WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 10, 2024 – 05:09 PM EDT

PDB ID	:	8UKV
Title	:	Crystal structure of nanobody/VHH domain of 34E5 in complex with the extracellular region of the epidermal growth factor variant III (EGFRvIII)
Authors	:	Stayrook, S.E.; Ferguson, K.M.; Bagchi, A.
Deposited on	:	2023-10-15
Resolution	:	2.94 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ 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:

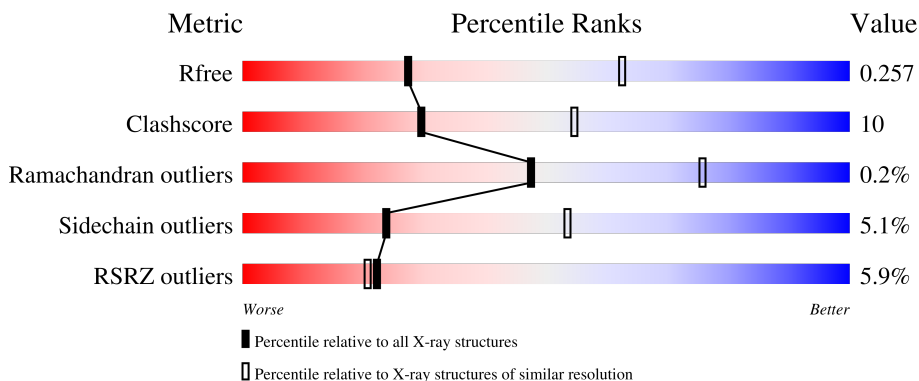|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.94 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		0.257
Clashscore		10
Ramachandran outliers		0.2%
Sidechain outliers		5.1%
RSRZ outliers		5.9%

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	2969 (2.98-2.90)
Clashscore	141614	3218 (2.98-2.90)
Ramachandran outliers	138981	3122 (2.98-2.90)
Sidechain outliers	138945	3124 (2.98-2.90)
RSRZ outliers	127900	2902 (2.98-2.90)

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	357	11% 67% 20% · 10%
1	B	357	2% 73% 14% · 13%
2	C	141	% 60% 23% · 16%
2	D	141	3% 52% 30% · 16%
3	E	2	100%

Continued on next page...

WORLDWIDE
PDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	2	100%
3	G	2	100%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 6739 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 Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	321	Total 2446	C 1511	N 438	O 467	S 30	0	0	0
1	B	310	Total 2352	C 1453	N 424	O 447	S 28	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	273	GLY	VAL	engineered mutation	UNP P00533
A	619	HIS	-	expression tag	UNP P00533
A	620	HIS	-	expression tag	UNP P00533
A	621	HIS	-	expression tag	UNP P00533
A	622	HIS	-	expression tag	UNP P00533
A	623	HIS	-	expression tag	UNP P00533
A	624	HIS	-	expression tag	UNP P00533
B	273	GLY	VAL	engineered mutation	UNP P00533
B	619	HIS	-	expression tag	UNP P00533
B	620	HIS	-	expression tag	UNP P00533
B	621	HIS	-	expression tag	UNP P00533
B	622	HIS	-	expression tag	UNP P00533
B	623	HIS	-	expression tag	UNP P00533
B	624	HIS	-	expression tag	UNP P00533

- Molecule 2 is a protein called Nanobody/VHH domain 34E5.

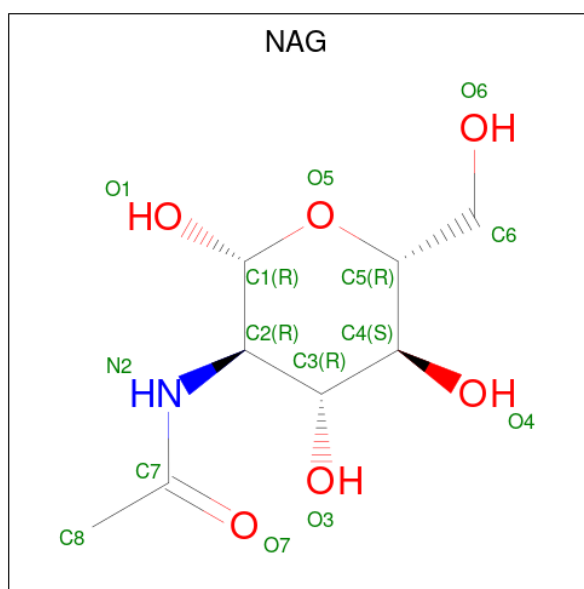| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	119	Total 924	C 574	N 163	O 182	S 5	0	0	0
2	D	118	Total 919	C 571	N 162	O 181	S 5	0	0	0

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

cetamido-2-deoxy-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	E	2	Total	C	N	O	0	0	0
			28	16	2	10			
3	F	2	Total	C	N	O	0	0	0
			28	16	2	10			
3	G	2	Total	C	N	O	0	0	0
			28	16	2	10			

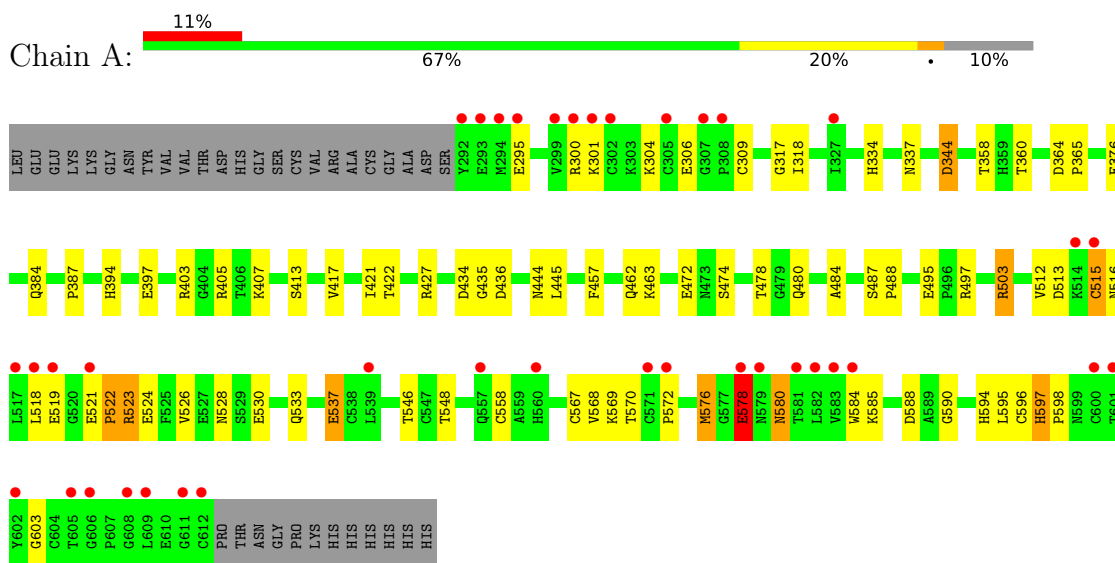- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

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

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.

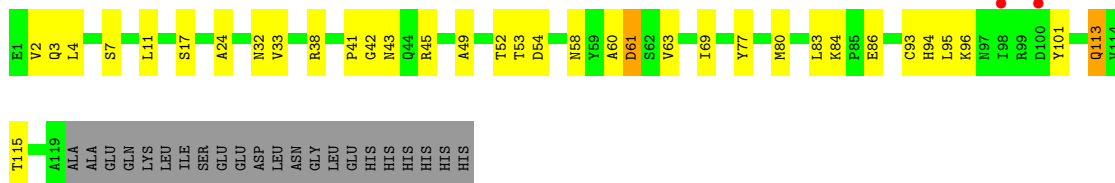• Molecule 1: Epidermal growth factor receptor
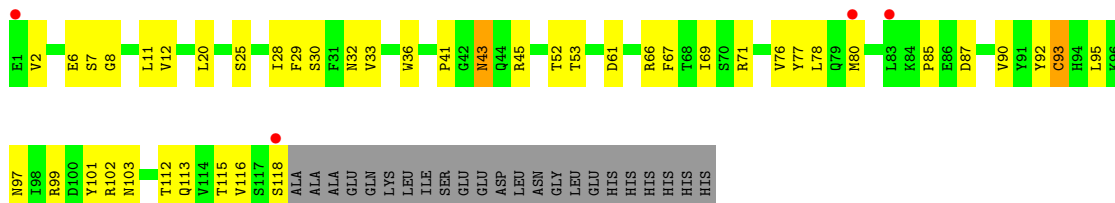

• Molecule 1: Epidermal growth factor receptor
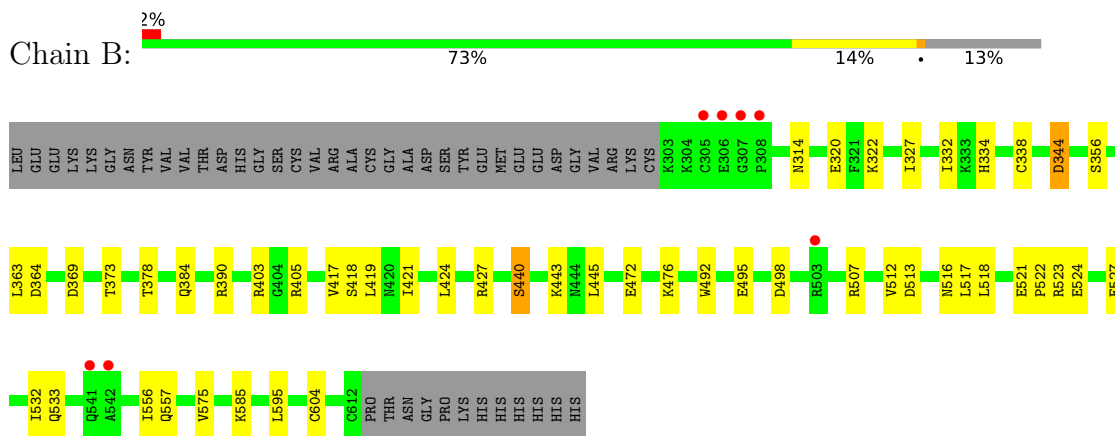
• Molecule 2: Nanobody/VHH domain 34E5

• Molecule 2: Nanobody/VHH domain 34E5

Chain D:

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

Chain E:

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

Chain F:

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

Chain G:

4   Data and refinement statistics   ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	66.29Å 80.36Å 91.84Å 106.92° 103.50° 104.56°	Depositor
Resolution (Å)	46.55 – 2.94 46.58 – 2.94	Depositor EDS
% Data completeness (in resolution range)	98.2 (46.55-2.94) 98.2 (46.58-2.94)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.92 (at 2.96Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, R_{free}	0.215 , 0.252 0.227 , 0.257	Depositor DCC
R_{free} test set	1682 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	76.2	Xtriage
Anisotropy	0.504	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 38.6	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	6739	wwPDB-VP
Average B, all atoms (Å^2)	93.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.66% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

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

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.33	0/2496	0.60	0/3377
1	B	0.31	0/2401	0.59	0/3253
2	C	0.36	0/942	0.64	0/1276
2	D	0.34	0/937	0.66	0/1269
All	All	0.33	0/6776	0.61	0/9175

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	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	578	GLU	Peptide
1	B	523	ARG	Sidechain

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	2446	0	2358	52	0
1	B	2352	0	2262	30	0
2	C	924	0	889	20	0
2	D	919	0	886	27	0
3	E	28	0	25	0	0
3	F	28	0	25	0	0
3	G	28	0	25	0	0
4	A	14	0	13	1	0
All	All	6739	0	6483	126	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 126 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)
2:D:99:ARG:HD2	2:D:103:ASN:HB3	1.28	1.08
1:A:484:ALA:O	1:A:503:ARG:NH1	2.06	0.88
1:B:521:GLU:HB3	1:B:522:PRO:HD3	1.64	0.79
2:D:28:ILE:HA	2:D:97:ASN:HD21	1.47	0.78
1:A:521:GLU:HB3	1:A:522:PRO:HD3	1.67	0.77

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	319/357 (89%)	288 (90%)	30 (9%)	1 (0%)	41	69
1	B	308/357 (86%)	274 (89%)	33 (11%)	1 (0%)	41	69
2	C	117/141 (83%)	111 (95%)	6 (5%)	0	100	100
2	D	116/141 (82%)	108 (93%)	8 (7%)	0	100	100
All	All	860/996 (86%)	781 (91%)	77 (9%)	2 (0%)	47	76

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	443	LYS
1	A	522	PRO

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	276/307 (90%)	259 (94%)	17 (6%)	18	45
1	B	265/307 (86%)	259 (98%)	6 (2%)	50	78
2	C	100/119 (84%)	92 (92%)	8 (8%)	12	32
2	D	100/119 (84%)	93 (93%)	7 (7%)	15	39
All	All	741/852 (87%)	703 (95%)	38 (5%)	24	53

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	61	ASP
2	D	93	CYS
2	C	101	TYR
2	D	25	SER
2	D	113	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	39	GLN
2	D	97	ASN
2	D	113	GLN
1	B	346	HIS
1	B	516	ASN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5   Carbohydrates ⓘ

6 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	E	1	3,1	14,14,15	0.49	0	17,19,21	0.63	0
3	NAG	E	2	3	14,14,15	0.20	0	17,19,21	0.53	0
3	NAG	F	1	3,1	14,14,15	0.54	0	17,19,21	0.62	0
3	NAG	F	2	3	14,14,15	0.27	0	17,19,21	0.47	0
3	NAG	G	1	3,1	14,14,15	0.35	0	17,19,21	0.59	0
3	NAG	G	2	3	14,14,15	0.20	0	17,19,21	0.44	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	E	1	3,1	-	0/6/23/26	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	E	2	3	-	2/6/23/26	0/1/1/1
3	NAG	F	1	3,1	-	0/6/23/26	0/1/1/1
3	NAG	F	2	3	-	2/6/23/26	0/1/1/1
3	NAG	G	1	3,1	-	2/6/23/26	0/1/1/1
3	NAG	G	2	3	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

| Mol	Chain	Res	Type	Atoms
3	G	2	NAG	O5-C5-C6-O6
3	G	1	NAG	O5-C5-C6-O6
3	G	2	NAG	C4-C5-C6-O6
3	G	1	NAG	C4-C5-C6-O6
3	G	2	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 E

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain F

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain G

Bond lengths

Bond angles

Torsions

Rings
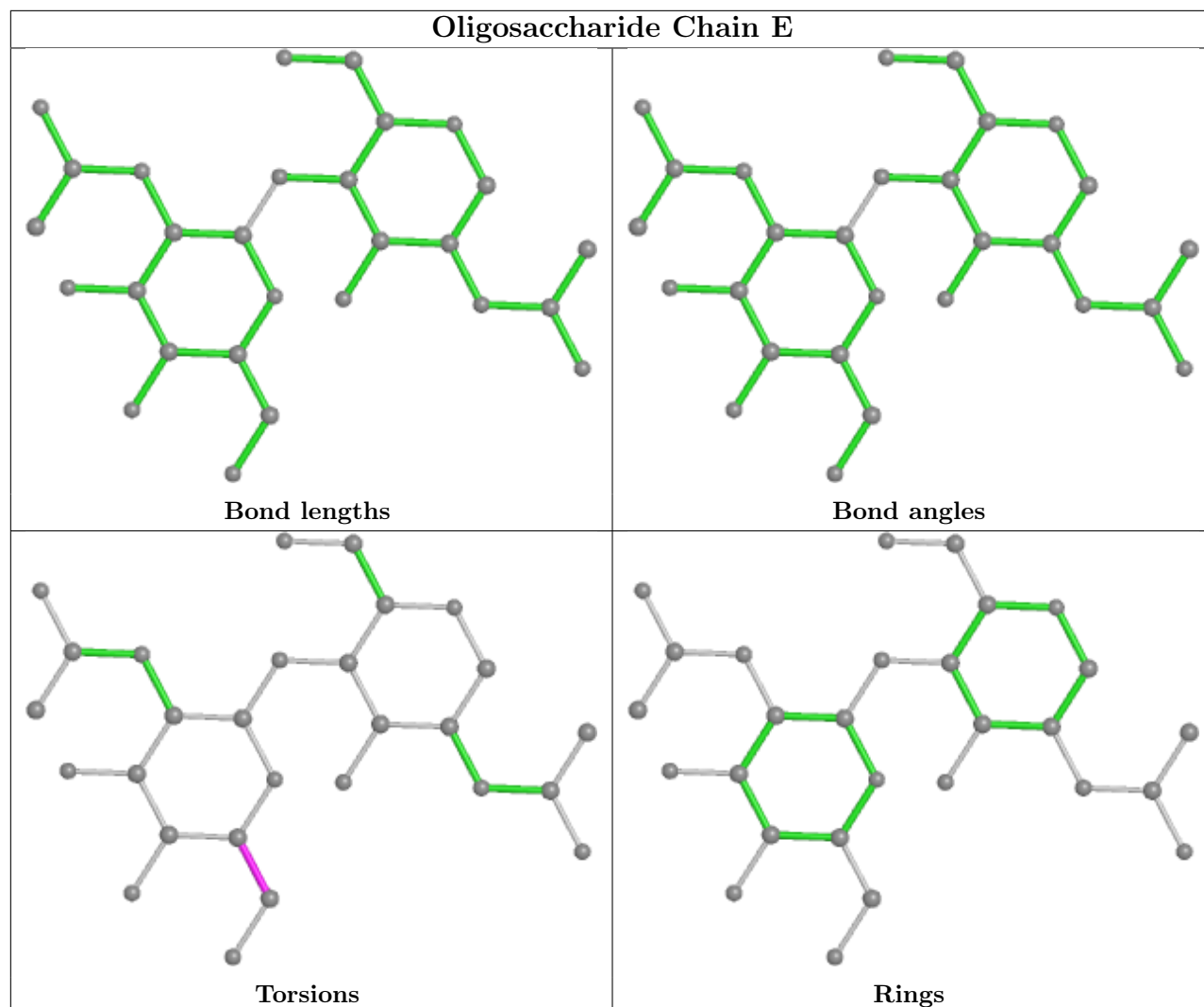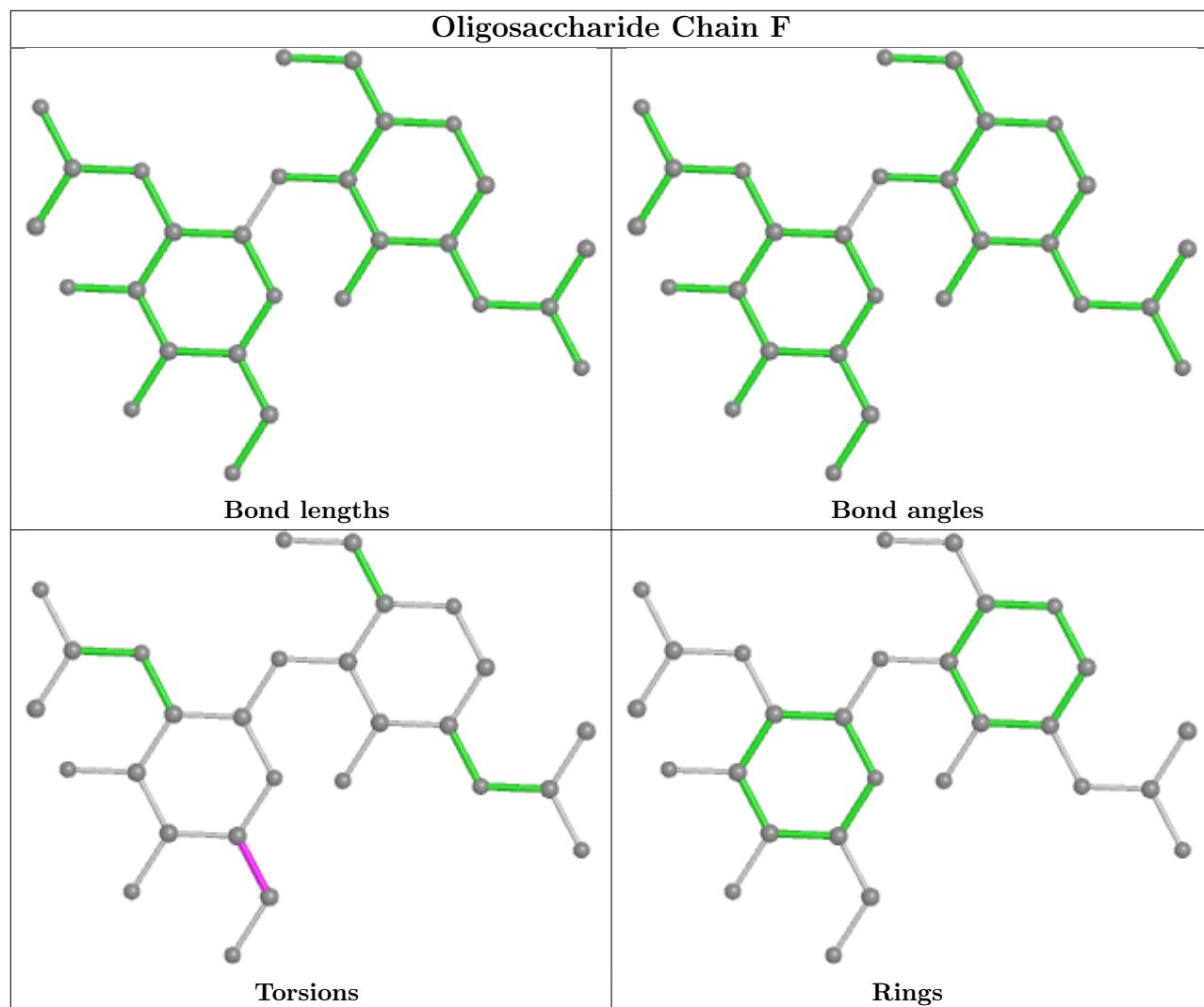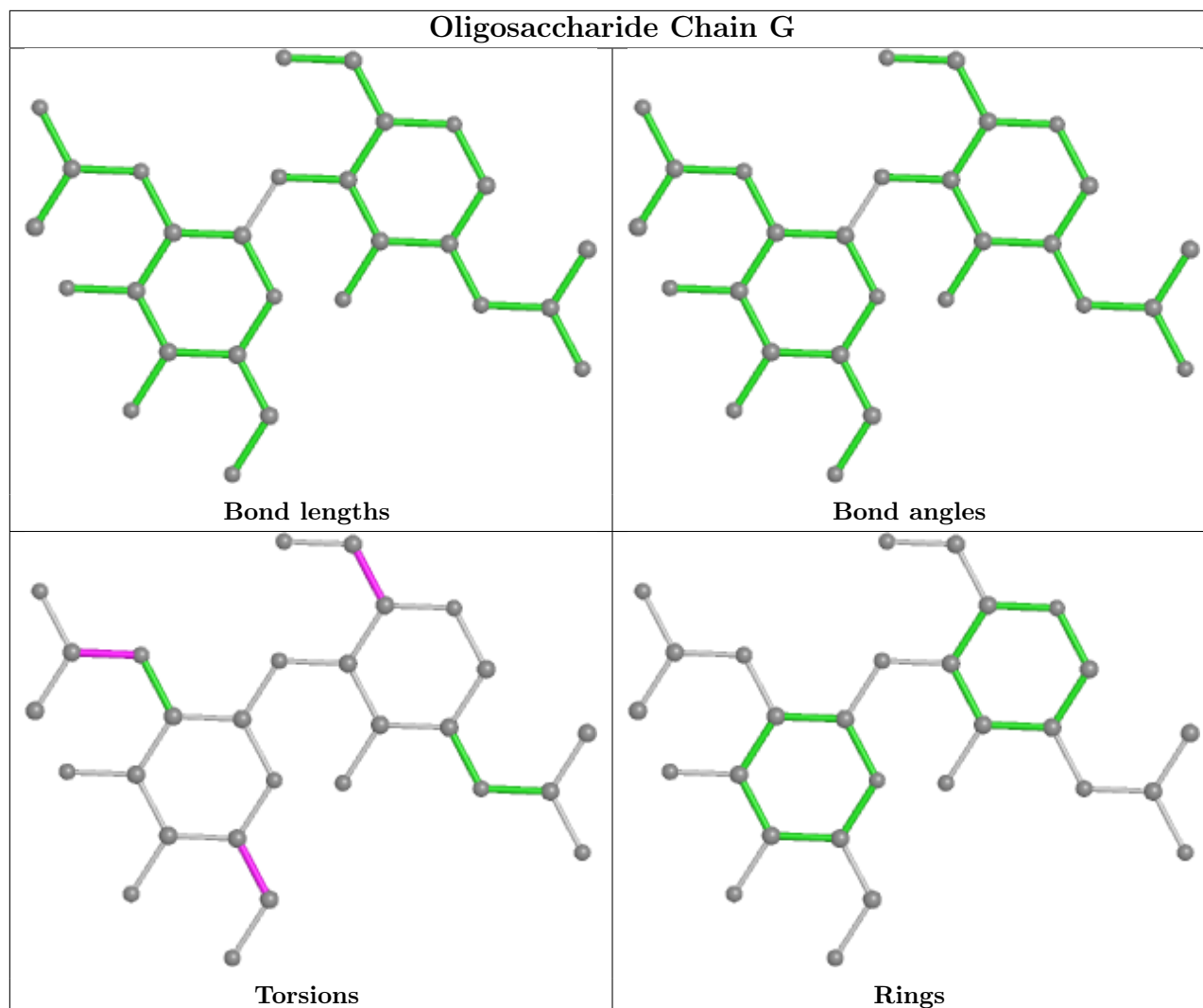
5.6 Ligand geometry ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	NAG	A	701	1	14,14,15	0.29	0	17,19,21	0.57	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
4	NAG	A	701	1	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	701	NAG	O5-C5-C6-O6
4	A	701	NAG	C8-C7-N2-C2
4	A	701	NAG	O7-C7-N2-C2
4	A	701	NAG	C4-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	701	NAG	1	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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	321/357 (89%)	0.59	38 (11%)	4	3	53, 95, 180, 219	0
1	B	310/357 (86%)	0.18	7 (2%)	60	61	47, 82, 137, 190	0
2	C	119/141 (84%)	0.36	2 (1%)	70	71	57, 79, 117, 176	0
2	D	118/141 (83%)	0.29	4 (3%)	45	43	55, 86, 116, 155	0
All	All	868/996 (87%)	0.37	51 (5%)	22	20	47, 87, 161, 219	0

The worst 5 of 51 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	609	LEU	9.6
2	D	1	GLU	6.8
1	A	612	CYS	6.7
1	A	292	TYR	6.3
1	A	294	MET	6.0

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.
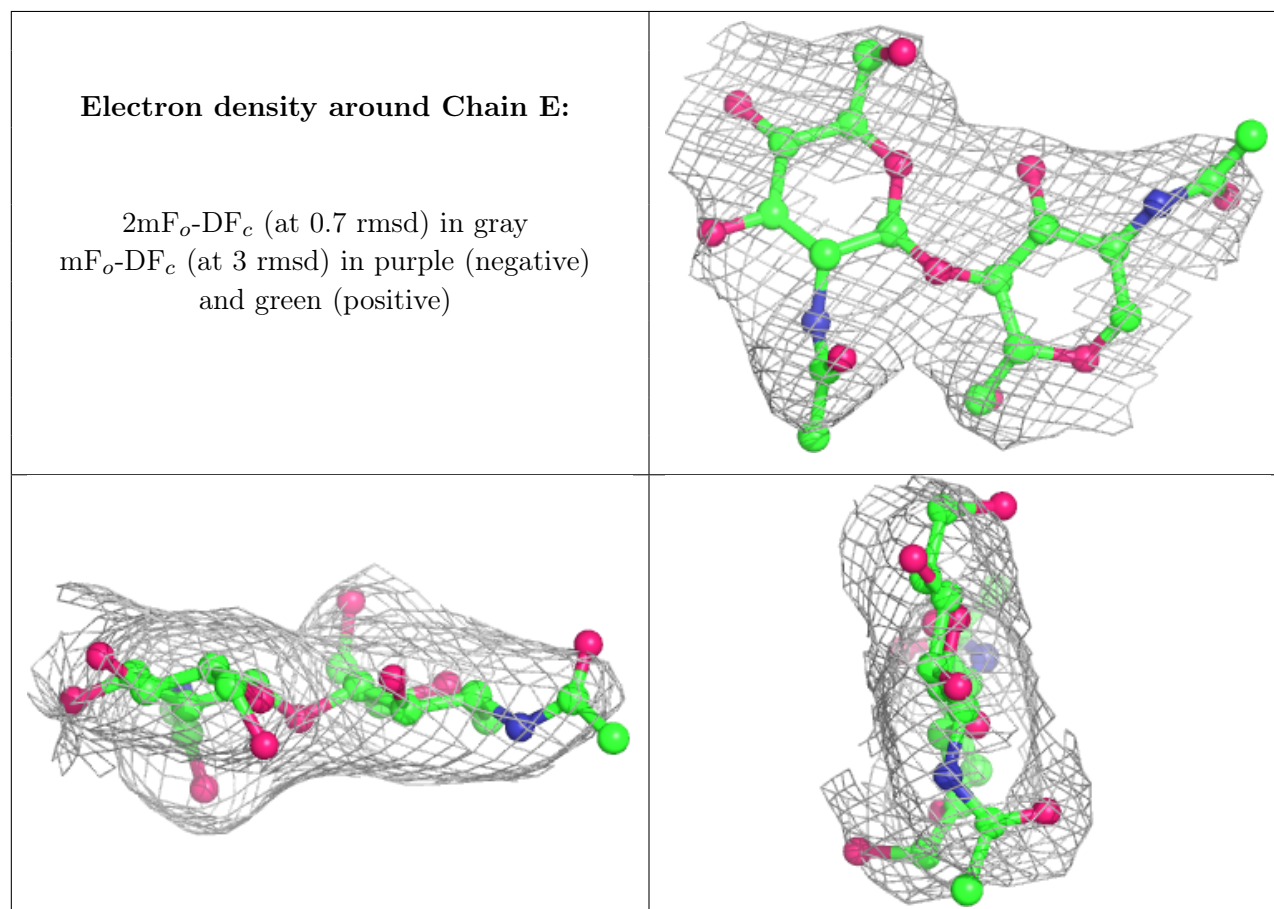
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
3	NAG	F	2	14/15	0.88	0.21	123,127,129,130	0
3	NAG	G	1	14/15	0.91	0.20	117,117,124,127	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
3	NAG	E	2	14/15	0.94	0.16	121,127,129,130	0
3	NAG	G	2	14/15	0.94	0.32	129,131,132,132	0
3	NAG	F	1	14/15	0.95	0.19	117,117,117,123	0
3	NAG	E	1	14/15	0.97	0.23	117,117,117,123	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.
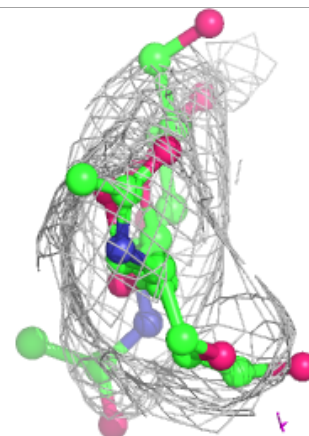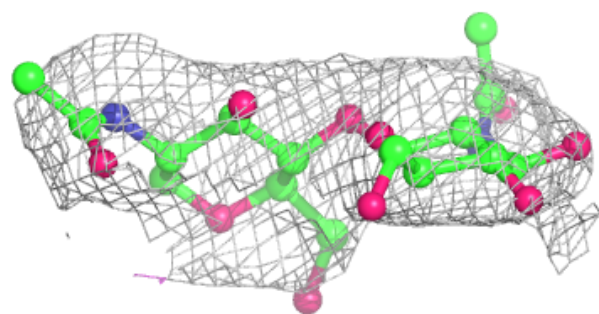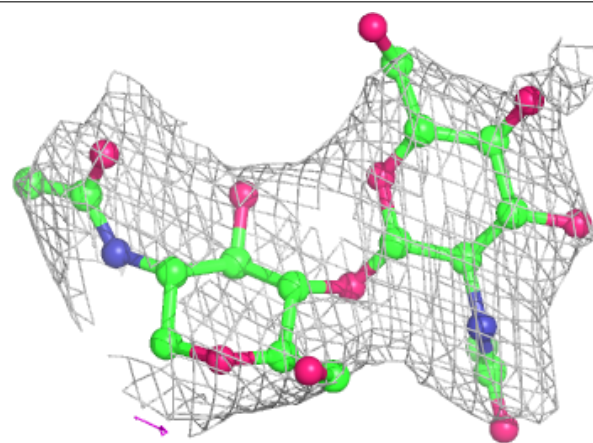

Electron density around Chain E:

$2mF_o$-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 F:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)
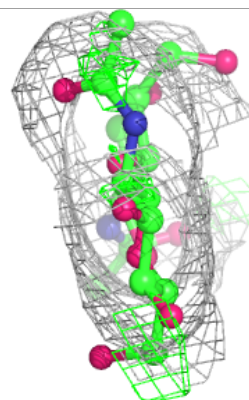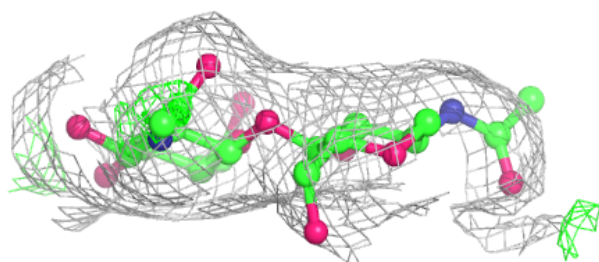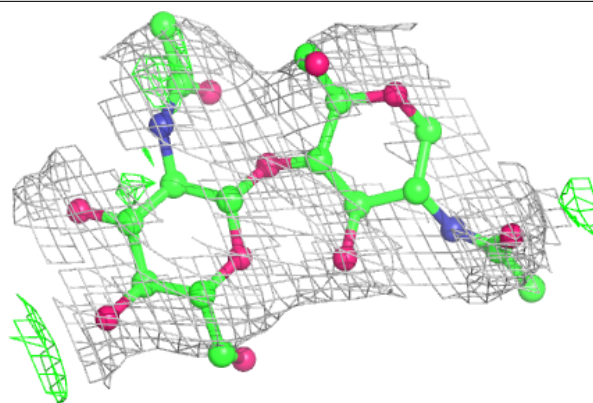
Electron density around Chain G:

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

6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	NAG	A	701	14/15	0.84	0.19	120,124,129,129	0

6.5 Other polymers ⓘ

There are no such residues in this entry.