



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 24, 2025 – 03:09 pm GMT

PDB ID : 9EZD  
Title : BsmI (Bottom Nicking mutant) crystallized with Mg<sup>2+</sup> and cognate dsDNA (Post-reactive complex)  
Authors : Sieskind, R.; Missouri, S.; Madru, C.; Commenge, I.; Niogret, G.; Rondelez, Y.; Haouz, A.; Legrand, P.; Sauguet, L.; Delarue, M.  
Deposited on : 2024-04-11  
Resolution : 2.63 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.41.5

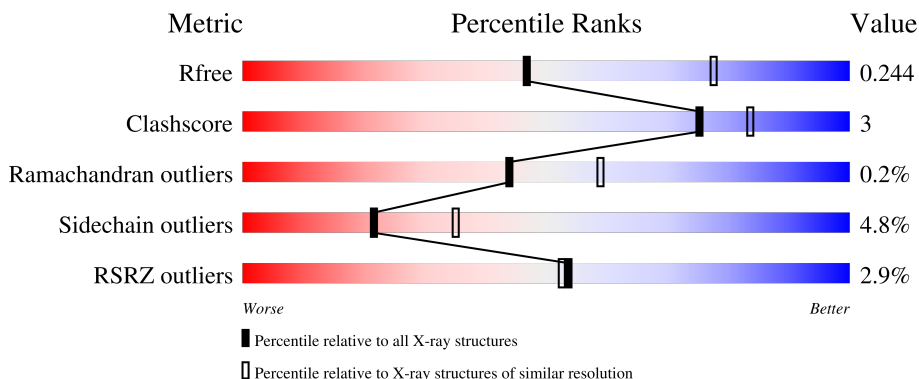
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.63 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))
$R_{free}$	164625	1851 (2.66-2.62)
Clashscore	180529	1953 (2.66-2.62)
Ramachandran outliers	177936	1929 (2.66-2.62)
Sidechain outliers	177891	1929 (2.66-2.62)
RSRZ outliers	164620	1850 (2.66-2.62)

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	676	<div> <div>3%</div> <div>88%</div> <div>10%</div> <div>..</div> </div>
2	E	8	<div> <div>75%</div> <div>25%</div> </div>
3	F	13	<div> <div>38%</div> <div>62%</div> </div>
4	G	5	<div> <div>20%</div> <div>80%</div> <div>20%</div> </div>

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 6013 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 BsmI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	667	Total	C	N	O	S	0	0	0
			5422	3499	914	995	14			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	507	ASP	ARG	engineered mutation	UNP Q8RLN4
A	509	VAL	GLY	engineered mutation	UNP Q8RLN4
A	546	VAL	GLU	engineered mutation	UNP Q8RLN4

- Molecule 2 is a DNA chain called DNA (Bottom strand - 5'-part).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	8	Total	C	N	O	P	0	0	0
			157	76	23	50	8			

- Molecule 3 is a DNA chain called DNA (Top strand).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	13	Total	C	N	O	P	0	0	0
			270	128	58	72	12			

- Molecule 4 is a DNA chain called DNA (Bottom strand - 3'-part).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	G	5	Total	C	N	O	P	0	0	0
			99	49	17	29	4			

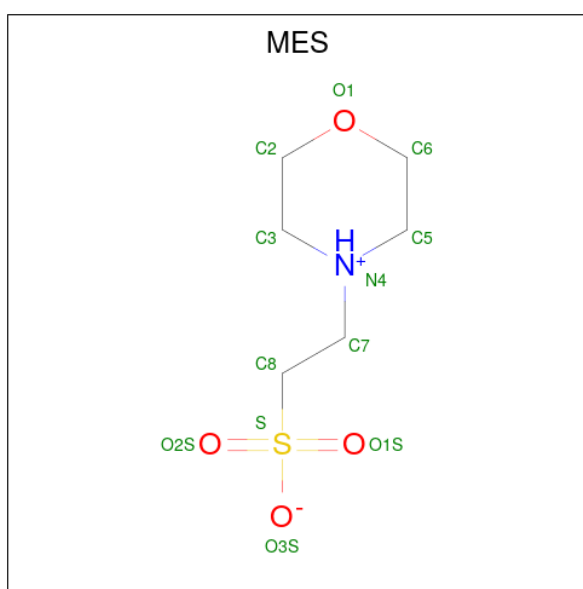
- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

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

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Cl	0	0
			1	1		

- Molecule 7 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

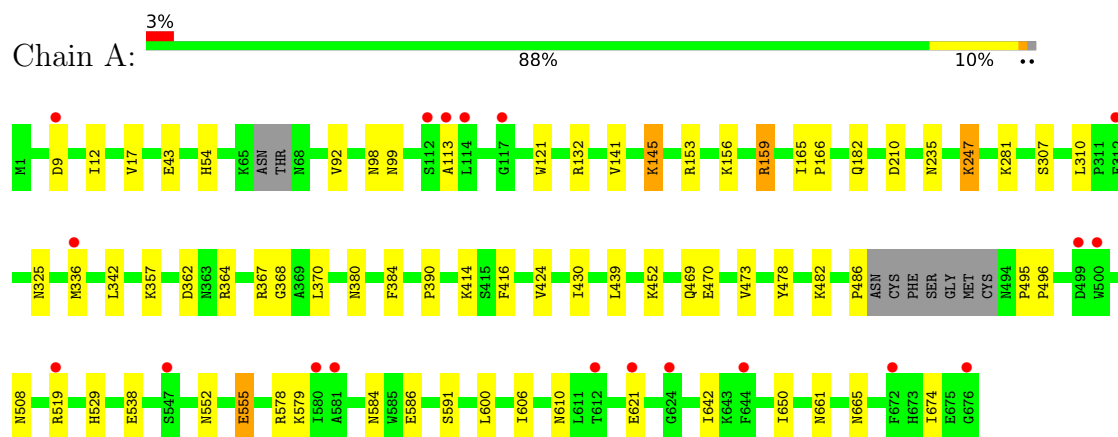
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	45	Total	O	0	0
			45	45		
8	F	4	Total	O	0	0
			4	4		
8	G	2	Total	O	0	0
			2	2		

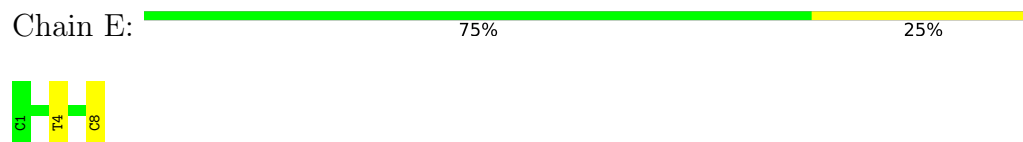
### 3 Residue-property plots [i](#)

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.

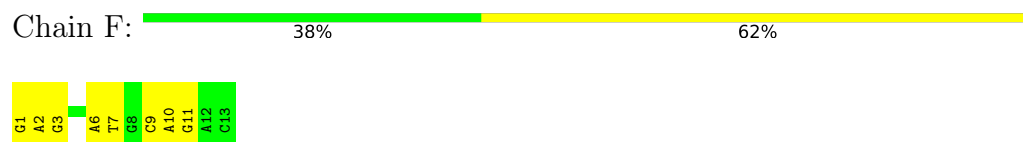
- Molecule 1: BsmI



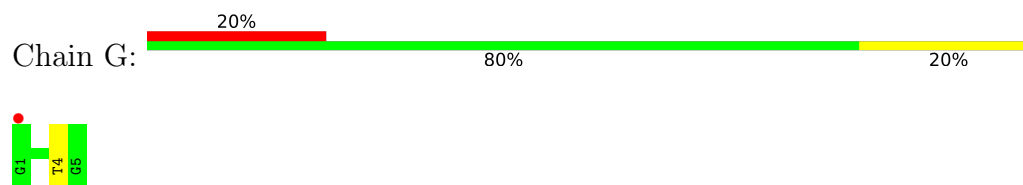
- Molecule 2: DNA (Bottom strand - 5'-part)



- Molecule 3: DNA (Top strand)



- Molecule 4: DNA (Bottom strand - 3'-part)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	98.21Å 130.42Å 62.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	78.45 – 2.63 78.45 – 2.63	Depositor EDS
% Data completeness (in resolution range)	52.1 (78.45-2.63) 52.1 (78.45-2.63)	Depositor EDS
$R_{merge}$	0.52	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.72 (at 2.65Å)	Xtriage
Refinement program	BUSTER 2.10.4	Depositor
R, $R_{free}$	0.219 , 0.252 0.215 , 0.244	Depositor DCC
$R_{free}$ test set	653 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	33.2	Xtriage
Anisotropy	0.397	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 56.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	6013	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.23% 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: CL, MES, 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).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.34	0/5545	0.54	0/7490
2	E	0.59	0/173	0.90	0/263
3	F	0.69	0/305	0.86	0/470
4	G	0.79	0/110	0.84	0/169
All	All	0.39	0/6133	0.59	0/8392

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	5422	0	5414	30	0
2	E	157	0	92	2	0
3	F	270	0	146	9	0
4	G	99	0	56	1	0
5	A	1	0	0	0	0
6	A	1	0	0	0	0
7	A	12	0	13	1	0
8	A	45	0	0	0	0
8	F	4	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	G	2	0	0	0	0
All	All	6013	0	5721	34	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 3.

All (34) 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)
3:F:6:DA:H2''	3:F:7:DT:H5''	1.44	0.99
1:A:12:ILE:HD11	1:A:529:HIS:HE1	1.39	0.86
1:A:486:PRO:HD2	1:A:661:ASN:HB3	1.59	0.81
1:A:17:VAL:HG21	1:A:92:VAL:HG11	1.71	0.71
1:A:156:LYS:NZ	2:E:4:DT:OP2	2.25	0.70
1:A:357:LYS:HG2	3:F:7:DT:H5'	1.78	0.65
1:A:247:LYS:NZ	3:F:3:DG:OP1	2.28	0.57
1:A:600:LEU:HD11	1:A:650:ILE:HG22	1.86	0.57
1:A:12:ILE:HD11	1:A:529:HIS:CE1	2.29	0.57
1:A:367:ARG:NH2	4:G:4:DT:OP2	2.37	0.57
1:A:390:PRO:HB2	1:A:430:ILE:HD11	1.87	0.56
1:A:486:PRO:CD	1:A:661:ASN:HB3	2.35	0.54
1:A:342:LEU:HD11	1:A:439:LEU:HD21	1.89	0.53
1:A:165:ILE:HB	1:A:166:PRO:HD3	1.91	0.53
1:A:384:PHE:CZ	1:A:424:VAL:HG21	2.46	0.51
1:A:281:LYS:HG2	2:E:8:DC:H5'	1.93	0.51
1:A:43:GLU:OE2	1:A:54:HIS:NE2	2.38	0.49
1:A:307:SER:HB2	1:A:310:LEU:HG	1.95	0.48
1:A:113:ALA:HB2	1:A:145:LYS:HE3	1.94	0.48
1:A:159:ARG:NH1	3:F:3:DG:H2''	2.30	0.47
1:A:478:TYR:O	1:A:482:LYS:HB2	2.15	0.45
1:A:495:PRO:HA	1:A:496:PRO:HD3	1.88	0.44
3:F:10:DA:H2''	3:F:11:DG:C8	2.53	0.43
3:F:9:DC:H2''	3:F:10:DA:C8	2.53	0.43
1:A:121:TRP:O	1:A:368:GLY:HA3	2.19	0.42
1:A:642:ILE:HB	1:A:674:ILE:HG22	2.01	0.42
1:A:370:LEU:HD12	1:A:416:PHE:HA	2.02	0.42
1:A:555:GLU:H	1:A:555:GLU:HG2	1.63	0.42
1:A:325:ASN:HB2	7:A:703:MES:H21	2.02	0.42
1:A:473:VAL:HG22	1:A:606:ILE:HG13	2.01	0.42
1:A:159:ARG:HH12	3:F:3:DG:H2''	1.85	0.41
3:F:1:DG:H2''	3:F:2:DA:C8	2.56	0.41

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:182:GLN:H	1:A:235:ASN:ND2	2.19	0.41
1:A:519:ARG:CZ	3:F:10:DA:H5''	2.51	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	661/676 (98%)	645 (98%)	15 (2%)	1 (0%)	44 59

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	584	ASN

### 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	600/614 (98%)	571 (95%)	29 (5%)	21 35

All (29) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	9	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	98	ASN
1	A	99	ASN
1	A	132	ARG
1	A	141	VAL
1	A	145	LYS
1	A	153	ARG
1	A	159	ARG
1	A	210	ASP
1	A	247	LYS
1	A	336	MET
1	A	362	ASP
1	A	364	ARG
1	A	380	ASN
1	A	414	LYS
1	A	452	LYS
1	A	469	GLN
1	A	470	GLU
1	A	508	ASN
1	A	538	GLU
1	A	552	ASN
1	A	555	GLU
1	A	578	ARG
1	A	579	LYS
1	A	586	GLU
1	A	591	SER
1	A	610	ASN
1	A	621	GLU
1	A	665	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	28	GLN
1	A	149	ASN
1	A	235	ASN
1	A	251	ASN
1	A	296	ASN
1	A	458	ASN
1	A	610	ASN
1	A	665	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 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 oligosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 2 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).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
7	MES	A	703	-	12,12,12	0.75	0	14,16,16	0.41	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
7	MES	A	703	-	-	4/6/14/14	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	A	703	MES	C7-C8-S-O2S
7	A	703	MES	C7-C8-S-O3S
7	A	703	MES	C7-C8-S-O1S
7	A	703	MES	C8-C7-N4-C3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	A	703	MES	1	0

## 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<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	667/676 (98%)	0.09	19 (2%) 55 54	16, 39, 63, 81	0
2	E	8/8 (100%)	-0.00	0 100 100	50, 54, 57, 60	0
3	F	13/13 (100%)	0.01	0 100 100	33, 47, 72, 74	0
4	G	5/5 (100%)	0.29	1 (20%) 3 4	50, 53, 67, 71	0
All	All	693/702 (98%)	0.09	20 (2%) 54 52	16, 39, 65, 81	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	114	LEU	5.0
1	A	581	ALA	3.4
1	A	500	TRP	2.9
1	A	113	ALA	2.7
1	A	672	PHE	2.7
1	A	624	GLY	2.5
1	A	676	GLY	2.5
1	A	621	GLU	2.3
1	A	312	PHE	2.3
1	A	117	GLY	2.3
1	A	644	PHE	2.3
1	A	112	SER	2.3
1	A	9	ASP	2.2
1	A	336	MET	2.2
1	A	580	ILE	2.2
1	A	612	THR	2.1
4	G	1	DG	2.1
1	A	547	SER	2.1
1	A	519	ARG	2.0
1	A	499	ASP	2.0

## 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](#)

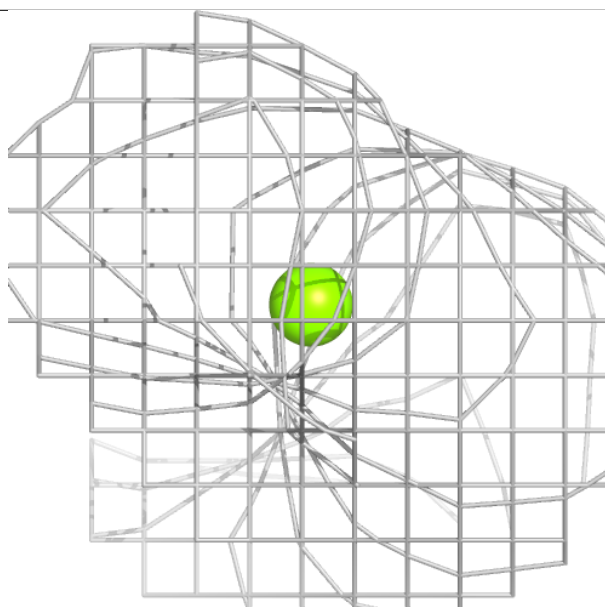
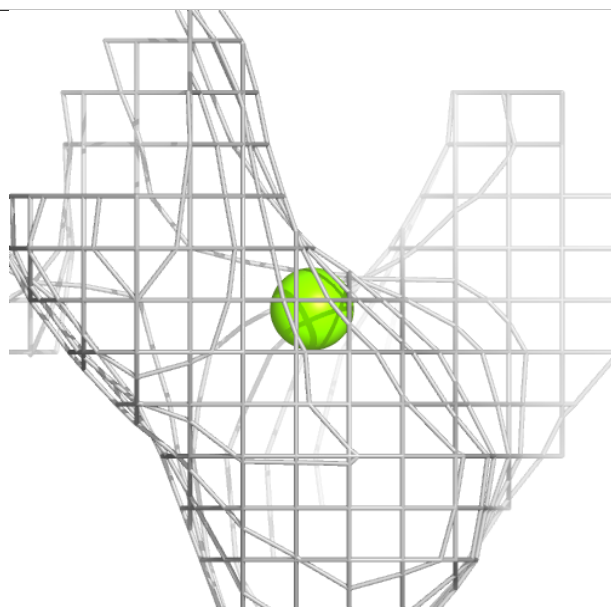
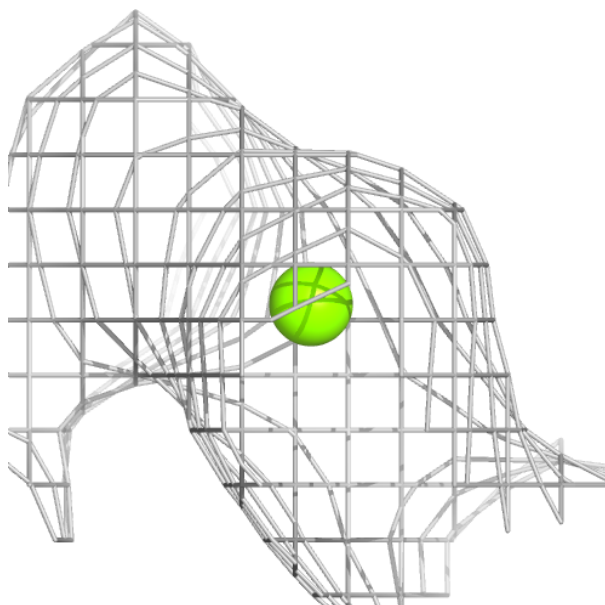
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 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
5	MG	A	701	1/1	0.87	0.22	202,202,202,202	0
7	MES	A	703	12/12	0.92	0.12	63,64,64,65	0
6	CL	A	702	1/1	0.96	0.07	47,47,47,47	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 701:**

$2mF_o-DF_c$  (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.