

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

Nov 21, 2023 – 04:48 AM JST

PDB ID : 7EOJ

Title : Crystal structure of the Pepper aptamer in complex with HBC, cesium soak

Authors: Huang, K.Y.; Ren, A.M.

Deposited on : 2021-04-22

Resolution : 1.77 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

EDS: 2.36

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)

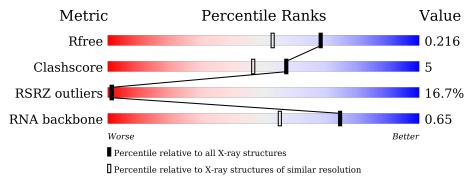
Validation Pipeline (wwPDB-VP) : 2.36

## 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 1.77 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	Similar resolution
Metric	$(\# \mathrm{Entries})$	$(\#  ext{Entries},  ext{ resolution range}( ext{Å}))$
$R_{free}$	130704	9185 (1.80-1.76)
Clashscore	141614	10184 (1.80-1.76)
RSRZ outliers	127900	9032 (1.80-1.76)
RNA backbone	3102	1069 (2.40-1.16)

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		
			16%		
1	A	49	73%	18%	6% •



# 2 Entry composition (i)

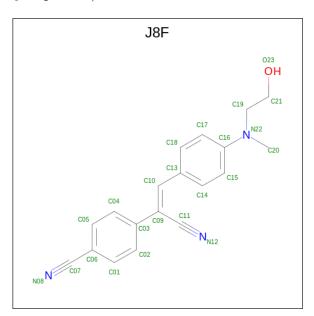
There are 5 unique types of molecules in this entry. The entry contains 1239 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 RNA chain called Pepper (49-MER).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace			
1	A	49	Total 1050	C 464	N 185	O 351	P 50	0	0	0

• Molecule 2 is 4-[( $\{Z\}$ )-1-cyano-2-[4-[2-hydroxyethyl(methyl)amino]phenyl]ethenyl]benzene carbonitrile (three-letter code: J8F) (formula:  $C_{19}H_{17}N_3O$ ) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	А	1	Total	С	N	О	0	0
	11	1	23	19	3	1		

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	9	Total Mg 9 9	0	0



• Molecule 4 is CESIUM ION (three-letter code: CS) (formula: Cs) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	8	Total Cs 8 8	0	0

• Molecule 5 is water.

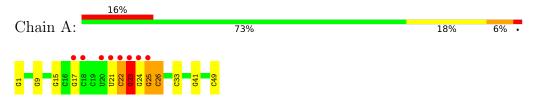
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	149	Total O 149 149	0	0



# 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: Pepper (49-MER)





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants	92.28Å 34.75Å 58.12Å	Donositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $127.93^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	45.85 - 1.77	Depositor
Resolution (A)	45.85 - 1.77	EDS
% Data completeness	96.6 (45.85-1.77)	Depositor
(in resolution range)	92.6 (45.85-1.77)	EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.35  (at  1.77Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
$R, R_{free}$	0.207 , $0.216$	Depositor
it, it <sub>free</sub>	0.207 , $0.216$	DCC
$R_{free}$ test set	735 reflections $(5.29%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.2	Xtriage
Anisotropy	0.597	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.25, 40.8	EDS
L-test for twinning <sup>2</sup>	$< L > = 0.49, < L^2> = 0.33$	Xtriage
Estimated twinning fraction	0.031 for -h-2*l,-k,l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	1239	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 10.49% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



<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: J8F, GDP, CS, MG

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).

Mal	Chain	Bond lengths		Bond angles	
IVIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5
1	A	0.47	0/1139	0.99	2/1774~(0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^{o})$
1	A	9	G	O4'-C1'-N9	5.32	112.46	108.20
1	A	23	G	P-O3'-C3'	5.19	125.93	119.70

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	1050	0	531	3	0
2	A	23	0	0	0	0
3	A	9	0	0	1	0
4	A	8	0	0	0	0
5	A	149	0	0	2	1
All	All	1239	0	531	4	1

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.



All (4) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	$egin{aligned}  ext{Clash} \  ext{overlap } ( ext{Å}) \end{aligned}$
3:A:106:MG:MG	5:A:250:HOH:O	1.13	0.89
1:A:22:C:H4'	1:A:23:G:H5'	1.60	0.83
1:A:49:C:O2'	5:A:201:HOH:O	1.99	0.80
1:A:25:G:O2'	1:A:26:C:OP1	2.30	0.49

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	$egin{aligned} &  ext{Interatomic} \ &  ext{distance} \ &  ext{(Å)} \end{aligned}$	$egin{aligned}  ext{Clash} \  ext{overlap } ( ext{Å}) \end{aligned}$
5:A:311:HOH:O	5:A:320:HOH:O[4_555]	2.10	0.10

## 5.3 Torsion angles (i)

#### 5.3.1 Protein backbone (i)

There are no protein molecules in this entry.

#### 5.3.2 Protein sidechains (i)

There are no protein molecules in this entry.

#### 5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	47/49 (95%)	9 (19%)	3 (6%)

5 of 9 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	15	G
1	A	17	G
1	A	21	U
1	A	22	С
1	A	23	G

All (3) RNA pucker outliers are listed below:



Mol	Chain	Res	Type
1	A	23	G
1	A	25	G
1	A	41	G

### 5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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	Mol Type Chain Res Lin			Link	Bo	ond leng	$ ag{ths}$	Bond angles		
MIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
1	GDP	A	1	1	24,30,30	0.99	1 (4%)	30,47,47	0.97	2 (6%)

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
1	GDP	A	1	1	-	4/12/32/32	0/3/3/3

#### All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	Observed(A)	Ideal(A)
1	A	1	GDP	C6-N1	-2.11	1.34	1.37

#### All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
1	A	1	GDP	C8-N7-C5	2.38	107.52	102.99
1	A	1	GDP	C5-C6-N1	2.19	117.83	113.95

There are no chirality outliers.

All (4) torsion outliers are listed below:



Mol	Chain	Res	Type	Atoms
1	A	1	GDP	C5'-O5'-PA-O2A
1	A	1	GDP	C5'-O5'-PA-O3A
1	A	1	GDP	C5'-O5'-PA-O1A
1	A	1	GDP	PA-O3A-PB-O2B

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry (i)

Of 18 ligands modelled in this entry, 17 are monoatomic - leaving 1 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).

_	Aol	Type	Chain	Res	Link	Bo	Bond lengths			Bond angles		
1	101	туре	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z >2	
	2	J8F	A	101	-	24,24,24	1.71	4 (16%)	31,31,31	1.04	2 (6%)	

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	$\mathbf{Type}$	Chain	$\operatorname{Res}$	Link	Chirals	Torsions	Rings
2	J8F	A	101	-	-	1/19/19/19	0/2/2/2

All (4) bond length outliers are listed below:

$\mathbf{N}$	/Iol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\text{\AA})$	$\operatorname{Ideal}(\text{\AA})$
	2	A	101	J8F	C11-C09	5.65	1.52	1.44
	2	A	101	J8F	C06-C07	3.81	1.53	1.44

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(A)	$Ideal(\AA)$
2	A	101	J8F	C13-C10	3.54	1.53	1.46
2	A	101	J8F	C16-N22	2.36	1.45	1.39

#### All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
2	A	101	J8F	C04-C03-C09	-2.48	118.69	121.18
2	A	101	J8F	C03-C09-C10	-2.40	119.51	123.99

There are no chirality outliers.

All (1) torsion outliers are listed below:

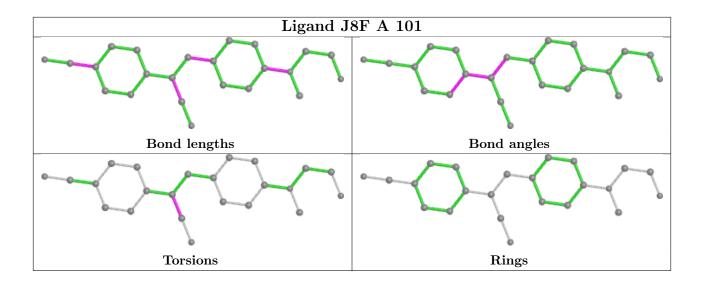
Mol	Chain	Res	Type	Atoms
2	A	101	J8F	C03-C09-C11-N12

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.





# 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 (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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(Å^2)$	Q < 0.9
1	A	48/49 (97%)	1.08	8 (16%) 1 1	15, 20, 78, 92	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	23	G	11.8
1	A	22	С	7.6
1	A	24	G	6.8
1	A	21	U	6.3
1	A	25	G	3.8

## 6.2 Non-standard residues in protein, DNA, RNA chains (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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
1	GDP	A	1	28/28	0.81	0.16	15,19,36,40	0

## 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,



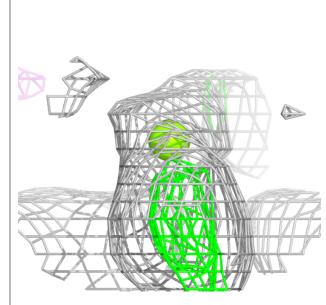
median,  $95^{th}$  percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

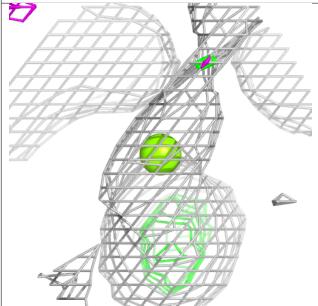
Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
3	MG	A	109	1/1	0.78	0.18	43,43,43,43	0
3	MG	A	110	1/1	0.88	0.26	33,33,33,33	0
2	J8F	A	101	23/23	0.91	0.13	14,17,20,29	0
3	MG	A	107	1/1	0.95	0.08	38,38,38,38	0
3	MG	A	108	1/1	0.95	0.07	36,36,36,36	0
3	MG	A	103	1/1	0.97	0.09	14,14,14,14	0
4	CS	A	115	1/1	0.97	0.12	66,66,66,66	0
3	MG	A	106	1/1	0.98	0.10	27,27,27,27	0
4	CS	A	111	1/1	0.98	0.06	35,35,35,35	0
3	MG	A	104	1/1	0.98	0.07	16,16,16,16	0
4	CS	A	117	1/1	0.98	0.05	31,31,31,31	0
4	CS	A	112	1/1	0.99	0.05	34,34,34,34	0
4	CS	A	114	1/1	0.99	0.10	40,40,40,40	0
3	MG	A	102	1/1	0.99	0.06	14,14,14,14	0
4	CS	A	116	1/1	0.99	0.10	50,50,50,50	0
3	MG	A	105	1/1	0.99	0.06	16,16,16,16	0
4	CS	A	118	1/1	0.99	0.09	39,39,39,39	0
4	CS	A	113	1/1	1.00	0.06	31,31,31,31	0

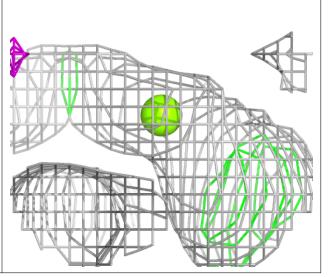
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



#### Electron density around MG A 109:

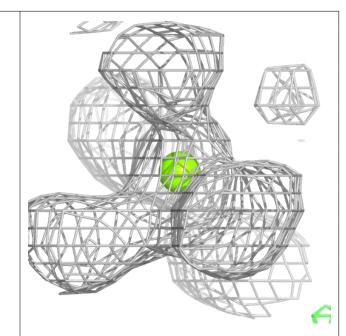


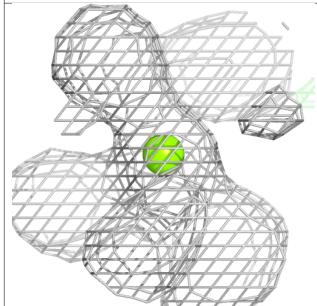


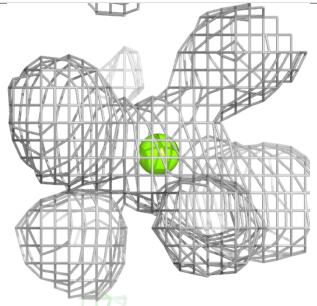




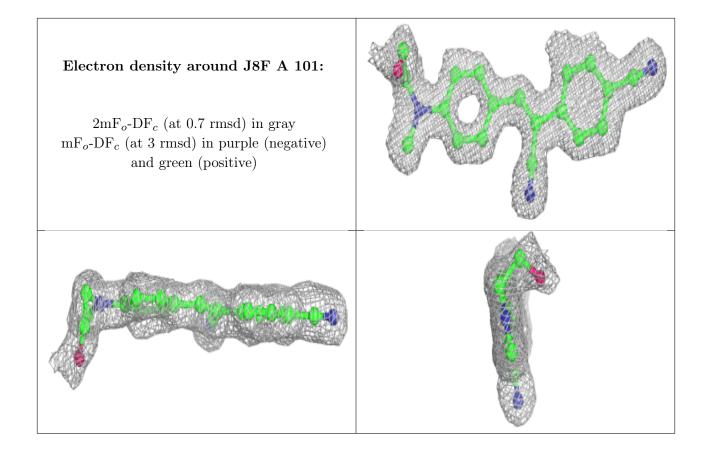
#### Electron density around MG A 110:





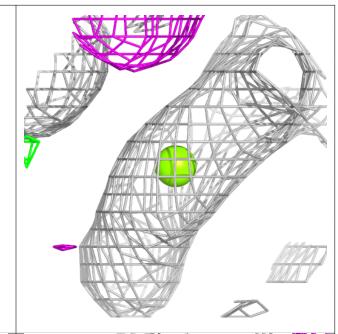


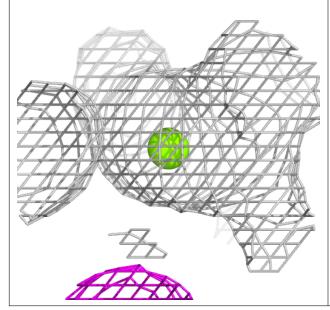






#### Electron density around MG A 107:

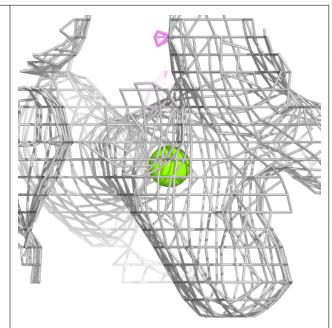


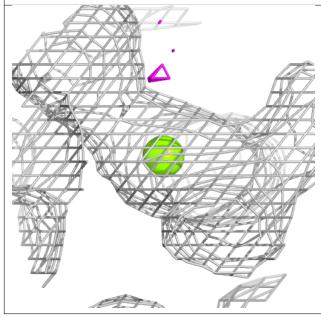


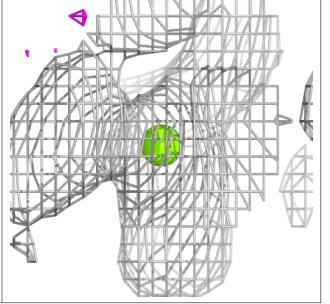




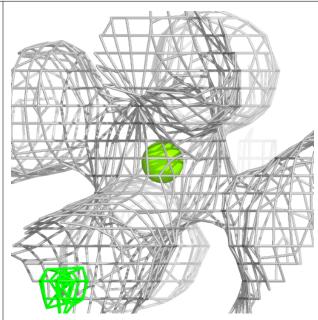
# Electron density around MG A 108: $\,$

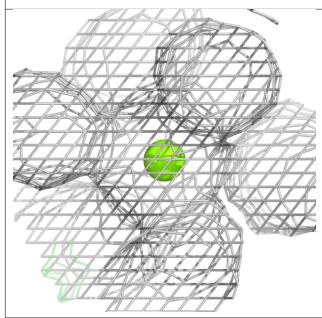


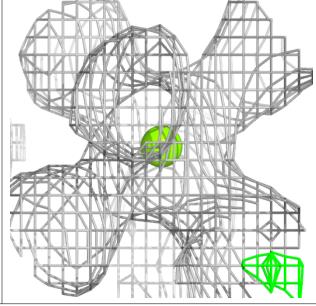


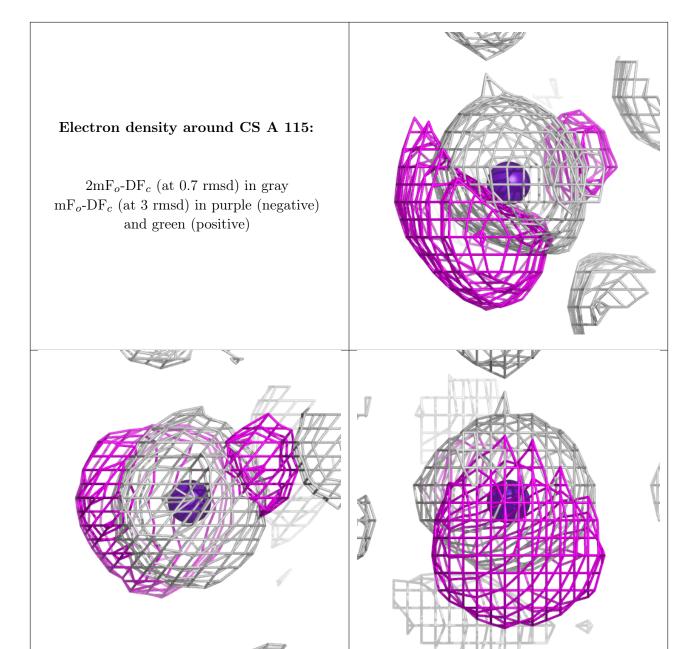


#### Electron density around MG A 103:



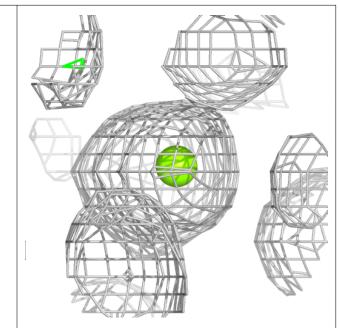


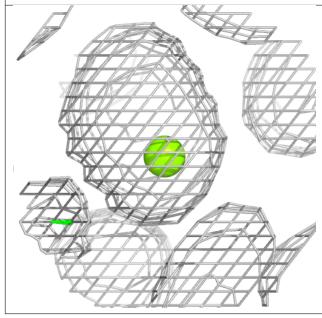


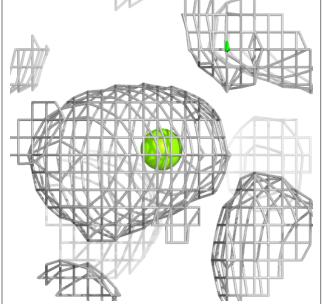




#### Electron density around MG A 106:





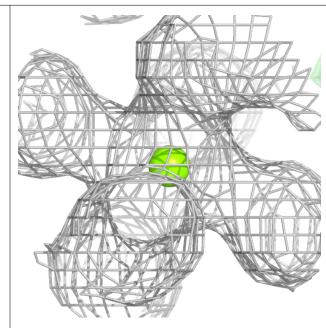


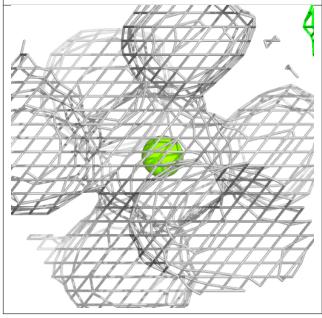


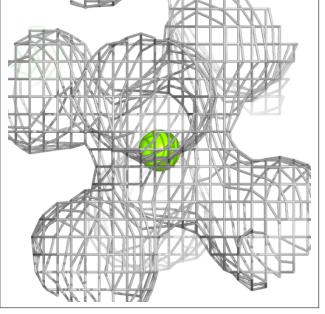
# 



#### Electron density around MG A 104:





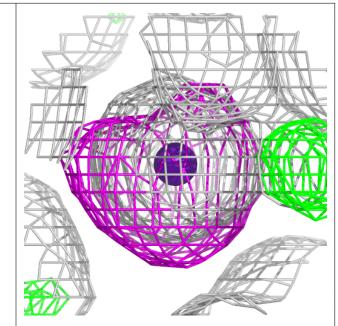


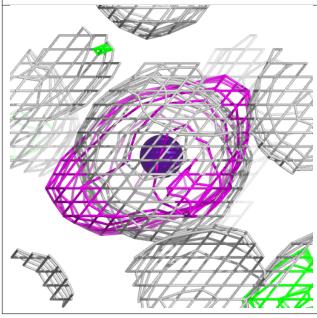
# Electron density around CS A 117: $2 \mathrm{mF}_o\text{-}\mathrm{DF}_c$ (at 0.7 rmsd) in gray ${\rm mF}_o\text{-}{\rm DF}_c$ (at 3 rmsd) in purple (negative) and green (positive)

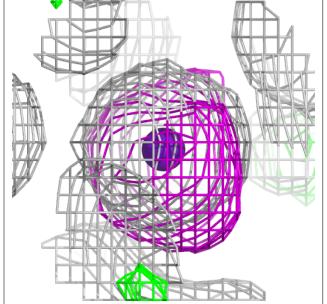


#### Electron density around CS A 112:

 $2 \text{mF}_o\text{-DF}_c$  (at 0.7 rmsd) in gray  $\text{mF}_o\text{-DF}_c$  (at 3 rmsd) in purple (negative) and green (positive)

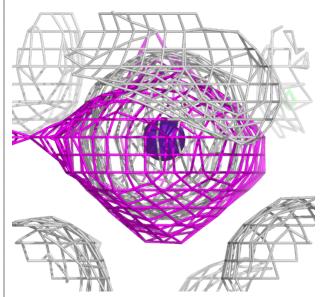


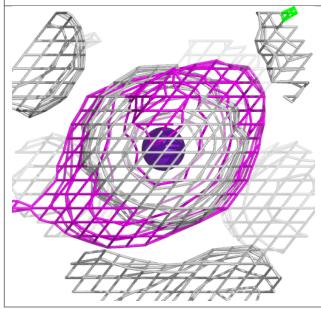


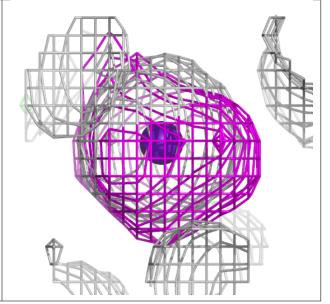




### Electron density around CS A 114:

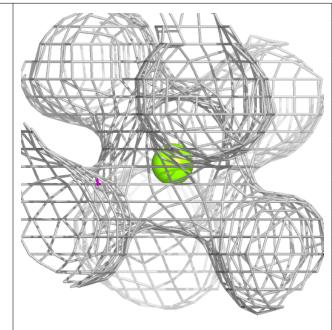


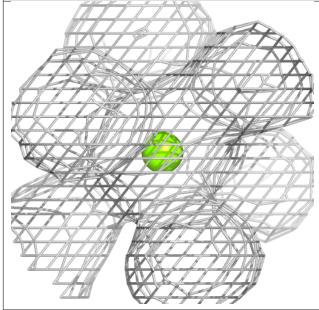


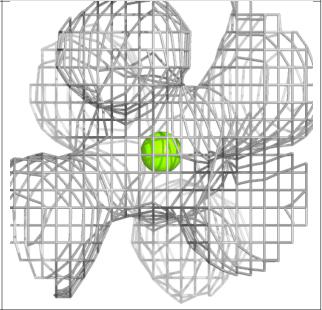




#### Electron density around MG A 102:





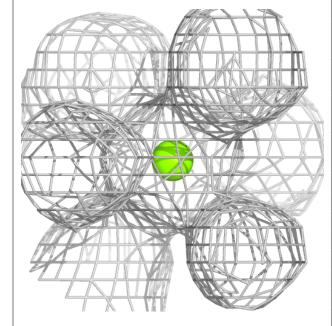


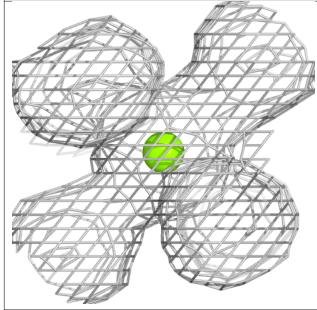


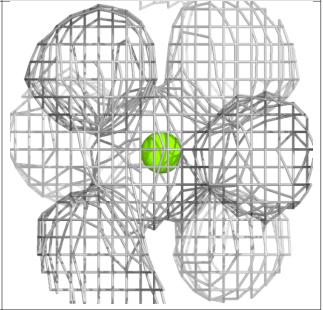
# 



#### Electron density around MG A 105:



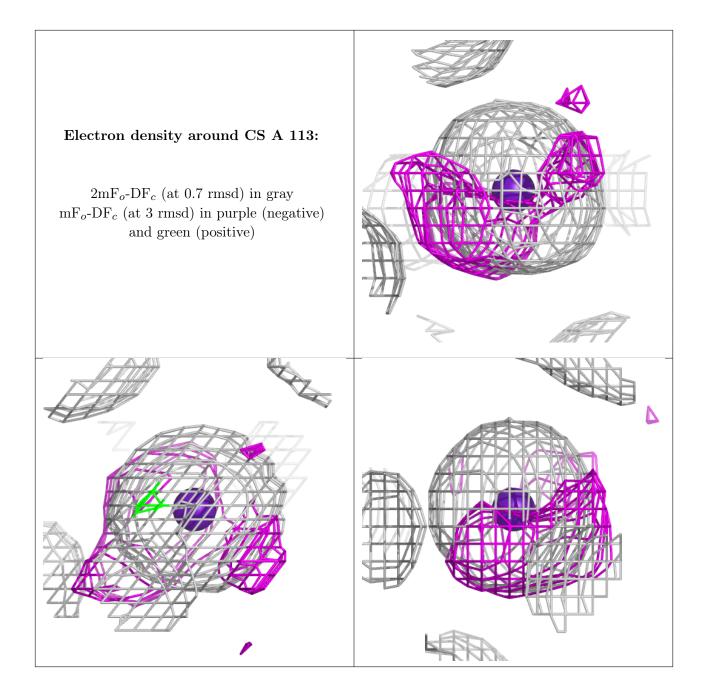






# Electron density around CS A 118: $2 \mathrm{mF}_o\text{-}\mathrm{DF}_c$ (at 0.7 rmsd) in gray $\mathrm{mF}_o\text{-}\mathrm{DF}_c$ (at 3 rmsd) in purple (negative) and green (positive)





# 6.5 Other polymers (i)

There are no such residues in this entry.

