wwPDB X-ray Structure Validation Summary Report (i)

Apr 5, 2022 – 10:06 am BST

PDB ID	:	7O45
Title	:	Crystal structure of ADD domain of the human DNMT3B methyltransferase
Authors	:	Boyko, K.M.; Nikolaeva, A.Y.; Bonchuk, A.N.; Georgiev, P.G.; Popov, V.O.
Deposited on	:	2021-04-05
Resolution	:	2.10 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

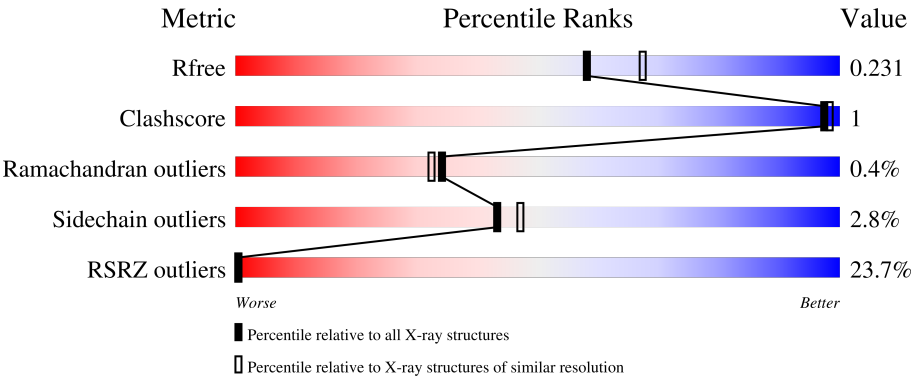
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.27
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	146	10% 95%
1	B	146	14% 92% 5%
1	C	146	10% 88% 8%
1	D	146	49% 58% 42%

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
2	ZN	D	601	-	-	-	X

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4065 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 Isoform 6 of DNA (cytosine-5)-methyltransferase 3B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	141	Total	C	N	O	S	0	1	0
			1105	680	194	212	19			
1	B	141	Total	C	N	O	S	0	2	0
			1109	681	194	214	20			
1	C	141	Total	C	N	O	S	0	3	0
			1111	682	196	214	19			
1	D	84	Total	C	N	O	S	0	0	0
			596	367	110	105	14			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	409	SER	-	expression tag	UNP Q9UBC3
A	410	GLY	-	expression tag	UNP Q9UBC3
A	411	SER	-	expression tag	UNP Q9UBC3
B	409	SER	-	expression tag	UNP Q9UBC3
B	410	GLY	-	expression tag	UNP Q9UBC3
B	411	SER	-	expression tag	UNP Q9UBC3
C	409	SER	-	expression tag	UNP Q9UBC3
C	410	GLY	-	expression tag	UNP Q9UBC3
C	411	SER	-	expression tag	UNP Q9UBC3
D	409	SER	-	expression tag	UNP Q9UBC3
D	410	GLY	-	expression tag	UNP Q9UBC3
D	411	SER	-	expression tag	UNP Q9UBC3

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Zn	0	0
			3	3		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	3	Total Zn 3 3	0	0
2	C	3	Total Zn 3 3	0	0
2	D	3	Total Zn 3 3	0	0

- Molecule 3 is BROMIDE ION (three-letter code: BR) (formula: Br).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total Br 2 2	0	0
3	C	1	Total Br 1 1	0	0

- Molecule 4 is water.

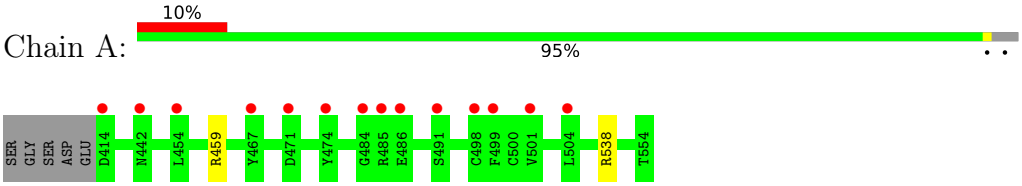
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	46	Total O 46 46	0	0
4	B	38	Total O 38 38	0	0
4	C	45	Total O 45 45	0	0

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.

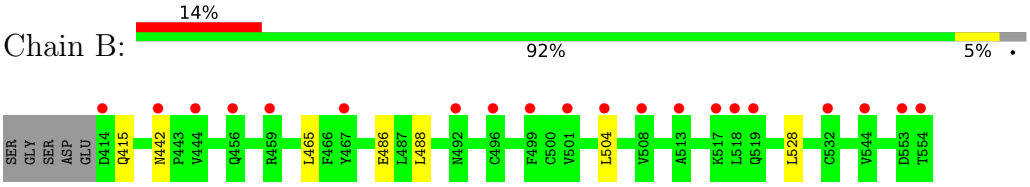
• Molecule 1: Isoform 6 of DNA (cytosine-5)-methyltransferase 3B

Chain A:



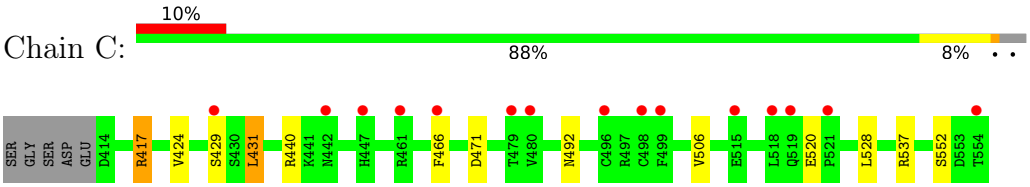
• Molecule 1: Isoform 6 of DNA (cytosine-5)-methyltransferase 3B

Chain B:



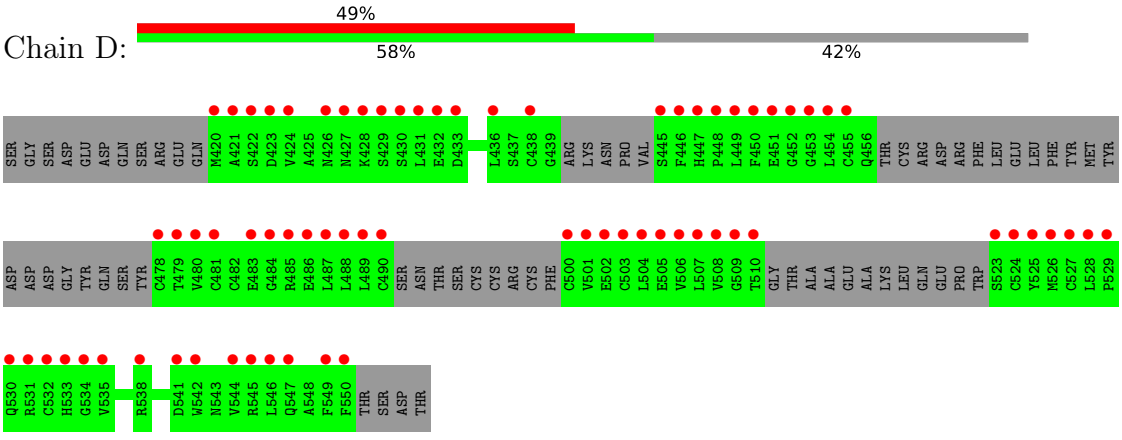
• Molecule 1: Isoform 6 of DNA (cytosine-5)-methyltransferase 3B

Chain C:



• Molecule 1: Isoform 6 of DNA (cytosine-5)-methyltransferase 3B

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	80.28Å 89.94Å 92.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	64.38 – 2.10 64.38 – 2.10	Depositor EDS
% Data completeness (in resolution range)	92.0 (64.38-2.10) 91.8 (64.38-2.10)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.69 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.200 , 0.228 0.202 , 0.231	Depositor DCC
R_{free} test set	1811 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	42.0	Xtriage
Anisotropy	0.581	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.018 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4065	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.43% 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: BR, 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	0.77	0/1130	0.91	1/1523 (0.1%)
1	B	0.78	0/1138	0.94	0/1534
1	C	0.77	0/1144	0.92	1/1544 (0.1%)
1	D	0.92	0/601	0.93	0/806
All	All	0.80	0/4013	0.93	2/5407 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	538	ARG	NE-CZ-NH1	5.16	122.88	120.30
1	C	537	ARG	NE-CZ-NH2	-5.16	117.72	120.30

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	1105	0	1026	1	0
1	B	1109	0	1027	1	0
1	C	1111	0	1025	4	0
1	D	596	0	531	0	0
2	A	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	3	0	0	0	0
2	C	3	0	0	0	0
2	D	3	0	0	0	0
3	A	2	0	0	0	0
3	C	1	0	0	0	0
4	A	46	0	0	1	0
4	B	38	0	0	0	0
4	C	45	0	0	1	0
All	All	4065	0	3609	6	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 1.

The worst 5 of 6 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:C:492[A]:ASN:OD1	4:C:702:HOH:O	1.84	0.95
1:C:424:VAL:HG21	1:C:431:LEU:HD13	1.89	0.55
1:A:459:ARG:NH1	4:A:702:HOH:O	2.45	0.49
1:B:488:LEU:HD13	1:B:504:LEU:HD22	1.96	0.47
1:C:424:VAL:HG11	1:C:431:LEU:HD13	1.95	0.47

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	140/146 (96%)	137 (98%)	3 (2%)	0	100 100
1	B	141/146 (97%)	133 (94%)	7 (5%)	1 (1%)	22 18
1	C	142/146 (97%)	139 (98%)	2 (1%)	1 (1%)	22 18

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	D	74/146 (51%)	69 (93%)	5 (7%)	0	100 100
All	All	497/584 (85%)	478 (96%)	17 (3%)	2 (0%)	34 32

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	415	GLN
1	C	466	PHE

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	124/130 (95%)	124 (100%)	0	100 100
1	B	126/130 (97%)	122 (97%)	4 (3%)	39 41
1	C	126/130 (97%)	118 (94%)	8 (6%)	18 15
1	D	60/130 (46%)	60 (100%)	0	100 100
All	All	436/520 (84%)	424 (97%)	12 (3%)	43 47

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	440	ARG
1	C	471	ASP
1	C	552	SER
1	C	520	GLU
1	B	528	LEU

Sometimes 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 15 ligands modelled in this entry, 15 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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

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	141/146 (96%)	1.20	14 (9%)	7	9	35, 47, 71, 81	0
1	B	141/146 (96%)	1.32	20 (14%)	2	3	34, 50, 80, 96	0
1	C	141/146 (96%)	1.13	15 (10%)	6	7	36, 45, 70, 105	0
1	D	84/146 (57%)	5.13	71 (84%)	0	0	66, 106, 140, 152	0
All	All	507/584 (86%)	1.87	120 (23%)	0	0	34, 51, 113, 152	0

The worst 5 of 120 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	484	GLY	20.3
1	D	504	LEU	15.8
1	D	489	LEU	14.9
1	D	428	LYS	13.1
1	D	500	CYS	12.9

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

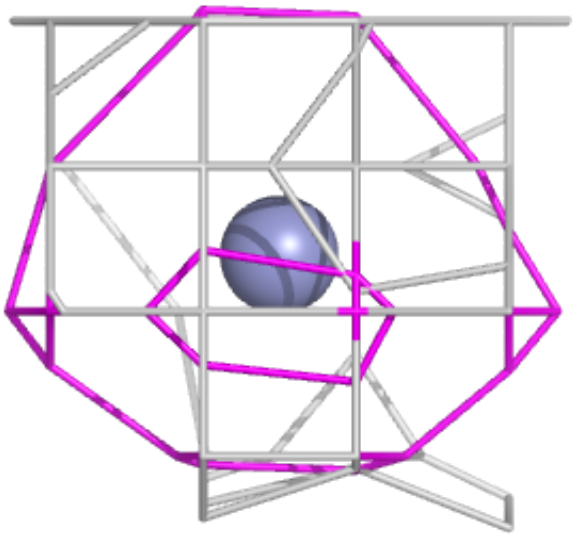
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

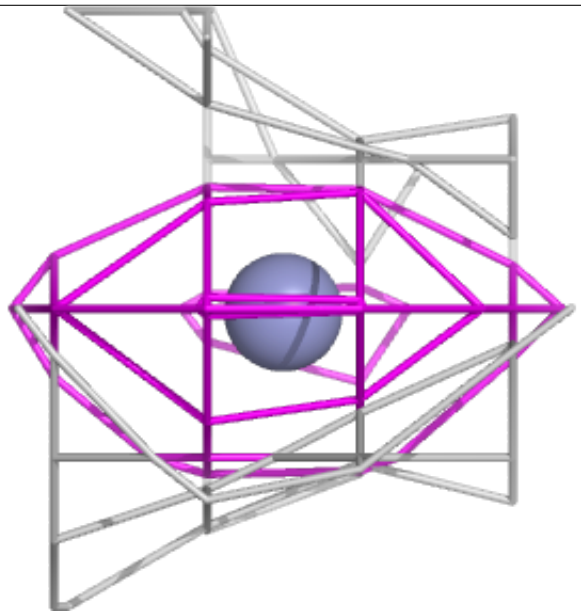
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	ZN	D	601	1/1	0.49	0.83	89,89,89,89	1
3	BR	A	604	1/1	0.51	0.09	107,107,107,107	0
2	ZN	D	602	1/1	0.88	0.15	77,77,77,77	1
2	ZN	B	602	1/1	0.91	0.05	45,45,45,45	0
2	ZN	B	603	1/1	0.92	0.07	43,43,43,43	0
2	ZN	B	601	1/1	0.93	0.07	47,47,47,47	0
3	BR	C	604	1/1	0.93	0.10	86,86,86,86	0
3	BR	A	605	1/1	0.94	0.12	78,78,78,78	0
2	ZN	A	601	1/1	0.96	0.07	43,43,43,43	0
2	ZN	C	601	1/1	0.96	0.11	40,40,40,40	0
2	ZN	D	603	1/1	0.96	0.19	80,80,80,80	1
2	ZN	A	602	1/1	0.97	0.07	42,42,42,42	0
2	ZN	A	603	1/1	0.97	0.09	36,36,36,36	0
2	ZN	C	603	1/1	0.97	0.08	38,38,38,38	0
2	ZN	C	602	1/1	0.98	0.05	40,40,40,40	0

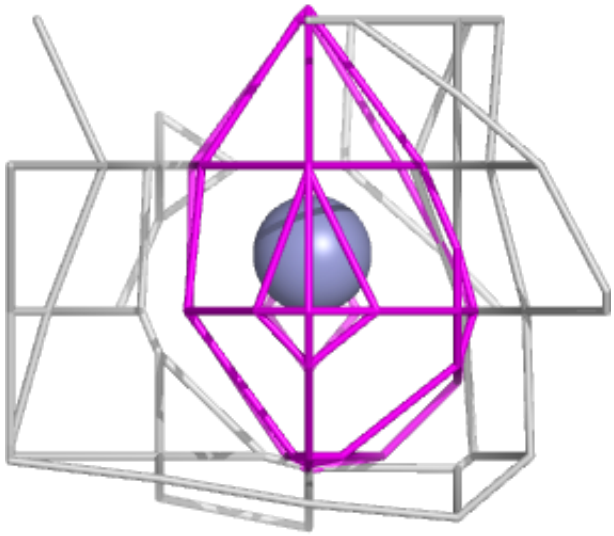
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around ZN D 601:

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

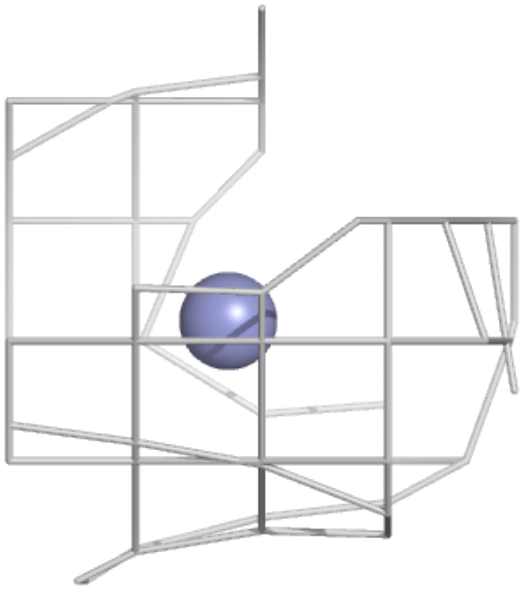


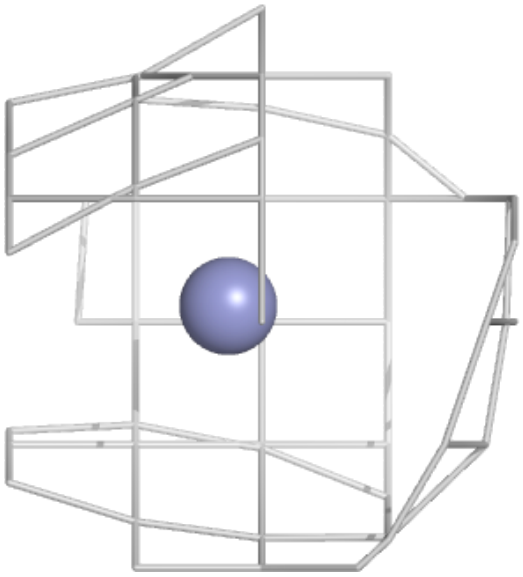


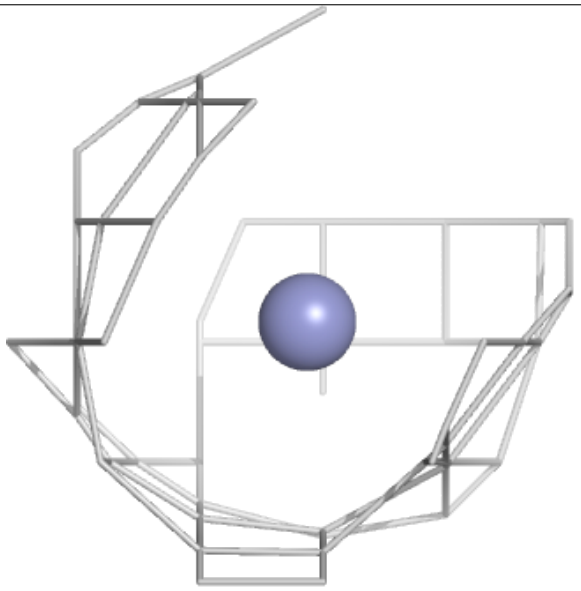


Electron density around ZN D 602:

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

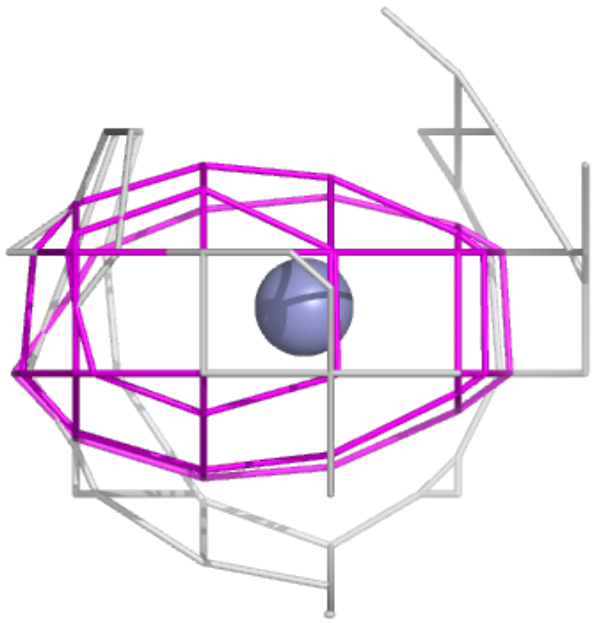


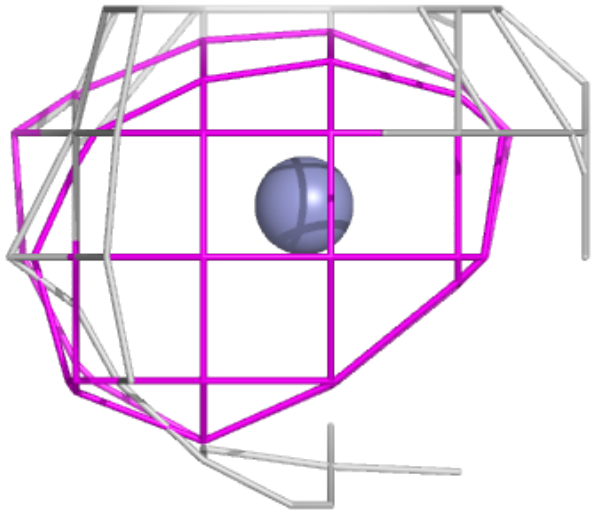


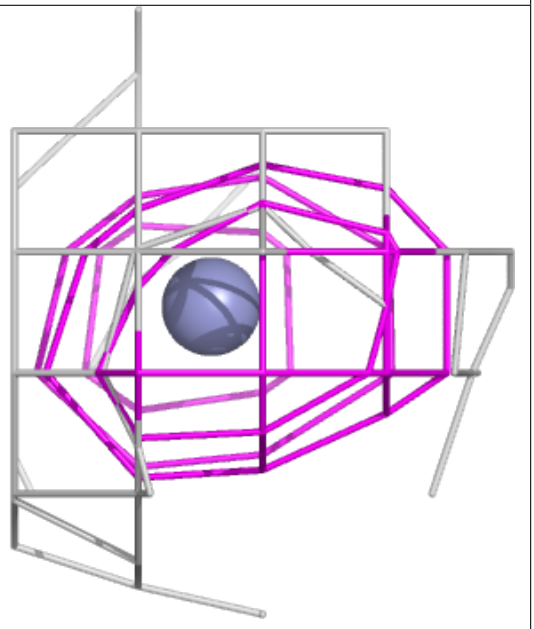


Electron density around ZN B 602:

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

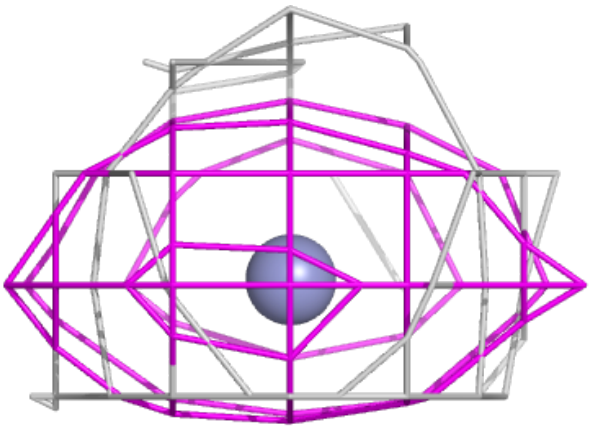


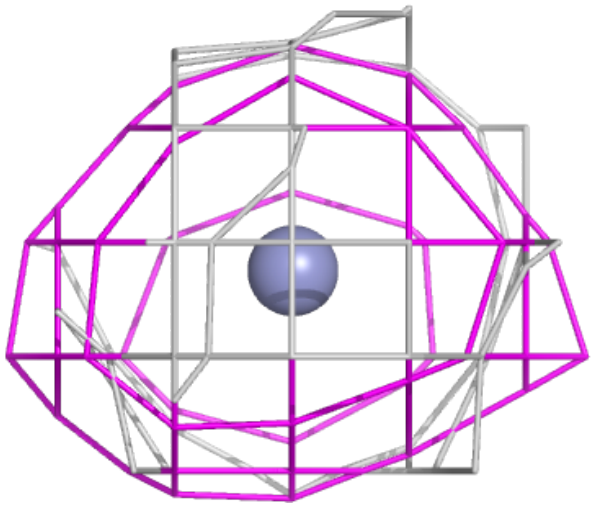


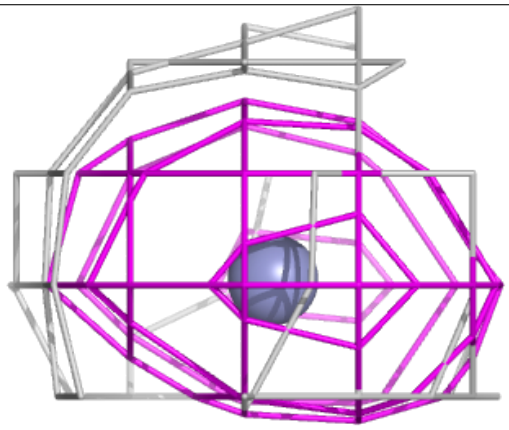


Electron density around ZN B 603:

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

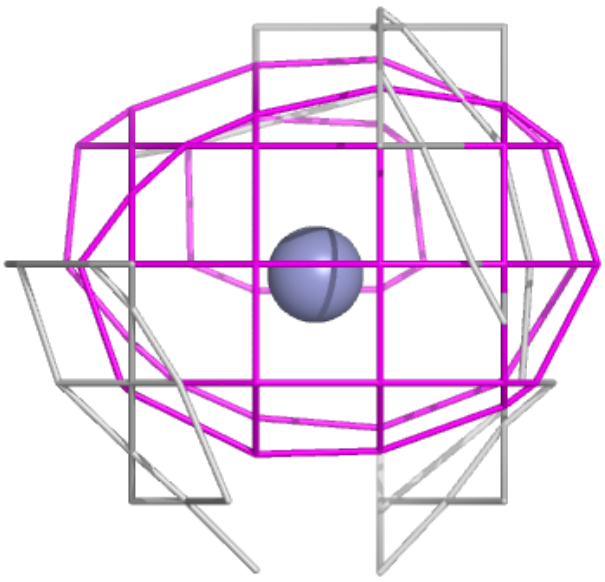


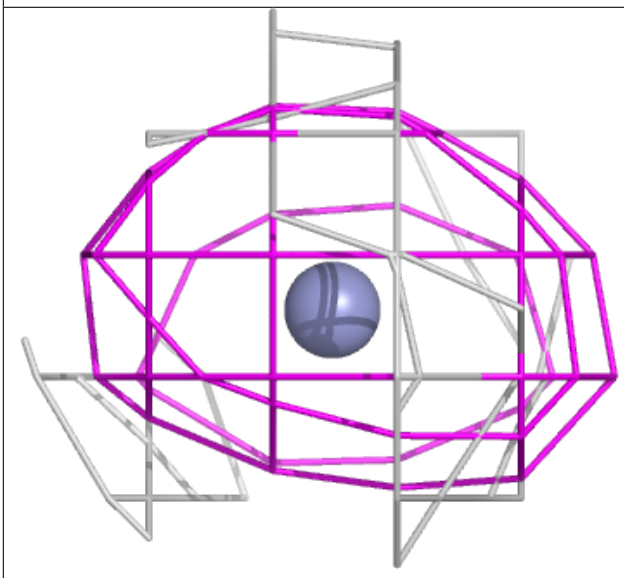


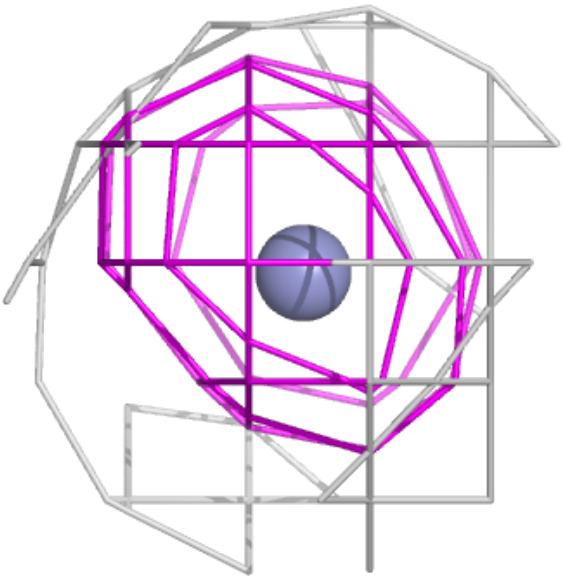


Electron density around ZN B 601:

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

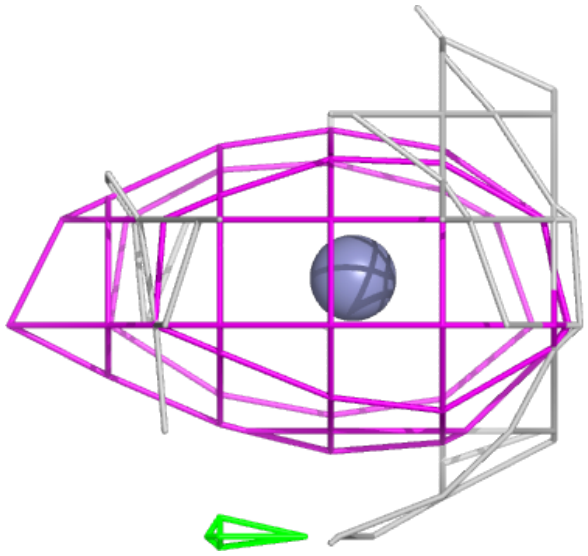


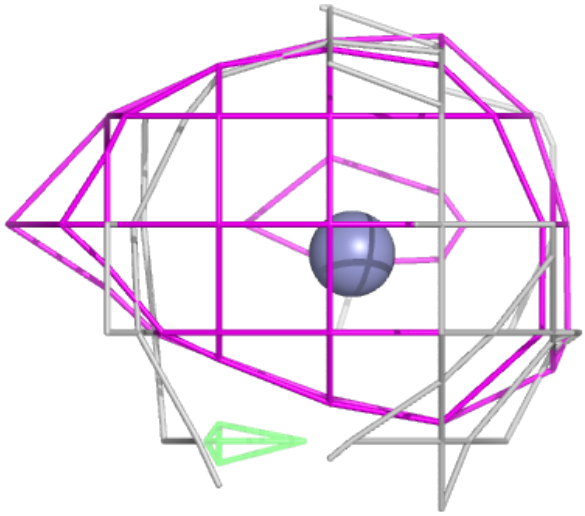


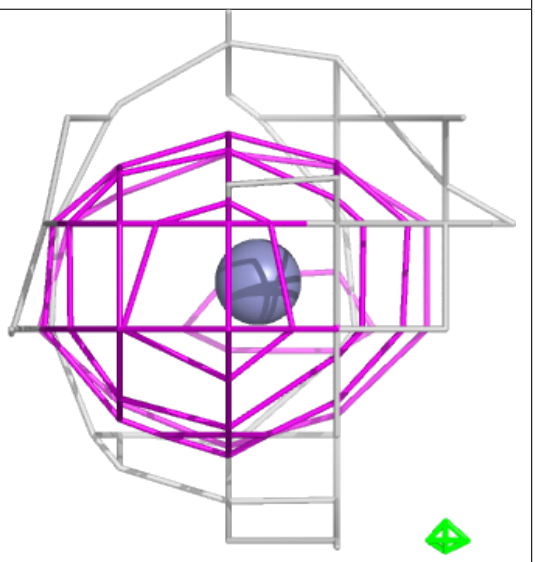


Electron density around ZN A 601:

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

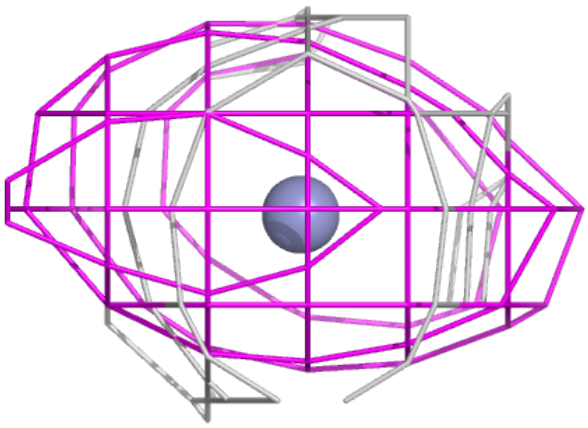


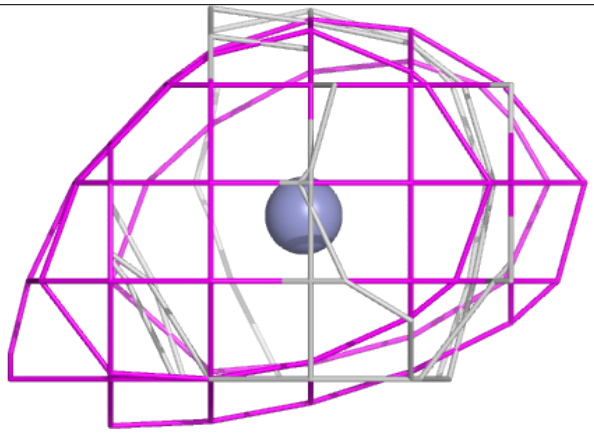


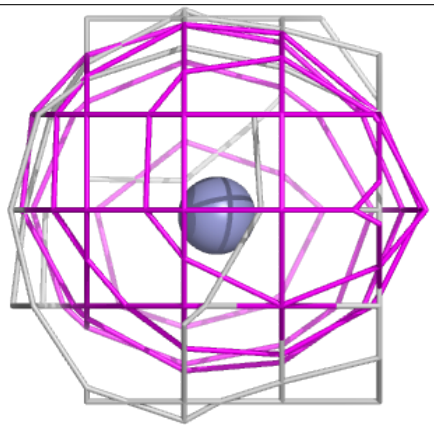


Electron density around ZN C 601:

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

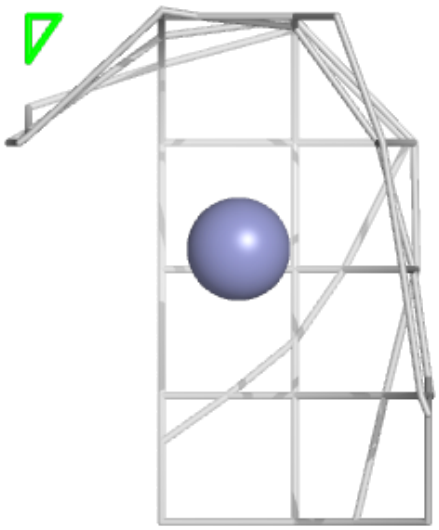


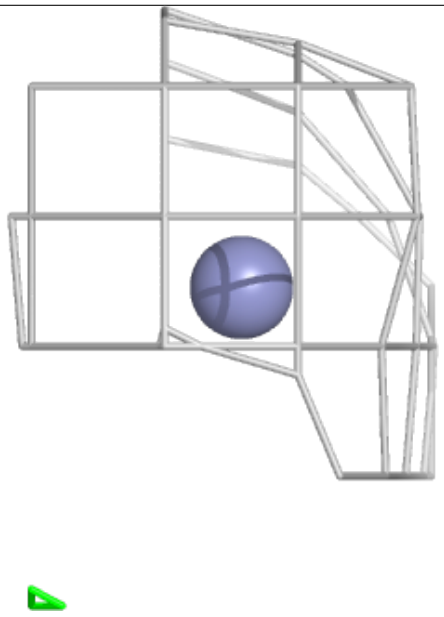


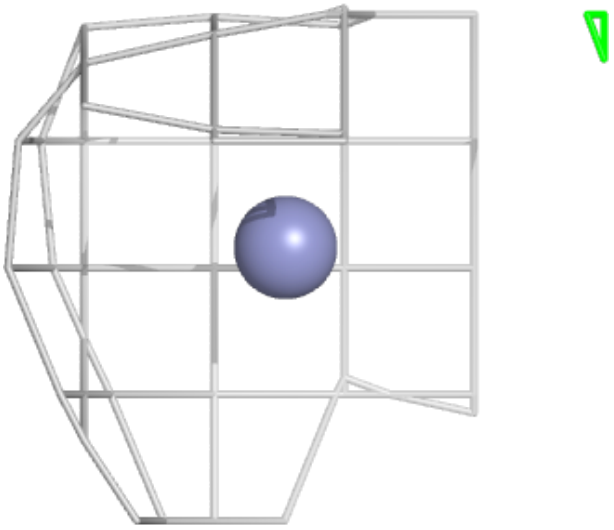


Electron density around ZN D 603:

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

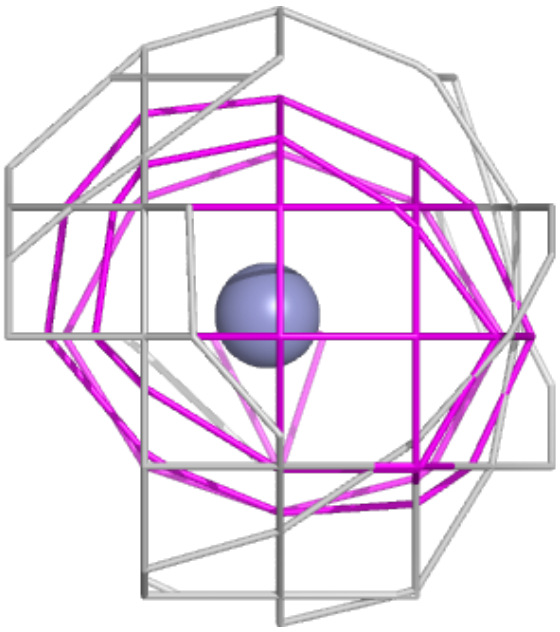


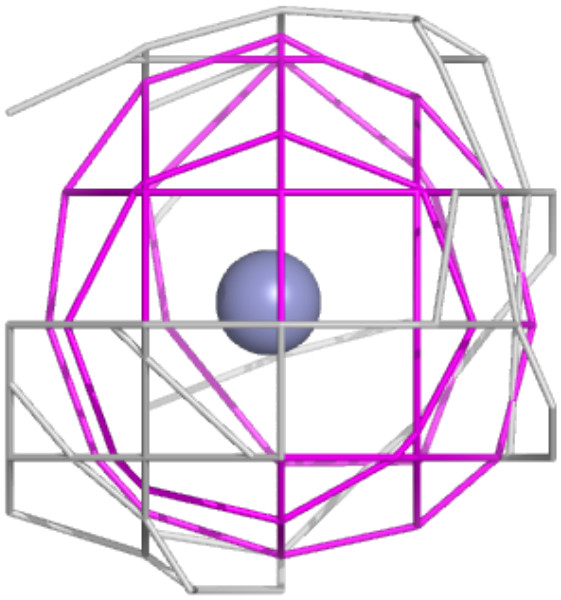


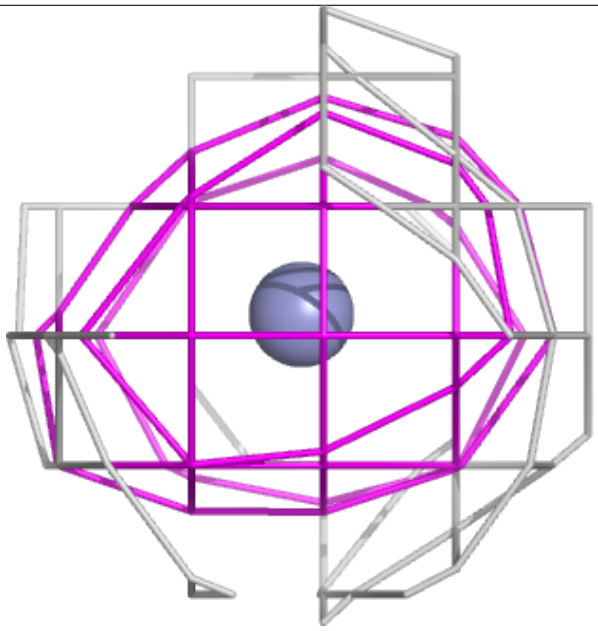


Electron density around ZN A 602:

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

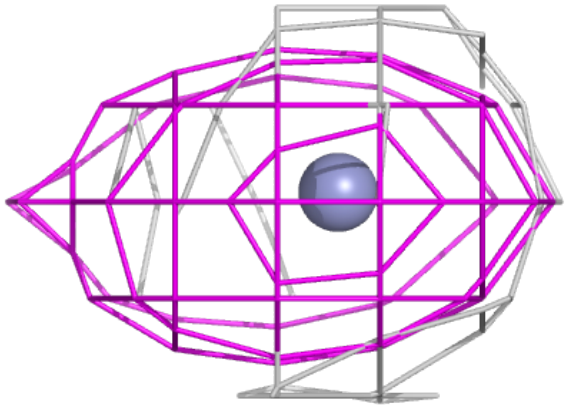


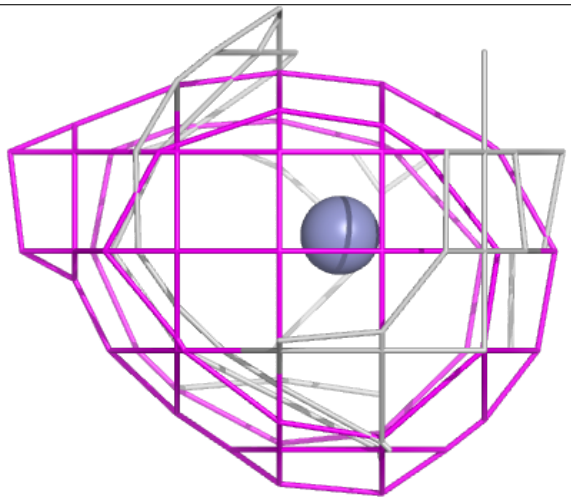


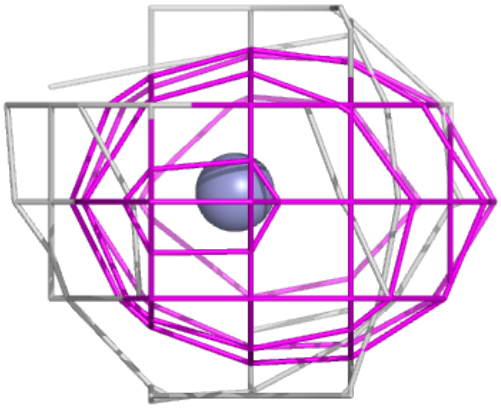


Electron density around ZN A 603:

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

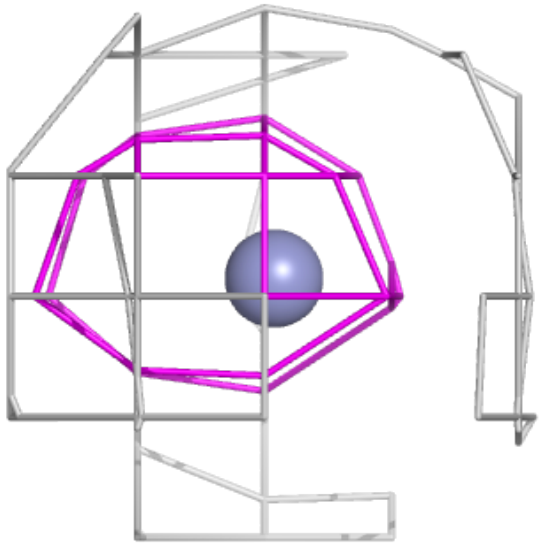


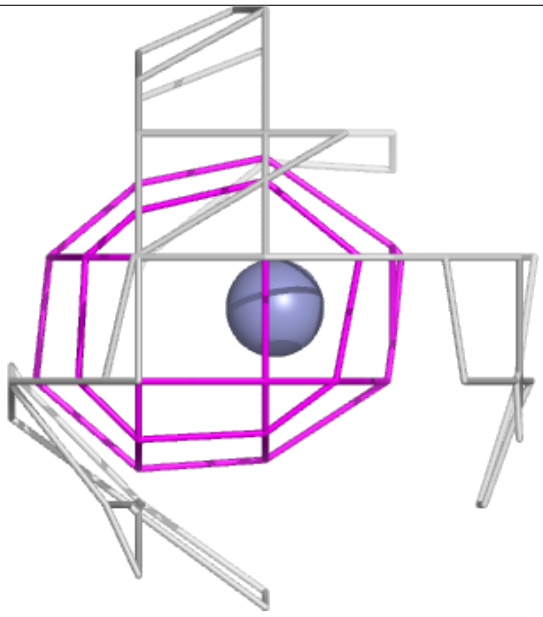


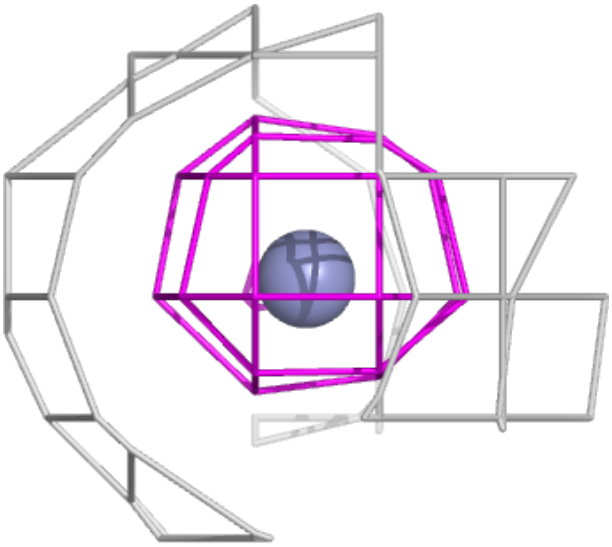


Electron density around ZN C 603:

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

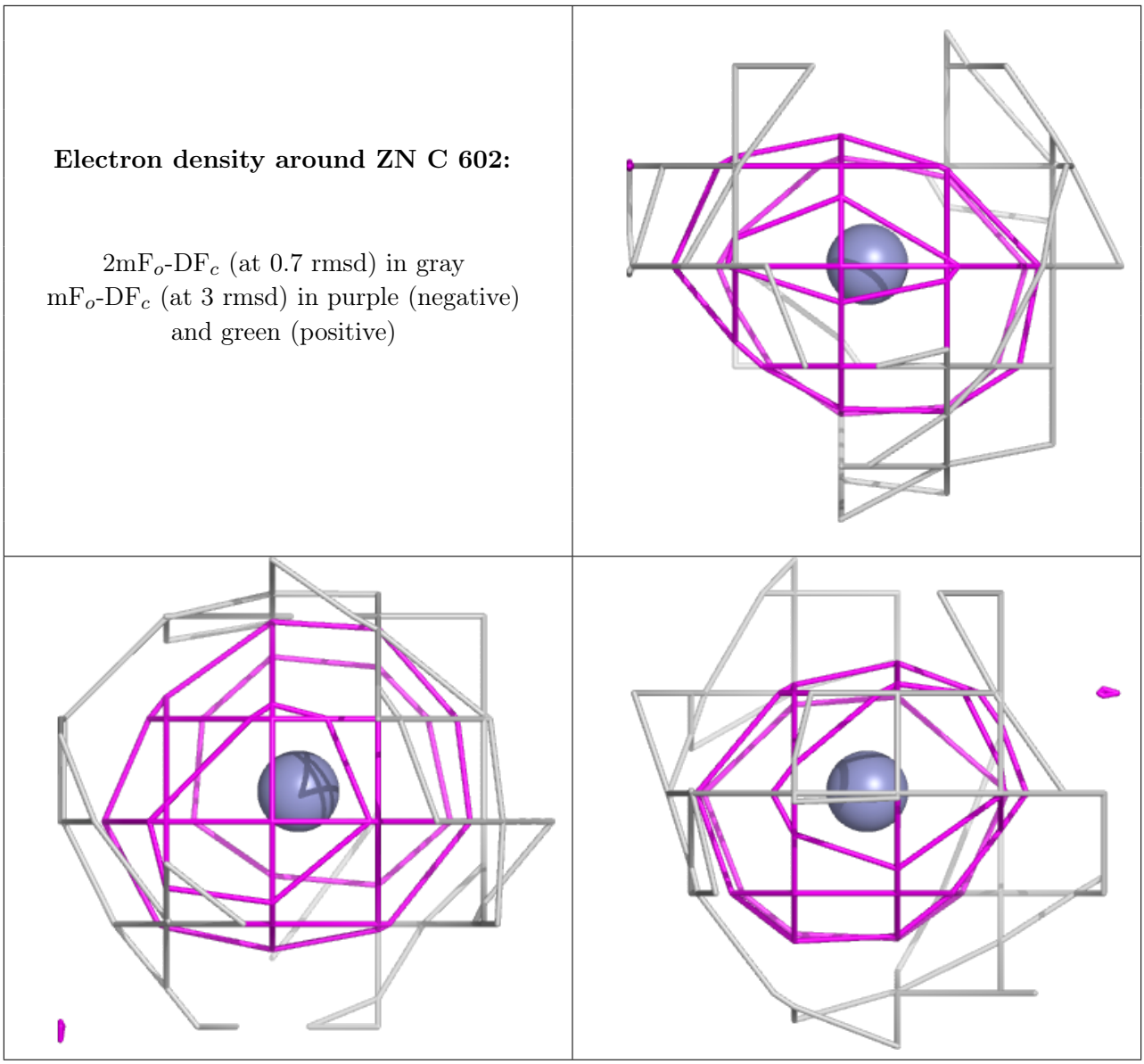






Electron density around ZN C 602:

$2mF_o$-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.