



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 28, 2016 – 02:38 AM EST

PDB ID : 5KYO  
Title : Crystal Structure of CYP101J2  
Authors : Unterweger, B.; Drinkwater, N.; Dumsday, G.J.; McGowan, S.  
Deposited on : 2016-07-22  
Resolution : 1.80 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20028442  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20028442

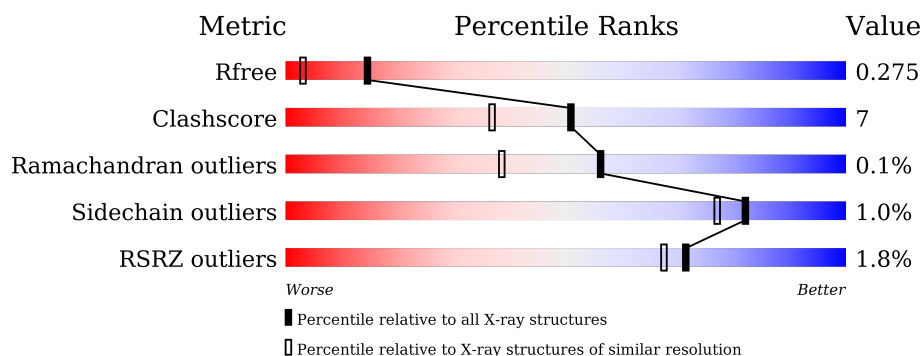
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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))
$R_{free}$	91344	4533 (1.80-1.80)
Clashscore	102246	5383 (1.80-1.80)
Ramachandran outliers	100387	5320 (1.80-1.80)
Sidechain outliers	100360	5319 (1.80-1.80)
RSRZ outliers	91569	4547 (1.80-1.80)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	430	
1	B	430	
1	C	430	
1	D	430	
1	E	430	
1	F	430	

## 2 Entry composition [i](#)

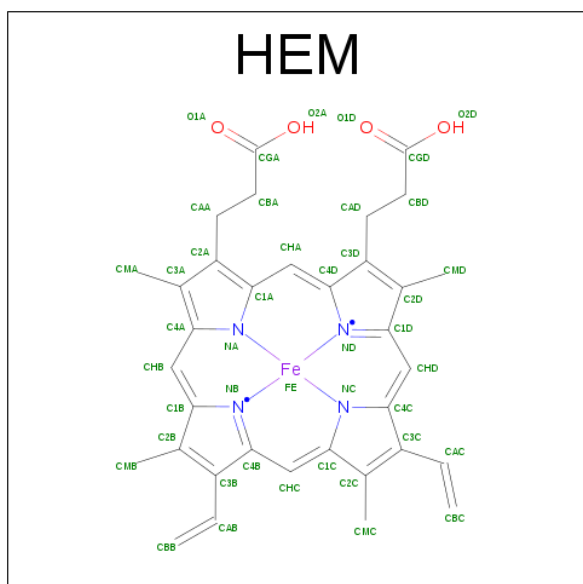
There are 3 unique types of molecules in this entry. The entry contains 20855 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 CYP101J2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	394	Total	C	N	O	S	0	11	0
			3158	2024	554	563	17			
1	B	393	Total	C	N	O	S	0	10	0
			3133	2009	552	556	16			
1	C	390	Total	C	N	O	S	0	6	0
			3093	1980	549	549	15			
1	D	384	Total	C	N	O	S	0	6	0
			3046	1949	541	540	16			
1	E	390	Total	C	N	O	S	0	4	0
			3054	1953	547	539	15			
1	F	383	Total	C	N	O	S	0	7	0
			2993	1916	528	533	16			

- 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
2	C	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
2	D	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
2	E	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
2	F	1	Total 43	C 34	Fe 1	N 4	O 4	0	0

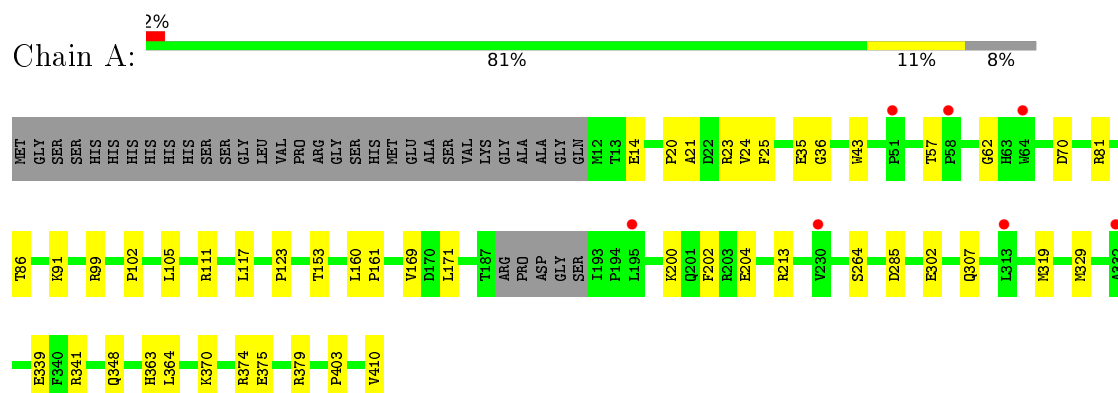
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	443	Total 443	O 443	0	0
3	B	426	Total 426	O 426	0	1
3	C	418	Total 418	O 418	0	0
3	D	312	Total 312	O 312	0	0
3	E	280	Total 280	O 280	0	0
3	F	241	Total 241	O 241	0	0

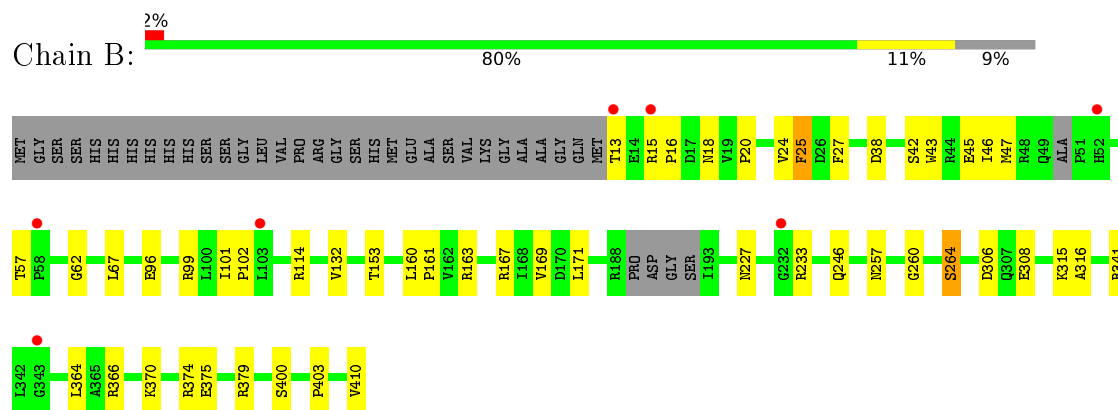
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

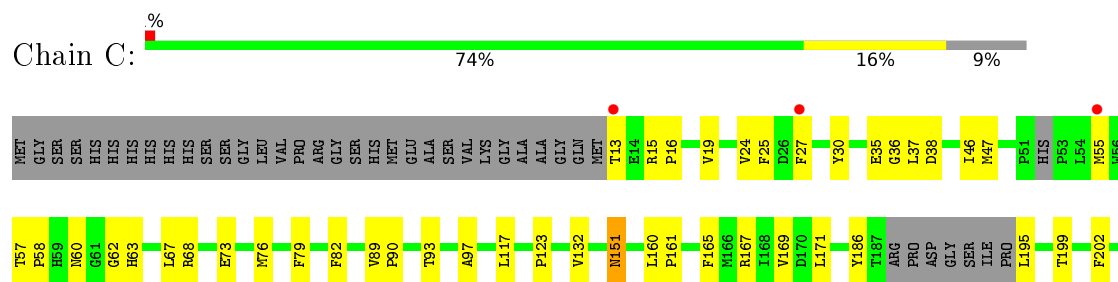
#### • Molecule 1: CYP101J2



#### • Molecule 1: CYP101J2

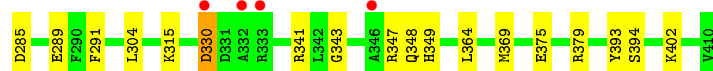
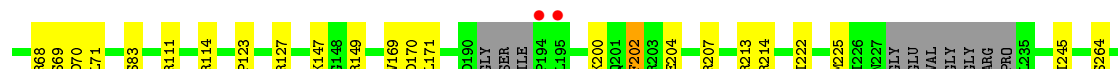
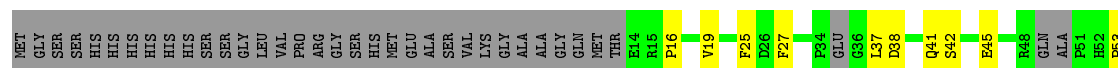
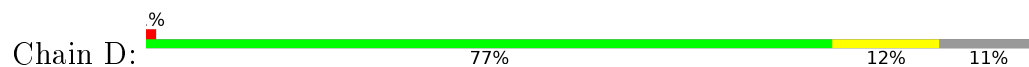


#### • Molecule 1: CYP101J2

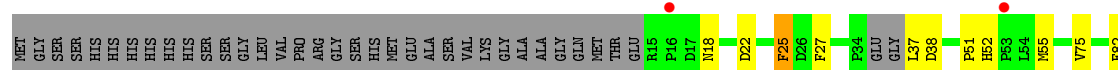
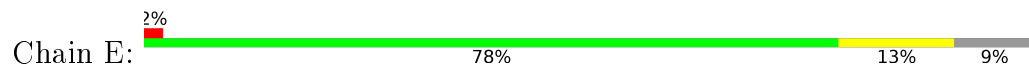




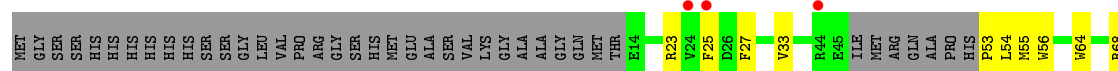
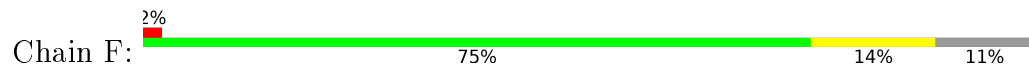
• Molecule 1: CYP101J2



• Molecule 1: CYP101J2



• Molecule 1: CYP101J2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.43Å 100.03Å 99.96Å 63.46° 63.66° 63.67°	Depositor
Resolution (Å)	42.60 – 1.80 52.60 – 1.80	Depositor EDS
% Data completeness (in resolution range)	95.2 (42.60-1.80) 83.2 (52.60-1.80)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.89 (at 1.80Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.223 , 0.276 0.224 , 0.275	Depositor DCC
$R_{free}$ test set	11754 reflections (4.53%)	DCC
Wilson B-factor (Å <sup>2</sup> )	21.1	Xtriage
Anisotropy	0.280	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 45.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.409 for l,h,k 0.409 for k,l,h 0.053 for -l,-k,-h 0.054 for -k,-h,-l 0.053 for -h,-l,-k	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	20855	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 16.68% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

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	0.40	0/3262	0.54	0/4431
1	B	0.41	0/3240	0.55	0/4404
1	C	0.44	1/3185 (0.0%)	0.56	0/4325
1	D	0.38	0/3137	0.54	0/4261
1	E	0.38	0/3143	0.55	0/4276
1	F	0.37	0/3087	0.53	1/4200 (0.0%)
All	All	0.40	1/19054 (0.0%)	0.54	1/25897 (0.0%)

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	D	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	410	VAL	CB-CG1	-6.57	1.39	1.52

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	287	LEU	CA-CB-CG	5.59	128.15	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	D	348	GLN	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	3158	0	3116	37	0
1	B	3133	0	3080	44	0
1	C	3093	0	3045	56	0
1	D	3046	0	2976	35	0
1	E	3054	0	2974	38	0
1	F	2993	0	2883	39	0
2	A	43	0	30	3	0
2	B	43	0	30	3	0
2	C	43	0	30	5	0
2	D	43	0	30	3	0
2	E	43	0	30	3	0
2	F	43	0	30	4	0
3	A	443	0	0	10	0
3	B	426	0	0	16	0
3	C	418	0	0	14	0
3	D	312	0	0	7	0
3	E	280	0	0	11	0
3	F	241	0	0	3	0
All	All	20855	0	18254	254	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 7.

The worst 5 of 254 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:D:375:GLU:OE1	1:D:379:ARG:NH1	2.01	0.93
1:F:111:ARG:HH11	1:F:114:ARG:HH22	1.19	0.91
1:A:410:VAL:HG21	1:B:410:VAL:H	1.40	0.86
1:F:375:GLU:OE1	1:F:379:ARG:NH1	2.10	0.85
1:F:247:VAL:HG13	2:F:501:HEM:HBC1	1.60	0.83

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	401/430 (93%)	390 (97%)	11 (3%)	0	100	100
1	B	397/430 (92%)	390 (98%)	7 (2%)	0	100	100
1	C	390/430 (91%)	378 (97%)	11 (3%)	1 (0%)	46	29
1	D	380/430 (88%)	368 (97%)	12 (3%)	0	100	100
1	E	386/430 (90%)	373 (97%)	13 (3%)	0	100	100
1	F	382/430 (89%)	368 (96%)	13 (3%)	1 (0%)	46	29
All	All	2336/2580 (90%)	2267 (97%)	67 (3%)	2 (0%)	56	38

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	348	GLN
1	F	235	LEU

### 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	338/372 (91%)	333 (98%)	5 (2%)	72	62
1	B	333/372 (90%)	328 (98%)	5 (2%)	72	62
1	C	327/372 (88%)	325 (99%)	2 (1%)	90	88

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	322/372 (87%)	317 (98%)	5 (2%)	70	59
1	E	319/372 (86%)	316 (99%)	3 (1%)	84	80
1	F	310/372 (83%)	308 (99%)	2 (1%)	90	88
All	All	1949/2232 (87%)	1927 (99%)	22 (1%)	82	74

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	308	GLU
1	D	202	PHE
1	F	196	ASP
1	C	151	ASN
1	C	202	PHE

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

Mol	Chain	Res	Type
1	C	151	ASN

### 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](#)

6 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	24,50,50	2.22	8 (33%)	16,82,82	1.24	2 (12%)
2	HEM	B	501	1	24,50,50	2.10	8 (33%)	16,82,82	1.63	4 (25%)
2	HEM	C	501	1,3	24,50,50	2.26	5 (20%)	16,82,82	1.92	5 (31%)
2	HEM	D	501	1,3	24,50,50	2.26	7 (29%)	16,82,82	1.62	5 (31%)
2	HEM	E	501	1,3	24,50,50	2.25	7 (29%)	16,82,82	1.34	3 (18%)
2	HEM	F	501	1,3	24,50,50	2.41	8 (33%)	16,82,82	1.56	4 (25%)

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	-	0/6/54/54	0/0/8/8
2	HEM	B	501	1	-	0/6/54/54	0/0/8/8
2	HEM	C	501	1,3	-	0/6/54/54	0/0/8/8
2	HEM	D	501	1,3	-	0/6/54/54	0/0/8/8
2	HEM	E	501	1,3	-	0/6/54/54	0/0/8/8
2	HEM	F	501	1,3	-	0/6/54/54	0/0/8/8

The worst 5 of 43 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	501	HEM	C3B-C2B	-5.67	1.33	1.40
2	C	501	HEM	C3B-C2B	-5.24	1.33	1.40
2	F	501	HEM	C3C-C2C	-4.84	1.34	1.40
2	C	501	HEM	C3C-C2C	-4.80	1.34	1.40
2	D	501	HEM	C3B-C2B	-4.57	1.34	1.40

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	501	HEM	CMA-C3A-C4A	-3.23	122.82	128.31
2	C	501	HEM	CAD-CBD-CGD	-3.20	106.55	112.78
2	B	501	HEM	CAD-CBD-CGD	-3.09	106.77	112.78
2	C	501	HEM	C3C-CAC-CBC	-3.02	120.32	126.40

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	501	HEM	CBD-CAD-C3D	-2.86	107.45	112.47

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 21 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	HEM	3	0
2	B	501	HEM	3	0
2	C	501	HEM	5	0
2	D	501	HEM	3	0
2	E	501	HEM	3	0
2	F	501	HEM	4	0

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	394/430 (91%)	0.07	7 (1%) 71 67	14, 24, 45, 66	8 (2%)
1	B	393/430 (91%)	0.18	7 (1%) 71 67	14, 24, 44, 67	15 (3%)
1	C	390/430 (90%)	0.13	5 (1%) 79 76	14, 24, 44, 64	8 (2%)
1	D	384/430 (89%)	0.15	6 (1%) 74 71	17, 28, 50, 70	6 (1%)
1	E	390/430 (90%)	0.21	9 (2%) 64 59	17, 27, 49, 82	8 (2%)
1	F	383/430 (89%)	0.21	9 (2%) 64 59	17, 27, 51, 75	7 (1%)
All	All	2334/2580 (90%)	0.16	43 (1%) 71 67	14, 26, 47, 82	52 (2%)

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	343	GLY	11.4
1	E	343	GLY	6.1
1	B	343	GLY	5.0
1	B	232	GLY	4.2
1	D	330	ASP	3.6

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	HEM	D	501	43/43	0.95	0.12	0.44	15,18,27,29	0
2	HEM	F	501	43/43	0.94	0.12	0.43	16,18,26,36	0
2	HEM	E	501	43/43	0.95	0.11	-0.07	15,17,24,34	0
2	HEM	C	501	43/43	0.95	0.11	-0.09	14,17,23,35	0
2	HEM	B	501	43/43	0.95	0.11	-0.24	14,17,25,30	0
2	HEM	A	501	43/43	0.96	0.10	-0.30	14,17,22,29	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.