Full wwPDB X-ray Structure Validation Report (i)

Nov 13, 2024 – 03:29 pm GMT

PDB ID	:	9F5J
Title	:	SARS-CoV-2 Nucleocapsid N-terminal domain (NTD) mutant Q58I
Authors	:	Dhamotharan, K.; Schlundt, A.
Deposited on	:	2024-04-29
Resolution	:	2.20 Å(reported)

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

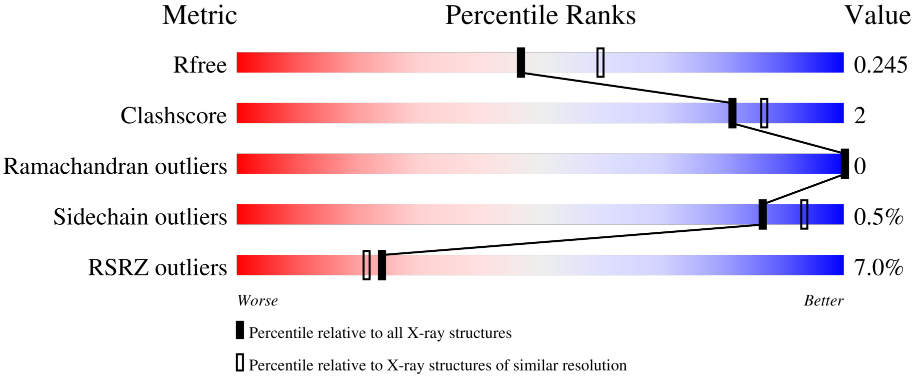
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

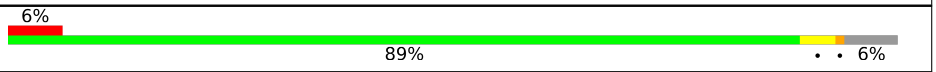
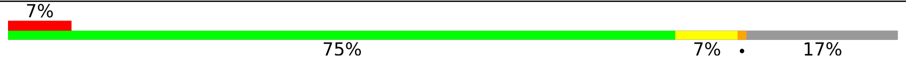
The reported resolution of this entry is 2.20 Å.

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	5791 (2.20-2.20)
Clashscore	180529	6634 (2.20-2.20)
Ramachandran outliers	177936	6560 (2.20-2.20)
Sidechain outliers	177891	6561 (2.20-2.20)
RSRZ outliers	164620	5791 (2.20-2.20)

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	138	
1	B	138	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3821 atoms, of which 1785 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 Nucleoprotein.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	A	130	Total C H N O S 1968 638 959 183 187 1	0	0	0
1	B	114	Total C H N O 1711 566 826 153 166	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	43	GLY	-	expression tag	UNP P0DTC9
A	58	ILE	GLN	engineered mutation	UNP P0DTC9
B	43	GLY	-	expression tag	UNP P0DTC9
B	58	ILE	GLN	engineered mutation	UNP P0DTC9

- Molecule 2 is CADMIUM ION (three-letter code: CD) (formula: Cd) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total Cd 2 2	0	0
2	B	2	Total Cd 2 2	0	0

- Molecule 3 is water.

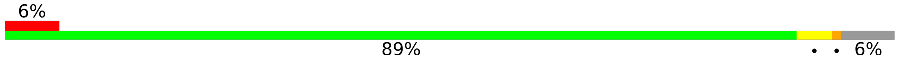
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	93	Total O 93 93	0	0
3	B	45	Total O 45 45	0	0

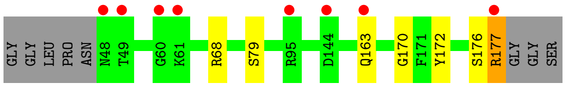
3 Residue-property plots ⓘ

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: Nucleoprotein

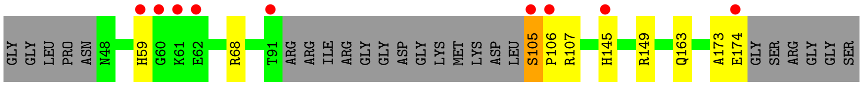
Chain A:





- Molecule 1: Nucleoprotein

Chain B:



4 Data and refinement statistics ⓘ

Property	Value
Space group	P 21 21 21
Cell constants a, b, c, α, β, γ	47.78Å 49.72Å 114.25Å 90.00° 90.00° 90.00°
Resolution (Å)	57.12 – 2.20 57.12 – 2.20
% Data completeness (in resolution range)	99.9 (57.12-2.20) 99.9 (57.12-2.20)
R_{merge}	0.10
R_{sym}	(Not available)
$< I/\sigma(I) >$ [1]	5.73 (at 2.20Å)
Refinement program	REFMAC 5.8.0430 (refmacat 0.4.88), BUSTER 2.10.4, PDB-REDO 8.1
R, R_{free}	0.202 , 0.242 0.206 , 0.245
R_{free} test set	719 reflections (4.98%)
Wilson B-factor (Å^2)	41.6
Anisotropy	0.619
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 46.3
L-test for twinning[2]	$< \lvert L \rvert > = 0.51$, $< L^2 > = 0.34$
Estimated twinning fraction	0.017 for k,h,-l
F_o,F_c correlation	0.95
Total number of atoms	3821
Average B, all atoms (Å^2)	50.0

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.93% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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: CD

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	0.68	1/1037 (0.1%)	0.93	3/1409 (0.2%)
1	B	0.62	0/912	1.02	4/1246 (0.3%)
All	All	0.65	1/1949 (0.1%)	0.97	7/2655 (0.3%)

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	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	79	SER	CB-OG	5.28	1.49	1.42

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	177	ARG	NE-CZ-NH2	6.87	123.73	120.30
1	B	68	ARG	NE-CZ-NH1	6.60	123.60	120.30
1	A	177	ARG	NE-CZ-NH1	6.44	123.52	120.30
1	A	177	ARG	NH1-CZ-NH2	-6.07	112.73	119.40
1	B	105	SER	CB-CA-C	5.37	120.30	110.10
1	B	149	ARG	NE-CZ-NH1	5.14	122.87	120.30
1	B	68	ARG	NE-CZ-NH2	-5.08	117.76	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	68	ARG	Sidechain
1	B	107	ARG	Sidechain

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	1009	959	981	5	0
1	B	885	826	848	5	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	93	0	0	0	0
3	B	45	0	0	2	0
All	All	2036	1785	1829	9	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 (9) 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:170:GLY:H	1:A:177:ARG:HH12	1.26	0.80
1:A:176:SER:OG	1:A:177:ARG:N	2.27	0.68
1:A:170:GLY:H	1:A:177:ARG:NH1	2.02	0.52
1:B:59:HIS:HB2	1:B:173:ALA:HB3	1.96	0.47
1:A:163:GLN:HA	1:B:163:GLN:HA	1.99	0.45
1:A:172:TYR:HB2	1:A:177:ARG:HD3	2.00	0.44
1:B:145:HIS:CE1	3:B:335:HOH:O	2.70	0.44
1:B:173:ALA:O	1:B:174:GLU:HB2	2.18	0.43
1:B:106:PRO:HA	3:B:331:HOH:O	2.19	0.42

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	128/138 (93%)	124 (97%)	4 (3%)	0	100	100
1	B	110/138 (80%)	108 (98%)	2 (2%)	0	100	100
All	All	238/276 (86%)	232 (98%)	6 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	102/106 (96%)	102 (100%)	0	100	100
1	B	90/106 (85%)	89 (99%)	1 (1%)	70	82
All	All	192/212 (91%)	191 (100%)	1 (0%)	86	93

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	105	SER

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 (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 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, 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	130/138 (94%)	0.18	8 (6%)	28	25	31, 43, 74, 88	0
1	B	114/138 (82%)	0.32	9 (7%)	20	18	32, 47, 84, 111	0
All	All	244/276 (88%)	0.25	17 (6%)	24	21	31, 45, 78, 111	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	105	SER	6.0
1	B	106	PRO	4.3
1	A	49	THR	3.5
1	B	91	THR	3.2
1	A	48	ASN	3.2
1	B	145	HIS	2.9
1	A	177	ARG	2.8
1	A	61	LYS	2.5
1	A	60	GLY	2.5
1	B	59	HIS	2.4
1	B	62	GLU	2.3
1	A	144	ASP	2.2
1	A	163	GLN	2.2
1	B	60	GLY	2.2
1	B	61	LYS	2.2
1	A	95	ARG	2.1
1	B	174	GLU	2.0

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.

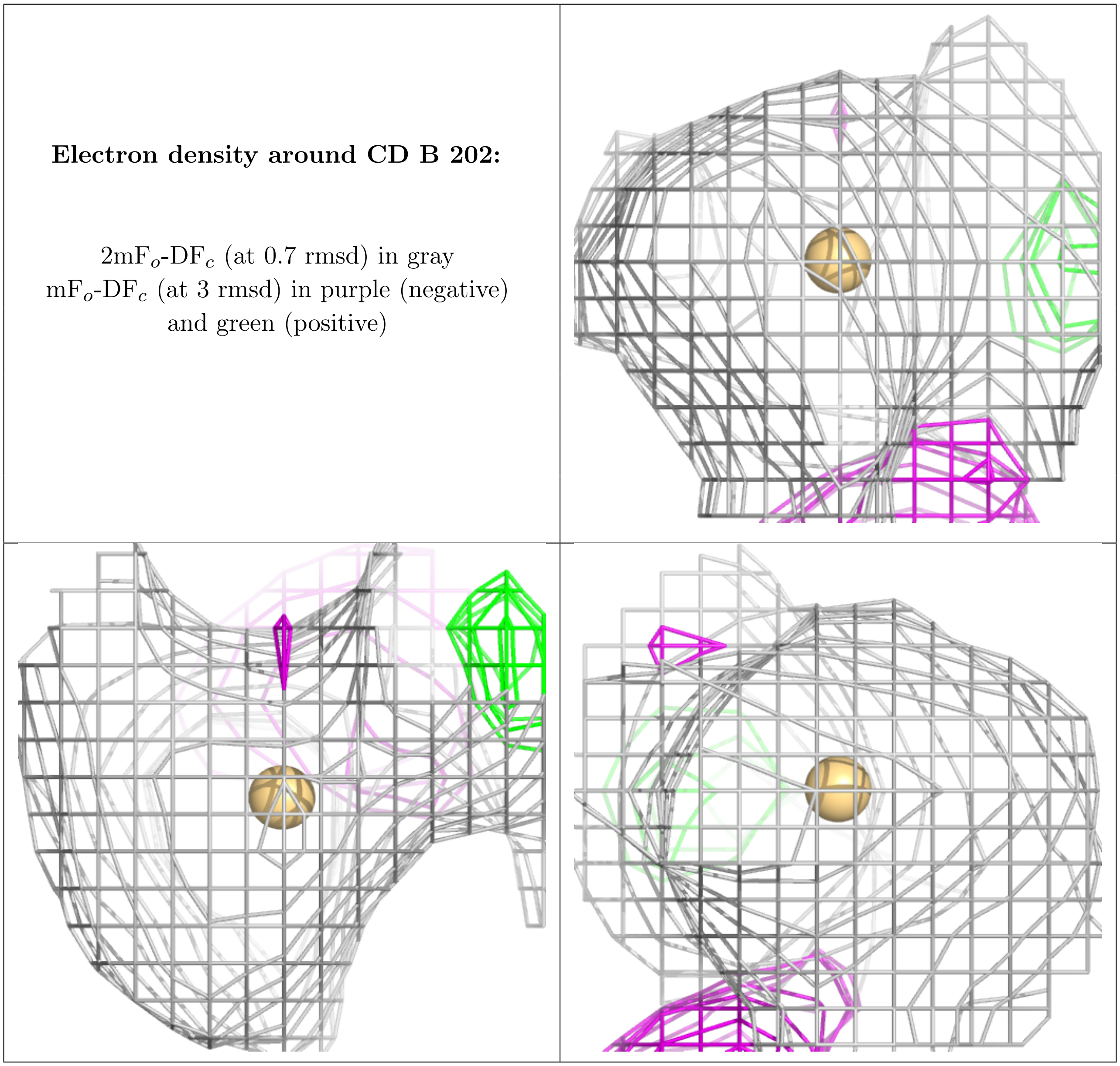
6.4 Ligands ⓘ

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	CD	B	202	1/1	0.82	0.12	146,146,146,146	0
2	CD	B	201	1/1	0.86	0.09	159,159,159,159	0
2	CD	A	202	1/1	0.98	0.04	51,51,51,51	0
2	CD	A	201	1/1	1.00	0.05	36,36,36,36	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.

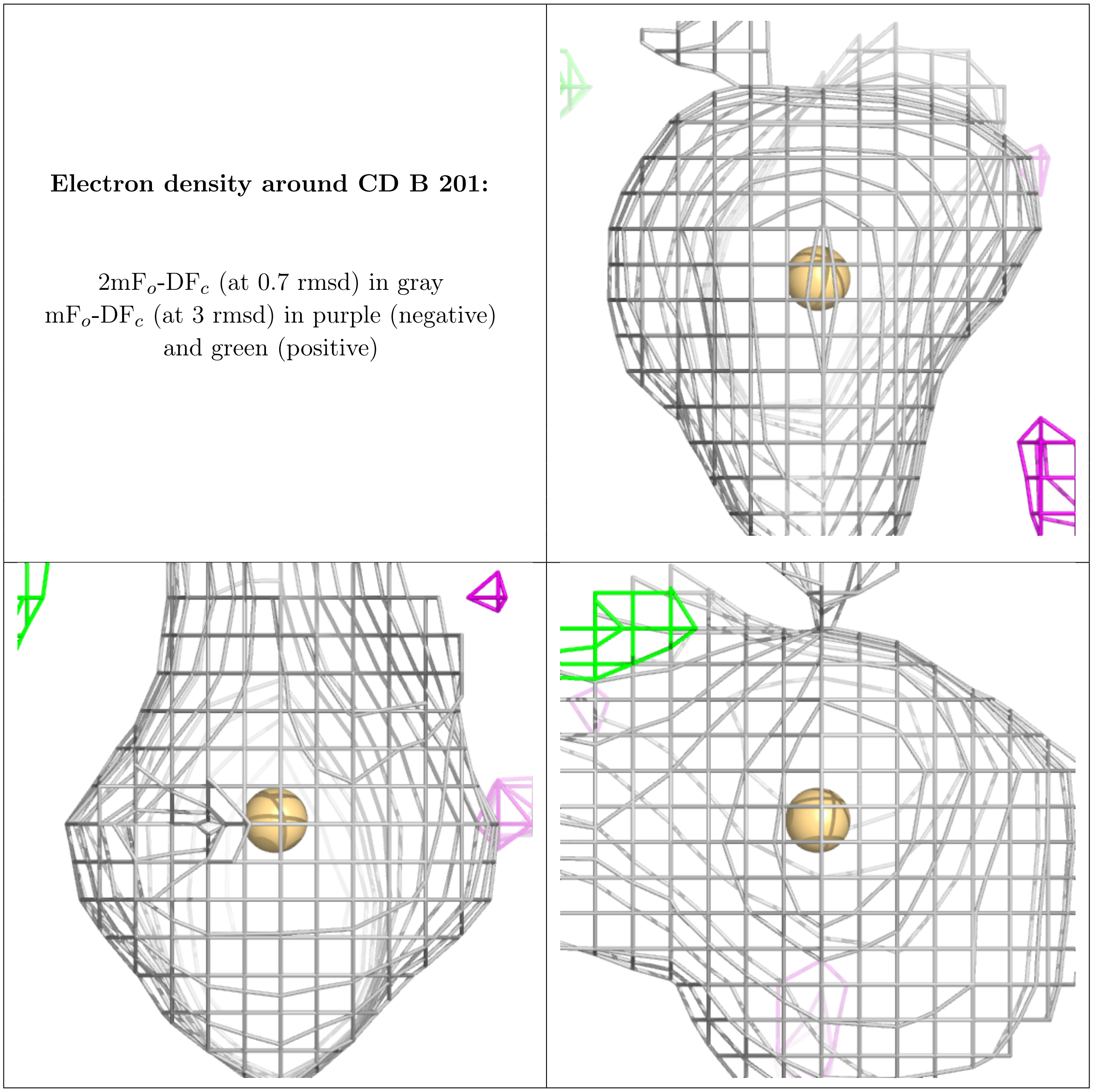
Electron density around CD B 202:



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

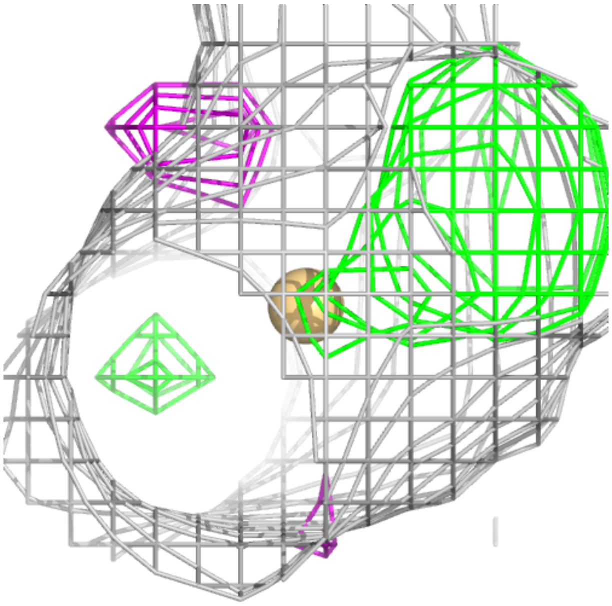
Electron density around CD B 201:

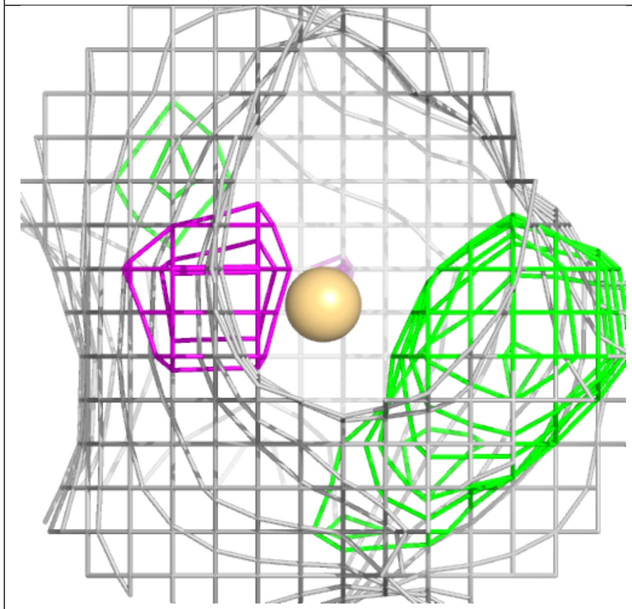
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

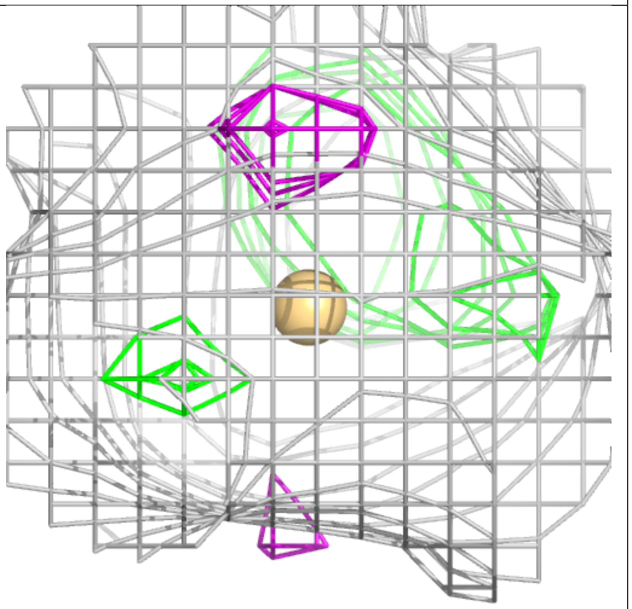


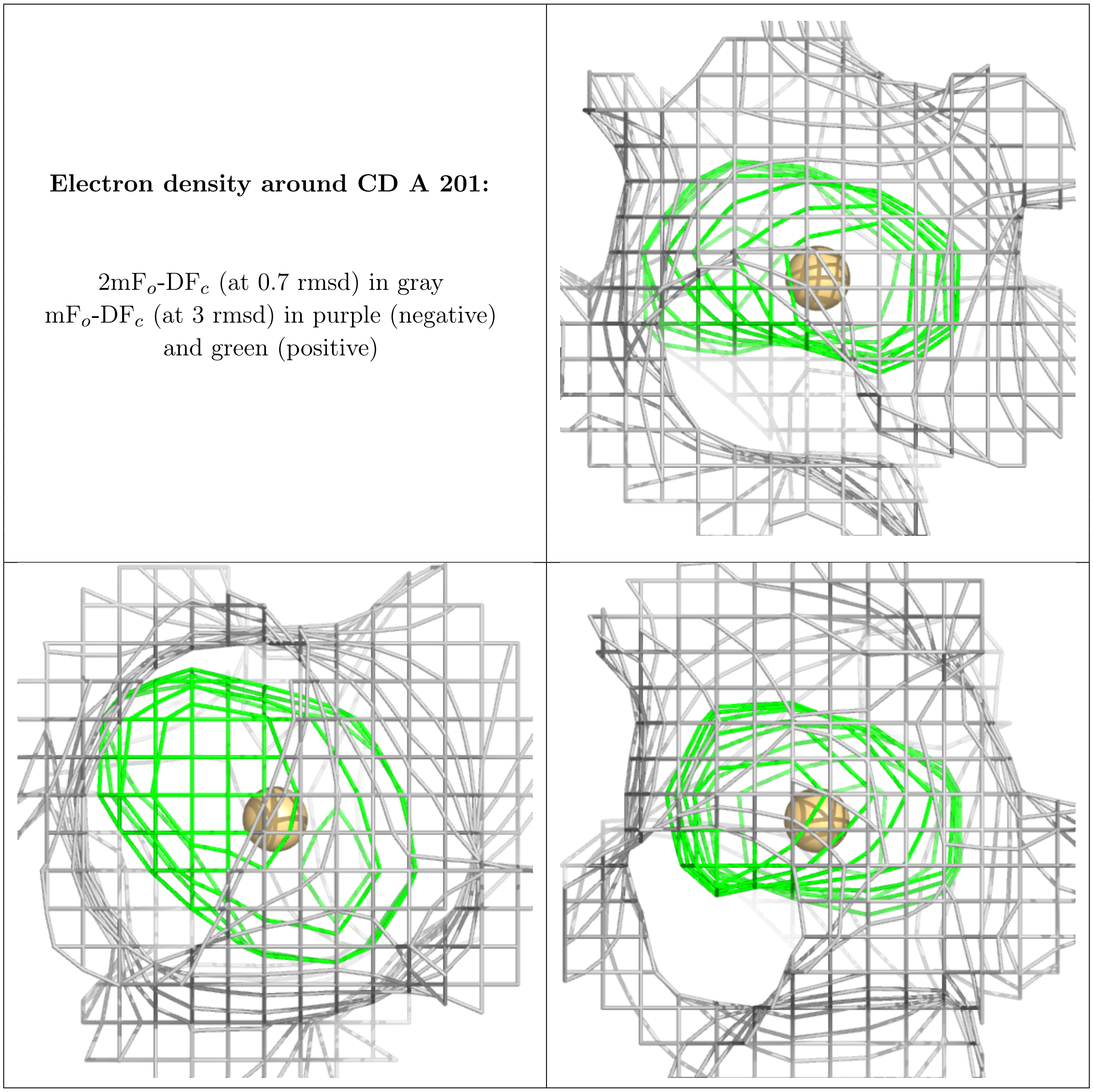
Electron density around CD A 202:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)









Electron density around CD A 201:

$2mF_o$-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.