wwPDB X-ray Structure Validation Summary Report (i)

Dec 17, 2023 – 02:07 am GMT

PDB ID	:	4APY
Title	:	Ethylene glycol-bound form of P450 CYP125A3 from Mycobacterium smegmatis
Authors	:	Frank, D.J.; Garcia Fernandez, E.; Kells, P.M.; Garcia Lopez, J.L.; Podust, L.M.; Ortiz de Montellano, P.R.
Deposited on	:	2012-04-09
Resolution	:	2.00 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 (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:

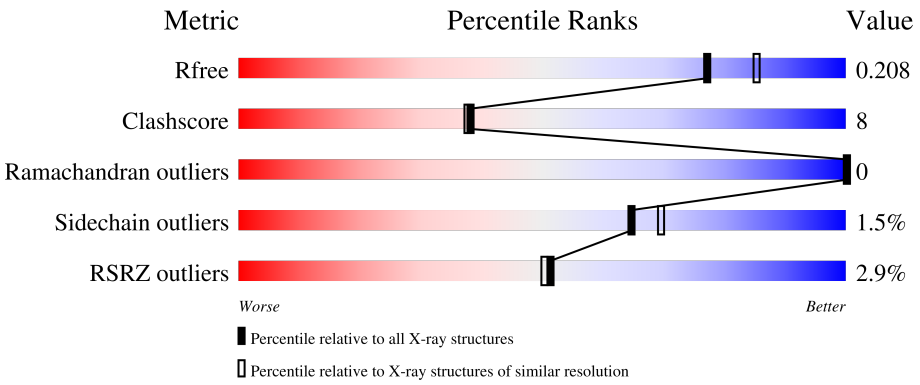
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	433	3% 83% 12% •

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	EDO	A	1425	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3686 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.

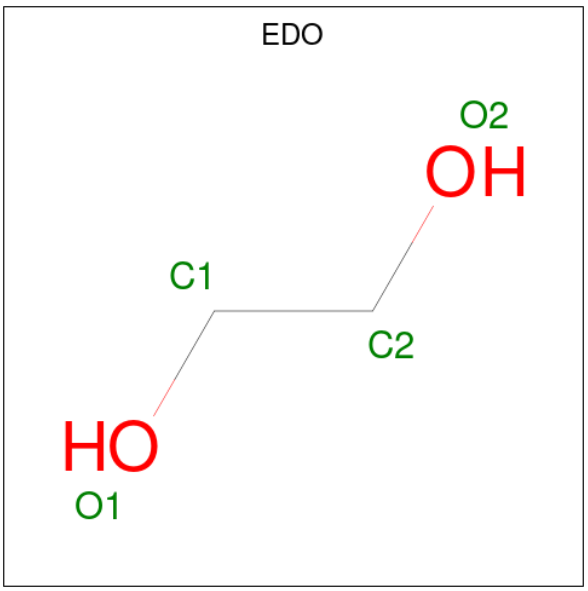
- Molecule 1 is a protein called P450 HEME-THIOLATE PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	414	3259	2049	565	631	14	0	6	0

There are 6 discrepancies between the modelled and reference sequences:

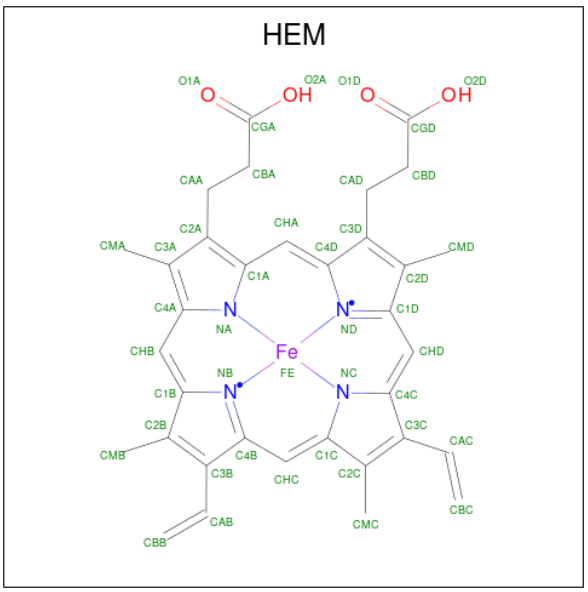
Chain	Residue	Modelled	Actual	Comment	Reference
A	428	HIS	-	expression tag	UNP A0R4Y3
A	429	HIS	-	expression tag	UNP A0R4Y3
A	430	HIS	-	expression tag	UNP A0R4Y3
A	431	HIS	-	expression tag	UNP A0R4Y3
A	432	HIS	-	expression tag	UNP A0R4Y3
A	433	HIS	-	expression tag	UNP A0R4Y3

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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

- Molecule 3 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		
3	A	1	43	34	1	4	4	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
4	A	1	1	1	0	0

- Molecule 5 is water.

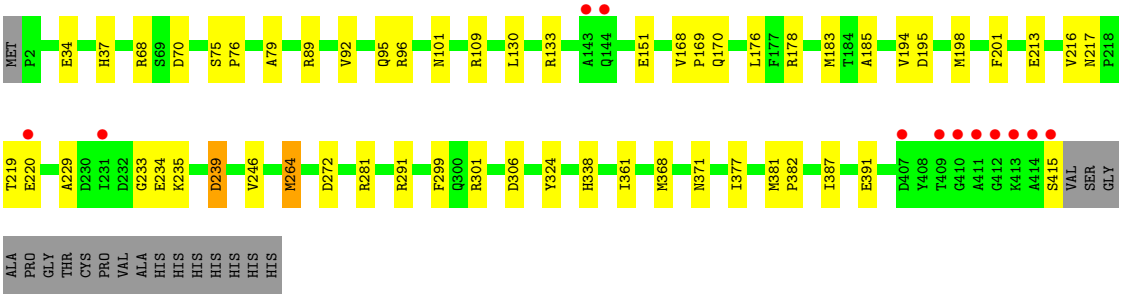
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	339	339	339	0	0

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: P450 HEME-THIOLATE PROTEIN

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 41	Depositor
Cell constants a, b, c, α, β, γ	108.19Å 108.19Å 120.23Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	76.50 – 2.00 54.09 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.2 (76.50-2.00) 99.3 (54.09-2.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.89 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.164 , 0.206 0.166 , 0.208	Depositor DCC
R_{free} test set	2341 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	38.5	Xtriage
Anisotropy	0.088	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 44.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.035 for -h,k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3686	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 3.33% 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: HEM, CL, EDO

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	1.14	3/3332 (0.1%)	1.01	10/4523 (0.2%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	324	TYR	CD1-CE1	6.66	1.49	1.39
1	A	201	PHE	CB-CG	6.04	1.61	1.51
1	A	151	GLU	CB-CG	5.44	1.62	1.52

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	68	ARG	NE-CZ-NH1	7.29	123.94	120.30
1	A	239	ASP	CB-CG-OD1	-7.08	111.93	118.30
1	A	89	ARG	NE-CZ-NH2	-6.94	116.83	120.30
1	A	178	ARG	NE-CZ-NH1	6.11	123.36	120.30
1	A	291	ARG	NE-CZ-NH2	-6.10	117.25	120.30

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	3259	0	3138	49	0
2	A	44	0	65	9	0
3	A	43	0	30	2	0
4	A	1	0	0	0	0
5	A	339	0	0	6	0
All	All	3686	0	3233	50	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 8.

The worst 5 of 50 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:219:THR:HG22	1:A:220:GLU:H	1.23	1.00
1:A:219:THR:HG22	1:A:220:GLU:N	1.84	0.90
1:A:195:ASP:HB3	2:A:1425:EDO:H12	1.55	0.87
3:A:1418:HEM:O1A	5:A:2277:HOH:O	1.91	0.87
1:A:264:MET:HE3	1:A:377:ILE:HG21	1.59	0.84

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	418/433 (96%)	404 (97%)	14 (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	341/357 (96%)	336 (98%)	5 (2%)	65 69

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	169	PRO
1	A	235	LYS
1	A	239	ASP
1	A	299	PHE
1	A	415	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	217	ASN
1	A	364	ASN
1	A	371	ASN
1	A	101	ASN
1	A	37	HIS

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 13 ligands modelled in this entry, 1 is monoatomic - leaving 12 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	EDO	A	1425	-	3,3,3	0.54	0	2,2,2	0.21	0
2	EDO	A	1420	-	3,3,3	0.35	0	2,2,2	1.98	1 (50%)
2	EDO	A	1419	-	3,3,3	1.17	0	2,2,2	0.67	0
2	EDO	A	1421	-	3,3,3	0.43	0	2,2,2	0.60	0
2	EDO	A	1417	-	3,3,3	0.96	0	2,2,2	0.68	0
2	EDO	A	1424	-	3,3,3	0.63	0	2,2,2	0.58	0
3	HEM	A	1418	2,1	41,50,50	2.17	10 (24%)	45,82,82	1.84	10 (22%)
2	EDO	A	1416	-	3,3,3	0.84	0	2,2,2	0.13	0
2	EDO	A	1427	-	3,3,3	0.51	0	2,2,2	0.49	0
2	EDO	A	1423	3	3,3,3	0.93	0	2,2,2	0.77	0
2	EDO	A	1428	-	3,3,3	0.56	0	2,2,2	0.14	0
2	EDO	A	1422	-	3,3,3	0.39	0	2,2,2	1.72	1 (50%)

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	EDO	A	1425	-	-	1/1/1/1	-
2	EDO	A	1420	-	-	1/1/1/1	-
2	EDO	A	1419	-	-	0/1/1/1	-
2	EDO	A	1421	-	-	1/1/1/1	-
2	EDO	A	1417	-	-	1/1/1/1	-
2	EDO	A	1424	-	-	1/1/1/1	-
3	HEM	A	1418	2,1	-	0/12/54/54	-
2	EDO	A	1416	-	-	1/1/1/1	-
2	EDO	A	1427	-	-	1/1/1/1	-
2	EDO	A	1423	3	-	0/1/1/1	-
2	EDO	A	1428	-	-	1/1/1/1	-
2	EDO	A	1422	-	-	0/1/1/1	-

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1418	HEM	C3C-C2C	-6.46	1.31	1.40
3	A	1418	HEM	C3D-C2D	6.33	1.50	1.36
3	A	1418	HEM	CAA-C2A	4.00	1.57	1.52
3	A	1418	HEM	CAB-C3B	3.75	1.57	1.47
3	A	1418	HEM	CMB-C2B	3.21	1.57	1.50

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	1418	HEM	CMA-C3A-C4A	-5.56	119.92	128.46
3	A	1418	HEM	C4D-ND-C1D	4.74	109.97	105.07
3	A	1418	HEM	C4A-C3A-C2A	4.08	109.83	107.00
3	A	1418	HEM	CBA-CAA-C2A	-3.35	106.90	112.62
3	A	1418	HEM	O2D-CGD-CBD	2.82	123.08	114.03

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1416	EDO	O1-C1-C2-O2
2	A	1420	EDO	O1-C1-C2-O2
2	A	1424	EDO	O1-C1-C2-O2
2	A	1417	EDO	O1-C1-C2-O2
2	A	1421	EDO	O1-C1-C2-O2

There are no ring outliers.

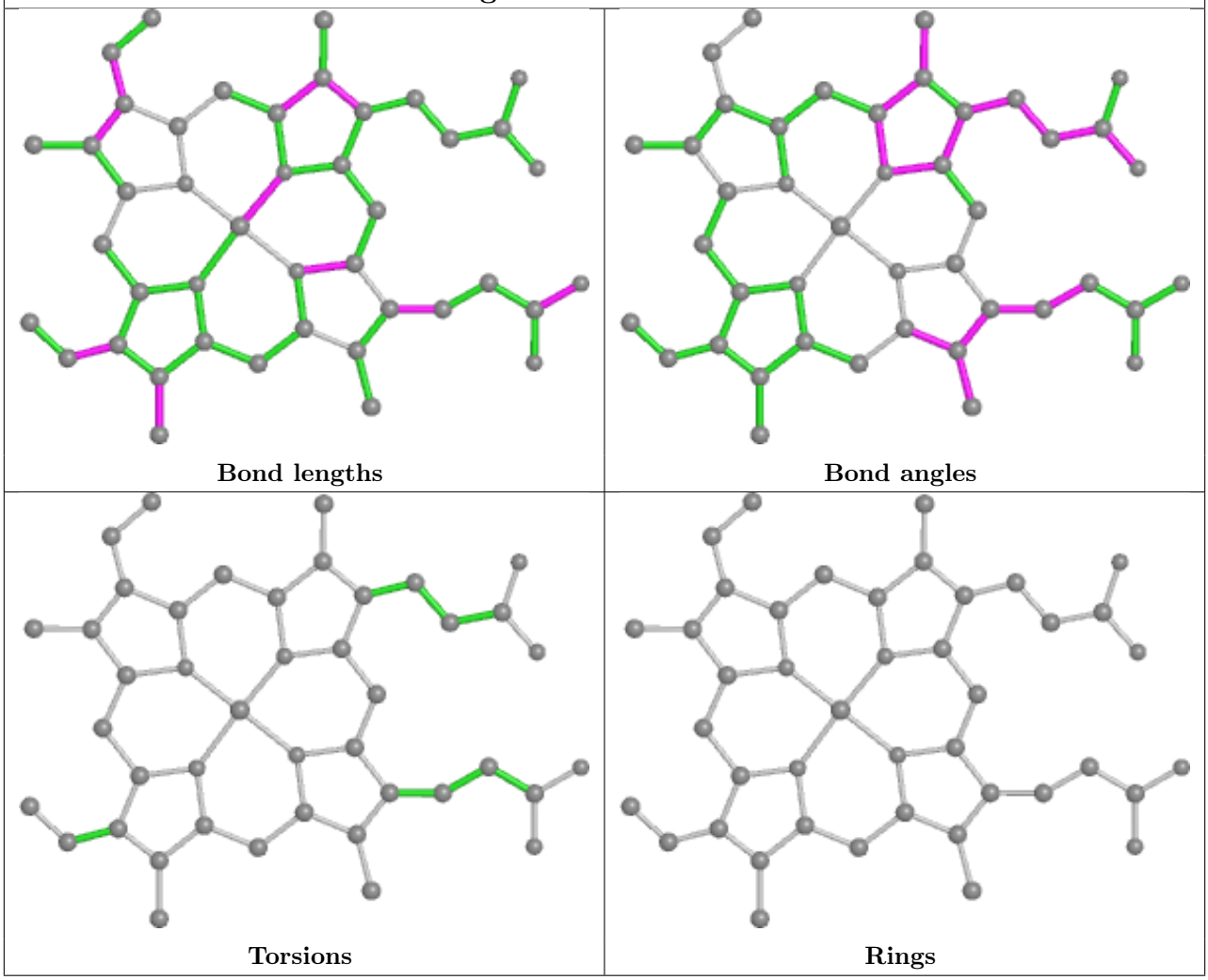
5 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1425	EDO	5	0
2	A	1420	EDO	1	0
2	A	1417	EDO	1	0
3	A	1418	HEM	2	0
2	A	1422	EDO	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 1418



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	414/433 (95%)	-0.35	12 (2%) 51 50	26, 36, 59, 88	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	414	ALA	6.5
1	A	411	ALA	5.8
1	A	413	LYS	4.5
1	A	412	GLY	4.3
1	A	415	SER	4.2

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, 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(Å^2)	Q<0.9
2	EDO	A	1419	4/4	0.78	0.17	51,58,58,62	0
2	EDO	A	1424	4/4	0.78	0.18	70,73,74,75	0

Continued on next page...

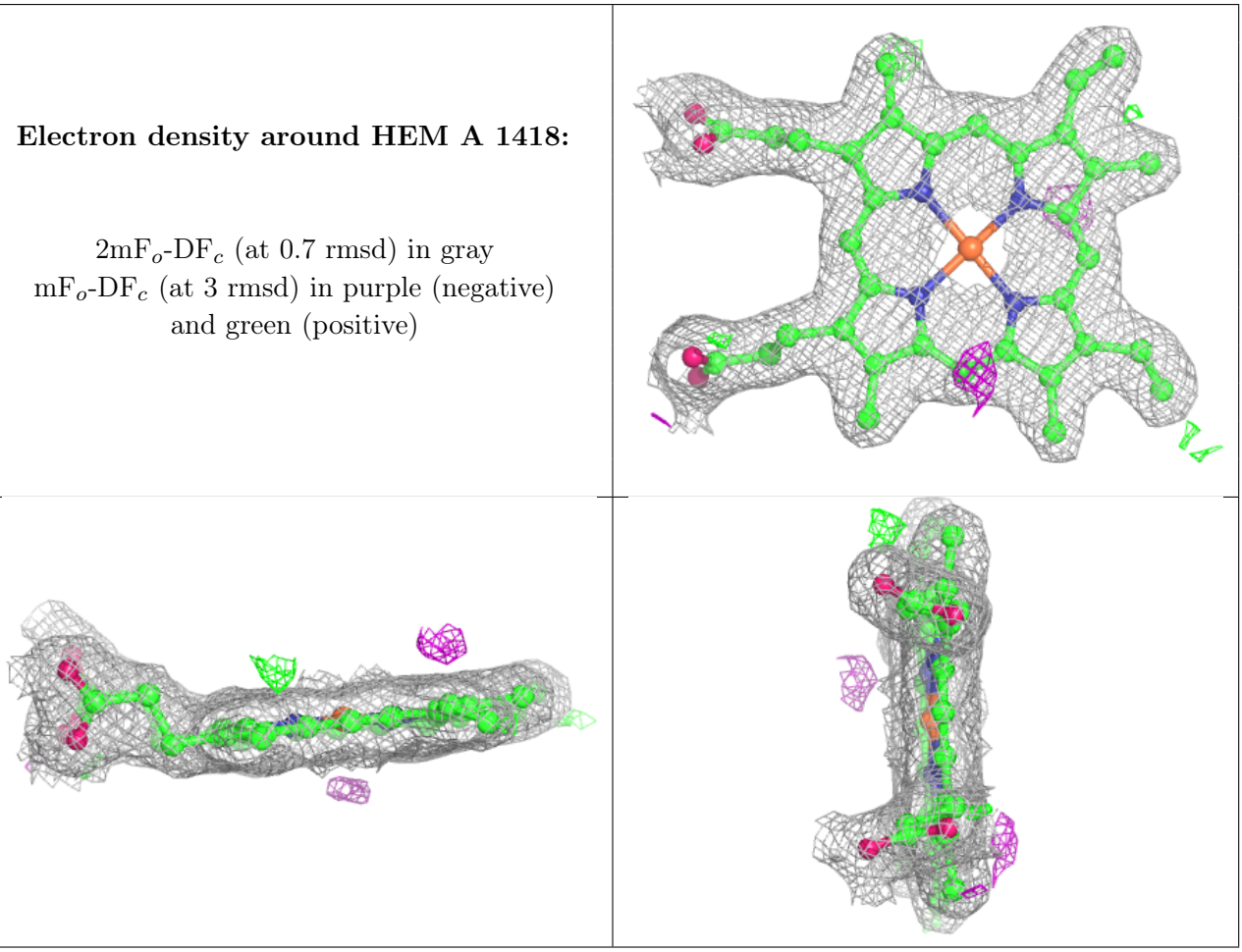
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	EDO	A	1422	4/4	0.80	0.19	54,57,59,63	0
2	EDO	A	1417	4/4	0.85	0.13	44,47,49,51	0
2	EDO	A	1425	4/4	0.86	0.18	29,41,42,49	4
2	EDO	A	1420	4/4	0.89	0.15	49,56,56,62	0
2	EDO	A	1428	4/4	0.89	0.31	57,62,64,68	0
2	EDO	A	1416	4/4	0.92	0.14	58,58,60,60	0
2	EDO	A	1421	4/4	0.93	0.20	48,52,53,53	0
2	EDO	A	1427	4/4	0.96	0.13	43,46,52,53	0
2	EDO	A	1423	4/4	0.96	0.10	39,40,41,46	0
4	CL	A	1426	1/1	0.96	0.23	67,67,67,67	1
3	HEM	A	1418	43/43	0.99	0.12	23,27,31,39	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 1418:

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.