Full wwPDB X-ray Structure Validation Report (i)

May 13, 2021 – 10:09 AM JST

PDB ID	:	7C7R
Title	:	Biofilm associated protein - B domain
Authors	:	Ma, J.F.; Xu, Z.H.; Zhang, Y.K.; Cheng, X.; Fan, S.L.; Wang, J.W.; Fang, X.Y.
Deposited on	:	2020-05-26
Resolution	:	3.07 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 following versions of software and data (see references (i)) were used in the production of this report:

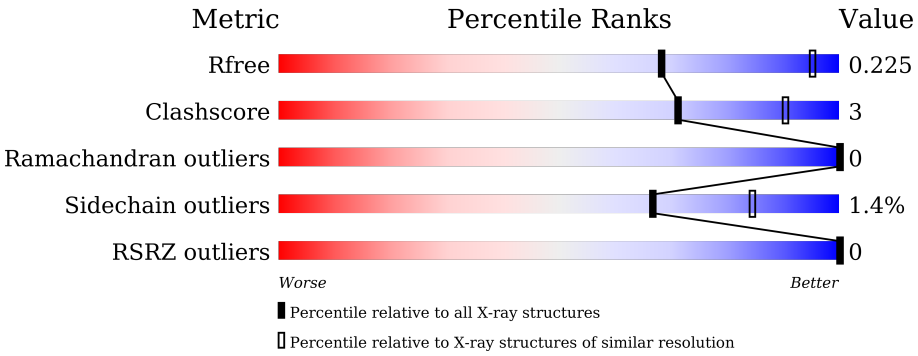
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.18
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.18

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 3.07 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	1447 (3.10-3.06)
Clashscore	141614	1546 (3.10-3.06)
Ramachandran outliers	138981	1487 (3.10-3.06)
Sidechain outliers	138945	1486 (3.10-3.06)
RSRZ outliers	127900	1416 (3.10-3.06)

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	449	84% 8% 8%
1	B	449	82% 11% 8%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6528 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 Biofilm-associated surface protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	415	Total	C	N	O	S	0	1	0
			3248	2016	549	680	3			
1	A	415	Total	C	N	O	S	0	0	0
			3240	2012	547	678	3			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	335	MET	-	initiating methionine	UNP Q79LN3
B	336	GLY	-	expression tag	UNP Q79LN3
B	337	SER	-	expression tag	UNP Q79LN3
B	338	SER	-	expression tag	UNP Q79LN3
B	339	HIS	-	expression tag	UNP Q79LN3
B	340	HIS	-	expression tag	UNP Q79LN3
B	341	HIS	-	expression tag	UNP Q79LN3
B	342	HIS	-	expression tag	UNP Q79LN3
B	343	HIS	-	expression tag	UNP Q79LN3
B	344	HIS	-	expression tag	UNP Q79LN3
B	345	SER	-	expression tag	UNP Q79LN3
B	346	SER	-	expression tag	UNP Q79LN3
B	347	GLY	-	expression tag	UNP Q79LN3
B	348	LEU	-	expression tag	UNP Q79LN3
B	349	VAL	-	expression tag	UNP Q79LN3
B	350	PRO	-	expression tag	UNP Q79LN3
B	351	ARG	-	expression tag	UNP Q79LN3
B	352	GLY	-	expression tag	UNP Q79LN3
B	353	SER	-	expression tag	UNP Q79LN3
B	354	HIS	-	expression tag	UNP Q79LN3
B	355	MET	-	expression tag	UNP Q79LN3
B	356	GLU	-	expression tag	UNP Q79LN3
B	357	ASN	-	expression tag	UNP Q79LN3
B	358	LEU	-	expression tag	UNP Q79LN3
B	359	TYR	-	expression tag	UNP Q79LN3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	360	PHE	-	expression tag	UNP Q79LN3
A	335	MET	-	initiating methionine	UNP Q79LN3
A	336	GLY	-	expression tag	UNP Q79LN3
A	337	SER	-	expression tag	UNP Q79LN3
A	338	SER	-	expression tag	UNP Q79LN3
A	339	HIS	-	expression tag	UNP Q79LN3
A	340	HIS	-	expression tag	UNP Q79LN3
A	341	HIS	-	expression tag	UNP Q79LN3
A	342	HIS	-	expression tag	UNP Q79LN3
A	343	HIS	-	expression tag	UNP Q79LN3
A	344	HIS	-	expression tag	UNP Q79LN3
A	345	SER	-	expression tag	UNP Q79LN3
A	346	SER	-	expression tag	UNP Q79LN3
A	347	GLY	-	expression tag	UNP Q79LN3
A	348	LEU	-	expression tag	UNP Q79LN3
A	349	VAL	-	expression tag	UNP Q79LN3
A	350	PRO	-	expression tag	UNP Q79LN3
A	351	ARG	-	expression tag	UNP Q79LN3
A	352	GLY	-	expression tag	UNP Q79LN3
A	353	SER	-	expression tag	UNP Q79LN3
A	354	HIS	-	expression tag	UNP Q79LN3
A	355	MET	-	expression tag	UNP Q79LN3
A	356	GLU	-	expression tag	UNP Q79LN3
A	357	ASN	-	expression tag	UNP Q79LN3
A	358	LEU	-	expression tag	UNP Q79LN3
A	359	TYR	-	expression tag	UNP Q79LN3
A	360	PHE	-	expression tag	UNP Q79LN3

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	5	Total Ca 5 5	0	0
2	A	5	Total Ca 5 5	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	9	Total O 9 9	0	0

Continued on next page...

Continued from previous page...

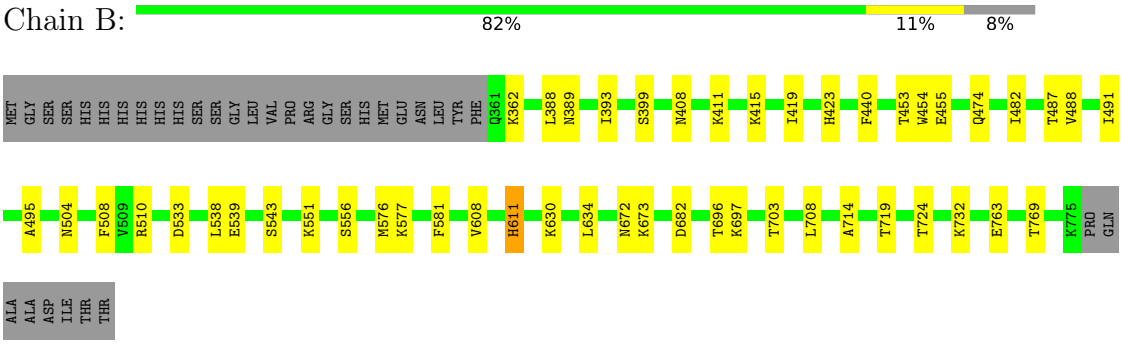
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	21	Total	O	0	0
			21	21		

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.

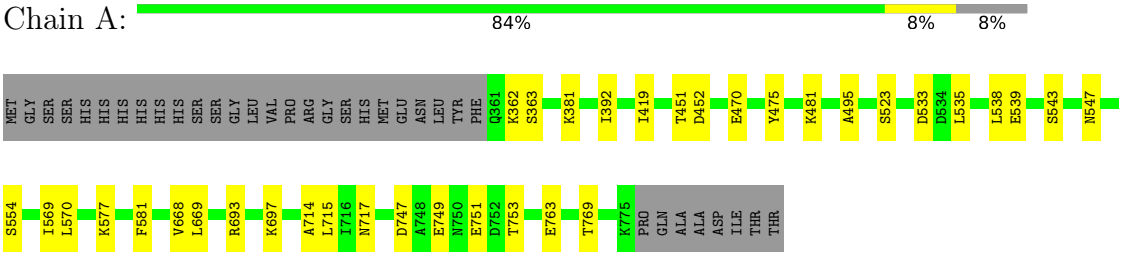
• Molecule 1: Biofilm-associated surface protein

Chain B:



• Molecule 1: Biofilm-associated surface protein

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	171.17Å 171.17Å 102.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.10 – 3.07 49.85 – 3.07	Depositor EDS
% Data completeness (in resolution range)	96.5 (49.10-3.07) 92.5 (49.85-3.07)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.74 (at 3.07Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.192 , 0.224 0.192 , 0.225	Depositor DCC
R_{free} test set	1399 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	71.2	Xtriage
Anisotropy	0.064	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 39.9	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	0.94	EDS
Total number of atoms	6528	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.67% 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: CA

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	0.25	0/3289	0.45	0/4449
1	B	0.25	0/3297	0.47	0/4460
All	All	0.25	0/6586	0.46	0/8909

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	3240	0	3117	20	0
1	B	3248	0	3122	24	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
3	A	21	0	0	0	0
3	B	9	0	0	1	0
All	All	6528	0	6239	44	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 (44) 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:B:504:ASN:ND2	3:B:902:HOH:O	2.25	0.68
1:B:577:LYS:HE2	1:B:581:PHE:HB3	1.74	0.68
1:A:577:LYS:HE2	1:A:581:PHE:HB3	1.76	0.68
1:B:474:GLN:NE2	1:B:577:LYS:O	2.27	0.68
1:A:538:LEU:HD12	1:A:715:LEU:HB2	1.76	0.66
1:B:455:GLU:OE2	1:B:510:ARG:NH2	2.30	0.65
1:A:569:ILE:HD11	1:A:669:LEU:HD12	1.78	0.64
1:A:538:LEU:H	1:A:538:LEU:HD23	1.66	0.60
1:B:419:ILE:HD13	1:B:495:ALA:HB1	1.85	0.59
1:B:708:LEU:HB2	1:B:719:THR:HG21	1.89	0.53
1:A:763:GLU:OE1	1:A:769:THR:OG1	2.25	0.53
1:B:408:ASN:O	1:B:411:LYS:NZ	2.33	0.53
1:B:608:VAL:HG13	1:B:634:LEU:HB2	1.91	0.53
1:A:543:SER:HB2	1:A:714:ALA:HB3	1.90	0.52
1:A:419:ILE:HD13	1:A:495:ALA:HB1	1.91	0.52
1:A:392:ILE:HD13	1:A:481:LYS:HD2	1.93	0.49
1:B:672:ASN:OD1	1:B:673:LYS:HG3	2.14	0.47
1:A:535:LEU:HA	1:A:538:LEU:HD21	1.96	0.47
1:B:703:THR:HG22	1:B:724:THR:HG22	1.96	0.47
1:B:543:SER:HB2	1:B:714:ALA:HB3	1.96	0.46
1:A:470:GLU:HB3	1:A:475:TYR:CZ	2.51	0.46
1:A:693:ARG:CZ	1:A:751:GLU:HG2	2.45	0.46
1:B:440:PHE:HB3	1:B:454:TRP:HB3	1.98	0.46
1:A:747:ASP:H	1:A:753:THR:HG21	1.81	0.45
1:B:533:ASP:OD2	1:B:533:ASP:N	2.50	0.45
1:A:533:ASP:N	1:A:533:ASP:OD1	2.50	0.44
1:B:508:PHE:HE1	1:B:510:ARG:HG3	1.81	0.44
1:B:415:LYS:HG3	1:B:453:THR:HG22	2.00	0.44
1:B:697:LYS:HE2	1:B:697:LYS:HB3	1.87	0.43
1:A:381:LYS:HD2	1:A:539:GLU:HG3	2.00	0.43
1:B:551:LYS:HD2	1:B:576:MET:HE1	2.00	0.43
1:A:362:LYS:HD3	1:A:749:GLU:HG2	2.01	0.43
1:A:451:THR:HG22	1:A:452:ASP:H	1.84	0.43
1:A:697:LYS:HE2	1:A:697:LYS:HB2	1.74	0.43
1:B:611:HIS:CE1	1:B:630:LYS:HB2	2.55	0.42
1:A:547:ASN:HA	1:A:717:ASN:HB2	2.02	0.42
1:B:388:LEU:HD12	1:B:488:VAL:HG23	2.02	0.41
1:B:487:THR:O	1:B:491:ILE:HG13	2.19	0.41
1:B:362:LYS:HE3	1:B:362:LYS:HB2	1.83	0.41
1:A:570:LEU:HD22	1:A:668:VAL:HG22	2.02	0.41
1:B:696:THR:O	1:B:732:LYS:NZ	2.53	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:362:LYS:NZ	1:A:363:SER:OG	2.49	0.41
1:B:393:ILE:HB	1:B:482:ILE:HB	2.03	0.41
1:B:763:GLU:HG3	1:B:769:THR:OG1	2.22	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	413/449 (92%)	395 (96%)	18 (4%)	0	100	100
1	B	414/449 (92%)	395 (95%)	19 (5%)	0	100	100
All	All	827/898 (92%)	790 (96%)	37 (4%)	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	359/388 (92%)	357 (99%)	2 (1%)	86	93
1	B	360/388 (93%)	352 (98%)	8 (2%)	52	76
All	All	719/776 (93%)	709 (99%)	10 (1%)	67	84

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	389	ASN
1	B	399	SER
1	B	423	HIS
1	B	538	LEU
1	B	539	GLU
1	B	556	SER
1	B	611	HIS
1	B	682	ASP
1	A	523	SER
1	A	554	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 10 ligands modelled in this entry, 10 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 (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	415/449 (92%)	-0.56	0	100	100	44, 66, 97, 120	0
1	B	415/449 (92%)	-0.43	0	100	100	45, 71, 109, 148	0
All	All	830/898 (92%)	-0.49	0	100	100	44, 68, 106, 148	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
2	CA	A	801	1/1	0.91	0.27	56,56,56,56	0
2	CA	B	801	1/1	0.94	0.26	65,65,65,65	0
2	CA	B	804	1/1	0.95	0.11	89,89,89,89	0
2	CA	B	805	1/1	0.95	0.14	88,88,88,88	0
2	CA	B	803	1/1	0.95	0.08	76,76,76,76	0
2	CA	B	802	1/1	0.98	0.07	76,76,76,76	0
2	CA	A	802	1/1	0.99	0.11	50,50,50,50	0

Continued on next page...

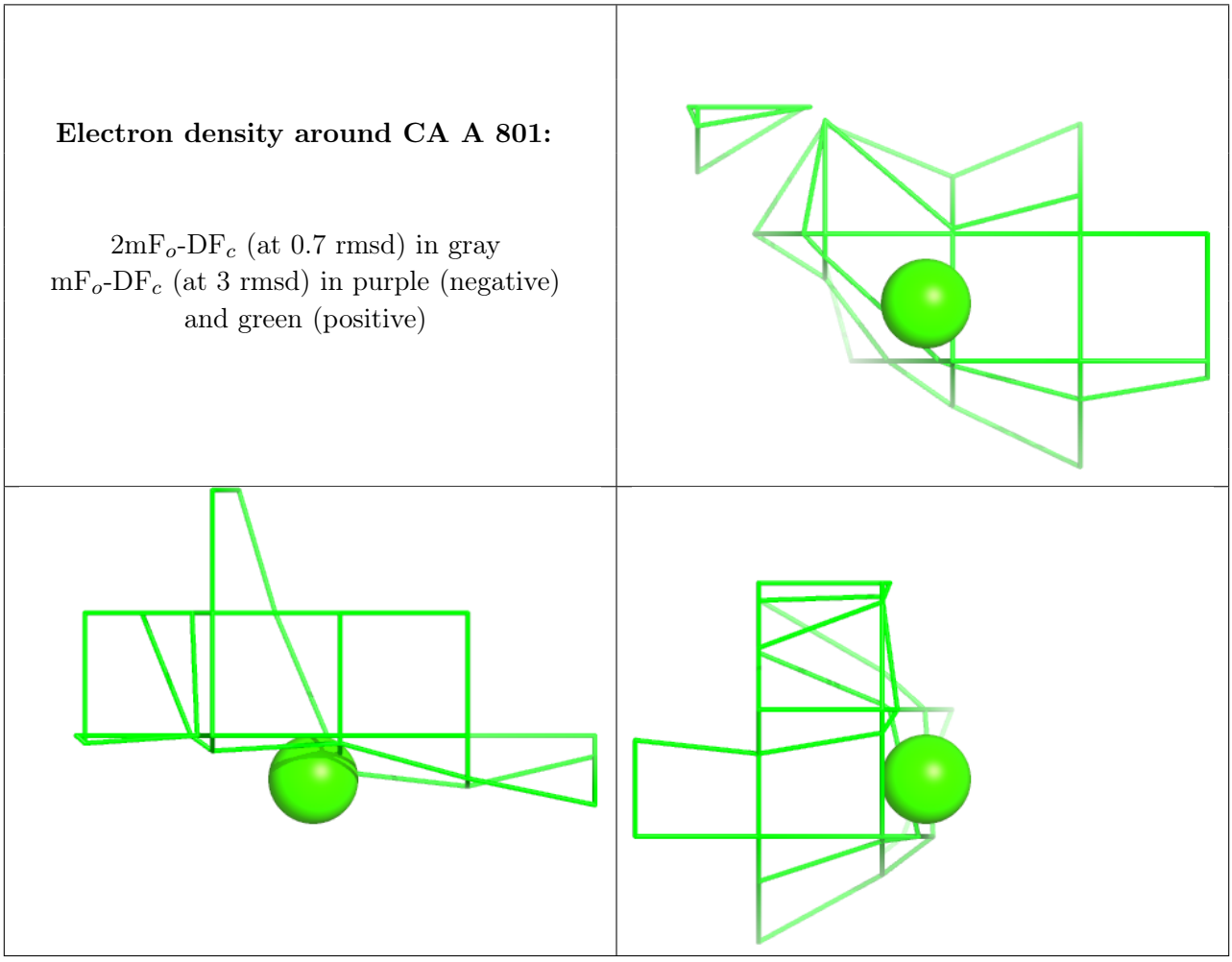
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	CA	A	803	1/1	0.99	0.14	62,62,62,62	0
2	CA	A	804	1/1	0.99	0.11	63,63,63,63	0
2	CA	A	805	1/1	0.99	0.15	52,52,52,52	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.

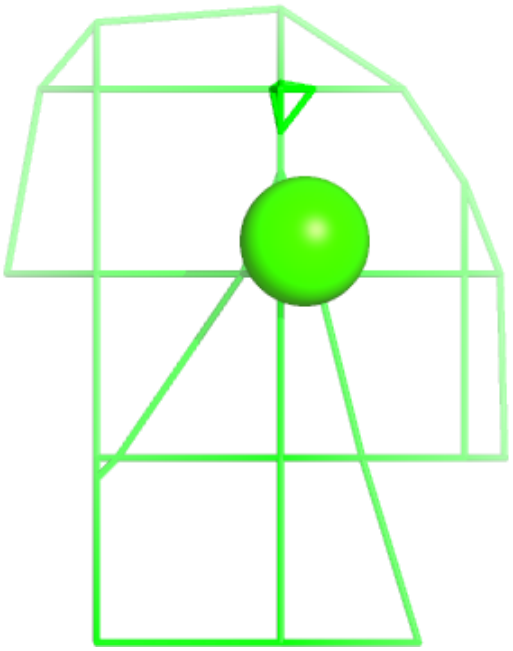
Electron density around CA A 801:

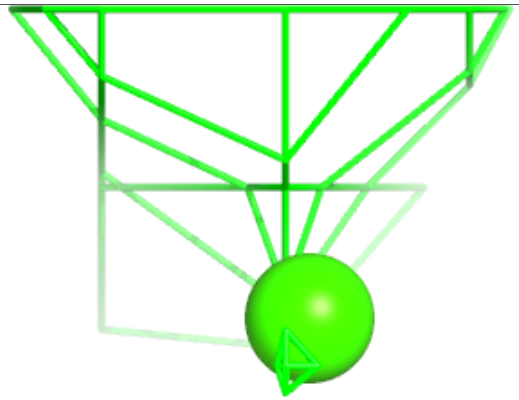
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

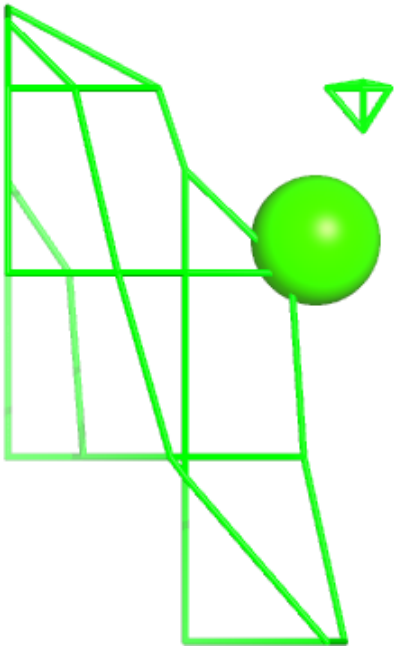


Electron density around CA B 801:

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

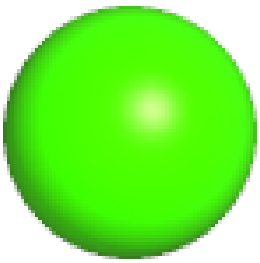


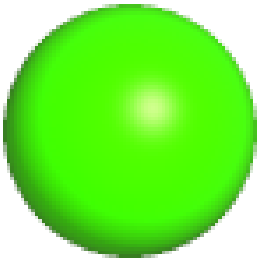


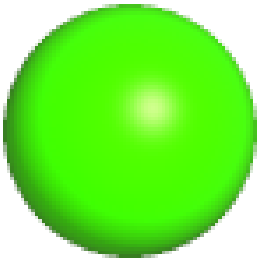


Electron density around CA B 804:

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

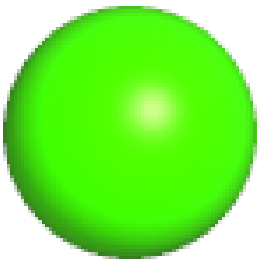


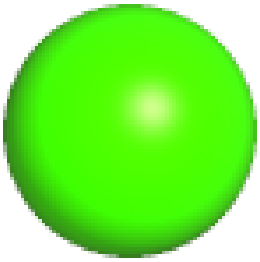


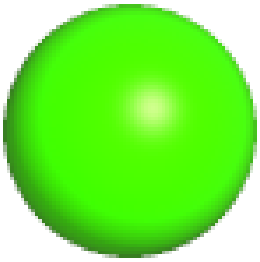


Electron density around CA B 805:

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

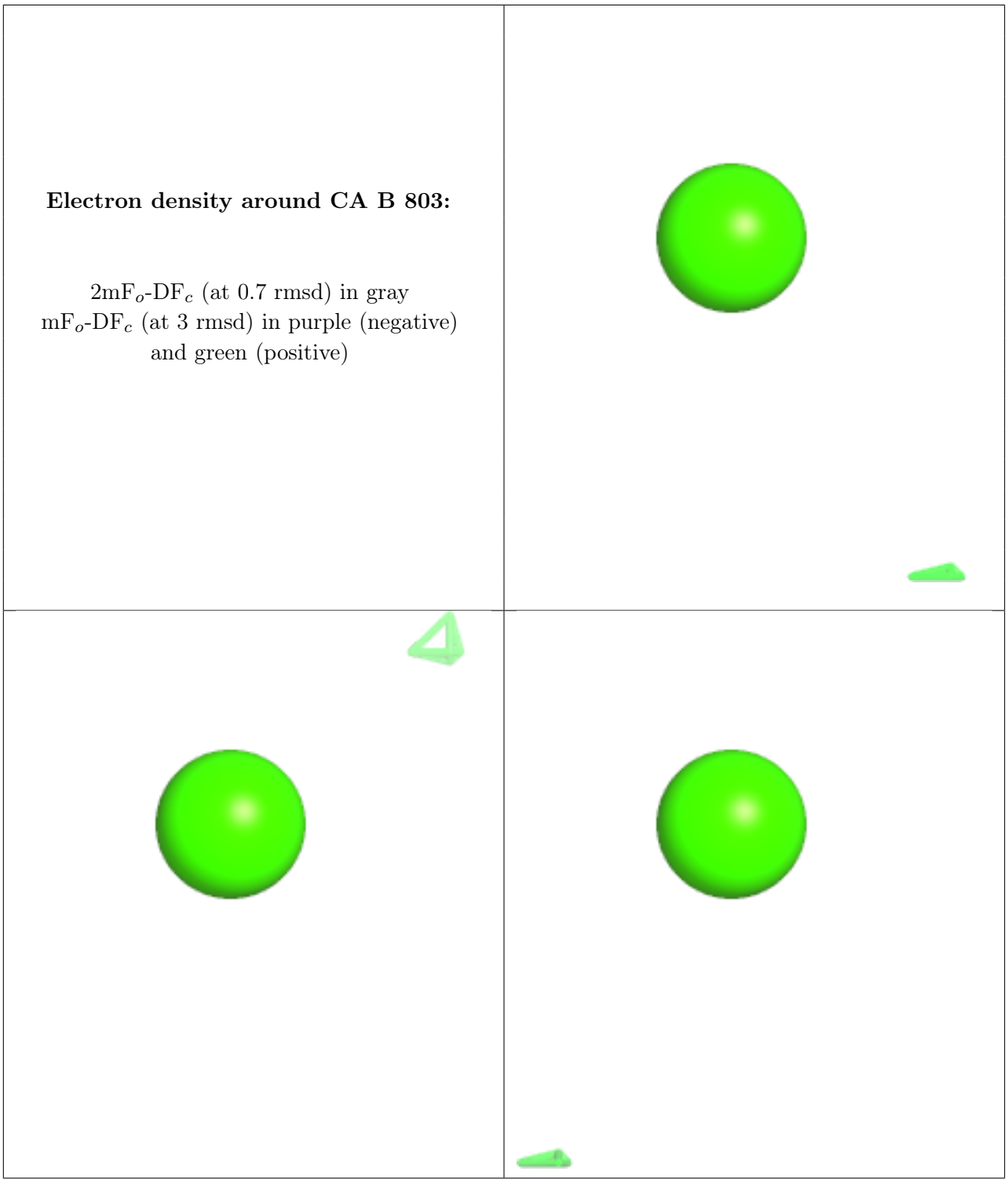






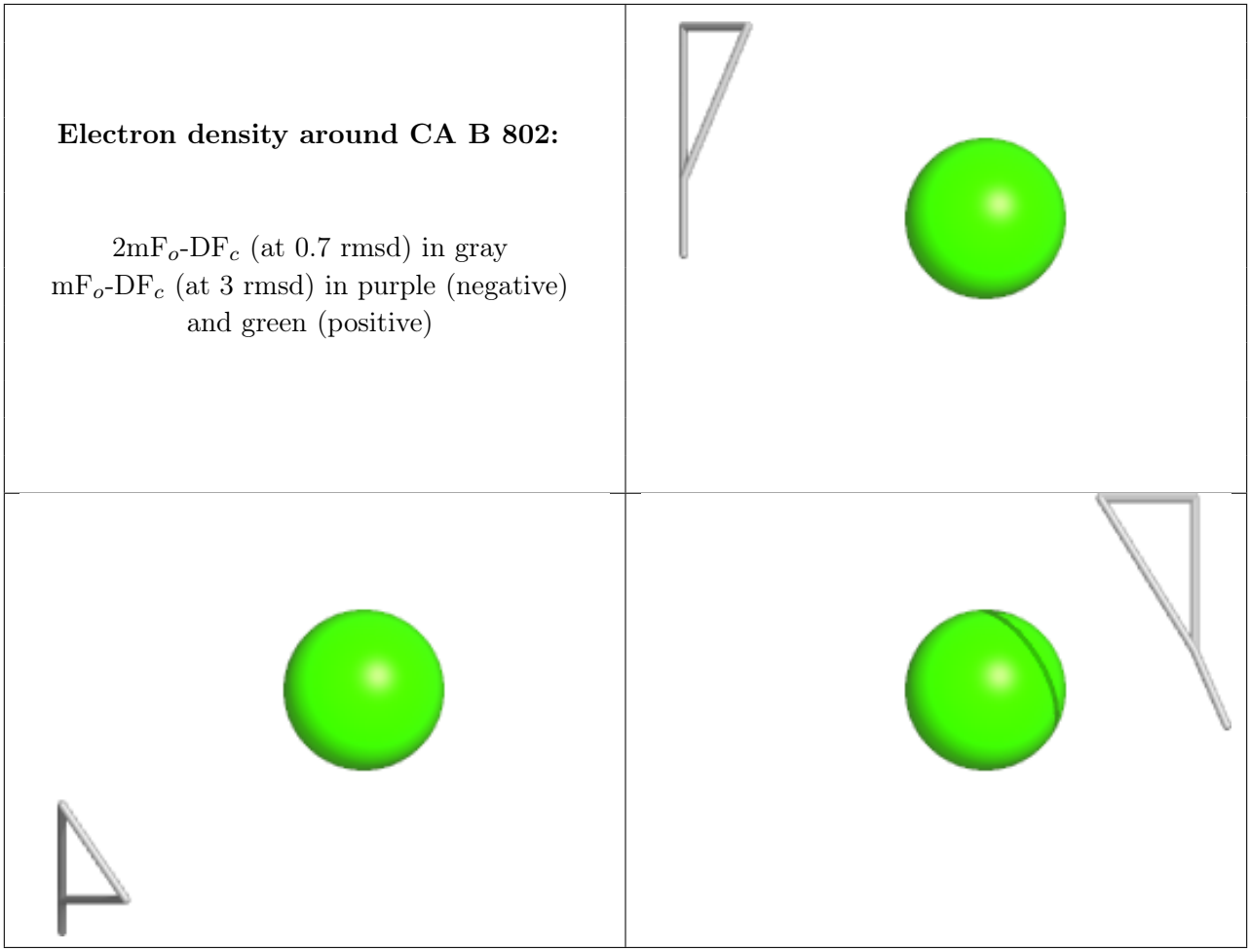
Electron density around CA B 803:

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



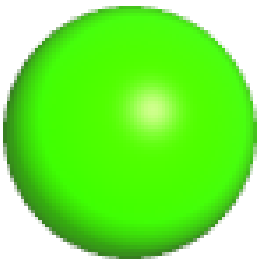
Electron density around CA B 802:

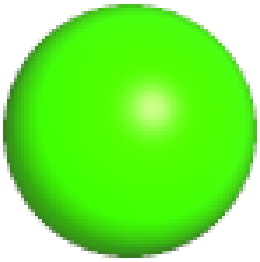
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

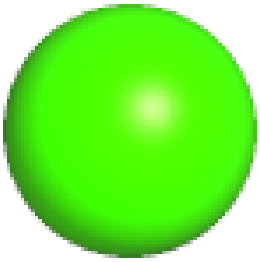


Electron density around CA A 802:

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

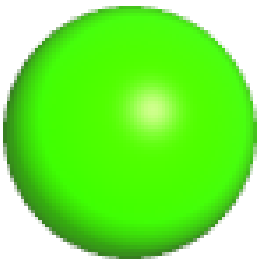


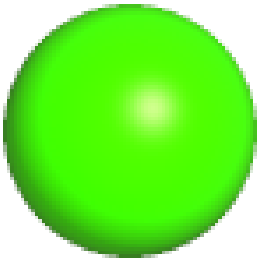


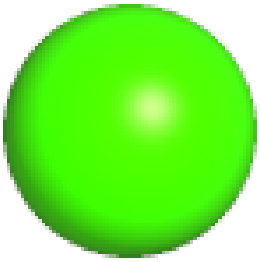


Electron density around CA A 803:

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

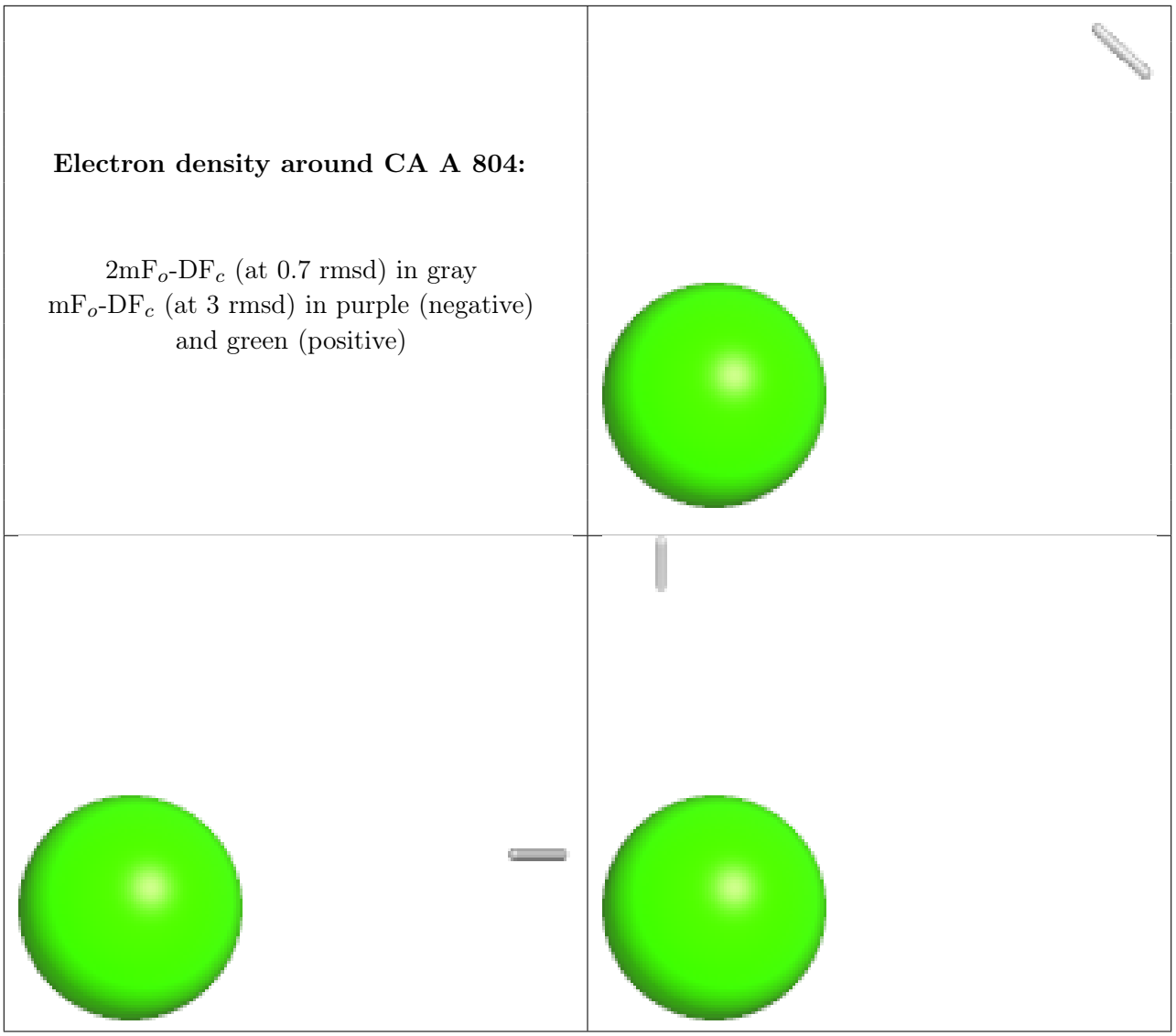






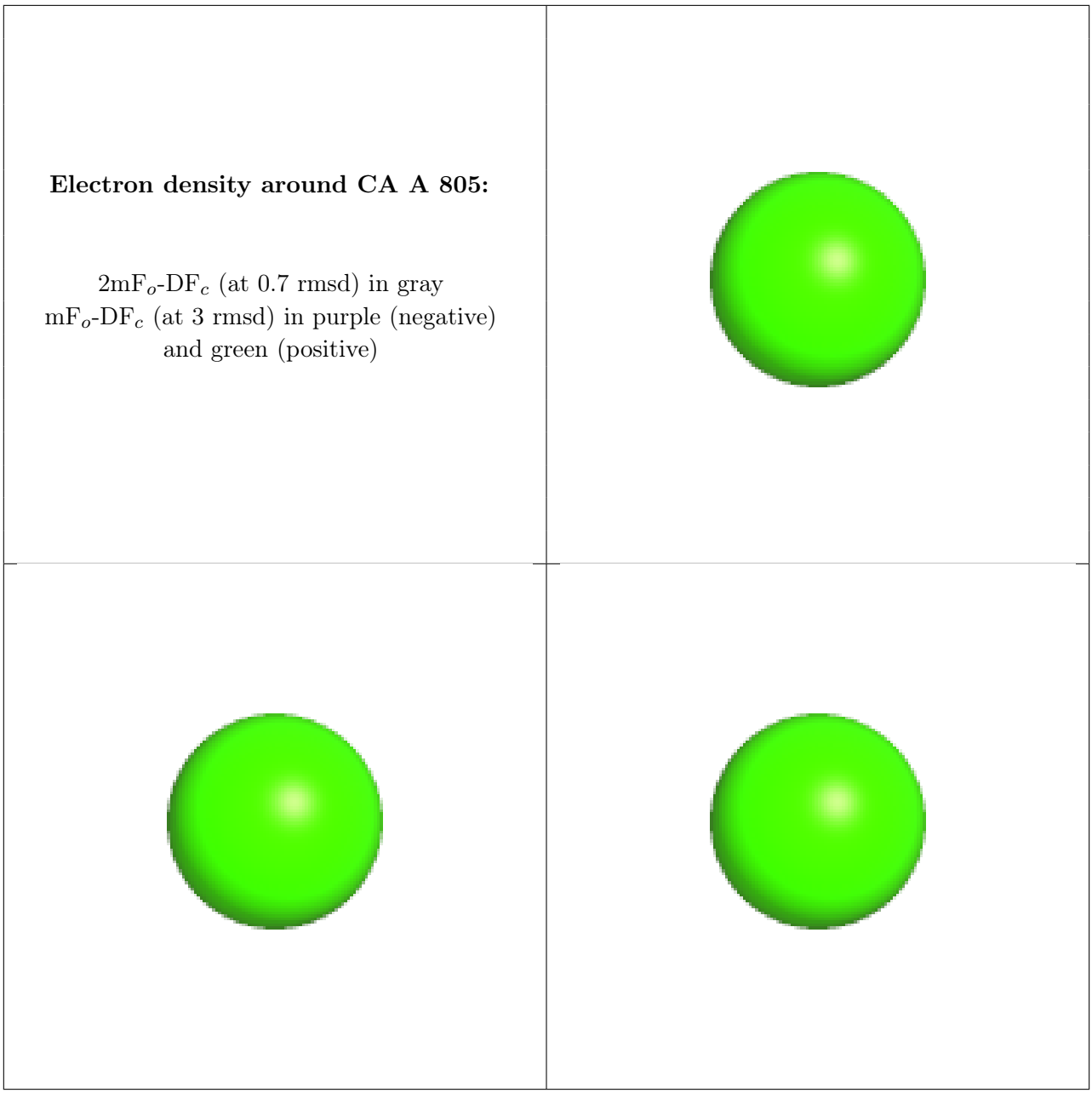
Electron density around CA A 804:

$2mF_o$-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 805:

$2mF_o$-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.