Full wwPDB X-ray Structure Validation Report (i)

Oct 23, 2021 – 09:00 AM EDT

PDB ID	:	1AEK
Title	:	SPECIFICITY OF LIGAND BINDING TO A BURIED POLAR CAVITY AT THE ACTIVE SITE OF CYTOCHROME C PEROXIDASE (INDOLINE)
Authors	:	Musah, R.A.; Jensen, G.M.; Fitzgerald, M.M.; Mcree, D.E.; Goodin, D.B.
Deposited on	:	1997-02-25
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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

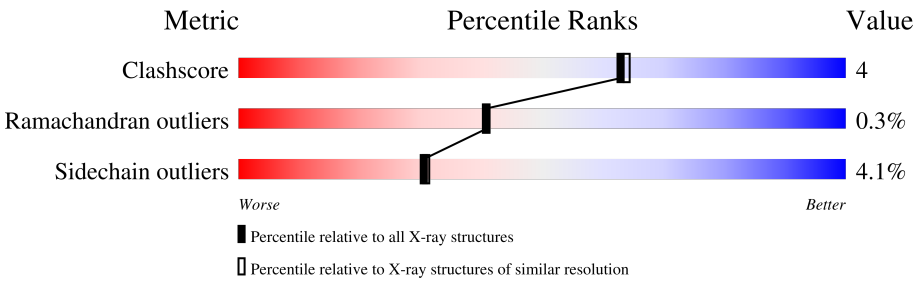
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.10 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))
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (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 >=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%

Mol	Chain	Length	Quality of chain
1	A	294	83% 13% • •

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
3	IDM	A	296	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2993 atoms, of which 503 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.

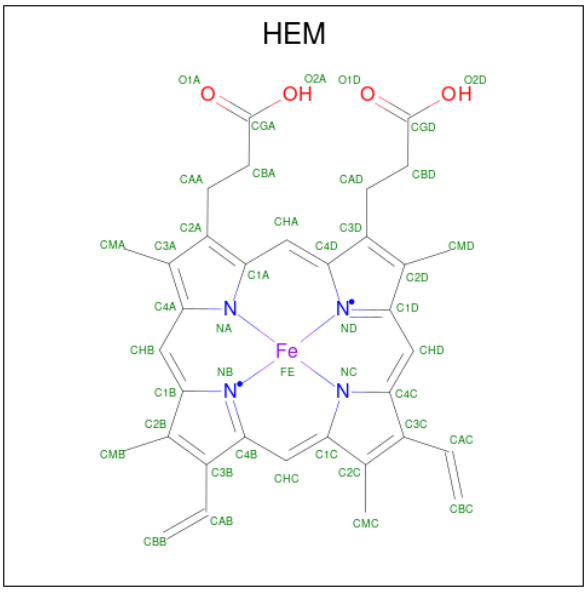
- Molecule 1 is a protein called CYTOCHROME C PEROXIDASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	291	2841	1492	503	388	452	6	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

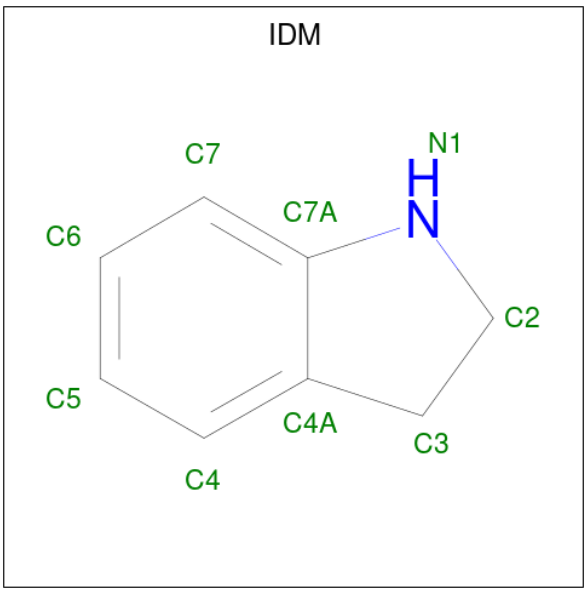
Chain	Residue	Modelled	Actual	Comment	Reference
A	53	ILE	THR	variant	UNP P00431
A	152	GLY	ASP	variant	UNP P00431
A	191	GLY	TRP	engineered mutation	UNP P00431
A	272	ASP	ASN	conflict	UNP P00431

- 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
			Total	C	Fe	N	O		
2	A	1	43	34	1	4	4	0	0

- Molecule 3 is INDOLINE (three-letter code: IDM) (formula: C_8H_9N).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	N		
3	A	1	9	8	1	0	0

- Molecule 4 is water.

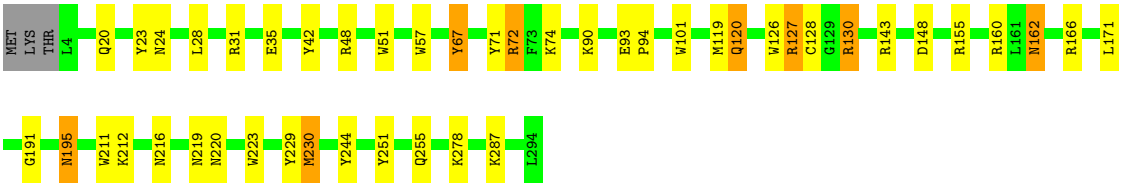
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	100	100	100	0	0

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. 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. 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: CYTOCHROME C PEROXIDASE

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	108.00Å 77.30Å 51.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	7.00 – 2.10 46.71 – 2.01	Depositor EDS
% Data completeness (in resolution range)	79.0 (7.00-2.10) 55.2 (46.71-2.01)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.92 (at 2.01Å)	Xtriage
Refinement program	XTALVIEW	Depositor
R, R_{free}	(Not available) , (Not available) 0.208 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	21.2	Xtriage
Anisotropy	0.559	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 60.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	2993	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.28% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: IDM, HEM

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	0.79	0/2402	1.57	50/3250 (1.5%)

There are no bond length outliers.

All (50) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	166	ARG	NE-CZ-NH2	-11.23	114.68	120.30
1	A	48	ARG	NE-CZ-NH1	9.71	125.16	120.30
1	A	31	ARG	NE-CZ-NH2	-9.16	115.72	120.30
1	A	101	TRP	CD1-CG-CD2	8.47	113.08	106.30
1	A	126	TRP	CD1-CG-CD2	8.22	112.88	106.30
1	A	127	ARG	NE-CZ-NH2	-8.03	116.29	120.30
1	A	57	TRP	CD1-CG-CD2	7.86	112.59	106.30
1	A	127	ARG	NE-CZ-NH1	7.67	124.14	120.30
1	A	223	TRP	CD1-CG-CD2	7.63	112.40	106.30
1	A	126	TRP	CE2-CD2-CG	-7.62	101.21	107.30
1	A	71	TYR	CB-CG-CD2	-7.57	116.45	121.00
1	A	51	TRP	CE2-CD2-CG	-7.44	101.35	107.30
1	A	127	ARG	CB-CG-CD	-7.37	92.44	111.60
1	A	223	TRP	CE2-CD2-CG	-7.29	101.46	107.30
1	A	51	TRP	CD1-CG-CD2	7.23	112.08	106.30
1	A	51	TRP	CB-CG-CD1	-7.22	117.61	127.00
1	A	101	TRP	CE2-CD2-CG	-7.21	101.53	107.30
1	A	162	ASN	CA-C-N	-6.99	101.83	117.20
1	A	51	TRP	CG-CD2-CE3	6.76	139.99	133.90
1	A	57	TRP	CE2-CD2-CG	-6.75	101.90	107.30
1	A	130	ARG	NE-CZ-NH2	-6.62	116.99	120.30
1	A	31	ARG	NE-CZ-NH1	6.57	123.58	120.30
1	A	211	TRP	CD1-CG-CD2	6.50	111.50	106.30
1	A	211	TRP	CE2-CD2-CG	-6.49	102.11	107.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	162	ASN	N-CA-C	6.32	128.06	111.00
1	A	229	TYR	CB-CG-CD2	-6.26	117.24	121.00
1	A	72	ARG	NE-CZ-NH1	6.17	123.39	120.30
1	A	130	ARG	CG-CD-NE	-6.03	99.13	111.80
1	A	101	TRP	CG-CD1-NE1	-5.79	104.31	110.10
1	A	126	TRP	CG-CD2-CE3	5.77	139.09	133.90
1	A	130	ARG	NE-CZ-NH1	5.67	123.14	120.30
1	A	48	ARG	NE-CZ-NH2	-5.58	117.51	120.30
1	A	143	ARG	NE-CZ-NH2	-5.56	117.52	120.30
1	A	155	ARG	NE-CZ-NH1	5.50	123.05	120.30
1	A	126	TRP	CG-CD1-NE1	-5.45	104.65	110.10
1	A	57	TRP	CG-CD1-NE1	-5.42	104.68	110.10
1	A	42	TYR	CB-CG-CD1	-5.34	117.80	121.00
1	A	90	LYS	CA-CB-CG	-5.32	101.70	113.40
1	A	244	TYR	CB-CG-CD2	-5.32	117.81	121.00
1	A	160	ARG	NE-CZ-NH1	5.30	122.95	120.30
1	A	255	GLN	CG-CD-NE2	5.24	129.28	116.70
1	A	166	ARG	CG-CD-NE	-5.23	100.82	111.80
1	A	223	TRP	CG-CD2-CE3	5.14	138.53	133.90
1	A	251	TYR	CB-CG-CD2	-5.14	117.92	121.00
1	A	148	ASP	CB-CG-OD1	5.08	122.88	118.30
1	A	160	ARG	NE-CZ-NH2	-5.08	117.76	120.30
1	A	255	GLN	OE1-CD-NE2	-5.08	110.22	121.90
1	A	67	TYR	CB-CG-CD1	-5.04	117.97	121.00
1	A	162	ASN	CA-C-O	5.03	130.66	120.10
1	A	230	MET	CA-CB-CG	5.02	121.84	113.30

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	2338	503	2209	19	0
2	A	43	0	30	1	0
3	A	9	0	9	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	100	0	0	4	0
All	All	2490	503	2248	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 4.

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:128:CYS:HA	4:A:326:HOH:O	1.62	0.98
1:A:191:GLY:HA3	3:A:296:IDM:C3	2.18	0.73
1:A:191:GLY:HA3	3:A:296:IDM:H31	1.72	0.71
1:A:130:ARG:NE	4:A:442:HOH:O	2.37	0.57
1:A:20:GLN:HE22	1:A:287:LYS:H	1.51	0.57
1:A:191:GLY:CA	3:A:296:IDM:C3	2.85	0.54
1:A:195:ASN:HD22	1:A:195:ASN:H	1.54	0.54
1:A:130:ARG:CZ	4:A:442:HOH:O	2.56	0.53
1:A:67:TYR:HA	1:A:130:ARG:HG2	1.90	0.53
1:A:230:MET:O	3:A:296:IDM:H22	2.11	0.51
1:A:119:MET:O	1:A:120:GLN:HG2	2.11	0.49
1:A:130:ARG:NH2	4:A:442:HOH:O	2.45	0.49
1:A:191:GLY:CA	3:A:296:IDM:H32	2.42	0.49
1:A:93:GLU:HB3	1:A:94:PRO:HD2	1.96	0.46
1:A:195:ASN:H	1:A:195:ASN:ND2	2.14	0.44
2:A:295:HEM:C1A	3:A:296:IDM:H6	2.52	0.44
1:A:191:GLY:N	3:A:296:IDM:H32	2.33	0.43
1:A:216:ASN:HD21	1:A:220:ASN:HB2	1.83	0.43
1:A:23:TYR:HD1	1:A:24:ASN:HD22	1.66	0.43
1:A:127:ARG:HH11	1:A:127:ARG:HD3	1.77	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	289/294 (98%)	283 (98%)	5 (2%)	1 (0%)	41	41

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	162	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	246/251 (98%)	236 (96%)	10 (4%)	30	31

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	28	LEU
1	A	35	GLU
1	A	72	ARG
1	A	74	LYS
1	A	120	GLN
1	A	171	LEU
1	A	195	ASN
1	A	212	LYS
1	A	219	ASN
1	A	278	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	20	GLN
1	A	24	ASN
1	A	78	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	195	ASN
1	A	220	ASN
1	A	240	GLN
1	A	292	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

2 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
3	IDM	A	296	-	9,10,10	2.33	2 (22%)	10,13,13	1.54	2 (20%)
2	HEM	A	295	1,4	27,50,50	1.45	4 (14%)	17,82,82	1.72	3 (17%)

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
3	IDM	A	296	-	-	-	0/2/2/2
2	HEM	A	295	1,4	-	0/6/54/54	-

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	296	IDM	C7A-N1	5.61	1.49	1.39
3	A	296	IDM	C2-N1	3.91	1.54	1.46
2	A	295	HEM	C3B-CAB	-3.71	1.40	1.47
2	A	295	HEM	C3C-CAC	-3.11	1.41	1.47
2	A	295	HEM	CBC-CAC	2.63	1.46	1.29
2	A	295	HEM	CBB-CAB	2.52	1.45	1.29

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	295	HEM	CMB-C2B-C3B	4.19	132.51	124.68
2	A	295	HEM	CMC-C2C-C3C	3.44	131.10	124.68
3	A	296	IDM	C4-C4A-C7A	-3.05	118.15	119.67
2	A	295	HEM	C1D-C2D-C3D	-2.72	105.10	107.00
3	A	296	IDM	C7-C7A-C4A	2.13	124.88	122.19

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

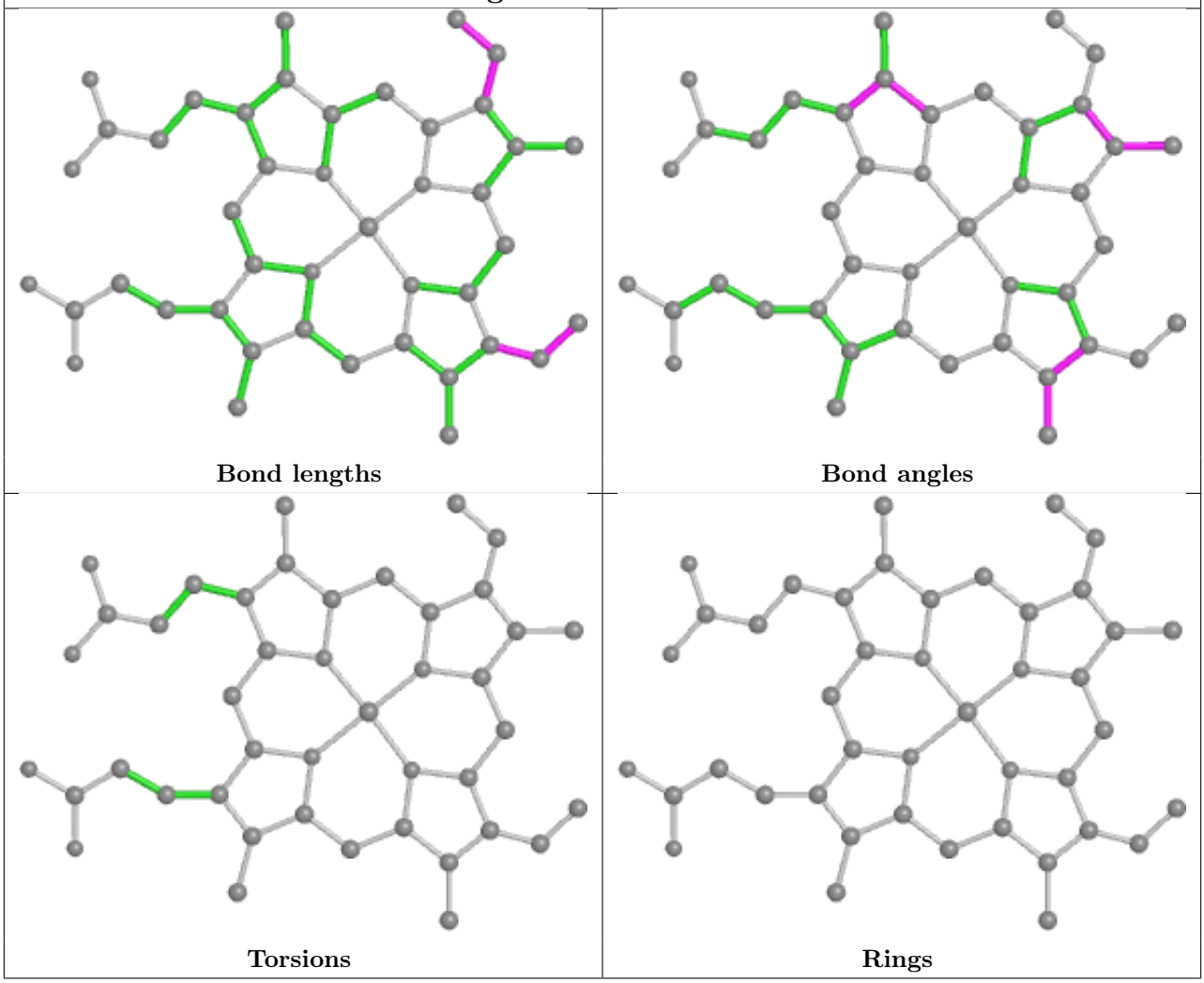
2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	296	IDM	7	0
2	A	295	HEM	1	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 then 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.

Ligand HEM A 295



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)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

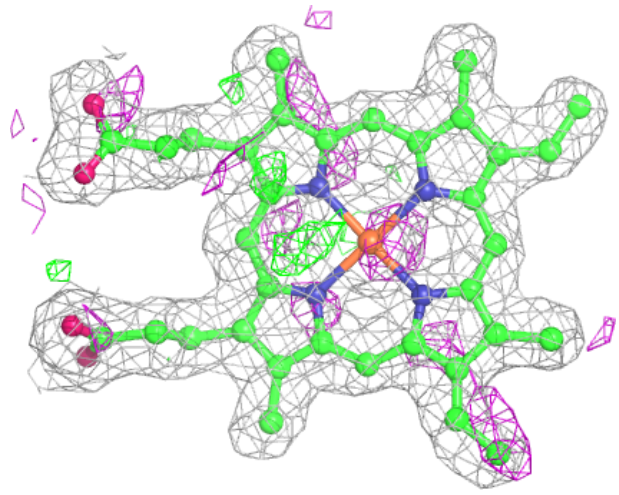
6.4 Ligands (i)

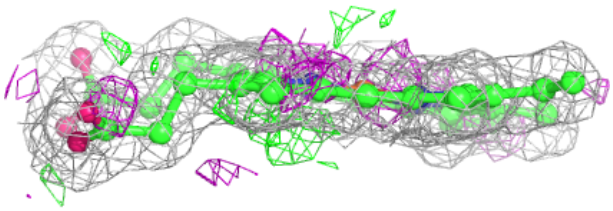
Unable to reproduce the depositors R factor - this section is therefore empty.

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 295:

$2mF_o$-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)

Unable to reproduce the depositors R factor - this section is therefore empty.