Full wwPDB X-ray Structure Validation Report (i)

Feb 26, 2025 – 01:48 AM JST

PDB ID	:	8YHN
Title	:	Crystal structure of Cytochrome P450 107P2 from streptomyces avermitilis
Authors	:	Jeong, E.S.; Kim, V.C.; Kim, C.M.; Lee, Y.B.
Deposited on	:	2024-02-28
Resolution	:	1.99 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:

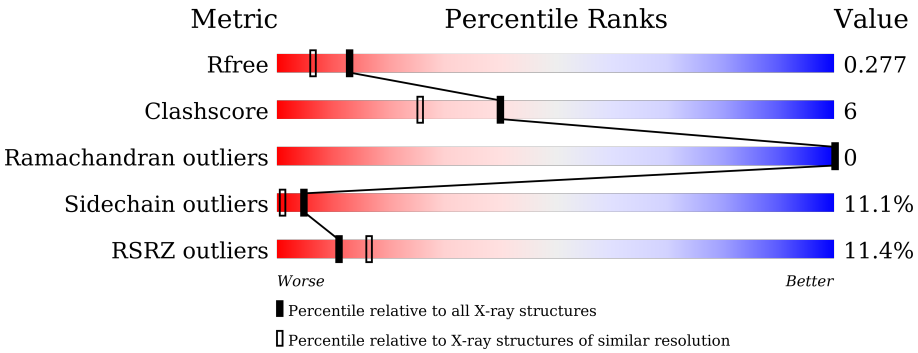
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.41.2

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 1.99 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	1356 (1.98-1.98)
Clashscore	180529	1437 (1.98-1.98)
Ramachandran outliers	177936	1426 (1.98-1.98)
Sidechain outliers	177891	1426 (1.98-1.98)
RSRZ outliers	164620	1356 (1.98-1.98)

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	411	11% 79% 15%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3084 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.

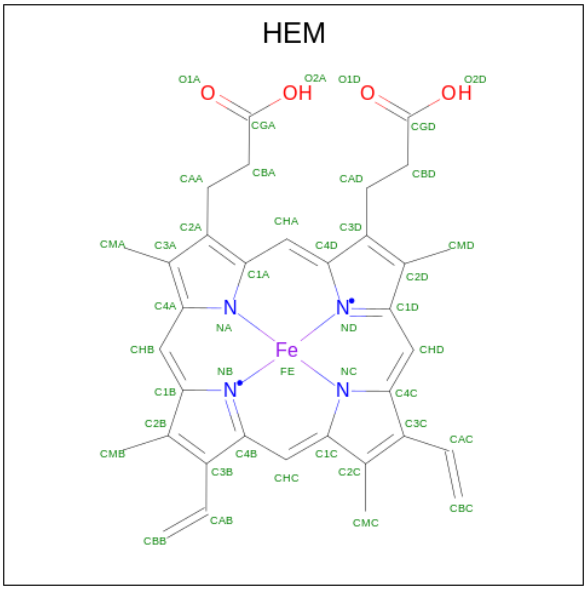
- Molecule 1 is a protein called Cytochrome P450.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	396	2984	1895	516	563	10	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

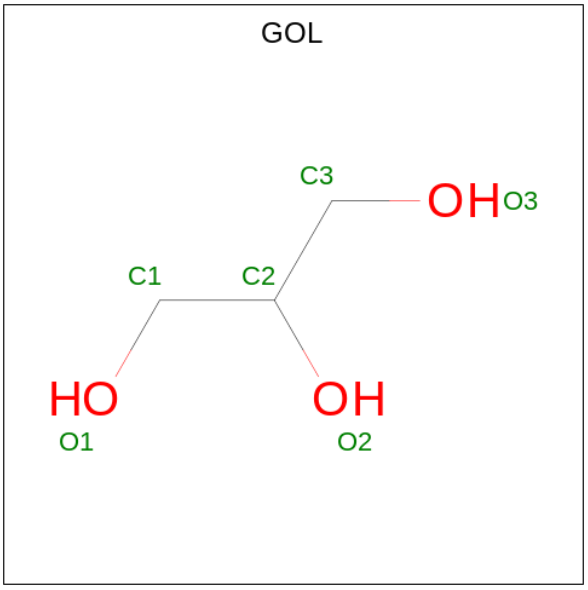
Chain	Residue	Modelled	Actual	Comment	Reference
A	406	HIS	-	expression tag	UNP A0A4D4MSR2
A	407	HIS	-	expression tag	UNP A0A4D4MSR2
A	408	HIS	-	expression tag	UNP A0A4D4MSR2
A	409	HIS	-	expression tag	UNP A0A4D4MSR2
A	410	HIS	-	expression tag	UNP A0A4D4MSR2
A	411	HIS	-	expression tag	UNP A0A4D4MSR2

- 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	C	Fe	N	O	0	0
			43	34	1	4	4		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

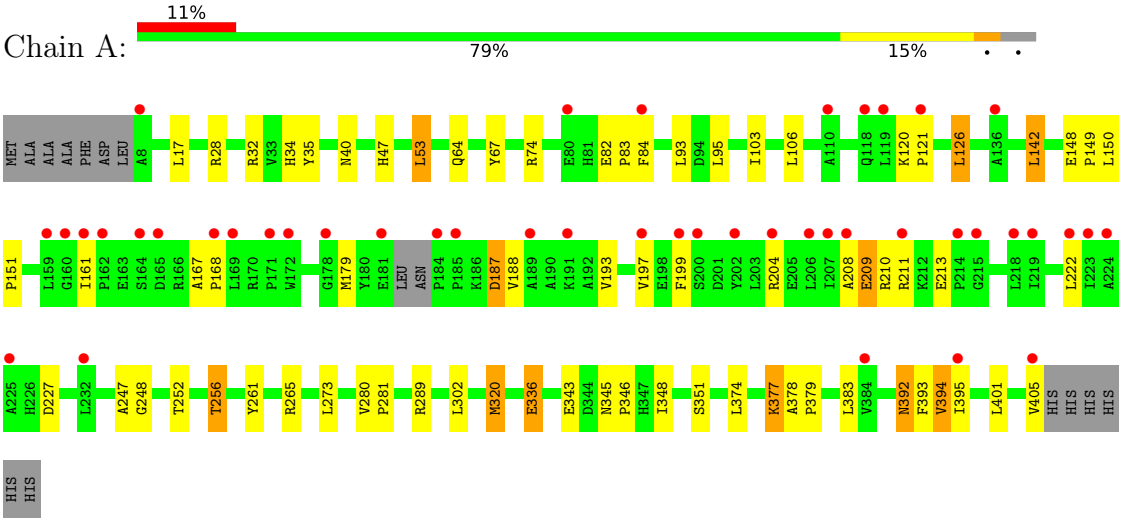
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	51	Total	O	0	0
			51	51		

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: Cytochrome P450



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	59.13Å 66.44Å 111.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.21 – 1.99 44.21 – 1.99	Depositor EDS
% Data completeness (in resolution range)	98.5 (44.21-1.99) 98.5 (44.21-1.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.20 (at 1.98Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.223 , 0.272 0.229 , 0.277	Depositor DCC
R_{free} test set	1493 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å^2)	29.6	Xtriage
Anisotropy	0.028	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 41.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3084	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.61% 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: HEM, GOL

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	0.68	0/3063	0.79	1/4199 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	28	ARG	NE-CZ-NH2	-5.66	117.47	120.30

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	2984	0	2812	33	0
2	A	43	0	30	7	0
3	A	6	0	8	3	0
4	A	51	0	0	2	1
All	All	3084	0	2850	38	1

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 6.

All (38) 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:209:GLU:O	1:A:213:GLU:HB3	1.99	0.62
1:A:252:THR:O	1:A:256:THR:HG22	2.01	0.60
1:A:261:TYR:CZ	1:A:265:ARG:HD2	2.38	0.58
1:A:35:TYR:OH	1:A:40:ASN:ND2	2.34	0.57
1:A:187:ASP:OD1	1:A:187:ASP:N	2.34	0.56
2:A:501:HEM:HMC2	2:A:501:HEM:HBC2	1.90	0.53
2:A:501:HEM:C1A	3:A:502:GOL:H11	2.43	0.53
2:A:501:HEM:NA	3:A:502:GOL:H11	2.24	0.52
1:A:392:ASN:ND2	1:A:394:VAL:H	2.09	0.51
1:A:148:GLU:N	1:A:149:PRO:HD2	2.26	0.51
1:A:208:ALA:O	1:A:211:ARG:N	2.42	0.51
1:A:150:LEU:HB3	1:A:151:PRO:HD3	1.93	0.51
2:A:501:HEM:HMB1	2:A:501:HEM:HBB2	1.91	0.51
2:A:501:HEM:HBA2	2:A:501:HEM:HMA1	1.93	0.50
1:A:392:ASN:HD22	1:A:393:PHE:N	2.10	0.49
1:A:392:ASN:HD22	1:A:394:VAL:H	1.60	0.49
1:A:289:ARG:HD3	1:A:345:ASN:ND2	2.28	0.49
1:A:53:LEU:HD23	1:A:320:MET:HE1	1.95	0.49
1:A:126:LEU:HD13	1:A:150:LEU:HD12	1.95	0.48
1:A:167:ALA:N	1:A:168:PRO:CD	2.78	0.47
1:A:84:PHE:CE2	1:A:179:MET:HG2	2.50	0.47
1:A:248:GLY:HA2	2:A:501:HEM:C2C	2.51	0.46
1:A:351:SER:HB2	2:A:501:HEM:HMA3	1.98	0.45
1:A:47:HIS:HE1	1:A:346:PRO:O	1.99	0.45
1:A:34:HIS:HD2	4:A:649:HOH:O	1.99	0.45
1:A:377:LYS:C	1:A:379:PRO:HD3	2.37	0.44
1:A:247:ALA:HB1	3:A:502:GOL:H12	2.00	0.44
1:A:401:LEU:HD12	1:A:401:LEU:HA	1.93	0.43
1:A:120:LYS:N	1:A:121:PRO:HD2	2.33	0.43
1:A:64:GLN:HA	1:A:67:TYR:O	2.19	0.42
1:A:167:ALA:N	1:A:168:PRO:HD2	2.34	0.42
1:A:142:LEU:HD23	1:A:142:LEU:HA	1.91	0.42
1:A:280:VAL:HB	1:A:281:PRO:HD3	2.02	0.42
1:A:193:VAL:O	1:A:197:VAL:HG23	2.20	0.41
1:A:336:GLU:CD	1:A:336:GLU:H	2.24	0.41
1:A:82:GLU:N	1:A:83:PRO:CD	2.84	0.41
1:A:378:ALA:N	1:A:379:PRO:HD3	2.36	0.41
1:A:47:HIS:HD2	4:A:646:HOH:O	2.03	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:A:602:HOH:O	4:A:603:HOH:O[3_554]	1.38	0.82

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	392/411 (95%)	381 (97%)	11 (3%)	0	100 100

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	296/335 (88%)	263 (89%)	33 (11%)	5 1

All (33) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	17	LEU
1	A	32	ARG
1	A	53	LEU
1	A	74	ARG
1	A	93	LEU
1	A	95	LEU
1	A	103	ILE
1	A	106	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	126	LEU
1	A	142	LEU
1	A	161	ILE
1	A	187	ASP
1	A	188	VAL
1	A	199	PHE
1	A	204	ARG
1	A	209	GLU
1	A	210	ARG
1	A	222	LEU
1	A	227	ASP
1	A	256	THR
1	A	273	LEU
1	A	302	LEU
1	A	320	MET
1	A	336	GLU
1	A	343	GLU
1	A	348	ILE
1	A	374	LEU
1	A	377	LYS
1	A	383	LEU
1	A	392	ASN
1	A	394	VAL
1	A	395	ILE
1	A	405	VAL

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	40	ASN
1	A	46	HIS
1	A	47	HIS
1	A	90	HIS
1	A	226	HIS
1	A	254	ASN
1	A	345	ASN
1	A	392	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)

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	$\#\|Z\| > 2$
2	HEM	A	501	1,3	41,50,50	1.69	7 (17%)	45,82,82	2.11	15 (33%)
3	GOL	A	502	2	5,5,5	0.09	0	5,5,5	0.40	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
2	HEM	A	501	1,3	-	0/12/54/54	-
3	GOL	A	502	2	-	4/4/4/4	-

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	501	HEM	C1B-NB	-4.98	1.31	1.40
2	A	501	HEM	C4B-NB	-4.23	1.30	1.38
2	A	501	HEM	C4D-ND	-3.81	1.33	1.40
2	A	501	HEM	C4D-C3D	3.24	1.50	1.45
2	A	501	HEM	FE-NB	2.82	2.10	1.96
2	A	501	HEM	C1D-ND	-2.38	1.33	1.38

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	501	HEM	O2A-CGA	-2.24	1.23	1.30

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	501	HEM	CHC-C4B-NB	5.89	130.83	124.43
2	A	501	HEM	C1B-NB-C4B	5.58	110.84	105.07
2	A	501	HEM	O2A-CGA-O1A	-4.07	113.15	123.30
2	A	501	HEM	CHD-C1D-C2D	-3.32	119.79	124.98
2	A	501	HEM	C3B-C2B-C1B	-2.96	104.29	106.49
2	A	501	HEM	C2D-C1D-ND	2.96	113.42	109.88
2	A	501	HEM	CHA-C4D-ND	2.79	127.83	124.38
2	A	501	HEM	C1D-C2D-C3D	-2.70	104.12	106.96
2	A	501	HEM	C2C-C3C-C4C	2.49	108.64	106.90
2	A	501	HEM	CHA-C4D-C3D	-2.37	120.88	125.33
2	A	501	HEM	O2A-CGA-CBA	2.34	121.54	114.03
2	A	501	HEM	CBA-CAA-C2A	2.23	116.43	112.62
2	A	501	HEM	CMA-C3A-C4A	-2.15	125.15	128.46
2	A	501	HEM	CHD-C1D-ND	2.08	126.69	124.43
2	A	501	HEM	CMB-C2B-C1B	2.04	128.14	125.04

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	502	GOL	O1-C1-C2-C3
3	A	502	GOL	C1-C2-C3-O3
3	A	502	GOL	O1-C1-C2-O2
3	A	502	GOL	O2-C2-C3-O3

There are no ring outliers.

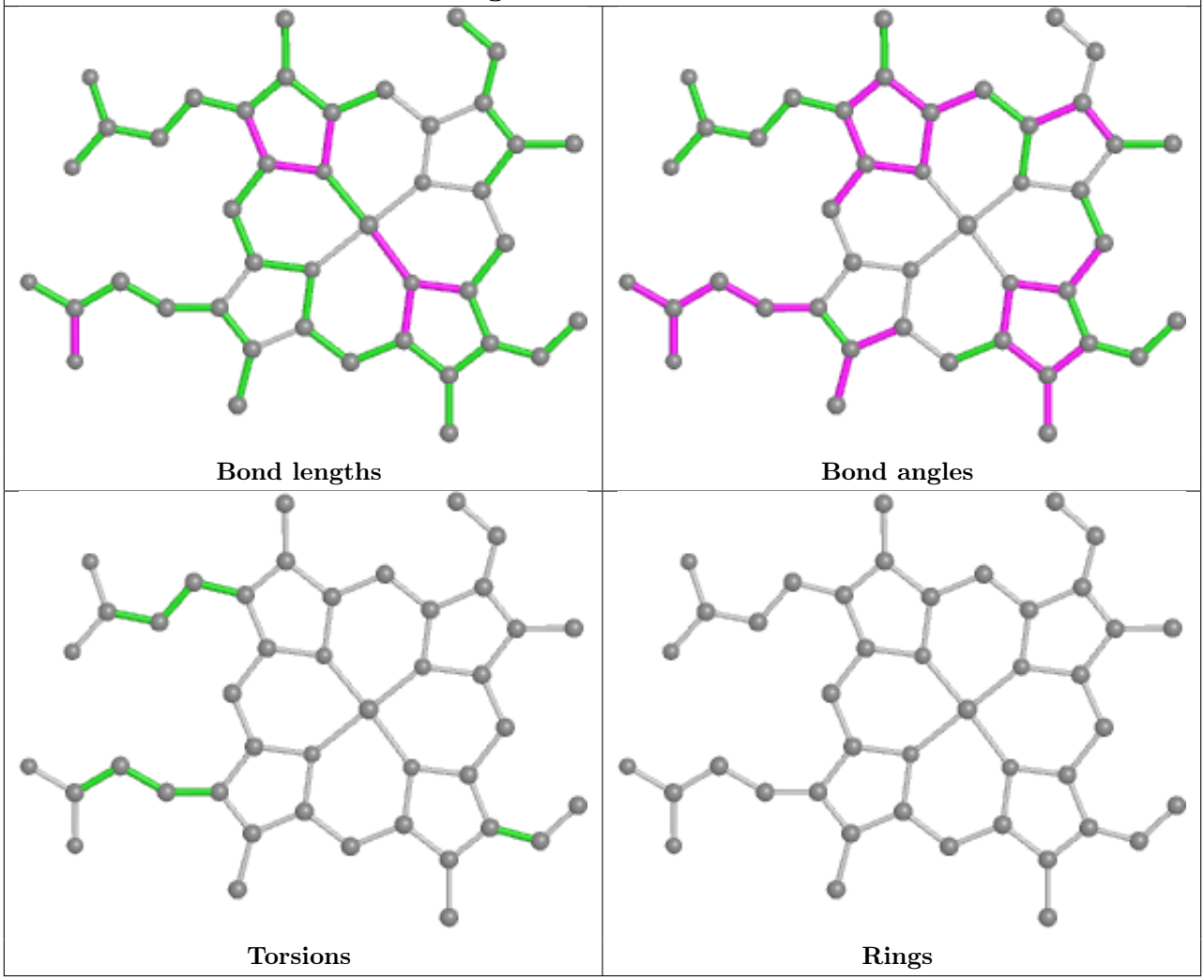
2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	HEM	7	0
3	A	502	GOL	3	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 501



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^{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	396/411 (96%)	0.70	45 (11%)	11	16	19, 39, 73, 107	0

All (45) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	184	PRO	5.3
1	A	165	ASP	4.5
1	A	405	VAL	4.4
1	A	161	ILE	4.2
1	A	160	GLY	3.9
1	A	185	PRO	3.9
1	A	215	GLY	3.7
1	A	222	LEU	3.4
1	A	202	TYR	3.4
1	A	162	PRO	3.4
1	A	206	LEU	3.4
1	A	223	ILE	3.3
1	A	207	ILE	3.2
1	A	219	ILE	2.9
1	A	199	PHE	2.9
1	A	197	VAL	2.9
1	A	214	PRO	2.9
1	A	181	GLU	2.9
1	A	169	LEU	2.8
1	A	171	PRO	2.8
1	A	118	GLN	2.8
1	A	208	ALA	2.7
1	A	164	SER	2.7
1	A	136	ALA	2.6
1	A	395	ILE	2.6
1	A	159	LEU	2.6
1	A	172	TRP	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	204	ARG	2.5
1	A	200	SER	2.5
1	A	119	LEU	2.5
1	A	224	ALA	2.4
1	A	168	PRO	2.3
1	A	189	ALA	2.2
1	A	211	ARG	2.2
1	A	121	PRO	2.2
1	A	225	ALA	2.2
1	A	84	PHE	2.1
1	A	8	ALA	2.1
1	A	232	LEU	2.1
1	A	218	LEU	2.1
1	A	191	LYS	2.1
1	A	384	VAL	2.0
1	A	80	GLU	2.0
1	A	178	GLY	2.0
1	A	110	ALA	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^{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.

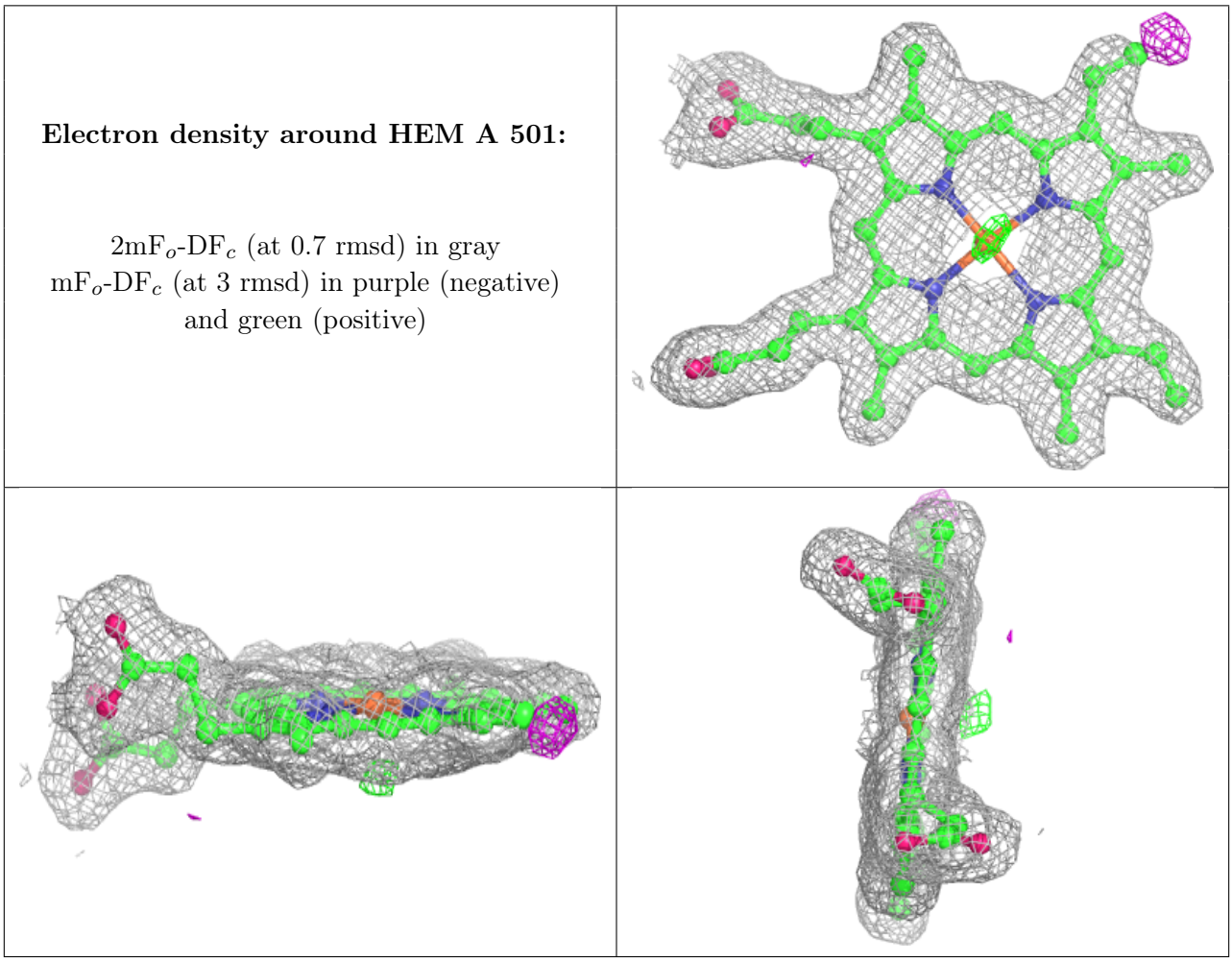
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	GOL	A	502	6/6	0.86	0.18	46,53,56,57	0
2	HEM	A	501	43/43	0.98	0.06	16,23,28,34	0

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 501:

$2mF_o$-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.