	107	SER	CYS	conflict	UNP E2ESL7
D	107	SER	CYS	conflict	UNP E2ESL7
E	107	SER	CYS	conflict	UNP E2ESL7
F	107	SER	CYS	conflict	UNP E2ESL7
G	107	SER	CYS	conflict	UNP E2ESL7
H	107	SER	CYS	conflict	UNP E2ESL7
I	107	SER	CYS	conflict	UNP E2ESL7

*Continued on next page...*

Continued from previous page...

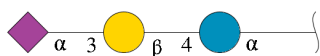
Chain	Residue	Modelled	Actual	Comment	Reference
J	107	SER	CYS	conflict	UNP E2ESL7

- Molecule 2 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			
2	K	2	Total	C	N	O	0	0	0
			32	17	1	14			
2	L	2	Total	C	N	O	0	0	0
			32	17	1	14			
2	M	2	Total	C	N	O	0	0	0
			32	17	1	14			

- Molecule 3 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-alpha-D-glucopyranose.



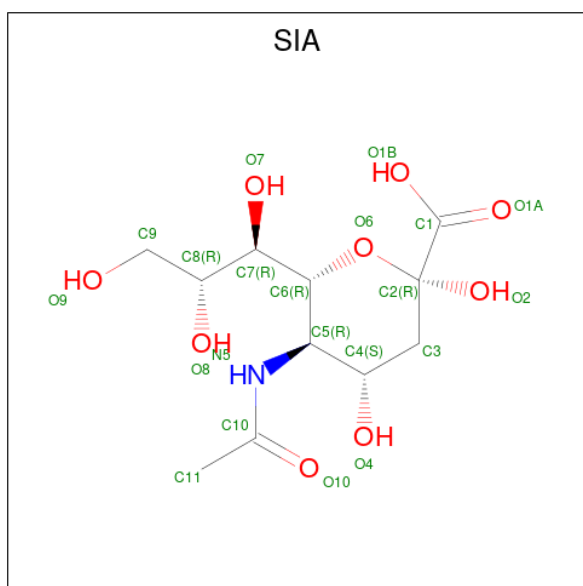
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	N	3	Total	C	N	O	0	0	0
			43	23	1	19			
3	O	3	Total	C	N	O	0	0	0
			43	23	1	19			

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	C	1	Total	C	O	0	0
			6	3	3		
4	H	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is N-acetyl-alpha-neuraminic acid (three-letter code: SIA) (formula:  $C_{11}H_{19}NO_9$ ) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	D	1	Total	C	N	O	0	0
			21	11	1	9		

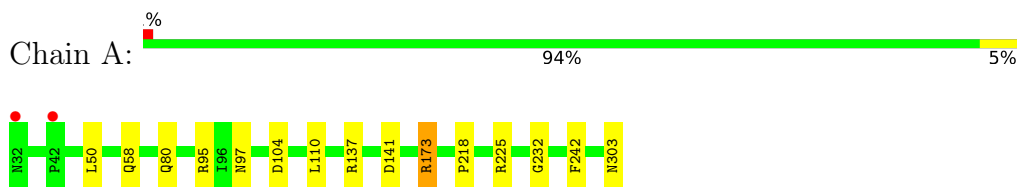
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	232	Total O 234 234	0	2
6	B	228	Total O 229 229	0	1
6	C	213	Total O 215 215	0	2
6	D	248	Total O 249 249	0	1
6	E	247	Total O 250 250	0	3
6	F	251	Total O 251 251	0	0
6	G	214	Total O 214 214	0	0
6	H	220	Total O 220 220	0	0
6	I	239	Total O 240 240	0	1
6	J	257	Total O 258 258	0	1

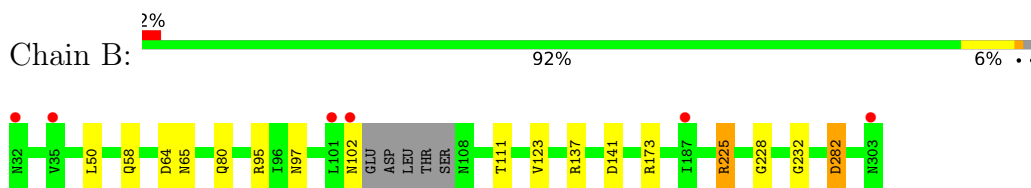
### 3 Residue-property plots [i](#)

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.

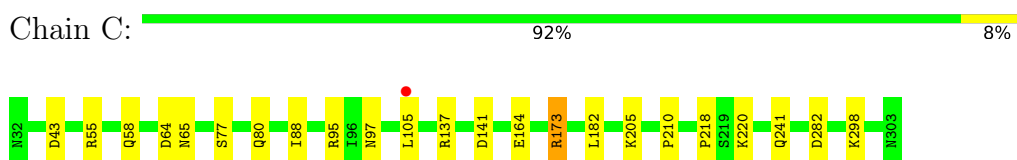
- Molecule 1: Capsid protein VP1



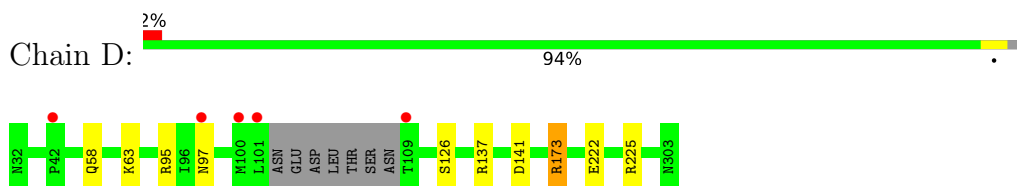
- Molecule 1: Capsid protein VP1



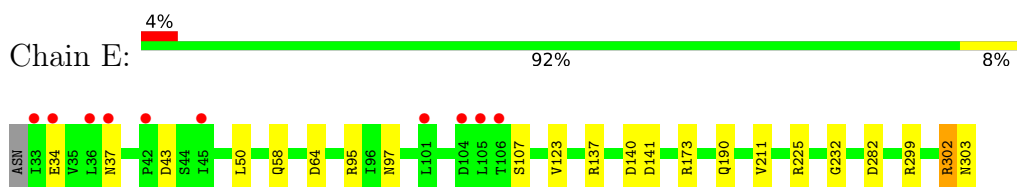
- Molecule 1: Capsid protein VP1



- Molecule 1: Capsid protein VP1

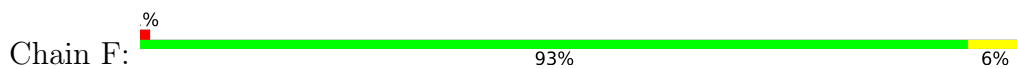


- Molecule 1: Capsid protein VP1



- Molecule 1: Capsid protein VP1





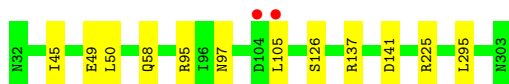
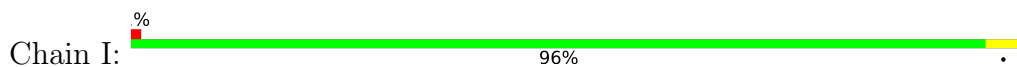
- Molecule 1: Capsid protein VP1



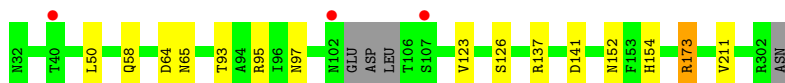
- Molecule 1: Capsid protein VP1



- Molecule 1: Capsid protein VP1



- Molecule 1: Capsid protein VP1



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



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




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

Chain M:  100%


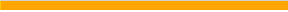
GAL1  
SIA2

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

Chain N:  100%

GLC1  
GAL2  
SIA3

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

Chain O:  67%  33%

GLC1  
GAL2  
SIA3

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	139.28Å 146.16Å 150.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.05 – 1.98 49.05 – 1.98	Depositor EDS
% Data completeness (in resolution range)	99.8 (49.05-1.98) 99.8 (49.05-1.98)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.64 (at 1.98Å)	Xtrriage
Refinement program	REFMAC 5.8.0230	Depositor
R, $R_{free}$	0.185 , 0.230 0.192 , 0.236	Depositor DCC
$R_{free}$ test set	2122 reflections (1.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.9	Xtrriage
Anisotropy	0.465	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 48.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.012 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	23268	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.91% 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, GAL, GLC, GOL

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	0.53	0/2139	0.70	1/2911 (0.0%)
1	B	0.50	0/2101	0.70	1/2860 (0.0%)
1	C	0.54	0/2150	0.74	1/2928 (0.0%)
1	D	0.52	0/2067	0.70	0/2813
1	E	0.53	0/2136	0.73	1/2909 (0.0%)
1	F	0.54	1/2121 (0.0%)	0.70	0/2891
1	G	0.54	2/2133 (0.1%)	0.73	1/2906 (0.0%)
1	H	0.50	0/2072	0.68	0/2822
1	I	0.54	1/2128 (0.0%)	0.70	0/2898
1	J	0.52	0/2098	0.70	0/2859
All	All	0.53	4/21145 (0.0%)	0.71	5/28797 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	2
1	C	0	2
1	D	0	2
1	E	0	2
1	F	0	1
1	G	0	4
1	I	0	1
1	J	0	1
All	All	0	16

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	161	GLU	CD-OE2	-5.78	1.19	1.25
1	G	161	GLU	CD-OE1	5.60	1.31	1.25
1	F	161	GLU	CD-OE1	-5.46	1.19	1.25
1	I	49	GLU	CD-OE2	5.04	1.31	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	225	ARG	NE-CZ-NH2	-6.68	116.96	120.30
1	A	225	ARG	NE-CZ-NH2	-5.78	117.41	120.30
1	E	299	ARG	NE-CZ-NH1	-5.76	117.42	120.30
1	B	225	ARG	NE-CZ-NH2	-5.39	117.60	120.30
1	C	173	ARG	CG-CD-NE	-5.32	100.64	111.80

There are no chirality outliers.

5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	173	ARG	Sidechain
1	B	173	ARG	Sidechain
1	B	225	ARG	Sidechain
1	C	173	ARG	Sidechain
1	C	55	ARG	Sidechain

## 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	2094	0	2016	9	0
1	B	2057	0	1977	10	0
1	C	2102	0	2010	17	0
1	D	2023	0	1950	6	0
1	E	2088	0	2009	13	0
1	F	2076	0	1985	12	0
1	G	2088	0	1994	5	0
1	H	2028	0	1931	5	0
1	I	2083	0	1995	6	0
1	J	2054	0	1958	9	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	K	32	0	28	0	0
2	L	32	0	28	0	0
2	M	32	0	28	0	0
3	N	43	0	37	0	0
3	O	43	0	37	1	0
4	C	6	0	8	0	0
4	H	6	0	8	0	0
5	D	21	0	18	0	0
6	A	234	0	0	4	0
6	B	229	0	0	2	0
6	C	215	0	0	6	0
6	D	249	0	0	2	0
6	E	250	0	0	4	0
6	F	251	0	0	1	0
6	G	214	0	0	3	0
6	H	220	0	0	2	0
6	I	240	0	0	2	0
6	J	258	0	0	2	0
All	All	23268	0	20017	83	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.

The worst 5 of 83 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:164[B]:GLU:HG3	6:C:505:HOH:O	1.62	0.98
1:G:71:LYS:CG	6:G:608:HOH:O	2.32	0.76
1:B:282:ASP:OD1	6:B:501:HOH:O	2.05	0.75
1:E:282[A]:ASP:OD1	6:E:401:HOH:O	2.04	0.74
1:C:164[B]:GLU:CG	6:C:505:HOH:O	2.28	0.74

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	270/272 (99%)	260 (96%)	10 (4%)	0	100	100
1	B	263/272 (97%)	254 (97%)	9 (3%)	0	100	100
1	C	272/272 (100%)	264 (97%)	8 (3%)	0	100	100
1	D	261/272 (96%)	253 (97%)	8 (3%)	0	100	100
1	E	270/272 (99%)	261 (97%)	9 (3%)	0	100	100
1	F	269/272 (99%)	260 (97%)	9 (3%)	0	100	100
1	G	270/272 (99%)	261 (97%)	9 (3%)	0	100	100
1	H	263/272 (97%)	255 (97%)	8 (3%)	0	100	100
1	I	270/272 (99%)	263 (97%)	7 (3%)	0	100	100
1	J	265/272 (97%)	257 (97%)	8 (3%)	0	100	100
All	All	2673/2720 (98%)	2588 (97%)	85 (3%)	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	231/241 (96%)	229 (99%)	2 (1%)	78	77
1	B	228/241 (95%)	225 (99%)	3 (1%)	69	64
1	C	234/241 (97%)	232 (99%)	2 (1%)	78	77
1	D	221/241 (92%)	220 (100%)	1 (0%)	88	87
1	E	231/241 (96%)	228 (99%)	3 (1%)	69	64
1	F	229/241 (95%)	226 (99%)	3 (1%)	69	64
1	G	229/241 (95%)	226 (99%)	3 (1%)	69	64
1	H	219/241 (91%)	217 (99%)	2 (1%)	78	77
1	I	228/241 (95%)	226 (99%)	2 (1%)	78	77

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	J	226/241 (94%)	224 (99%)	2 (1%)	78	77
All	All	2276/2410 (94%)	2253 (99%)	23 (1%)	76	73

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	43	ASP
1	H	50	LEU
1	G	58	GLN
1	H	58	GLN
1	C	241	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	102	ASN
1	G	102	ASN
1	I	80	GLN
1	J	80	GLN
1	J	152	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](#)

12 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
2	GAL	K	1	2	12,12,12	0.73	0	17,17,17	0.75	0
2	SIA	K	2	2	20,20,21	1.56	3 (15%)	24,28,31	1.78	8 (33%)
2	GAL	L	1	2	12,12,12	0.65	0	17,17,17	1.34	3 (17%)
2	SIA	L	2	2	20,20,21	1.11	1 (5%)	24,28,31	1.68	5 (20%)
2	GAL	M	1	2	12,12,12	0.69	0	17,17,17	1.11	1 (5%)
2	SIA	M	2	2	20,20,21	0.86	0	24,28,31	1.16	3 (12%)
3	GLC	N	1	3	12,12,12	0.62	0	17,17,17	1.16	1 (5%)
3	GAL	N	2	3	11,11,12	0.69	0	15,15,17	1.58	2 (13%)
3	SIA	N	3	3	20,20,21	1.19	1 (5%)	24,28,31	1.79	7 (29%)
3	GLC	O	1	3	12,12,12	0.94	0	17,17,17	1.49	5 (29%)
3	GAL	O	2	3	11,11,12	0.82	0	15,15,17	1.61	2 (13%)
3	SIA	O	3	3	20,20,21	1.01	1 (5%)	24,28,31	1.25	2 (8%)

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	GAL	K	1	2	-	2/2/22/22	0/1/1/1
2	SIA	K	2	2	-	6/18/34/38	0/1/1/1
2	GAL	L	1	2	-	2/2/22/22	0/1/1/1
2	SIA	L	2	2	-	7/18/34/38	0/1/1/1
2	GAL	M	1	2	-	2/2/22/22	0/1/1/1
2	SIA	M	2	2	-	4/18/34/38	0/1/1/1
3	GLC	N	1	3	-	0/2/22/22	0/1/1/1
3	GAL	N	2	3	-	0/2/19/22	0/1/1/1
3	SIA	N	3	3	-	5/18/34/38	0/1/1/1
3	GLC	O	1	3	-	2/2/22/22	0/1/1/1
3	GAL	O	2	3	-	2/2/19/22	0/1/1/1
3	SIA	O	3	3	-	5/18/34/38	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	K	2	SIA	C2-C1	3.80	1.55	1.52
3	N	3	SIA	C2-C1	3.22	1.55	1.52

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	K	2	SIA	C7-C6	2.79	1.56	1.53
2	K	2	SIA	C4-C5	2.64	1.55	1.53
2	L	2	SIA	C2-C1	2.32	1.54	1.52

The worst 5 of 39 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	N	3	SIA	O6-C2-C1	4.99	117.49	107.70
2	K	2	SIA	O6-C2-C1	4.46	116.44	107.70
3	N	2	GAL	C3-C4-C5	3.90	117.19	110.24
2	L	2	SIA	O6-C2-C1	3.84	115.24	107.70
2	L	2	SIA	O6-C2-C3	-3.53	105.60	110.46

There are no chirality outliers.

5 of 37 torsion outliers are listed below:

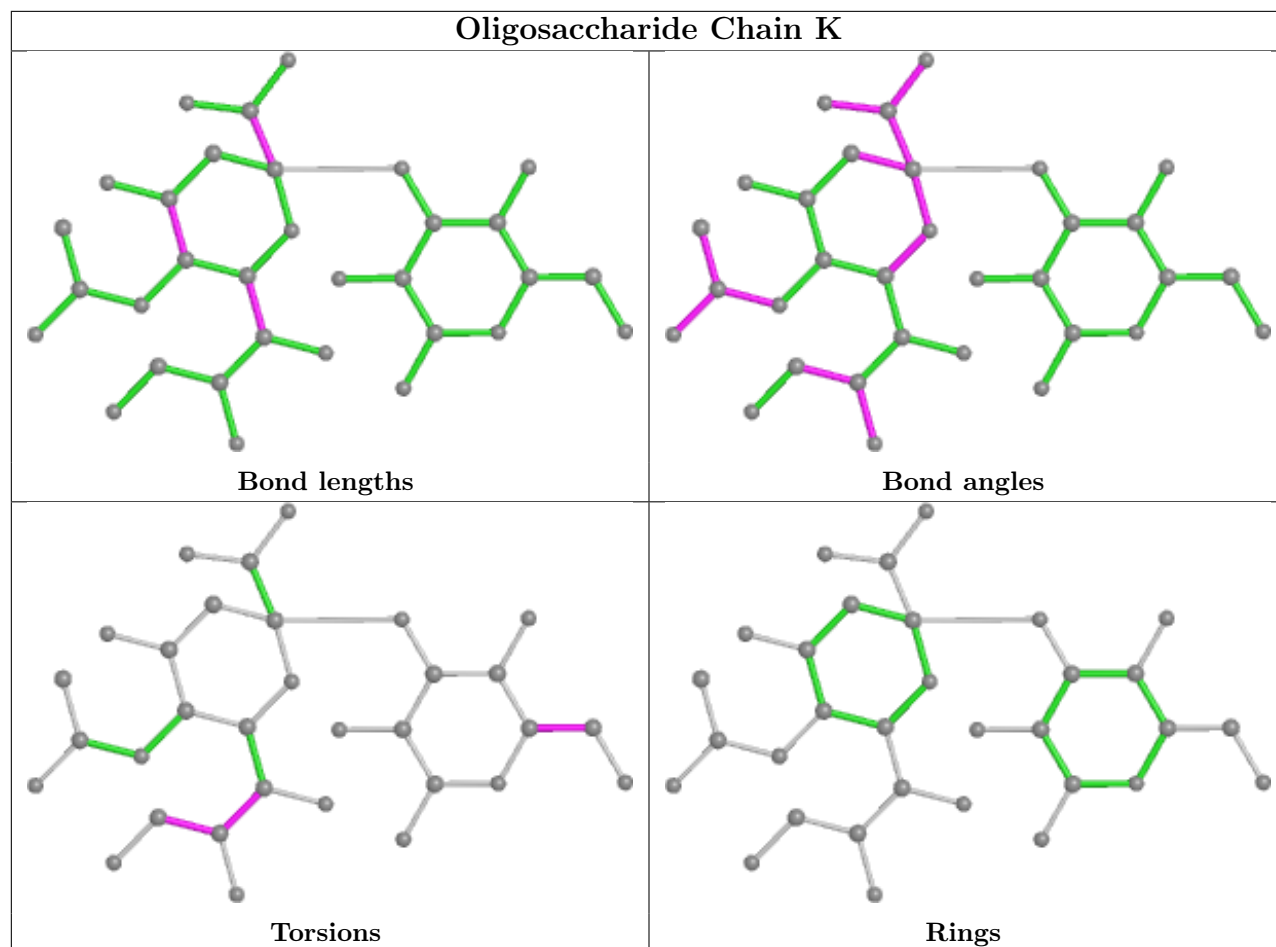
Mol	Chain	Res	Type	Atoms
2	K	2	SIA	C6-C7-C8-C9
2	K	2	SIA	O7-C7-C8-C9
2	K	2	SIA	O8-C8-C9-O9
2	L	2	SIA	O8-C8-C9-O9
2	K	2	SIA	C7-C8-C9-O9

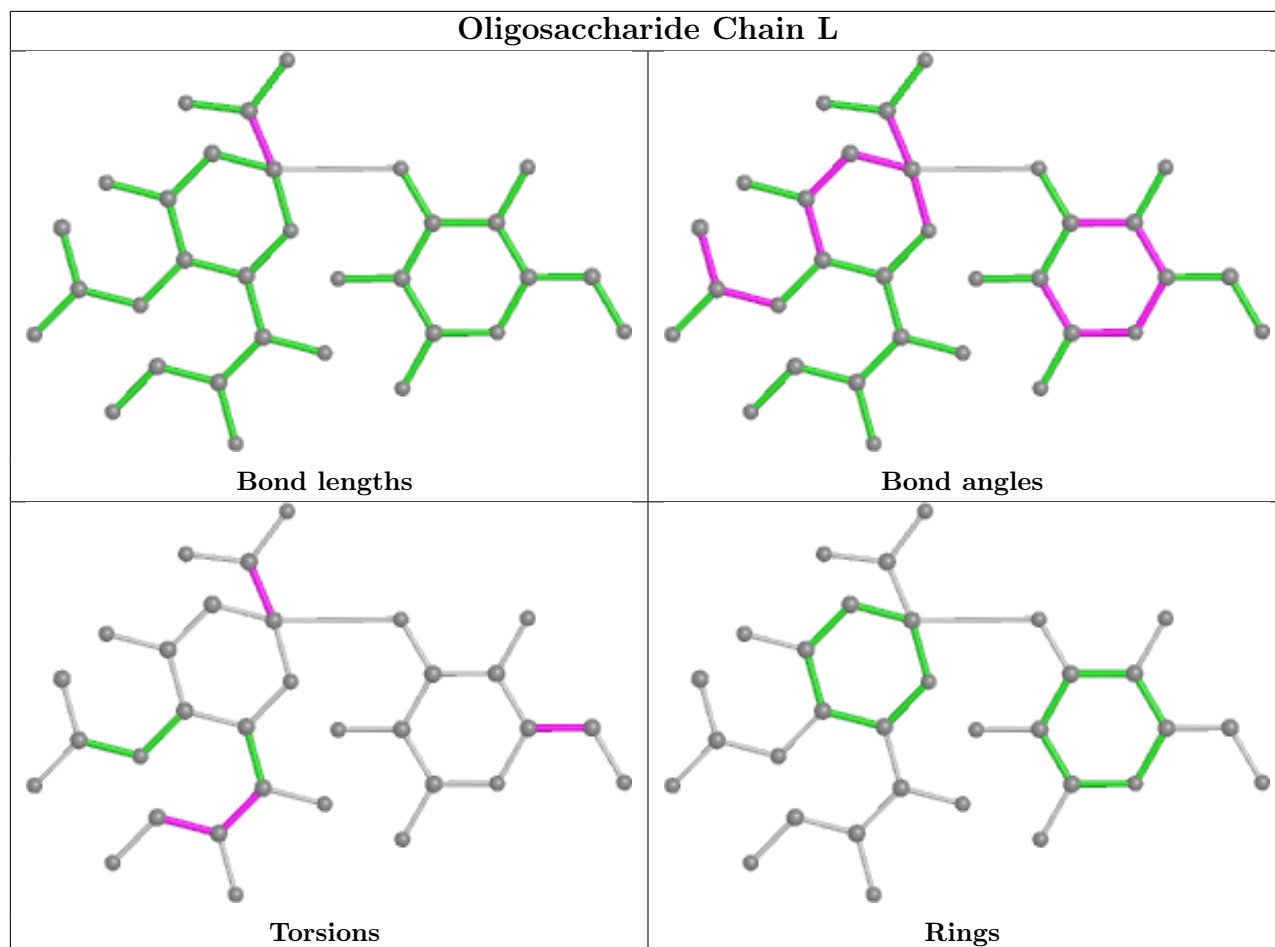
There are no ring outliers.

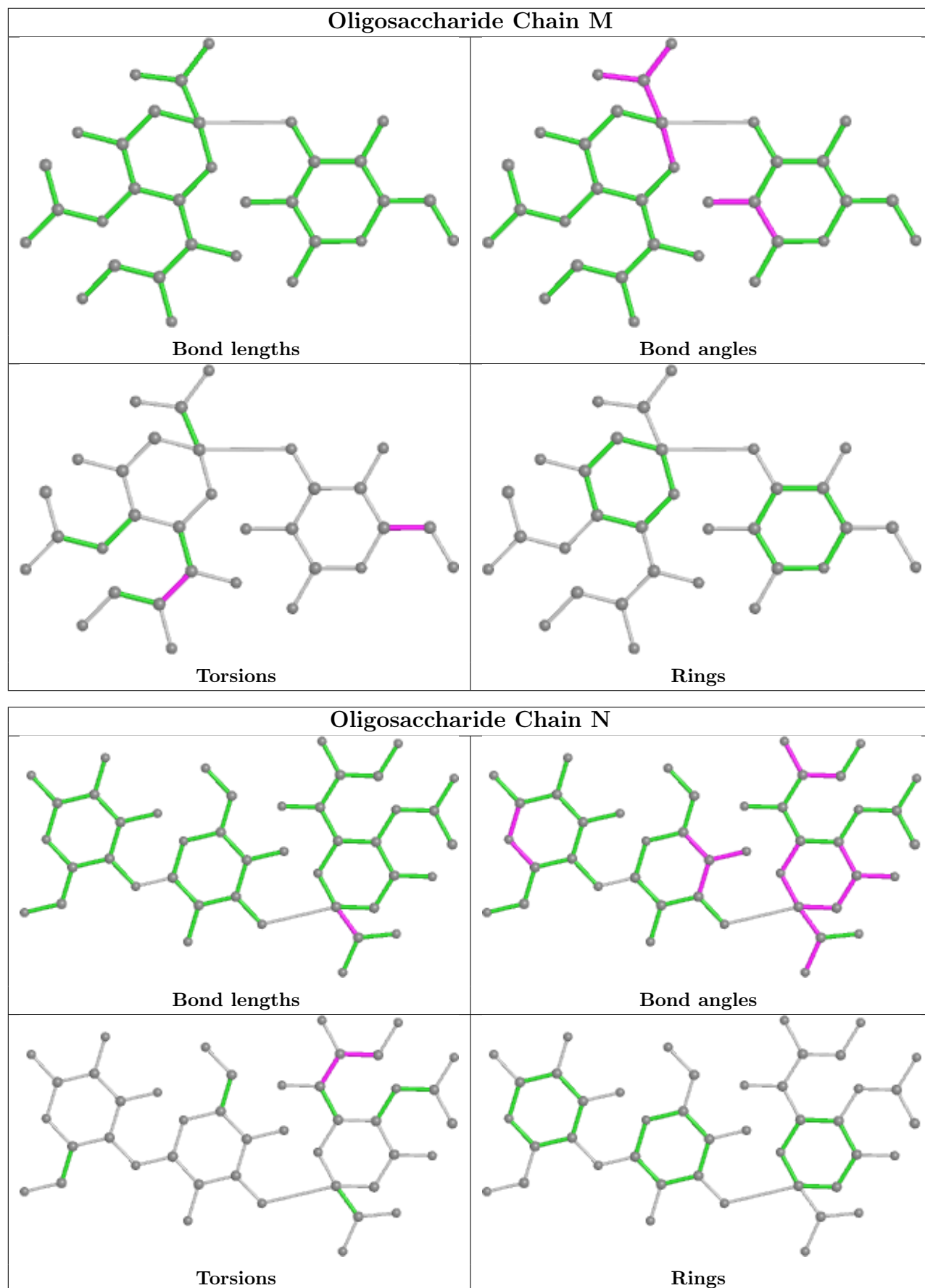
1 monomer is involved in 1 short contact:

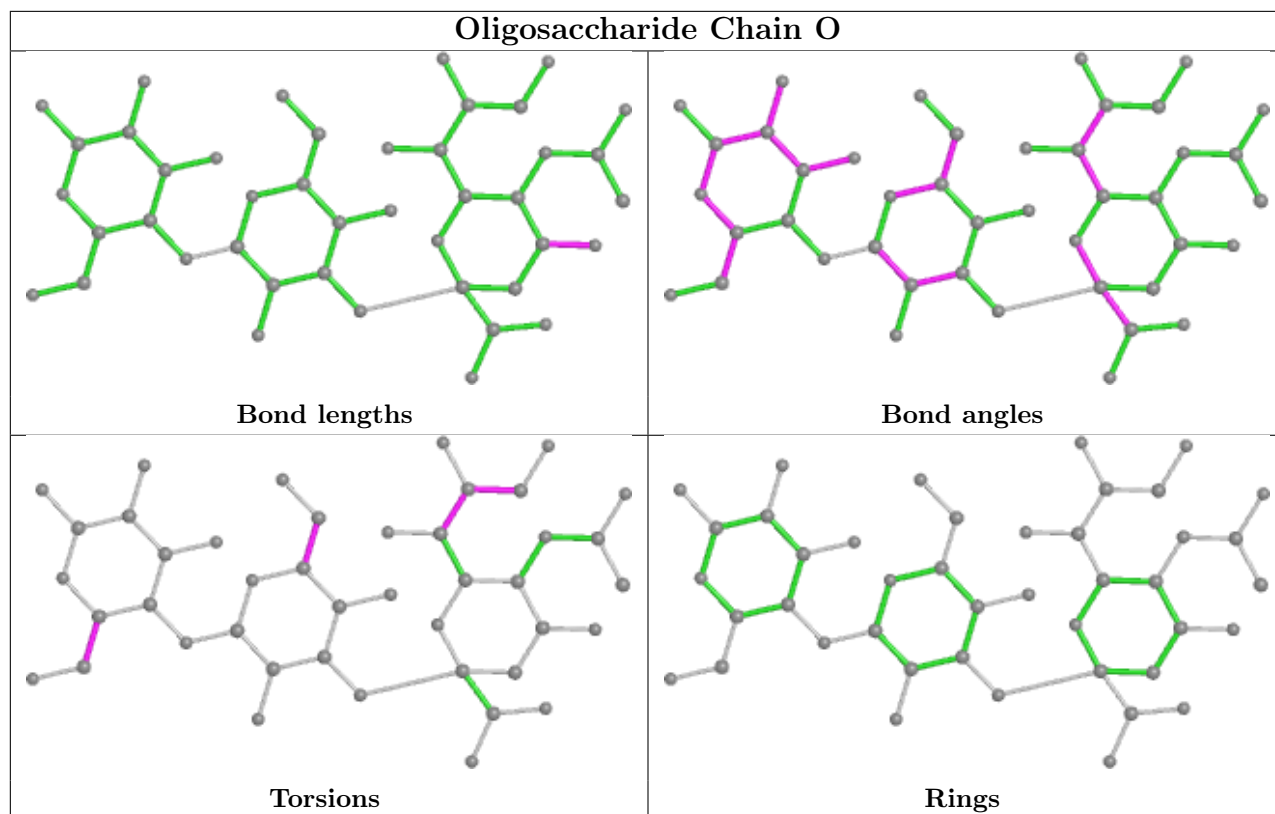
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	O	3	SIA	1	0

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](#)

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	SIA	D	401	-	21,21,21	1.84	6 (28%)	25,31,31	1.56	4 (16%)
4	GOL	C	403	-	5,5,5	0.35	0	5,5,5	1.13	0
4	GOL	H	401	-	5,5,5	0.30	0	5,5,5	0.73	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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	SIA	D	401	-	-	6/20/38/38	0/1/1/1
4	GOL	C	403	-	-	1/4/4/4	-
4	GOL	H	401	-	-	0/4/4/4	-

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	D	401	SIA	O2-C2	4.98	1.46	1.39
5	D	401	SIA	C4-C5	3.19	1.55	1.53
5	D	401	SIA	C7-C6	2.50	1.56	1.53
5	D	401	SIA	C8-C7	2.22	1.57	1.53
5	D	401	SIA	C5-N5	2.19	1.49	1.45

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	D	401	SIA	O1A-C1-C2	-3.51	118.28	123.59
5	D	401	SIA	O8-C8-C9	-3.16	101.73	109.14
5	D	401	SIA	O7-C7-C8	2.53	114.93	108.81
5	D	401	SIA	C4-C5-N5	2.36	115.04	110.38

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	D	401	SIA	C6-C7-C8-C9
5	D	401	SIA	O7-C7-C8-C9
5	D	401	SIA	C6-C7-C8-O8
5	D	401	SIA	O7-C7-C8-O8
5	D	401	SIA	O1A-C1-C2-O6

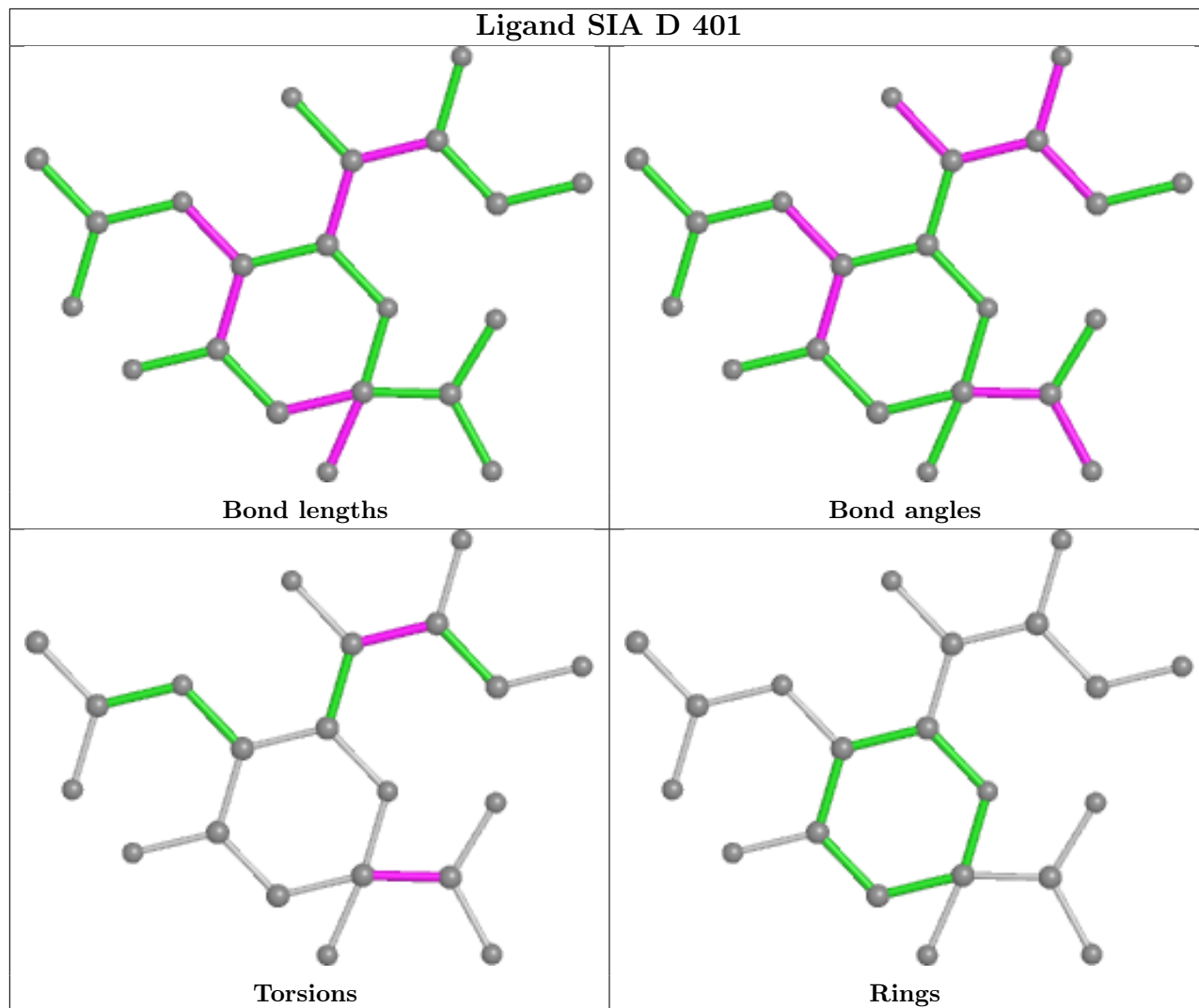
There are no ring outliers.

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 all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and



any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



## 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	272/272 (100%)	-0.01	2 (0%) 87 88	15, 25, 47, 87	0
1	B	267/272 (98%)	0.04	6 (2%) 62 63	17, 25, 49, 79	0
1	C	272/272 (100%)	-0.04	1 (0%) 92 93	15, 23, 44, 81	0
1	D	265/272 (97%)	0.00	5 (1%) 66 68	15, 23, 46, 69	0
1	E	271/272 (99%)	0.08	10 (3%) 41 44	17, 25, 48, 77	0
1	F	271/272 (99%)	0.06	3 (1%) 80 82	17, 25, 41, 72	0
1	G	271/272 (99%)	0.04	0 100 100	17, 27, 47, 69	0
1	H	267/272 (98%)	0.11	11 (4%) 37 39	16, 26, 49, 94	0
1	I	272/272 (100%)	-0.04	2 (0%) 87 88	16, 23, 43, 106	0
1	J	268/272 (98%)	-0.02	3 (1%) 80 82	16, 23, 44, 66	0
All	All	2696/2720 (99%)	0.02	43 (1%) 72 73	15, 25, 47, 106	0

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	303	ASN	5.4
1	D	101	LEU	4.2
1	E	33	ILE	3.9
1	J	102	ASN	3.6
1	E	104	ASP	3.3

### 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](#)

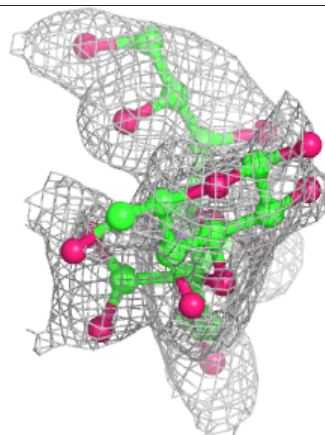
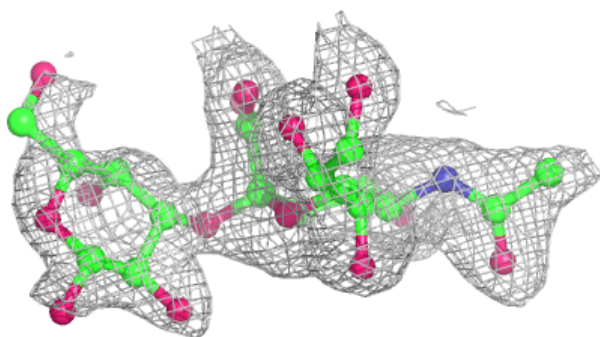
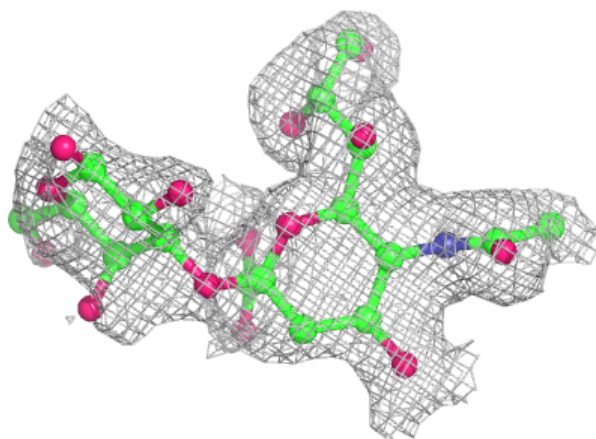
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q<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
3	GAL	N	2	11/12	0.77	0.44	60,64,90,92	0
3	GLC	O	1	12/12	0.78	0.29	41,52,63,64	0
3	SIA	N	3	20/21	0.79	0.22	37,47,57,63	0
3	GLC	N	1	12/12	0.79	0.37	48,73,80,83	0
2	GAL	L	1	12/12	0.80	0.30	54,67,75,93	0
2	GAL	M	1	12/12	0.84	0.28	52,62,67,71	0
2	SIA	K	2	20/21	0.85	0.20	36,49,59,60	0
2	SIA	L	2	20/21	0.86	0.16	31,41,47,49	0
2	GAL	K	1	12/12	0.86	0.34	68,84,90,90	0
2	SIA	M	2	20/21	0.89	0.19	28,39,52,53	0
3	SIA	O	3	20/21	0.93	0.14	25,30,42,43	0
3	GAL	O	2	11/12	0.94	0.24	31,34,48,62	0

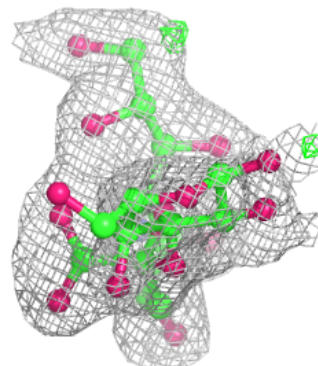
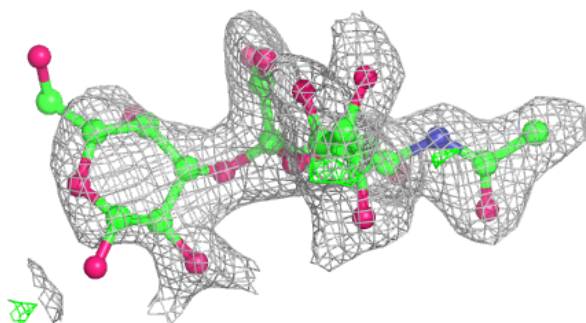
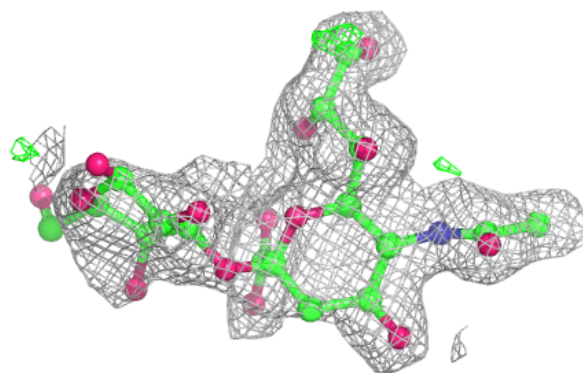
The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

**Electron density around Chain K:**

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

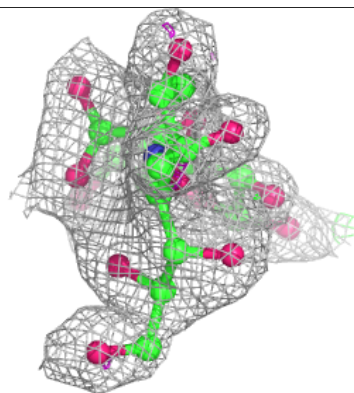
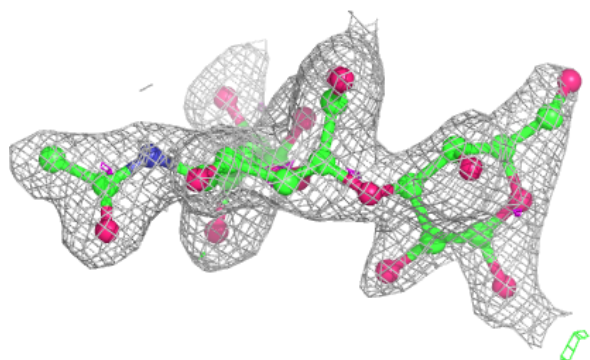
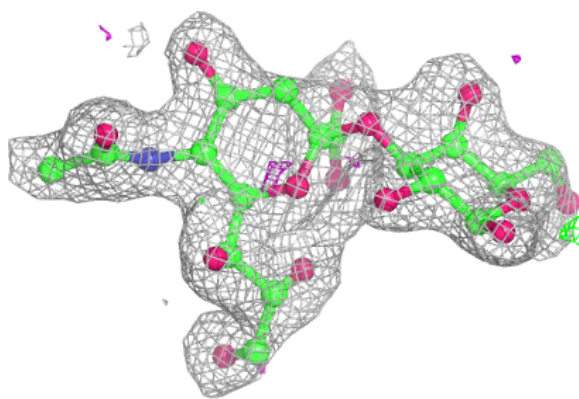
**Electron density around Chain L:**

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

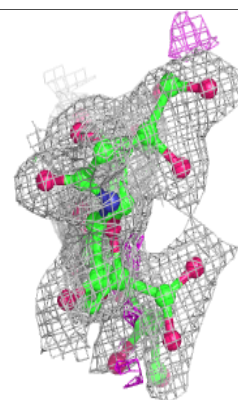
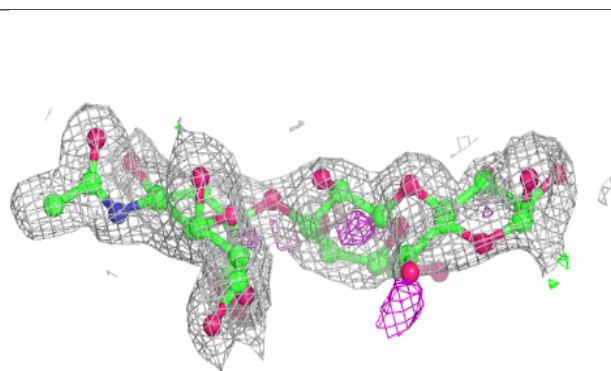
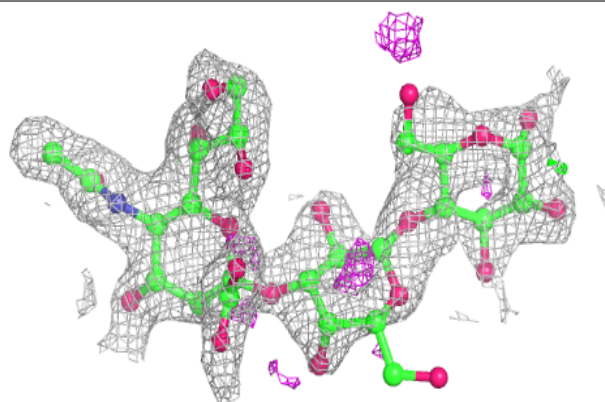


**Electron density around Chain M:**

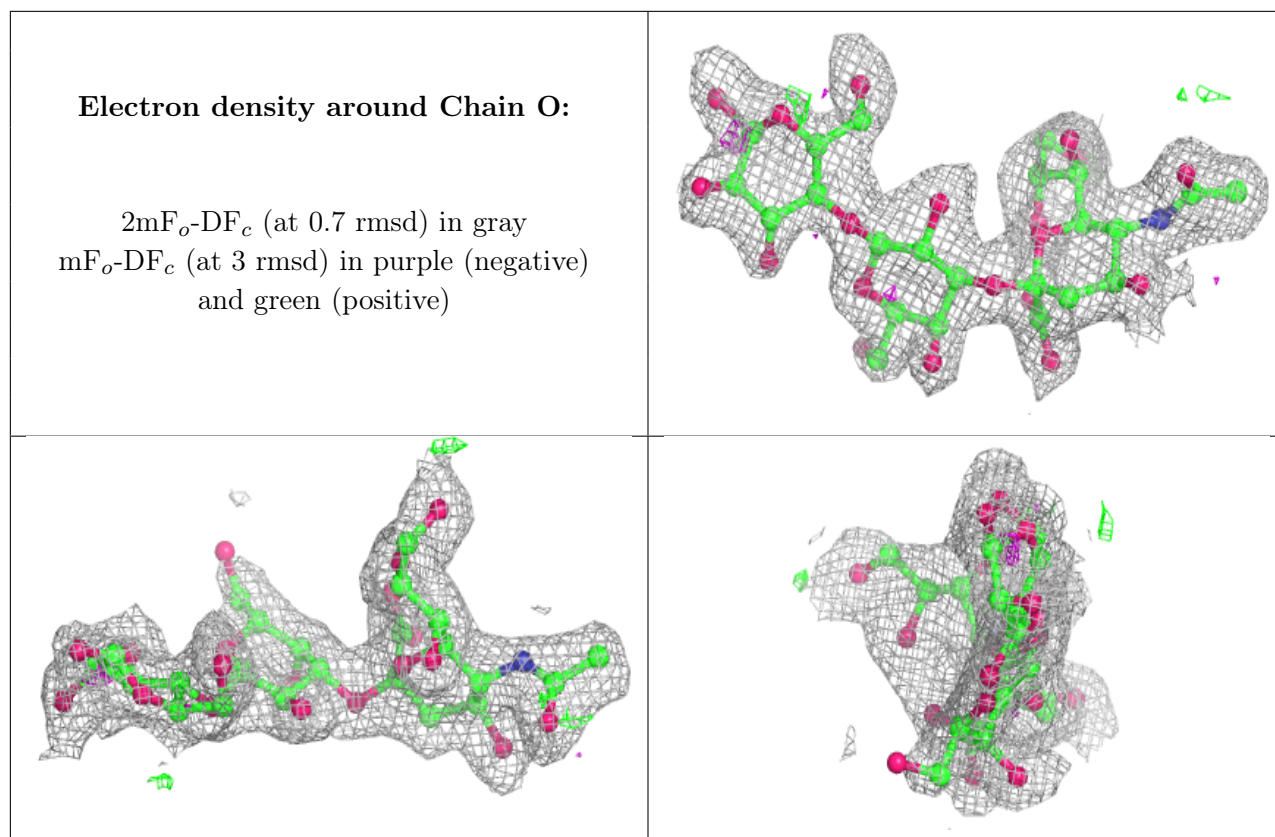
$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)

**Electron density around Chain N:**

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





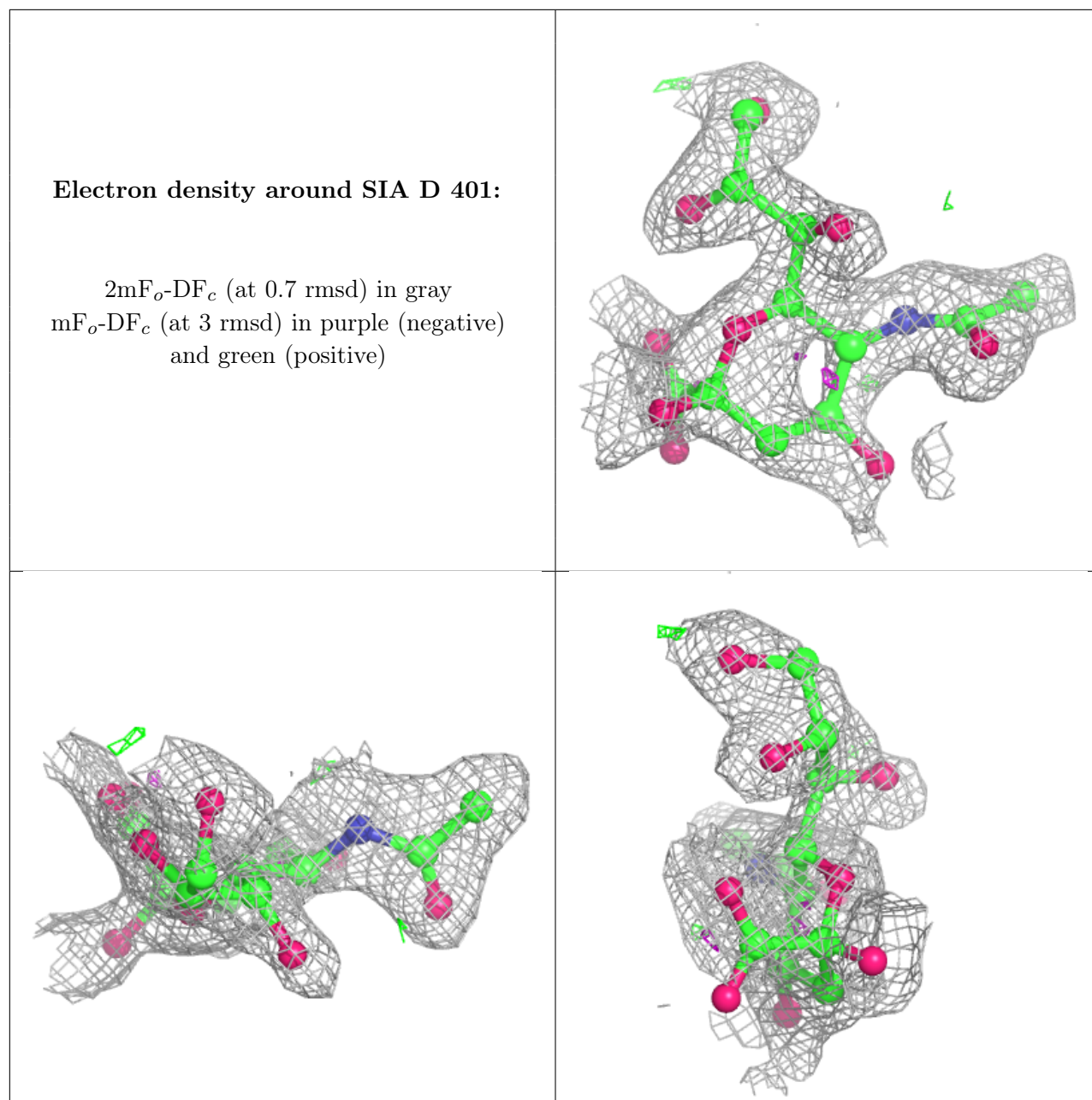


## 6.4 Ligands [i](#)

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	SIA	D	401	21/21	0.77	0.28	33,60,72,75	0
4	GOL	H	401	6/6	0.92	0.14	30,40,46,46	0
4	GOL	C	403	6/6	0.93	0.12	24,32,38,40	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



## 6.5 Other polymers [i](#)

There are no such residues in this entry.