Full wwPDB X-ray Structure Validation Report (i)

Mar 24, 2025 – 03:10 pm GMT

PDB ID	:	9EZ7
Title	:	BsmI (Nicking top mutant) crystallized with Ca2+ and cognate dsDNA
Authors	:	Sieskind, R.; Missoury, S.; Madru, C.; Commenge, I.; Niogret, G.; Hollenstein, M.; Rondelez, Y.; Haouz, A.; Legrand, P.; Sauguet, L.; Delarue, M.
Deposited on	:	2024-04-10
Resolution	:	2.03 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*
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 (i)) were used in the production of this report:

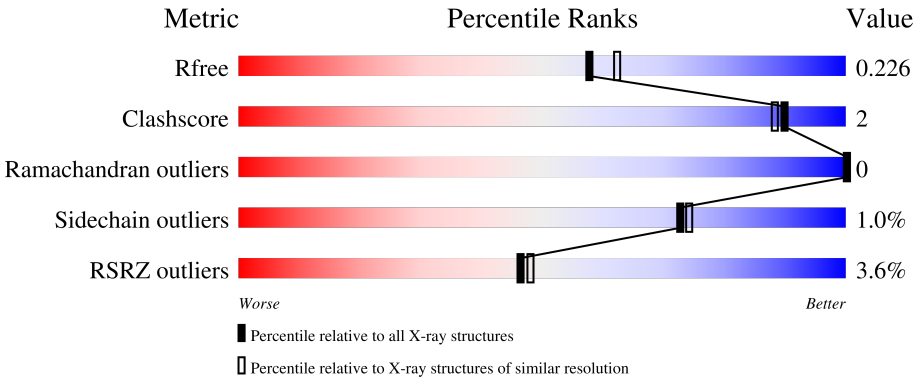
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.03 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	2096 (2.04-2.04)
Clashscore	180529	2229 (2.04-2.04)
Ramachandran outliers	177936	2217 (2.04-2.04)
Sidechain outliers	177891	2217 (2.04-2.04)
RSRZ outliers	164620	2096 (2.04-2.04)

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	669	4% 94% 6%
2	E	13	100%
3	F	13	8% 100%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 6435 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	669	Total	C	N	O	S	0	0	0
			5417	3502	914	988	13			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	109	VAL	GLU	engineered mutation	UNP Q8RLN4
A	?	-	CYS	deletion	UNP Q8RLN4
A	?	-	PHE	deletion	UNP Q8RLN4
A	?	-	SER	deletion	UNP Q8RLN4
A	?	-	GLY	deletion	UNP Q8RLN4
A	?	-	MET	deletion	UNP Q8RLN4
A	?	-	CYS	deletion	UNP Q8RLN4
A	?	-	GLY	deletion	UNP Q8RLN4

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	13	Total	C	N	O	P	0	0	0
			257	125	40	80	12			

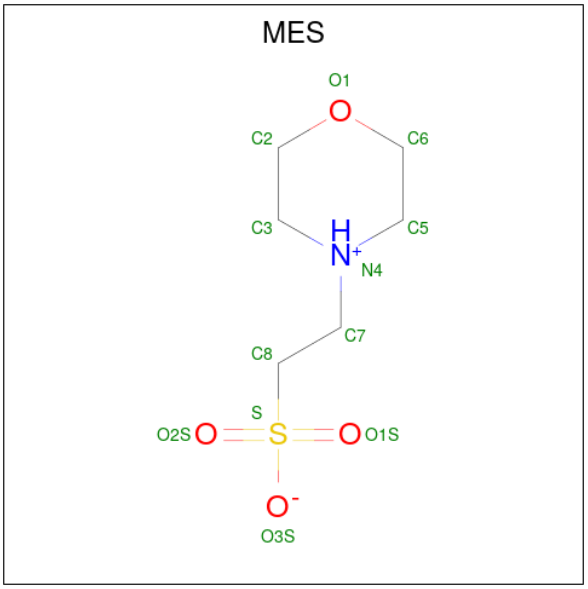
- Molecule 3 is a DNA chain called DNA (Top strand).

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

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

- Molecule 5 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



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

- Molecule 6 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	2	Total	Ca	0	0
			2	2		

- Molecule 7 is water.

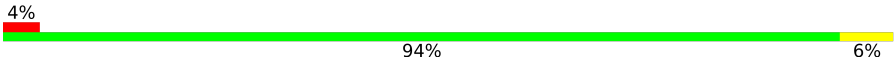
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	421	Total	O	0	0
			421	421		
7	E	28	Total	O	0	0
			28	28		
7	F	27	Total	O	0	0
			27	27		

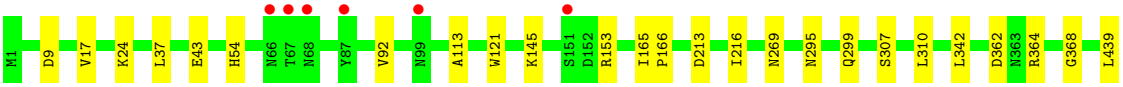
3 Residue-property plots (i)

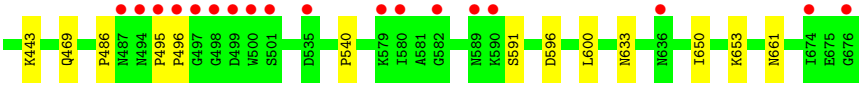
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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

Chain A:







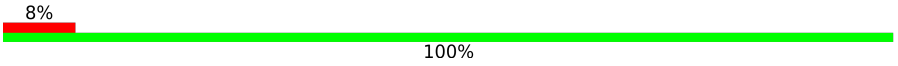
- Molecule 2: DNA (Bottom strand)

Chain E:

There are no outlier residues recorded for this chain.

- Molecule 3: DNA (Top strand)

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	98.55Å 132.99Å 63.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.45 – 2.03 29.45 – 2.03	Depositor EDS
% Data completeness (in resolution range)	73.1 (29.45-2.03) 73.1 (29.45-2.03)	Depositor EDS
R_{merge}	0.30	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.63 (at 2.03Å)	Xtriage
Refinement program	BUSTER 2.10.4	Depositor
R, R_{free}	0.199 , 0.233 0.194 , 0.226	Depositor DCC
R_{free} test set	1880 reflections (4.72%)	wwPDB-VP
Wilson B-factor (Å^2)	22.4	Xtriage
Anisotropy	0.085	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 56.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6435	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.38% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] 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.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CL, MES, CA

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	0.39	0/5543	0.58	0/7494
2	E	0.86	0/285	0.93	0/437
3	F	0.79	0/305	0.88	0/470
All	All	0.45	0/6133	0.62	0/8401

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

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	5417	0	5380	20	0
2	E	257	0	150	0	0
3	F	270	0	146	0	0
4	A	1	0	0	0	0
5	A	12	0	13	0	0
6	A	2	0	0	0	0
7	A	421	0	0	3	0
7	E	28	0	0	0	0
7	F	27	0	0	0	0
All	All	6435	0	5689	20	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 (20) 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:A:17:VAL:HG21	1:A:92:VAL:HG11	1.77	0.64
1:A:24:LYS:HD3	1:A:216:ILE:HD12	1.83	0.59
1:A:486:PRO:HG2	1:A:661:ASN:HD22	1.68	0.59
1:A:486:PRO:CD	1:A:661:ASN:HB3	2.35	0.57
1:A:600:LEU:HD11	1:A:650:ILE:HG22	1.91	0.52
1:A:269:ASN:ND2	7:A:812:HOH:O	2.45	0.50
1:A:113:ALA:HB2	1:A:145:LYS:HE3	1.94	0.48
1:A:37:LEU:HD13	1:A:633:ASN:HB2	1.94	0.48
1:A:165:ILE:HB	1:A:166:PRO:HD3	1.96	0.47
1:A:653:LYS:NZ	7:A:814:HOH:O	2.49	0.46
1:A:307:SER:HB2	1:A:310:LEU:HG	1.99	0.45
1:A:540:PRO:HD2	1:A:596:ASP:O	2.18	0.44
1:A:342:LEU:HD11	1:A:439:LEU:HD21	2.00	0.42
1:A:24:LYS:HD2	1:A:213:ASP:OD1	2.20	0.42
1:A:121:TRP:O	1:A:368:GLY:HA3	2.20	0.42
1:A:495:PRO:HA	1:A:496:PRO:HD3	1.93	0.41
1:A:443:LYS:HD2	7:A:840:HOH:O	2.21	0.41
1:A:486:PRO:HD2	1:A:661:ASN:HB3	2.03	0.41
1:A:43:GLU:OE2	1:A:54:HIS:NE2	2.48	0.40
1:A:295:ASN:O	1:A:299:GLN:HG3	2.22	0.40

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	663/669 (99%)	649 (98%)	14 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains

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	592/608 (97%)	586 (99%)	6 (1%)	73 74

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	9	ASP
1	A	153	ARG
1	A	362	ASP
1	A	364	ARG
1	A	469	GLN
1	A	591	SER

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	149	ASN
1	A	235	ASN
1	A	251	ASN
1	A	269	ASN
1	A	296	ASN
1	A	457	GLN
1	A	494	ASN
1	A	661	ASN

5.3.3 RNA

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 3 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	$\#\|Z\| > 2$
5	MES	A	702	-	12,12,12	0.79	0	14,16,16	0.55	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	MES	A	702	-	-	0/6/14/14	0/1/1/1

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 (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	A	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	487:ASN	C	494:ASN	N	5.99
1	A	582:GLY	C	584:ASN	N	4.01

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^{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	OWAB(Å^2)	Q<0.9
1	A	669/669 (100%)	-0.17	24 (3%) 46 48	12, 23, 44, 70	0
2	E	13/13 (100%)	-0.34	0 100 100	22, 27, 39, 42	0
3	F	13/13 (100%)	-0.36	1 (7%) 21 23	12, 22, 42, 48	0
All	All	695/695 (100%)	-0.18	25 (3%) 46 48	12, 23, 44, 70	0

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	676	GLY	4.3
1	A	495	PRO	4.2
1	A	497	GLY	4.2
1	A	87	TYR	3.7
1	A	498	GLY	3.5
1	A	67	THR	3.5
1	A	582	GLY	3.5
1	A	496	PRO	3.3
1	A	674	ILE	3.3
1	A	499	ASP	3.2
1	A	66	ASN	3.0
1	A	500	TRP	3.0
1	A	589	ASN	2.7
1	A	636	ASN	2.7
1	A	501	SER	2.6
1	A	535	ASP	2.5
1	A	151	SER	2.4
1	A	494	ASN	2.4
1	A	68	ASN	2.3
1	A	579	LYS	2.3
1	A	590	LYS	2.3
1	A	99	ASN	2.3
1	A	487	ASN	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	F	13	DC	2.1
1	A	580	ILE	2.1

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)

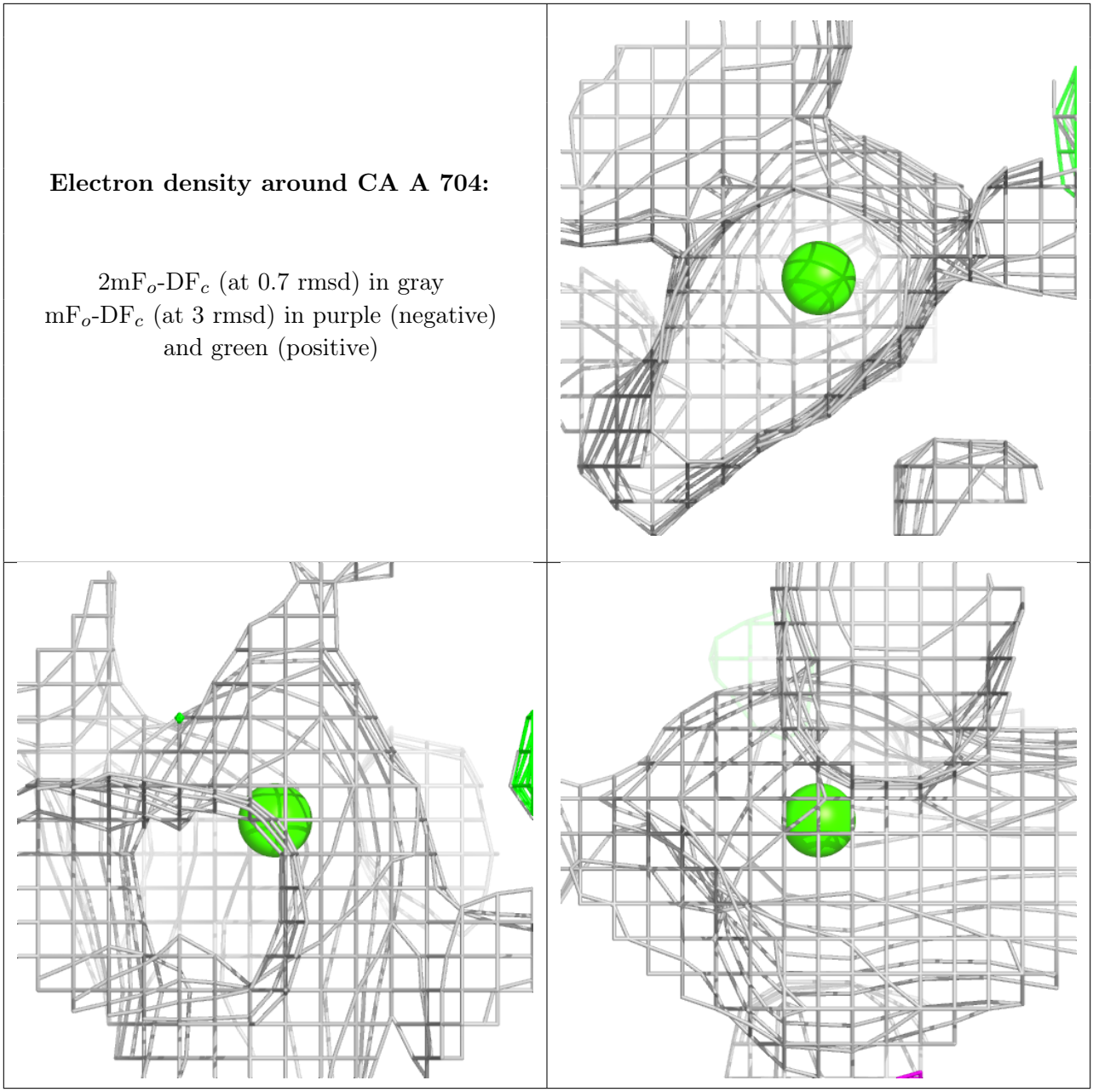
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	MES	A	702	12/12	0.95	0.08	23,24,26,27	0
6	CA	A	704	1/1	0.95	0.07	59,59,59,59	0
4	CL	A	701	1/1	0.99	0.03	24,24,24,24	0
6	CA	A	703	1/1	1.00	0.01	23,23,23,23	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.

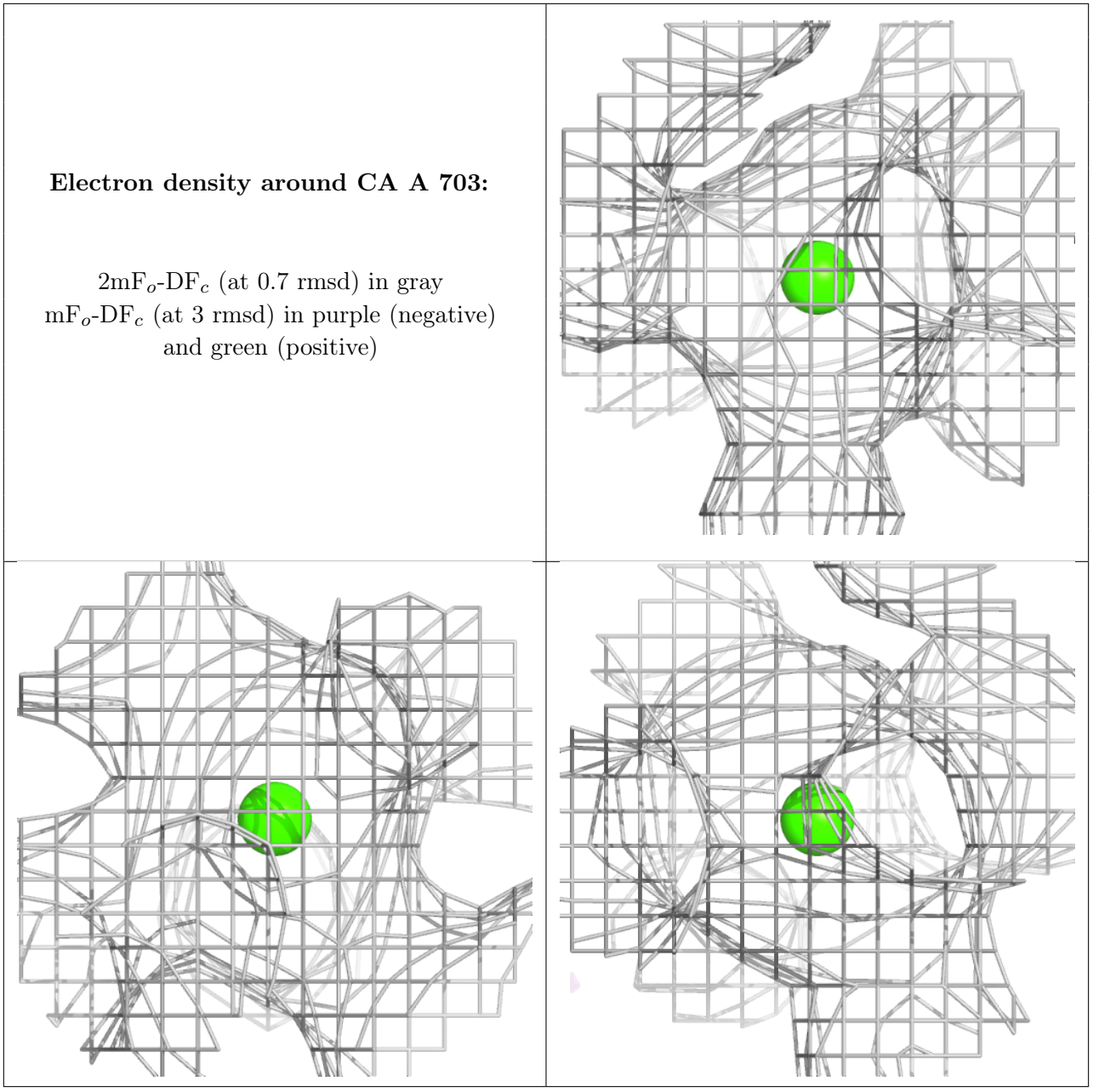
Electron density around CA A 704:

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



Electron density around CA A 703:

$2mF_o$-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.