



# Full wwPDB X-ray Structure Validation Report ⓘ

Apr 18, 2022 – 12:04 PM JST

PDB ID : 7VGM  
Title : Crystal structure of Phenylalanine hydroxylase from *Bacillus cereus* ATCC 14579  
Authors : Park, J.; Kim, K.-J.  
Deposited on : 2021-09-17  
Resolution : 2.27 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtrriage (Phenix) : 1.13  
EDS : 2.27  
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) : 2.27

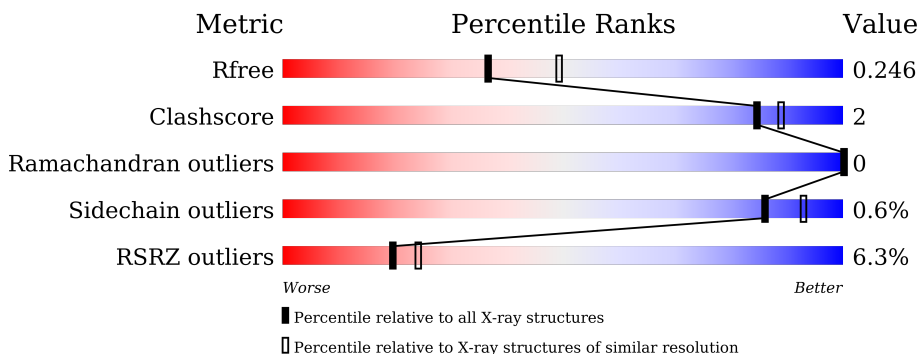
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.27 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))
$R_{free}$	130704	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)
RSRZ outliers	127900	6849 (2.30-2.26)

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	592	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	ACY	A	601	-	-	X	-

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 4856 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.

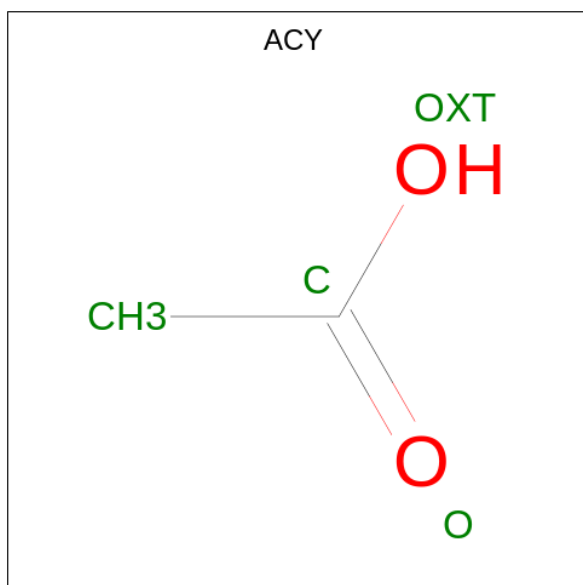
- Molecule 1 is a protein called Phenylalanine-4-hydroxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	571	4425	2818	731	862	14	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	585	LEU	-	cloning artifact	UNP Q818B4
A	586	GLU	-	cloning artifact	UNP Q818B4
A	587	HIS	-	expression tag	UNP Q818B4
A	588	HIS	-	expression tag	UNP Q818B4
A	589	HIS	-	expression tag	UNP Q818B4
A	590	HIS	-	expression tag	UNP Q818B4
A	591	HIS	-	expression tag	UNP Q818B4
A	592	HIS	-	expression tag	UNP Q818B4

- Molecule 2 is ACETIC ACID (three-letter code: ACY) (formula: C<sub>2</sub>H<sub>4</sub>O<sub>2</sub>).

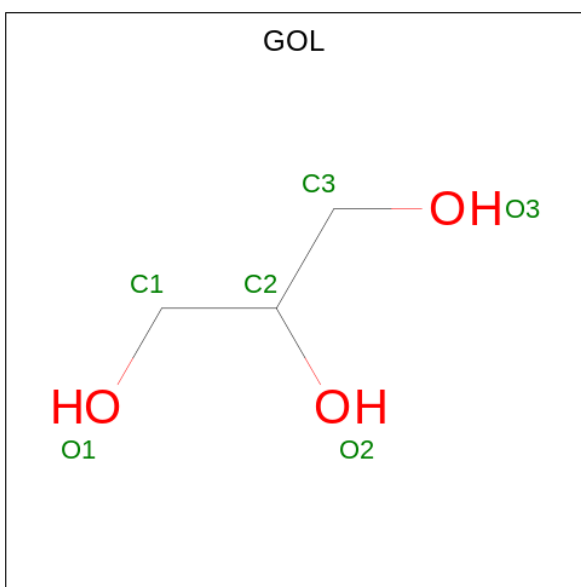


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0

- Molecule 3 is FE (III) ION (three-letter code: FE) (formula: Fe) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Fe 1 1	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 6 3 3	0	0
4	A	1	Total C O 6 3 3	0	0
4	A	1	Total C O 6 3 3	0	0

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0

- Molecule 6 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total C O 7 4 3	0	0
6	A	1	Total C O 7 4 3	0	0
6	A	1	Total C O 7 4 3	0	0

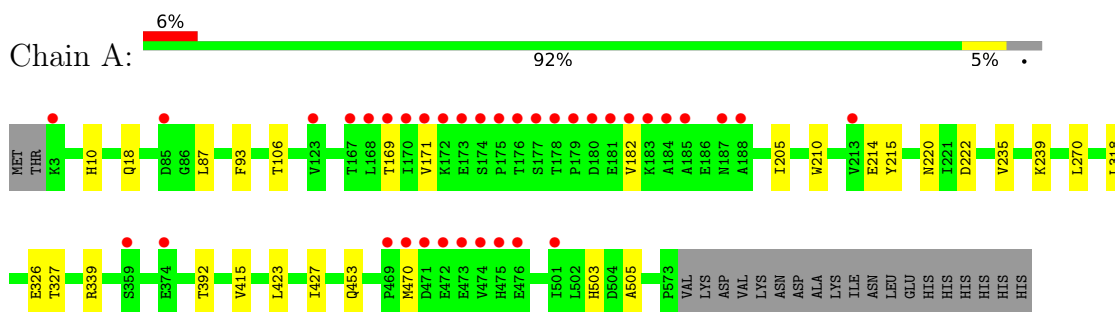
- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	351	Total O 351 351	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: Phenylalanine-4-hydroxylase



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	77.27Å 93.70Å 95.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.49 – 2.27 33.46 – 2.27	Depositor EDS
% Data completeness (in resolution range)	98.8 (33.49-2.27) 98.9 (33.46-2.27)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.97 (at 2.27Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.173 , 0.241 0.196 , 0.246	Depositor DCC
$R_{free}$ test set	1570 reflections (4.87%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	30.1	Xtrriage
Anisotropy	0.123	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 38.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.023 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4856	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.24% 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: GOL, ACY, FE, EDO, PEG

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	0.57	0/4521	0.76	0/6151

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	4425	0	4354	18	0
2	A	12	0	9	5	0
3	A	1	0	0	0	0
4	A	18	0	24	0	0
5	A	28	0	42	0	0
6	A	21	0	30	2	0
7	A	351	0	0	2	0
All	All	4856	0	4459	21	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 (21) 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:93:PHE:HE1	2:A:601:ACY:H3	1.30	0.95
1:A:93:PHE:CE1	2:A:601:ACY:H3	2.03	0.94
2:A:601:ACY:H2	7:A:733:HOH:O	1.90	0.70
1:A:93:PHE:HE1	2:A:601:ACY:CH3	2.07	0.61
2:A:601:ACY:CH3	7:A:733:HOH:O	2.53	0.49
1:A:453:GLN:HB3	6:A:608:PEG:H22	1.95	0.47
1:A:171:VAL:HG12	1:A:182:VAL:HG22	1.97	0.46
1:A:503:HIS:CE1	1:A:505:ALA:HB3	2.50	0.46
1:A:220:ASN:OD1	1:A:222:ASP:HB2	2.17	0.44
1:A:87:LEU:HD11	1:A:169:THR:HG21	2.00	0.44
1:A:326:GLU:HG2	1:A:327:THR:N	2.34	0.43
1:A:235:VAL:CG2	1:A:239:LYS:HE3	2.49	0.42
1:A:18:GLN:HB2	1:A:106:THR:HG21	2.01	0.42
1:A:10:HIS:HA	1:A:470:MET:SD	2.60	0.42
1:A:215:TYR:HB3	1:A:270:LEU:HD22	2.02	0.41
1:A:235:VAL:HG22	1:A:239:LYS:HE3	2.03	0.41
1:A:415:VAL:HB	1:A:423:LEU:HB3	2.01	0.41
1:A:210:TRP:HA	1:A:214:GLU:HB2	2.03	0.41
6:A:609:PEG:H31	6:A:609:PEG:H12	1.83	0.41
1:A:427:ILE:HD12	1:A:427:ILE:HA	1.92	0.40
1:A:205:ILE:HD13	1:A:205:ILE:HA	1.98	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	569/592 (96%)	556 (98%)	13 (2%)	0	<a href="#">100</a> <a href="#">100</a>

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	Rotameric	Outliers	Percentiles
1	A	480/500 (96%)	477 (99%)	3 (1%)	86 93

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	318	LEU
1	A	339	ARG
1	A	392	THR

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 17 ligands modelled in this entry, 1 is monoatomic - leaving 16 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
2	ACY	A	615	-	1,3,3	3.38	1 (100%)	0,3,3	-	-
6	PEG	A	608	-	6,6,6	0.13	0	5,5,5	0.10	0
2	ACY	A	601	3	1,3,3	3.78	1 (100%)	0,3,3	-	-
4	GOL	A	604	-	5,5,5	0.08	0	5,5,5	0.38	0
5	EDO	A	605	-	3,3,3	0.09	0	2,2,2	0.39	0
5	EDO	A	606	-	3,3,3	0.12	0	2,2,2	0.31	0
5	EDO	A	610	-	3,3,3	0.11	0	2,2,2	0.19	0
5	EDO	A	611	-	3,3,3	0.14	0	2,2,2	0.24	0
5	EDO	A	613	-	3,3,3	0.11	0	2,2,2	0.20	0
5	EDO	A	616	-	3,3,3	0.07	0	2,2,2	0.24	0
4	GOL	A	603	-	5,5,5	0.11	0	5,5,5	0.26	0
6	PEG	A	607	-	6,6,6	0.15	0	5,5,5	0.12	0
6	PEG	A	609	-	6,6,6	0.11	0	5,5,5	0.11	0
4	GOL	A	617	-	5,5,5	0.09	0	5,5,5	0.39	0
2	ACY	A	614	3	1,3,3	2.47	1 (100%)	0,3,3	-	-
5	EDO	A	612	-	3,3,3	0.09	0	2,2,2	0.50	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
6	PEG	A	608	-	-	1/4/4/4	-
5	EDO	A	605	-	-	0/1/1/1	-
4	GOL	A	604	-	-	2/4/4/4	-
5	EDO	A	606	-	-	0/1/1/1	-
5	EDO	A	610	-	-	0/1/1/1	-
5	EDO	A	611	-	-	0/1/1/1	-
5	EDO	A	613	-	-	0/1/1/1	-
5	EDO	A	616	-	-	0/1/1/1	-
4	GOL	A	603	-	-	1/4/4/4	-
6	PEG	A	607	-	-	2/4/4/4	-
6	PEG	A	609	-	-	1/4/4/4	-
4	GOL	A	617	-	-	1/4/4/4	-
5	EDO	A	612	-	-	0/1/1/1	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	ACY	CH3-C	3.78	1.53	1.48
2	A	615	ACY	CH3-C	3.38	1.53	1.48
2	A	614	ACY	CH3-C	2.47	1.51	1.48

There are no bond angle outliers.

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	604	GOL	O1-C1-C2-C3
6	A	607	PEG	C4-C3-O2-C2
4	A	617	GOL	C1-C2-C3-O3
4	A	604	GOL	O1-C1-C2-O2
6	A	607	PEG	O1-C1-C2-O2
6	A	608	PEG	C4-C3-O2-C2
6	A	609	PEG	C1-C2-O2-C3
4	A	603	GOL	C1-C2-C3-O3

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	608	PEG	1	0
2	A	601	ACY	5	0
6	A	609	PEG	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	571/592 (96%)	0.22	36 (6%) <span style="border: 1px solid red; padding: 2px;">20</span> <span style="border: 1px solid red; padding: 2px;">24</span>	19, 33, 71, 121	0

All (36) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	174	SER	6.4
1	A	474	VAL	6.4
1	A	179	PRO	6.2
1	A	180	ASP	5.9
1	A	471	ASP	5.7
1	A	173	GLU	5.1
1	A	171	VAL	4.9
1	A	177	SER	4.7
1	A	168	LEU	4.5
1	A	472	GLU	4.4
1	A	182	VAL	4.2
1	A	473	GLU	4.1
1	A	175	PRO	4.1
1	A	184	ALA	4.1
1	A	476	GLU	3.8
1	A	172	LYS	3.8
1	A	178	THR	3.7
1	A	470	MET	3.6
1	A	475	HIS	3.5
1	A	176	THR	3.3
1	A	123	VAL	3.2
1	A	169	THR	3.2
1	A	183	LYS	3.1
1	A	469	PRO	2.8
1	A	170	ILE	2.8
1	A	187	ASN	2.7
1	A	213	VAL	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	85	ASP	2.5
1	A	3	LYS	2.5
1	A	167	THR	2.5
1	A	359	SER	2.4
1	A	188	ALA	2.4
1	A	374	GLU	2.3
1	A	501	ILE	2.2
1	A	185	ALA	2.1
1	A	181	GLU	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<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(Å <sup>2</sup> )	Q<0.9
5	EDO	A	612	4/4	0.66	0.30	39,46,47,51	0
6	PEG	A	609	7/7	0.69	0.27	50,54,58,58	0
4	GOL	A	604	6/6	0.71	0.25	47,50,53,56	0
4	GOL	A	617	6/6	0.71	0.23	55,56,60,60	0
5	EDO	A	616	4/4	0.72	0.20	48,57,58,61	0
5	EDO	A	605	4/4	0.72	0.28	44,49,50,50	0
5	EDO	A	610	4/4	0.77	0.16	56,62,64,66	0
5	EDO	A	606	4/4	0.81	0.32	58,62,63,65	0
5	EDO	A	611	4/4	0.83	0.18	40,52,55,56	0
6	PEG	A	608	7/7	0.86	0.20	40,54,55,56	0
6	PEG	A	607	7/7	0.86	0.30	45,47,55,56	0
2	ACY	A	601	4/4	0.87	0.21	20,20,20,20	0
4	GOL	A	603	6/6	0.88	0.20	38,44,48,52	0
5	EDO	A	613	4/4	0.88	0.20	52,58,59,60	0

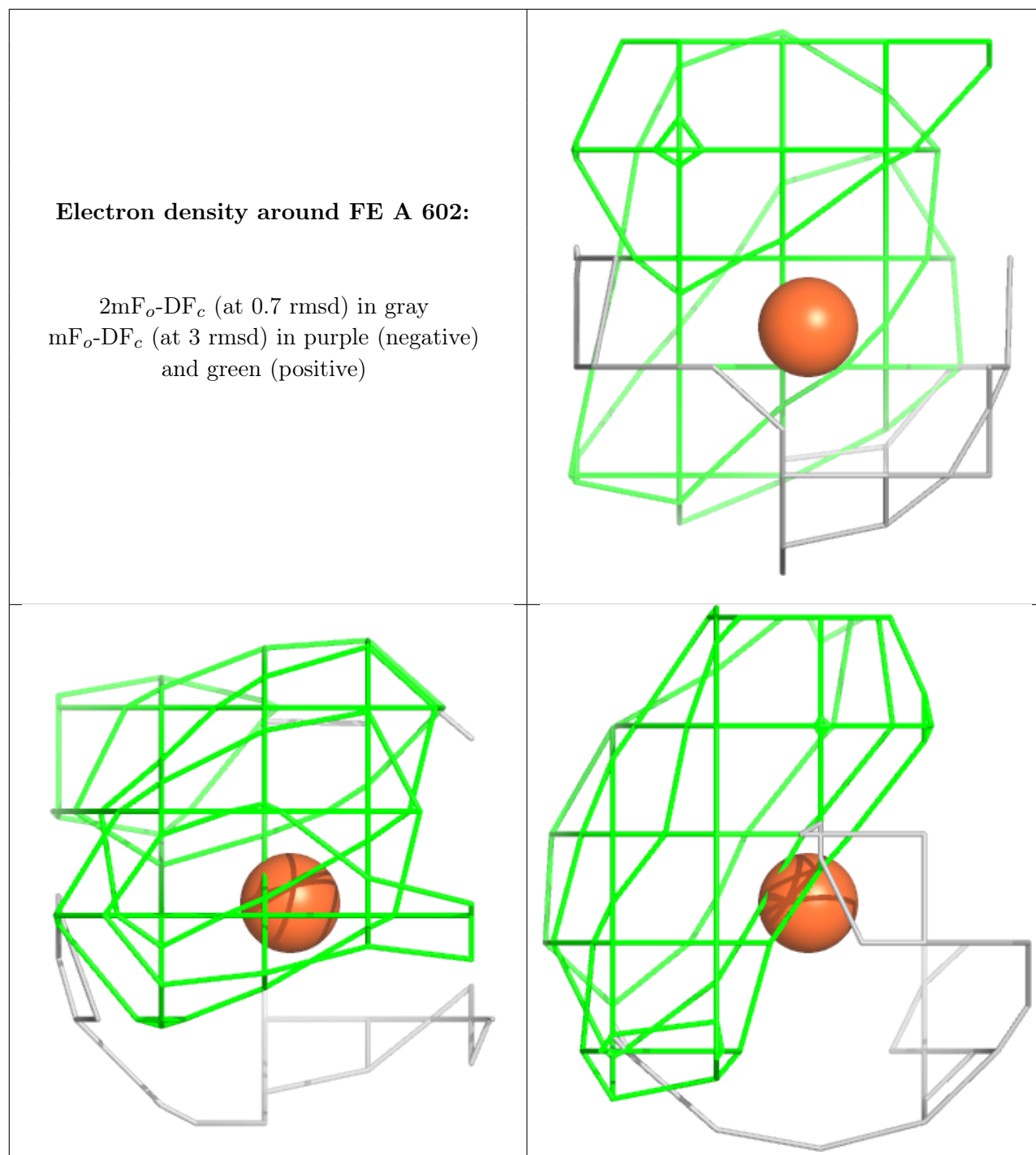
*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Type</b>	<b>Chain</b>	<b>Res</b>	<b>Atoms</b>	<b>RSCC</b>	<b>RSR</b>	<b>B-factors(<math>\text{\AA}^2</math>)</b>	<b>Q&lt;0.9</b>
2	ACY	A	615	4/4	0.91	0.14	42,45,51,52	0
2	ACY	A	614	4/4	0.94	0.31	36,37,40,46	0
3	FE	A	602	1/1	0.99	0.17	17,17,17,17	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.