



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 04:11 pm GMT

PDB ID : 1THB  
Title : REFINEMENT OF A PARTIALLY OXYGENATED T STATE  
HAEMOGLOBIN AT 1.5 ANGSTROMS RESOLUTION  
Authors : Waller, D.A.; Liddington, R.C.  
Deposited on : 1990-01-23  
Resolution : 1.50 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 : trunk28620  
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) : recalc28949

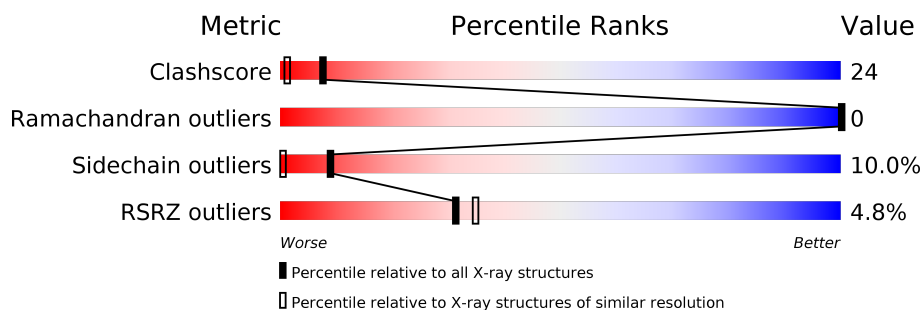
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.50 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))
Clashscore	112137	2503 (1.50-1.50)
Ramachandran outliers	110173	2445 (1.50-1.50)
Sidechain outliers	110143	2443 (1.50-1.50)
RSRZ outliers	101464	2305 (1.50-1.50)

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	<div> <div>2%</div> <div>55% 34% 10% .</div> </div>
1	C	141	<div> <div>3%</div> <div>43% 38% 15% 5%</div> </div>
2	B	146	<div> <div>%</div> <div>44% 42% 12% .</div> </div>
2	D	146	<div> <div>6%</div> <div>36% 37% 23% .</div> </div>

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
5	IHP	B	315	-	-	-	X

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 4910 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.

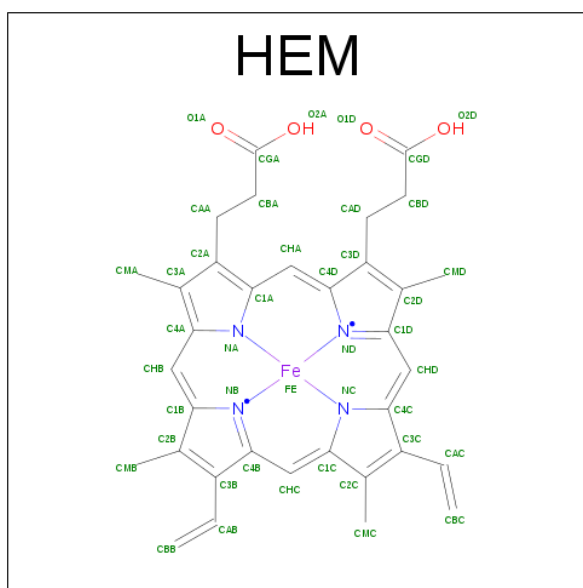
- Molecule 1 is a protein called HEMOGLOBIN A (OXY) (ALPHA CHAIN).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	141	Total	C	N	O	S	0	0	0
			1069	685	187	194	3			
1	C	141	Total	C	N	O	S	0	0	0
			1069	685	187	194	3			

- Molecule 2 is a protein called HEMOGLOBIN A (DEOXY) (BETA CHAIN).

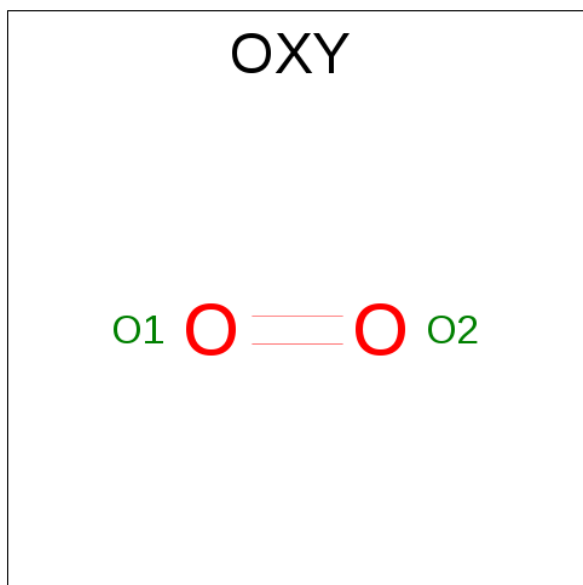
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	146	Total	C	N	O	S	0	0	0
			1123	724	195	201	3			
2	D	146	Total	C	N	O	S	0	0	0
			1123	724	195	201	3			

- Molecule 3 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula:  $C_{34}H_{32}FeN_4O_4$ ).



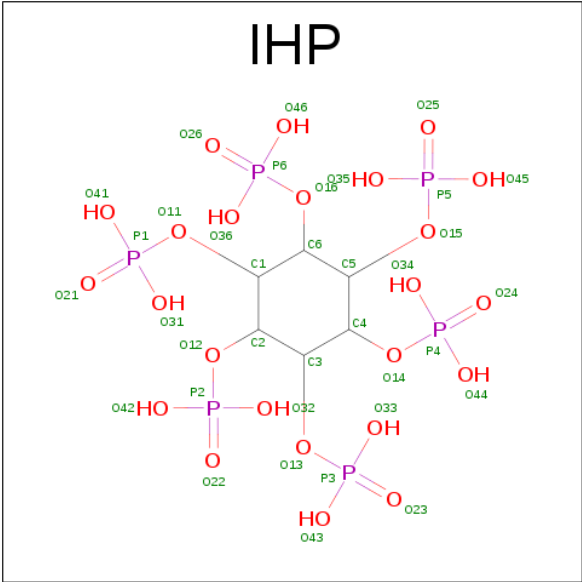
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
3	B	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
3	C	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
3	D	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		

- Molecule 4 is OXYGEN MOLECULE (three-letter code: OXY) (formula: O<sub>2</sub>).



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

- Molecule 5 is INOSITOL HEXAKISPHOSPHATE (three-letter code: IHP) (formula: C<sub>6</sub>H<sub>18</sub>O<sub>24</sub>P<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	B	1	Total	C	O	P	0	0
			36	6	24	6		

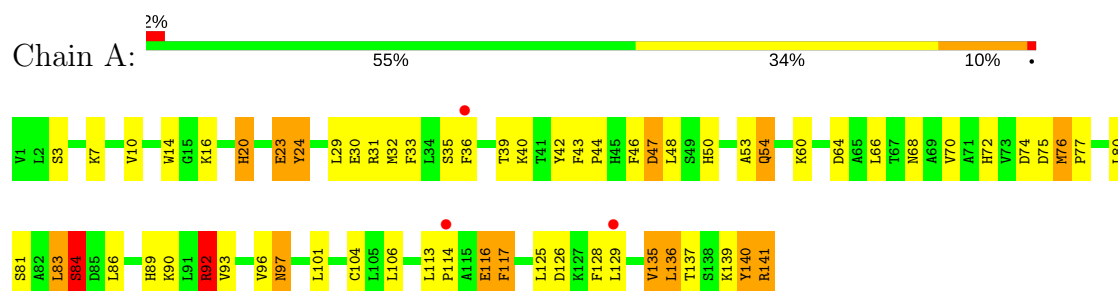
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	91	Total	O	0	0
			91	91		
6	B	91	Total	O	0	0
			91	91		
6	C	76	Total	O	0	0
			76	76		
6	D	56	Total	O	0	0
			56	56		

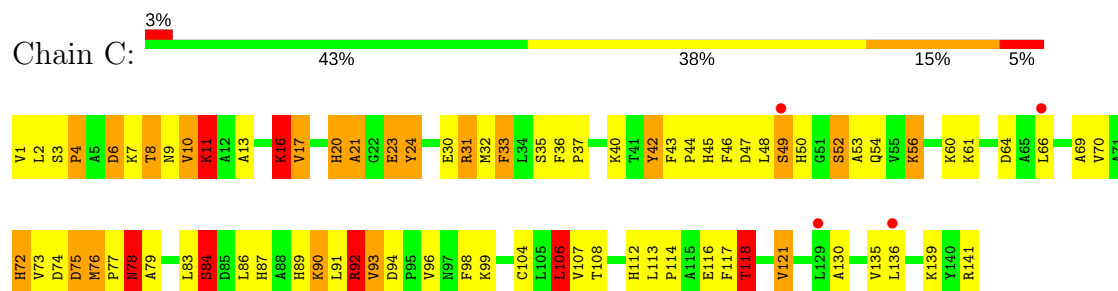
### 3 Residue-property plots

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.

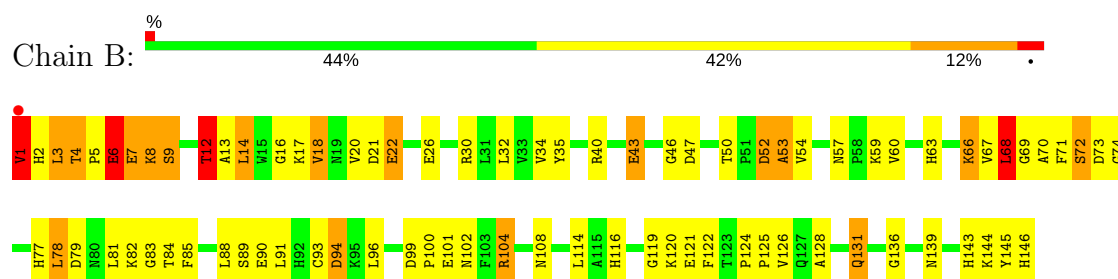
#### • Molecule 1: HEMOGLOBIN A (OXY) (ALPHA CHAIN)



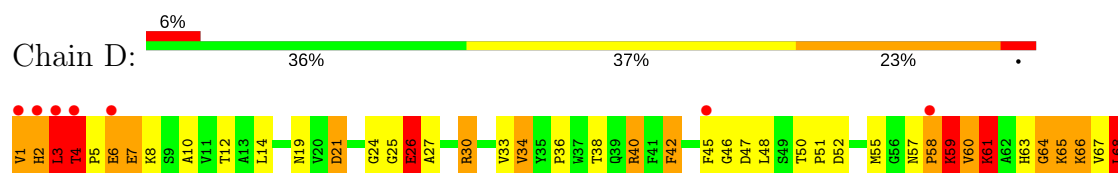
#### • Molecule 1: HEMOGLOBIN A (OXY) (ALPHA CHAIN)

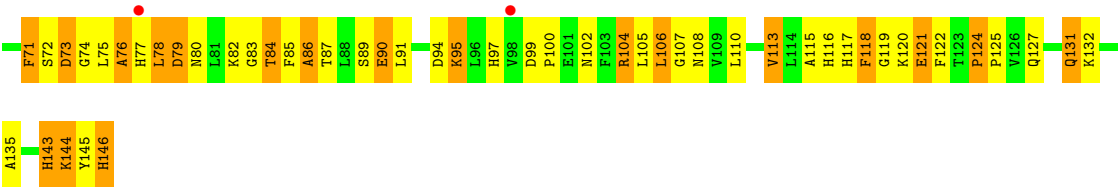


#### • Molecule 2: HEMOGLOBIN A (DEOXY) (BETA CHAIN)



#### • Molecule 2: HEMOGLOBIN A (DEOXY) (BETA CHAIN)







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	95.80Å 97.80Å 65.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	10.00 – 1.50 27.54 – 1.50	Depositor EDS
% Data completeness (in resolution range)	(Not available) (10.00-1.50) 81.3 (27.54-1.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.15 (at 1.50Å)	Xtriage
Refinement program	PROLSQ	Depositor
R, $R_{free}$	0.196 , (Not available) 0.203 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	21.5	Xtriage
Anisotropy	0.203	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 70.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.011 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4910	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.02% 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, IHP, OXY

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	1.34	3/1097 (0.3%)	2.40	66/1491 (4.4%)
1	C	1.22	0/1097	2.93	97/1491 (6.5%)
2	B	1.38	4/1153 (0.3%)	2.66	100/1566 (6.4%)
2	D	1.23	0/1153	2.82	99/1566 (6.3%)
All	All	1.29	7/4500 (0.2%)	2.71	362/6114 (5.9%)

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	A	0	1
1	C	0	1
2	B	0	1
2	D	1	2
All	All	1	5

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	6	GLU	CD-OE2	6.27	1.32	1.25
2	B	100	PRO	N-CD	6.14	1.56	1.47
2	B	35	TYR	CG-CD1	5.59	1.46	1.39
2	B	69	GLY	N-CA	5.44	1.54	1.46
1	A	23	GLU	CD-OE2	5.24	1.31	1.25

The worst 5 of 362 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	92	ARG	NE-CZ-NH2	27.82	134.21	120.30
2	D	2	HIS	CB-CA-C	23.96	158.31	110.40
1	C	30	GLU	OE1-CD-OE2	20.59	148.01	123.30
2	D	2	HIS	CA-CB-CG	18.75	145.48	113.60
1	C	141	ARG	NE-CZ-NH2	-18.66	110.97	120.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	D	2	HIS	CA

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	92	ARG	Sidechain
2	B	104	ARG	Sidechain
1	C	92	ARG	Sidechain
2	D	104	ARG	Sidechain
2	D	40	ARG	Sidechain

## 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	1069	0	1073	35	0
1	C	1069	0	1073	48	0
2	B	1123	0	1118	37	0
2	D	1123	0	1116	110	0
3	A	43	0	30	3	0
3	B	43	0	30	1	0
3	C	43	0	30	0	0
3	D	43	0	30	8	0
4	A	2	0	0	0	0
4	C	2	0	0	0	0
5	B	36	0	6	2	0
6	A	91	0	0	1	1
6	B	91	0	0	1	0
6	C	76	0	0	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	D	56	0	0	7	0
All	All	4910	0	4506	222	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 24.

The worst 5 of 222 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:1:VAL:CG2	2:D:3:LEU:HD12	1.57	1.33
2:D:60:VAL:HG12	6:D:191:HOH:O	0.95	1.12
2:D:4:THR:HG23	2:D:6:GLU:OE2	1.52	1.08
2:D:3:LEU:HD23	2:D:7:GLU:HB3	1.40	1.04
2:D:1:VAL:HG21	2:D:3:LEU:HD12	1.03	1.02

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)
6:A:234:HOH:O	6:A:234:HOH:O[2_665]	1.76	0.44

## 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	139/141 (99%)	136 (98%)	3 (2%)	0	100	100
1	C	139/141 (99%)	133 (96%)	6 (4%)	0	100	100
2	B	144/146 (99%)	141 (98%)	3 (2%)	0	100	100
2	D	144/146 (99%)	140 (97%)	4 (3%)	0	100	100
All	All	566/574 (99%)	550 (97%)	16 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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	113/113 (100%)	110 (97%)	3 (3%)	50	17
1	C	113/113 (100%)	101 (89%)	12 (11%)	8	0
2	B	118/118 (100%)	102 (86%)	16 (14%)	4	0
2	D	118/118 (100%)	103 (87%)	15 (13%)	5	0
All	All	462/462 (100%)	416 (90%)	46 (10%)	9	0

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	10	VAL
1	C	73	VAL
2	D	95	LYS
1	C	11	LYS
1	C	52	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	116	HIS
1	C	97	ASN
2	D	77	HIS
2	B	102	ASN
2	D	63	HIS

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

7 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	142	1,4	28,50,50	2.33	10 (35%)	17,82,82	2.87	12 (70%)
4	OXY	A	143	3	1,1,1	0.20	0	0,0,0	0.00	-
3	HEM	B	147	2	28,50,50	1.81	8 (28%)	17,82,82	2.83	7 (41%)
5	IHP	B	315	-	36,36,36	1.23	5 (13%)	54,60,60	1.06	2 (3%)
3	HEM	C	142	1,4	28,50,50	1.71	7 (25%)	17,82,82	2.12	7 (41%)
4	OXY	C	143	3	1,1,1	0.34	0	0,0,0	0.00	-
3	HEM	D	147	2	28,50,50	2.15	11 (39%)	17,82,82	4.01	8 (47%)

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	142	1,4	-	0/6/54/54	0/0/8/8
4	OXY	A	143	3	-	0/0/0/0	0/0/0/0
3	HEM	B	147	2	-	0/6/54/54	0/0/8/8
5	IHP	B	315	-	-	0/30/54/54	0/1/1/1
3	HEM	C	142	1,4	-	0/6/54/54	0/0/8/8
4	OXY	C	143	3	-	0/0/0/0	0/0/0/0
3	HEM	D	147	2	-	0/6/54/54	0/0/8/8

The worst 5 of 41 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	142	HEM	C3C-C2C	-5.30	1.33	1.40
3	C	142	HEM	C3B-C2B	-3.54	1.35	1.40
3	C	142	HEM	C3C-C2C	-3.38	1.35	1.40
3	A	142	HEM	C3B-C2B	-2.73	1.36	1.40
3	B	147	HEM	C3C-C2C	-2.66	1.36	1.40

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	147	HEM	CMA-C3A-C4A	-6.79	118.03	128.46
3	D	147	HEM	CBA-CAA-C2A	-6.73	99.61	112.48
3	D	147	HEM	CAA-CBA-CGA	-5.91	102.56	112.66
3	D	147	HEM	CMD-C2D-C1D	-5.63	119.81	128.46
3	D	147	HEM	CAD-CBD-CGD	-5.62	103.06	112.66

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 14 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	142	HEM	3	0
3	B	147	HEM	1	0
5	B	315	IHP	2	0
3	D	147	HEM	8	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	31/141 (21%)	1.16	3 (9%) 8 10	17, 20, 41, 48	0
1	C	86/141 (60%)	0.14	4 (4%) 32 36	16, 25, 47, 68	0
2	B	119/146 (81%)	-0.10	1 (0%) 86 88	14, 22, 51, 75	0
2	D	115/146 (78%)	0.39	9 (7%) 14 16	18, 31, 58, 75	0
All	All	351/574 (61%)	0.23	17 (4%) 31 35	14, 24, 54, 75	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	114	PRO	18.5
1	A	129	LEU	17.5
2	D	1	VAL	12.6
2	B	1	VAL	8.5
1	C	136	LEU	8.1

### 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( $\text{\AA}^2$ )	Q<0.9
5	IHP	B	315	36/36	0.66	0.69	3.17	20,20,20,20	0
4	OXY	C	143	2/2	0.94	0.12	0.79	21,21,21,32	1
3	HEM	B	147	43/43	0.98	0.08	0.13	11,16,33,53	0
3	HEM	C	142	43/43	0.98	0.08	-0.38	15,20,41,71	0
3	HEM	D	147	43/43	0.98	0.08	-0.68	17,27,76,77	0
4	OXY	A	143	2/2	-	-	-	25,25,25,36	1
3	HEM	A	142	43/43	0.88	0.31	-	17,23,49,77	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.