Full wwPDB X-ray Structure Validation Report ⓘ

Oct 17, 2022 – 01:56 pm BST

PDB ID	:	7O9R
Title	:	Crystal structure of holo-H44A mutant of Hydroxy ketone aldolase (SwHKA) from Sphingomonas wittichii RW1
Authors	:	Laustsen, J.; Justo, I.; Marsden, S.R.; Hanefeld, U.; Bento, I.
Deposited on	:	2021-04-16
Resolution	:	1.85 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

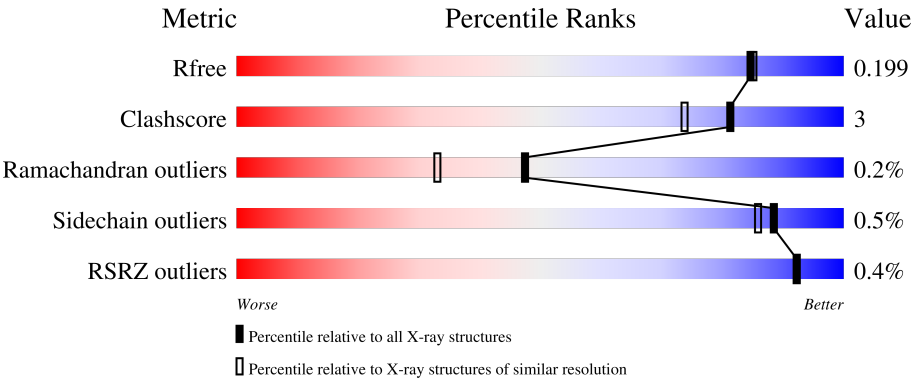
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	AAA	252	92% 7%
1	BBB	252	% 94% 6%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 7840 atoms, of which 3820 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 HpcH/HpaI aldolase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	250	Total	C	H	N	O	S	43	4	0
			3771	1185	1892	334	351	9			
1	BBB	252	Total	C	H	N	O	S	43	5	0
			3824	1202	1918	340	355	9			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	0	HIS	-	expression tag	UNP A5VH82
AAA	1	HIS	-	expression tag	UNP A5VH82
AAA	44	ALA	HIS	engineered mutation	UNP A5VH82
BBB	0	HIS	-	expression tag	UNP A5VH82
BBB	1	HIS	-	expression tag	UNP A5VH82
BBB	44	ALA	HIS	engineered mutation	UNP A5VH82

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	AAA	1	Total	Mg	0	0
			1	1		
2	BBB	1	Total	Mg	0	0
			1	1		

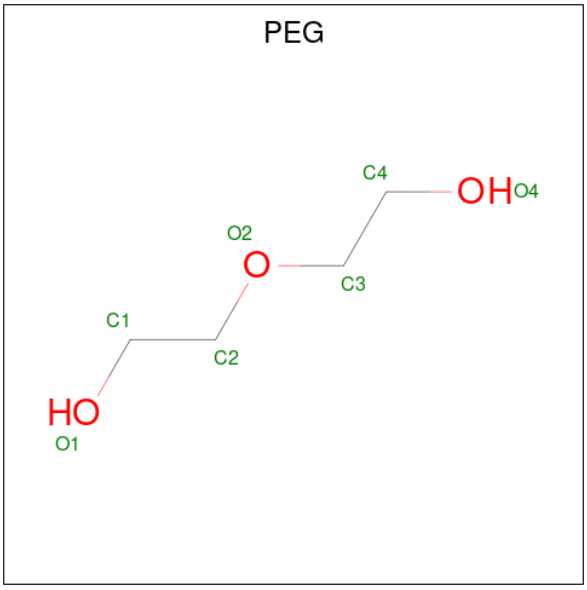
- Molecule 3 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	1	Total	K	0	0
			1	1		
3	BBB	2	Total	K	0	0
			2	2		

- Molecule 4 is BROMIDE ION (three-letter code: BR) (formula: Br).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	4	Total Br 4 4	0	0
4	BBB	2	Total Br 2 2	0	0

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	BBB	1	Total C H O 18 4 10 4	2	1

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	AAA	122	Total O 122 122	0	2
6	BBB	94	Total O 94 94	0	1

3 Residue-property plots (i)

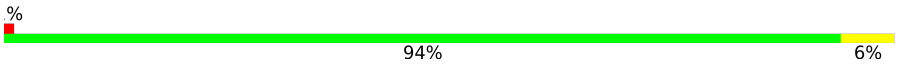
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density.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

Chain AAA:

• Molecule 1: HpcH/HpaI aldolase

Chain BBB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	70.88Å 70.88Å 223.18Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	59.19 – 1.85 59.19 – 1.85	Depositor EDS
% Data completeness (in resolution range)	93.4 (59.19-1.85) 76.3 (59.19-1.85)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.39 (at 1.86Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.158 , 0.196 0.167 , 0.199	Depositor DCC
R_{free} test set	1399 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å^2)	26.1	Xtriage
Anisotropy	0.036	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.038 for -h-k,k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	7840	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.32% 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: MG, BR, K, PEG

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	AAA	0.68	0/1926	0.78	0/2618
1	BBB	0.68	0/1951	0.80	0/2652
All	All	0.68	0/3877	0.79	0/5270

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	AAA	1879	1892	1892	10	0
1	BBB	1906	1918	1913	10	0
2	AAA	1	0	0	0	0
2	BBB	1	0	0	0	0
3	AAA	1	0	0	0	0
3	BBB	2	0	0	0	0
4	AAA	4	0	0	1	0
4	BBB	2	0	0	0	0
5	BBB	8	10	10	0	0
6	AAA	122	0	0	0	1
6	BBB	94	0	0	3	0
All	All	4020	3820	3815	20	1

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 3.

All (20) 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:AAA:170[B]:THR:OG1	4:AAA:303:BR:BR	2.69	0.55
1:BBB:178:HIS:ND1	6:BBB:402:HOH:O	2.35	0.49
1:BBB:108:HIS:O	1:BBB:113:GLY:HA3	2.14	0.48
1:AAA:49:PHE:O	1:AAA:52:VAL:HG12	2.14	0.47
1:AAA:249:ARG:O	1:AAA:250:ALA:HB3	2.16	0.47
1:BBB:197:ARG:NH2	6:BBB:411:HOH:O	2.49	0.46
1:AAA:208:GLY:HA2	1:AAA:227:PHE:O	2.15	0.46
1:AAA:142:ALA:O	1:AAA:165:VAL:HA	2.17	0.45
1:BBB:49:PHE:O	1:BBB:52:VAL:HG12	2.16	0.45
1:BBB:208:GLY:HA2	1:BBB:227:PHE:O	2.17	0.44
1:BBB:16:ALA:O	1:BBB:229:THR:HA	2.17	0.43
1:AAA:199:ARG:NE	1:AAA:199:ARG:HA	2.34	0.43
1:BBB:110:SER:HA	1:BBB:113:GLY:O	2.19	0.42
1:BBB:197:ARG:CZ	6:BBB:460:HOH:O	2.66	0.42
1:AAA:16:ALA:O	1:AAA:229:THR:HA	2.20	0.41
1:AAA:143:MET:HA	1:AAA:166:PHE:O	2.19	0.41
1:AAA:92:PRO:HA	1:AAA:143:MET:HB2	2.02	0.41
1:BBB:230:ALA:HA	1:BBB:231:ALA:HA	1.92	0.41
1:BBB:40:ILE:HG21	1:BBB:52:VAL:HG23	2.03	0.40
1:AAA:40:ILE:HG21	1:AAA:52:VAL:HG23	2.02	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:AAA:447:HOH:O	6:AAA:447:HOH:O[3_545]	1.96	0.24

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	AAA	252/252 (100%)	244 (97%)	8 (3%)	0	100 100
1	BBB	255/252 (101%)	244 (96%)	10 (4%)	1 (0%)	34 19
All	All	507/504 (101%)	488 (96%)	18 (4%)	1 (0%)	47 33

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	BBB	118	GLY

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	AAA	190/188 (101%)	188 (99%)	2 (1%)	73 65
1	BBB	192/188 (102%)	191 (100%)	1 (0%)	88 86
All	All	382/376 (102%)	379 (99%)	3 (1%)	88 76

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	112[A]	ARG
1	AAA	112[B]	ARG
1	BBB	116	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 13 ligands modelled in this entry, 11 are monoatomic - leaving 2 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	$\#\|Z\| > 2$
5	PEG	BBB	301[A]	-	3,3,6	0.20	0	2,2,5	0.25	0
5	PEG	BBB	301[B]	-	3,3,6	0.18	0	2,2,5	0.20	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
5	PEG	BBB	301[A]	-	-	1/1/1/4	-
5	PEG	BBB	301[B]	-	-	1/1/1/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	BBB	301[A]	PEG	O1-C1-C2-O2
5	BBB	301[B]	PEG	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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

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, 95^{th} 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	AAA	250/252 (99%)	-0.57	0 100 100	20, 26, 40, 50	1 (0%)
1	BBB	252/252 (100%)	-0.42	2 (0%) 86 86	20, 28, 52, 72	0
All	All	502/504 (99%)	-0.50	2 (0%) 92 92	20, 27, 45, 72	1 (0%)

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	131	LEU	3.6
1	BBB	114	LEU	2.0

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

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, 95^{th} 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
2	MG	AAA	301	1/1	0.67	0.15	50,50,50,50	1
2	MG	BBB	302	1/1	0.71	0.14	57,57,57,57	1
5	PEG	BBB	301[A]	4/7	0.91	0.12	32,43,43,43	9

Continued on next page...

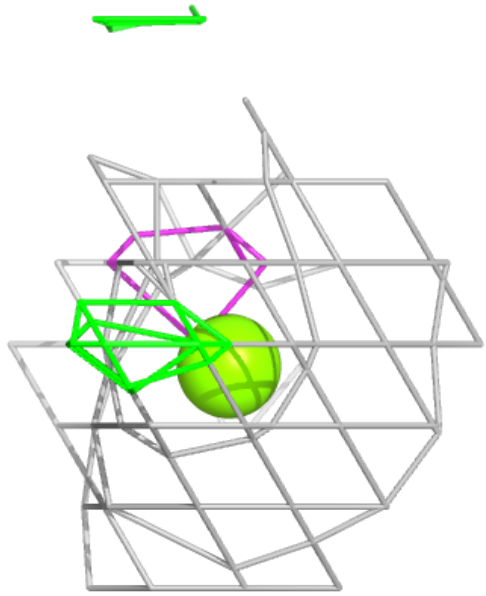
Continued from previous page...

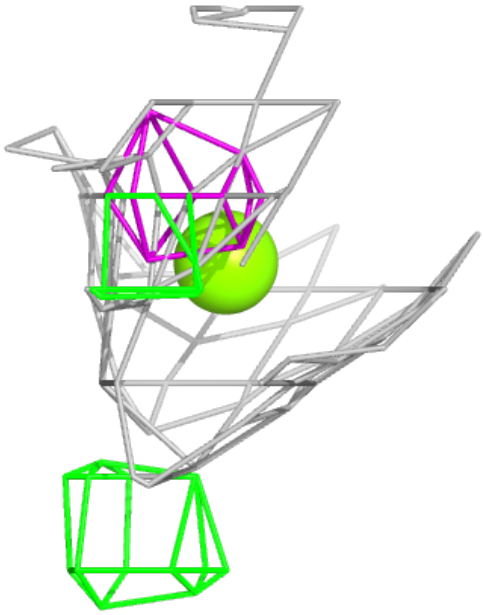
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	PEG	BBB	301[B]	4/7	0.91	0.12	36,38,40,41	9
4	BR	BBB	305	1/1	0.97	0.03	37,37,37,37	1
3	K	BBB	304	1/1	0.98	0.16	74,74,74,74	1
4	BR	BBB	306	1/1	0.98	0.04	58,58,58,58	0
4	BR	AAA	303	1/1	0.98	0.05	40,40,40,40	1
4	BR	AAA	305	1/1	0.98	0.03	39,39,39,39	1
3	K	BBB	303	1/1	0.99	0.06	26,26,26,26	0
4	BR	AAA	306	1/1	0.99	0.04	35,35,35,35	1
4	BR	AAA	304	1/1	0.99	0.06	27,27,27,27	1
3	K	AAA	302	1/1	1.00	0.08	25,25,25,25	0

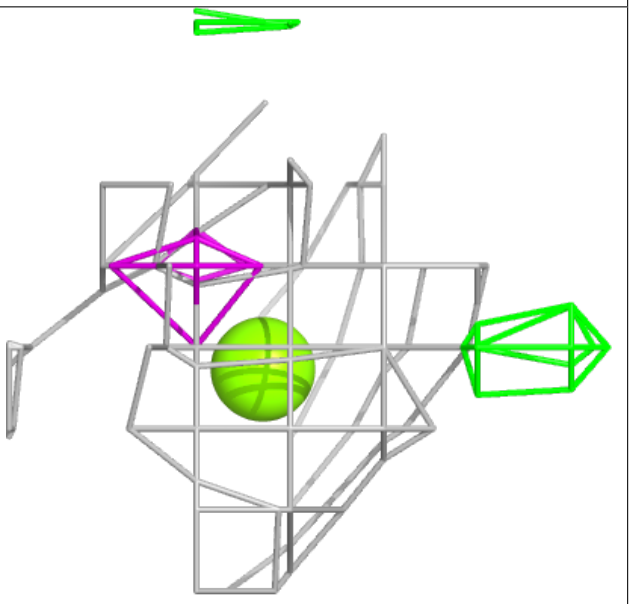
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around MG AAA 301:

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

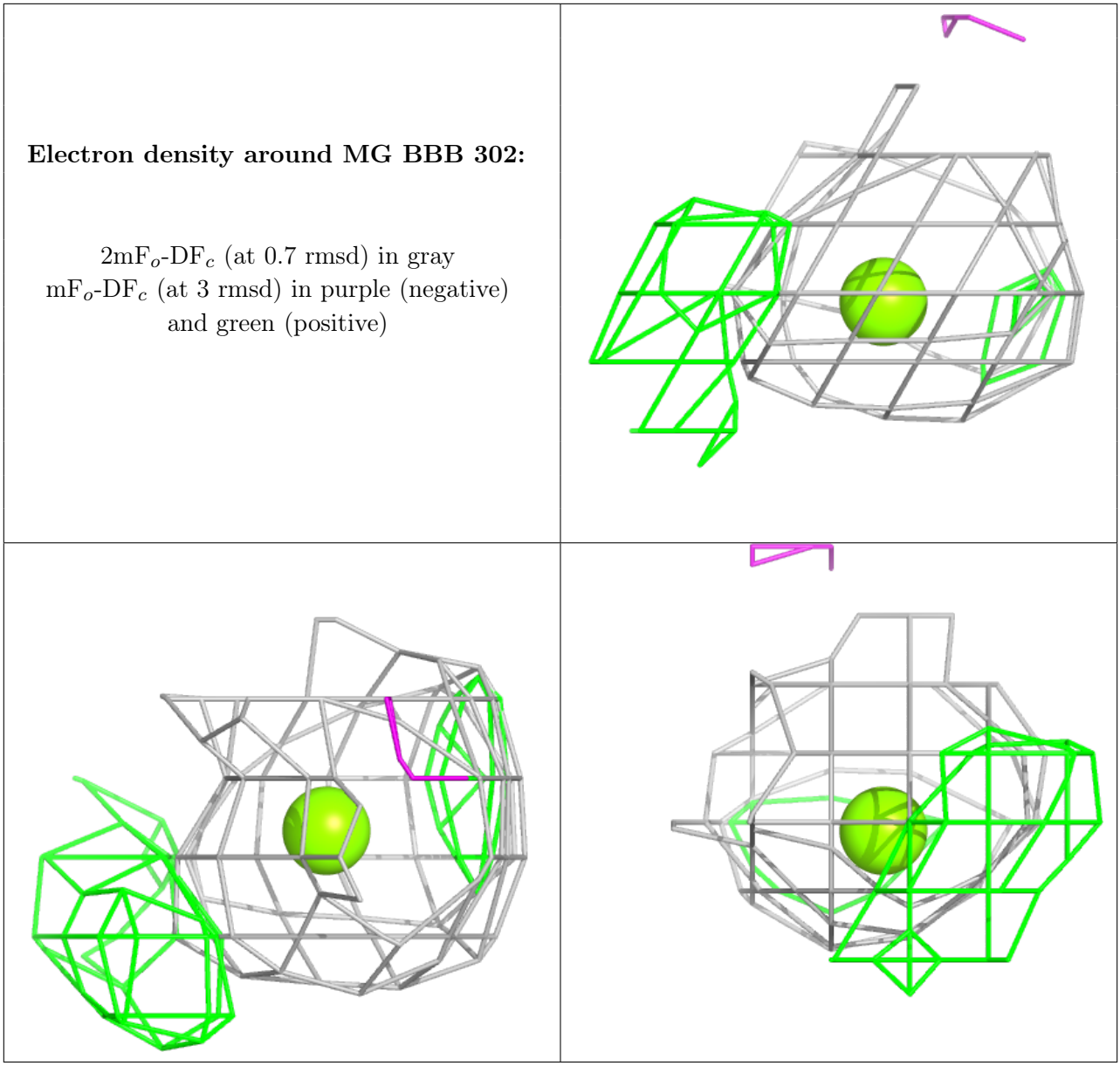






Electron density around MG BBB 302:

$2mF_o$-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 (i)

There are no such residues in this entry.