



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 04:19 AM GMT

PDB ID : 2MGJ  
Title : HIGH RESOLUTION CRYSTAL STRUCTURES OF FIVE DISTAL HISTIDINE MUTANTS OF SPERM WHALE MYOGLOBIN  
Authors : Quillin, M.L.; Arduini, R.M.; Phillips Jr., G.N.  
Deposited on : 1993-05-10  
Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

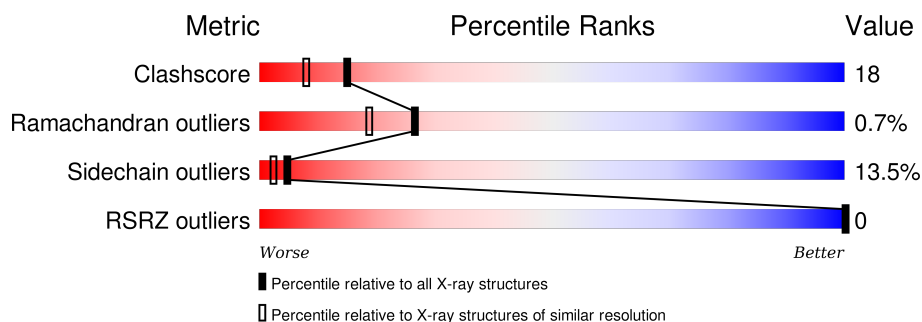
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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))
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (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 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	154	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 1392 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 MYOGLOBIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	154	Total	C	N	O	S	0	0	0
			1222	787	216	216	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	64	VAL	HIS	CONFLICT	UNP P02185
A	122	ASN	ASP	CONFLICT	UNP P02185

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 3 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: C<sub>34</sub>H<sub>32</sub>FeN<sub>4</sub>O<sub>4</sub>).



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

- Molecule 4 is water.

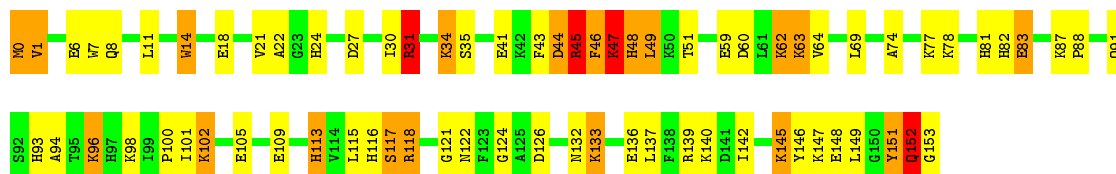
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	122	Total	O	0	0
			122	122		

### 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 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 ( $\text{RSRZ} > 2$ ). 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: MYOGLOBIN

Chain A: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 6	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	91.20 Å 91.20 Å 45.87 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	5.00 – 2.00 5.00 – 2.00	Depositor EDS
% Data completeness (in resolution range)	(Not available) (5.00-2.00) 95.1 (5.00-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$	-	Xtriage
Refinement program	X-PLOR	Depositor
R, $R_{free}$	0.156 , (Not available) 0.153 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	19.5	Xtriage
Anisotropy	0.066	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.80 , 145.7	EDS
Estimated twinning fraction	0.053 for h,-h-k,-l	Xtriage
L-test for twinning <sup>1</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 13216 reflections	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	1392	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows:

<sup>1</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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, SO4

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	0.93	4/1249 (0.3%)	1.36	17/1676 (1.0%)

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	A	0	37

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	81	HIS	CG-ND1	6.36	1.52	1.38
1	A	24	HIS	CG-CD2	5.74	1.45	1.35
1	A	113	HIS	CG-CD2	5.62	1.45	1.35
1	A	113	HIS	CB-CG	5.19	1.59	1.50

All (17) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	14	TRP	CD1-CG-CD2	8.60	113.18	106.30
1	A	118	ARG	NE-CZ-NH2	-7.99	116.30	120.30
1	A	7	TRP	CD1-CG-CD2	7.91	112.62	106.30
1	A	14	TRP	CE2-CD2-CG	-7.71	101.13	107.30
1	A	45	ARG	NE-CZ-NH1	-7.61	116.50	120.30
1	A	139	ARG	NE-CZ-NH2	-7.48	116.56	120.30
1	A	31	ARG	NE-CZ-NH1	7.10	123.85	120.30
1	A	7	TRP	CE2-CD2-CG	-6.97	101.72	107.30
1	A	31	ARG	NE-CZ-NH2	-6.74	116.93	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	21	VAL	CG1-CB-CG2	-6.61	100.33	110.90
1	A	48	HIS	ND1-CG-CD2	6.30	117.63	108.80
1	A	139	ARG	NE-CZ-NH1	5.93	123.27	120.30
1	A	81	HIS	CG-CD2-NE2	-5.60	98.55	109.20
1	A	48	HIS	CG-CD2-NE2	-5.53	98.70	109.20
1	A	14	TRP	CG-CD1-NE1	-5.43	104.67	110.10
1	A	117	SER	CB-CA-C	-5.31	100.01	110.10
1	A	14	TRP	CB-CG-CD1	-5.06	120.42	127.00

There are no chirality outliers.

All (37) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	0	MET	Mainchain
1	A	1	VAL	Mainchain
1	A	100	PRO	Mainchain
1	A	105	GLU	Sidechain
1	A	109	GLU	Sidechain
1	A	11	LEU	Mainchain
1	A	113	HIS	Sidechain
1	A	118	ARG	Sidechain,Mainchain
1	A	121	GLY	Mainchain
1	A	122	ASN	Mainchain
1	A	124	GLY	Mainchain
1	A	126	ASP	Sidechain
1	A	148	GLU	Sidechain,Mainchain
1	A	151	TYR	Sidechain,Mainchain
1	A	18	GLU	Sidechain
1	A	27	ASP	Sidechain
1	A	34	LYS	Mainchain
1	A	35	SER	Mainchain
1	A	43	PHE	Mainchain
1	A	44	ASP	Sidechain,Mainchain
1	A	45	ARG	Mainchain
1	A	46	PHE	Mainchain
1	A	47	LYS	Mainchain
1	A	48	HIS	Mainchain
1	A	49	LEU	Mainchain
1	A	60	ASP	Mainchain
1	A	69	LEU	Mainchain
1	A	74	ALA	Mainchain
1	A	77	LYS	Mainchain

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	78	LYS	Mainchain
1	A	82	HIS	Mainchain
1	A	94	ALA	Mainchain
1	A	96	LYS	Mainchain

## 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	1222	0	1255	44	0
2	A	5	0	0	1	0
3	A	43	0	30	2	0
4	A	122	0	0	13	1
All	All	1392	0	1285	45	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 18.

All (45) 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:102:LYS:HB3	4:A:312:HOH:O	1.41	1.17
1:A:14:TRP:HZ3	4:A:303:HOH:O	1.34	1.11
1:A:8:GLN:HG2	4:A:306:HOH:O	1.51	1.10
1:A:136:GLU:HB3	4:A:295:HOH:O	1.65	0.94
1:A:0:MET:HG3	1:A:137:LEU:HD13	1.48	0.93
1:A:101:ILE:HG12	1:A:153:GLY:OXT	1.76	0.83
1:A:132:ASN:O	1:A:136:GLU:HG3	1.80	0.81
1:A:145:LYS:HD3	1:A:145:LYS:N	1.95	0.79
1:A:152:GLN:C	4:A:266:HOH:O	2.28	0.72
1:A:83:GLU:CD	1:A:83:GLU:H	1.93	0.71
1:A:22:ALA:HB1	1:A:62:LYS:HD2	1.73	0.70
1:A:59:GLU:O	1:A:63:LYS:HG2	1.92	0.70
1:A:83:GLU:OE1	4:A:262:HOH:O	2.10	0.69
1:A:151:TYR:O	1:A:152:GLN:O	2.12	0.68
1:A:152:GLN:HA	1:A:152:GLN:OE1	1.89	0.63

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:6:GLU:OE1	1:A:133:LYS:HE2	2.01	0.61
1:A:45:ARG:HB2	4:A:241:HOH:O	2.02	0.59
1:A:14:TRP:CZ3	4:A:303:HOH:O	2.24	0.56
1:A:116:HIS:HD2	4:A:260:HOH:O	1.91	0.54
1:A:151:TYR:O	1:A:152:GLN:C	2.45	0.54
1:A:31:ARG:O	1:A:31:ARG:HD3	2.09	0.52
1:A:152:GLN:NE2	4:A:266:HOH:O	2.42	0.52
1:A:45:ARG:HH12	1:A:64:VAL:CG2	2.22	0.51
1:A:31:ARG:HD3	1:A:31:ARG:C	2.30	0.51
1:A:87:LYS:HB2	1:A:88:PRO:HD3	1.94	0.49
1:A:0:MET:HG3	1:A:137:LEU:CD1	2.31	0.49
1:A:91:GLN:HA	1:A:149:LEU:HD13	1.94	0.48
1:A:0:MET:C	1:A:1:VAL:HG23	2.33	0.48
3:A:154:HEM:HMC2	3:A:154:HEM:HBC2	1.96	0.47
1:A:0:MET:CG	1:A:137:LEU:HD13	2.33	0.47
1:A:146:TYR:HB2	1:A:152:GLN:OE1	2.14	0.47
1:A:41:GLU:HA	1:A:47:LYS:HG3	1.97	0.47
1:A:136:GLU:CD	4:A:295:HOH:O	2.52	0.47
1:A:83:GLU:CD	1:A:83:GLU:N	2.66	0.45
1:A:51:THR:HB	2:A:156:SO4:O4	2.16	0.44
1:A:47:LYS:HG2	4:A:274:HOH:O	2.17	0.44
1:A:116:HIS:HE1	4:A:223:HOH:O	2.00	0.43
1:A:45:ARG:NH1	3:A:154:HEM:CGD	2.82	0.42
1:A:59:GLU:O	1:A:63:LYS:CG	2.63	0.42
1:A:46:PHE:HB2	1:A:49:LEU:CD1	2.50	0.42
1:A:147:LYS:HG3	1:A:152:GLN:HG2	2.02	0.41
1:A:30:ILE:O	1:A:34:LYS:HG3	2.20	0.41
1:A:93:HIS:N	1:A:93:HIS:HD1	2.19	0.41
1:A:46:PHE:HB2	1:A:49:LEU:HD11	2.04	0.40
1:A:102:LYS:O	1:A:102:LYS:HG3	2.20	0.40

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)
4:A:236:HOH:O	4:A:236:HOH:O[2_655]	2.10	0.10

## 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	152/154 (99%)	143 (94%)	8 (5%)	1 (1%)	26	19

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	152	GLN

### 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	126/126 (100%)	109 (86%)	17 (14%)	5	2

All (17) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	31	ARG
1	A	44	ASP
1	A	45	ARG
1	A	47	LYS
1	A	62	LYS
1	A	63	LYS
1	A	83	GLU
1	A	96	LYS
1	A	98	LYS
1	A	102	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	115	LEU
1	A	117	SER
1	A	133	LYS
1	A	140	LYS
1	A	142	ILE
1	A	145	LYS
1	A	152	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	81	HIS
1	A	113	HIS
1	A	116	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

2 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	154	1	30,50,50	3.11	13 (43%)	24,82,82	2.36	8 (33%)
2	SO4	A	156	-	4,4,4	1.31	0	6,6,6	0.17	0

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	154	1	-	0/10/54/54	0/0/8/8
2	SO4	A	156	-	-	0/0/0/0	0/0/0/0

All (13) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	154	HEM	C3B-C4B	-9.35	1.43	1.51
3	A	154	HEM	C2D-C3D	-6.77	1.34	1.54
3	A	154	HEM	C1C-NC	-4.75	1.30	1.36
3	A	154	HEM	C3B-CAB	-4.62	1.42	1.51
3	A	154	HEM	C2C-C1C	-4.07	1.44	1.52
3	A	154	HEM	C2D-C1D	-3.48	1.40	1.51
3	A	154	HEM	C3D-C4D	-3.21	1.47	1.51
3	A	154	HEM	C2B-C1B	-2.21	1.44	1.51
3	A	154	HEM	CBB-CAB	2.61	1.44	1.29
3	A	154	HEM	CBC-CAC	2.69	1.44	1.29
3	A	154	HEM	CHC-C1C	2.85	1.43	1.36
3	A	154	HEM	CAA-C2A	3.67	1.58	1.52
3	A	154	HEM	C4C-NC	4.08	1.41	1.36

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	154	HEM	C1D-CHD-C4C	-2.46	121.71	125.82
3	A	154	HEM	CMD-C2D-C3D	2.62	125.96	114.35
3	A	154	HEM	CBA-CAA-C2A	3.47	118.74	112.53
3	A	154	HEM	CAD-C3D-C2D	3.75	124.00	113.22
3	A	154	HEM	CAD-C3D-C4D	3.79	125.84	112.47
3	A	154	HEM	C2D-C3D-C4D	4.20	108.63	101.50
3	A	154	HEM	CMC-C2C-C3C	4.72	128.32	116.53
3	A	154	HEM	CMB-C2B-C3B	4.85	128.65	116.53

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	154	HEM	2	0
2	A	156	SO4	1	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	154/154 (100%)	-0.96	0	100 100	11, 20, 37, 50	0

There are no RSRZ outliers to report.

### 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	SO4	A	156	5/5	0.97	0.09	0.44	51,51,51,51	0
3	HEM	A	154	43/43	0.98	0.06	-0.05	15,18,27,31	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.