W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Aug 7, 2023 – 11:22 AM EDT

PDB ID	:	7UU5
Title	:	Crystal structure of APOBEC3G complex with 5'-Overhang dsRNA
Authors	:	Yang, H.; Li, S.; Chen, X.S.
Deposited on	:	2022-04-28
Resolution	:	2.90 Å(reported)

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

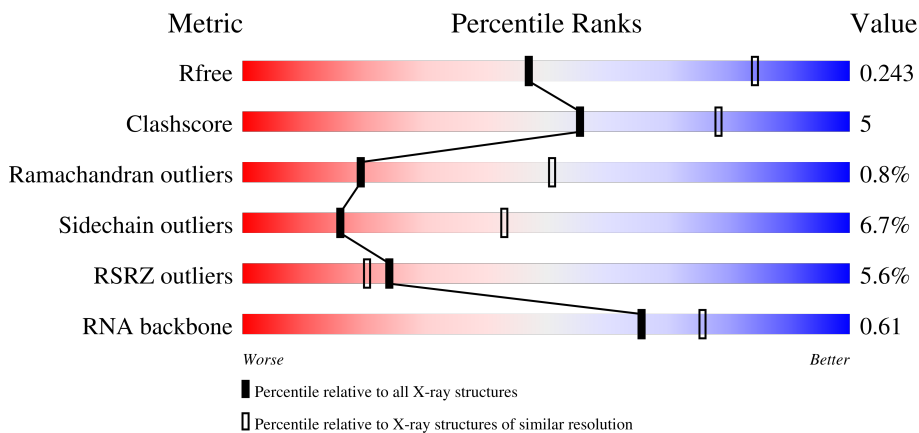|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	: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

1 Overall quality at a glance ⓘ

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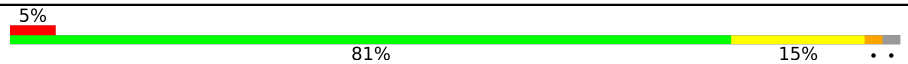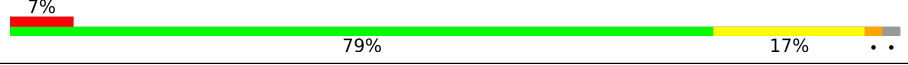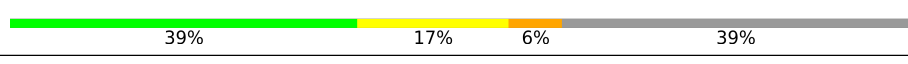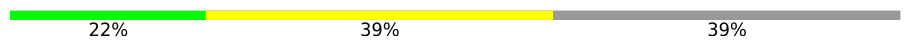*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	Percentile Ranks	Value
Rfree		0.243
Clashscore		5
Ramachandran outliers		0.8%
Sidechain outliers		6.7%
RSRZ outliers		5.6%
RNA backbone		0.61

Worse ← → *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)
RNA backbone	3102	1007 (3.16-2.64)

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	386	5% / 81% / 15% / ··
1	B	386	7% / 79% / 17% / ··
2	M	18	39% / 17% / 6% / 39%
3	N	18	22% / 39% / 39%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 6771 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 DNA dC->dU-editing enzyme APOBEC-3G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	379	Total 3148	C 2023	N 558	O 545	S 22	0	0	0
1	B	379	Total 3148	C 2023	N 558	O 545	S 22	0	0	0

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	GLY	-	expression tag	UNP M1GSK9
A	-5	PRO	-	expression tag	UNP M1GSK9
A	-4	GLY	-	expression tag	UNP M1GSK9
A	-3	GLY	-	expression tag	UNP M1GSK9
A	-2	SER	-	expression tag	UNP M1GSK9
A	-1	GLY	-	expression tag	UNP M1GSK9
A	0	GLY	-	expression tag	UNP M1GSK9
A	128	ASP	LYS	conflict	UNP M1GSK9
A	139	ALA	CYS	conflict	UNP M1GSK9
A	140	GLU	GLN	conflict	UNP M1GSK9
A	141	ALA	LYS	conflict	UNP M1GSK9
A	142	GLY	ARG	conflict	UNP M1GSK9
A	?	-	ASP	deletion	UNP M1GSK9
A	?	-	GLY	deletion	UNP M1GSK9
A	?	-	PRO	deletion	UNP M1GSK9
A	?	-	HIS	deletion	UNP M1GSK9
A	259	ALA	GLU	conflict	UNP M1GSK9
B	-6	GLY	-	expression tag	UNP M1GSK9
B	-5	PRO	-	expression tag	UNP M1GSK9
B	-4	GLY	-	expression tag	UNP M1GSK9
B	-3	GLY	-	expression tag	UNP M1GSK9
B	-2	SER	-	expression tag	UNP M1GSK9
B	-1	GLY	-	expression tag	UNP M1GSK9
B	0	GLY	-	expression tag	UNP M1GSK9
B	128	ASP	LYS	conflict	UNP M1GSK9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	139	ALA	CYS	conflict	UNP M1GSK9
B	140	GLU	GLN	conflict	UNP M1GSK9
B	141	ALA	LYS	conflict	UNP M1GSK9
B	142	GLY	ARG	conflict	UNP M1GSK9
B	?	-	ASP	deletion	UNP M1GSK9
B	?	-	GLY	deletion	UNP M1GSK9
B	?	-	PRO	deletion	UNP M1GSK9
B	?	-	HIS	deletion	UNP M1GSK9
B	259	ALA	GLU	conflict	UNP M1GSK9

- Molecule 2 is a RNA chain called RNA (5'-R(P*UP*AP*AP*CP*GP*CP*UP*GP*CP*GP *G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	M	11	236	105	43	77	11	0	0	0

- Molecule 3 is a RNA chain called RNA (5'-R(P*UP*AP*AP*CP*CP*GP*CP*AP*GP*CP *G)-3').

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	N	11	235	105	44	75	11	0	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
4	A	2	2	2	0	0
			Total	Zn		
4	B	2	2	2	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.
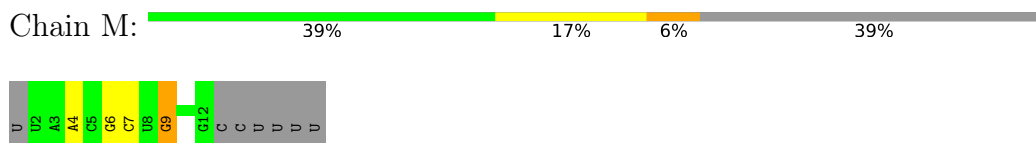
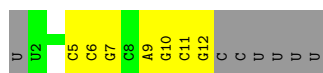• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3G
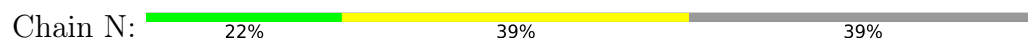
Chain A:

• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3G

Chain B:
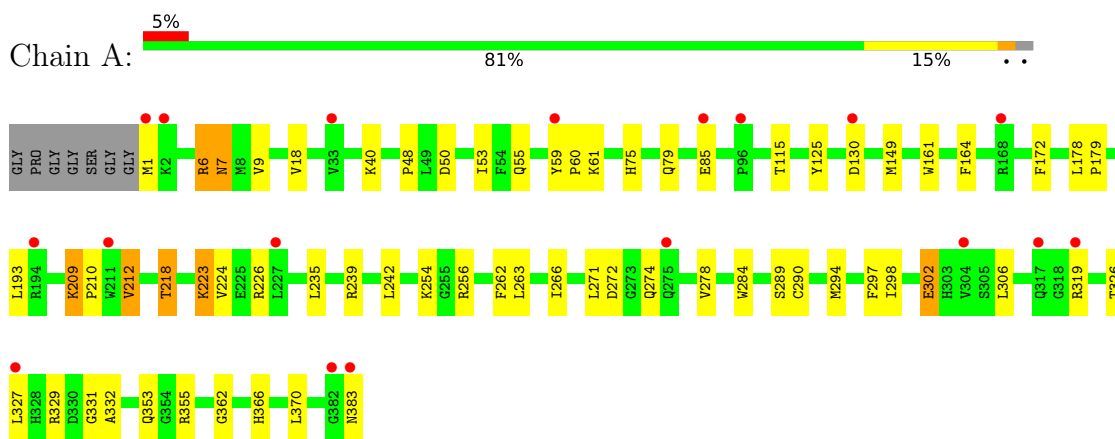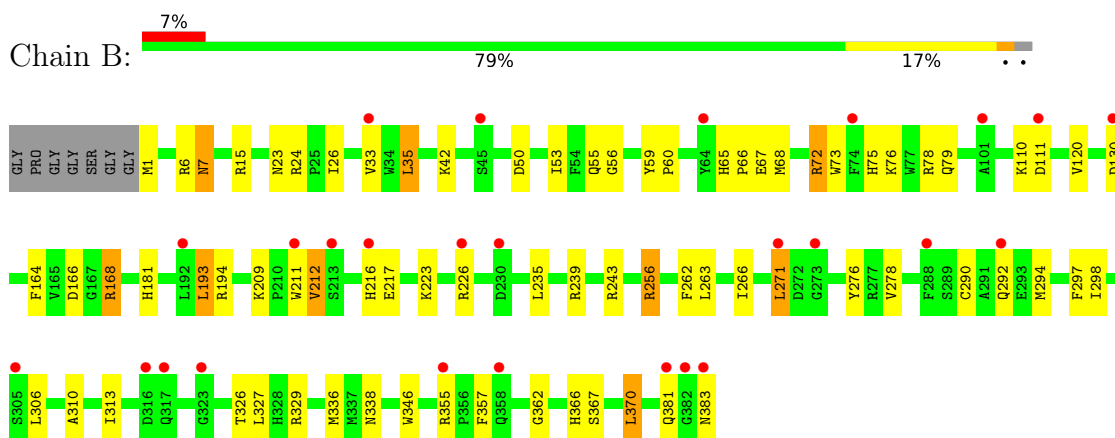
• Molecule 2: RNA (5'-R(P*UP*AP*AP*CP*GP*CP*UP*GP*CP*GP*G)-3')

Chain M:

• Molecule 3: RNA (5'-R(P*UP*AP*AP*CP*CP*GP*CP*AP*GP*CP*G)-3')

Chain N:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	61.30Å 71.56Å 72.71Å 119.39° 111.00° 90.08°	Depositor
Resolution (Å)	46.51 – 2.90 46.51 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.2 (46.51-2.90) 97.8 (46.51-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.84 (at 2.91Å)	Xtriage
Refinement program	PHENIX 1.16_3549	Depositor
R, R_{free}	0.187 , 0.243 0.187 , 0.243	Depositor DCC
R_{free} test set	1335 reflections (6.26%)	wwPDB-VP
Wilson B-factor (Å^2)	71.2	Xtriage
Anisotropy	0.495	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 58.2	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.379 for -h,k,-k-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6771	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.95% 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: 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	0.27	0/3253	0.46	0/4416
1	B	0.27	0/3253	0.49	1/4416 (0.0%)
2	M	0.26	0/263	0.97	0/408
3	N	0.24	0/262	0.89	0/406
All	All	0.27	0/7031	0.53	1/9646 (0.0%)

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

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	35	LEU	CA-CB-CG	5.11	127.05	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	319	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	3148	0	3028	27	0
1	B	3148	0	3028	32	0
2	M	236	0	120	3	0
3	N	235	0	121	7	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
All	All	6771	0	6297	66	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 5.

The worst 5 of 66 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:256:ARG:HH11	1:B:256:ARG:HG2	1.49	0.76
1:B:310:ALA:HB3	1:B:336:MET:HE1	1.71	0.72
1:B:217:GLU:OE1	1:B:243:ARG:NH2	2.27	0.69
1:A:362:GLY:O	1:A:366:HIS:ND1	2.26	0.68
1:B:298:ILE:HD11	1:B:306:LEU:HB2	1.78	0.65

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	377/386 (98%)	360 (96%)	15 (4%)	2 (0%)	29	61

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	377/386 (98%)	357 (95%)	16 (4%)	4 (1%)	14	42
All	All	754/772 (98%)	717 (95%)	31 (4%)	6 (1%)	19	51

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	7	ASN
1	B	42	LYS
1	B	212	VAL
1	A	7	ASN
1	A	212	VAL

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	337/339 (99%)	316 (94%)	21 (6%)	18	47
1	B	337/339 (99%)	313 (93%)	24 (7%)	14	40
All	All	674/678 (99%)	629 (93%)	45 (7%)	16	43

5 of 45 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	120	VAL
1	B	223	LYS
1	B	130	ASP
1	B	194	ARG
1	B	256	ARG

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	M	10/18 (55%)	1 (10%)	0
3	N	10/18 (55%)	0	0
All	All	20/36 (55%)	1 (5%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	M	9	G

There are no RNA pucker outliers to report.

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 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 ⓘ

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, 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	379/386 (98%)	0.34	18 (4%)	31	28	45, 76, 132, 189	0
1	B	379/386 (98%)	0.40	26 (6%)	16	13	47, 76, 136, 183	0
2	M	11/18 (61%)	-0.14	0	100	100	80, 143, 213, 215	0
3	N	11/18 (61%)	-0.07	0	100	100	80, 131, 192, 197	0
All	All	780/808 (96%)	0.35	44 (5%)	24	20	45, 77, 141, 215	0

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	382	GLY	8.3
1	B	216	HIS	5.8
1	A	59	TYR	5.7
1	B	383	ASN	5.1
1	B	230	ASP	4.5

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

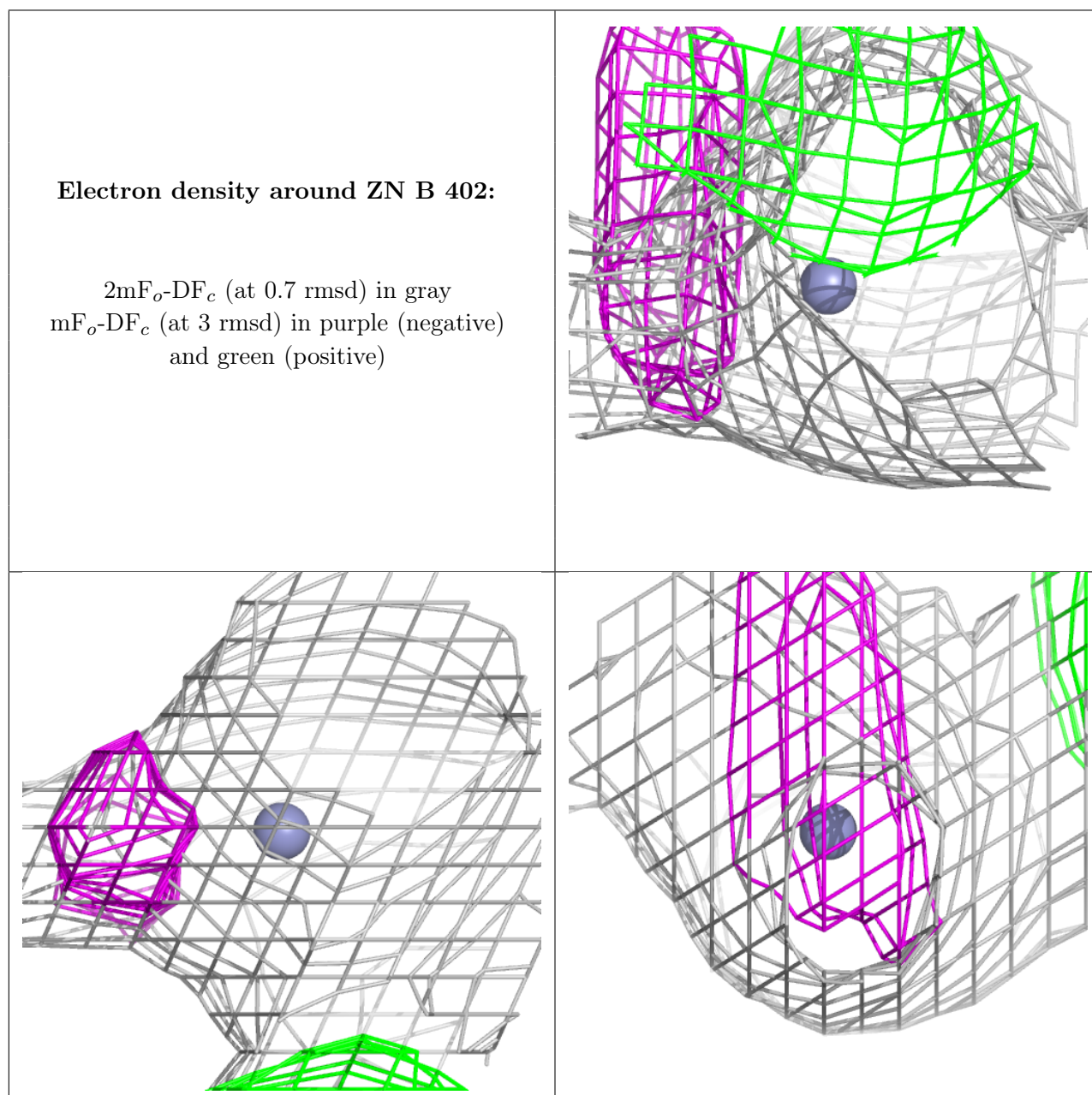There are no non-standard protein/DNA/RNA residues in this entry.

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, 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
4	ZN	B	402	1/1	0.68	0.11	82,82,82,82	1
4	ZN	B	401	1/1	0.75	0.11	124,124,124,124	0
4	ZN	A	402	1/1	0.76	0.14	96,96,96,96	1
4	ZN	A	401	1/1	0.86	0.06	129,129,129,129	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.
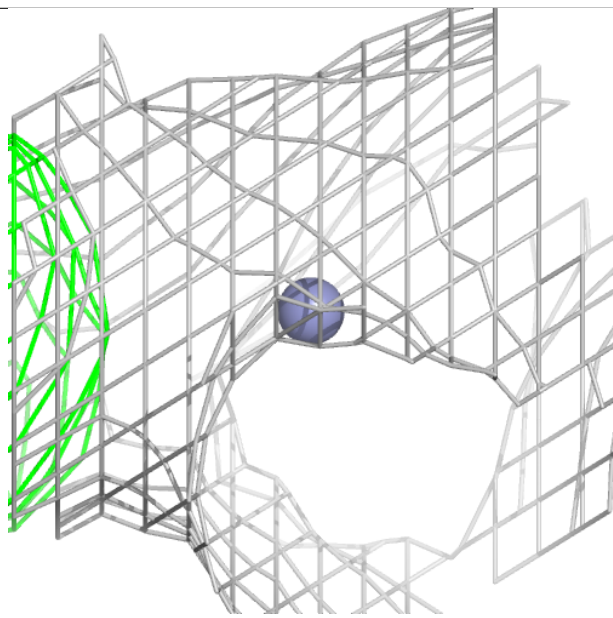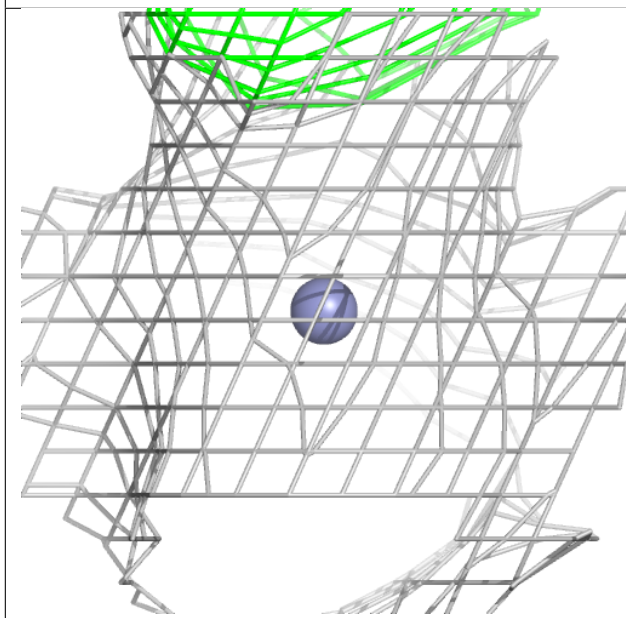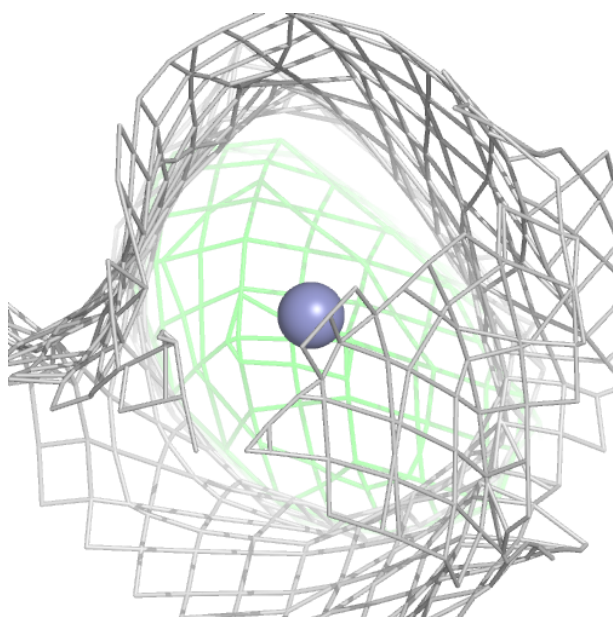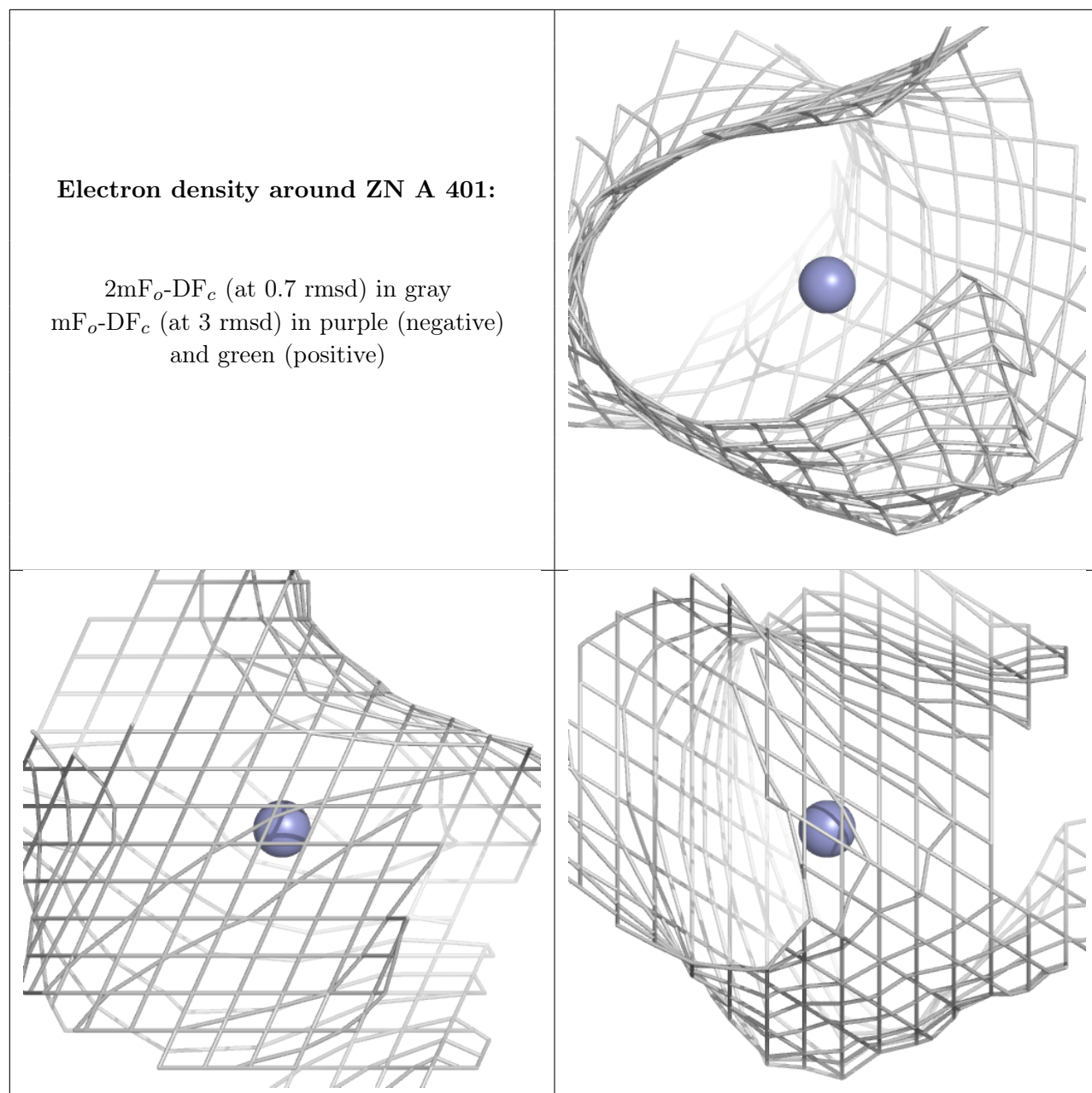

Electron density around ZN B 402:

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

Electron density around ZN 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)
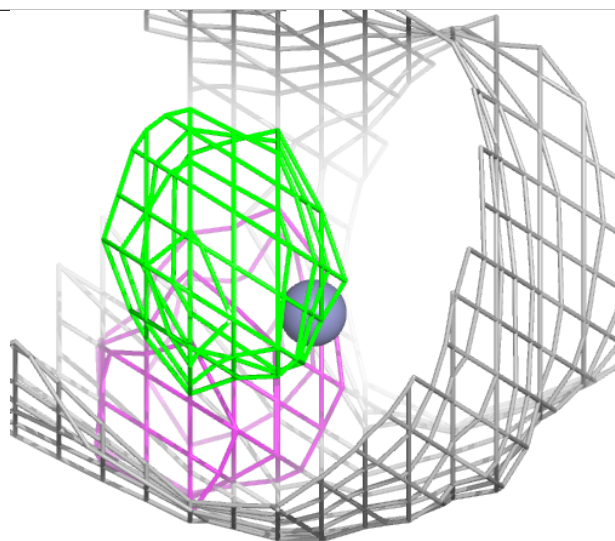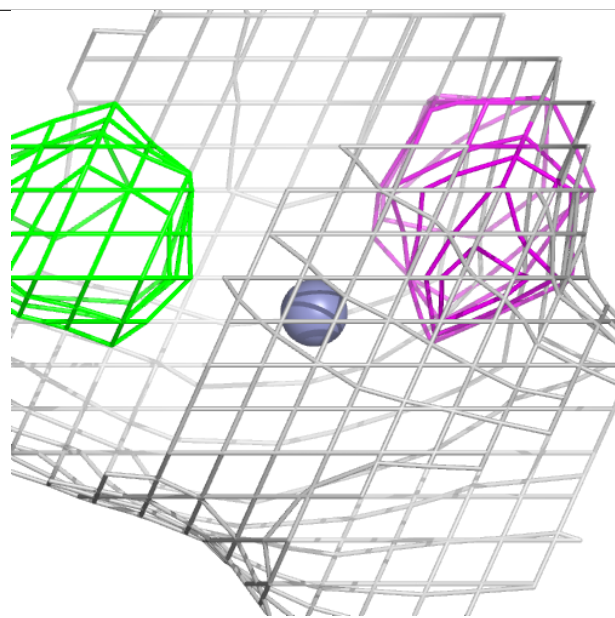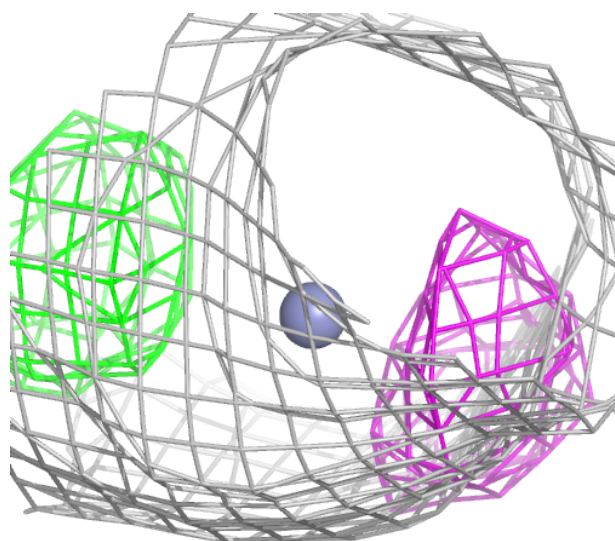
Electron density around ZN A 402:

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

Electron density around ZN 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)

6.5 Other polymers ⓘ

There are no such residues in this entry.