wwPDB X-ray Structure Validation Summary Report (i)

Mar 6, 2026 – 09:14 PM UTC

PDB ID	:	6KGF / pdb_00006kgf
Title	:	Crystal structure of CaDoc0917(R16D)-CaCohA2 complex at pH 8.2
Authors	:	Feng, Y.; Yao, X.
Deposited on	:	2019-07-11
Resolution	:	2.30 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

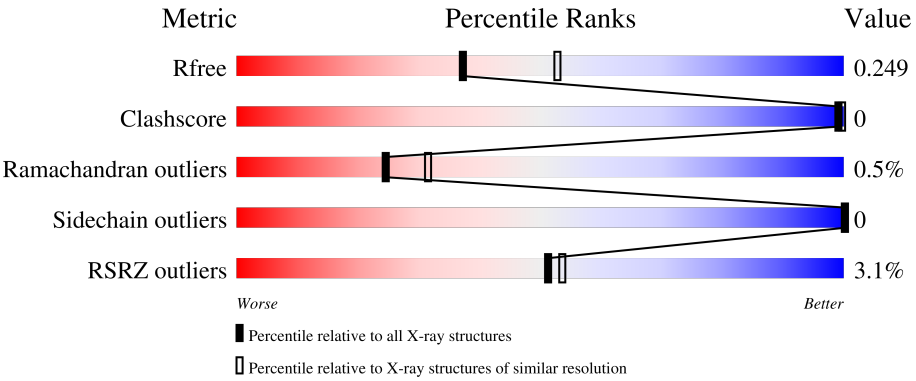
MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

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 2.30 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}	180053	6319 (2.30-2.30)
Clashscore	190562	6919 (2.30-2.30)
Ramachandran outliers	187476	6854 (2.30-2.30)
Sidechain outliers	187428	6854 (2.30-2.30)
RSRZ outliers	180081	6325 (2.30-2.30)

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	150	3% 93% • 5%
1	C	150	3% 97% ••
2	B	62	5% 98% •
2	D	62	2% 98% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6272 atoms, of which 3039 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 Probably cellulosomal scaffolding protein, secreted cellulose-binding and cohesin domain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	C	147	Total	C	H	N	O	S	0	0	0
			2164	678	1086	168	230	2			
1	A	143	Total	C	H	N	O	S	0	0	0
			2090	658	1046	162	222	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	0	SER	-	expression tag	UNP Q977Y4
A	0	SER	-	expression tag	UNP Q977Y4

- Molecule 2 is a protein called And cellulose-binding endoglucanase family 9 CelL ortholog dockerin domain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	D	62	Total	C	H	N	O	S	0	0	0
			915	276	457	81	98	3			
2	B	61	Total	C	H	N	O	S	0	0	0
			902	273	450	80	96	3			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	SER	-	expression tag	UNP Q97KK2
D	16	ASP	ARG	engineered mutation	UNP Q97KK2
B	0	SER	-	expression tag	UNP Q97KK2
B	16	ASP	ARG	engineered mutation	UNP Q97KK2

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

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

- Molecule 4 is water.

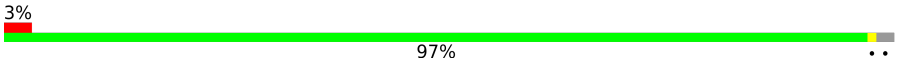
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	90	Total O 90 90	0	0
4	D	39	Total O 39 39	0	0
4	A	43	Total O 43 43	0	0
4	B	25	Total O 25 25	0	0

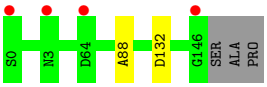
3 Residue-property plots (i)

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: Probably cellulosomal scaffolding protein, secreted cellulose-binding and cohesin domain

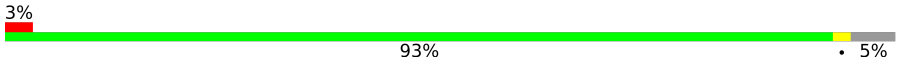
Chain C:

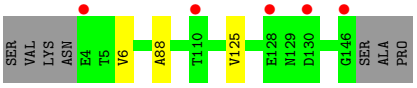




• Molecule 1: Probably cellulosomal scaffolding protein, secreted cellulose-binding and cohesin domain

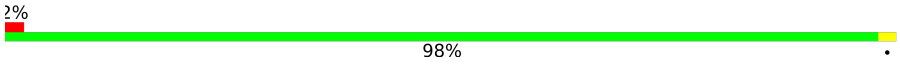
Chain A:





• Molecule 2: And cellulose-binding endoglucanase family 9 CelL ortholog dockerin domain

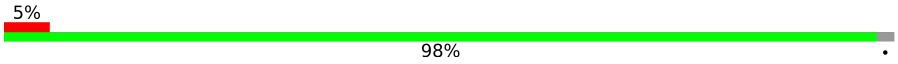
Chain D:

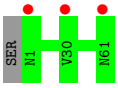




• Molecule 2: And cellulose-binding endoglucanase family 9 CelL ortholog dockerin domain

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	41.65Å 75.97Å 120.33Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.16 – 2.30 47.16 – 2.30	Depositor EDS
% Data completeness (in resolution range)	98.7 (47.16-2.30) 90.5 (47.16-2.30)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.65 (at 2.29Å)	Xtriage
Refinement program	PHENIX (1.13_2998: ???)	Depositor
R, R_{free}	0.201 , 0.249 0.203 , 0.249	Depositor DCC
R_{free} test set	1736 reflections (9.89%)	wwPDB-VP
Wilson B-factor (Å^2)	30.6	Xtriage
Anisotropy	0.712	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 38.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6272	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.70% 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	0.09	0/1059	0.25	0/1453
1	C	0.10	0/1093	0.28	0/1498
2	B	0.14	0/452	0.24	0/606
2	D	0.09	0/458	0.23	0/614
All	All	0.10	0/3062	0.26	0/4171

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	1044	1046	1046	1	0
1	C	1078	1086	1086	1	0
2	B	452	450	452	0	0
2	D	458	457	457	1	0
3	B	2	0	0	0	0
3	D	2	0	0	0	0
4	A	43	0	0	0	0
4	B	25	0	0	0	0
4	C	90	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	39	0	0	0	0
All	All	3233	3039	3041	2	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 0.

All (2) 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:132:ASP:OD1	2:D:49:ARG:NH2	2.28	0.66
1:A:6:VAL:HG21	1:A:125:VAL:HG11	2.00	0.43

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	141/150 (94%)	136 (96%)	4 (3%)	1 (1%)	18 23
1	C	145/150 (97%)	140 (97%)	4 (3%)	1 (1%)	18 23
2	B	59/62 (95%)	57 (97%)	2 (3%)	0	100 100
2	D	60/62 (97%)	59 (98%)	1 (2%)	0	100 100
All	All	405/424 (96%)	392 (97%)	11 (3%)	2 (0%)	24 31

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	88	ALA
1	C	88	ALA

5.3.2 Protein sidechains

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	123/130 (95%)	123 (100%)	0	100	100
1	C	128/130 (98%)	128 (100%)	0	100	100
2	B	50/51 (98%)	50 (100%)	0	100	100
2	D	51/51 (100%)	51 (100%)	0	100	100
All	All	352/362 (97%)	352 (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. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	108	ASN

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 4 ligands modelled in this entry, 4 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, 95th 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	143/150 (95%)	0.66	5 (3%)	47	49	32, 52, 85, 122	0
1	C	147/150 (98%)	0.35	4 (2%)	56	58	29, 43, 71, 85	0
2	B	61/62 (98%)	0.51	3 (4%)	35	36	36, 48, 70, 99	0
2	D	62/62 (100%)	0.35	1 (1%)	70	72	30, 44, 60, 80	0
All	All	413/424 (97%)	0.48	13 (3%)	51	53	29, 46, 79, 122	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	146	GLY	5.5
1	A	146	GLY	5.1
1	C	0	SER	4.1
1	A	4	GLU	3.3
1	A	128	GLU	3.3

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

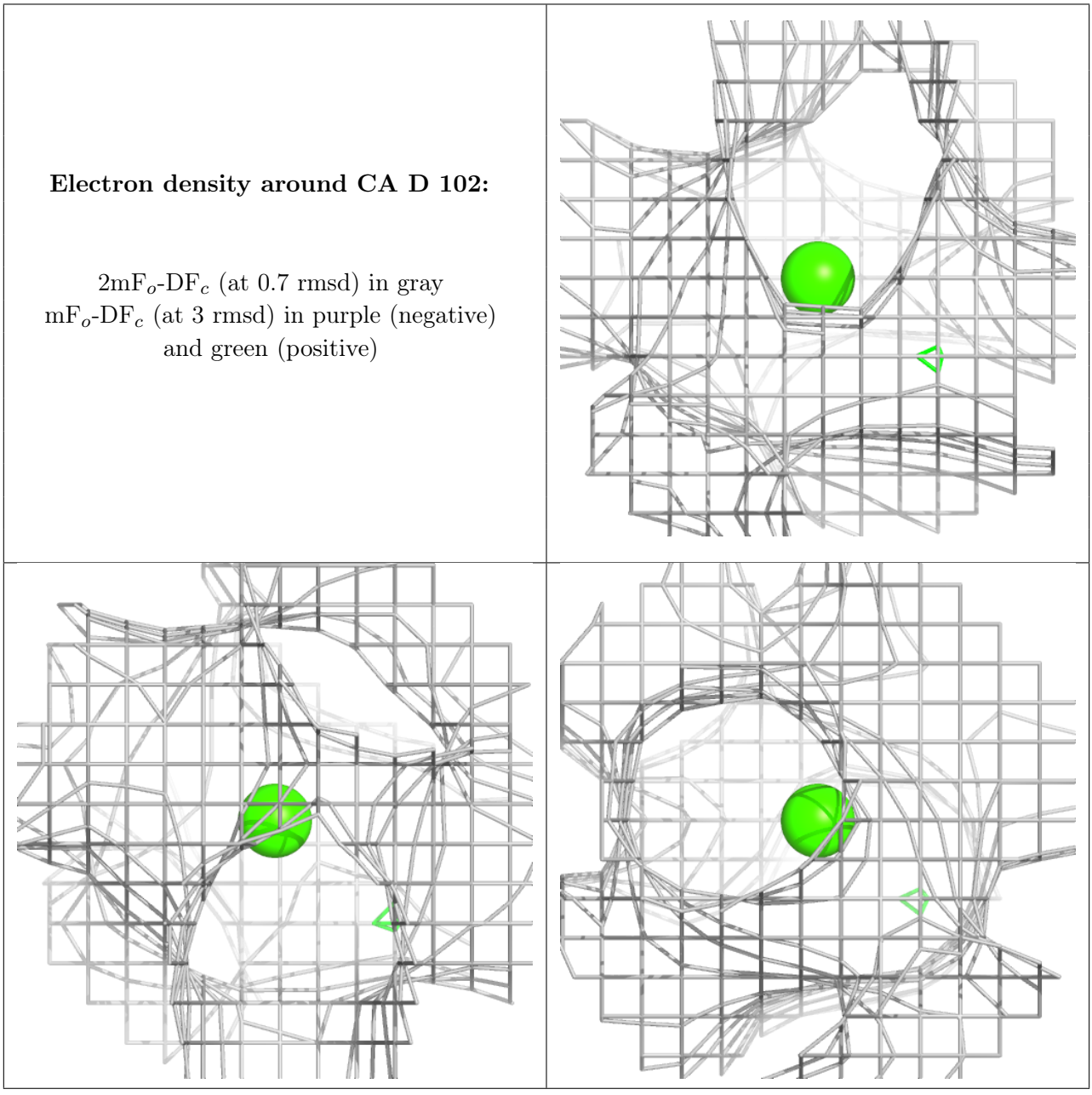
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	CA	D	102	1/1	0.95	0.04	28,28,28,28	0
3	CA	D	101	1/1	0.97	0.04	33,33,33,33	0
3	CA	B	101	1/1	0.97	0.04	36,36,36,36	0
3	CA	B	102	1/1	0.97	0.03	32,32,32,32	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.

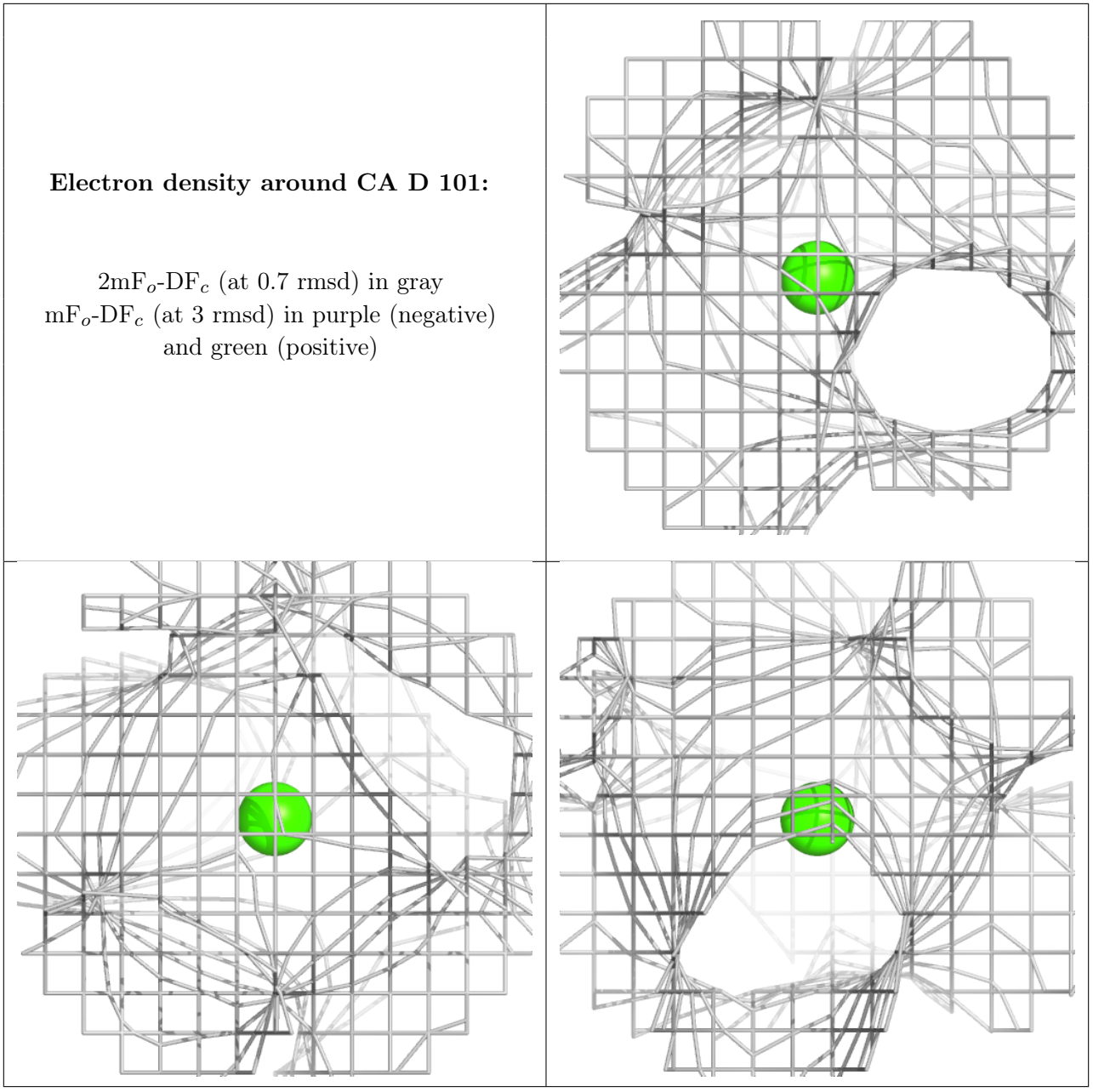
Electron density around CA D 102:

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



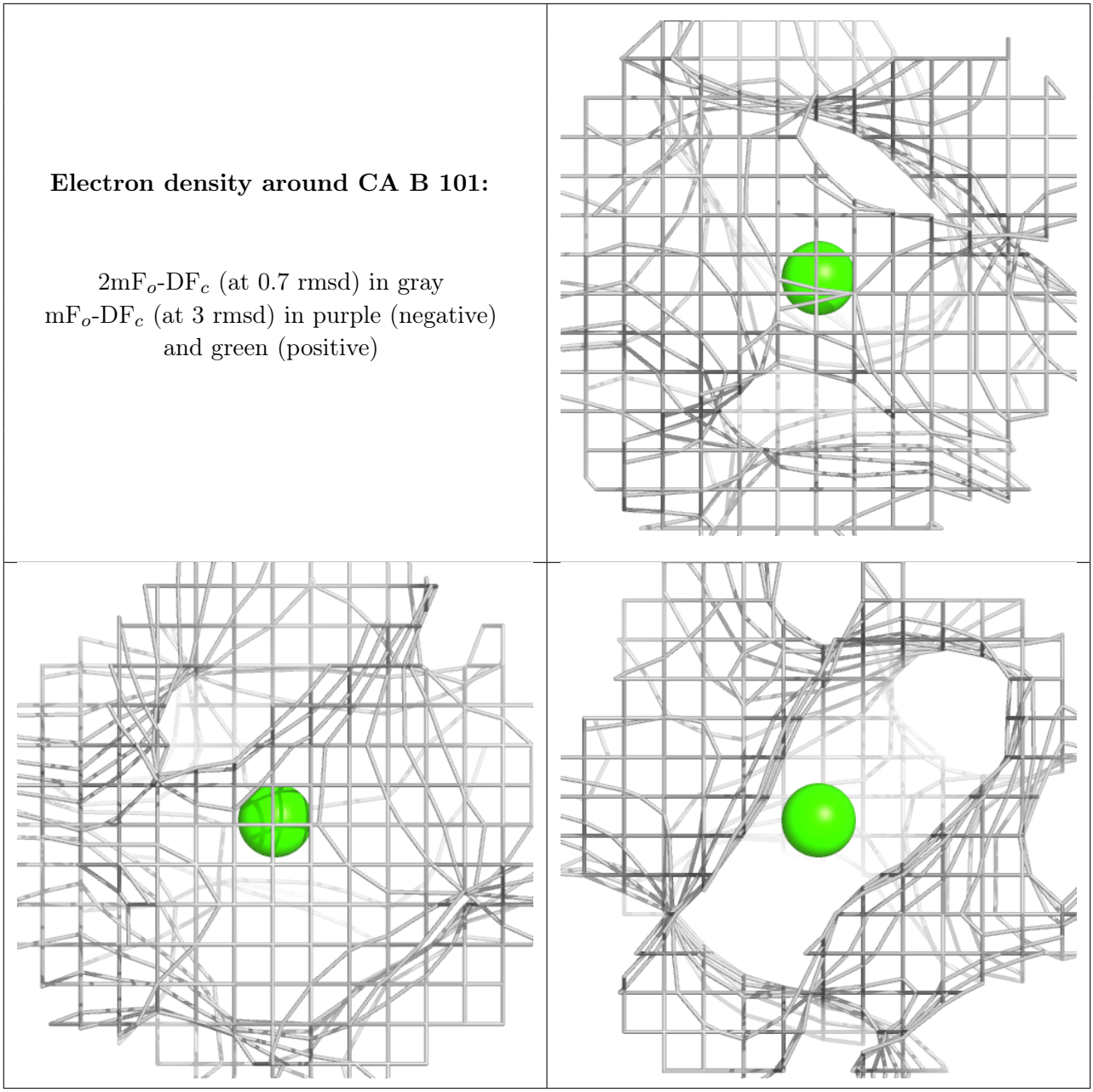
Electron density around CA D 101:

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



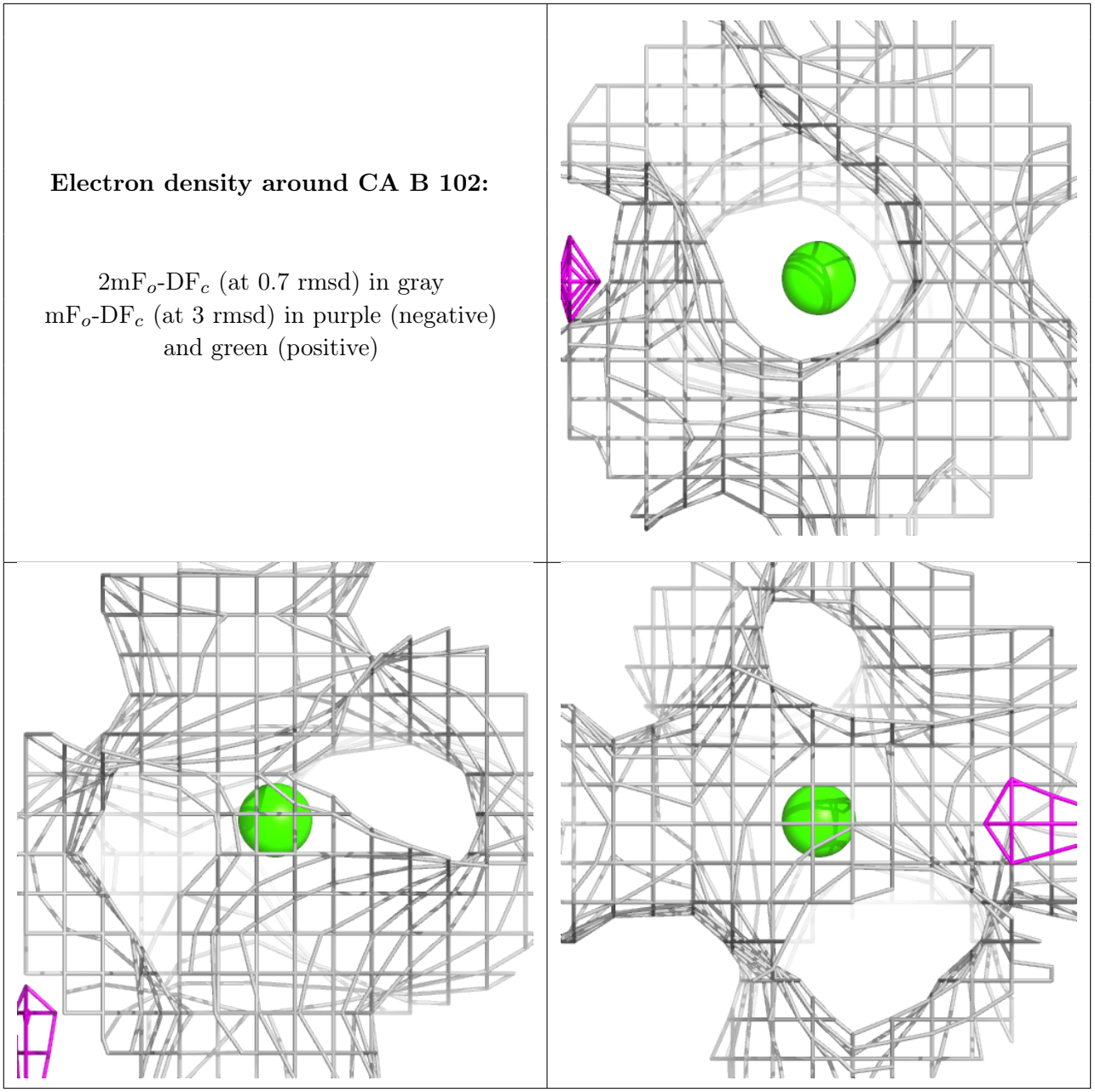
Electron density around CA B 101:

$2mF_o$-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 102:

$2mF_o$-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.