



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 19, 2025 – 03:10 pm GMT

PDB ID : 9GYZ  
Title : Crystal structure of domain-of-unknown-function DUF4867 from *Bacillus megaterium* (unmodelled additional ligand density at active site)  
Authors : Sharma, M.; Davies, G.J.  
Deposited on : 2024-10-03  
Resolution : 2.10 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  
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

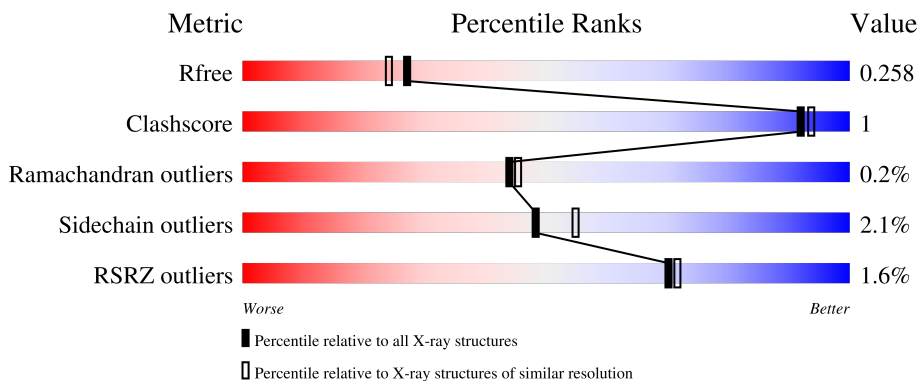
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	6234 (2.10-2.10)
Clashscore	180529	6893 (2.10-2.10)
Ramachandran outliers	177936	6839 (2.10-2.10)
Sidechain outliers	177891	6840 (2.10-2.10)
RSRZ outliers	164620	6234 (2.10-2.10)

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	243	 % 82% 8% 9%
1	B	243	 2% 82% 9% 9%

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 3598 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 DUF4867 domain-containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	221	Total 1742	C 1120	N 282	O 332	S 8	0	1	0
1	A	221	Total 1752	C 1126	N 287	O 331	S 8	0	1	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-19	MET	-	initiating methionine	UNP D5E1T1
B	-18	GLY	-	expression tag	UNP D5E1T1
B	-17	SER	-	expression tag	UNP D5E1T1
B	-16	SER	-	expression tag	UNP D5E1T1
B	-15	HIS	-	expression tag	UNP D5E1T1
B	-14	HIS	-	expression tag	UNP D5E1T1
B	-13	HIS	-	expression tag	UNP D5E1T1
B	-12	HIS	-	expression tag	UNP D5E1T1
B	-11	HIS	-	expression tag	UNP D5E1T1
B	-10	HIS	-	expression tag	UNP D5E1T1
B	-9	SER	-	expression tag	UNP D5E1T1
B	-8	SER	-	expression tag	UNP D5E1T1
B	-7	GLY	-	expression tag	UNP D5E1T1
B	-6	LEU	-	expression tag	UNP D5E1T1
B	-5	VAL	-	expression tag	UNP D5E1T1
B	-4	PRO	-	expression tag	UNP D5E1T1
B	-3	ARG	-	expression tag	UNP D5E1T1
B	-2	GLY	-	expression tag	UNP D5E1T1
B	-1	SER	-	expression tag	UNP D5E1T1
B	0	HIS	-	expression tag	UNP D5E1T1
A	-19	MET	-	initiating methionine	UNP D5E1T1
A	-18	GLY	-	expression tag	UNP D5E1T1
A	-17	SER	-	expression tag	UNP D5E1T1
A	-16	SER	-	expression tag	UNP D5E1T1
A	-15	HIS	-	expression tag	UNP D5E1T1

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-14	HIS	-	expression tag	UNP D5E1T1
A	-13	HIS	-	expression tag	UNP D5E1T1
A	-12	HIS	-	expression tag	UNP D5E1T1
A	-11	HIS	-	expression tag	UNP D5E1T1
A	-10	HIS	-	expression tag	UNP D5E1T1
A	-9	SER	-	expression tag	UNP D5E1T1
A	-8	SER	-	expression tag	UNP D5E1T1
A	-7	GLY	-	expression tag	UNP D5E1T1
A	-6	LEU	-	expression tag	UNP D5E1T1
A	-5	VAL	-	expression tag	UNP D5E1T1
A	-4	PRO	-	expression tag	UNP D5E1T1
A	-3	ARG	-	expression tag	UNP D5E1T1
A	-2	GLY	-	expression tag	UNP D5E1T1
A	-1	SER	-	expression tag	UNP D5E1T1
A	0	HIS	-	expression tag	UNP D5E1T1

- Molecule 2 is FE (III) ION (three-letter code: FE) (formula: Fe) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Fe 1 1	0	0
2	A	1	Total Fe 1 1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	51	Total O 51 51	0	0
3	A	51	Total O 51 51	0	0



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	82.51Å 82.51Å 136.19Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.44 – 2.10 45.44 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.5 (45.44-2.10) 99.5 (45.44-2.10)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.99 (at 2.10Å)	Xtrriage
Refinement program	REFMAC 5.8.0430	Depositor
R, $R_{free}$	0.208 , 0.255 0.215 , 0.258	Depositor DCC
$R_{free}$ test set	1592 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.1	Xtrriage
Anisotropy	0.389	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 41.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.021 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3598	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.09% 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: FE

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	0.82	0/1796	1.22	14/2438 (0.6%)
1	B	0.77	1/1784 (0.1%)	1.14	7/2423 (0.3%)
All	All	0.80	1/3580 (0.0%)	1.18	21/4861 (0.4%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	63	GLU	CD-OE1	-7.21	1.17	1.25

All (21) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	182	ARG	NE-CZ-NH1	-9.75	115.43	120.30
1	A	173	ASP	CB-CG-OD2	-7.35	111.69	118.30
1	A	73	GLU	CG-CD-OE2	6.77	131.83	118.30
1	B	111	MET	CG-SD-CE	-6.64	89.58	100.20
1	A	33	GLU	CB-CA-C	5.99	122.37	110.40
1	B	5	GLN	CB-CA-C	5.85	122.11	110.40
1	A	29	LYS	CB-CA-C	-5.84	98.72	110.40
1	B	215	LYS	CA-CB-CG	5.84	126.24	113.40
1	B	26	HIS	CA-CB-CG	-5.72	103.87	113.60
1	A	202	ARG	NE-CZ-NH2	-5.61	117.50	120.30

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	100	LYS	CB-CA-C	5.48	121.35	110.40
1	A	76	TYR	CB-CG-CD2	5.44	124.27	121.00
1	A	192	LYS	CB-CA-C	5.35	121.10	110.40
1	A	223	GLN	CA-C-O	-5.32	108.93	120.10
1	A	182	ARG	NH1-CZ-NH2	5.31	125.24	119.40
1	B	223	GLN	CA-C-O	-5.29	109.00	120.10
1	A	202	ARG	NE-CZ-NH1	5.22	122.91	120.30
1	B	129	ASN	CB-CA-C	-5.15	100.10	110.40
1	B	29	LYS	CA-CB-CG	5.11	124.64	113.40
1	A	203	GLU	OE1-CD-OE2	-5.08	117.20	123.30
1	A	43	MET	CG-SD-CE	-5.02	92.17	100.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	202	ARG	Sidechain

## 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	1752	0	1687	4	0
1	B	1742	0	1669	6	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	51	0	0	0	0
3	B	51	0	0	0	0
All	All	3598	0	3356	10	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 1.

All (10) 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:B:76:TYR:CD2	1:B:81:ILE:HG21	2.42	0.53
1:A:76:TYR:CD2	1:A:81:ILE:HG21	2.45	0.51
1:B:33:GLU:N	1:B:33:GLU:OE1	2.44	0.48
1:A:15:PRO:HG2	1:A:132:ILE:HD13	1.96	0.47
1:B:159:GLU:CD	1:B:159:GLU:H	2.18	0.47
1:B:171:THR:HG23	1:B:193:ASN:HB3	1.97	0.46
1:A:103:GLU:HB2	1:A:167:LEU:HD21	1.99	0.44
1:B:104:ILE:HB	1:B:166:ILE:HB	1.99	0.43
1:A:104:ILE:HB	1:A:166:ILE:HB	2.01	0.42
1:B:51:ASP:OD1	1:B:51:ASP:N	2.53	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	220/243 (90%)	205 (93%)	14 (6%)	1 (0%)	25	23
1	B	220/243 (90%)	209 (95%)	11 (5%)	0	100	100
All	All	440/486 (90%)	414 (94%)	25 (6%)	1 (0%)	44	45

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	218	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	190/219 (87%)	187 (98%)	3 (2%)	58	65
1	B	187/219 (85%)	182 (97%)	5 (3%)	40	44
All	All	377/438 (86%)	369 (98%)	8 (2%)	48	55

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	3	SER
1	B	8	LYS
1	B	89	PRO
1	B	204	GLN
1	B	223	GLN
1	A	3	SER
1	A	37	ASN
1	A	223	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	123	ASN
1	B	178	ASN
1	A	69	ASN

### 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 oligosaccharides 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	221/243 (90%)	0.15	3 (1%) 73 74	30, 47, 64, 85	1 (0%)
1	B	221/243 (90%)	0.21	4 (1%) 67 69	26, 48, 67, 84	1 (0%)
All	All	442/486 (90%)	0.18	7 (1%) 70 71	26, 48, 66, 85	2 (0%)

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	26	HIS	2.6
1	A	26	HIS	2.4
1	A	3	SER	2.3
1	B	6	TYR	2.2
1	A	14	ILE	2.2
1	B	214	ILE	2.1
1	B	3	SER	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](#)

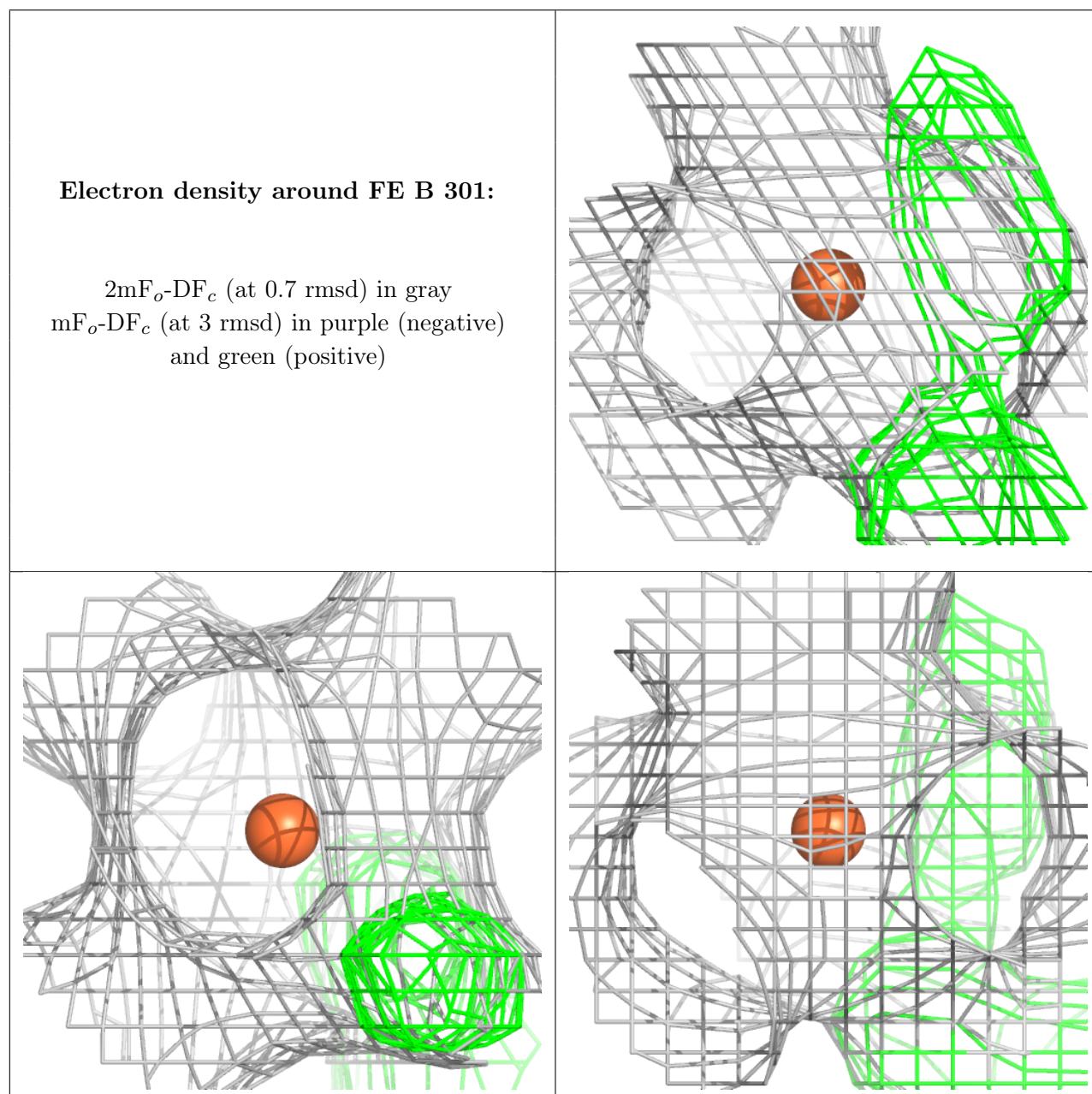
There are no monosaccharides in this entry.

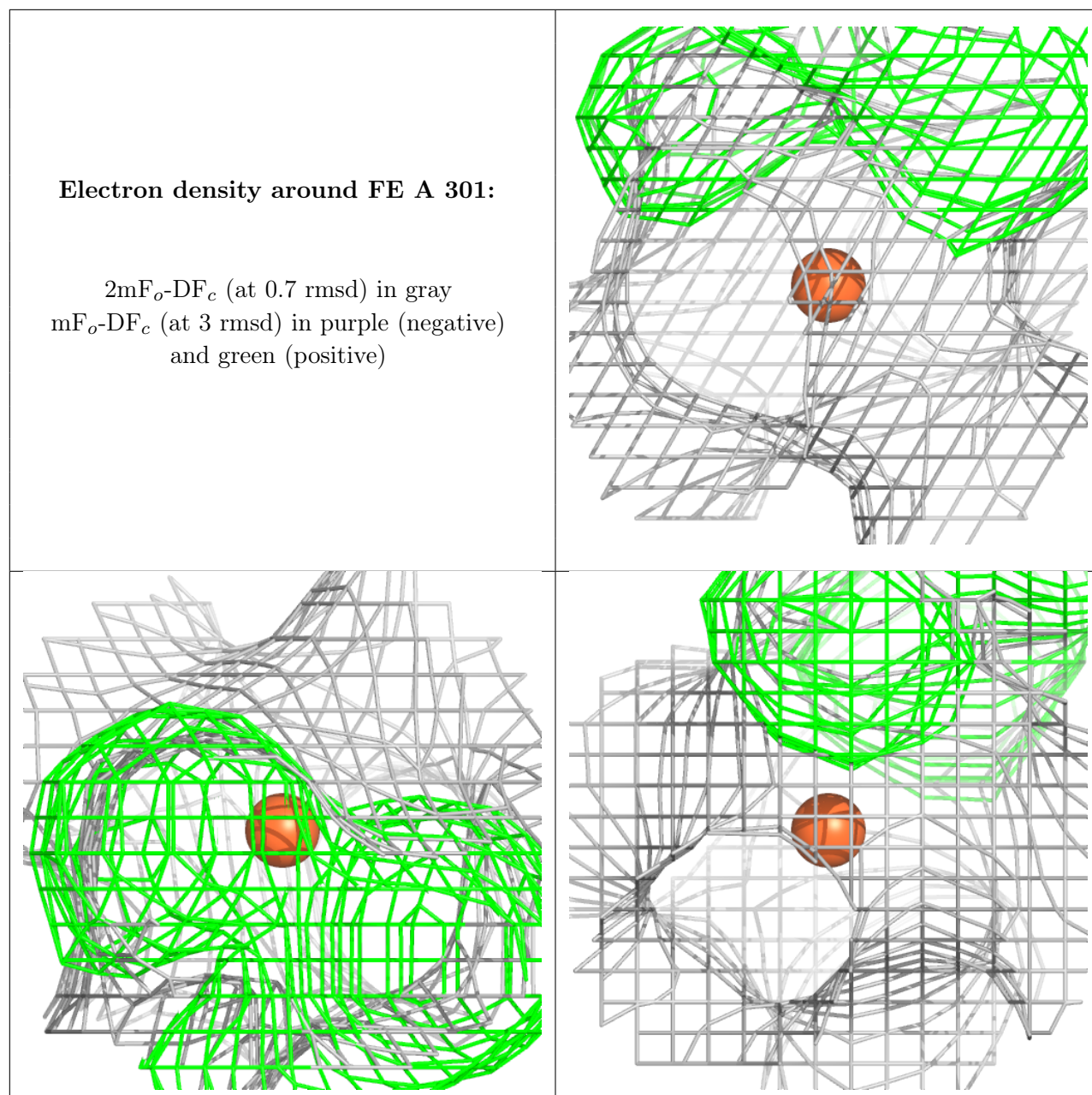
### 6.4 Ligands [i](#)

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
2	FE	B	301	1/1	0.97	0.05	53,53,53,53	0
2	FE	A	301	1/1	0.97	0.05	48,48,48,48	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.