Full wwPDB X-ray Structure Validation Report ⓘ

Sep 7, 2020 – 02:53 PM BST

PDB ID	:	6WO0
Title	:	human Artemis/SNM1C catalytic domain, crystal form 1
Authors	:	Karim, F.; Liu, S.; Laciak, A.R.; Volk, L.; Rosenblum, M.; Curtis, R.; Huang, N.; Carr, G.; Zhu, G.
Deposited on	:	2020-04-23
Resolution	:	1.97 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 following versions of software and data (see references ⓘ) were used in the production of this report:

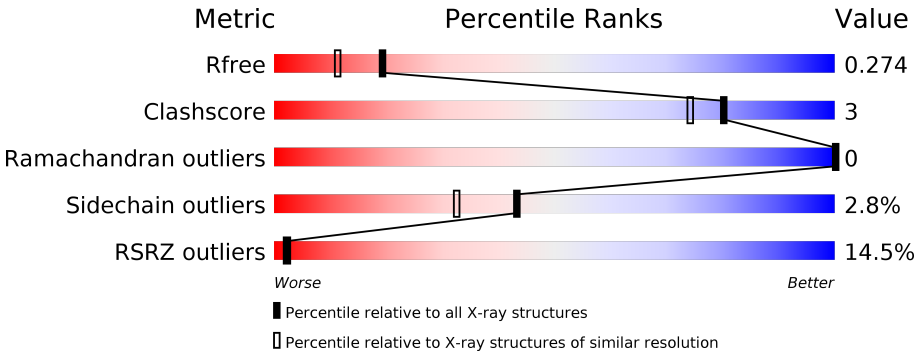
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.14.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.2

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 1.97 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 on 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	375	14% 88% 7%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3040 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 Protein artemis.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	359	Total	C	N	O	S	0	2	0
			2921	1865	510	526	20			

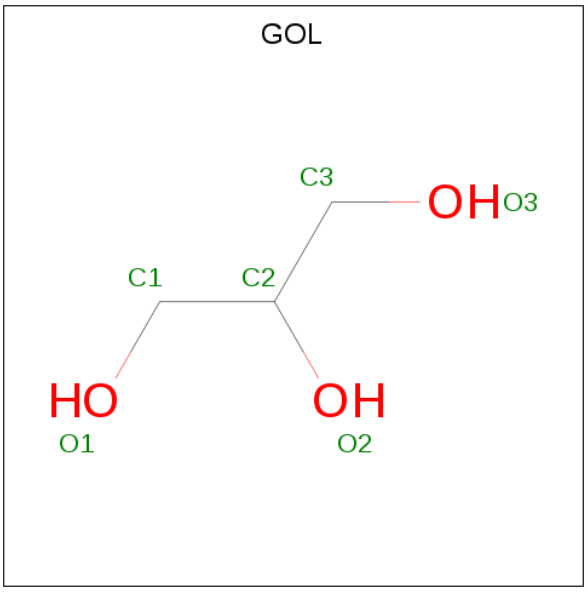
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	369	LYS	-	expression tag	UNP Q96SD1
A	370	GLY	-	expression tag	UNP Q96SD1
A	371	GLU	-	expression tag	UNP Q96SD1
A	372	ASN	-	expression tag	UNP Q96SD1
A	373	LEU	-	expression tag	UNP Q96SD1
A	374	TYR	-	expression tag	UNP Q96SD1
A	375	PHE	-	expression tag	UNP Q96SD1
A	376	GLN	-	expression tag	UNP Q96SD1

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by author).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Zn	0	0
			2	2		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 4 is water.

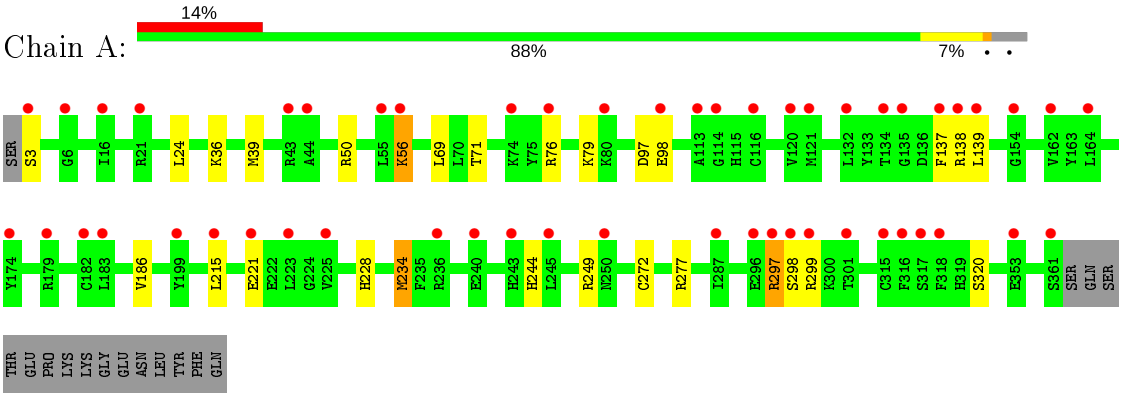
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	105	Total O 105 105	0	0

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: Protein artemis

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	72.69Å 111.01Å 55.16Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.54 – 1.97 34.54 – 1.97	Depositor EDS
% Data completeness (in resolution range)	99.5 (34.54-1.97) 99.6 (34.54-1.97)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.71 (at 1.97Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.218 , 0.270 0.225 , 0.274	Depositor DCC
R_{free} test set	1560 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	42.8	Xtriage
Anisotropy	0.525	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 45.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3040	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.43% 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: GOL, ZN

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	0.69	0/3002	0.83	0/4065

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	2921	0	2912	16	0
2	A	2	0	0	0	0
3	A	12	0	15	1	0
4	A	105	0	0	0	0
All	All	3040	0	2927	16	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 3.

All (16) 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:228:HIS:CE1	1:A:249:ARG:HG2	2.35	0.61
1:A:24:LEU:O	1:A:50:ARG:HD3	2.04	0.58
1:A:36:LYS:HB2	3:A:404:GOL:H2	1.85	0.57
1:A:76:ARG:O	1:A:79:LYS:HB2	2.06	0.56
1:A:39:MET:HE3	1:A:69:LEU:HD13	1.88	0.56
1:A:221:GLU:HG2	1:A:244:HIS:CE1	2.40	0.55
1:A:97:ASP:OD1	1:A:97:ASP:C	2.51	0.49
1:A:56:LYS:HD3	1:A:56:LYS:HA	1.50	0.48
1:A:228:HIS:HE1	1:A:272:CYS:CB	2.29	0.46
1:A:39:MET:CE	1:A:69:LEU:HD13	2.46	0.45
1:A:297:ARG:O	1:A:297:ARG:HG3	2.17	0.44
1:A:228:HIS:HE1	1:A:272:CYS:HB3	1.84	0.43
1:A:186:VAL:HG11	1:A:215:LEU:HD11	2.02	0.41
1:A:71:THR:HG21	1:A:234:MET:HA	2.02	0.41
1:A:97:ASP:OD1	1:A:98:GLU:N	2.54	0.41
1:A:137:PHE:HD2	1:A:139:LEU:HG	1.85	0.40

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	359/375 (96%)	351 (98%)	8 (2%)	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	328/341 (96%)	319 (97%)	9 (3%)	44 35

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3	SER
1	A	56	LYS
1	A	138	ARG
1	A	234	MET
1	A	277	ARG
1	A	297	ARG
1	A	298	SER
1	A	299	ARG
1	A	320	SER

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

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

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 4 ligands modelled in this entry, 2 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	$\#\|Z\| > 2$
3	GOL	A	403	2	5,5,5	0.19	0	5,5,5	0.42	0
3	GOL	A	404	-	5,5,5	0.11	0	5,5,5	0.25	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	GOL	A	403	2	-	4/4/4/4	-
3	GOL	A	404	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	404	GOL	O1-C1-C2-O2
3	A	404	GOL	O1-C1-C2-C3
3	A	403	GOL	O1-C1-C2-C3
3	A	403	GOL	C1-C2-C3-O3
3	A	403	GOL	O2-C2-C3-O3
3	A	403	GOL	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	404	GOL	1	0

5.7 Other polymers (i)

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, 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	A	359/375 (95%)	0.93	52 (14%) 2 2	40, 54, 82, 113	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	297	ARG	6.7
1	A	299	ARG	5.2
1	A	301	THR	4.2
1	A	76	ARG	4.2
1	A	361	SER	3.8
1	A	298	SER	3.7
1	A	113	ALA	3.7
1	A	182	CYS	3.7
1	A	250	ASN	3.6
1	A	240	GLU	3.5
1	A	243[A]	HIS	3.5
1	A	80	LYS	3.5
1	A	318	PHE	3.5
1	A	315	CYS	3.4
1	A	245	LEU	3.3
1	A	317	SER	3.3
1	A	183	LEU	3.2
1	A	179	ARG	3.1
1	A	138	ARG	3.1
1	A	120	VAL	2.9
1	A	225	VAL	2.9
1	A	164	LEU	2.8
1	A	116	CYS	2.7
1	A	287	ILE	2.7
1	A	316	PHE	2.7
1	A	74	LYS	2.7
1	A	236	ARG	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	56	LYS	2.6
1	A	139	LEU	2.5
1	A	44	ALA	2.5
1	A	174	TYR	2.5
1	A	114	GLY	2.5
1	A	121	MET	2.5
1	A	154	GLY	2.5
1	A	221	GLU	2.4
1	A	135	GLY	2.4
1	A	162	VAL	2.4
1	A	134	THR	2.4
1	A	199	TYR	2.4
1	A	353	GLU	2.4
1	A	21	ARG	2.3
1	A	137	PHE	2.3
1	A	3	SER	2.2
1	A	296	GLU	2.2
1	A	55	LEU	2.2
1	A	132	LEU	2.2
1	A	98	GLU	2.2
1	A	6	GLY	2.2
1	A	215	LEU	2.1
1	A	223	LEU	2.1
1	A	43	ARG	2.0
1	A	16	ILE	2.0

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)

There are no monosaccharides in this entry.

6.4 Ligands (i)

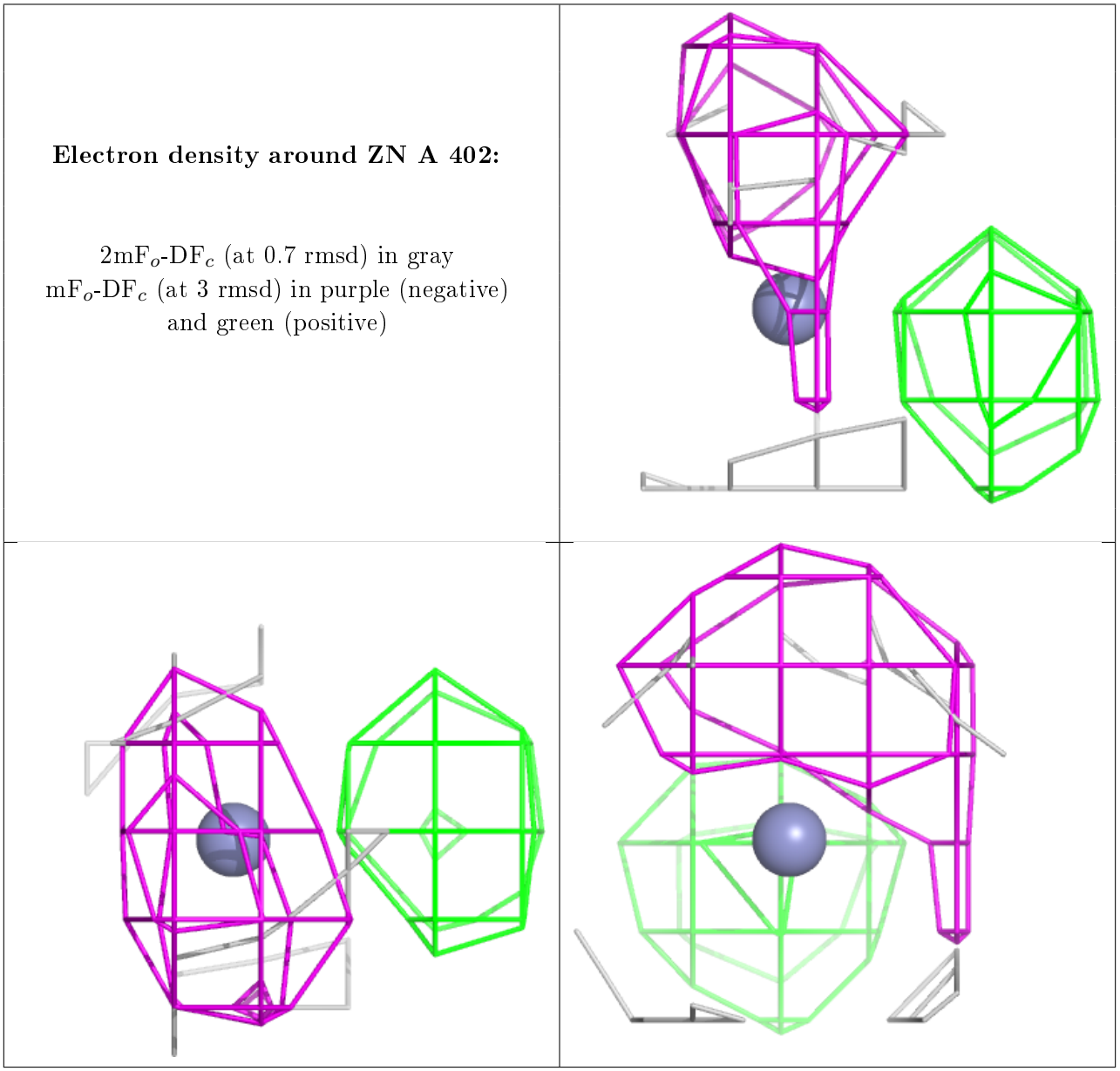
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
3	GOL	A	404	6/6	0.79	0.23	48,55,64,67	0
2	ZN	A	402	1/1	0.94	0.09	45,45,45,45	0
3	GOL	A	403	6/6	0.94	0.18	36,45,48,52	0
2	ZN	A	401	1/1	0.97	0.09	43,43,43,43	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.

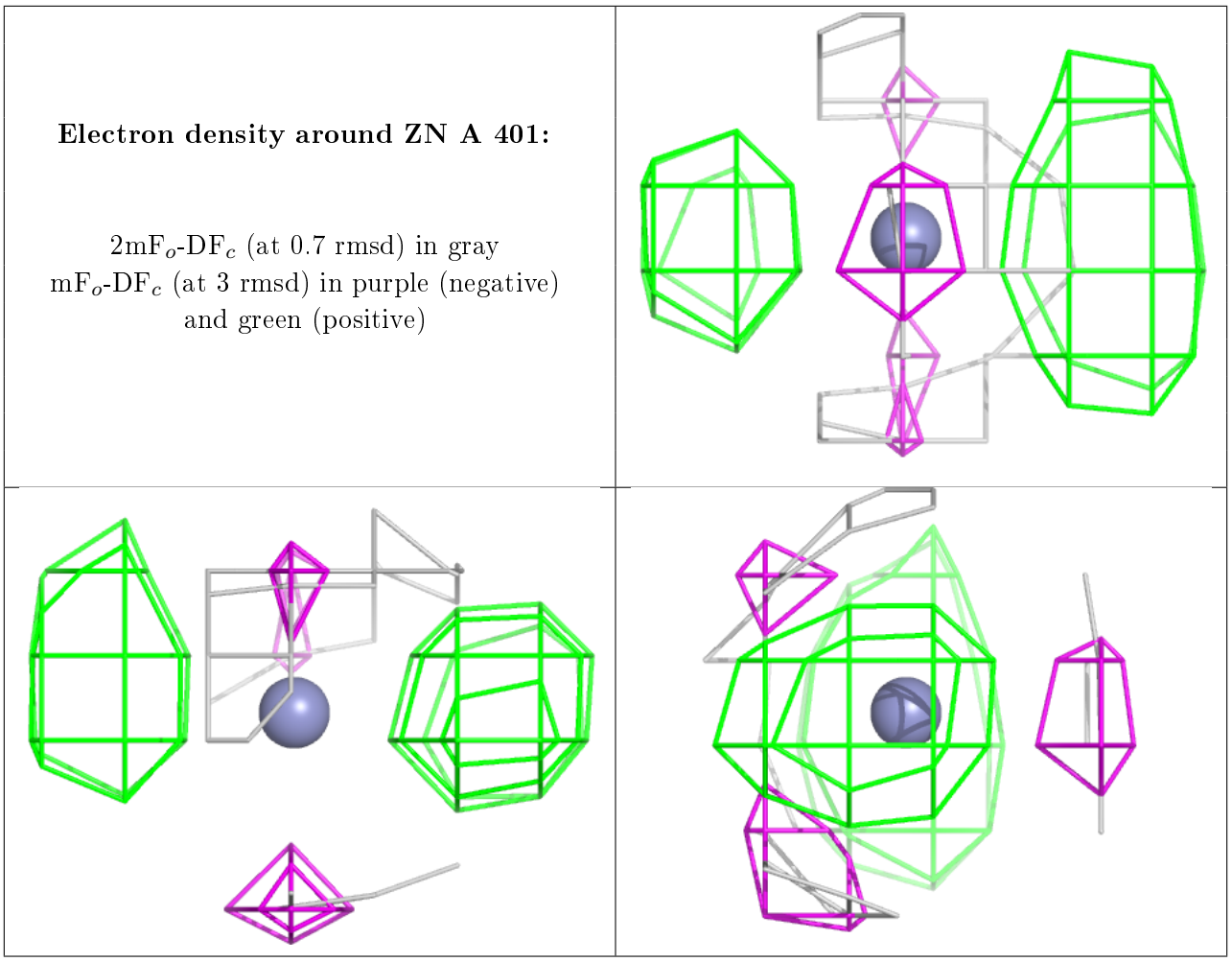
Electron density around ZN A 402:

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



Electron density around ZN A 401:

$2mF_o$-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.