



# wwPDB X-ray Structure Validation Summary Report

Oct 24, 2017 – 03:35 PM EDT

PDB ID : 3GYS  
Title : Crystal structure determination of cat (*Felis silvestris catus*) hemoglobin at 2.9 angstrom resolution  
Authors : Balasubramanian, M.; Sathya Moorthy, Pon.; Neelagandan, K.; Ponnuswamy, M.N.  
Deposited on : unknown  
Resolution : 2.90 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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-20030345

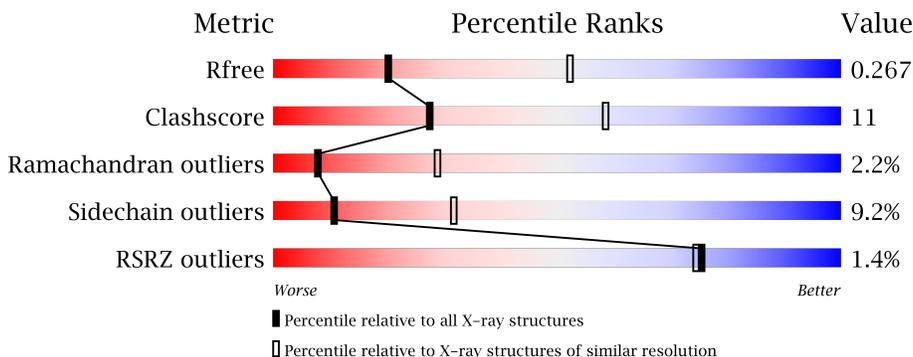
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

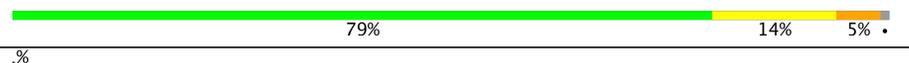
The reported resolution of this entry is 2.90 Å.

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))
$R_{free}$	100719	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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	141	
1	C	141	
1	E	141	
1	G	141	
2	B	145	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	D	145	 <p>.% 78% 20% .</p>
2	F	145	 <p>2% 61% 30% 8% .</p>
2	H	145	 <p>3% 77% 18% 5% .</p>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 9038 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.

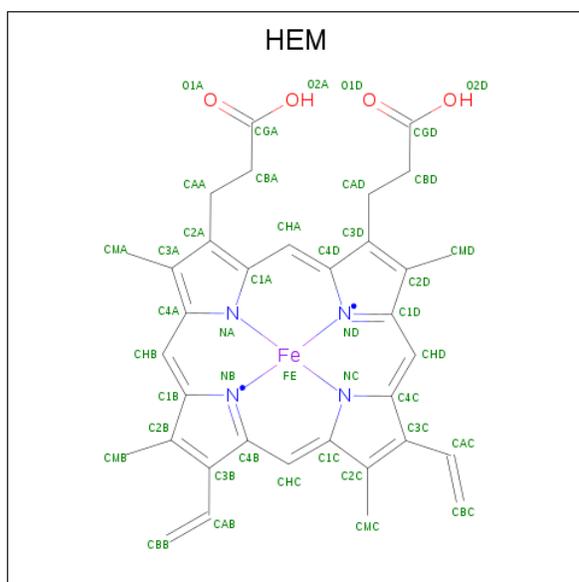
- Molecule 1 is a protein called Hemoglobin subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	138	1046	668	179	193	6	0	0	0
1	C	139	1055	674	181	194	6	0	0	0
1	E	138	1046	668	179	193	6	0	0	0
1	G	138	1046	668	179	193	6	0	0	0

- Molecule 2 is a protein called Hemoglobin subunit beta-A/B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	145	1123	721	197	202	3	0	0	0
2	D	145	1123	721	197	202	3	0	0	0
2	F	145	1123	721	197	202	3	0	0	0
2	H	145	1123	721	197	202	3	0	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	Total 43	C 34	Fe 1	N 4	O 4	0	0
3	B	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
3	C	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
3	D	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
3	E	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
3	F	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
3	G	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
3	H	1	Total 43	C 34	Fe 1	N 4	O 4	0	0

- Molecule 4 is water.

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

*Continued on next page...*

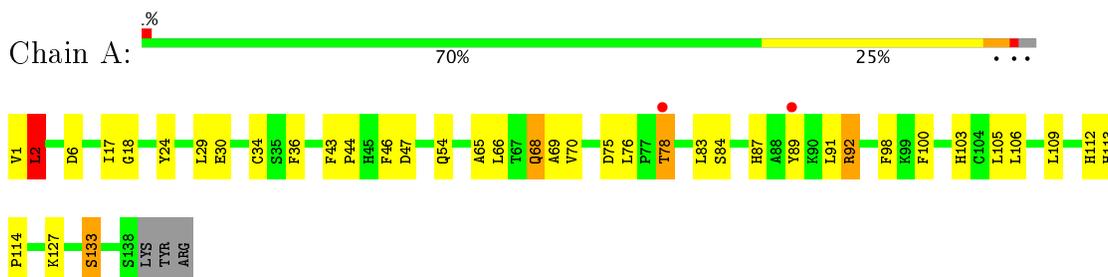
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
4	F	2	Total O 2 2	0	0
4	G	2	Total O 2 2	0	0
4	H	1	Total O 1 1	0	0

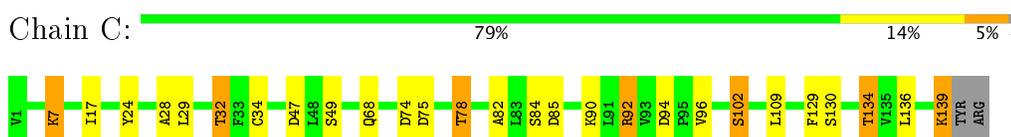
### 3 Residue-property plots [i](#)

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

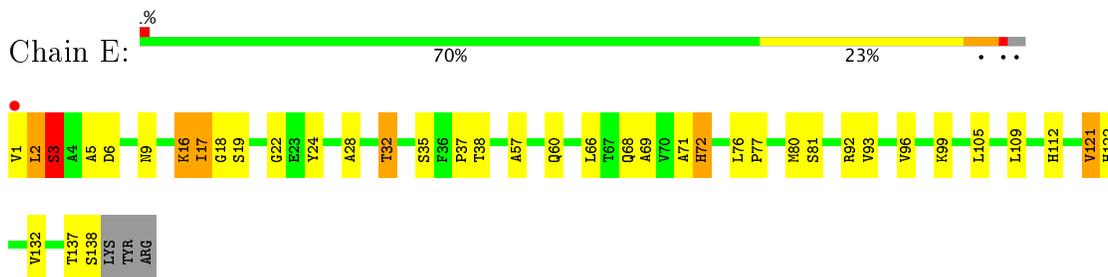
- Molecule 1: Hemoglobin subunit alpha



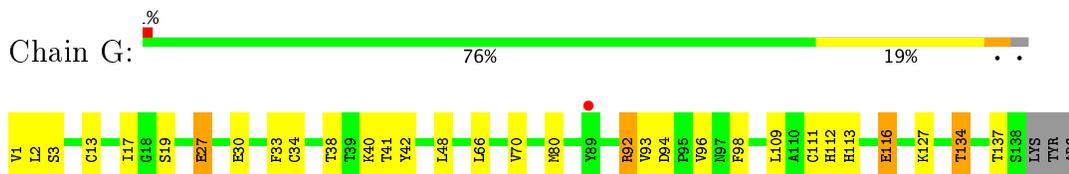
- Molecule 1: Hemoglobin subunit alpha



- Molecule 1: Hemoglobin subunit alpha

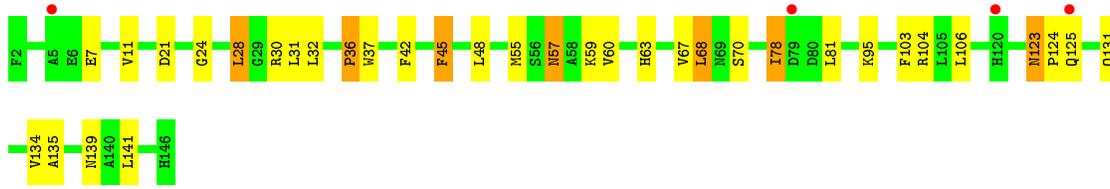


- Molecule 1: Hemoglobin subunit alpha

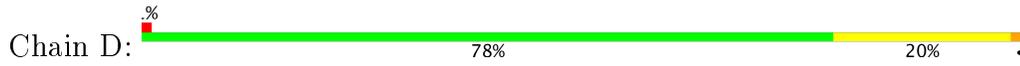


- Molecule 2: Hemoglobin subunit beta-A/B

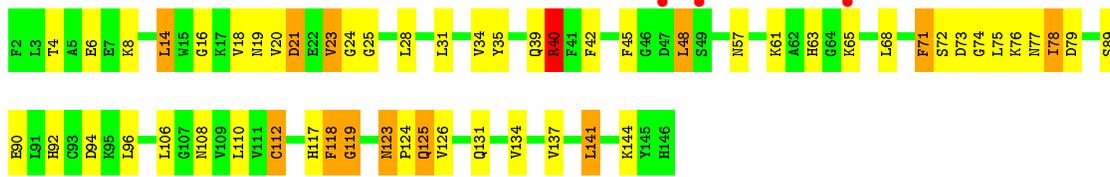




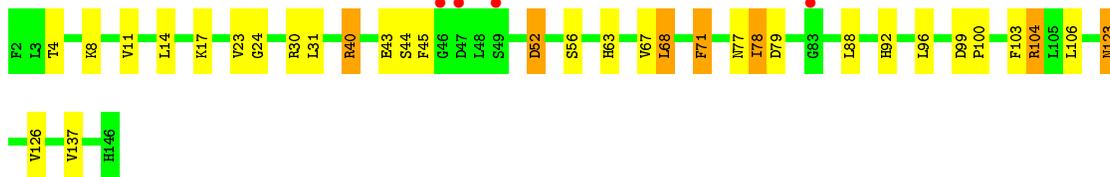
• Molecule 2: Hemoglobin subunit beta-A/B



• Molecule 2: Hemoglobin subunit beta-A/B



• Molecule 2: Hemoglobin subunit beta-A/B



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	86.02Å 96.47Å 146.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.13 – 2.90 24.13 – 2.90	Depositor EDS
% Data completeness (in resolution range)	96.0 (24.13-2.90) 96.1 (24.13-2.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.28 (at 2.89Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.187 , 0.271 0.185 , 0.267	Depositor DCC
$R_{free}$ test set	2656 reflections (11.12%)	DCC
Wilson B-factor (Å <sup>2</sup> )	55.2	Xtrriage
Anisotropy	0.080	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 45.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9038	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.88% 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	0.69	0/1075	0.74	0/1462
1	C	0.79	1/1084 (0.1%)	0.79	0/1473
1	E	0.66	1/1075 (0.1%)	0.69	0/1462
1	G	0.71	0/1075	0.75	0/1462
2	B	0.69	0/1149	0.76	0/1551
2	D	0.68	0/1149	0.71	0/1551
2	F	0.63	0/1149	0.67	0/1551
2	H	0.64	0/1149	0.75	2/1551 (0.1%)
All	All	0.69	2/8905 (0.0%)	0.73	2/12063 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	34	CYS	CB-SG	-7.22	1.70	1.82
1	E	1	VAL	CB-CG2	5.17	1.63	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	99	ASP	CB-CG-OD2	5.92	123.63	118.30
2	H	123	ASN	CB-CA-C	-5.26	99.87	110.40

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	1046	0	1022	30	0
1	C	1055	0	1035	19	0
1	E	1046	0	1022	25	0
1	G	1046	0	1022	20	0
2	B	1123	0	1109	26	0
2	D	1123	0	1109	21	0
2	F	1123	0	1109	35	0
2	H	1123	0	1109	25	0
3	A	43	0	30	3	0
3	B	43	0	30	0	0
3	C	43	0	30	2	0
3	D	43	0	30	8	0
3	E	43	0	30	4	0
3	F	43	0	30	4	0
3	G	43	0	30	2	0
3	H	43	0	30	4	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
4	E	1	0	0	0	0
4	F	2	0	0	0	0
4	G	2	0	0	0	0
4	H	1	0	0	0	0
All	All	9038	0	8777	202	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 11.

The worst 5 of 202 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)
2:D:123:ASN:HB2	2:D:126:VAL:HG23	1.30	1.06
1:E:2:LEU:O	1:E:3:SER:HB2	1.61	0.98
1:A:75:ASP:OD2	1:A:78:THR:HB	1.70	0.91
2:H:24:GLY:HA2	2:H:68:LEU:HG	1.55	0.86
2:F:21:ASP:HB2	2:F:65:LYS:HG2	1.58	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	136/141 (96%)	126 (93%)	8 (6%)	2 (2%)	12	39
1	C	137/141 (97%)	131 (96%)	5 (4%)	1 (1%)	25	60
1	E	136/141 (96%)	117 (86%)	13 (10%)	6 (4%)	3	11
1	G	136/141 (96%)	123 (90%)	10 (7%)	3 (2%)	8	29
2	B	143/145 (99%)	128 (90%)	14 (10%)	1 (1%)	25	60
2	D	143/145 (99%)	125 (87%)	15 (10%)	3 (2%)	8	30
2	F	143/145 (99%)	125 (87%)	12 (8%)	6 (4%)	3	12
2	H	143/145 (99%)	122 (85%)	18 (13%)	3 (2%)	8	30
All	All	1117/1144 (98%)	997 (89%)	95 (8%)	25 (2%)	8	29

5 of 25 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2	LEU
1	E	3	SER
1	E	17	ILE
2	B	21	ASP
1	E	18	GLY

### 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	112/115 (97%)	103 (92%)	9 (8%)	14	38

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	113/115 (98%)	105 (93%)	8 (7%)	17	44
1	E	112/115 (97%)	100 (89%)	12 (11%)	8	23
1	G	112/115 (97%)	105 (94%)	7 (6%)	21	51
2	B	117/117 (100%)	106 (91%)	11 (9%)	10	30
2	D	117/117 (100%)	107 (92%)	10 (8%)	12	35
2	F	117/117 (100%)	101 (86%)	16 (14%)	4	12
2	H	117/117 (100%)	106 (91%)	11 (9%)	10	30
All	All	917/928 (99%)	833 (91%)	84 (9%)	11	32

5 of 84 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	141	LEU
1	E	96	VAL
2	H	52	ASP
1	E	3	SER
1	E	68	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	131	GLN
2	F	117	HIS
2	H	131	GLN
1	E	9	ASN
1	E	60	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

8 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	HEM	A	143	1,4	28,50,50	2.25	8 (28%)	17,82,82	1.79	1 (5%)
3	HEM	B	148	2	28,50,50	2.39	11 (39%)	17,82,82	1.58	4 (23%)
3	HEM	C	143	1,4	28,50,50	2.49	8 (28%)	17,82,82	3.30	8 (47%)
3	HEM	D	148	2	28,50,50	2.43	9 (32%)	17,82,82	1.92	4 (23%)
3	HEM	E	143	1,4	28,50,50	2.22	8 (28%)	17,82,82	1.97	5 (29%)
3	HEM	F	148	2,4	28,50,50	2.30	9 (32%)	17,82,82	1.81	6 (35%)
3	HEM	G	143	1,4	28,50,50	2.24	7 (25%)	17,82,82	1.67	4 (23%)
3	HEM	H	148	2	28,50,50	2.33	8 (28%)	17,82,82	2.01	5 (29%)

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	HEM	A	143	1,4	-	0/6/54/54	0/0/8/8
3	HEM	B	148	2	-	0/6/54/54	0/0/8/8
3	HEM	C	143	1,4	-	0/6/54/54	0/0/8/8
3	HEM	D	148	2	-	0/6/54/54	0/0/8/8
3	HEM	E	143	1,4	-	0/6/54/54	0/0/8/8
3	HEM	F	148	2,4	-	0/6/54/54	0/0/8/8
3	HEM	G	143	1,4	-	0/6/54/54	0/0/8/8
3	HEM	H	148	2	-	0/6/54/54	0/0/8/8

The worst 5 of 68 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	148	HEM	C3B-C2B	-5.74	1.32	1.40
3	C	143	HEM	C3C-C2C	-5.70	1.32	1.40
3	B	148	HEM	C3B-C2B	-5.11	1.33	1.40
3	F	148	HEM	C3B-C2B	-5.00	1.33	1.40
3	G	143	HEM	C3C-C2C	-4.87	1.33	1.40

The worst 5 of 37 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	143	HEM	CAA-CBA-CGA	-8.78	97.65	112.66
3	A	143	HEM	CAA-CBA-CGA	-5.81	102.73	112.66
3	C	143	HEM	CAD-CBD-CGD	-5.54	103.20	112.66
3	E	143	HEM	CAA-CBA-CGA	-5.35	103.53	112.66
3	D	148	HEM	CAA-CBA-CGA	-5.03	104.07	112.66

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

7 monomers are involved in 27 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	143	HEM	3	0
3	C	143	HEM	2	0
3	D	148	HEM	8	0
3	E	143	HEM	4	0
3	F	148	HEM	4	0
3	G	143	HEM	2	0
3	H	148	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	138/141 (97%)	-0.18	2 (1%) 75 74	31, 44, 55, 61	0
1	C	139/141 (98%)	-0.36	0 100 100	19, 30, 45, 57	0
1	E	138/141 (97%)	-0.13	1 (0%) 87 86	37, 55, 66, 71	0
1	G	138/141 (97%)	-0.27	1 (0%) 87 86	26, 40, 57, 60	0
2	B	145/145 (100%)	-0.11	4 (2%) 53 48	28, 51, 66, 73	0
2	D	145/145 (100%)	-0.13	1 (0%) 87 86	24, 47, 62, 66	0
2	F	145/145 (100%)	0.07	3 (2%) 64 60	38, 64, 84, 87	0
2	H	145/145 (100%)	0.11	4 (2%) 53 48	36, 56, 71, 73	0
All	All	1133/1144 (99%)	-0.12	16 (1%) 75 74	19, 48, 71, 87	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	49	SER	3.8
1	E	1	VAL	3.7
2	F	47	ASP	3.2
2	B	120	HIS	3.2
2	H	83	GLY	2.7

### 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
3	HEM	B	148	43/43	0.92	0.19	0.47	39,44,59,61	0
3	HEM	F	148	43/43	0.92	0.22	0.47	58,63,76,78	0
3	HEM	E	143	43/43	0.96	0.18	0.19	40,47,56,57	0
3	HEM	D	148	43/43	0.93	0.20	0.13	43,46,60,61	0
3	HEM	C	143	43/43	0.95	0.17	0.04	23,28,38,42	0
3	HEM	A	143	43/43	0.95	0.17	-0.09	32,39,52,58	0
3	HEM	G	143	43/43	0.97	0.16	-0.22	27,33,42,47	0
3	HEM	H	148	43/43	0.94	0.19	-0.33	62,67,74,75	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.