Full wwPDB X-ray Structure Validation Report (i)

Dec 9, 2024 – 06:05 pm GMT

PDB ID	:	9GBF
Title	:	X-RAY structure of PHDvC5HCH tandem domain of NSD2
Authors	:	Musco, G.; Cocomazzi, P.; Berardi, A.; Knapp, S.; Kramer, A.
Deposited on	:	2024-07-31
Resolution	:	1.76 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 (i) 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 (i)) were used in the production of this report:

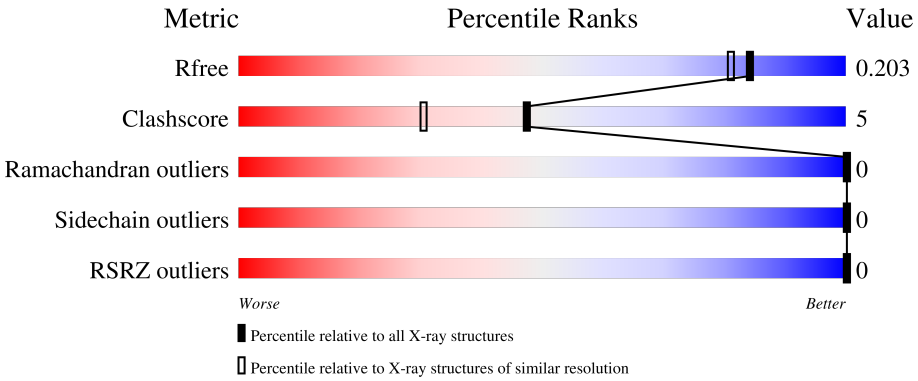
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 1.76 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}	164625	2888 (1.76-1.76)
Clashscore	180529	3097 (1.76-1.76)
Ramachandran outliers	177936	3072 (1.76-1.76)
Sidechain outliers	177891	3072 (1.76-1.76)
RSRZ outliers	164620	2887 (1.76-1.76)

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	103	61% 17% 21%
1	B	103	78% 10% 13%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2641 atoms, of which 1197 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 Histone-lysine N-methyltransferase NSD2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	81	Total	C	H	N	O	S	0	0	0
			1200	396	564	113	114	13			
1	B	90	Total	C	H	N	O	S	0	0	0
			1307	430	606	127	131	13			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1324	SER	CYS	engineered mutation	UNP O96028
B	1324	SER	CYS	engineered mutation	UNP O96028

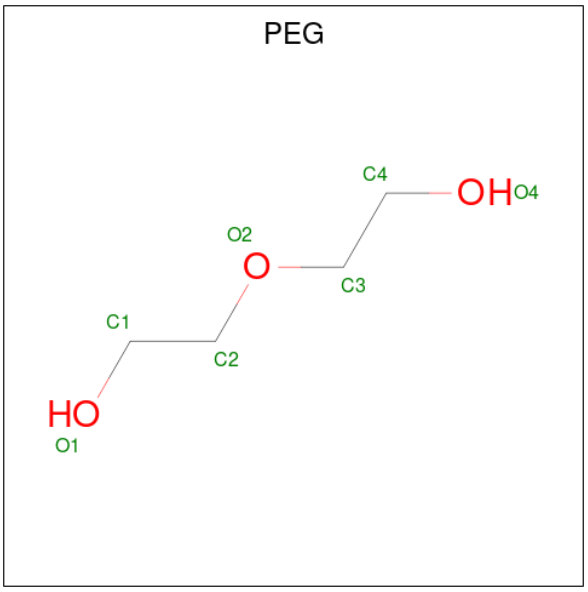
- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	4	Total	Zn	0	0
			4	4		
2	B	4	Total	Zn	0	0
			4	4		

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

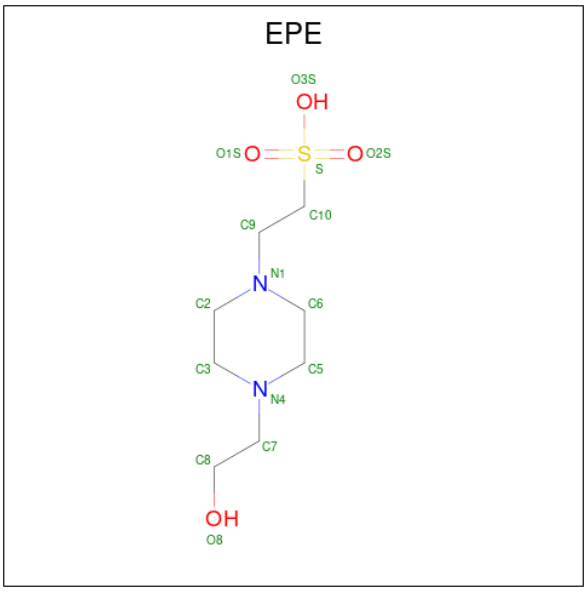
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Na	0	0
			1	1		
3	B	1	Total	Na	0	0
			1	1		

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
4	A	1	17	4	10	3	0	0

- Molecule 5 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	H	N	O	S		
5	B	1	32	8	17	2	4	1	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	25	Total O 25 25	0	0
6	B	50	Total O 50 50	0	0

3 Residue-property plots (i)

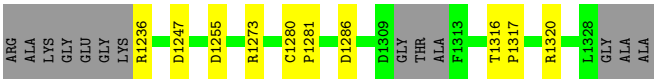
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Histone-lysine N-methyltransferase NSD2

Chain A:

• Molecule 1: Histone-lysine N-methyltransferase NSD2

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	36.13Å 45.25Å 60.65Å 90.00° 99.65° 90.00°	Depositor
Resolution (Å)	4.97 – 1.76 4.97 – 1.76	Depositor EDS
% Data completeness (in resolution range)	67.5 (4.97-1.76) 81.2 (4.97-1.76)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.07 (at 1.60Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.188 , 0.228 0.190 , 0.203	Depositor DCC
R_{free} test set	915 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	20.5	Xtriage
Anisotropy	0.022	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	1.28 , 105.1	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	0.96	EDS
Total number of atoms	2641	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.73% 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: PEG, NA, EPE, 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	0.97	1/655 (0.2%)	0.90	0/877
1	B	0.95	0/722	0.99	3/972 (0.3%)
All	All	0.96	1/1377 (0.1%)	0.95	3/1849 (0.2%)

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	2
All	All	0	3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1254	CYS	CB-SG	5.23	1.91	1.82

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	1247	ASP	CB-CG-OD2	-7.14	111.87	118.30
1	B	1286	ASP	CB-CG-OD1	5.44	123.20	118.30
1	B	1255	ASP	CB-CG-OD2	-5.17	113.64	118.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1273	ARG	Sidechain
1	B	1273	ARG	Sidechain
1	B	1320	ARG	Sidechain

5.2 Too-close contacts (i)

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	636	564	563	9	0
1	B	701	606	604	5	0
2	A	4	0	0	0	0
2	B	4	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	7	10	10	0	0
5	B	15	17	17	0	0
6	A	25	0	0	0	0
6	B	50	0	0	1	0
All	All	1444	1197	1194	14	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 5.

All (14) 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:1252:VAL:HG21	1:A:1265:LEU:HD21	1.76	0.65
1:B:1316:THR:HG22	1:B:1317:PRO:HD2	1.82	0.61
1:A:1255:ASP:O	1:A:1277:LYS:NZ	2.37	0.49
1:A:1252:VAL:HG21	1:A:1265:LEU:CD2	2.44	0.48
1:B:1316:THR:CG2	1:B:1317:PRO:HD2	2.44	0.46
1:B:1236:ARG:NH2	6:B:1504:HOH:O	2.47	0.46
1:A:1263:TYR:HB2	1:A:1268:LEU:HD11	1.99	0.45
1:B:1316:THR:HG22	1:B:1317:PRO:CD	2.47	0.44
1:A:1244:ARG:CZ	1:A:1300:PRO:HG2	2.49	0.42
1:A:1314:SER:O	1:A:1321:SER:HA	2.20	0.42
1:A:1242:CYS:HA	1:A:1262:ALA:O	2.20	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1284:HIS:HB2	1:A:1289:GLY:O	2.21	0.41
1:B:1280:CYS:HB2	1:B:1281:PRO:CD	2.51	0.41
1:A:1238:SER:OG	1:A:1253:LEU:CD2	2.69	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	75/103 (73%)	73 (97%)	2 (3%)	0	100	100
1	B	86/103 (84%)	85 (99%)	1 (1%)	0	100	100
All	All	161/206 (78%)	158 (98%)	3 (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	72/86 (84%)	72 (100%)	0	100	100
1	B	78/86 (91%)	78 (100%)	0	100	100
All	All	150/172 (87%)	150 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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 oligosaccharides in this entry.

5.6 Ligand geometry

Of 12 ligands modelled in this entry, 10 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	$\#\|Z\| > 2$
4	PEG	A	1406	-	6,6,6	0.30	0	5,5,5	0.56	0
5	EPE	B	1406	-	15,15,15	1.05	1 (6%)	18,20,20	2.17	4 (22%)

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	PEG	A	1406	-	-	1/4/4/4	-
5	EPE	B	1406	-	-	5/9/19/19	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	1406	EPE	C10-S	3.06	1.81	1.77

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	B	1406	EPE	O1S-S-C10	7.01	115.36	106.92
5	B	1406	EPE	C9-N1-C2	-3.22	102.99	111.23
5	B	1406	EPE	C2-C3-N4	2.68	116.14	110.64
5	B	1406	EPE	O2S-S-O1S	-2.13	106.58	113.95

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	B	1406	EPE	C8-C7-N4-C5
5	B	1406	EPE	C10-C9-N1-C6
5	B	1406	EPE	C8-C7-N4-C3
4	A	1406	PEG	O1-C1-C2-O2
5	B	1406	EPE	C9-C10-S-O3S
5	B	1406	EPE	C9-C10-S-O2S

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	81/103 (78%)	-0.61	0	100	100	18, 29, 56, 74	0
1	B	90/103 (87%)	-0.59	0	100	100	16, 28, 61, 71	0
All	All	171/206 (83%)	-0.60	0	100	100	16, 28, 59, 74	0

There are no RSRZ outliers to report.

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
4	PEG	A	1406	7/7	0.91	0.06	27,34,43,43	0
3	NA	A	1405	1/1	0.92	0.07	23,23,23,23	0
5	EPE	B	1406	15/15	0.95	0.06	22,32,40,42	0
2	ZN	A	1403	1/1	0.97	0.02	25,25,25,25	0
3	NA	B	1405	1/1	0.98	0.02	22,22,22,22	0
2	ZN	A	1402	1/1	0.99	0.01	25,25,25,25	0
2	ZN	B	1403	1/1	1.00	0.01	18,18,18,18	0

Continued on next page...

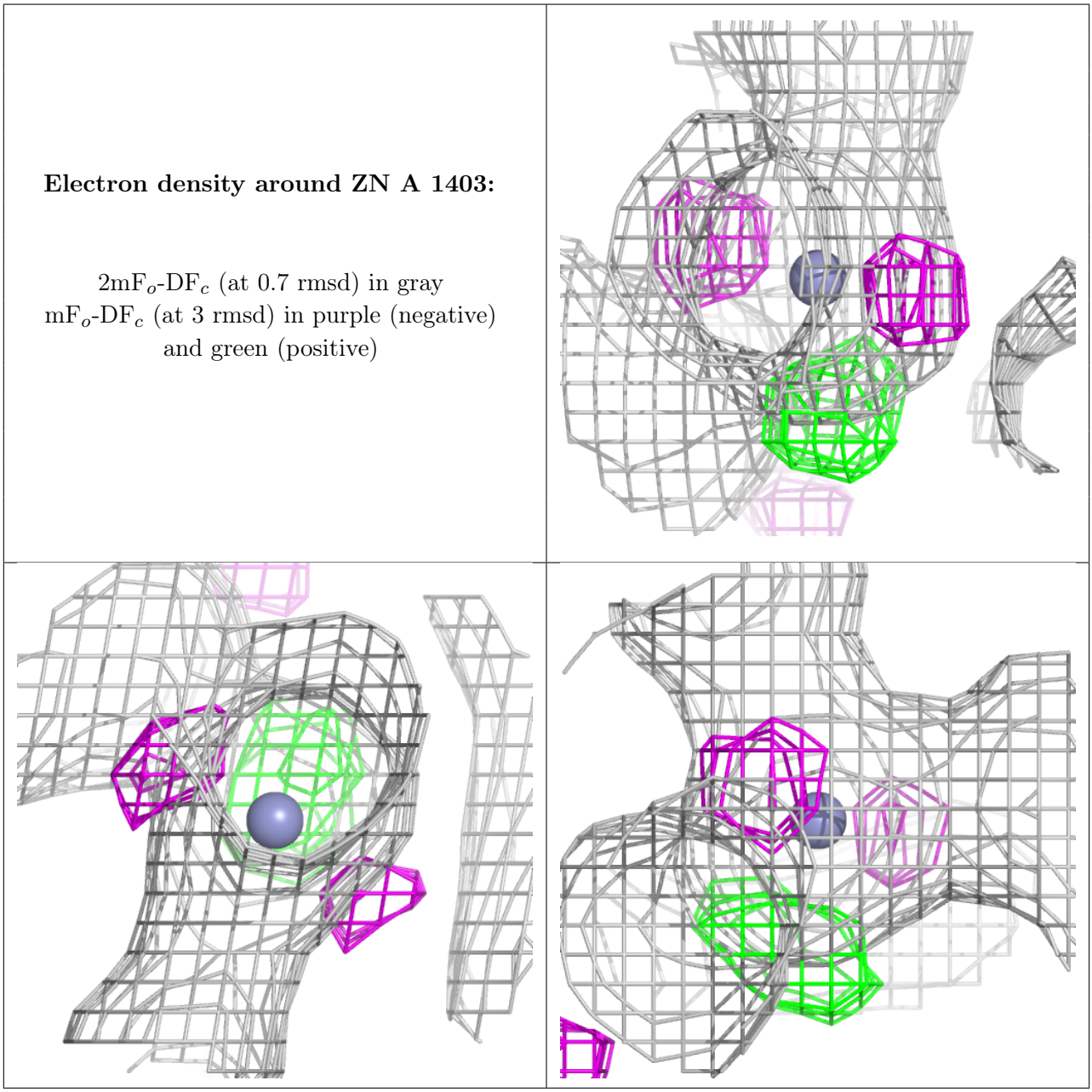
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	ZN	B	1404	1/1	1.00	0.01	18,18,18,18	0
2	ZN	A	1401	1/1	1.00	0.01	16,16,16,16	0
2	ZN	A	1404	1/1	1.00	0.01	18,18,18,18	0
2	ZN	B	1401	1/1	1.00	0.01	16,16,16,16	0
2	ZN	B	1402	1/1	1.00	0.01	14,14,14,14	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.

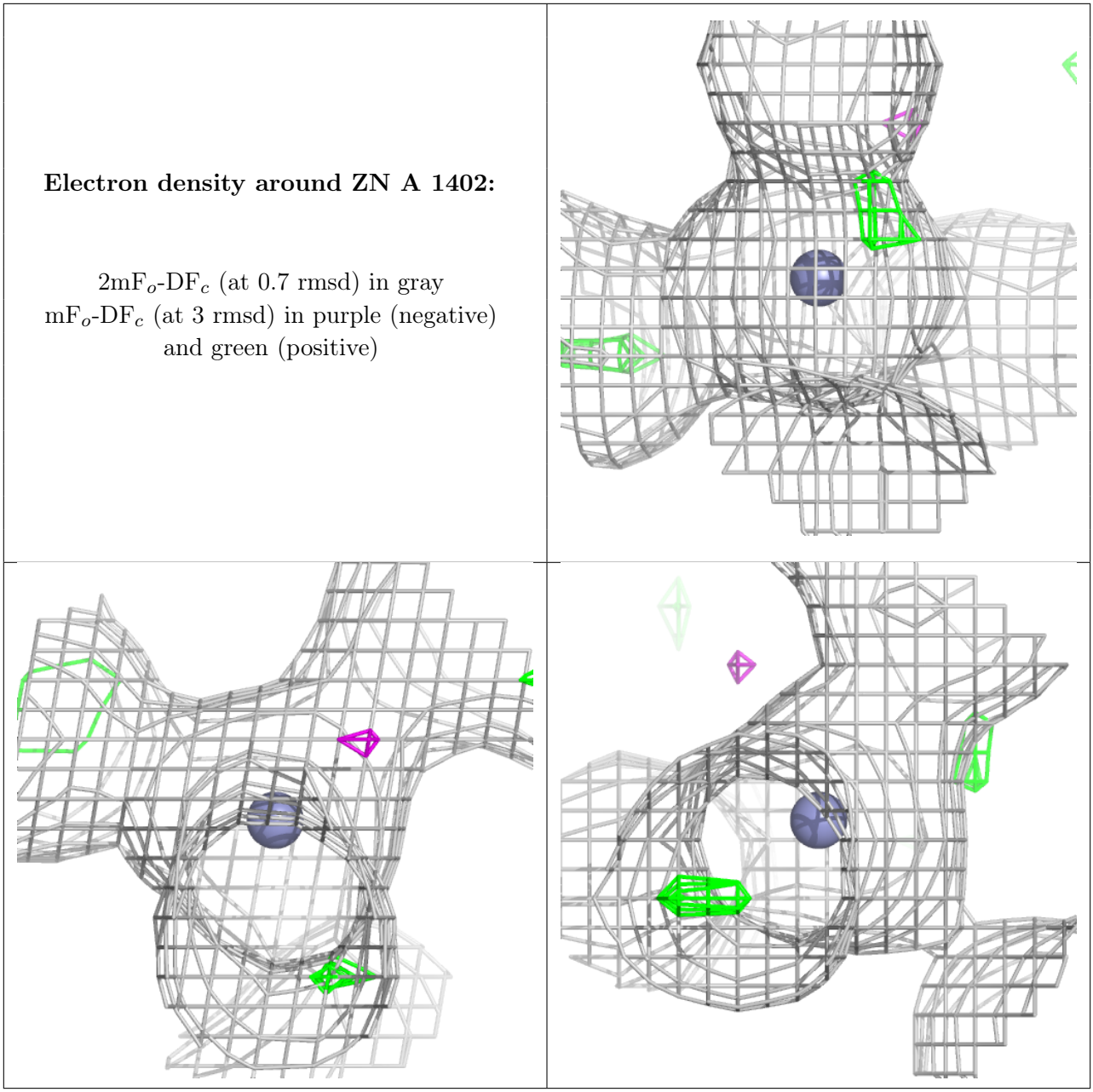
Electron density around ZN A 1403:

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



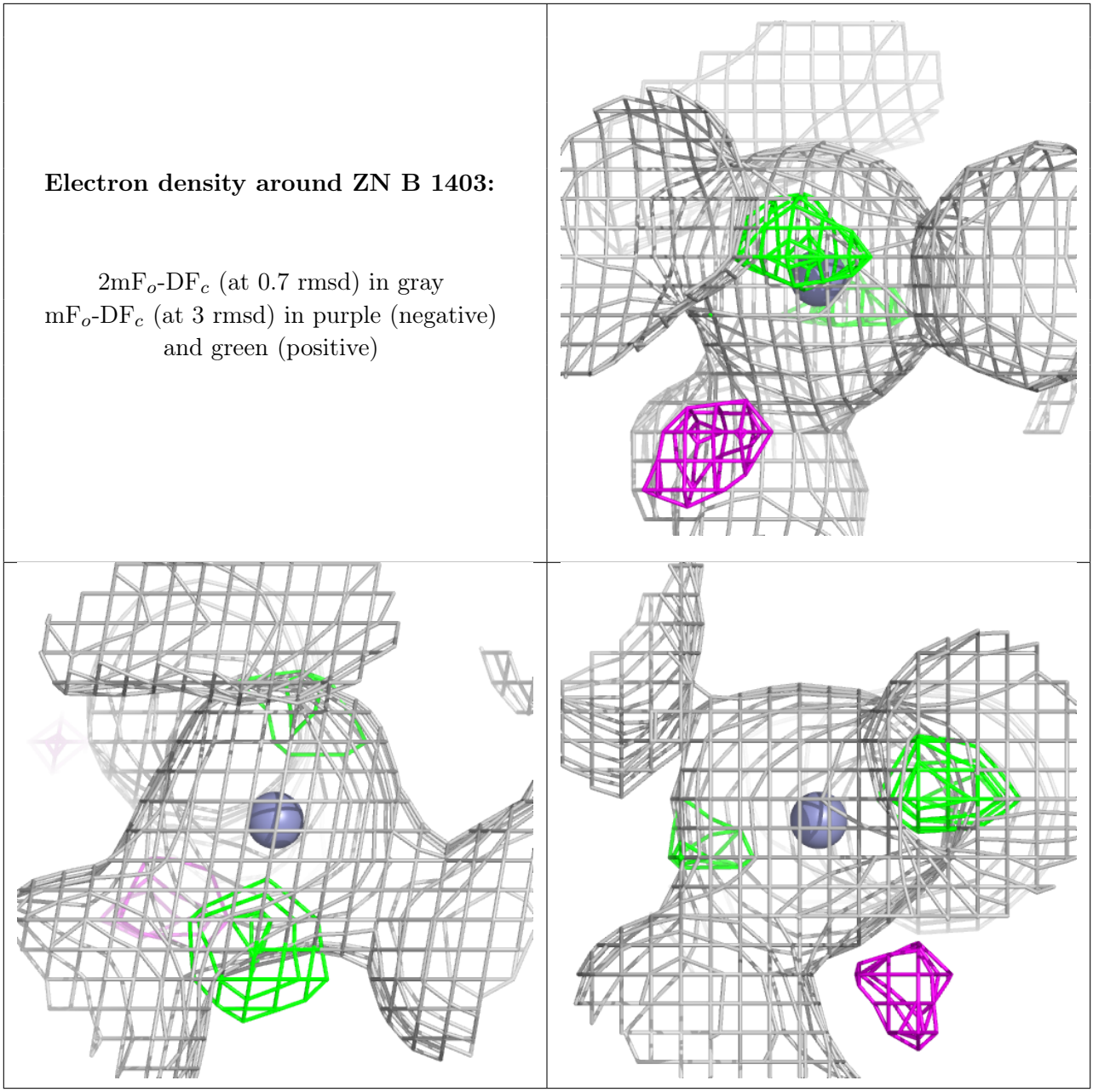
Electron density around ZN A 1402:

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



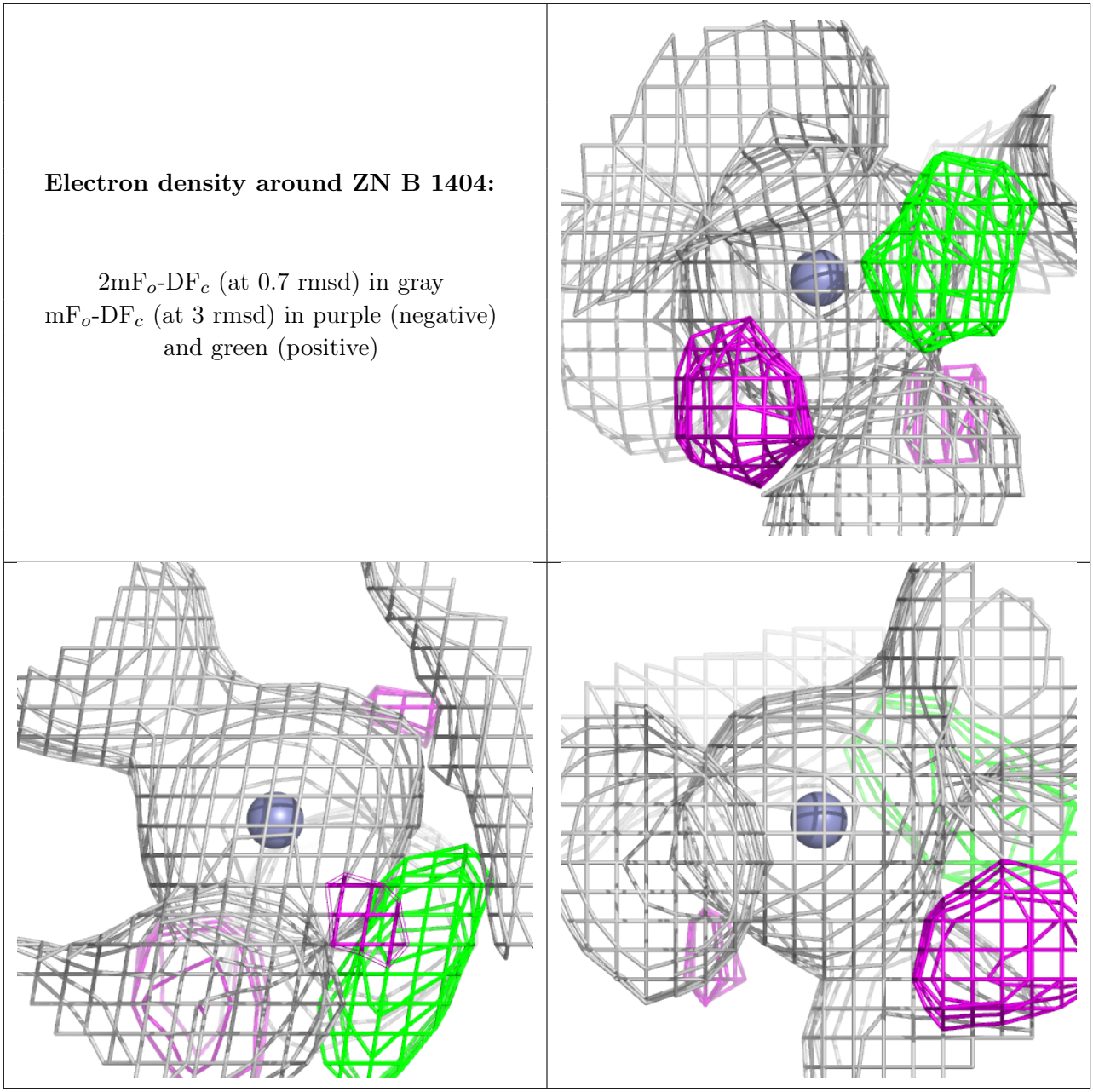
Electron density around ZN B 1403:

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



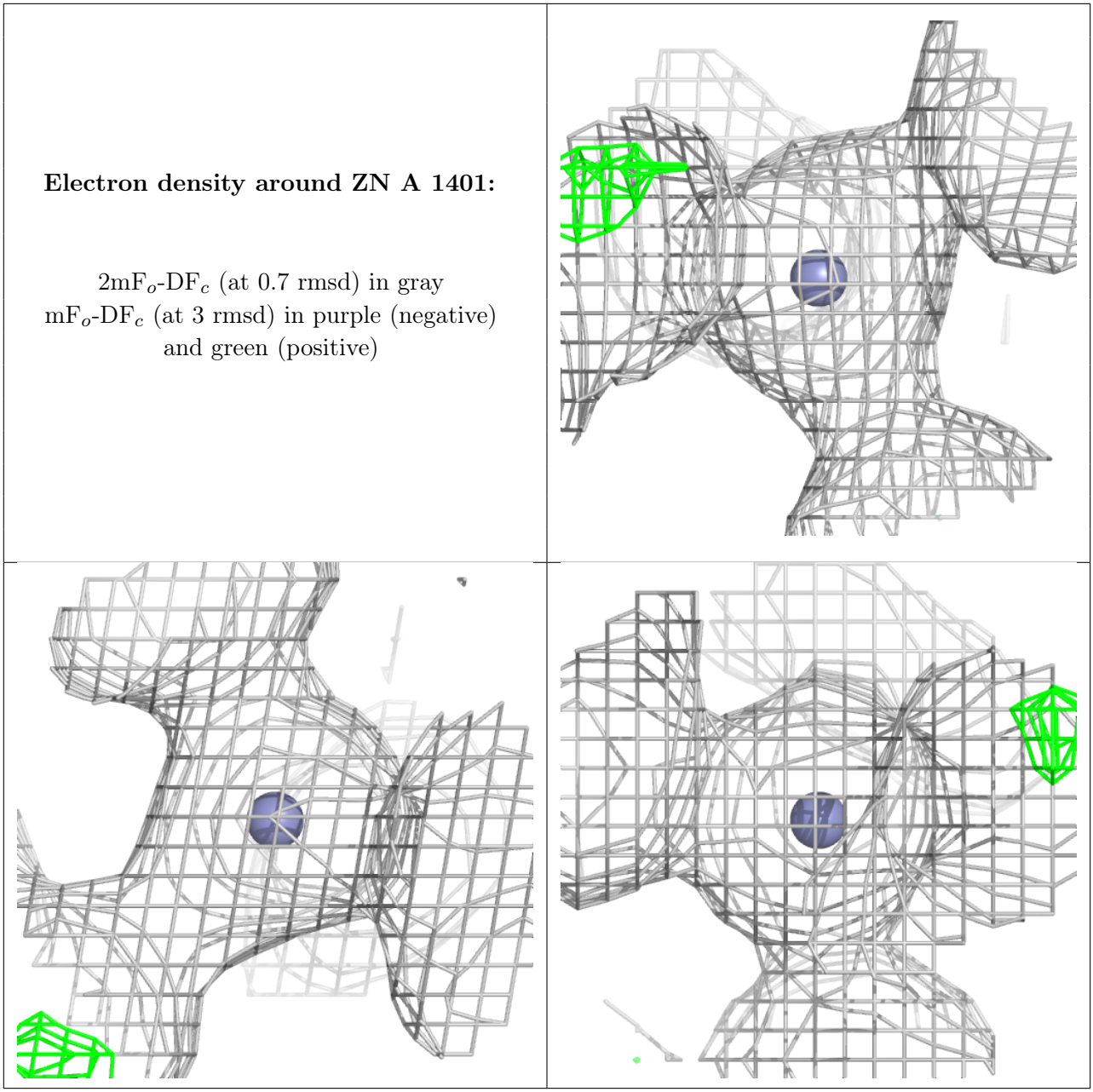
Electron density around ZN B 1404:

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



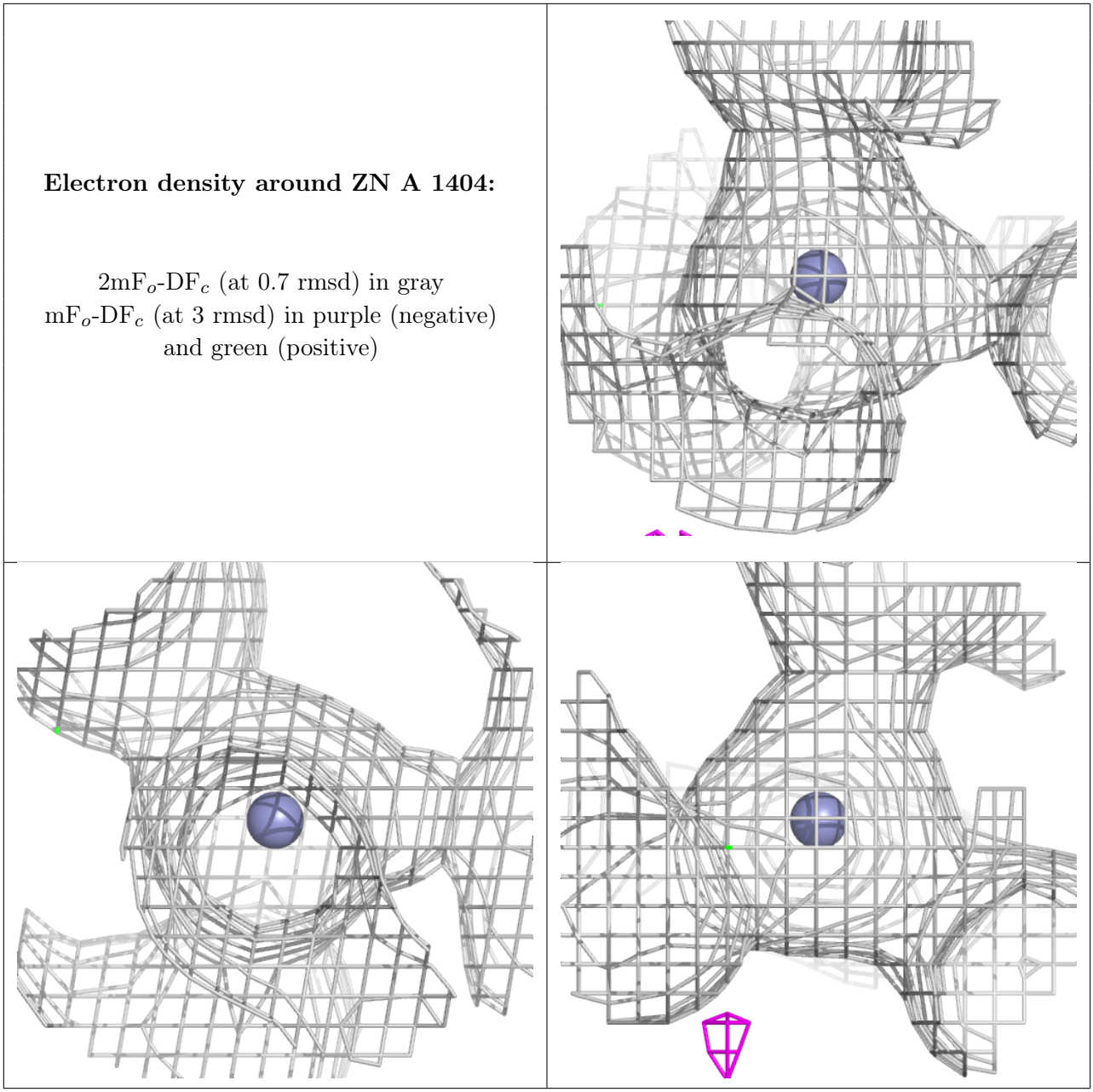
Electron density around ZN A 1401:

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



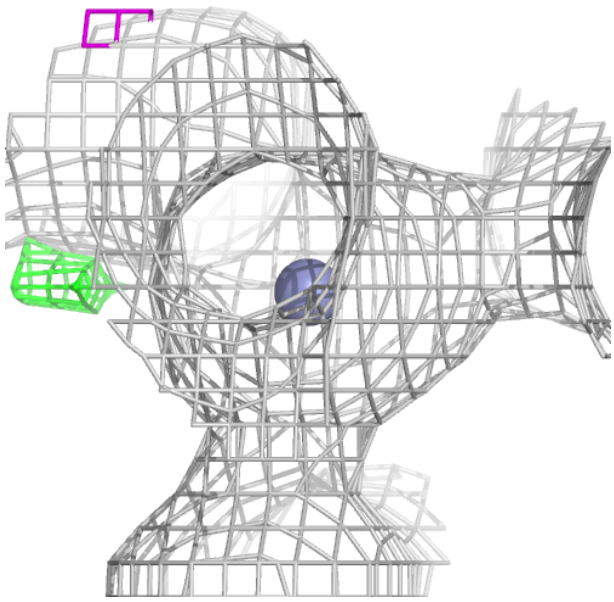
Electron density around ZN A 1404:

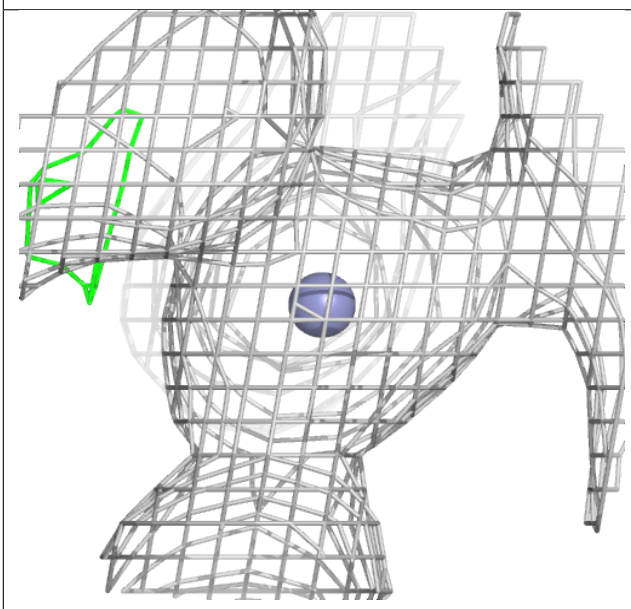
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

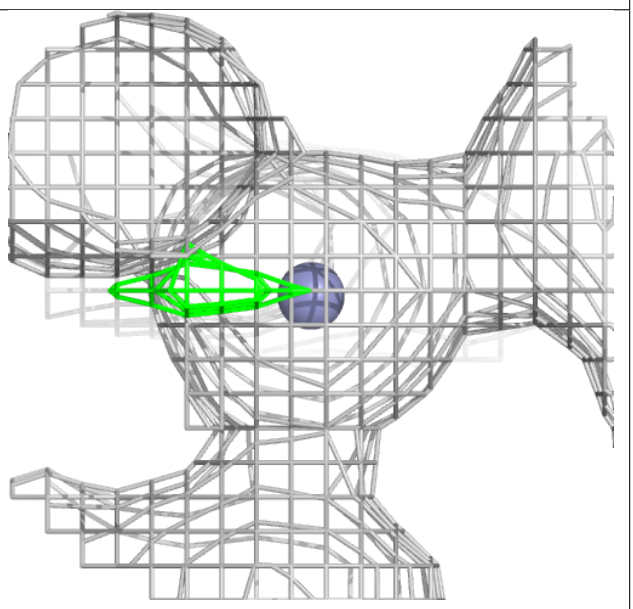


Electron density around ZN B 1401:

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

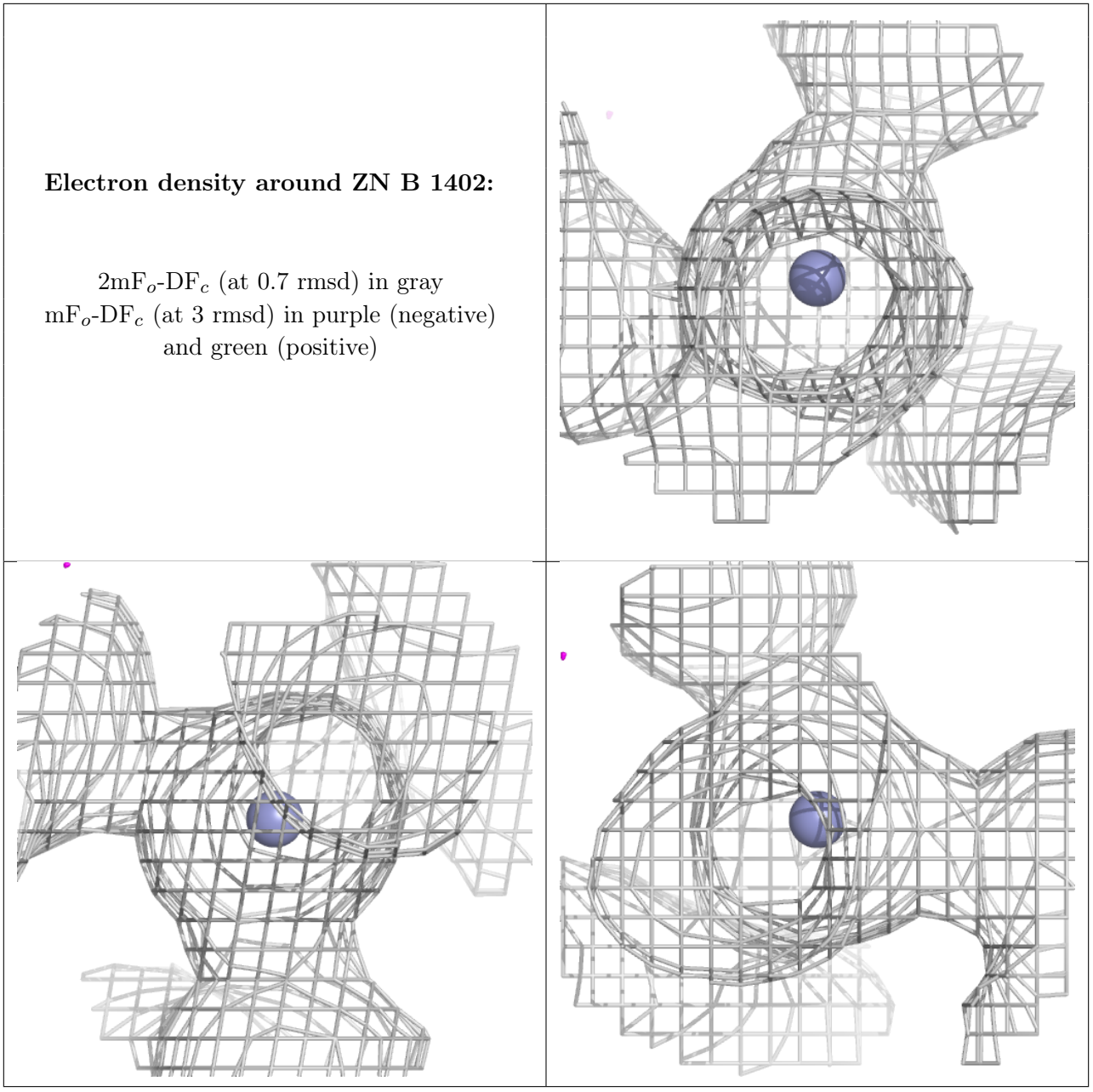






Electron density around ZN B 1402:

$2mF_o$-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.