Full wwPDB X-ray Structure Validation Report ⓘ

May 5, 2024 – 06:28 PM EDT

PDB ID	:	8UJM
Title	:	Crystal structure of human CTDNEP1-NEP1R1 protein phosphatase complex with magnesium
Authors	:	Gao, S.; Airola, M.V.
Deposited on	:	2023-10-11
Resolution	:	2.16 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 ⓘ 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 ⓘ) were used in the production of this report:

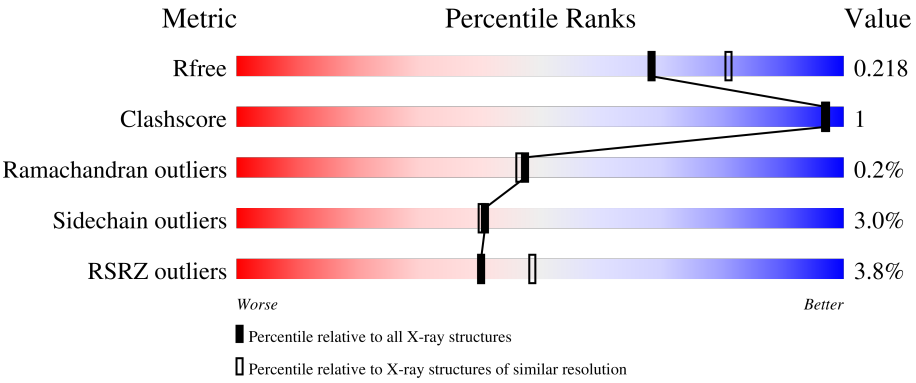
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.16 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	276	2% 82% . 15%
1	B	276	4% 80% 5% 14%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7836 atoms, of which 3782 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 CTD nuclear envelope phosphatase 1,Nuclear envelope phosphatase-regulatory subunit 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	235	Total	C	H	N	O	S	0	0	0
			3765	1197	1883	332	347	6			
1	B	237	Total	C	H	N	O	S	0	0	0
			3794	1205	1899	334	350	6			

There are 78 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP O95476
A	207	GLY	-	linker	UNP O95476
A	208	SER	-	linker	UNP O95476
A	209	ALA	-	linker	UNP O95476
A	210	LYS	-	linker	UNP O95476
A	211	GLY	-	linker	UNP O95476
A	212	SER	-	linker	UNP O95476
A	213	GLU	-	linker	UNP O95476
A	214	SER	-	linker	UNP O95476
A	215	ASN	-	linker	UNP O95476
A	216	SER	-	linker	UNP O95476
A	217	LEU	-	linker	UNP O95476
A	218	GLU	-	linker	UNP O95476
A	219	GLN	-	linker	UNP O95476
A	220	ALA	-	linker	UNP O95476
A	221	GLU	-	linker	UNP O95476
A	222	ASP	-	linker	UNP O95476
A	223	LEU	-	linker	UNP O95476
A	224	LYS	-	linker	UNP O95476
A	225	ALA	-	linker	UNP O95476
A	226	PHE	-	linker	UNP O95476
A	227	GLU	-	linker	UNP O95476
A	228	ARG	-	linker	UNP O95476
A	229	ARG	-	linker	UNP O95476

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	230	LEU	-	linker	UNP O95476
A	231	THR	-	linker	UNP O95476
A	232	GLU	-	linker	UNP O95476
A	233	TYR	-	linker	UNP O95476
A	234	ILE	-	linker	UNP O95476
A	235	HIS	-	linker	UNP O95476
A	236	CYS	-	linker	UNP O95476
A	237	LEU	-	linker	UNP O95476
A	238	GLN	-	linker	UNP O95476
A	271	HIS	-	expression tag	UNP H3BUT5
A	272	HIS	-	expression tag	UNP H3BUT5
A	273	HIS	-	expression tag	UNP H3BUT5
A	274	HIS	-	expression tag	UNP H3BUT5
A	275	HIS	-	expression tag	UNP H3BUT5
A	276	HIS	-	expression tag	UNP H3BUT5
B	1	MET	-	initiating methionine	UNP O95476
B	207	GLY	-	linker	UNP O95476
B	208	SER	-	linker	UNP O95476
B	209	ALA	-	linker	UNP O95476
B	210	LYS	-	linker	UNP O95476
B	211	GLY	-	linker	UNP O95476
B	212	SER	-	linker	UNP O95476
B	213	GLU	-	linker	UNP O95476
B	214	SER	-	linker	UNP O95476
B	215	ASN	-	linker	UNP O95476
B	216	SER	-	linker	UNP O95476
B	217	LEU	-	linker	UNP O95476
B	218	GLU	-	linker	UNP O95476
B	219	GLN	-	linker	UNP O95476
B	220	ALA	-	linker	UNP O95476
B	221	GLU	-	linker	UNP O95476
B	222	ASP	-	linker	UNP O95476
B	223	LEU	-	linker	UNP O95476
B	224	LYS	-	linker	UNP O95476
B	225	ALA	-	linker	UNP O95476
B	226	PHE	-	linker	UNP O95476
B	227	GLU	-	linker	UNP O95476
B	228	ARG	-	linker	UNP O95476
B	229	ARG	-	linker	UNP O95476
B	230	LEU	-	linker	UNP O95476
B	231	THR	-	linker	UNP O95476
B	232	GLU	-	linker	UNP O95476

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	233	TYR	-	linker	UNP O95476
B	234	ILE	-	linker	UNP O95476
B	235	HIS	-	linker	UNP O95476
B	236	CYS	-	linker	UNP O95476
B	237	LEU	-	linker	UNP O95476
B	238	GLN	-	linker	UNP O95476
B	271	HIS	-	expression tag	UNP H3BUT5
B	272	HIS	-	expression tag	UNP H3BUT5
B	273	HIS	-	expression tag	UNP H3BUT5
B	274	HIS	-	expression tag	UNP H3BUT5
B	275	HIS	-	expression tag	UNP H3BUT5
B	276	HIS	-	expression tag	UNP H3BUT5

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Mg 1 1	0	0
2	B	1	Total Mg 1 1	0	0

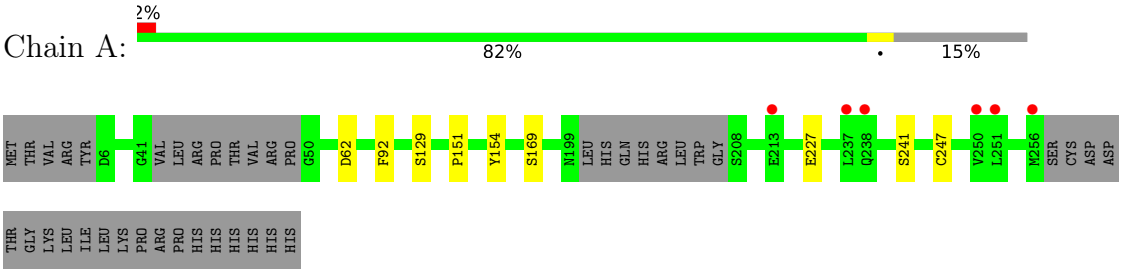
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	138	Total O 138 138	0	0
3	B	137	Total O 137 137	0	0

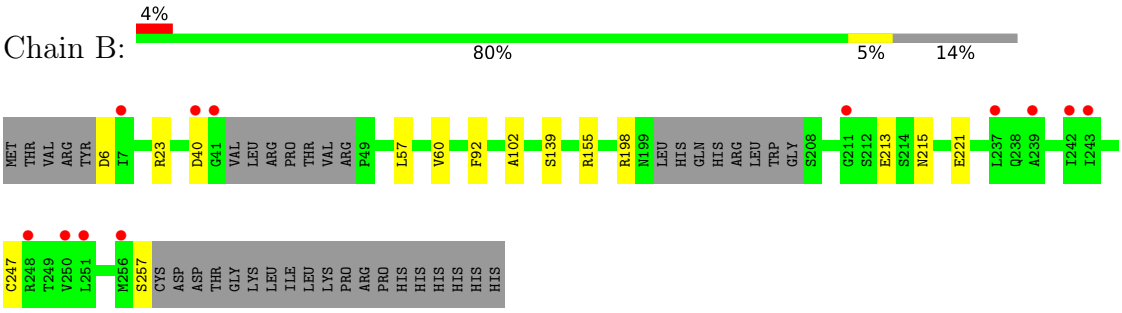
3 Residue-property plots (i)

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: CTD nuclear envelope phosphatase 1,Nuclear envelope phosphatase-regulatory subunit 1



• Molecule 1: CTD nuclear envelope phosphatase 1,Nuclear envelope phosphatase-regulatory subunit 1



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	56.41Å 102.95Å 57.24Å 90.00° 118.20° 90.00°	Depositor
Resolution (Å)	50.45 – 2.16 51.48 – 2.16	Depositor EDS
% Data completeness (in resolution range)	99.5 (50.45-2.16) 99.6 (51.48-2.16)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.62 (at 2.16Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.197 , 0.221 0.197 , 0.218	Depositor DCC
R_{free} test set	1491 reflections (4.85%)	wwPDB-VP
Wilson B-factor (Å^2)	33.7	Xtriage
Anisotropy	0.177	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 47.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.000 for -h-l,k,h 0.000 for l,k,-h-l 0.012 for h,-k,-h-l 0.009 for -h-l,-k,l 0.206 for l,-k,h	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7836	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.22% 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: MG

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	0.24	0/1922	0.50	0/2606
1	B	0.24	0/1936	0.49	0/2625
All	All	0.24	0/3858	0.50	0/5231

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	1882	1883	1888	1	1
1	B	1895	1899	1901	6	1
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	138	0	0	0	0
3	B	137	0	0	1	0
All	All	4054	3782	3789	7	1

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 1.

All (7) 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:198:ARG:NH1	1:B:213:GLU:OE1	2.39	0.56
1:B:57:LEU:CD2	1:B:102:ALA:HB1	2.37	0.55
1:B:57:LEU:HD21	1:B:102:ALA:HB1	1.93	0.51
1:B:60:VAL:HG23	1:B:60:VAL:O	2.18	0.44
1:B:215:ASN:ND2	3:B:510:HOH:O	2.51	0.44
1:A:151:PRO:HA	1:A:154:TYR:CZ	2.56	0.40
1:B:57:LEU:HD22	1:B:102:ALA:HB1	2.04	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:227:GLU:OE2	1:B:155:ARG:HE[1_655]	1.56	0.04

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	229/276 (83%)	220 (96%)	8 (4%)	1 (0%)	34 29
1	B	231/276 (84%)	223 (96%)	8 (4%)	0	100 100
All	All	460/552 (83%)	443 (96%)	16 (4%)	1 (0%)	47 46

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	169	SER

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	213/252 (84%)	208 (98%)	5 (2%)	50	53
1	B	215/252 (85%)	207 (96%)	8 (4%)	34	32
All	All	428/504 (85%)	415 (97%)	13 (3%)	41	40

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	62	ASP
1	A	92	PHE
1	A	129	SER
1	A	241	SER
1	A	247	CYS
1	B	6	ASP
1	B	23	ARG
1	B	40	ASP
1	B	92	PHE
1	B	139	SER
1	B	221	GLU
1	B	247	CYS
1	B	257	SER

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	39	HIS
1	B	215	ASN

5.3.3 RNA (i)

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 monosaccharides in this entry.

5.6 Ligand geometry

Of 2 ligands modelled in this entry, 2 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

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	235/276 (85%)	0.02	6 (2%)	56	64	25, 38, 93, 104	0
1	B	237/276 (85%)	0.13	12 (5%)	28	36	26, 39, 100, 141	0
All	All	472/552 (85%)	0.08	18 (3%)	40	49	25, 39, 96, 141	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	251	LEU	4.2
1	B	251	LEU	4.0
1	B	239	ALA	3.5
1	A	238	GLN	3.1
1	B	256	MET	3.1
1	B	40	ASP	3.0
1	B	250	VAL	2.9
1	A	213	GLU	2.9
1	A	250	VAL	2.9
1	A	256	MET	2.5
1	B	248	ARG	2.4
1	B	242	ILE	2.2
1	B	243	ILE	2.2
1	A	237	LEU	2.1
1	B	211	GLY	2.1
1	B	237	LEU	2.1
1	B	41	GLY	2.1
1	B	7	ILE	2.1

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 (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

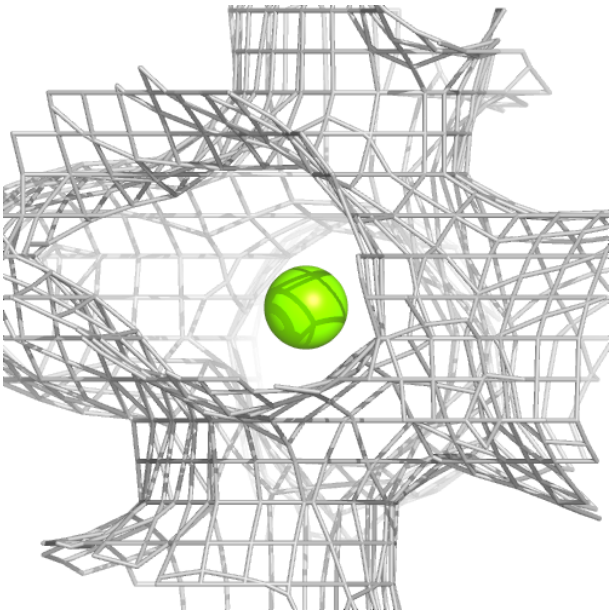
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

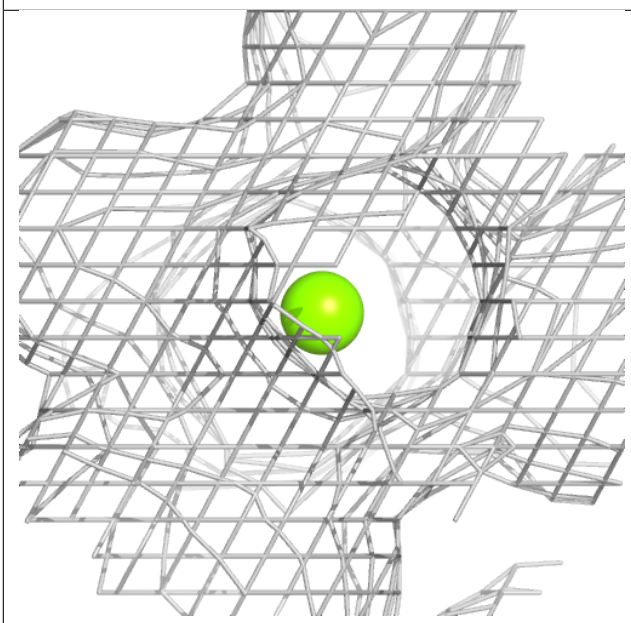
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	MG	A	401	1/1	0.99	0.09	34,34,34,34	0
2	MG	B	401	1/1	0.99	0.10	36,36,36,36	0

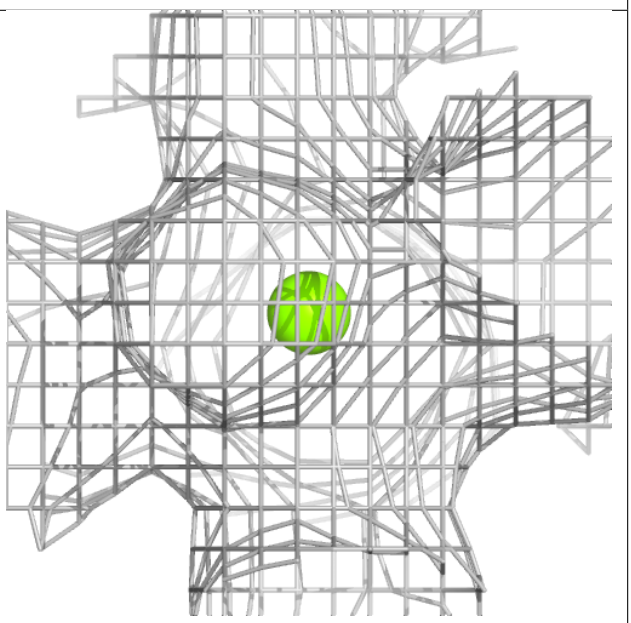
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around MG A 401:

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

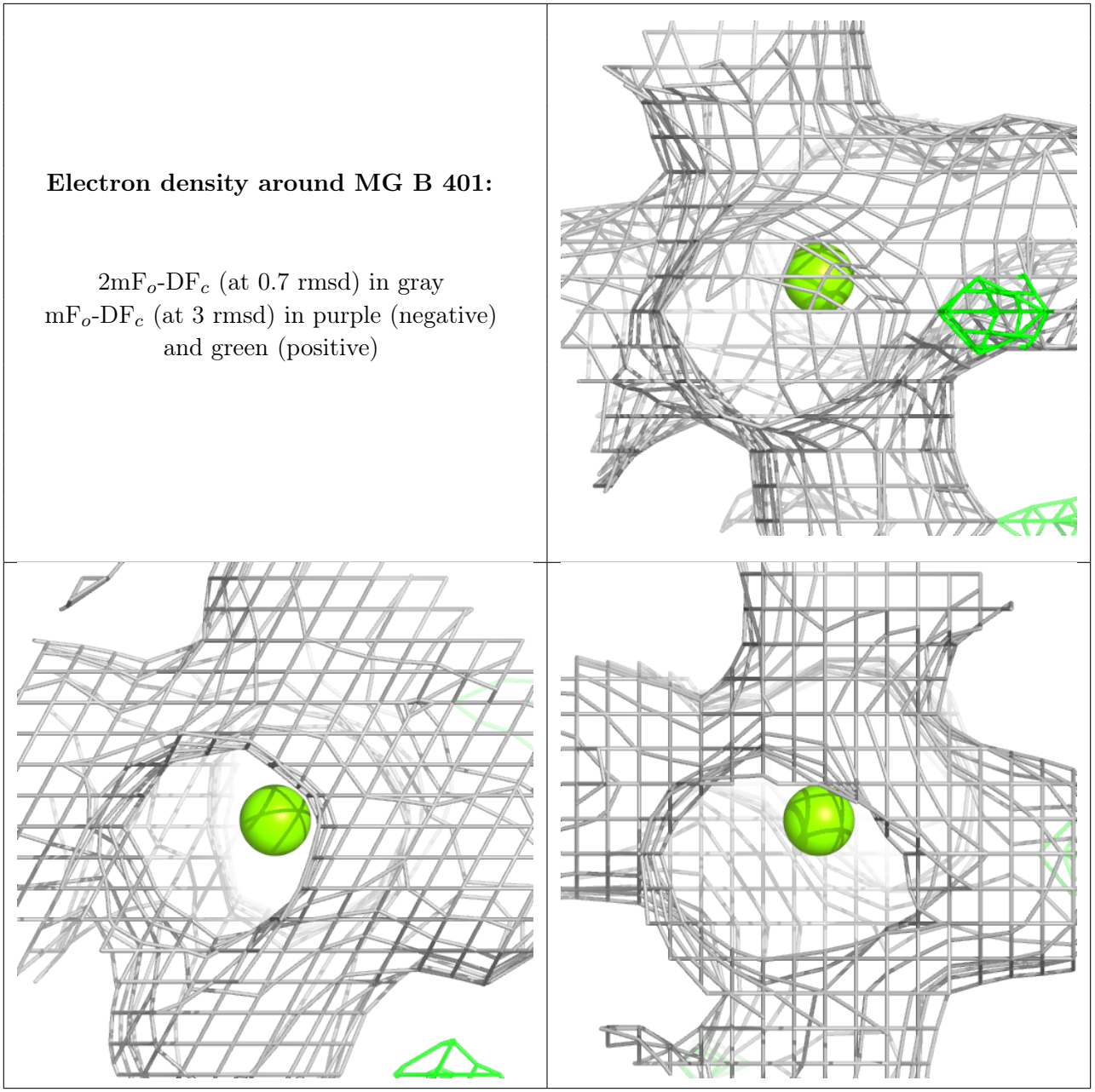






Electron density around MG B 401:

$2mF_o$-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.