Full wwPDB X-ray Structure Validation Report (i)

Dec 9, 2025 – 10:13 AM EST

PDB ID	:	9ZK1 / pdb_00009zk1
Title	:	Crystal structure of a calcium bound C2 domain containing protein from Trichomonas vaginalis (P21 form)
Authors	:	Seattle Structural Genomics Center for Infectious Disease (SSGCID)
Deposited on	:	2025-12-05
Resolution	:	1.51 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:

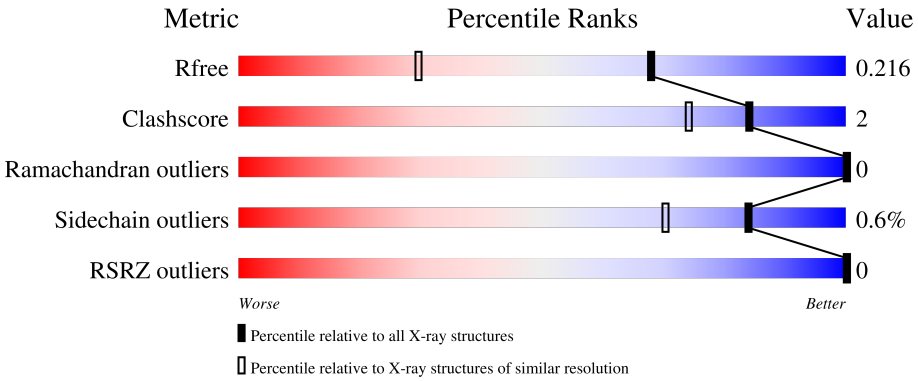
MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.47

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 1.51 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	5293 (1.54-1.50)
Clashscore	180529	5759 (1.54-1.50)
Ramachandran outliers	177936	5653 (1.54-1.50)
Sidechain outliers	177891	5650 (1.54-1.50)
RSRZ outliers	164620	5293 (1.54-1.50)

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	143	88% 5% • 6%
1	B	143	89% 5% 6%
1	C	143	90% 5% 6%
1	D	143	81% 6% 13%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 4755 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 C2 domain containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	134	Total	C	N	O	S	0	1	0
			1055	686	172	194	3			
1	B	134	Total	C	N	O	S	0	1	0
			1064	690	175	196	3			
1	C	135	Total	C	N	O	S	0	0	0
			1070	691	176	200	3			
1	D	125	Total	C	N	O	S	0	3	0
			1020	662	170	185	3			

There are 32 discrepancies between the modelled and reference sequences:

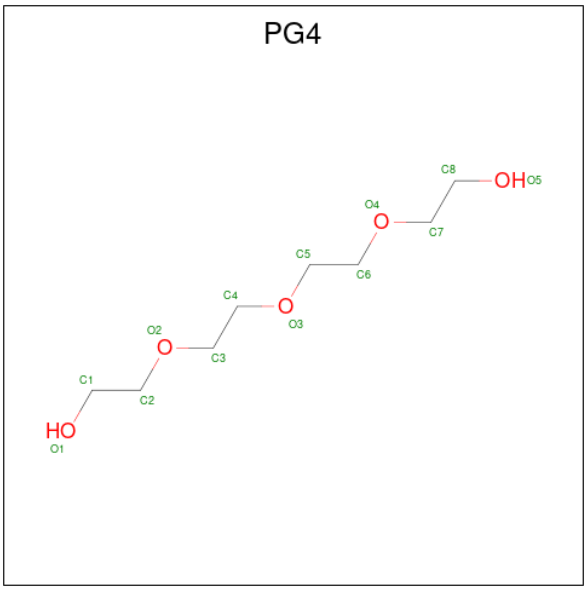
Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	MET	-	initiating methionine	UNP A2EZR3
A	-6	ALA	-	expression tag	UNP A2EZR3
A	-5	HIS	-	expression tag	UNP A2EZR3
A	-4	HIS	-	expression tag	UNP A2EZR3
A	-3	HIS	-	expression tag	UNP A2EZR3
A	-2	HIS	-	expression tag	UNP A2EZR3
A	-1	HIS	-	expression tag	UNP A2EZR3
A	0	HIS	-	expression tag	UNP A2EZR3
B	-7	MET	-	initiating methionine	UNP A2EZR3
B	-6	ALA	-	expression tag	UNP A2EZR3
B	-5	HIS	-	expression tag	UNP A2EZR3
B	-4	HIS	-	expression tag	UNP A2EZR3
B	-3	HIS	-	expression tag	UNP A2EZR3
B	-2	HIS	-	expression tag	UNP A2EZR3
B	-1	HIS	-	expression tag	UNP A2EZR3
B	0	HIS	-	expression tag	UNP A2EZR3
C	-7	MET	-	initiating methionine	UNP A2EZR3
C	-6	ALA	-	expression tag	UNP A2EZR3
C	-5	HIS	-	expression tag	UNP A2EZR3
C	-4	HIS	-	expression tag	UNP A2EZR3
C	-3	HIS	-	expression tag	UNP A2EZR3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-2	HIS	-	expression tag	UNP A2EZR3
C	-1	HIS	-	expression tag	UNP A2EZR3
C	0	HIS	-	expression tag	UNP A2EZR3
D	-7	MET	-	initiating methionine	UNP A2EZR3
D	-6	ALA	-	expression tag	UNP A2EZR3
D	-5	HIS	-	expression tag	UNP A2EZR3
D	-4	HIS	-	expression tag	UNP A2EZR3
D	-3	HIS	-	expression tag	UNP A2EZR3
D	-2	HIS	-	expression tag	UNP A2EZR3
D	-1	HIS	-	expression tag	UNP A2EZR3
D	0	HIS	-	expression tag	UNP A2EZR3

- Molecule 2 is TETRAETHYLENE GLYCOL (CCD ID: PG4) (formula: $C_8H_{18}O_5$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 13 8 5	0	0

- Molecule 3 is CALCIUM ION (CCD ID: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

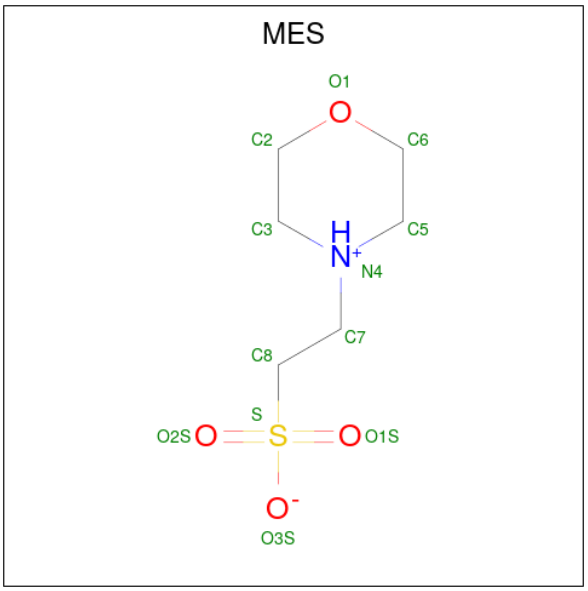
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total Ca 2 2	0	0
3	B	2	Total Ca 2 2	0	0

Continued on next page...

Continued from previous page...

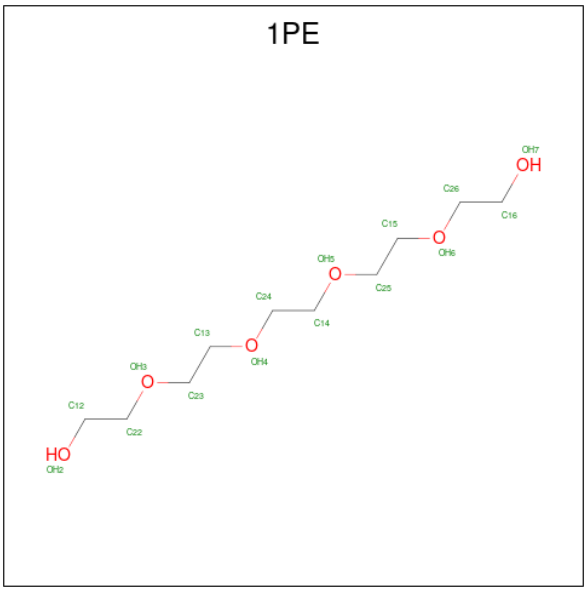
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	3	Total Ca 3 3	0	0
3	D	3	Total Ca 3 3	0	0

- Molecule 4 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (CCD ID: MES) (formula: $C_6H_{13}NO_4S$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C N O S 12 6 1 4 1	0	0

- Molecule 5 is PENTAETHYLENE GLYCOL (CCD ID: 1PE) (formula: $C_{10}H_{22}O_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	1	Total C O 16 10 6	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	119	Total O 119 119	0	0
6	B	138	Total O 138 138	0	0
6	C	104	Total O 104 104	0	0
6	D	134	Total O 134 134	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: C2 domain containing protein

Chain A:

• Molecule 1: C2 domain containing protein

Chain B:

• Molecule 1: C2 domain containing protein

Chain C:

• Molecule 1: C2 domain containing protein

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	35.95Å 88.06Å 97.20Å 90.00° 90.85° 90.00°	Depositor
Resolution (Å)	44.03 – 1.51 44.03 – 1.51	Depositor EDS
% Data completeness (in resolution range)	98.9 (44.03-1.51) 98.9 (44.03-1.51)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.96 (at 1.51Å)	Xtriage
Refinement program	PHENIX (2.0_5765: ???)	Depositor
R, R_{free}	0.188 , 0.214 0.192 , 0.216	Depositor DCC
R_{free} test set	4539 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	21.9	Xtriage
Anisotropy	0.839	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 31.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.077 for h,-k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	4755	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.52% 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: 1PE, PG4, CA, MES

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	0.31	0/1083	0.48	0/1474
1	B	0.30	0/1092	0.52	0/1484
1	C	0.27	0/1095	0.47	0/1489
1	D	0.28	0/1052	0.51	0/1427
All	All	0.29	0/4322	0.50	0/5874

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	1055	0	1070	6	0
1	B	1064	0	1084	4	0
1	C	1070	0	1076	4	0
1	D	1020	0	1058	5	0
2	A	13	0	18	1	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
3	C	3	0	0	0	0
3	D	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	12	0	13	0	0
5	C	16	0	22	0	0
6	A	119	0	0	0	0
6	B	138	0	0	0	0
6	C	104	0	0	0	0
6	D	134	0	0	0	0
All	All	4755	0	4341	19	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 2.

All (19) 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:90:ILE:HG23	1:A:123:ILE:HD11	1.76	0.68
1:D:85:LEU:C	1:D:85:LEU:HD12	2.25	0.61
1:A:90:ILE:CG2	1:A:123:ILE:HD11	2.30	0.61
1:D:71:LEU:HD12	1:D:105:LEU:HD13	1.82	0.60
1:D:59:LEU:HD13	1:D:62:VAL:HG12	1.86	0.57
1:C:71:LEU:HD12	1:C:105:LEU:HD13	1.87	0.57
1:A:28:ILE:HG23	1:A:37:ILE:HG23	1.88	0.56
1:C:85:LEU:C	1:C:85:LEU:HD12	2.30	0.56
1:A:85:LEU:HD12	1:A:85:LEU:C	2.30	0.56
1:B:23[A]:CYS:SG	1:B:83:ILE:HD11	2.49	0.52
1:C:66:THR:HG21	1:C:87:LYS:HE3	1.91	0.52
1:C:59:LEU:HD13	1:C:62:VAL:HG12	1.94	0.51
1:D:95:LEU:HD23	1:D:96:GLY:N	2.31	0.46
1:B:61:ASP:CG	1:B:64:ILE:HD12	2.43	0.43
1:A:86:ILE:HD13	1:A:103:TYR:HB3	2.01	0.43
1:A:93:PHE:CE1	2:A:201:PG4:H41	2.55	0.42
1:B:90:ILE:HG23	1:B:123:ILE:HD11	2.02	0.42
1:D:85:LEU:HD12	1:D:85:LEU:O	2.21	0.41
1:B:23[B]:CYS:HB2	1:B:46:THR:O	2.21	0.41

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	133/143 (93%)	128 (96%)	5 (4%)	0	100	100
1	B	133/143 (93%)	129 (97%)	4 (3%)	0	100	100
1	C	133/143 (93%)	130 (98%)	3 (2%)	0	100	100
1	D	126/143 (88%)	123 (98%)	3 (2%)	0	100	100
All	All	525/572 (92%)	510 (97%)	15 (3%)	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	118/130 (91%)	117 (99%)	1 (1%)	79	62
1	B	120/130 (92%)	120 (100%)	0	100	100
1	C	120/130 (92%)	120 (100%)	0	100	100
1	D	117/130 (90%)	115 (98%)	2 (2%)	56	28
All	All	475/520 (91%)	472 (99%)	3 (1%)	84	70

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

Mol	Chain	Res	Type
1	A	85	LEU
1	D	1	MET
1	D	37	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	92	GLN
1	C	132	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 13 ligands modelled in this entry, 10 are monoatomic - leaving 3 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	1PE	C	201	-	15,15,15	0.30	0	14,14,14	0.31	0
2	PG4	A	201	-	12,12,12	0.10	0	11,11,11	0.14	0
4	MES	B	301	-	12,12,12	1.10	1 (8%)	15,16,16	0.79	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	1PE	C	201	-	-	4/13/13/13	-
2	PG4	A	201	-	-	0/10/10/10	-
4	MES	B	301	-	-	0/6/14/14	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	301	MES	C8-S	2.86	1.81	1.77

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	C	201	1PE	OH5-C14-C24-OH4
5	C	201	1PE	C13-C23-OH3-C22
5	C	201	1PE	OH4-C13-C23-OH3
5	C	201	1PE	C14-C24-OH4-C13

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	201	PG4	1	0

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	A	134/143 (93%)	-1.03	0	100	100	15, 30, 57, 82	1 (0%)
1	B	134/143 (93%)	-1.02	0	100	100	14, 29, 57, 66	1 (0%)
1	C	135/143 (94%)	-0.78	0	100	100	22, 36, 60, 73	0
1	D	125/143 (87%)	-0.93	0	100	100	16, 28, 50, 69	3 (2%)
All	All	528/572 (92%)	-0.94	0	100	100	14, 31, 57, 82	5 (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 oligosaccharides 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	PG4	A	201	13/13	0.97	0.06	46,50,65,66	0
4	MES	B	301	12/12	0.97	0.05	39,47,76,81	0
5	1PE	C	201	16/16	0.98	0.05	34,42,55,56	0
3	CA	D	203	1/1	0.99	0.11	39,39,39,39	0

Continued on next page...

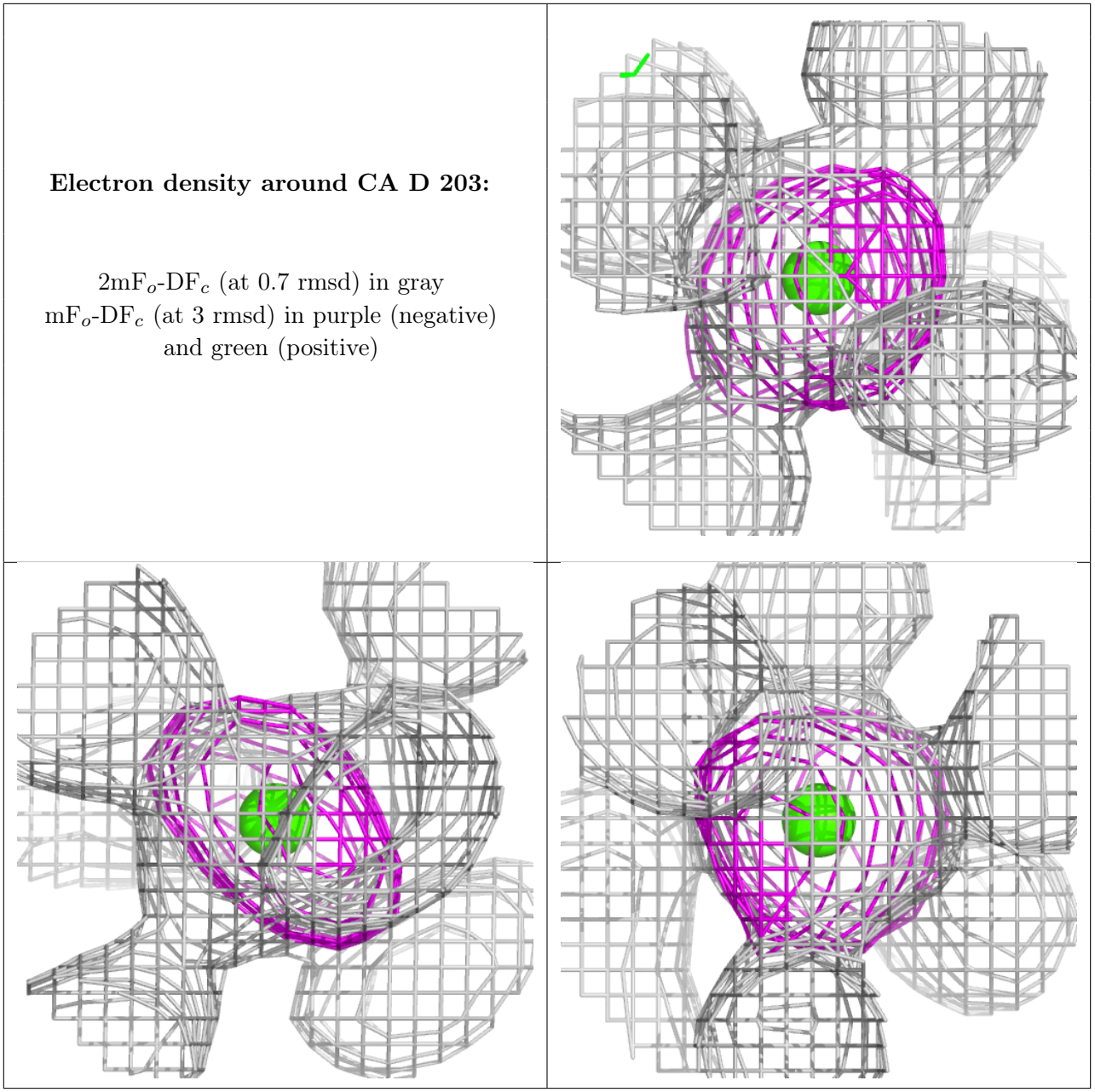
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	CA	C	202	1/1	0.99	0.04	30,30,30,30	0
3	CA	C	203	1/1	0.99	0.04	30,30,30,30	0
3	CA	A	203	1/1	1.00	0.01	21,21,21,21	0
3	CA	C	204	1/1	1.00	0.12	37,37,37,37	0
3	CA	D	201	1/1	1.00	0.02	22,22,22,22	0
3	CA	D	202	1/1	1.00	0.01	21,21,21,21	0
3	CA	B	302	1/1	1.00	0.01	20,20,20,20	0
3	CA	B	303	1/1	1.00	0.01	20,20,20,20	0
3	CA	A	202	1/1	1.00	0.01	22,22,22,22	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.

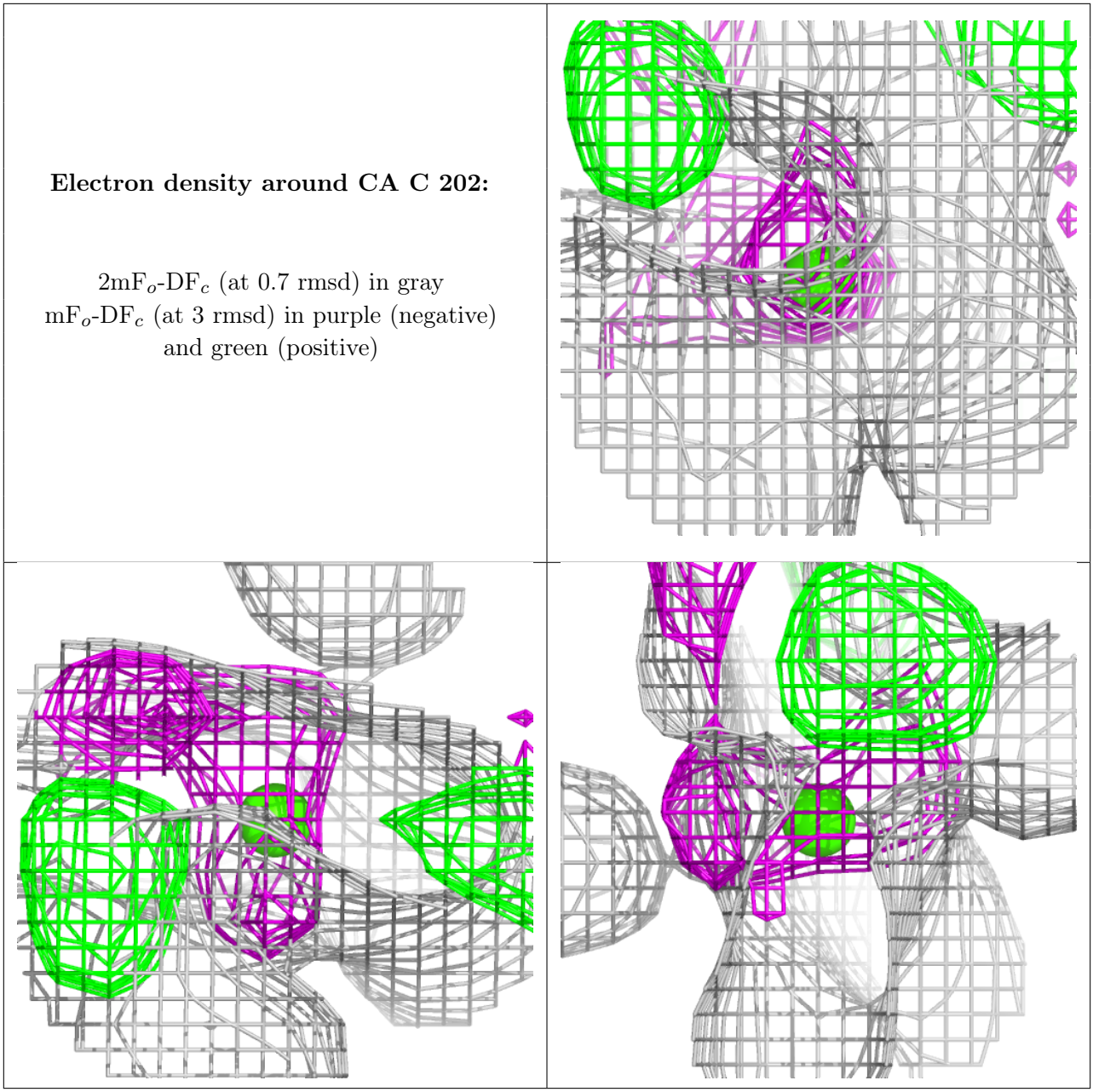
Electron density around CA D 203:

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



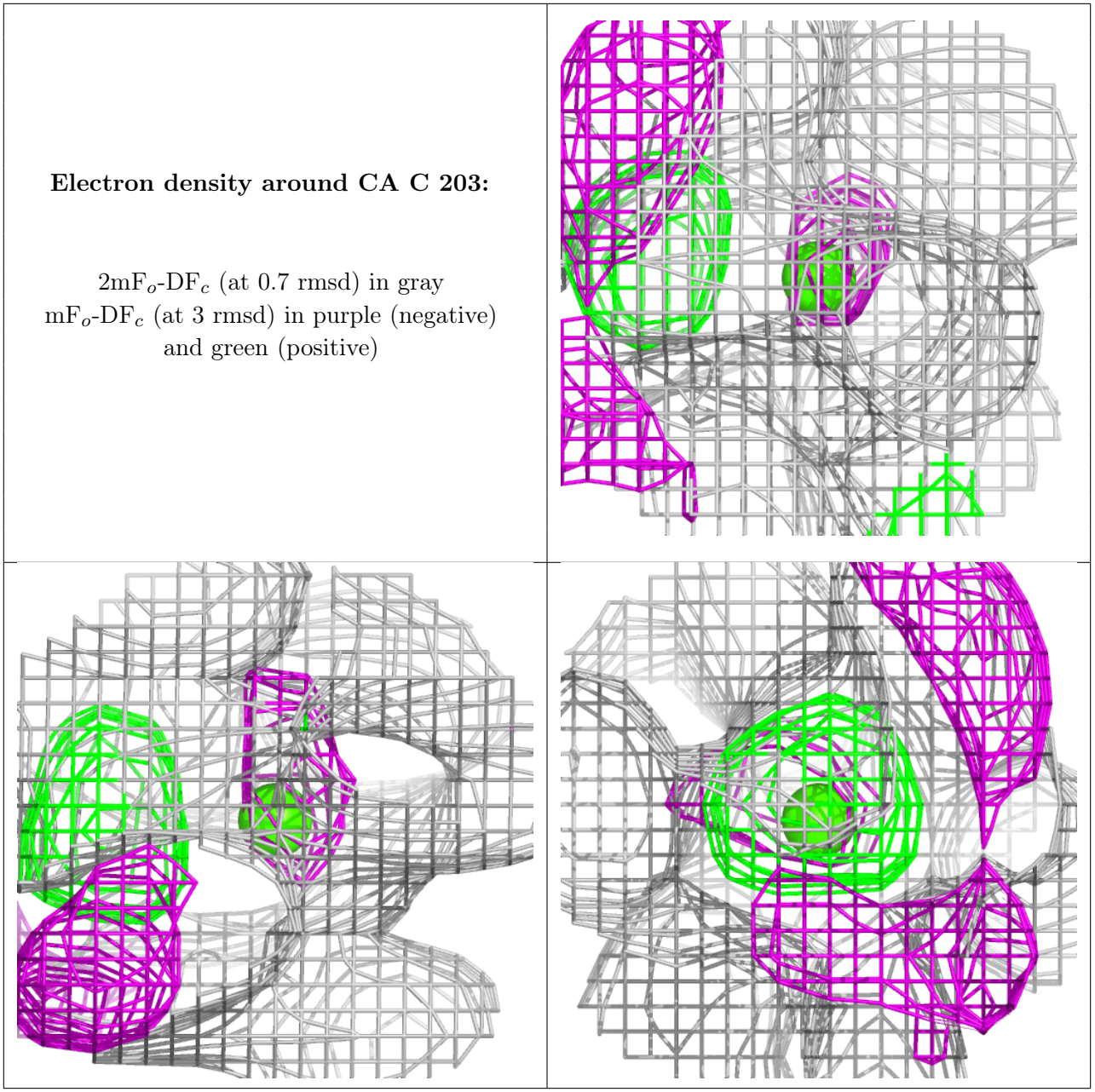
Electron density around CA C 202:

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



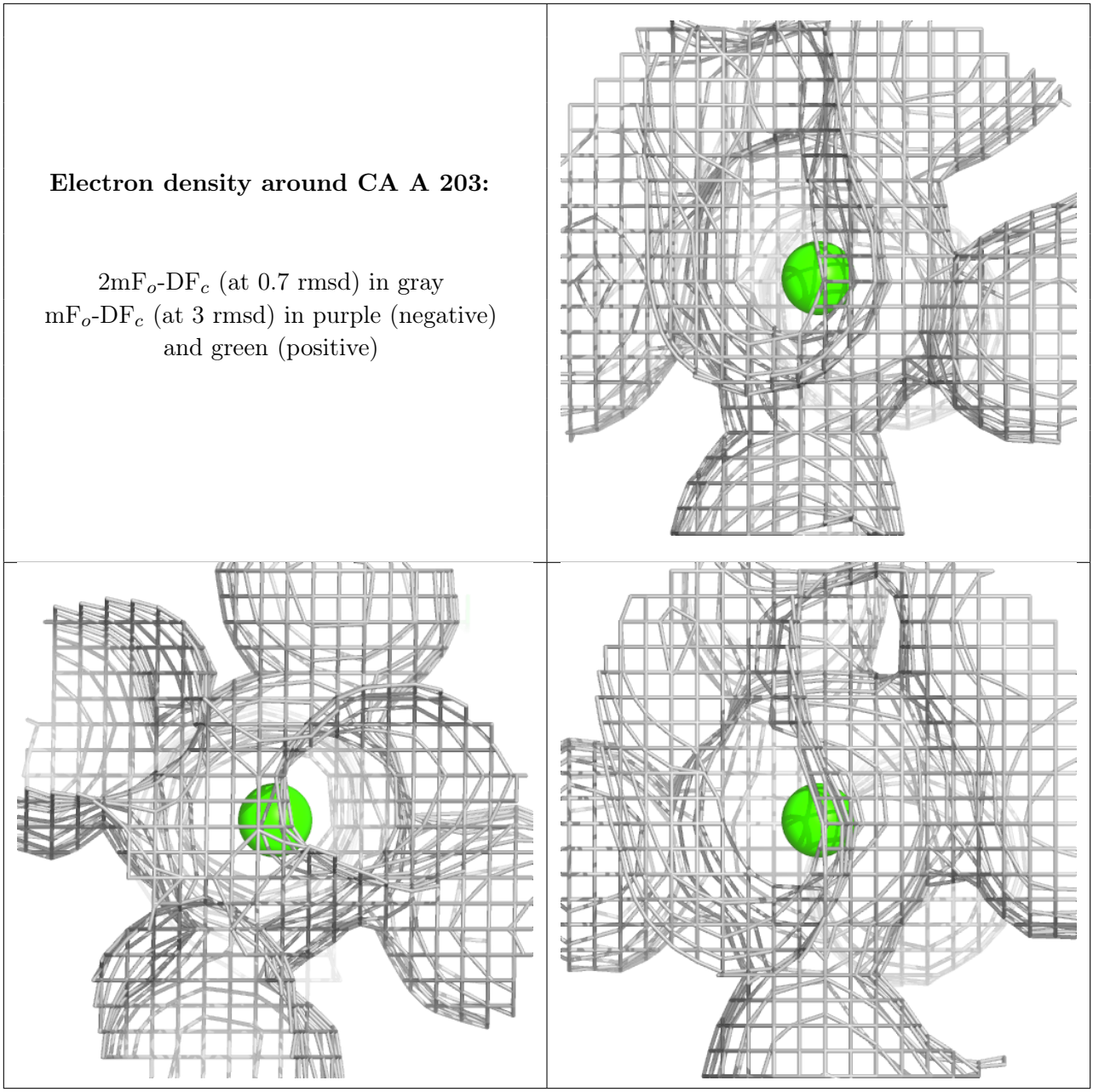
Electron density around CA C 203:

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



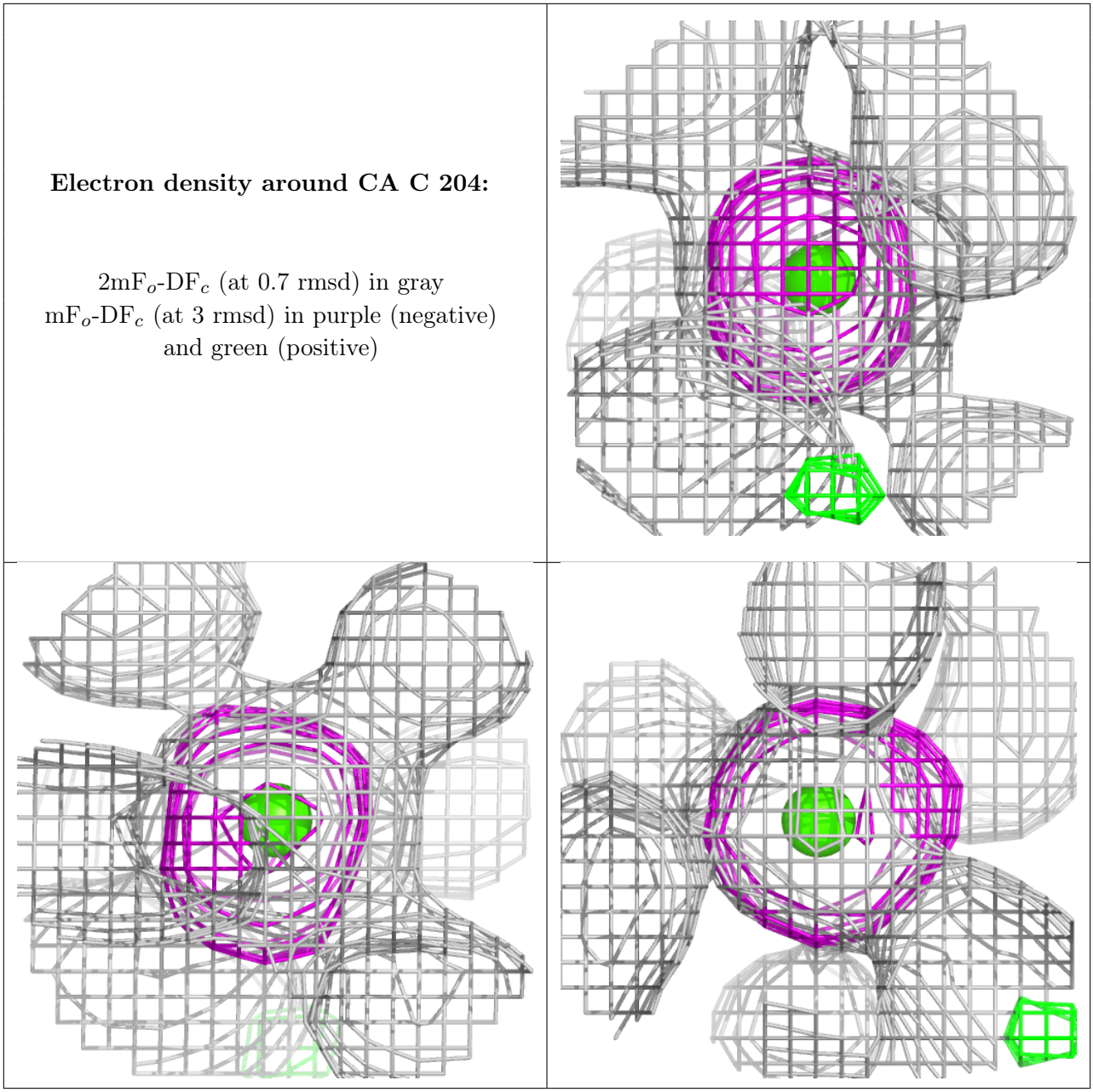
Electron density around CA A 203:

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



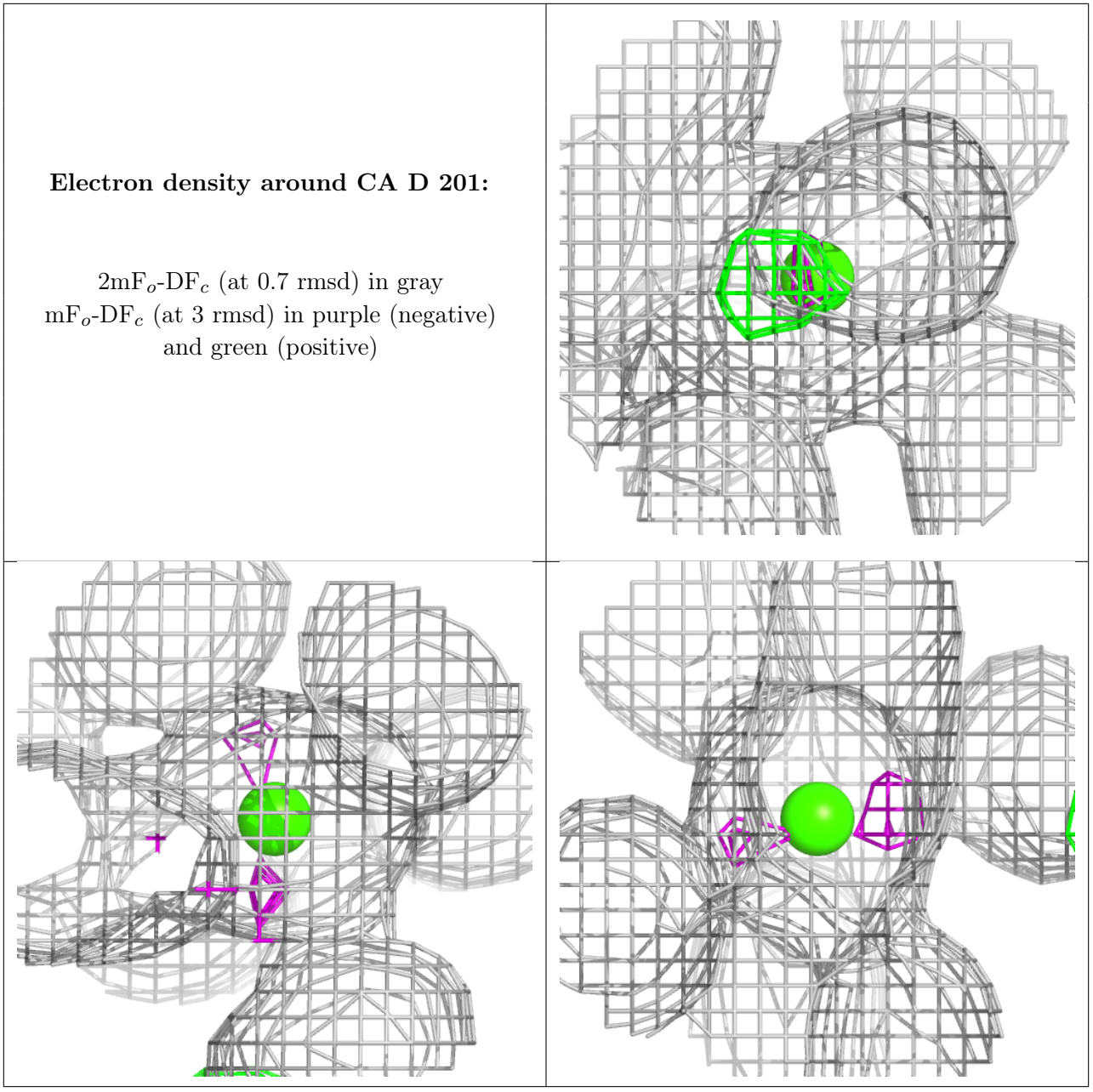
Electron density around CA C 204:

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



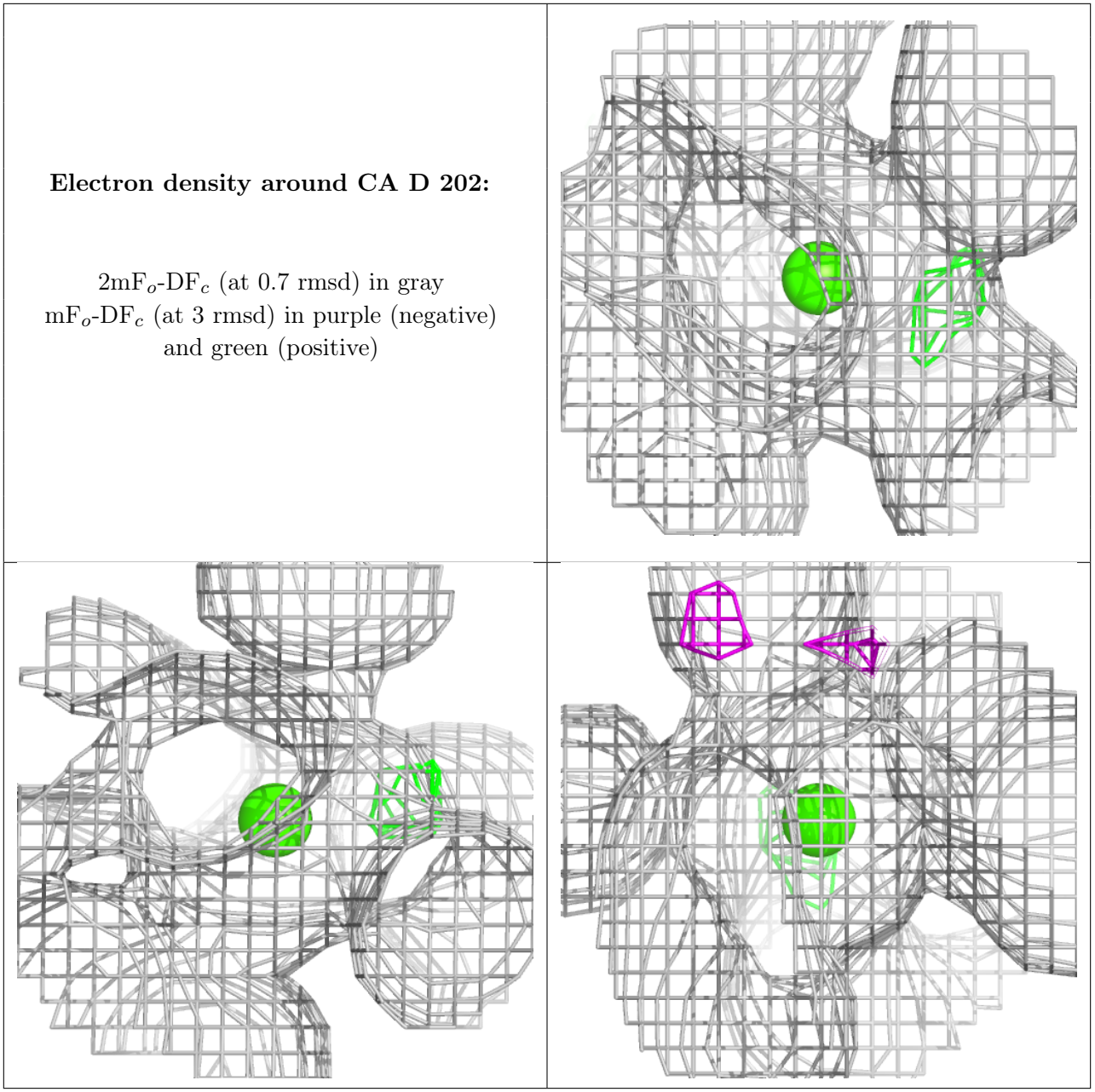
Electron density around CA D 201:

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



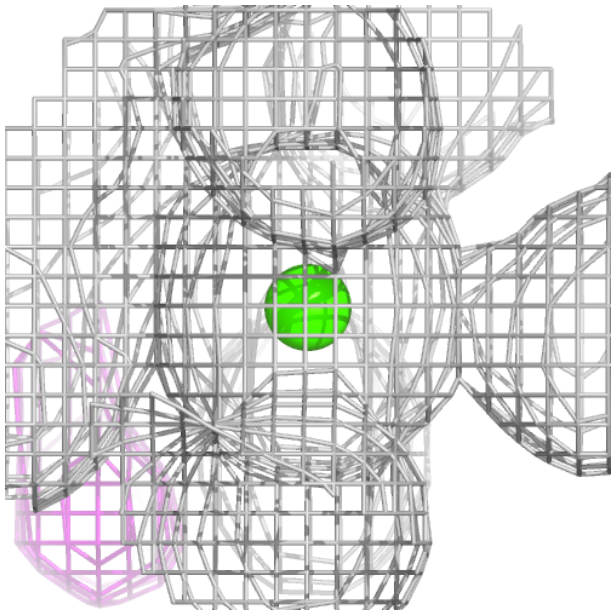
Electron density around CA D 202:

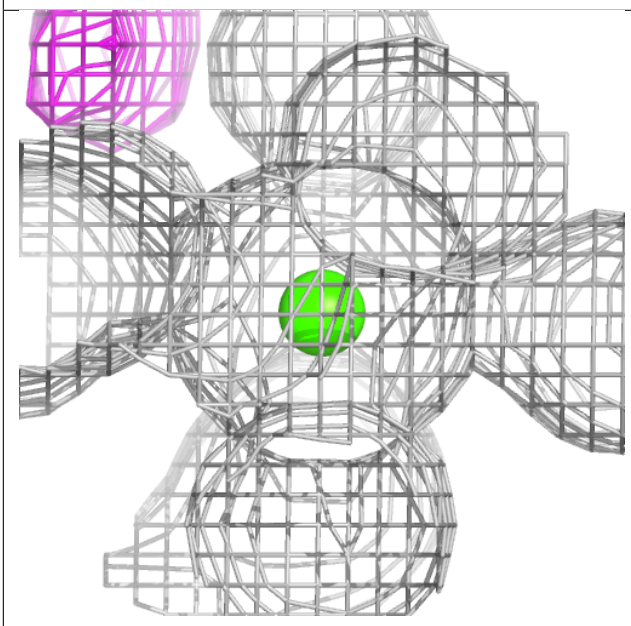
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

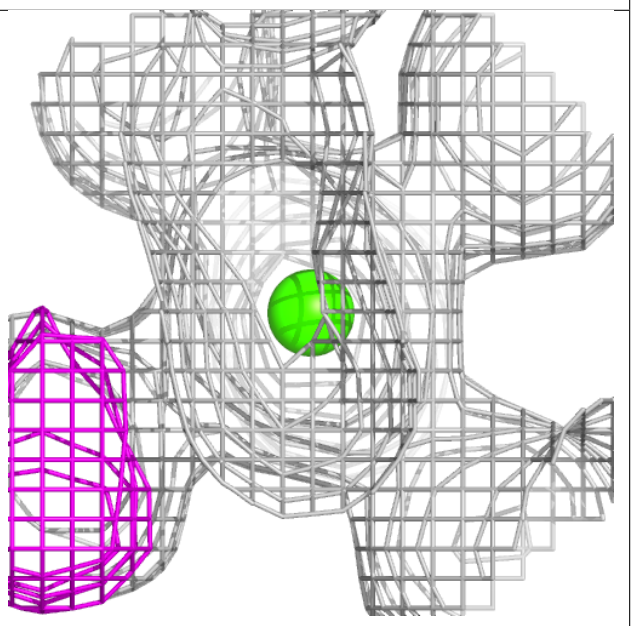


Electron density around CA B 302:

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

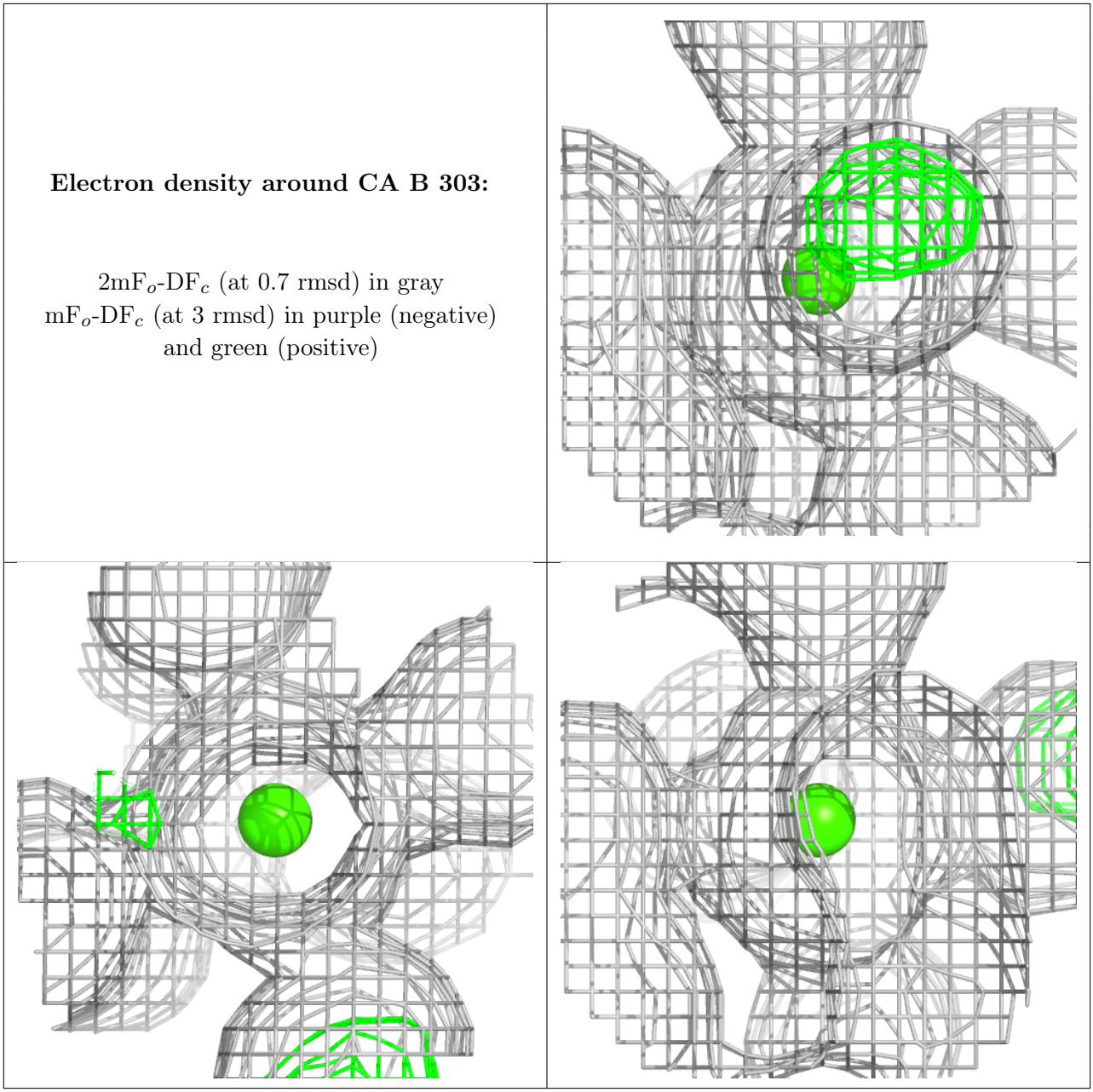






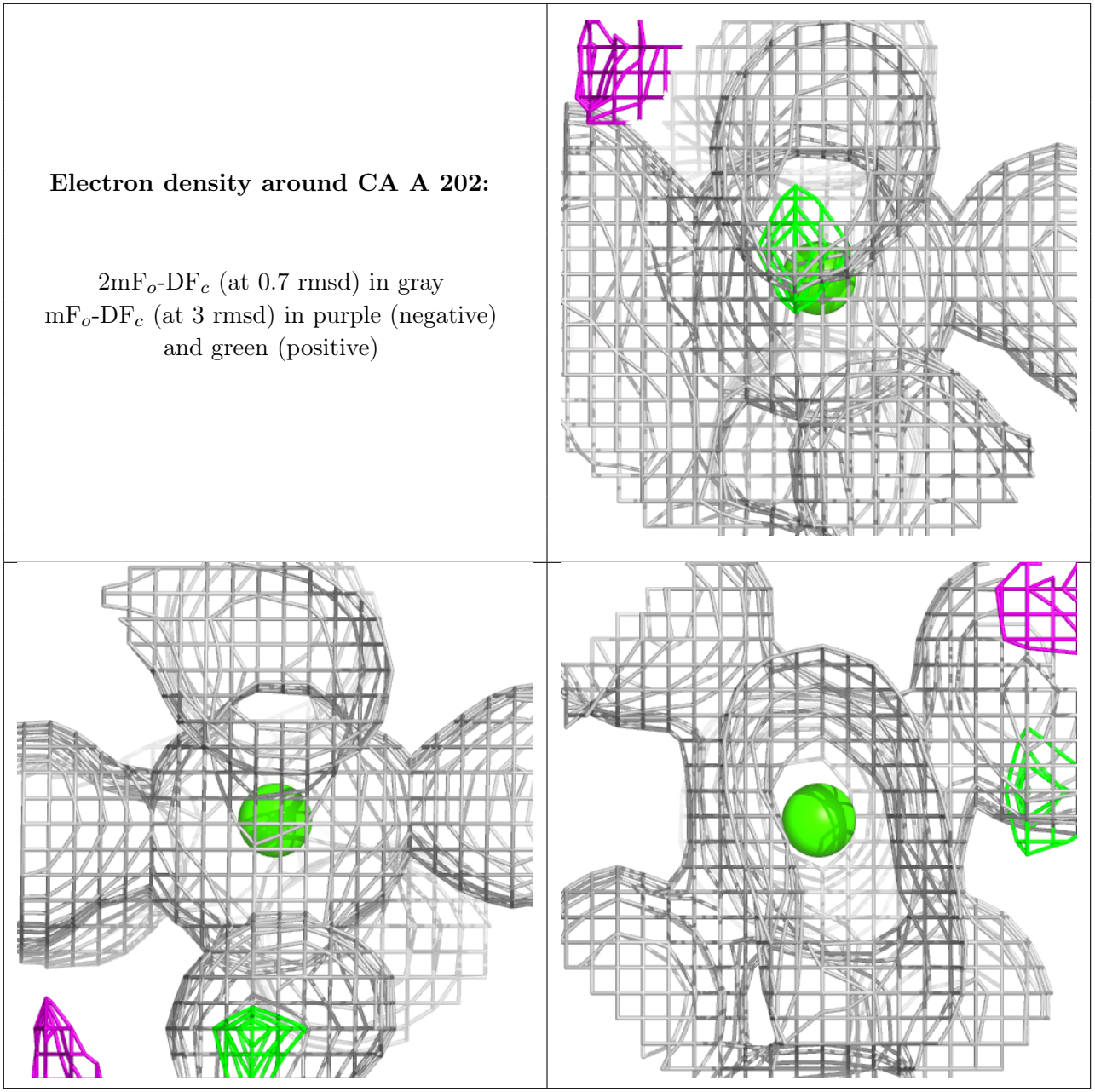
Electron density around CA B 303:

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



Electron density around CA A 202:

$2mF_o$-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.