

# wwPDB X-ray Structure Validation Summary Report (i)

#### Jul 3, 2024 – 11:00 AM EDT

PDB ID	:	8VMS
Title	:	Homing endonuclease I-PpoI-DNA complex: reaction at pH7.0 ( $K+MES$ ) with
		$500 \mathrm{~uM~Mg2+~for} 80 \mathrm{s}$
Authors	:	Chang, C.; Gao, Y.
Deposited on	:	2024-01-13
Resolution	:	1.42  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:

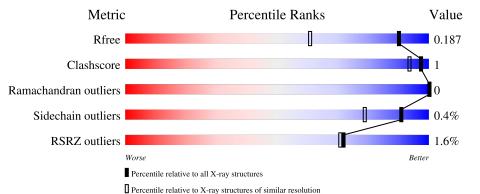
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 1.42 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
$R_{free}$	130704	2579 (1.44-1.40)
Clashscore	141614	2696 (1.44-1.40)
Ramachandran outliers	138981	2632(1.44-1.40)
Sidechain outliers	138945	2631 (1.44-1.40)
RSRZ outliers	127900	2528 (1.44-1.40)

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	С	21	90%	10%
1	D	21	81%	19%
2	А	162	100%	
2	В	162	4% 98%	•



#### 8VMS

# 2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 3876 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 DNA chain called DNA (5'-D(\*TP\*TP\*GP\*AP\*CP\*TP\*CP\*TP\*CP\*TP\* TP\*AP\*AP\*GP\*AP\*GP\*AP\*GP\*TP\*CP\*A)-3').

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	1 C	91	Total	С	Ν	0	Р	0	2	0
		21	471	226	86	137	22	0		
1	Л	21	Total	С	Ν	0	Р	0	0 2	0
	I D	21	471	226	86	137	22	0	Δ	0

• Molecule 2 is a protein called Intron-encoded endonuclease I-PpoI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
0	Λ	162	Total	С	Ν	0	S	0	0	0
		102	1245	786	232	219	8	0		
0	В	162	Total	С	Ν	0	S	0	1	0
	2 B	102	1249	789	232	220	8	0		0

• Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

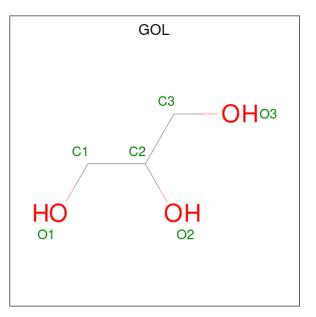
N	Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
	3	С	1	Total Mg 1 1	0	0
	3	D	1	Total Mg 1 1	0	0

• Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	С	1	Total Na 1 1	0	0
4	D	1	Total Na 1 1	0	0



• Molecule 5 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 6 & 3 & 3 \end{array}$	0	0
5	D	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 6 & 3 & 3 \end{array}$	0	0
5	А	1	$\begin{array}{ccc} \text{Total}  \text{C}  \text{O} \\ 6  3  3 \end{array}$	0	0
5	А	1	$\begin{array}{ccc} \text{Total}  \text{C}  \text{O} \\ 6  3  3 \end{array}$	0	0
5	В	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 6 & 3 & 3 \end{array}$	0	0

• Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	А	2	Total Zn 2 2	0	0
6	В	2	Total Zn 2 2	0	0

• Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	С	44	Total O 44 44	0	0
7	D	57	$\begin{array}{cc} \text{Total} & \text{O} \\ 57 & 57 \end{array}$	0	0

Continued on next page...



Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	А	150	Total O 150 150	0	1
7	В	151	Total O 151 151	0	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: DNA (5'-D(\*TP\*TP\*GP\*AP\*CP\*TP\*CP\*TP\*CP\*TP\*TP\*AP\*AP\*GP\*AP\*GP \*AP\*GP\*TP\*CP\*A)-3')

Chain C:	90%	10%
T 401 6413 6414 A421		
• Molecule 1: DNA (5'-D(*TP*TP* *AP*GP*TP*CP*A)-3')	GP*AP*CP*TP*CP*TP*CP	*TP*TP*AP*AP*GP*AP*GP
Chain D: 81	%	19%
1501 1507 1508 1513 1513 1513 1513 1513		
• Molecule 2: Intron-encoded endor	uclease I-PpoI	
Chain A:	100%	
There are no outlier residues recorde	ed for this chain.	
• Molecule 2: Intron-encoded endom	uclease I-PpoI	
Chain B:	98%	





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants	113.58Å 113.58Å 87.99Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $120.00^{\circ}$	Depositor
Resolution (Å)	42.93 - 1.42	Depositor
Resolution (A)	42.93 - 1.42	EDS
% Data completeness	99.6 (42.93-1.42)	Depositor
(in resolution range)	99.6(42.93-1.42)	EDS
R <sub>merge</sub>	0.09	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	1.32 (at 1.42 Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
P. P.	0.179 , $0.189$	Depositor
$R, R_{free}$	0.178 , $0.187$	DCC
$R_{free}$ test set	6180 reflections $(5.01%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	17.5	Xtriage
Anisotropy	0.173	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.35 , $36.4$	EDS
L-test for twinning <sup>2</sup>	$<  L  > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.023 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3876	wwPDB-VP
Average B, all atoms $(Å^2)$	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.36% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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: NA, GOL, MG, 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
	Unam	RMSZ   #  Z  > 5		RMSZ	# Z  > 5
1	С	0.84	0/529	1.00	0/812
1	D	0.83	0/529	0.99	0/812
2	А	0.34	0/1286	0.61	0/1760
2	В	0.40	0/1293	0.63	0/1770
All	All	0.55	0/3637	0.76	0/5154

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	С	471	0	262	1	0
1	D	471	0	262	3	0
2	А	1245	0	1191	0	0
2	В	1249	0	1198	1	0
3	С	1	0	0	0	0
3	D	1	0	0	0	0
4	С	1	0	0	0	0
4	D	1	0	0	0	0
5	А	12	0	16	0	0

Continued on next page...



	Mol     Chain     Non-H     H(model)     H(added)     Clashes     Symm-Clash								
Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes			
5	В	6	0	8	0	0			
5	D	12	0	16	1	0			
6	А	2	0	0	0	0			
6	В	2	0	0	0	0			
7	А	150	0	0	0	0			
7	В	151	0	0	0	0			
7	С	44	0	0	0	0			
7	D	57	0	0	0	0			
All	All	3876	0	2953	5	0			

Continued from previous page...

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 (5) 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:260:THR:HB	2:B:303[A]:THR:HG21	1.92	0.51
1:D:513[B]:DA:O3'	1:D:514[B]:DG:P	2.72	0.47
1:D:507:DC:H2'	1:D:508:DT:C6	2.51	0.46
1:D:514[B]:DG:H21	5:D:603:GOL:HO3	1.61	0.42
1:C:413[B]:DA:O3'	1:C:414[B]:DG:P	2.77	0.42

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	Perce	ntiles
2	А	160/162~(99%)	160 (100%)	0	0	100	100
2	В	161/162~(99%)	160 (99%)	1 (1%)	0	100	100
All	All	321/324~(99%)	320 (100%)	1 (0%)	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	Analysed Rotameric Outliers		Percentiles		
2	А	133/133~(100%)	133 (100%)	0	100 100		
2	В	134/133~(101%)	133~(99%)	1 (1%)	84 65		
All	All	267/266~(100%)	266 (100%)	1 (0%)	91 79		

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	В	203	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

#### 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 (i)

Of 13 ligands modelled in this entry, 8 are monoatomic - leaving 5 for Mogul analysis.



In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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| > 2 is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Tune	Chain	Dea Link		В	ond leng	gths	В	Bond ang	gles
	Type	Unam	Res	Link	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2
5	GOL	А	603	-	$5,\!5,\!5$	0.11	0	$5,\!5,\!5$	0.46	0
5	GOL	В	403	-	$5,\!5,\!5$	0.08	0	$5,\!5,\!5$	0.29	0
5	GOL	D	603	-	$5,\!5,\!5$	0.09	0	$5,\!5,\!5$	0.22	0
5	GOL	D	604	-	$5,\!5,\!5$	0.08	0	$5,\!5,\!5$	0.22	0
5	GOL	А	604	-	$5,\!5,\!5$	0.10	0	$5,\!5,\!5$	0.39	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	GOL	А	603	-	-	4/4/4/4	-
5	GOL	В	403	-	-	0/4/4/4	-
5	GOL	D	603	-	-	0/4/4/4	-
5	GOL	D	604	-	-	0/4/4/4	-
5	GOL	А	604	-	-	1/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	А	603	GOL	O1-C1-C2-C3
5	А	604	GOL	C1-C2-C3-O3
5	А	603	GOL	O1-C1-C2-O2
5	А	603	GOL	O2-C2-C3-O3
5	А	603	GOL	C1-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 1 short contact:



Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	D	603	GOL	1	0

### 5.7 Other polymers (i)

There are no such residues in this entry.

### 5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	С	1
1	D	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	С	413[B]:DA	O3'	414[B]:DG	Р	2.77
1	D	513[B]:DA	O3'	514[B]:DG	Р	2.72



# 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,  $95^{th}$  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	$\mathbf{OWAB}(\mathrm{\AA}^2)$	Q < 0.9
1	С	21/21~(100%)	-0.14	0 100 100	16, 23, 28, 30	0
1	D	21/21~(100%)	-0.19	0 100 100	16, 22, 29, 31	0
2	А	162/162~(100%)	-0.00	0 100 100	14, 18, 28, 36	0
2	В	162/162~(100%)	0.08	6 (3%) 41 41	14, 18, 36, 48	0
All	All	366/366~(100%)	0.02	6 (1%) 72 71	14, 19, 32, 48	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	В	203	LEU	5.0
2	В	202	ALA	4.2
2	В	290	ASN	2.7
2	В	205	ASN	2.3
2	В	269	ASN	2.2

## 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)

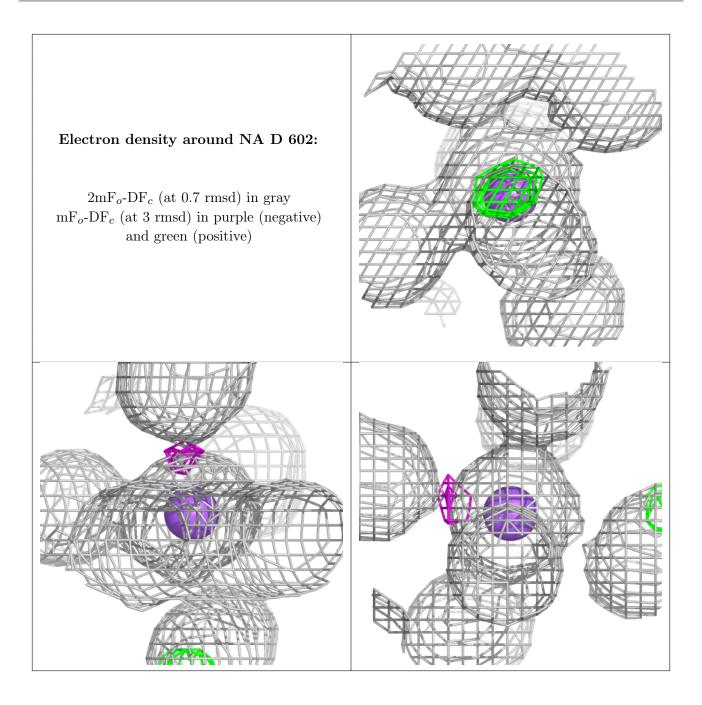
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median,  $95^{th}$  percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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



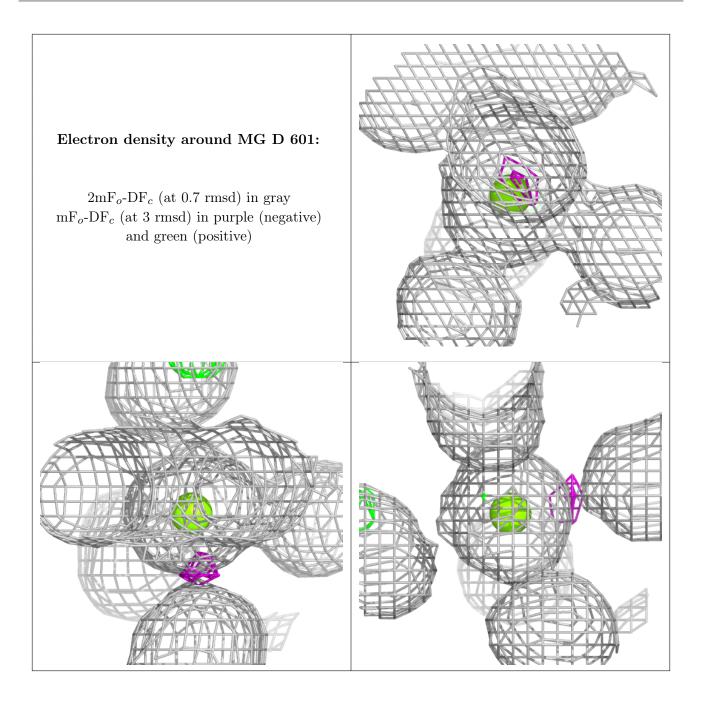
Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\operatorname{B-factors}(\operatorname{\AA}^2)$	Q < 0.9
5	GOL	А	604	6/6	0.84	0.27	33,34,38,45	0
5	GOL	D	604	6/6	0.85	0.25	23,33,35,38	0
5	GOL	D	603	6/6	0.87	0.18	23,29,35,36	0
5	GOL	А	603	6/6	0.93	0.14	$19,\!28,\!35,\!35$	0
5	GOL	В	403	6/6	0.95	0.11	$16,\!18,\!19,\!23$	0
4	NA	D	602	1/1	0.99	0.08	$15,\!15,\!15,\!15$	1
3	MG	D	601	1/1	0.99	0.08	$15,\!15,\!15,\!15$	1
4	NA	С	502	1/1	0.99	0.07	$15,\!15,\!15,\!15$	1
6	ZN	А	602	1/1	0.99	0.08	14,14,14,14	0
6	ZN	А	601	1/1	1.00	0.06	16,16,16,16	0
3	MG	С	501	1/1	1.00	0.07	$15,\!15,\!15,\!15$	1
6	ZN	В	401	1/1	1.00	0.07	$15,\!15,\!15,\!15$	0
6	ZN	В	402	1/1	1.00	0.08	14,14,14,14	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.

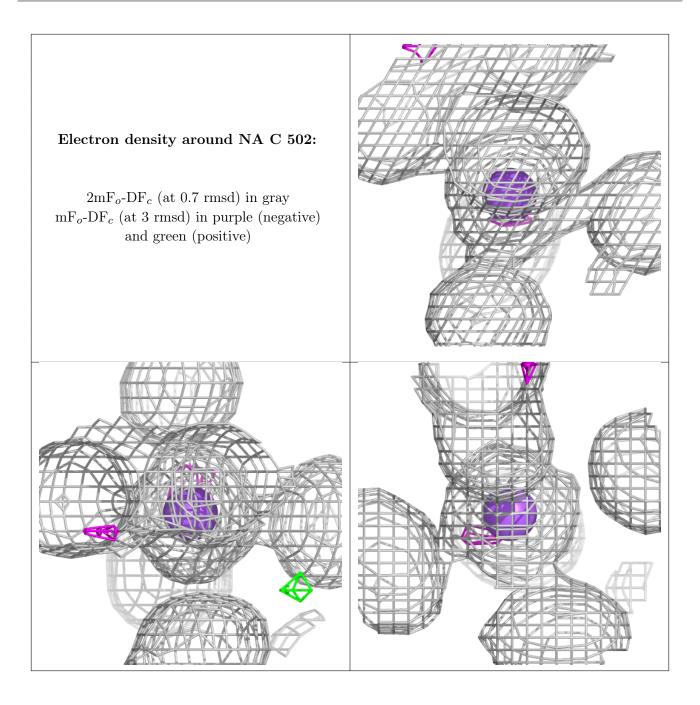




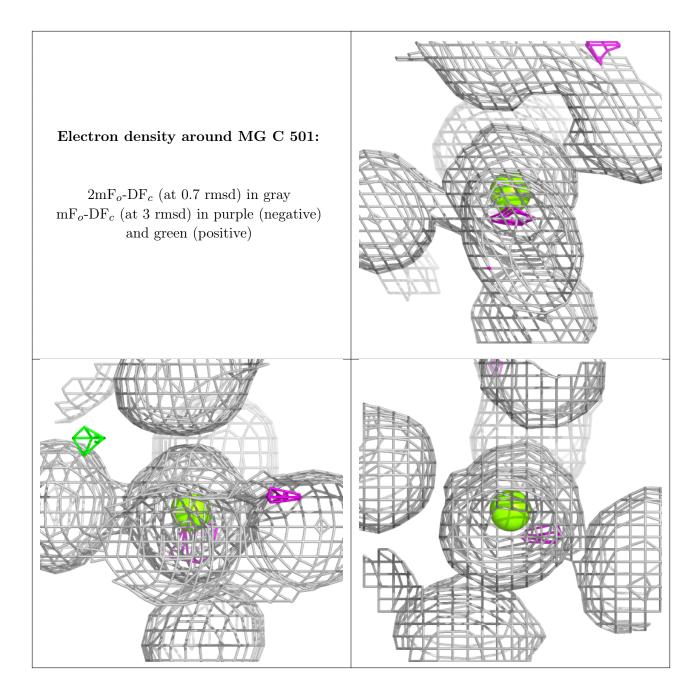












## 6.5 Other polymers (i)

There are no such residues in this entry.

