wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 21, 2023 – 05:10 PM JST

PDB ID	:	7W47
Title	:	Crystal structure of the gastric proton pump complexed with tegoprazan
Authors	:	Abe, K.; Tanaka, S.; Morita, M.; Yamagishi, T.
Deposited on	:	2021-11-26
Resolution	:	3.00 Å(reported)

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
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 3.00 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.
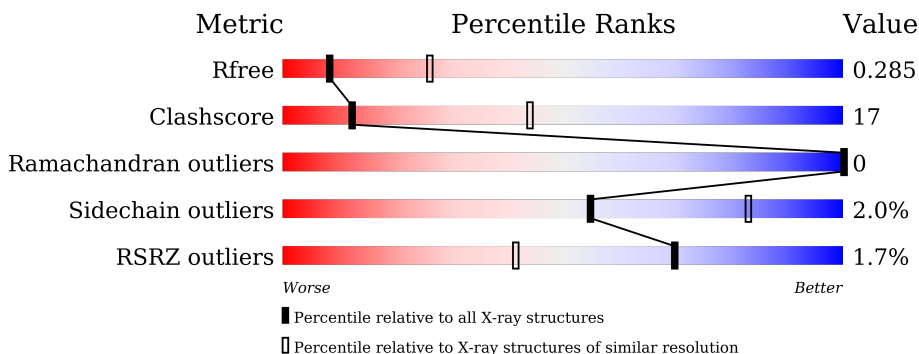
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	1034	64% 31% • 5%
2	B	290	60% 29% • 10%

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
6	NAG	B	302	-	-	-	X
6	NAG	B	303	-	-	-	X

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 9904 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 Potassium-transporting ATPase alpha chain 1.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	A	987	Total 7663	Be 1	C 4890	F 3	N 1292	O 1423	S 54	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	220	CYS	ARG	engineered mutation	UNP P19156
A	593	CYS	SER	engineered mutation	UNP P19156
A	1005	SER	GLY	engineered mutation	UNP P19156

- Molecule 2 is a protein called Potassium-transporting ATPase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	260	Total 2084	C 1353	N 346	O 374	S 11	0	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Mg 1	0	0

- Molecule 4 is RUBIDIUM ION (three-letter code: RB) (formula: Rb) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total 3	Rb 3	0	0

- Molecule 5 is Tegoprazan (three-letter code: 8BN) (formula: $C_{20}H_{19}F_2N_3O_3$) (labeled as "Ligand of Interest" by depositor).

8BN
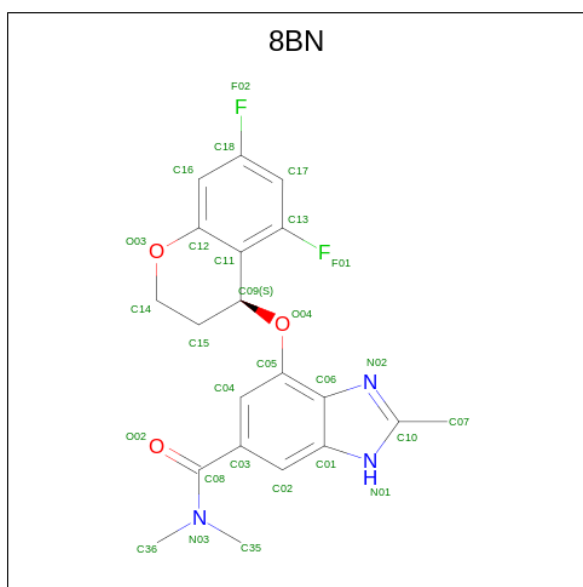
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	F	N	O		
5	A	1	28	20	2	3	3	0	0

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

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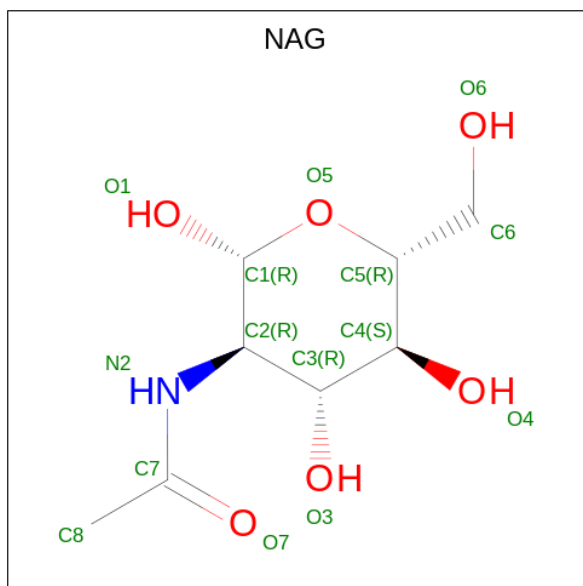
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
6	B	1	14	8	1	5	0	0
6	B	1	Total	C	N	O	0	0
			14	8	1	5		

Continued from previous page...

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

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	63	Total 63	O 63	0	0
7	B	20	Total 20	O 20	0	0

3 Residue-property plots ⓘ

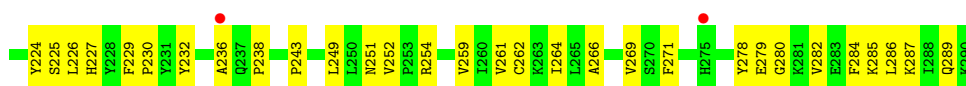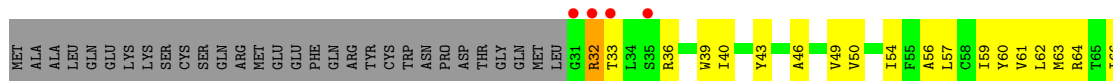These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Potassium-transporting ATPase alpha chain 1

G1024 S1025 W1026 W1027 D1028 Q1029 E1030 L1031 Y1032 Y1033
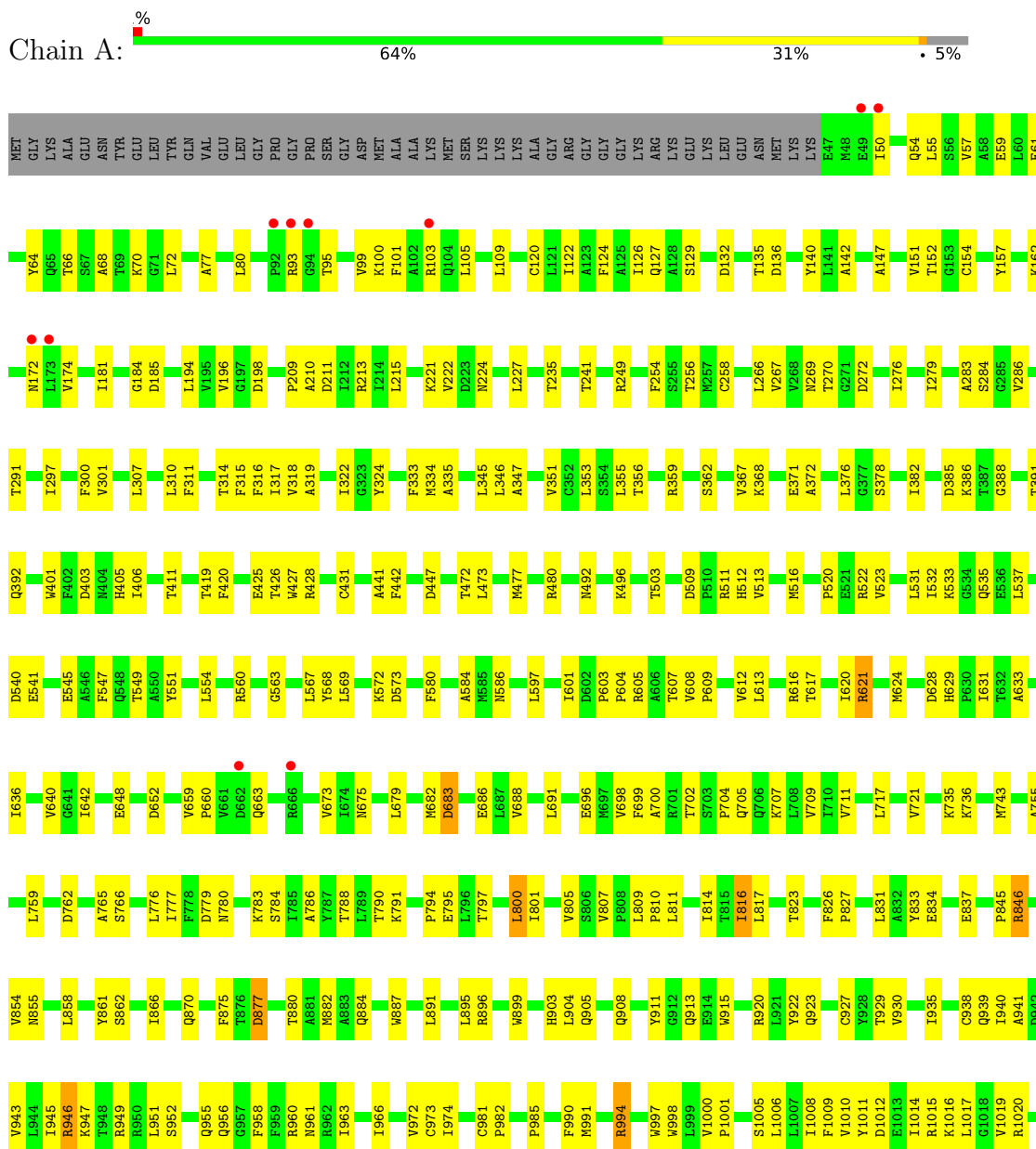
• Molecule 2: Potassium-transporting ATPase subunit beta

Chain B:

60%   29%   • 10%

2%

MET ALA ALA LEU GLN GLU LYS LYS SER CYS SER GLN ARG MET GLU GLU PHE GLN ARG TYR CYS TRP ASN PRO ASP THR GLY GLN MET LEU G31 R32 T33 L34 S35 R36 W39 I40 Y43 A46 V49 V50 I54 F55 A56 L57 C58 I59 Y60 V61 L62 M63 R64 T65 I66

P71 D72 L77 P80 G81 V82 T83 L84 K92 S97 R115 Q125 S128 C131 T132 S133 E134 F138 Q139 E140 S141 F142 L143 K149 F150 S151 C152 K153 F154 K176 I186 V187 K188 F189 L190 P191 A196 P197 A202 F203 R208 L213 Q214

Y224 S225 L226 H227 Y228 F229 P230 Y231 Y232 A236 Q237 P238 P243 L249 L250 N251 V252 P253 R254 V259 I260 V261 C262 K263 I264 L265 A266 V269 S270 F271 H275 Y278 E279 G280 K281 V282 E283 F284 K285 L286 K287 I288 Q289 K290

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	104.99Å 104.99Å 367.79Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.26 – 3.00 48.26 – 2.88	Depositor EDS
% Data completeness (in resolution range)	84.8 (48.26-3.00) 76.3 (48.26-2.88)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.50 (at 2.91Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.222 , 0.285 0.223 , 0.285	Depositor DCC
R_{free} test set	2028 reflections (4.90%)	wwPDB-VP
Wilson B-factor (Å^2)	67.3	Xtriage
Anisotropy	0.017	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 39.6	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.045 for -h,-k,l	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	9904	wwPDB-VP
Average B, all atoms (Å^2)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.78% 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: 8BN, BFD, NAG, RB, MG

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	0.52	1/7808 (0.0%)	0.72	2/10604 (0.0%)
2	B	0.49	0/2151	0.70	1/2926 (0.0%)
All	All	0.51	1/9959 (0.0%)	0.71	3/13530 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	973	CYS	CB-SG	-8.46	1.67	1.82

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	816	ILE	CG1-CB-CG2	-6.01	98.18	111.40
2	B	32	ARG	NE-CZ-NH2	-5.47	117.56	120.30
1	A	800	LEU	CB-CG-CD2	-5.16	102.23	111.00

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	7663	0	7683	263	0
2	B	2084	0	2025	79	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	1	0	0	0	0
4	A	3	0	0	0	0
5	A	28	0	0	1	0
6	B	42	0	39	1	0
7	A	63	0	0	9	0
7	B	20	0	0	5	0
All	All	9904	0	9747	330	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 17.

The worst 5 of 330 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:B:128:SER:HB2	2:B:153:LYS:HG3	1.44	0.99
2:B:139:GLN:HE22	2:B:150:PHE:HB2	1.27	0.96
1:A:985:PRO:HG3	1:A:991:MET:HG2	1.56	0.86
1:A:929:THR:HG21	1:A:991:MET:H	1.43	0.82
2:B:214:GLN:HG2	2:B:251:ASN:HB2	1.61	0.80

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	984/1034 (95%)	935 (95%)	49 (5%)	0	100	100
2	B	258/290 (89%)	239 (93%)	19 (7%)	0	100	100
All	All	1242/1324 (94%)	1174 (94%)	68 (6%)	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	833/869 (96%)	818 (98%)	15 (2%)	59	85
2	B	227/254 (89%)	221 (97%)	6 (3%)	46	78
All	All	1060/1123 (94%)	1039 (98%)	21 (2%)	55	83

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	994	ARG
2	B	203	PHE
2	B	254	ARG
2	B	208	ARG
2	B	64	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	923	GLN
2	B	111	HIS
2	B	139	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue 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	$\#|Z| > 2$
1	BFD	A	385	3,1	8,11,12	5.28	3 (37%)	3,15,17	1.07	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
1	BFD	A	385	3,1	-	2/5/11/13	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	385	BFD	F3-BE	9.08	1.76	1.54
1	A	385	BFD	F2-BE	8.36	1.74	1.54
1	A	385	BFD	F1-BE	8.19	1.74	1.54

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	385	BFD	CA-CB-CG-OD1
1	A	385	BFD	CA-CB-CG-OD2

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 for Mogul analysis.

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	B	302	2	14,14,15	0.28	0	17,19,21	0.52	0
5	8BN	A	1105	-	28,31,31	2.56	4 (14%)	29,46,46	2.54	10 (34%)
6	NAG	B	303	2	14,14,15	0.55	0	17,19,21	0.90	0
6	NAG	B	301	2	14,14,15	0.57	1 (7%)	17,19,21	1.16	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
6	NAG	B	302	2	-	2/6/23/26	0/1/1/1
5	8BN	A	1105	-	-	1/12/22/22	0/4/4/4
6	NAG	B	303	2	-	1/6/23/26	0/1/1/1
6	NAG	B	301	2	-	2/6/23/26	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1105	8BN	C11-C13	11.01	1.50	1.38
5	A	1105	8BN	C12-C11	5.04	1.47	1.39
5	A	1105	8BN	C05-C06	3.11	1.47	1.42
5	A	1105	8BN	C08-N03	-2.44	1.31	1.34
6	B	301	NAG	C1-C2	2.04	1.55	1.52

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	1105	8BN	C14-O03-C12	6.39	124.68	113.65
5	A	1105	8BN	F01-C13-C11	6.06	124.06	118.13
5	A	1105	8BN	C17-C13-C11	-4.73	118.69	124.63
5	A	1105	8BN	C13-C17-C18	4.31	121.14	116.62
5	A	1105	8BN	O03-C12-C16	3.28	122.10	117.05

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	B	302	NAG	C4-C5-C6-O6
6	B	302	NAG	O5-C5-C6-O6
5	A	1105	8BN	C11-C09-O04-C05
6	B	301	NAG	O5-C5-C6-O6
6	B	301	NAG	C4-C5-C6-O6

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1105	8BN	1	0
6	B	301	NAG	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 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 then 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.

Ligand 8BN A 1105

Bond lengths

Bond angles

Torsions

Rings
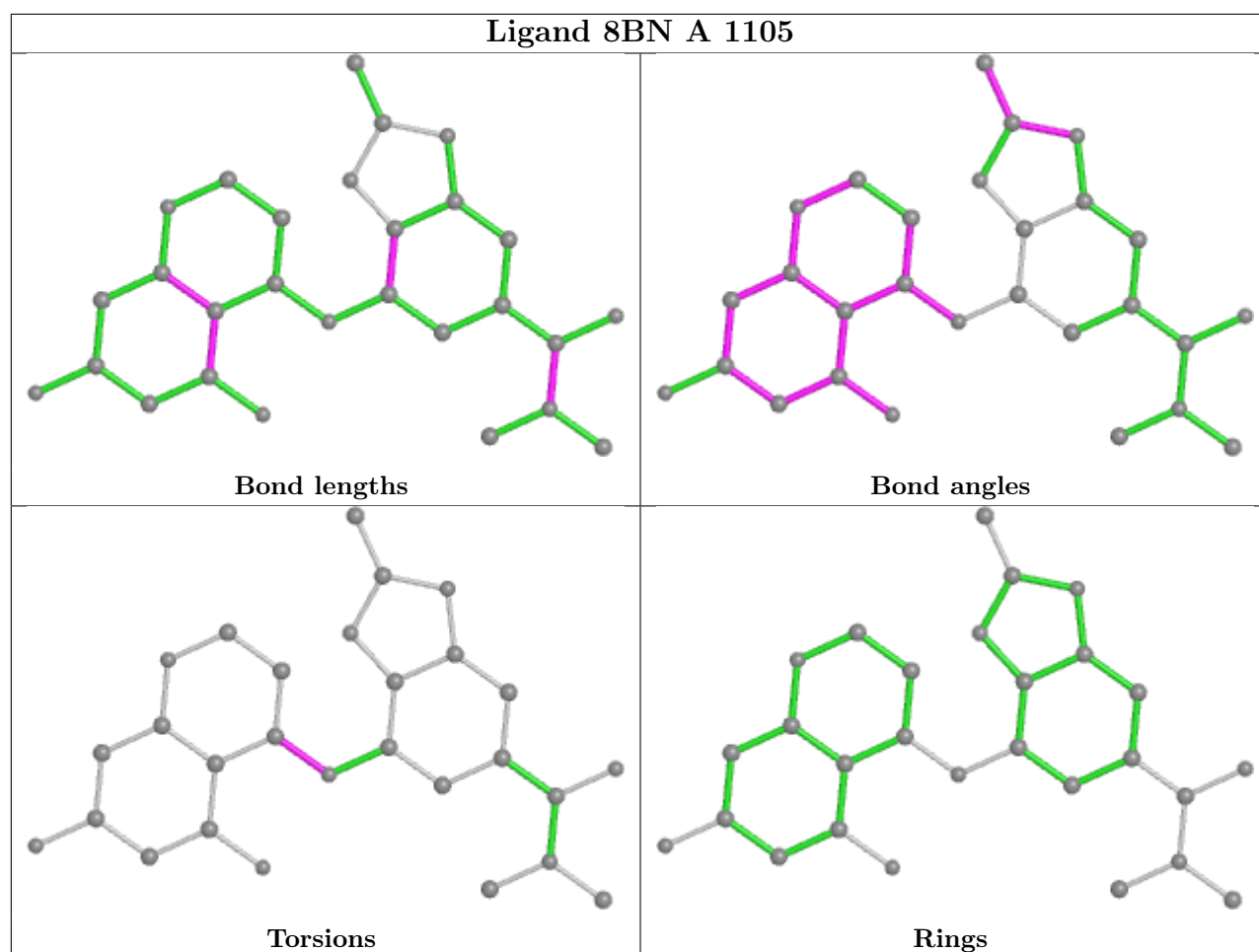
5.7 Other polymers ⓘ

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(Å^2)	Q<0.9
1	A	986/1034 (95%)	-0.54	14 (1%)	75	49	22, 52, 99, 166	0
2	B	260/290 (89%)	-0.34	7 (2%)	54	26	36, 73, 124, 191	0
All	All	1246/1324 (94%)	-0.50	21 (1%)	70	41	22, 57, 108, 191	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	33	THR	5.2
1	A	49	GLU	4.2
1	A	1025	SER	4.0
2	B	32	ARG	3.6
1	A	662	ASP	3.1

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

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.

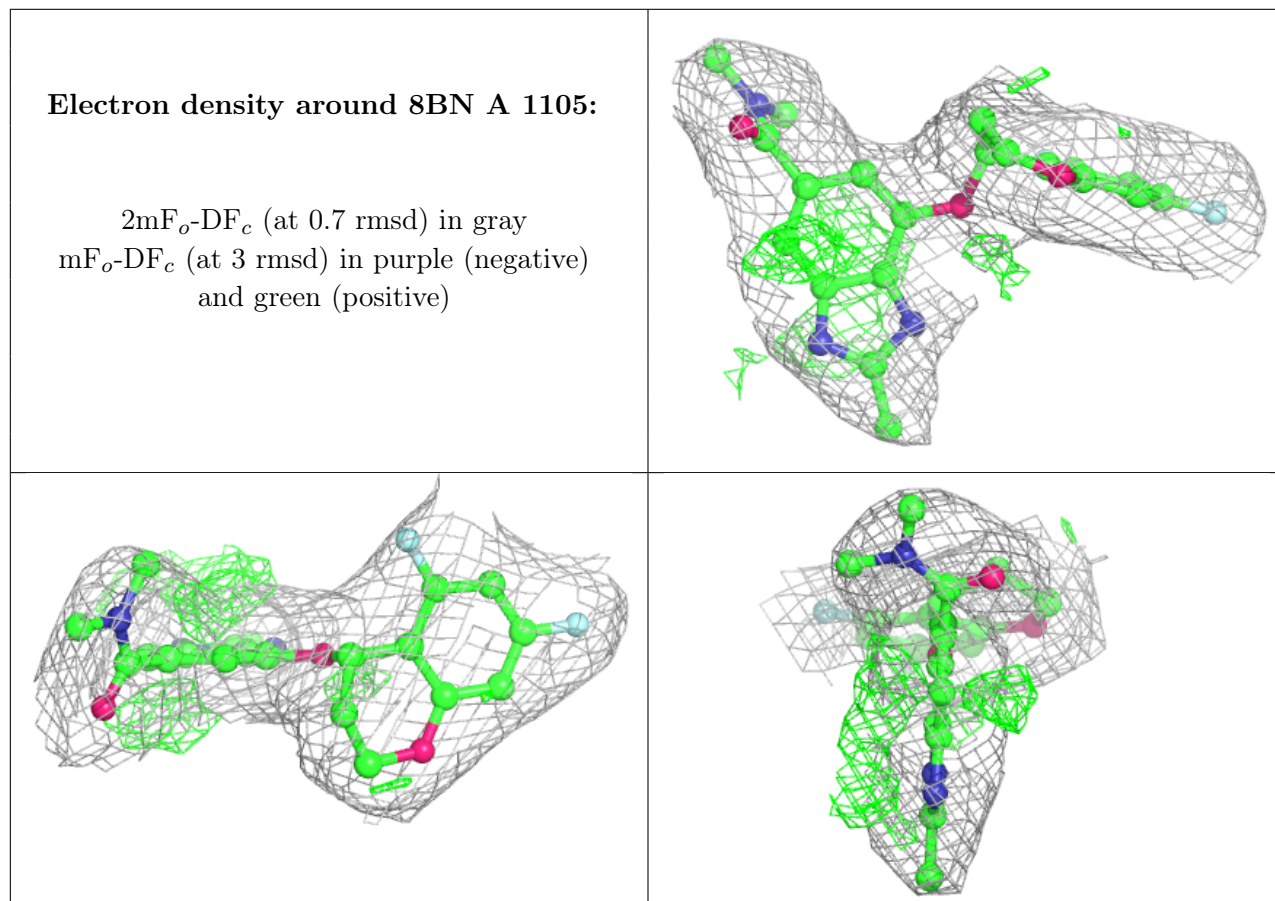| Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
1	BFD	A	385	12/13	0.94	0.22	28,39,45,58	0

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

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
6	NAG	B	302	14/15	0.74	0.48	110,122,131,132	0
6	NAG	B	303	14/15	0.78	0.46	86,112,121,121	0
3	MG	A	1101	1/1	0.92	0.24	21,21,21,21	0
5	8BN	A	1105	28/28	0.95	0.21	25,43,68,78	0
6	NAG	B	301	14/15	0.95	0.15	34,48,53,64	0
4	RB	A	1103	1/1	0.98	0.17	73,73,73,73	0
4	RB	A	1104	1/1	0.98	0.33	130,130,130,130	0
4	RB	A	1102	1/1	0.98	0.22	95,95,95,95	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.
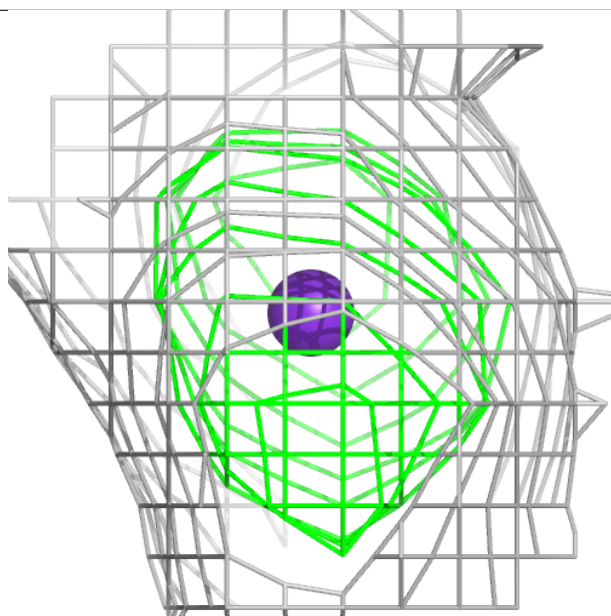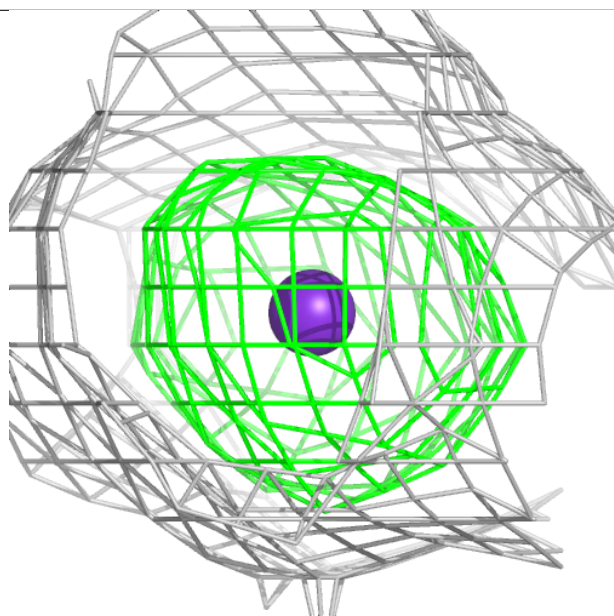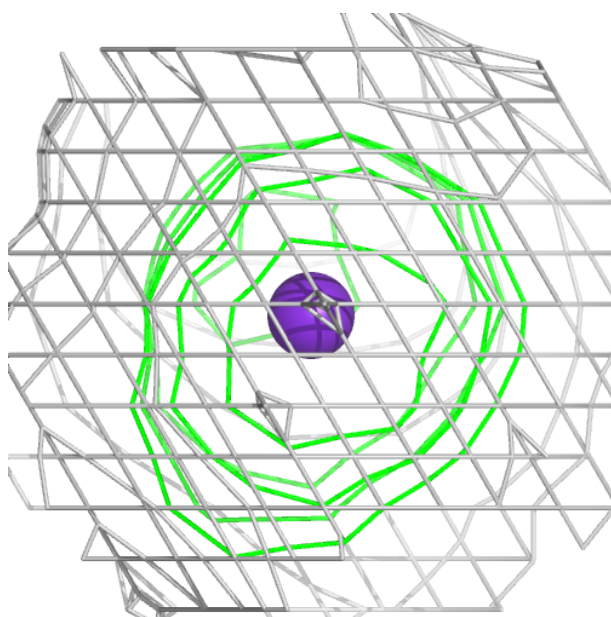
Electron density around 8BN A 1105:

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

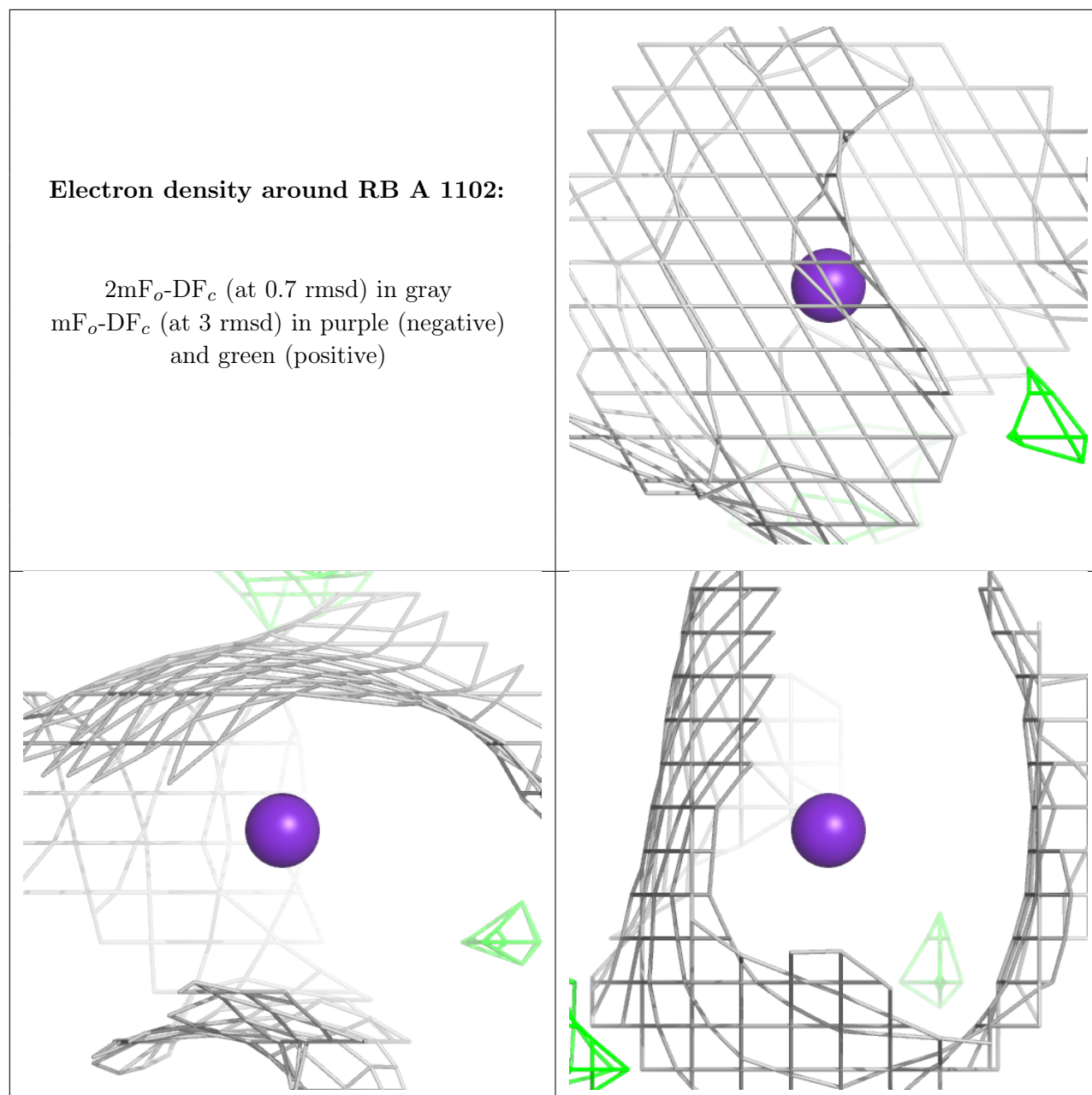**Electron density around RB A 1103:**

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)
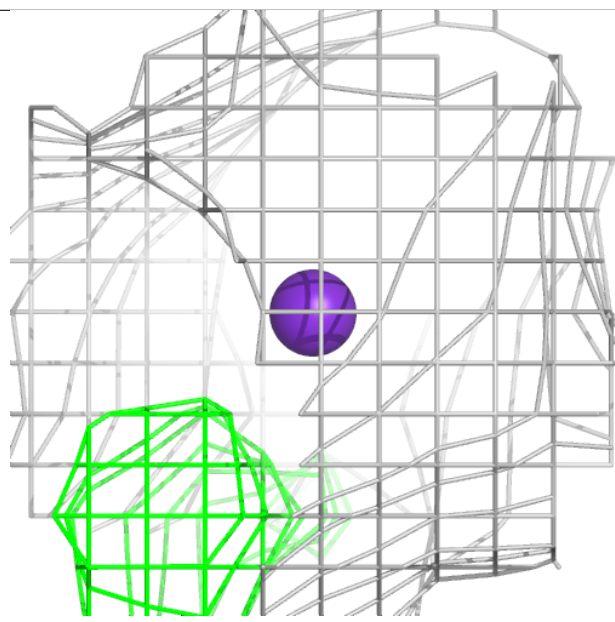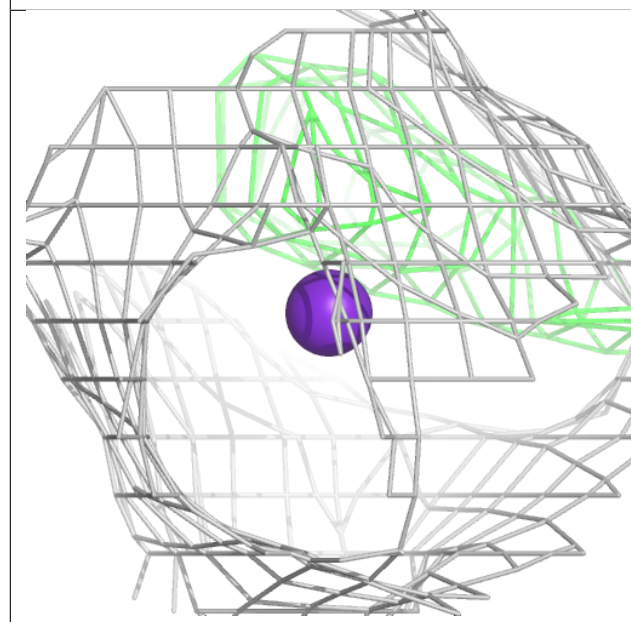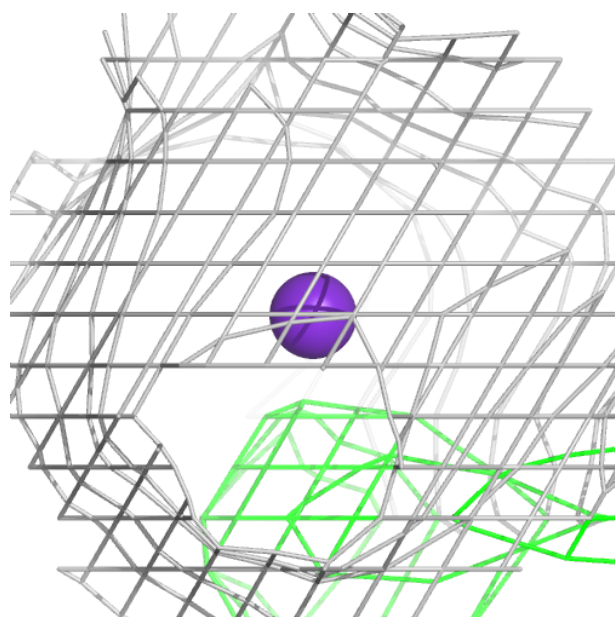
Electron density around RB A 1104:

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

Electron density around RB A 1102:

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

6.5 Other polymers ⓘ

There are no such residues in this entry.