



wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 07:16 pm BST

PDB ID : 4R20
Title : Zebra fish cytochrome P450 17A2 with Abiraterone
Authors : Pallan, P.S.; Egli, M.
Deposited on : 2014-08-08
Resolution : 2.86 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 ⓘ 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)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

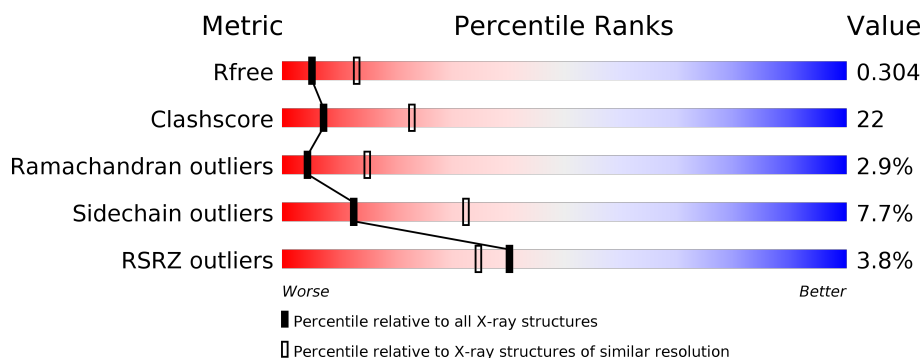
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.86 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	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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 on 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 $\leq 5\%$. 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	486	<div> <div>4%</div> <div> <div></div> <div>58%</div> <div>21%</div> <div>• •</div> <div>16%</div> </div> </div>
1	B	486	<div> <div>2%</div> <div> <div></div> <div>58%</div> <div>22%</div> <div>• •</div> <div>15%</div> </div> </div>

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
4	HG	B	604	-	-	X	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6555 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 Cytochrome P450 family 17 polypeptide 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	406	Total	C	N	O	S	18	0	0
			3171	2011	563	582	15			
1	B	414	Total	C	N	O	S	5	0	0
			3227	2052	570	590	15			

There are 32 discrepancies between the modelled and reference sequences:

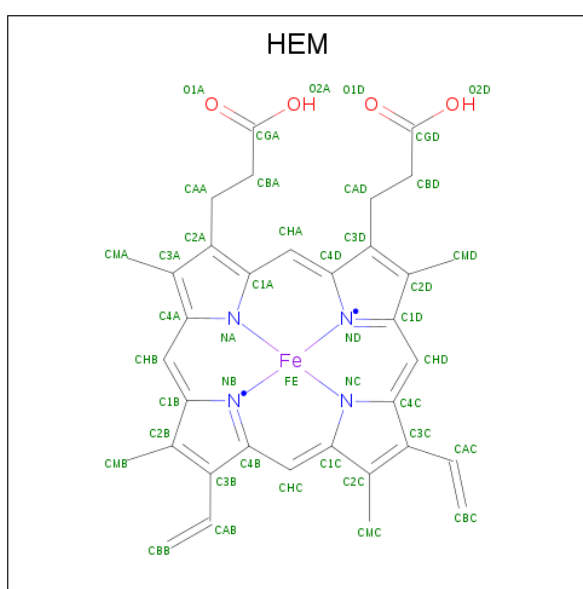
Chain	Residue	Modelled	Actual	Comment	Reference
A	16	MET	-	EXPRESSION TAG	UNP A7U483
A	17	ALA	-	EXPRESSION TAG	UNP A7U483
A	18	LYS	-	EXPRESSION TAG	UNP A7U483
A	19	LYS	-	EXPRESSION TAG	UNP A7U483
A	20	THR	-	EXPRESSION TAG	UNP A7U483
A	21	SER	-	EXPRESSION TAG	UNP A7U483
A	22	SER	-	EXPRESSION TAG	UNP A7U483
A	23	LYS	-	EXPRESSION TAG	UNP A7U483
A	24	GLY	-	EXPRESSION TAG	UNP A7U483
A	25	LYS	-	EXPRESSION TAG	UNP A7U483
A	496	HIS	-	EXPRESSION TAG	UNP A7U483
A	497	HIS	-	EXPRESSION TAG	UNP A7U483
A	498	HIS	-	EXPRESSION TAG	UNP A7U483
A	499	HIS	-	EXPRESSION TAG	UNP A7U483
A	500	HIS	-	EXPRESSION TAG	UNP A7U483
A	501	HIS	-	EXPRESSION TAG	UNP A7U483
B	16	MET	-	EXPRESSION TAG	UNP A7U483
B	17	ALA	-	EXPRESSION TAG	UNP A7U483
B	18	LYS	-	EXPRESSION TAG	UNP A7U483
B	19	LYS	-	EXPRESSION TAG	UNP A7U483
B	20	THR	-	EXPRESSION TAG	UNP A7U483
B	21	SER	-	EXPRESSION TAG	UNP A7U483
B	22	SER	-	EXPRESSION TAG	UNP A7U483
B	23	LYS	-	EXPRESSION TAG	UNP A7U483
B	24	GLY	-	EXPRESSION TAG	UNP A7U483

Continued on next page...

Continued from previous page...

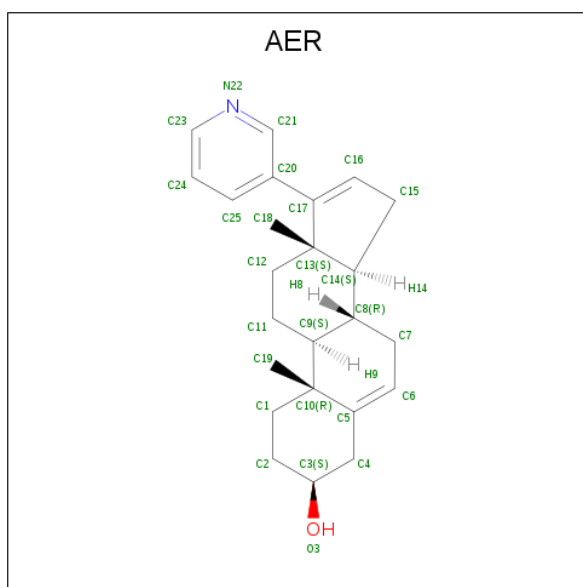
Chain	Residue	Modelled	Actual	Comment	Reference
B	25	LYS	-	EXPRESSION TAG	UNP A7U483
B	496	HIS	-	EXPRESSION TAG	UNP A7U483
B	497	HIS	-	EXPRESSION TAG	UNP A7U483
B	498	HIS	-	EXPRESSION TAG	UNP A7U483
B	499	HIS	-	EXPRESSION TAG	UNP A7U483
B	500	HIS	-	EXPRESSION TAG	UNP A7U483
B	501	HIS	-	EXPRESSION TAG	UNP A7U483

- Molecule 2 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: $C_{34}H_{32}FeN_4O_4$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
2	B	1	Total 43	C 34	Fe 1	N 4	O 4	0	0

- Molecule 3 is Abiraterone (three-letter code: AER) (formula: $C_{24}H_{31}NO$).



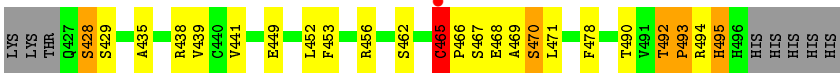
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			26	24	1	1		
3	B	1	Total	C	N	O	0	0
			26	24	1	1		

- Molecule 4 is MERCURY (II) ION (three-letter code: HG) (formula: Hg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	Hg	0	0
			2	2		
4	A	3	Total	Hg	0	0
			3	3		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	3	Total	O	0	0
			3	3		
5	B	11	Total	O	0	0
			11	11		



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	57.01Å 79.77Å 95.24Å 90.00° 92.73° 90.00°	Depositor
Resolution (Å)	42.32 – 2.86 42.32 – 2.86	Depositor EDS
% Data completeness (in resolution range)	99.3 (42.32-2.86) 99.3 (42.32-2.86)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.98 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.223 , 0.305 0.223 , 0.304	Depositor DCC
R_{free} test set	1005 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å ²)	79.5	Xtriage
Anisotropy	0.394	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 68.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.039 for h,-k,-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6555	wwPDB-VP
Average B, all atoms (Å ²)	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.24% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: HEM, AER, HG

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	0.50	1/3239 (0.0%)	0.79	5/4398 (0.1%)
1	B	0.54	1/3299 (0.0%)	0.83	4/4487 (0.1%)
All	All	0.52	2/6538 (0.0%)	0.81	9/8885 (0.1%)

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	A	0	7
1	B	0	3
All	All	0	10

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	465	CYS	CB-SG	6.84	1.93	1.82
1	A	187	CYS	CB-SG	6.55	1.93	1.82

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	267	ARG	NE-CZ-NH1	-18.72	110.94	120.30
1	B	267	ARG	CD-NE-CZ	11.68	139.95	123.60
1	A	438	ARG	NE-CZ-NH2	-11.15	114.72	120.30
1	A	438	ARG	NE-CZ-NH1	8.61	124.60	120.30
1	B	465	CYS	CA-CB-SG	7.18	126.93	114.00

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	145	GLY	Peptide
1	A	147	ASN	Peptide
1	A	149	LEU	Peptide
1	A	202	LEU	Peptide
1	A	236	LEU	Peptide

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	3171	0	3171	133	0
1	B	3227	0	3221	146	0
2	A	43	0	30	6	0
2	B	43	0	30	7	0
3	A	26	0	31	4	0
3	B	26	0	31	4	0
4	A	3	0	0	0	0
4	B	2	0	0	2	0
5	A	3	0	0	0	0
5	B	11	0	0	0	0
All	All	6555	0	6514	291	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 22.

The worst 5 of 291 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:205:VAL:HA	1:A:208:TYR:CE2	1.64	1.33
1:A:135:VAL:HG12	1:A:139:PHE:CE2	1.85	1.12
1:B:149:LEU:HB3	1:B:152:ILE:HG13	1.21	1.11
1:A:187:CYS:SG	1:A:300:PHE:HZ	1.77	1.06
1:B:465:CYS:SG	1:B:469:ALA:HB3	1.94	1.06

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	394/486 (81%)	367 (93%)	19 (5%)	8 (2%)	7	23
1	B	406/486 (84%)	365 (90%)	26 (6%)	15 (4%)	3	11
All	All	800/972 (82%)	732 (92%)	45 (6%)	23 (3%)	4	15

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	148	LYS
1	A	199	ASP
1	A	200	PRO
1	A	419	GLU
1	B	79	LEU

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	354/427 (83%)	331 (94%)	23 (6%)	17	41
1	B	357/427 (84%)	325 (91%)	32 (9%)	9	25
All	All	711/854 (83%)	656 (92%)	55 (8%)	13	32

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	62	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	149	LEU
1	B	452	LEU
1	B	65	SER
1	B	96	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	150	GLN
1	B	405	HIS
1	B	278	GLN
1	B	130	ASN
1	B	154	GLN

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 9 ligands modelled in this entry, 5 are monoatomic - leaving 4 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	HEM	B	601	1,3	27,50,50	1.14	2 (7%)	17,82,82	2.36	5 (29%)
2	HEM	A	601	1,3	27,50,50	1.14	2 (7%)	17,82,82	1.87	3 (17%)
3	AER	A	602	2	30,30,30	2.25	2 (6%)	44,47,47	2.55	16 (36%)
3	AER	B	602	2	30,30,30	2.32	2 (6%)	44,47,47	2.64	14 (31%)

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
2	HEM	B	601	1,3	-	2/6/54/54	-
2	HEM	A	601	1,3	-	0/6/54/54	-
3	AER	A	602	2	-	2/4/62/62	0/5/5/5
3	AER	B	602	2	-	1/4/62/62	0/5/5/5

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	602	AER	C16-C17	11.68	1.54	1.33
3	A	602	AER	C16-C17	11.52	1.54	1.33
2	B	601	HEM	C3B-C2B	-4.02	1.34	1.40
2	A	601	HEM	C3B-C2B	-2.95	1.36	1.40
3	B	602	AER	C13-C17	2.60	1.56	1.53

The worst 5 of 38 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	602	AER	C15-C16-C17	-10.96	103.86	112.87
3	A	602	AER	C15-C16-C17	-8.58	105.82	112.87
3	B	602	AER	C20-C17-C16	-8.55	113.46	125.19
3	A	602	AER	C20-C17-C16	-8.30	113.79	125.19
2	B	601	HEM	CBA-CAA-C2A	5.51	122.64	112.49

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	601	HEM	C1A-C2A-CAA-CBA
2	B	601	HEM	C3A-C2A-CAA-CBA

Continued on next page...

Continued from previous page...

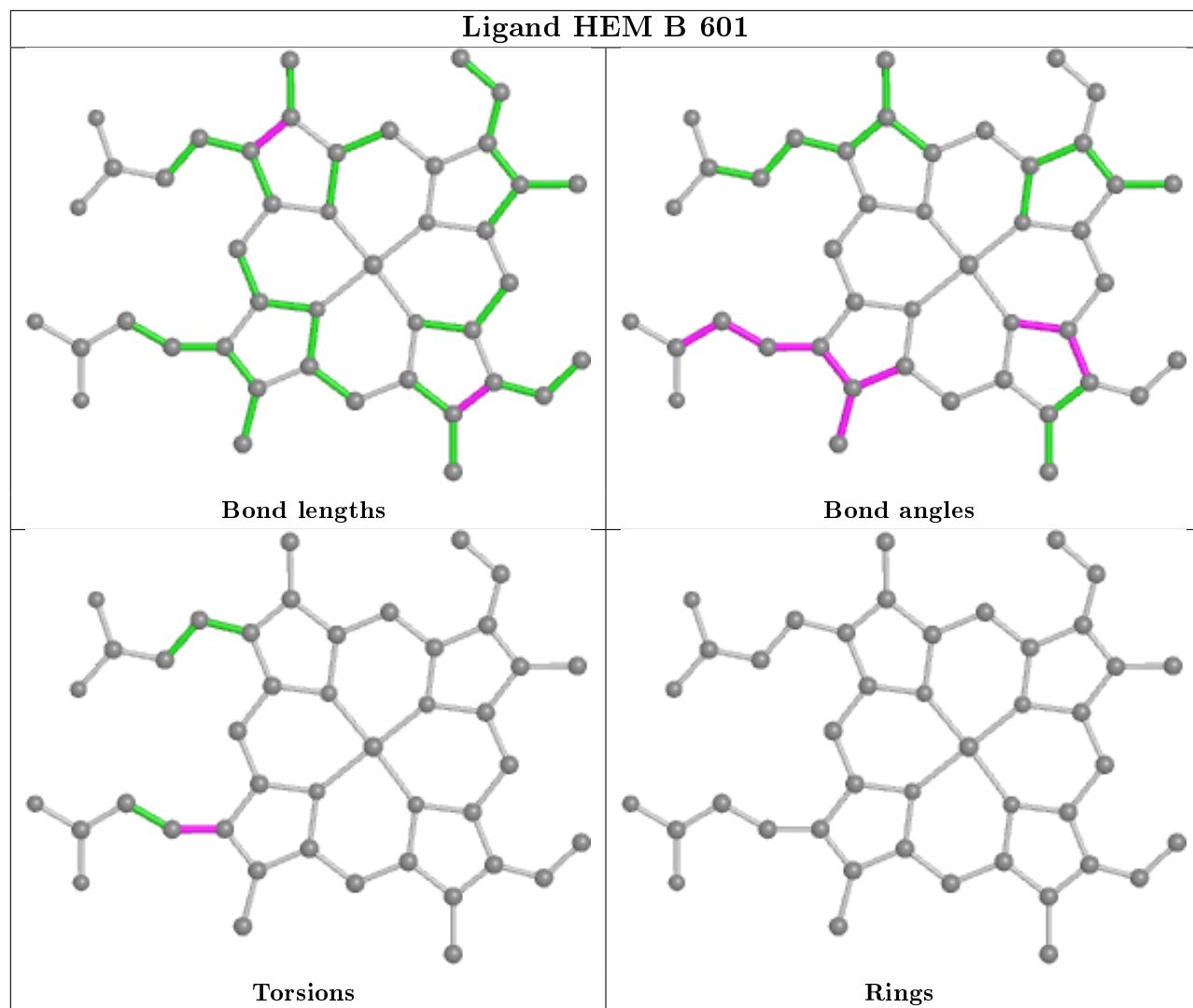
Mol	Chain	Res	Type	Atoms
3	A	602	AER	C13-C17-C20-C25
3	B	602	AER	C13-C17-C20-C25
3	A	602	AER	C13-C17-C20-C21

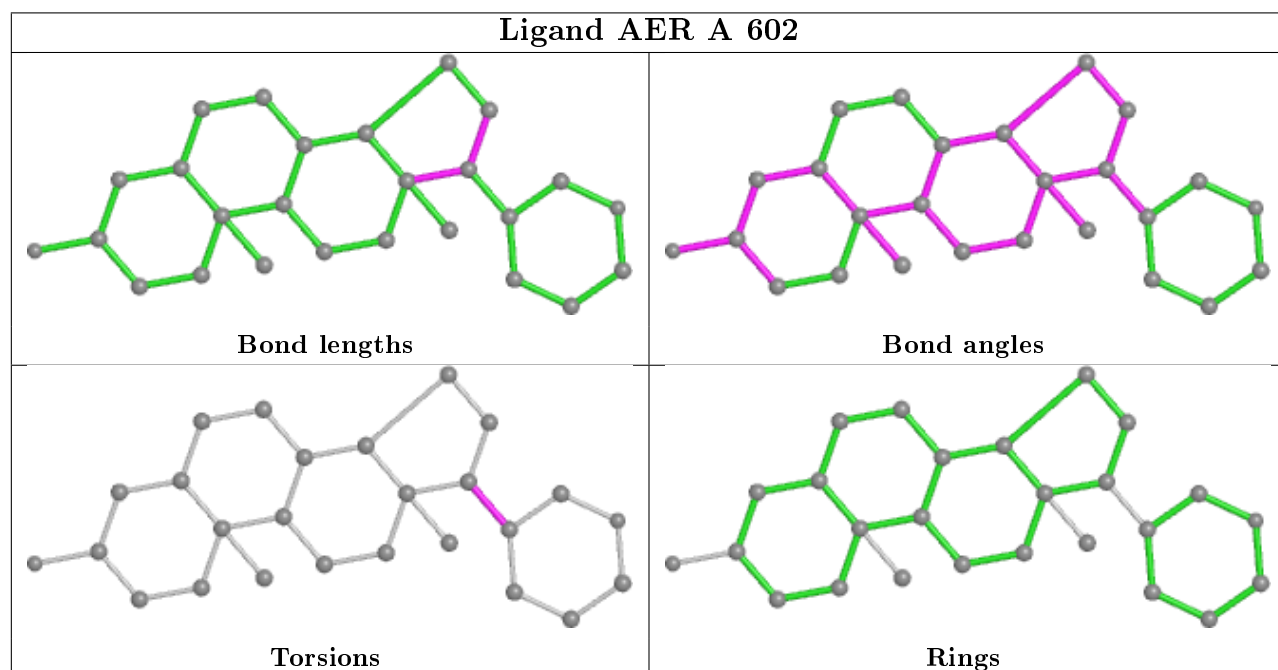
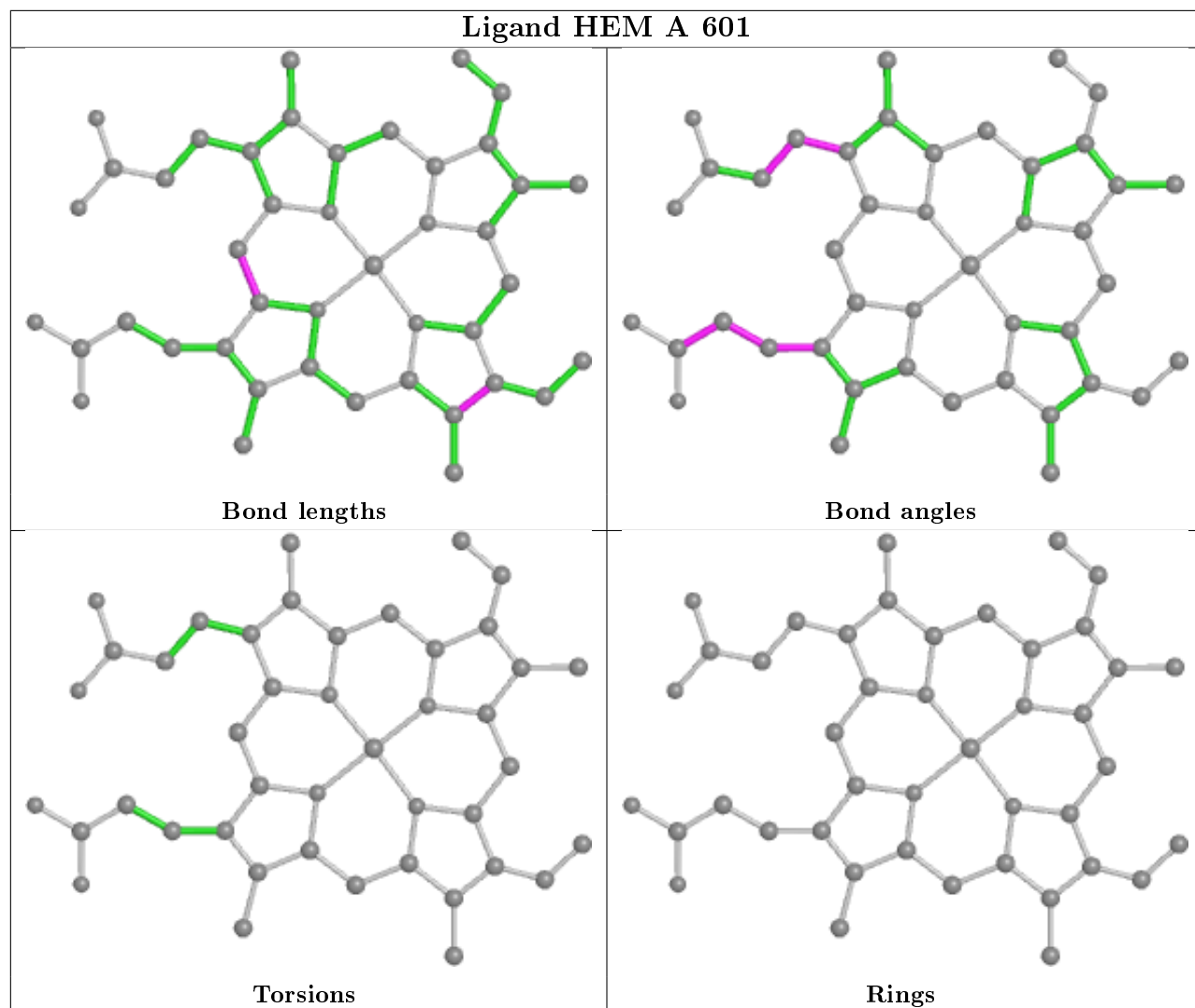
There are no ring outliers.

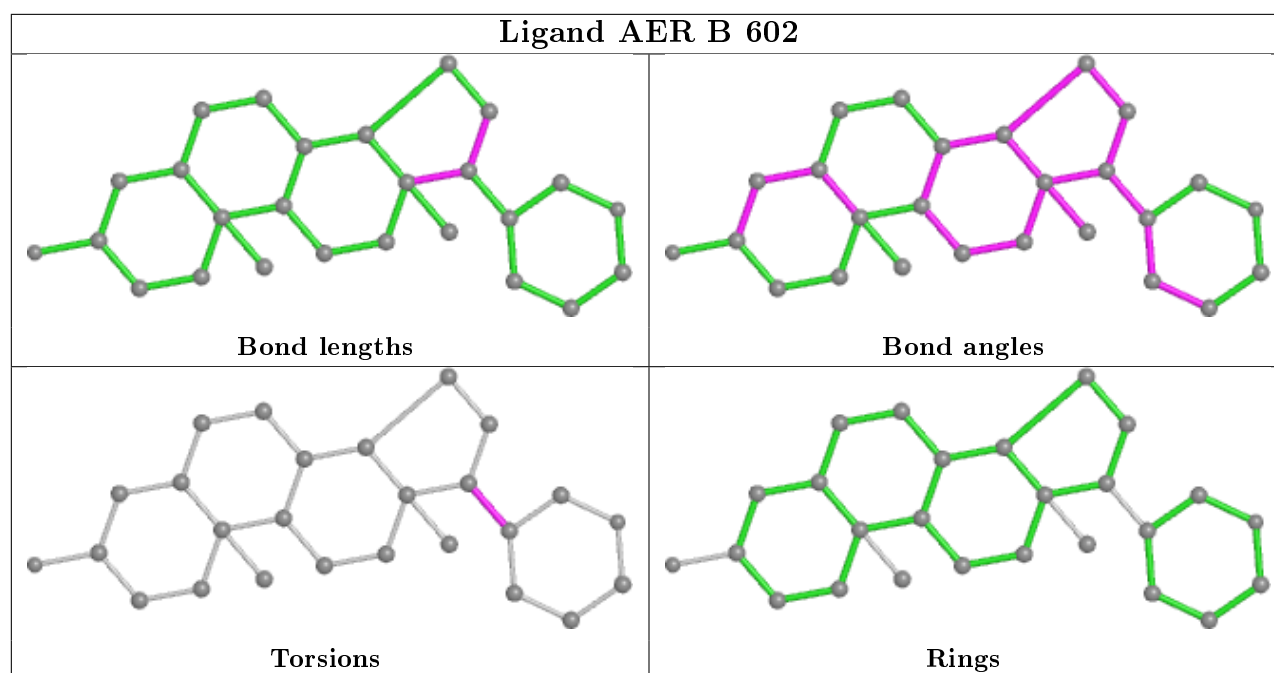
4 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	601	HEM	7	0
2	A	601	HEM	6	0
3	A	602	AER	4	0
3	B	602	AER	4	0

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 than 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, 95th 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	406/486 (83%)	0.16	21 (5%)	27 22	47, 90, 133, 160	4 (0%)
1	B	414/486 (85%)	0.03	10 (2%)	59 56	48, 82, 125, 161	1 (0%)
All	All	820/972 (84%)	0.09	31 (3%)	40 35	47, 86, 132, 161	5 (0%)

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	282	GLY	6.3
1	A	165	GLN	4.3
1	B	165	GLN	4.3
1	A	275	ILE	4.3
1	A	276	GLY	4.0

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](#)

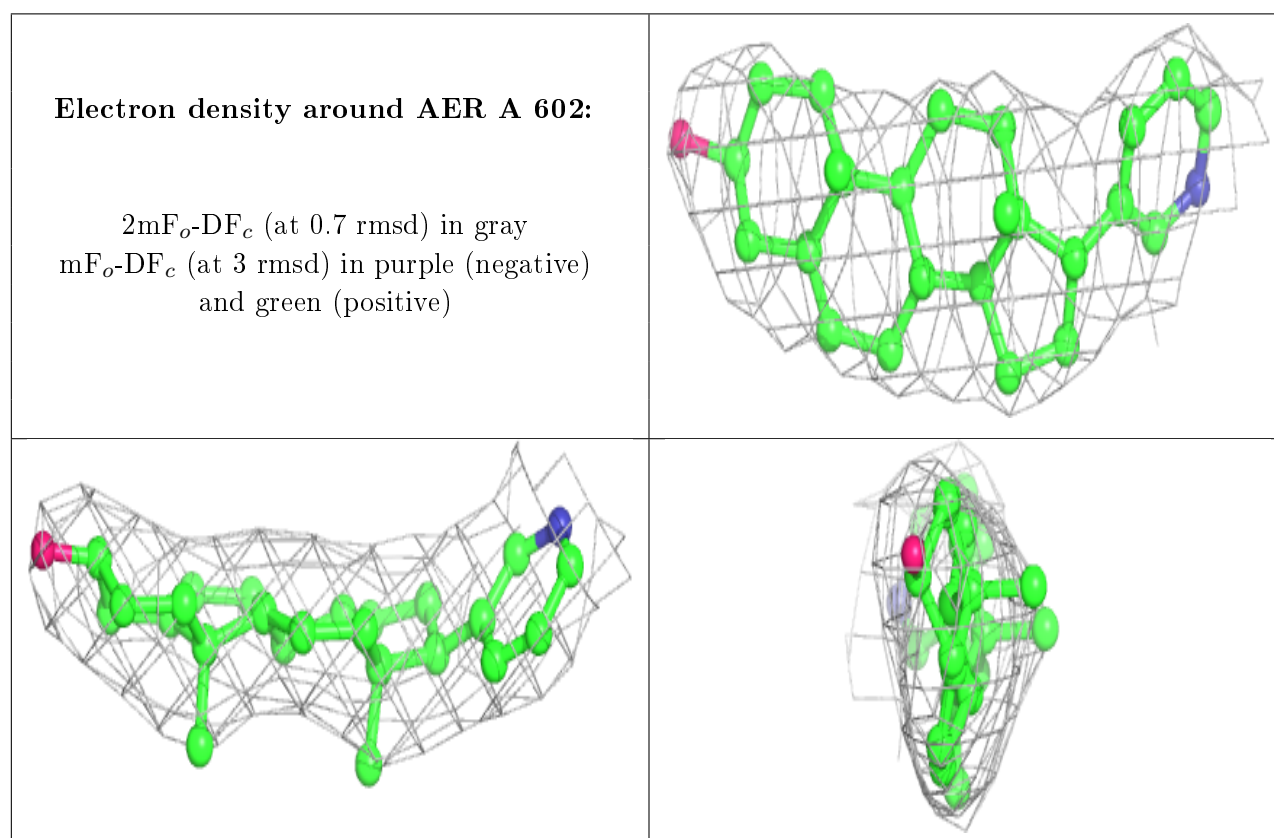
There are no carbohydrates in this entry.

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, 95th 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.

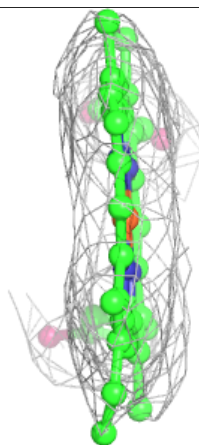
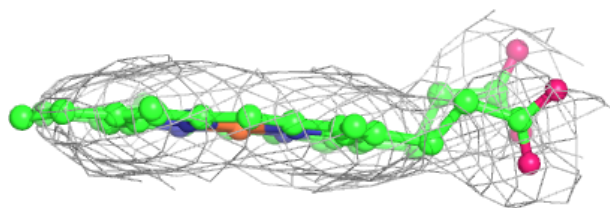
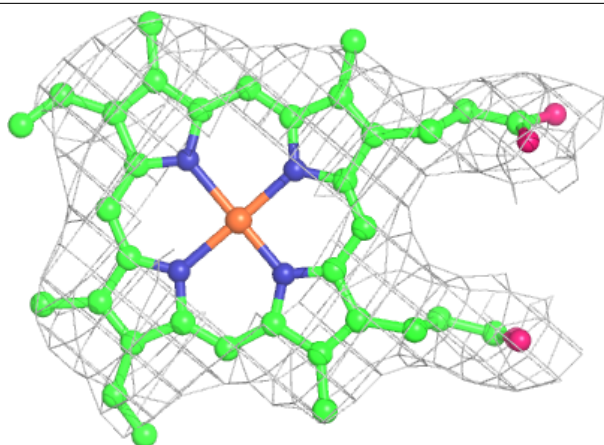
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	HG	B	603	1/1	0.82	0.11	152,152,152,152	1
4	HG	A	604	1/1	0.96	0.04	116,116,116,116	1
3	AER	A	602	26/26	0.97	0.19	55,66,72,75	0
4	HG	A	605	1/1	0.97	0.18	88,88,88,88	1
2	HEM	A	601	43/43	0.98	0.17	44,50,71,79	0
4	HG	B	604	1/1	0.98	0.09	114,114,114,114	1
3	AER	B	602	26/26	0.98	0.21	54,60,69,74	0
4	HG	A	603	1/1	0.98	0.03	126,126,126,126	1
2	HEM	B	601	43/43	0.99	0.17	38,44,53,58	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.

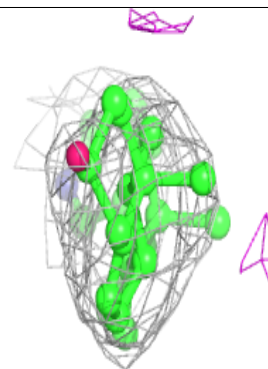
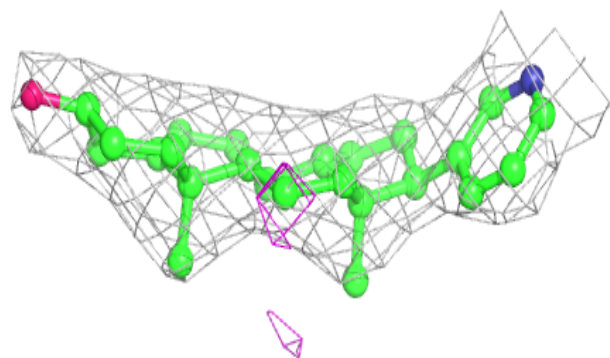
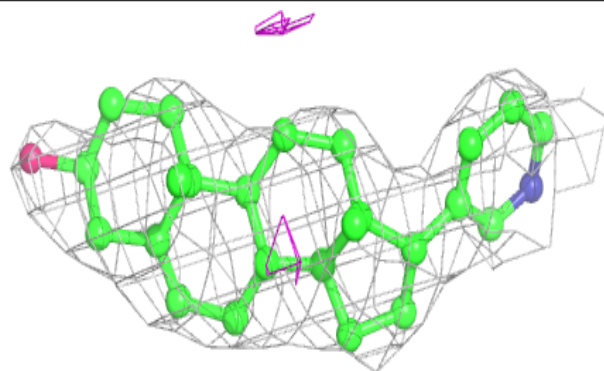


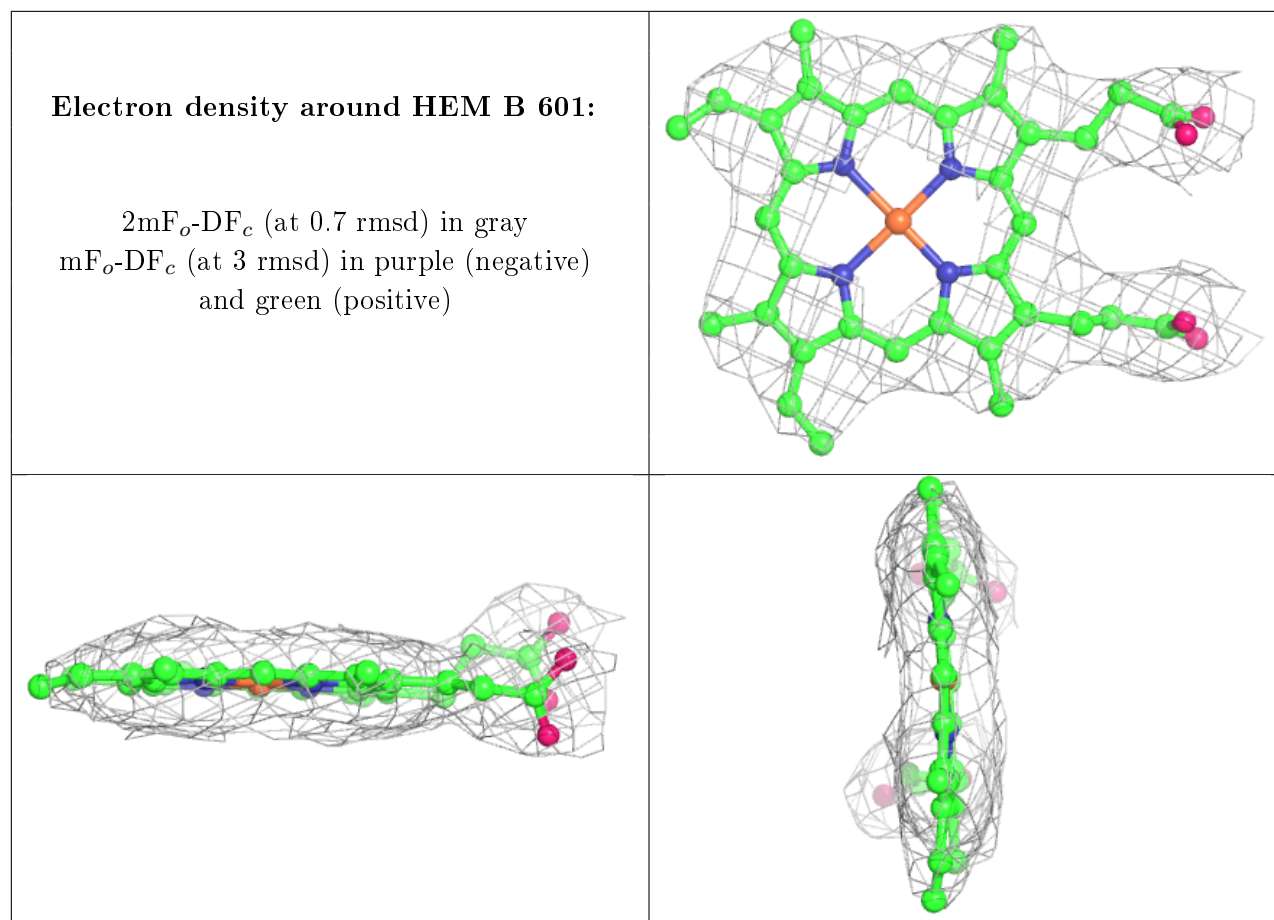
Electron density around HEM A 601:

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

**Electron density around AER B 602:**

$2mF_o-DF_c$ (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.