



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 10, 2022 – 12:24 PM JST

PDB ID : 7ESI  
Title : Crystal structure of the collagenase unit of a Vibrio collagenase from Vibrio harveyi VHJR7 at 1.8 angstrom resolution.  
Authors : Cao, H.Y.; Wang, Y.; Peng, M.; Zhang, Y.Z.  
Deposited on : 2021-05-11  
Resolution : 1.80 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](mailto: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 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 : 2.27  
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) : 2.27

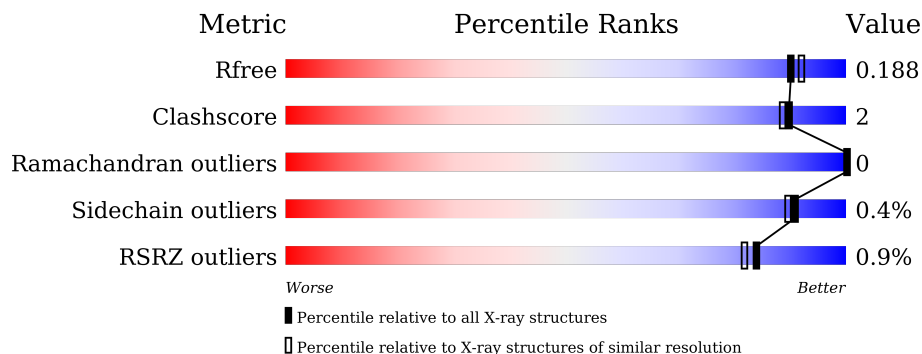
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	613	 85% 13%
2	B	14	 14% 71% 21% 7%
3	C	12	 8% 100%

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 5243 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 Collagenase unit (CU).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	532	4230	2682	700	832	16	0	0	0

- Molecule 2 is a protein called Peptide P1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	14	117	72	21	24	0	0	0

- Molecule 3 is a protein called Peptide P2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	12	96	63	15	18	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
4	A	1	Total	Zn	0	0
			1	1		

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Ca	0	0
			1	1		


- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	745	Total 745	O 745	0	0
6	B	23	Total 23	O 23	0	0
6	C	30	Total 30	O 30	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: Collagenase unit (CU)

Chain A:  85% 13%

MET  
GLU  
LEU  
LEU  
LYS  
ASN  
LEU  
THR  
VAL  
ALA  
ILE  
ALA  
THR  
PHE  
PHE  
SER  
SER  
THR  
ASN  
ALA  
LEU  
ALA  
LEU  
SER  
GLU  
PRO  
SER  
GLN  
GLN  
VAL  
THR  
GLU  
ILE  
TYR  
GLN  
HIS  
HIS  
ALA  
HIS  
GLN  
ASN  
GLY  
ASN  
ASP  
ARG  
ALA  
LEU  
PRO  
ASP  
TYR  
ALA  
PRO  
THR  
LYS  
LEU  
LEU  
PRO  
GLN  
GLN  
PRO  
LYS

PRO  
THR  
LEU  
ARG  
THR  
LEU  
SER  
THR  
ARG  
SER  
ALA  
GLU  
GLN  
ALA  
ALA  
ASN  
VAL  
C78  
T84  
I113  
R293  
S311  
E315  
Y323  
V359  
P452  
F458  
V459  
A460  
Y465  
F543  
D544  
V545  
Y582  
H609  
HIS  
HIS  
HIS  
HIS

- Molecule 2: Peptide P1

Chain B:  14% 71% 21% 7%

E1  
P2  
I9  
Y10  
Q11  
H12  
H13  
A14

- Molecule 3: Peptide P2

Chain C:  8% 100%

D1  
P12

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.44Å 50.07Å 90.68Å 90.00° 104.72° 90.00°	Depositor
Resolution (Å)	49.79 – 1.80 49.79 – 1.80	Depositor EDS
% Data completeness (in resolution range)	98.9 (49.79-1.80) 98.9 (49.79-1.80)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.62 (at 1.81Å)	Xtrriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.159 , 0.187 0.161 , 0.188	Depositor DCC
$R_{free}$ test set	2963 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.8	Xtrriage
Anisotropy	0.165	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 47.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	5243	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.44% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, 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	0.36	0/4342	0.50	0/5903
2	B	0.91	0/120	0.67	0/163
3	C	0.31	0/99	0.52	0/136
All	All	0.39	0/4561	0.50	0/6202

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 [i](#)

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	4230	0	3916	9	0
2	B	117	0	105	10	0
3	C	96	0	99	0	0
4	A	1	0	0	0	0
5	A	1	0	0	0	0
6	A	745	0	0	1	0
6	B	23	0	0	0	0
6	C	30	0	0	0	0
All	All	5243	0	4120	15	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 (15) 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)
2:B:1:GLU:N	2:B:2:PRO:HD2	1.90	0.86
2:B:1:GLU:H3	2:B:2:PRO:HD2	1.53	0.74
2:B:1:GLU:N	2:B:2:PRO:CD	2.54	0.70
2:B:1:GLU:H2	2:B:2:PRO:HD2	1.63	0.63
1:A:311:SER:O	1:A:315:GLU:HG3	2.05	0.57
2:B:1:GLU:H3	2:B:2:PRO:CD	2.17	0.56
1:A:293:ARG:HD3	2:B:1:GLU:HG2	1.88	0.55
1:A:465:TYR:HD2	2:B:10:TYR:HH	1.57	0.52
2:B:1:GLU:O	2:B:1:GLU:HG3	2.10	0.50
1:A:84:THR:HG21	1:A:113:ILE:HD13	1.93	0.49
1:A:545:VAL:HG21	2:B:12:HIS:NE2	2.31	0.45
1:A:323:TYR:CD1	1:A:359:VAL:HG22	2.52	0.44
1:A:293:ARG:CD	2:B:1:GLU:HG2	2.48	0.44
1:A:458:PHE:CZ	1:A:460:ALA:HB2	2.52	0.44
1:A:315:GLU:OE1	6:A:801:HOH:O	2.22	0.41

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	530/613 (86%)	518 (98%)	12 (2%)	0	100	100
2	B	12/14 (86%)	12 (100%)	0	0	100	100
3	C	10/12 (83%)	9 (90%)	1 (10%)	0	100	100
All	All	552/639 (86%)	539 (98%)	13 (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	440/509 (86%)	439 (100%)	1 (0%)	93	92
2	B	13/13 (100%)	12 (92%)	1 (8%)	13	4
3	C	11/11 (100%)	11 (100%)	0	100	100
All	All	464/533 (87%)	462 (100%)	2 (0%)	91	89

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	582	TYR
2	B	1	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	354	GLN

### 5.3.3 RNA [i](#)

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 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q < 0.9
1	A	532/613 (86%)	-0.50	2 (0%) 92 90	12, 19, 29, 39	0
2	B	14/14 (100%)	0.92	2 (14%) 2 1	21, 29, 40, 42	0
3	C	12/12 (100%)	-0.15	1 (8%) 11 8	18, 23, 38, 43	0
All	All	558/639 (87%)	-0.46	5 (0%) 84 82	12, 19, 30, 43	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	9	ILE	4.9
1	A	543	PHE	3.6
1	A	452	PRO	3.1
2	B	10	TYR	3.1
3	C	1	ASP	2.6

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.3 Carbohydrates [i](#)

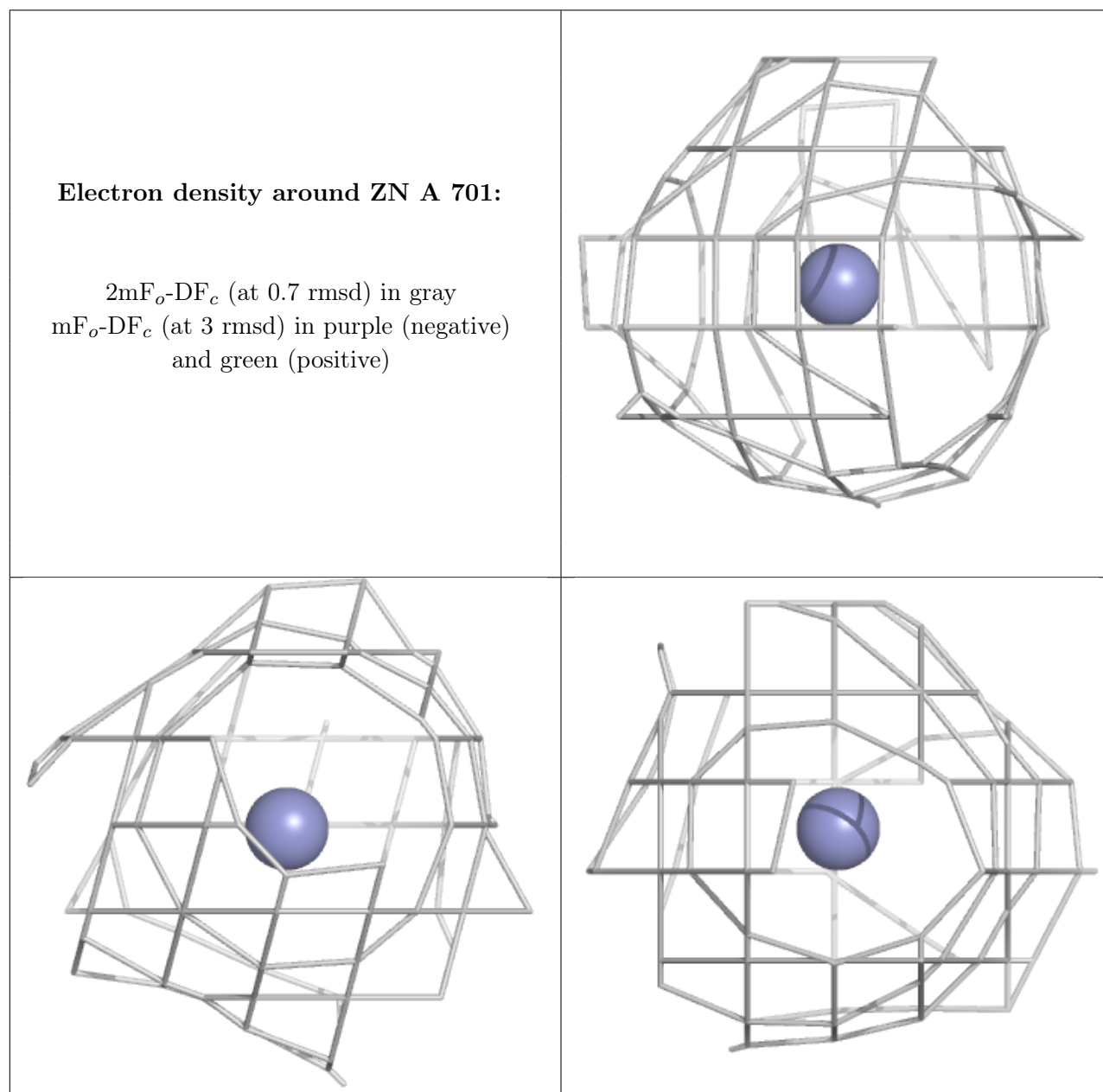
There are no monosaccharides in this entry.

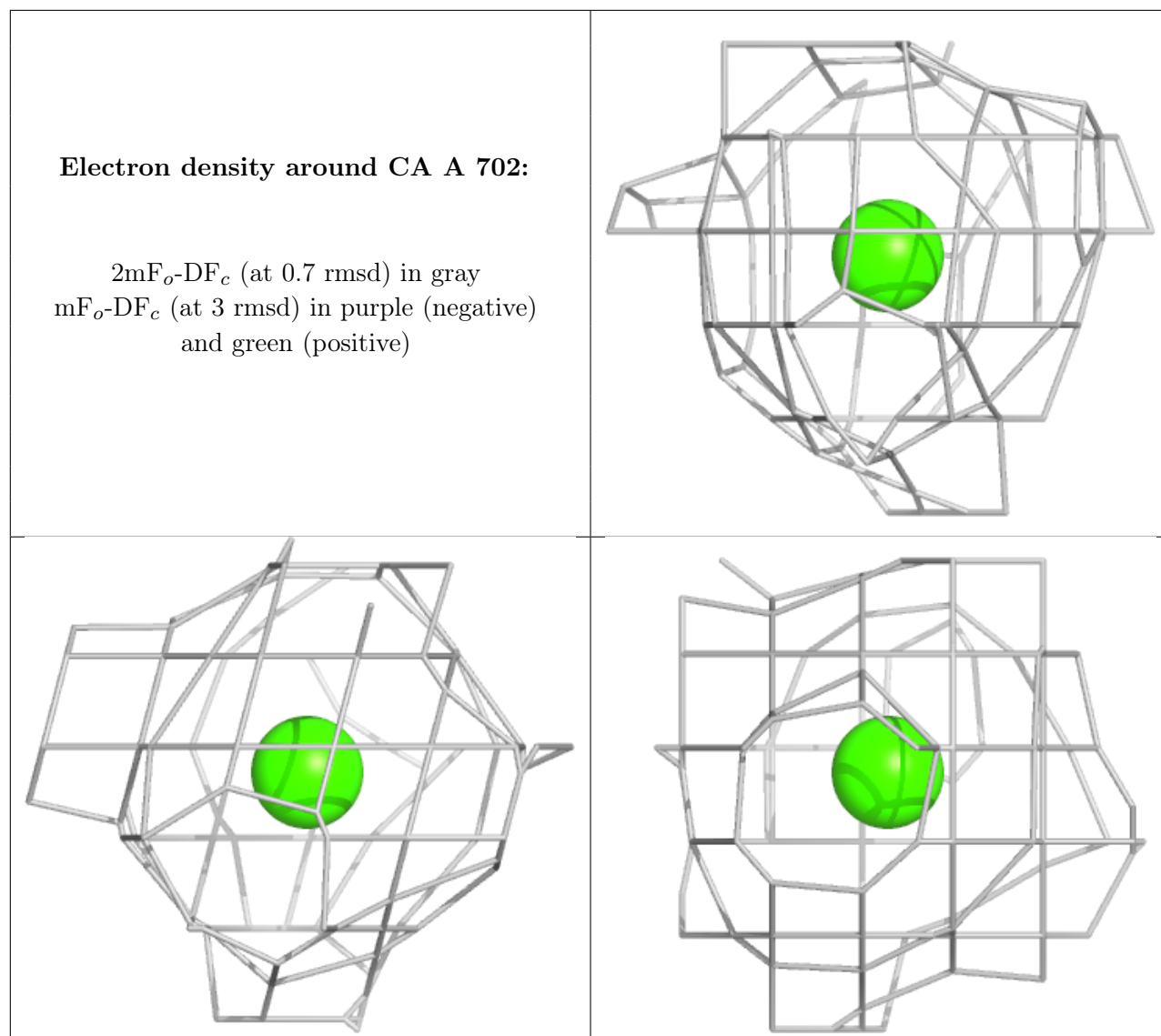
### 6.4 Ligands [i](#)

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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
4	ZN	A	701	1/1	1.00	0.06	14,14,14,14	0
5	CA	A	702	1/1	1.00	0.07	17,17,17,17	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.





## 6.5 Other polymers [i](#)

There are no such residues in this entry.