WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2023 – 11:57 am GMT

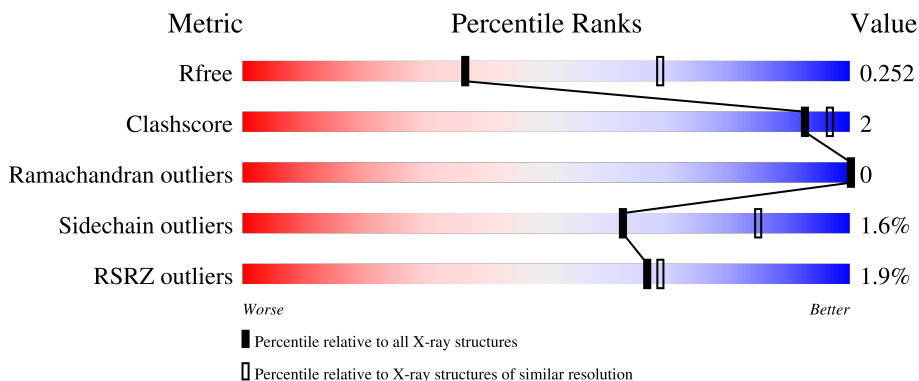|		
PDB ID	:	4CQZ
Title	:	Crystal Structure of H5 (VN1194) Gln196Arg Mutant Haemagglutinin
Authors	:	Liu, J.; Xiong, X.; Xiao, H.; Martin, S.R.; Coombs, P.J.; Collins, P.J.; Vachieri, S.G.; Walker, P.A.; Lin, Y.P.; McCauley, J.W.; Gamblin, S.J.; Skehel, J.J.
Deposited on	:	2014-02-21
Resolution	:	2.70 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 2.70 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		0.252
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		1.6%
RSRZ outliers		1.9%

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}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	326	2% / 92% / 6% ·
2	B	166	2% / 93% / · ·
3	C	3	100%
4	D	2	100%
5	E	2	50% / 50%

WORLDWIDE PDB
PROTEIN DATA BANK

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
4	NAG	D	1	X	-	-	-
5	NAG	E	1	X	-	-	-

2 Entry composition ⓘ

There are 8 unique types of molecules in this entry. The entry contains 4080 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.

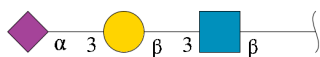| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	321	Total 2552	C 1612	N 441	O 484	S 15	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	192	ARG	GLN	engineered mutation	UNP Q6DQ34
A	325	THR	ARG	conflict	UNP Q6DQ34

- Molecule 2 is a protein called Hemagglutinin HA2.

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

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

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

- Molecule 4 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-3)-2-a cetamido-2-deoxy-beta-D-glucopyranose.

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

- Molecule 5 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.
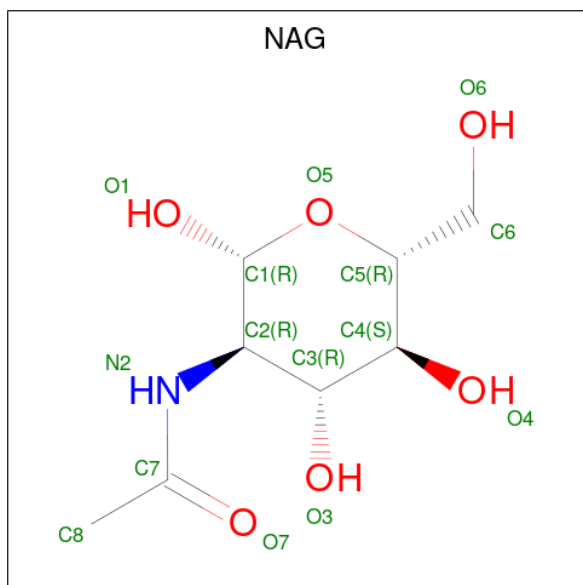

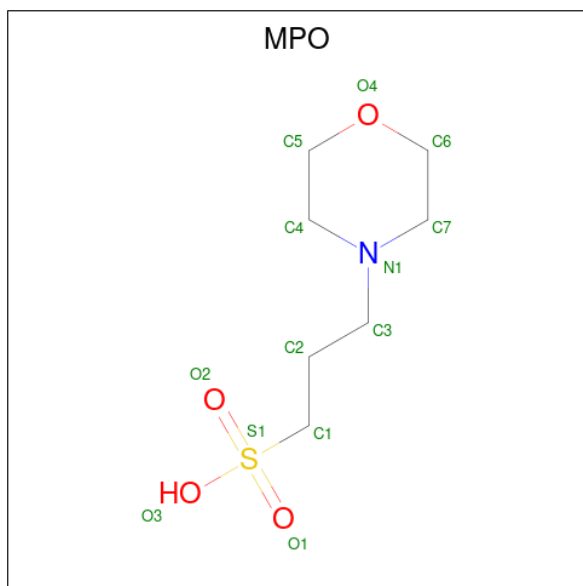| Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
5	E	2	28	16	2	10	0	0	0

- Molecule 6 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
6	A	1	14	8	1	5	0	0

- Molecule 7 is 3[N-MORPHOLINO]PROPANE SULFONIC ACID (three-letter code: MPO)

(formula: $C_7H_{15}NO_4S$).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	B	1	Total	C	N	O	S	0	0
			13	7	1	4	1		

• Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	55	Total	O	0	0
			55	55		
8	B	32	Total	O	0	0
			32	32		

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 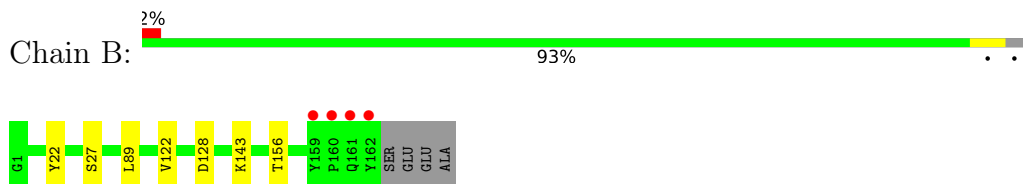graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates 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

Chain A:

2%

92%   6% •

D1 I5 Q15 R53 S124 G130 Y137 W149 K152 S155 I160 N166 T167 N168 R192 N193 K234 P235 E242 S243 N244 F247 I248 A249 Y252 Y305 R310 P321 GLN ARG GLU THR ARG
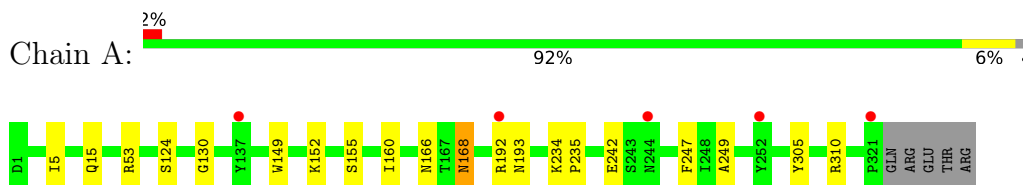
• Molecule 2: Hemagglutinin HA2

Chain B:

2%

93%   • •

G1 Y22 S27 L89 V122 D128 K143 T156 Y159 P160 Q161 Y162 SER GLU GLU ALA

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

Chain C: 100%

NAG1 GAL2 SIA3
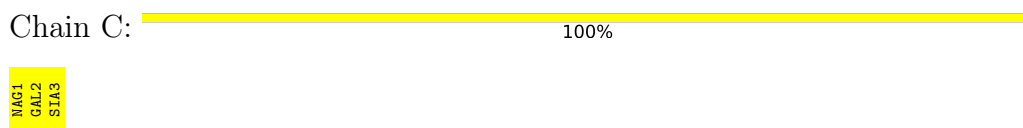
• Molecule 4: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-3)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain D: 100%

NAG1 NAG2
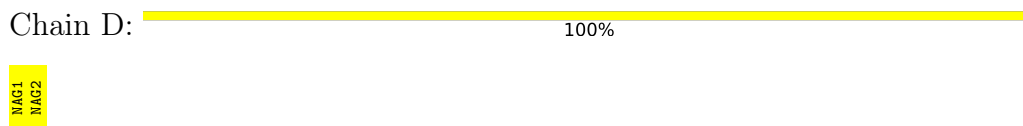
• Molecule 5: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain E: 50%   50%

NAG1 NAG2

4   Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	101.15Å 101.15Å 447.89Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	149.30 – 2.70 68.99 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.4 (149.30-2.70) 99.5 (68.99-2.70)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.80 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.8.0046	Depositor
R, R_{free}	0.217 , 0.252 0.220 , 0.252	Depositor DCC
R_{free} test set	1261 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	60.6	Xtriage
Anisotropy	0.585	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 39.6	EDS
L-test for twinning[2]	$< |L| > = 0.51, < L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4080	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.07% of the height of the origin peak. No significant pseudotranslation is detected.*

---

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

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: GAL, MPO, SIA, 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	0.30	0/2617	0.50	0/3555
2	B	0.32	0/1339	0.46	0/1802
All	All	0.31	0/3956	0.49	0/5357

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 ⓘ

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	2552	0	2500	10	0
2	B	1312	0	1218	4	0
3	C	46	0	40	0	0
4	D	28	0	25	0	0
5	E	28	0	25	0	0
6	A	14	0	13	0	0
7	B	13	0	15	0	0
8	A	55	0	0	1	0
8	B	32	0	0	0	0
All	All	4080	0	3836	12	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 12 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:166:ASN:OD1	1:A:168:ASN:ND2	2.22	0.72
1:A:124:SER:OG	8:A:2035:HOH:O	2.07	0.71
1:A:5:ILE:HD11	2:B:122:VAL:HG21	1.84	0.58
1:A:247:PHE:CE2	1:A:249:ALA:HB2	2.41	0.55
1:A:15:GLN:O	1:A:310:ARG:NH2	2.47	0.46

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	320/326 (98%)	308 (96%)	12 (4%)	0	100	100
2	B	160/166 (96%)	157 (98%)	3 (2%)	0	100	100
All	All	480/492 (98%)	465 (97%)	15 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	290/294 (99%)	286 (99%)	4 (1%)	67	86
2	B	138/141 (98%)	135 (98%)	3 (2%)	52	79
All	All	428/435 (98%)	421 (98%)	7 (2%)	62	85

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	193	ASN
2	B	22	TYR
2	B	143	LYS
2	B	27	SER
1	A	192	ARG

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

Mol	Chain	Res	Type
1	A	138	GLN
1	A	168	ASN
2	B	15	GLN

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 ⓘ

7 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	3	15,15,15	0.44	0	21,21,21	1.12	1 (4%)
3	GAL	C	2	3	11,11,12	0.34	0	15,15,17	1.10	1 (6%)
3	SIA	C	3	3	20,20,21	0.54	0	24,28,31	1.30	4 (16%)
4	NAG	D	1	4,1	14,14,15	0.60	0	17,19,21	1.55	4 (23%)
4	NAG	D	2	4	14,14,15	0.42	0	17,19,21	0.92	1 (5%)
5	NAG	E	1	2,5	14,14,15	0.57	0	17,19,21	1.48	4 (23%)
5	NAG	E	2	5	14,14,15	0.56	0	17,19,21	0.95	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	C	1	3	-	1/6/26/26	0/1/1/1
3	GAL	C	2	3	-	2/2/19/22	0/1/1/1
3	SIA	C	3	3	-	0/18/34/38	0/1/1/1
4	NAG	D	1	4,1	1/1/5/7	2/6/23/26	0/1/1/1
4	NAG	D	2	4	-	0/6/23/26	0/1/1/1
5	NAG	E	1	2,5	1/1/5/7	4/6/23/26	0/1/1/1
5	NAG	E	2	5	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	1	NAG	C1-C2-N2	-3.61	106.55	110.73
4	D	1	NAG	C8-C7-N2	3.24	121.59	116.10
4	D	1	NAG	C2-N2-C7	3.18	127.43	122.90
5	E	1	NAG	C1-O5-C5	3.12	116.42	112.19
3	C	2	GAL	O5-C5-C6	3.11	112.08	107.20

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	D	1	NAG	C1
5	E	1	NAG	C1

5 of 9 torsion outliers are listed below:

| Mol	Chain	Res	Type	Atoms
3	C	2	GAL	O5-C5-C6-O6
3	C	2	GAL	C4-C5-C6-O6
4	D	1	NAG	C8-C7-N2-C2
4	D	1	NAG	O7-C7-N2-C2
5	E	1	NAG	C8-C7-N2-C2

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

Oligosaccharide Chain D

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain E

Bond lengths

Bond angles

Torsions

Rings
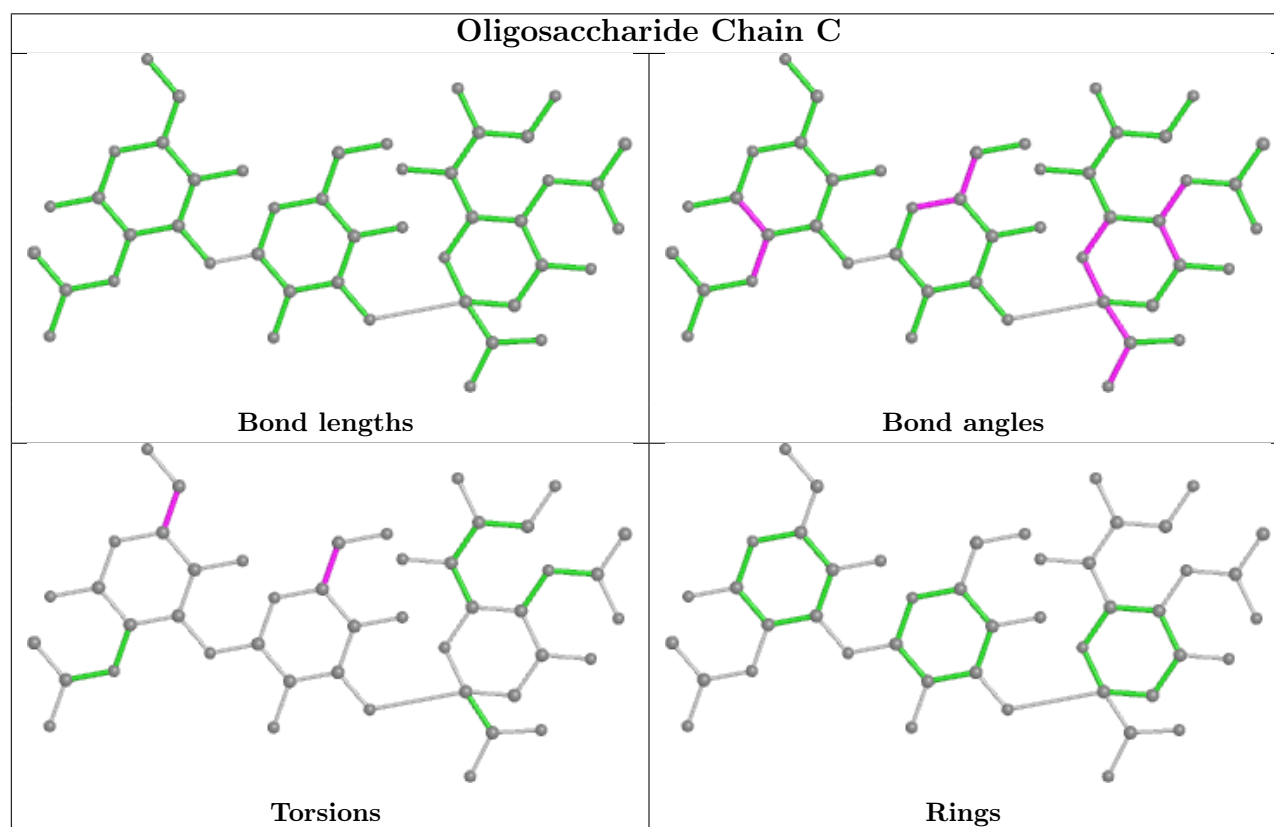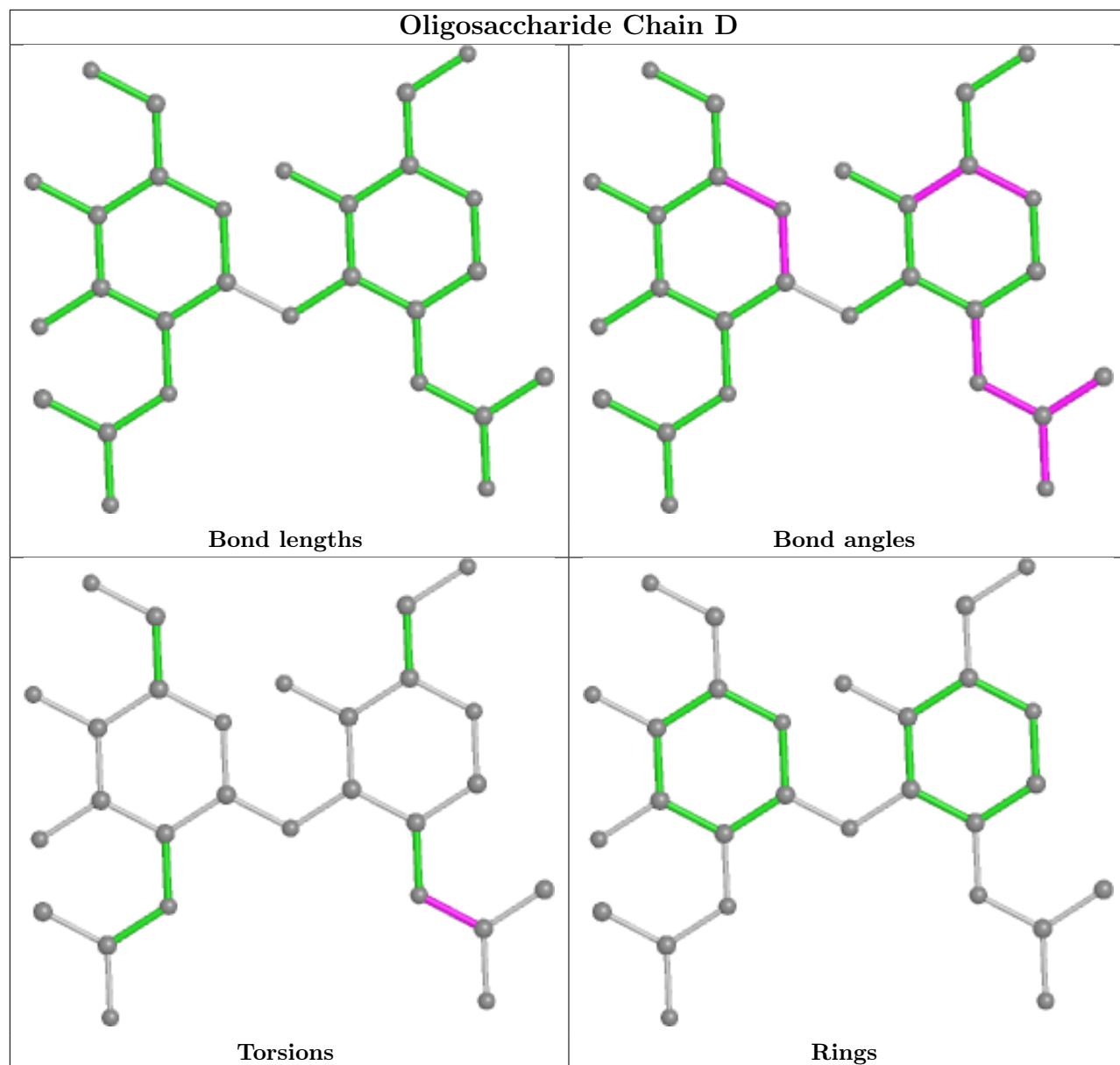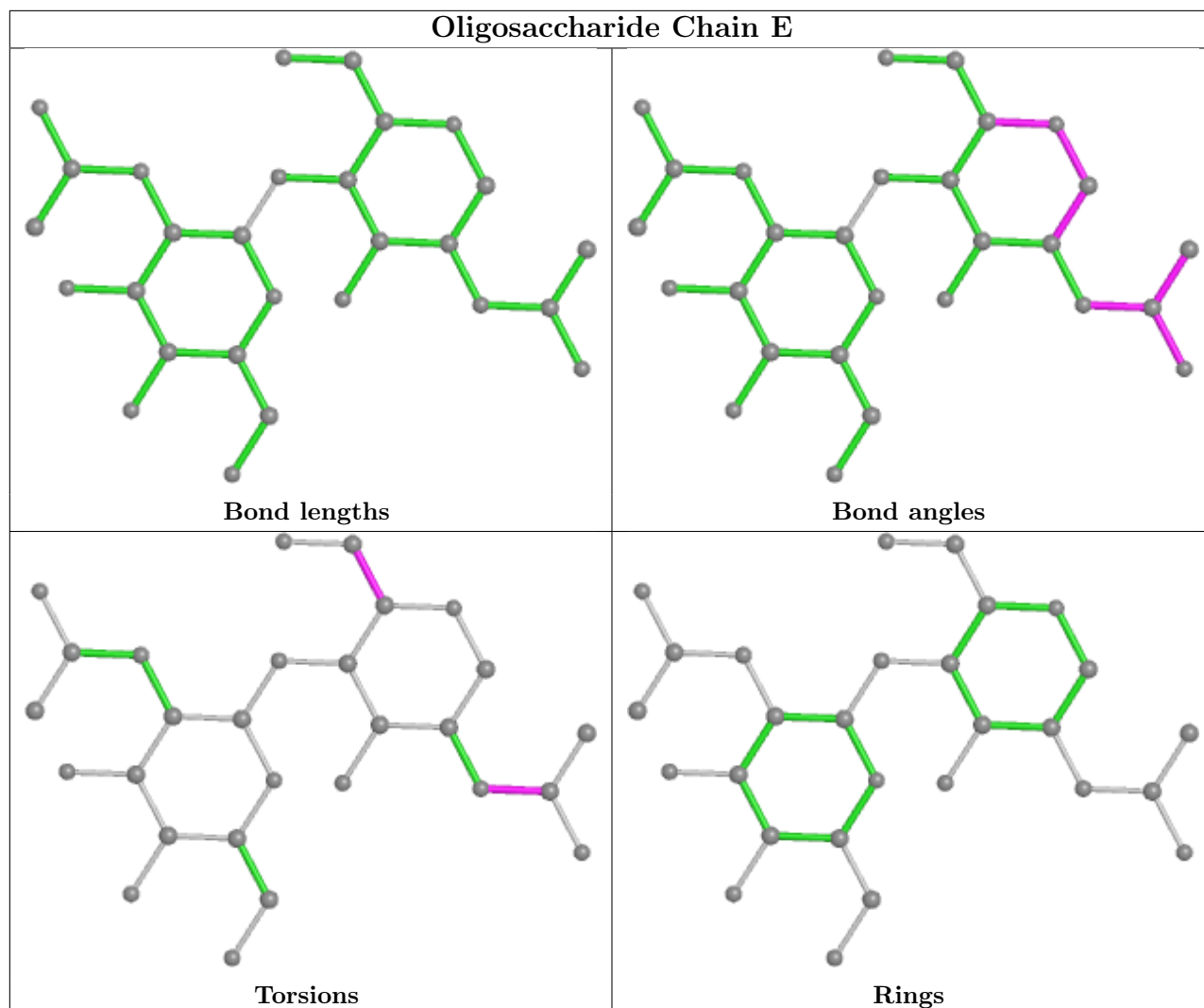
5.6 Ligand geometry ⓘ

2 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
6	NAG	A	1327	1	14,14,15	0.46	0	17,19,21	1.16	2 (11%)
7	MPO	B	1163	-	13,13,13	2.54	2 (15%)	17,17,17	1.26	3 (17%)

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
6	NAG	A	1327	1	-	2/6/23/26	0/1/1/1
7	MPO	B	1163	-	-	5/7/15/15	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	B	1163	MPO	C1-S1	-7.76	1.66	1.77
7	B	1163	MPO	O3-S1	4.73	1.64	1.47

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	B	1163	MPO	O1-S1-C1	2.37	109.77	106.92
6	A	1327	NAG	O5-C5-C6	2.36	110.90	107.20
7	B	1163	MPO	O3-S1-C1	2.31	109.51	105.77
7	B	1163	MPO	O2-S1-C1	2.01	109.33	106.92
6	A	1327	NAG	C3-C4-C5	-2.01	106.66	110.24

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	1327	NAG	O5-C5-C6-O6
7	B	1163	MPO	C2-C1-S1-O3
6	A	1327	NAG	C4-C5-C6-O6
7	B	1163	MPO	C2-C3-N1-C4
7	B	1163	MPO	C2-C3-N1-C7

There are no ring outliers.

No monomer is involved in short contacts.

5.7   Other polymers ⓘ

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	321/326 (98%)	0.14	5 (1%)	72	74	49, 78, 98, 112	0
2	B	162/166 (97%)	0.30	4 (2%)	57	59	44, 65, 103, 173	0
All	All	483/492 (98%)	0.19	9 (1%)	66	69	44, 75, 99, 173	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	160	PRO	5.0
2	B	162	TYR	4.2
2	B	161	GLN	4.0
1	A	321	PRO	3.4
2	B	159	TYR	3.3

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.
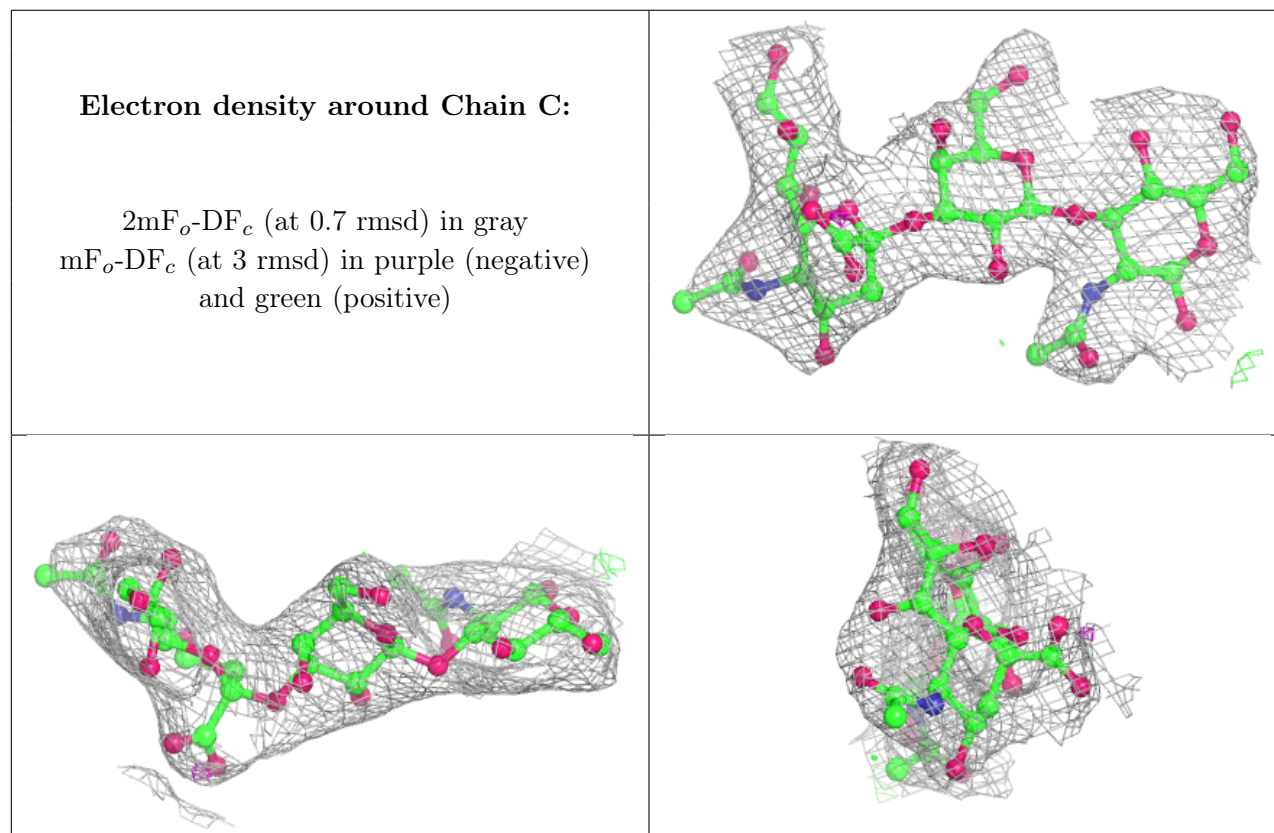
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
3	NAG	C	1	15/15	0.84	0.15	95,100,103,104	0
5	NAG	E	2	14/15	0.85	0.29	114,130,132,133	0
4	NAG	D	1	14/15	0.87	0.21	85,96,108,112	0
5	NAG	E	1	14/15	0.89	0.21	103,112,119,124	0
4	NAG	D	2	14/15	0.89	0.16	120,126,129,132	0

Continued on next page...

Continued from previous page...

| Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
3	GAL	C	2	11/12	0.93	0.13	86,91,95,96	0
3	SIA	C	3	20/21	0.96	0.22	73,81,91,93	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:

$2mF_o$-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 D:

$2mF_o$-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 E:

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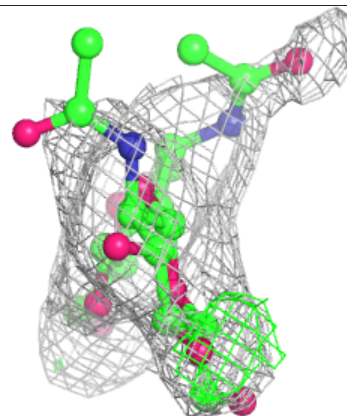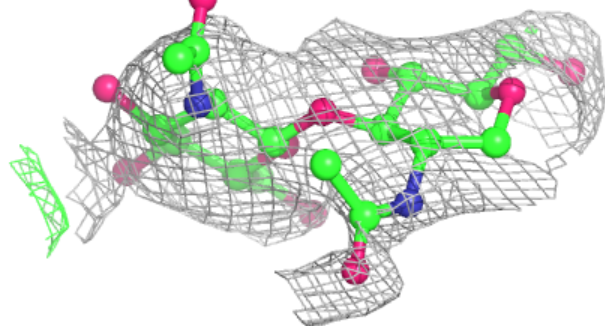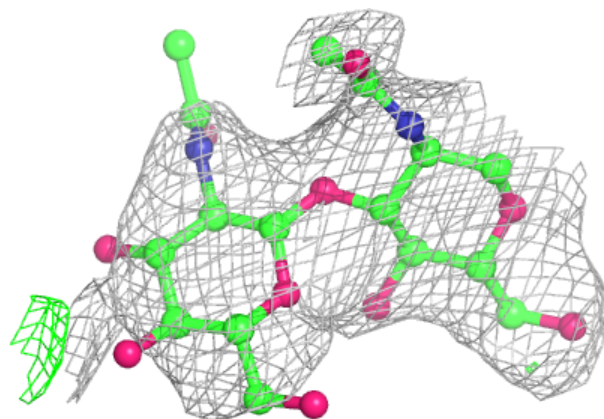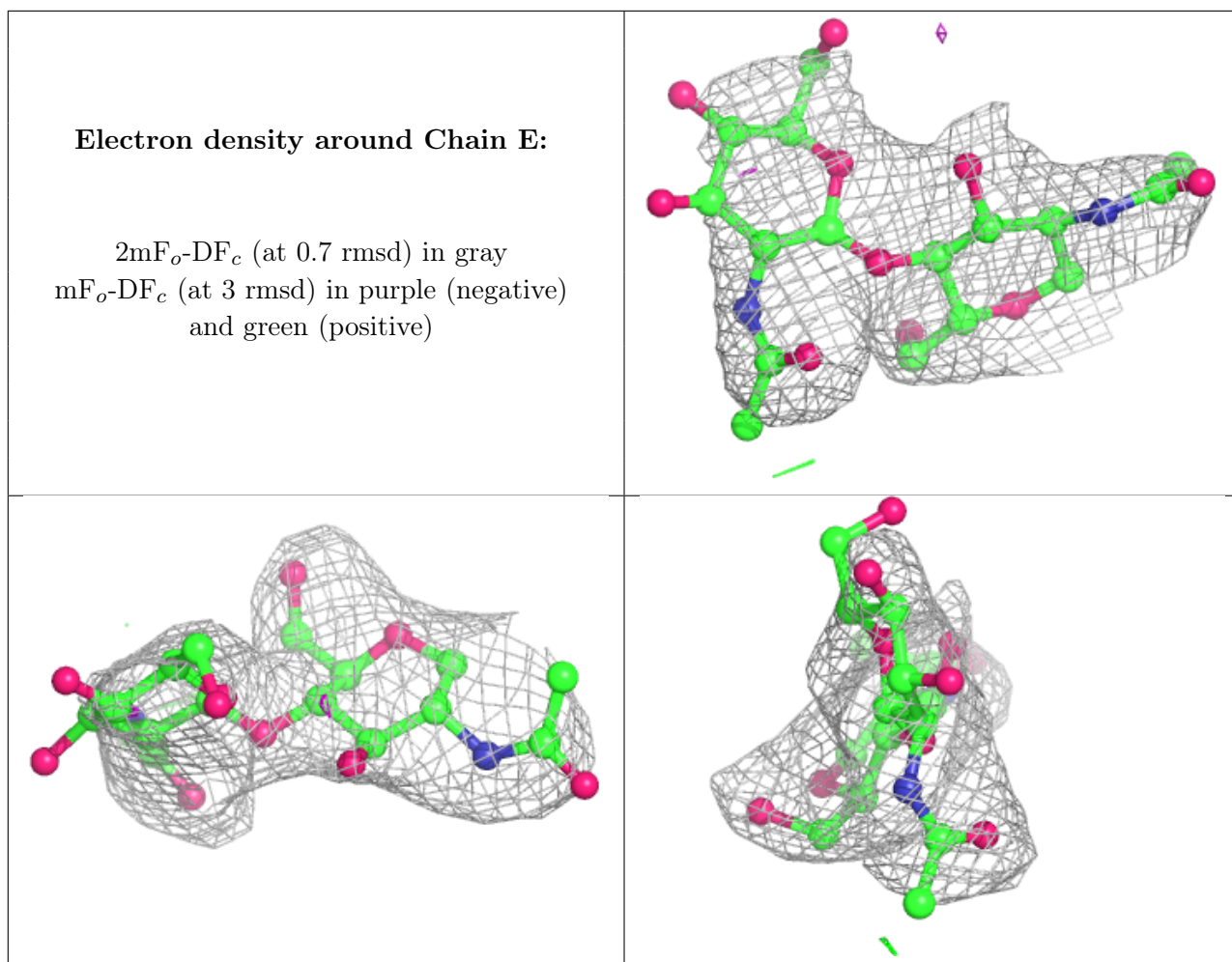
6.4 Ligands (i)

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, 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
7	MPO	B	1163	13/13	0.75	0.40	114,122,132,137	0
6	NAG	A	1327	14/15	0.92	0.11	76,84,86,90	0

6.5 Other polymers (i)

There are no such residues in this entry.