wwPDB X-ray Structure Validation Summary Report (i)

Aug 13, 2023 – 12:07 AM JST

PDB ID	:	8IDF
Title	:	Crystal structure of human TUT1 complexed with U6 snRNA
Authors	:	Yamashita, S.; Tomita, K.
Deposited on	:	2023-02-13
Resolution	:	3.70 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:

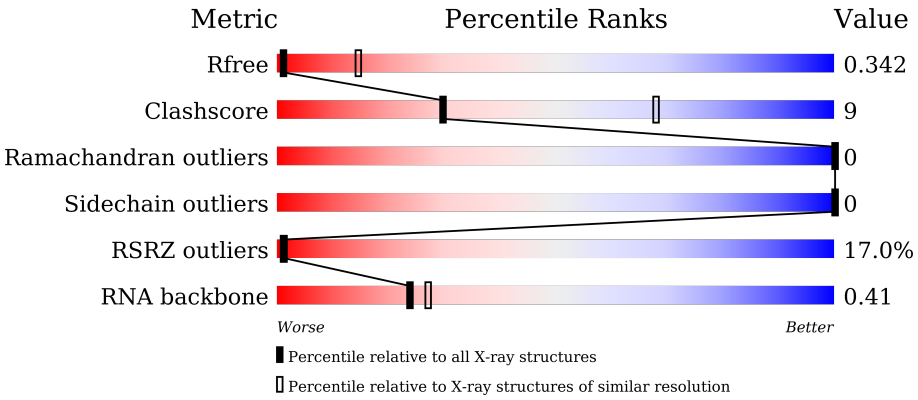
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.70 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	1049 (3.88-3.52)
Clashscore	141614	1027 (3.86-3.54)
Ramachandran outliers	138981	1069 (3.88-3.52)
Sidechain outliers	138945	1065 (3.88-3.52)
RSRZ outliers	127900	1578 (3.90-3.50)
RNA backbone	3102	1027 (4.40-3.00)

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	537	11% 69% 17% 15%
2	B	53	57% 26% 51% 23%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4658 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 Speckle targeted PIP5K1A-regulated poly(A) polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	459	3529	2213	658	645	13	0	0	0

There are 86 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	19	SER	CYS	engineered mutation	UNP Q9H6E5
A	218	ALA	ASP	engineered mutation	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	GLU	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	LEU	deletion	UNP Q9H6E5
A	?	-	ASP	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	LEU	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	LEU	deletion	UNP Q9H6E5
A	?	-	ASP	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	GLN	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	LEU	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	CYS	deletion	UNP Q9H6E5
A	?	-	THR	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	ASP	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	GLN	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	GLN	deletion	UNP Q9H6E5
A	?	-	ASP	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	GLU	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	LEU	deletion	UNP Q9H6E5
A	?	-	ASP	deletion	UNP Q9H6E5
A	?	-	PHE	deletion	UNP Q9H6E5
A	?	-	GLU	deletion	UNP Q9H6E5
A	?	-	THR	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	LEU	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	GLN	deletion	UNP Q9H6E5
A	?	-	THR	deletion	UNP Q9H6E5
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	ASP	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	LEU	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	?	-	GLU	deletion	UNP Q9H6E5
A	?	-	THR	deletion	UNP Q9H6E5
A	?	-	LEU	deletion	UNP Q9H6E5
A	?	-	ALA	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	PRO	deletion	UNP Q9H6E5
A	?	-	GLN	deletion	UNP Q9H6E5
A	?	-	SER	deletion	UNP Q9H6E5
A	372	ALA	CYS	engineered mutation	UNP Q9H6E5
A	399	ALA	CYS	engineered mutation	UNP Q9H6E5
A	415	ALA	CYS	engineered mutation	UNP Q9H6E5
A	501	ALA	CYS	engineered mutation	UNP Q9H6E5
A	504	SER	CYS	engineered mutation	UNP Q9H6E5
A	574	ALA	CYS	engineered mutation	UNP Q9H6E5
A	600	LEU	-	expression tag	UNP Q9H6E5
A	601	GLU	-	expression tag	UNP Q9H6E5
A	602	HIS	-	expression tag	UNP Q9H6E5
A	603	HIS	-	expression tag	UNP Q9H6E5
A	604	HIS	-	expression tag	UNP Q9H6E5
A	605	HIS	-	expression tag	UNP Q9H6E5
A	606	HIS	-	expression tag	UNP Q9H6E5
A	607	HIS	-	expression tag	UNP Q9H6E5

- Molecule 2 is a RNA chain called RNA (53-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	53	Total	C	N	O	P	0	0	0
			1128	505	199	371	53			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

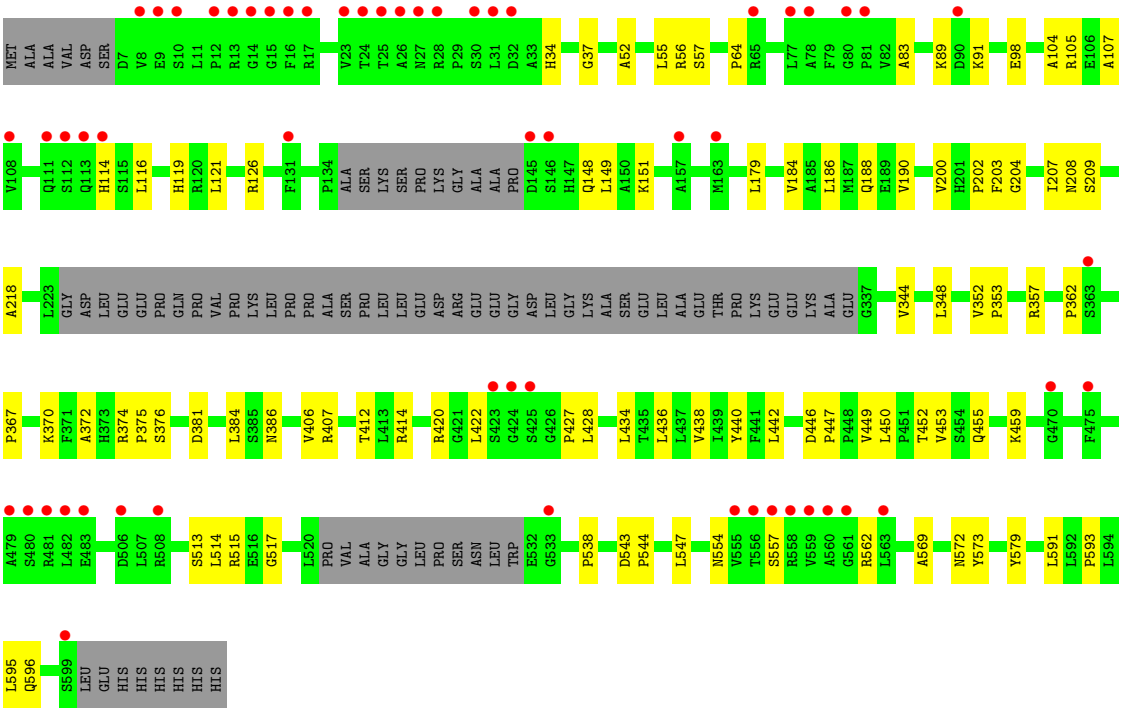
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		

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.

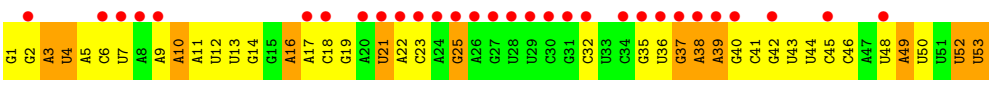
• Molecule 1: Speckle targeted PIP5K1A-regulated poly(A) polymerase

Chain A:



• Molecule 2: RNA (53-MER)

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	78.72Å 78.72Å 370.27Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.99 – 3.70 45.75 – 3.70	Depositor EDS
% Data completeness (in resolution range)	99.3 (19.99-3.70) 99.5 (45.75-3.70)	Depositor EDS
R_{merge}	0.53	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.89 (at 3.66Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.305 , 0.339 0.319 , 0.342	Depositor DCC
R_{free} test set	400 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	120.3	Xtriage
Anisotropy	0.624	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 142.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	4658	wwPDB-VP
Average B, all atoms (Å^2)	209.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.91% 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: ZN

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	0/3597	0.40	0/4867
2	B	0.17	0/1261	0.80	0/1962
All	All	0.22	0/4858	0.55	0/6829

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	3529	0	3537	63	0
2	B	1128	0	568	26	0
3	A	1	0	0	0	0
All	All	4658	0	4105	82	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 9.

The worst 5 of 82 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:204:GLY:HA2	2:B:53:U:H4'	1.69	0.73
1:A:209:SER:O	1:A:407:ARG:NH1	2.28	0.67
1:A:420:ARG:NH2	1:A:593:PRO:O	2.28	0.66
1:A:453:VAL:HB	1:A:515:ARG:HA	1.80	0.64
1:A:420:ARG:NH1	1:A:596:GLN:O	2.29	0.64

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	451/537 (84%)	429 (95%)	22 (5%)	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	380/442 (86%)	380 (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. There are no such sidechains identified.

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	52/53 (98%)	22 (42%)	2 (3%)

5 of 22 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	3	A
2	B	4	U
2	B	5	A
2	B	6	C
2	B	7	U

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	37	G
2	B	52	U

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 1 ligands modelled in this entry, 1 is 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	459/537 (85%)	0.60	57 (12%)	4	4	63, 169, 281, 395	0
2	B	53/53 (100%)	3.37	30 (56%)	0	0	159, 303, 506, 896	0
All	All	512/590 (86%)	0.89	87 (16%)	1	1	63, 177, 366, 896	0

The worst 5 of 87 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	29	U	13.4
2	B	34	C	9.2
2	B	30	C	8.2
2	B	9	A	7.5
2	B	26	A	7.3

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 monosaccharides in this entry.

6.4 Ligands

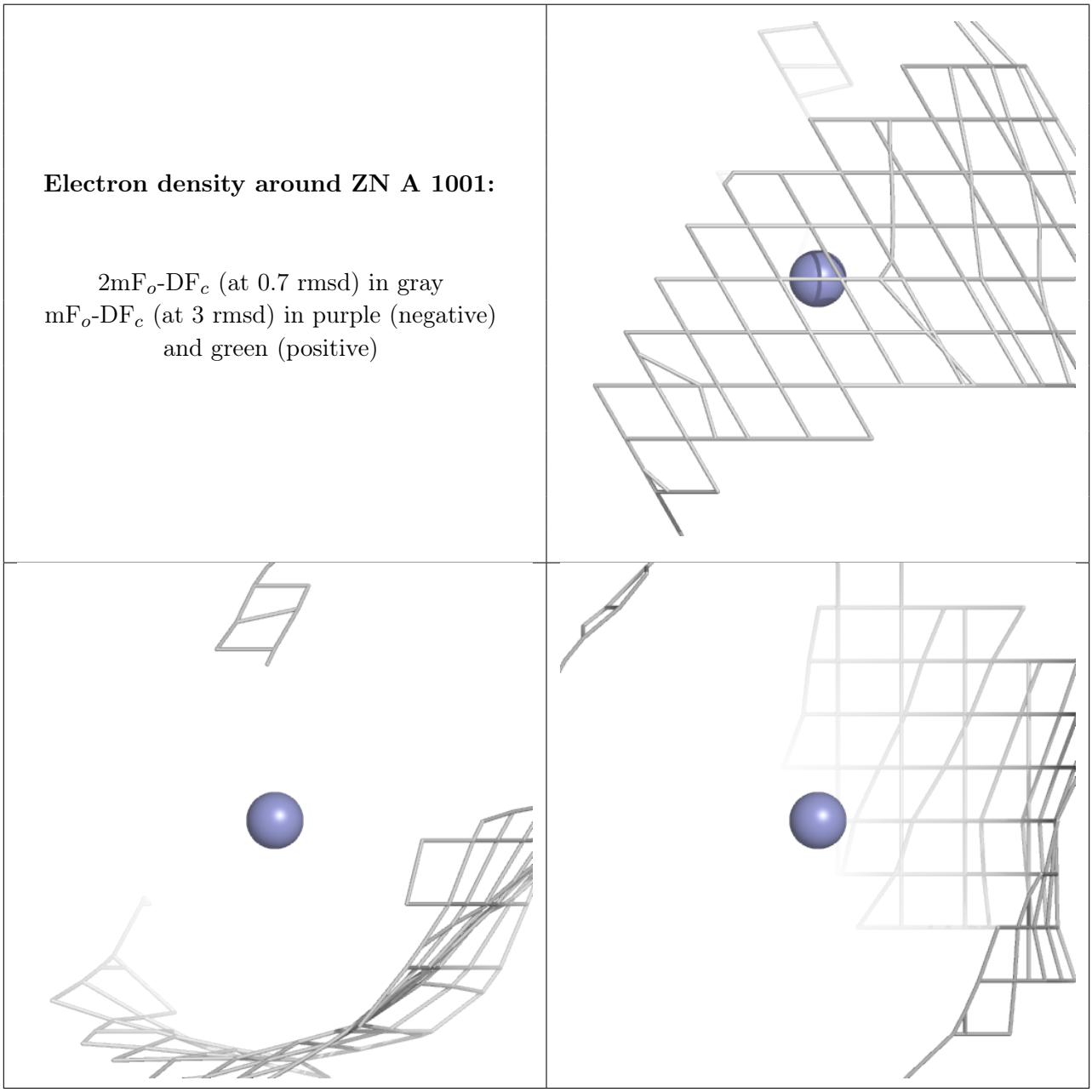
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
3	ZN	A	1001	1/1	0.95	0.03	170,170,170,170	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 ZN A 1001:

$2mF_o$-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 (i)

There are no such residues in this entry.