Full wwPDB X-ray Structure Validation Report (i)

Jun 24, 2024 – 04:35 AM EDT

PDB ID	:	6A7I
Title	:	CYP154C4 from Streptomyces sp. W2061
Authors	:	Lee, C.W.; Lee, J.H.
Deposited on	:	2018-07-03
Resolution	:	2.19 Å(reported)

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

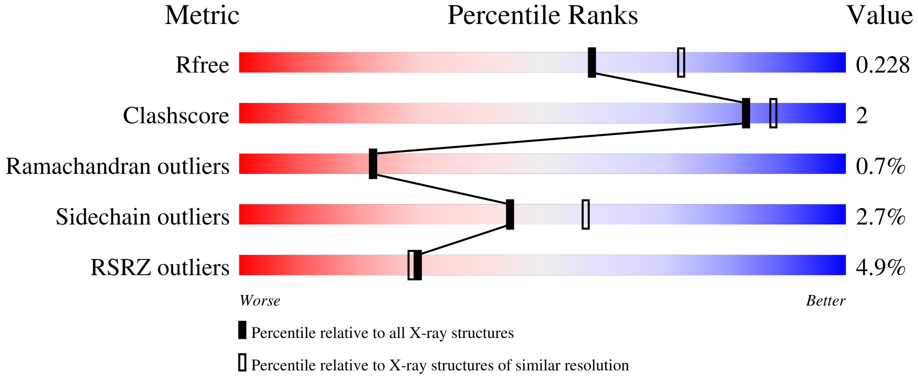
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

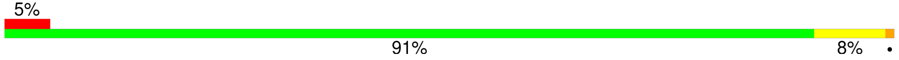
The reported resolution of this entry is 2.19 Å.

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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 3448 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.

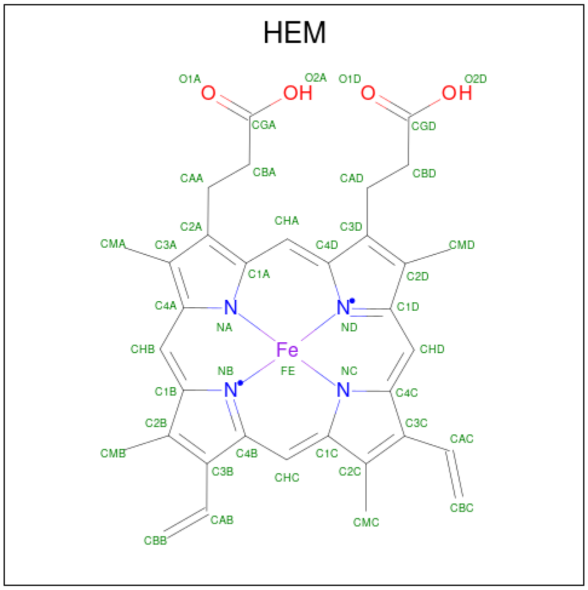
- Molecule 1 is a protein called Cytochrome P450.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	409	3137	1982	554	592	9	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ARG	-	expression tag	UNP A0A087KD84
A	0	LEU	-	expression tag	UNP A0A087KD84
A	60	ASP	ASN	conflict	UNP A0A087KD84
A	192	ALA	THR	conflict	UNP A0A087KD84
A	200	ALA	GLY	conflict	UNP A0A087KD84
A	203	THR	ALA	conflict	UNP A0A087KD84
A	270	GLY	PRO	conflict	UNP A0A087KD84
A	302	LYS	THR	conflict	UNP A0A087KD84

- 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 water.

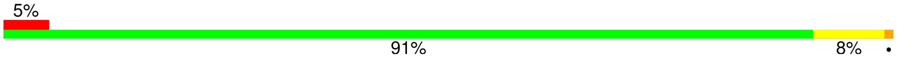
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	268	268	268	0	0

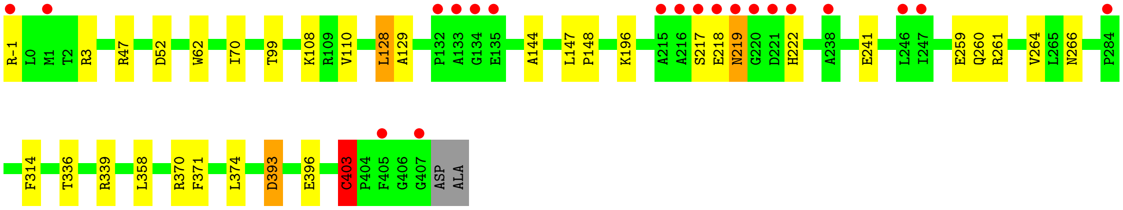
3 Residue-property plots

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

Chain A:





4 Data and refinement statistics

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	83.75Å 115.03Å 127.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.47 – 2.19 35.03 – 2.19	Depositor EDS
% Data completeness (in resolution range)	97.8 (46.47-2.19) 97.8 (35.03-2.19)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.42 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.8.0189	Depositor
R, R_{free}	0.173 , 0.222 0.182 , 0.228	Depositor DCC
R_{free} test set	1554 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	35.2	Xtriage
Anisotropy	0.005	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 41.3	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	0.95	EDS
Total number of atoms	3448	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.14% 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

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	0.97	2/3202 (0.1%)	0.99	7/4361 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	259	GLU	CD-OE1	5.93	1.32	1.25
1	A	259	GLU	CG-CD	5.49	1.60	1.51

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	339	ARG	NE-CZ-NH1	-7.85	116.38	120.30
1	A	3	ARG	NE-CZ-NH1	6.49	123.54	120.30
1	A	47	ARG	NE-CZ-NH1	5.95	123.27	120.30
1	A	52	ASP	CB-CG-OD1	5.36	123.12	118.30
1	A	52	ASP	CB-CG-OD2	-5.23	113.60	118.30
1	A	314	PHE	CB-CG-CD2	-5.22	117.14	120.80
1	A	370	ARG	NE-CZ-NH1	5.04	122.82	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	3137	0	3150	13	0
2	A	43	0	30	2	0
3	A	268	0	0	2	0
All	All	3448	0	3180	15	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 (15) 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:403:CYS:SG	3:A:853:HOH:O	2.39	0.81
1:A:128:LEU:O	1:A:129:ALA:HB3	2.04	0.58
1:A:266:ASN:ND2	3:A:608:HOH:O	2.39	0.54
1:A:110:VAL:HG12	1:A:358:LEU:HD22	1.89	0.54
1:A:62:TRP:CZ3	1:A:70:ILE:HD13	2.47	0.50
1:A:147:LEU:HB3	1:A:148:PRO:HD3	1.95	0.48
1:A:261:ARG:NH2	1:A:374:LEU:O	2.47	0.47
2:A:501:HEM:HBB2	2:A:501:HEM:HMB2	1.96	0.46
1:A:260:GLN:O	1:A:264:VAL:HG23	2.18	0.44
1:A:108:LYS:HE2	1:A:108:LYS:HB2	1.86	0.44
1:A:144:ALA:O	1:A:148:PRO:HD2	2.19	0.43
1:A:129:ALA:HB2	1:A:371:PHE:HZ	1.84	0.43
1:A:217:SER:OG	1:A:218:GLU:N	2.53	0.41
2:A:501:HEM:HHD	2:A:501:HEM:HBC2	2.02	0.41
1:A:241:GLU:OE2	1:A:393:ASP:OD2	2.38	0.41

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	407/411 (99%)	386 (95%)	18 (4%)	3 (1%)	22	22

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	219	ASN
1	A	403	CYS
1	A	336	THR

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	331/332 (100%)	322 (97%)	9 (3%)	44	57

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	-1	ARG
1	A	99	THR
1	A	128	LEU
1	A	196	LYS
1	A	219	ASN
1	A	222	HIS
1	A	393	ASP
1	A	396	GLU
1	A	403	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	103	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

1 ligand is 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	42,50,50	1.51	9 (21%)	46,82,82	1.82	13 (28%)

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	-	3/12/54/54	-

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	501	HEM	CHB-C1B	3.36	1.42	1.34
2	A	501	HEM	C4D-ND	-3.32	1.34	1.40
2	A	501	HEM	C1B-NB	-3.20	1.34	1.40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	501	HEM	C3C-C4C	2.67	1.45	1.41
2	A	501	HEM	CHA-C4D	2.59	1.40	1.34
2	A	501	HEM	C1D-ND	-2.50	1.33	1.38
2	A	501	HEM	C4B-NB	-2.25	1.34	1.38
2	A	501	HEM	FE-NB	2.17	2.10	1.98
2	A	501	HEM	C3B-C4B	2.10	1.49	1.44

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	501	HEM	C4B-CHC-C1C	4.20	128.10	122.56
2	A	501	HEM	C1B-NB-C4B	3.80	109.71	105.21
2	A	501	HEM	CHD-C1D-C2D	-3.61	119.32	125.03
2	A	501	HEM	CHD-C1D-ND	3.58	128.29	124.44
2	A	501	HEM	CMA-C3A-C4A	-2.96	124.12	128.46
2	A	501	HEM	O2D-CGD-CBD	2.67	122.42	114.00
2	A	501	HEM	CHC-C4B-NB	2.63	127.27	124.44
2	A	501	HEM	CAD-C3D-C4D	2.56	129.16	124.70
2	A	501	HEM	CHA-C4D-C3D	-2.45	120.72	125.23
2	A	501	HEM	O2A-CGA-CBA	2.19	120.91	114.00
2	A	501	HEM	C2D-C1D-ND	2.13	112.36	109.90
2	A	501	HEM	C4B-C3B-C2B	-2.12	105.33	107.28
2	A	501	HEM	C3D-C4D-ND	2.03	112.40	110.17

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	501	HEM	CAA-CBA-CGA-O2A
2	A	501	HEM	CAA-CBA-CGA-O1A
2	A	501	HEM	CAD-CBD-CGD-O2D

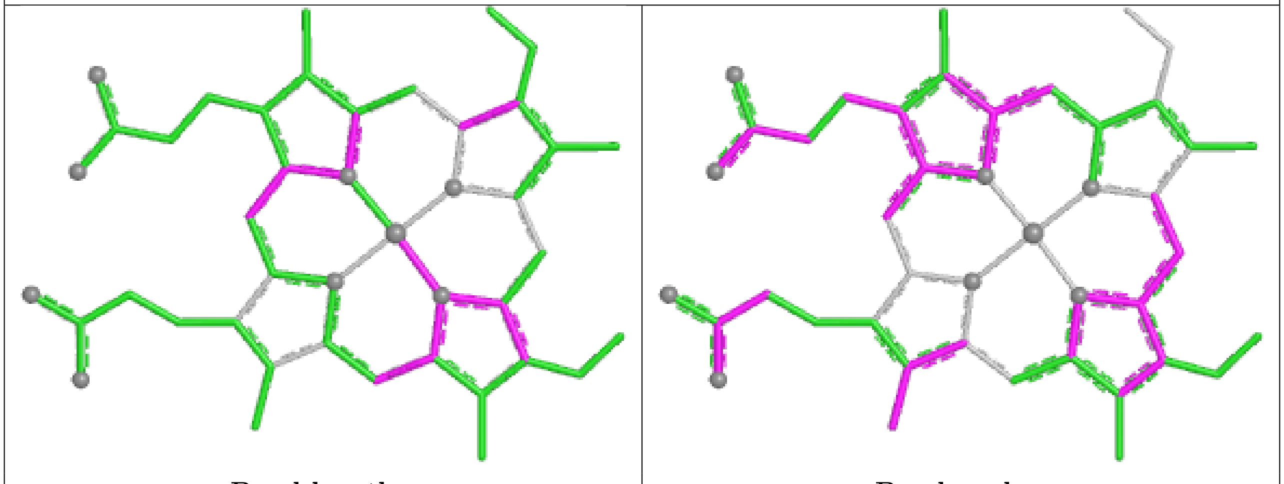
There are no ring outliers.

1 monomer is involved in 2 short contacts:

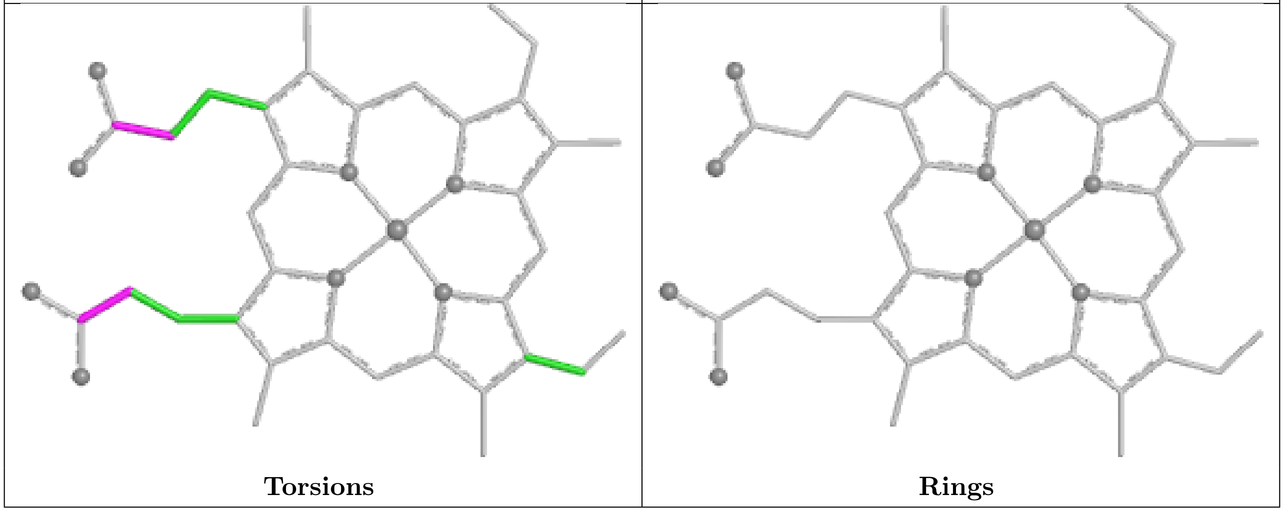
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	HEM	2	0

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



Bond lengths **Bond angles**



Torsions **Rings**

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 ⓘ

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	409/411 (99%)	0.14	20 (4%)	29	28	25, 40, 71, 144	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	217	SER	8.4
1	A	407	GLY	7.8
1	A	219	ASN	7.7
1	A	218	GLU	7.3
1	A	221	ASP	5.1
1	A	216	ALA	4.1
1	A	220	GLY	3.8
1	A	135	GLU	3.2
1	A	-1	ARG	3.2
1	A	215	ALA	3.2
1	A	405	PHE	3.0
1	A	247	ILE	3.0
1	A	132	PRO	3.0
1	A	238	ALA	2.9
1	A	246	LEU	2.8
1	A	134	GLY	2.8
1	A	1	MET	2.3
1	A	133	ALA	2.3
1	A	284	PRO	2.2
1	A	222	HIS	2.2

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

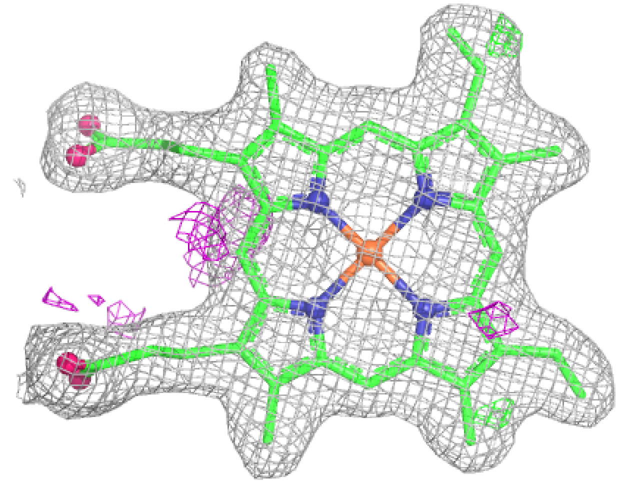
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

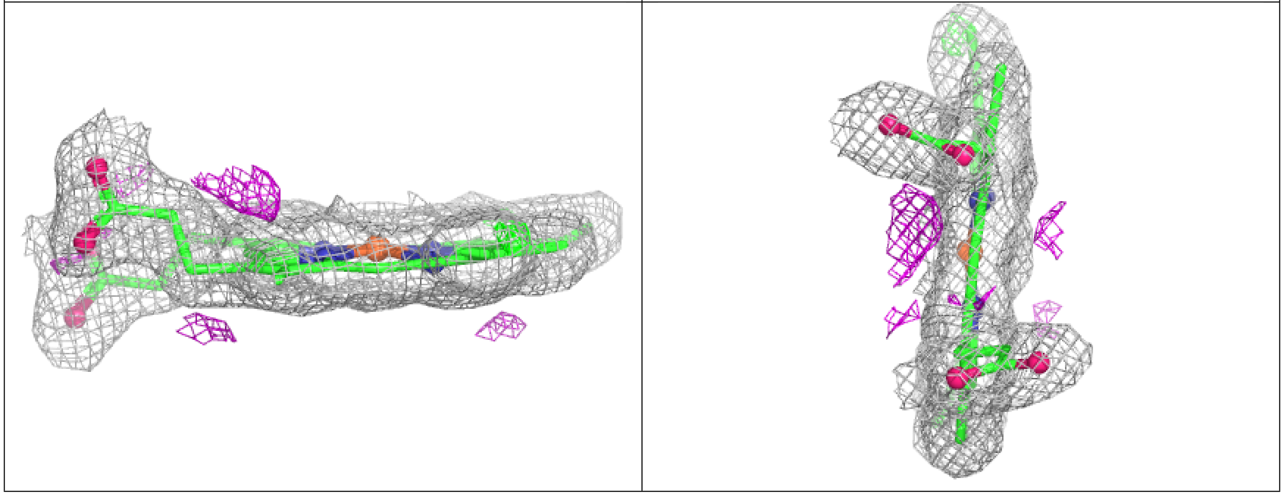
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\AA^2$)	Q<0.9
2	HEM	A	501	43/43	0.98	0.17	24,27,33,40	0

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 ⓘ

There are no such residues in this entry.