wwPDB X-ray Structure Validation Summary Report (i)

Aug 7, 2020 – 09:24 PM BST

PDB ID	:	6L7W
Title	:	Crystal structure of Cet1 from Trypanosoma cruzi in complex with manganese ion.
Authors	:	Kuwabara, N.; Ho, K.
Deposited on	:	2019-11-03
Resolution	:	2.60 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 following versions of software and data (see references (1)) were used in the production of this report:

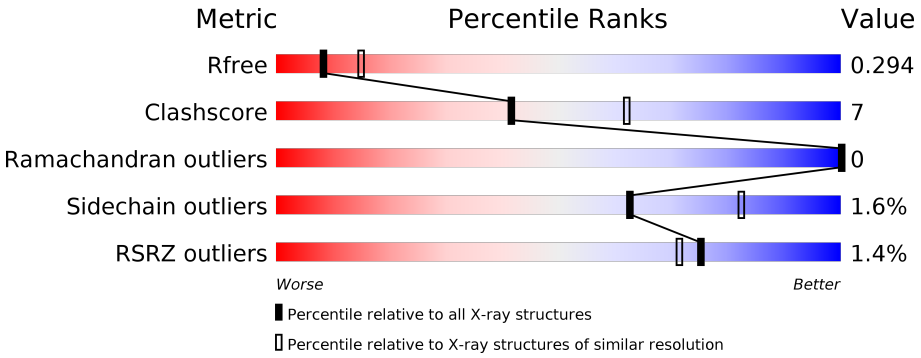
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 on 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	209	% 68% 20% 11%
1	B	209	% 74% 11% 15%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2893 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 mRNA_triPase domain-containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	185	Total	C	N	O	S	0	0	0
			1456	912	255	279	10			
1	B	177	Total	C	N	O	S	0	0	0
			1394	878	242	264	10			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	14	GLY	-	expression tag	UNP Q4E2I1
A	15	PRO	-	expression tag	UNP Q4E2I1
A	16	GLY	-	expression tag	UNP Q4E2I1
A	17	SER	-	expression tag	UNP Q4E2I1
A	?	-	ASP	deletion	UNP Q4E2I1
A	?	-	ALA	deletion	UNP Q4E2I1
A	?	-	LEU	deletion	UNP Q4E2I1
A	?	-	ARG	deletion	UNP Q4E2I1
A	?	-	GLU	deletion	UNP Q4E2I1
A	?	-	ASN	deletion	UNP Q4E2I1
A	?	-	THR	deletion	UNP Q4E2I1
A	?	-	SER	deletion	UNP Q4E2I1
A	?	-	ASP	deletion	UNP Q4E2I1
A	?	-	GLN	deletion	UNP Q4E2I1
A	?	-	ASN	deletion	UNP Q4E2I1
A	?	-	GLY	deletion	UNP Q4E2I1
A	?	-	ALA	deletion	UNP Q4E2I1
A	?	-	GLN	deletion	UNP Q4E2I1
A	?	-	THR	deletion	UNP Q4E2I1
A	?	-	ASN	deletion	UNP Q4E2I1
A	?	-	GLY	deletion	UNP Q4E2I1
A	?	-	LEU	deletion	UNP Q4E2I1
A	?	-	ARG	deletion	UNP Q4E2I1
A	?	-	ILE	deletion	UNP Q4E2I1
A	?	-	VAL	deletion	UNP Q4E2I1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	126	ASN	ASP	engineered mutation	UNP Q4E2I1
B	14	GLY	-	expression tag	UNP Q4E2I1
B	15	PRO	-	expression tag	UNP Q4E2I1
B	16	GLY	-	expression tag	UNP Q4E2I1
B	17	SER	-	expression tag	UNP Q4E2I1
B	?	-	ASP	deletion	UNP Q4E2I1
B	?	-	ALA	deletion	UNP Q4E2I1
B	?	-	LEU	deletion	UNP Q4E2I1
B	?	-	ARG	deletion	UNP Q4E2I1
B	?	-	GLU	deletion	UNP Q4E2I1
B	?	-	ASN	deletion	UNP Q4E2I1
B	?	-	THR	deletion	UNP Q4E2I1
B	?	-	SER	deletion	UNP Q4E2I1
B	?	-	ASP	deletion	UNP Q4E2I1
B	?	-	GLN	deletion	UNP Q4E2I1
B	?	-	ASN	deletion	UNP Q4E2I1
B	?	-	GLY	deletion	UNP Q4E2I1
B	?	-	ALA	deletion	UNP Q4E2I1
B	?	-	GLN	deletion	UNP Q4E2I1
B	?	-	THR	deletion	UNP Q4E2I1
B	?	-	ASN	deletion	UNP Q4E2I1
B	?	-	GLY	deletion	UNP Q4E2I1
B	?	-	LEU	deletion	UNP Q4E2I1
B	?	-	ARG	deletion	UNP Q4E2I1
B	?	-	ILE	deletion	UNP Q4E2I1
B	?	-	VAL	deletion	UNP Q4E2I1
B	126	ASN	ASP	engineered mutation	UNP Q4E2I1

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn) (labeled as "Ligand of Interest" by author).

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

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	19	Total O 19 19	0	0

Continued on next page...

Continued from previous page...

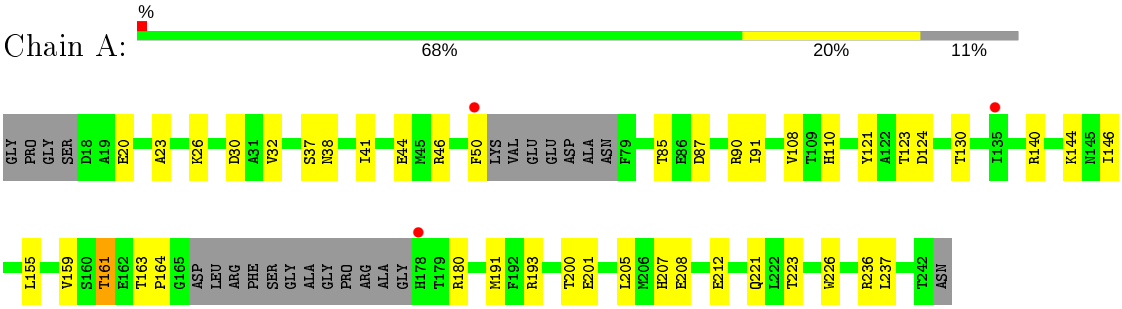
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	22	Total O 22 22	0	0

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.

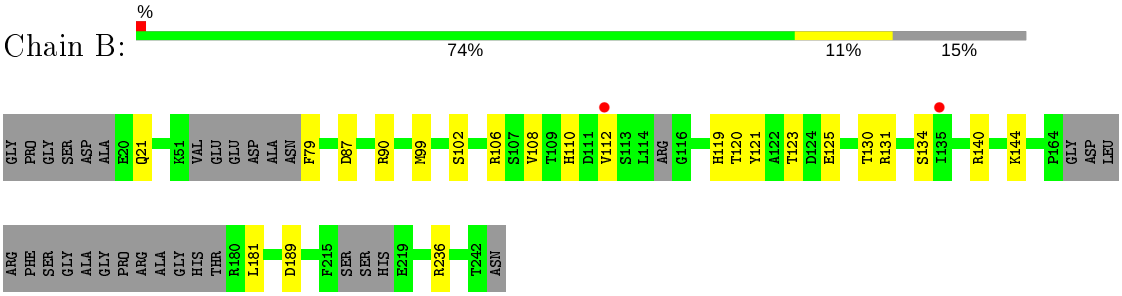
- Molecule 1: mRNA_triPase domain-containing protein

Chain A:



- Molecule 1: mRNA_triPase domain-containing protein

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	52.45Å 73.21Å 63.75Å 90.00° 114.06° 90.00°	Depositor
Resolution (Å)	47.89 – 2.60 47.89 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.89-2.60) 90.6 (47.89-2.60)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.33 (at 2.61Å)	Xtriage
Refinement program	PHENIX 1.13_2998, PHENIX 1.13_2998	Depositor
R, R_{free}	0.239 , 0.294 0.239 , 0.294	Depositor DCC
R_{free} test set	632 reflections (4.64%)	wwPDB-VP
Wilson B-factor (Å^2)	37.9	Xtriage
Anisotropy	0.685	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 56.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.023 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	2893	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.20% 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: MN

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	0.28	0/1480	0.47	0/2001
1	B	0.28	0/1414	0.47	0/1910
All	All	0.28	0/2894	0.47	0/3911

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	1456	0	1406	26	0
1	B	1394	0	1353	16	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	19	0	0	3	0
3	B	22	0	0	3	0
All	All	2893	0	2759	39	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 7.

The worst 5 of 39 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:44:GLU:HG2	1:A:212:GLU:HG2	1.66	0.77
1:A:44:GLU:OE1	1:A:46:ARG:NH1	2.24	0.71
1:A:200:THR:HG22	1:A:205:LEU:HD23	1.73	0.69
1:B:108:VAL:HB	1:B:123:THR:HG21	1.74	0.69
1:A:110:HIS:HB3	1:A:121:TYR:HB2	1.74	0.67

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	179/209 (86%)	174 (97%)	5 (3%)	0	100	100
1	B	167/209 (80%)	164 (98%)	3 (2%)	0	100	100
All	All	346/418 (83%)	338 (98%)	8 (2%)	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	157/182 (86%)	153 (98%)	4 (2%)	47	73
1	B	150/182 (82%)	149 (99%)	1 (1%)	84	94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	307/364 (84%)	302 (98%)	5 (2%)	62 82

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	20	GLU
1	A	37	SER
1	A	50	PHE
1	A	161	THR
1	B	236	ARG

Some 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

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	185/209 (88%)	0.25	3 (1%)	72	68	34, 54, 88, 115	0
1	B	177/209 (84%)	0.24	2 (1%)	80	78	34, 52, 93, 126	0
All	All	362/418 (86%)	0.25	5 (1%)	75	71	34, 53, 89, 126	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	50	PHE	4.1
1	A	178	HIS	3.0
1	B	135	ILE	2.5
1	B	112	VAL	2.4
1	A	135	ILE	2.2

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

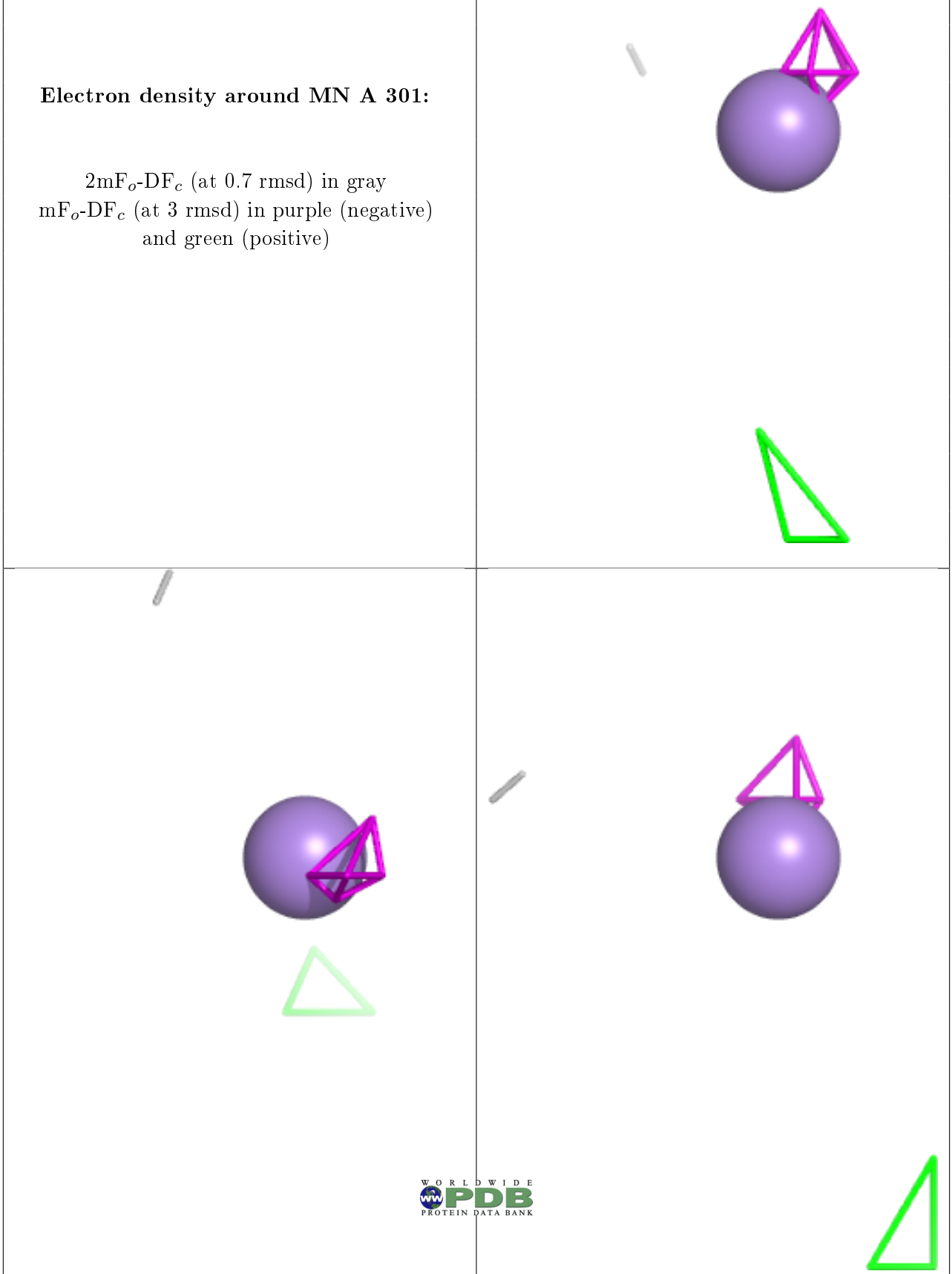
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	MN	A	301	1/1	0.91	0.07	59,59,59,59	0
2	MN	B	301	1/1	0.99	0.07	47,47,47,47	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.

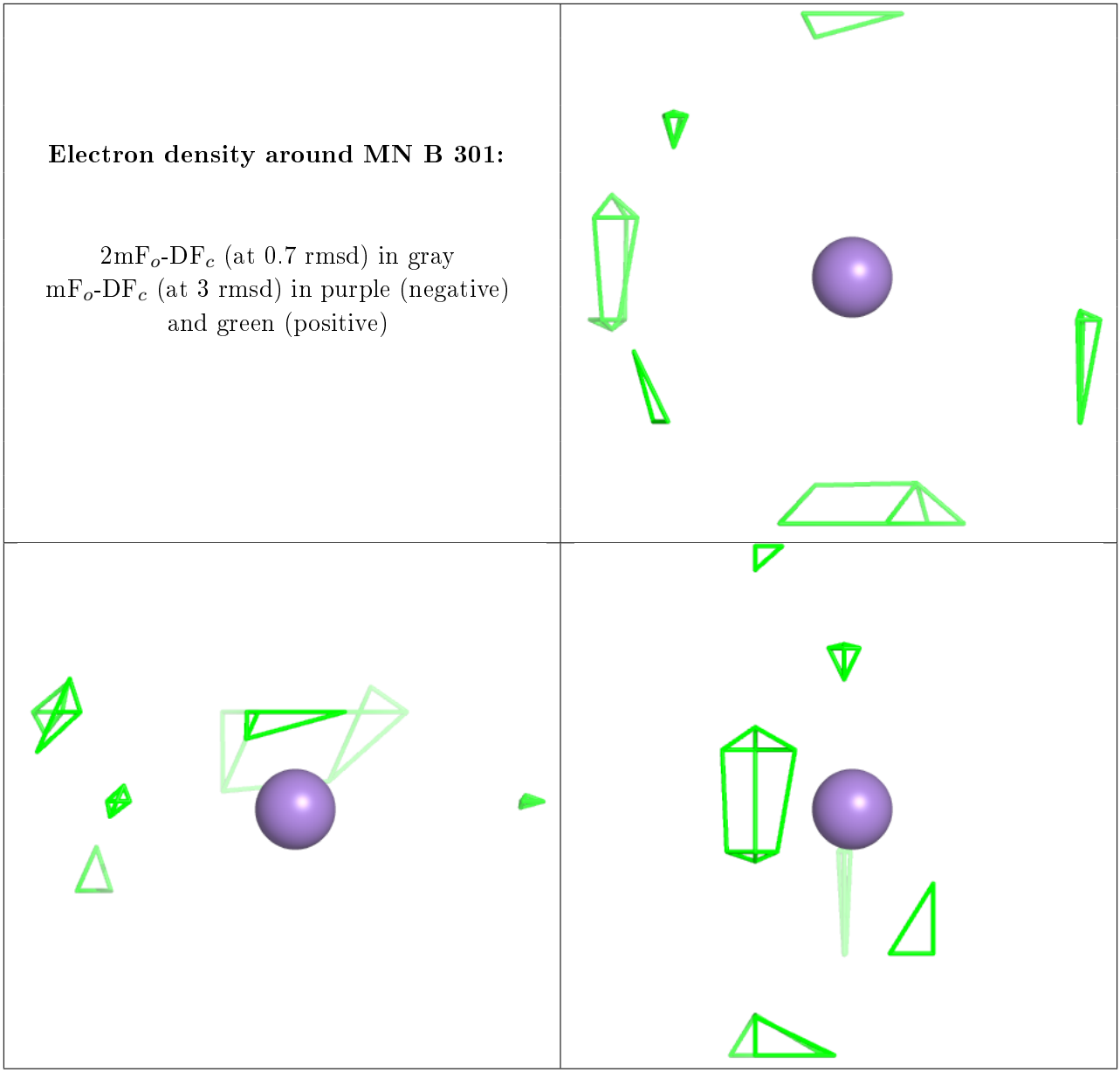
Electron density around MN A 301:

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



Electron density around MN B 301:

$2mF_o$-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.