



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 05:39 pm GMT

PDB ID : 2RI4  
Title : Crystal Structure determination of Goat Methemoglobin at 2.7 Angstrom  
Authors : Sathya Moorthy, P.; Neelagandan, K.; Balasubramanian, M.; Ponnuswamy, M.N.  
Deposited on : 2007-10-10  
Resolution : 2.70 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

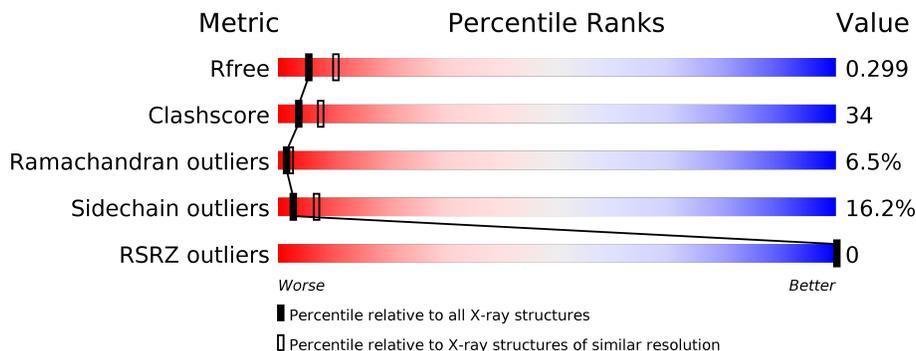
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (2.70-2.70)

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	I	141	
1	K	141	
2	B	145	
2	D	145	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	J	145	 45% 43% 10% ..
2	L	145	 41% 45% 14% .

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 8995 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-1/2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	139	1039	664	181	192	2	0	0	0
1	C	138	1032	659	180	191	2	0	0	0
1	I	137	1020	652	176	190	2	0	0	0
1	K	138	1032	659	180	191	2	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	142	1101	706	194	198	3	0	0	0
2	D	144	1121	720	196	201	4	0	0	0
2	J	144	1124	721	198	202	3	0	0	0
2	L	145	1131	726	199	202	4	0	0	0

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



*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
4	I	9	Total O 9 9	0	0
4	J	6	Total O 6 6	0	0
4	K	7	Total O 7 7	0	0
4	L	5	Total O 5 5	0	0



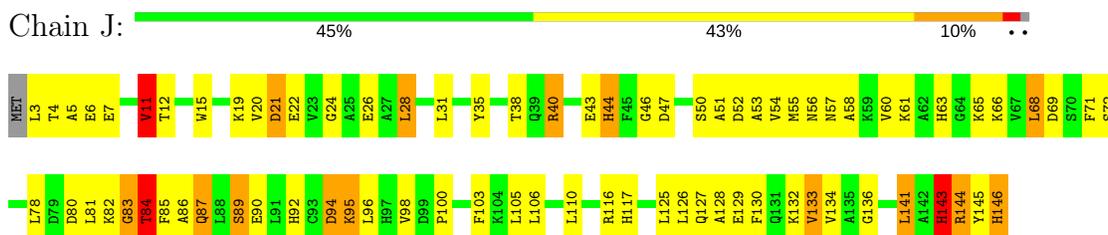
- Molecule 2: Hemoglobin subunit beta-A



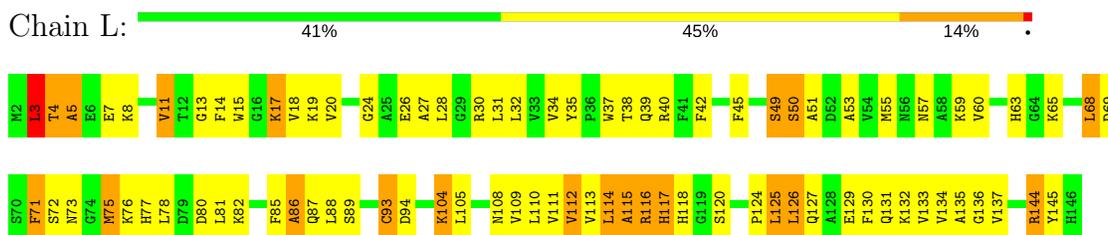
- Molecule 2: Hemoglobin subunit beta-A



- Molecule 2: Hemoglobin subunit beta-A



- Molecule 2: Hemoglobin subunit beta-A



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.69Å 68.33Å 95.60Å 110.64° 91.90° 108.94°	Depositor
Resolution (Å)	22.50 – 2.70 22.49 – 2.70	Depositor EDS
% Data completeness (in resolution range)	90.5 (22.50-2.70) 77.2 (22.49-2.70)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.33 (at 2.71Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.225 , 0.308 0.221 , 0.299	Depositor DCC
$R_{free}$ test set	2863 reflections (11.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	50.6	Xtrriage
Anisotropy	0.161	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 24.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	8995	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.47% 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	0.68	1/1064 (0.1%)	0.79	1/1446 (0.1%)
1	C	0.77	0/1057	0.85	0/1436
1	I	0.76	1/1043 (0.1%)	0.86	0/1417
1	K	0.63	0/1057	0.79	0/1436
2	B	0.71	0/1125	0.80	1/1520 (0.1%)
2	D	0.75	0/1146	0.85	0/1548
2	J	0.66	0/1150	0.78	0/1553
2	L	0.70	0/1157	0.79	1/1563 (0.1%)
All	All	0.71	2/8799 (0.0%)	0.81	3/11919 (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	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	I	111	CYS	CB-SG	-5.34	1.73	1.81
1	A	111	CYS	CB-SG	-5.29	1.73	1.81

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	L	3	LEU	CA-CB-CG	5.90	128.88	115.30
2	B	3	LEU	CA-CB-CG	5.17	127.18	115.30
1	A	34	LEU	CA-CB-CG	5.05	126.93	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1	VAL	Peptide

## 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	1039	0	1044	76	0
1	C	1032	0	1032	84	0
1	I	1020	0	1023	73	0
1	K	1032	0	1032	67	0
2	B	1101	0	1100	92	0
2	D	1121	0	1118	86	0
2	J	1124	0	1116	72	0
2	L	1131	0	1125	107	0
3	A	43	0	30	6	0
3	B	43	0	30	10	0
3	C	43	0	30	2	0
3	D	43	0	30	11	0
3	I	43	0	30	2	0
3	J	43	0	30	4	0
3	K	43	0	30	1	0
3	L	43	0	30	4	0
4	A	5	0	0	0	0
4	B	6	0	0	1	0
4	C	5	0	0	0	0
4	D	8	0	0	4	0
4	I	9	0	0	3	0
4	J	6	0	0	1	0
4	K	7	0	0	1	0
4	L	5	0	0	0	0
All	All	8995	0	8830	610	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 34.

The worst 5 of 610 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:B:39:GLN:HB3	1:C:92:ARG:NH1	1.53	1.22
2:B:59:LYS:HD2	2:B:59:LYS:N	1.54	1.19
3:D:147:HEM:HHA	3:D:147:HEM:HBD2	1.29	1.11
1:I:84:SER:HB3	1:I:138:SER:OG	1.55	1.04
2:L:71:PHE:HE2	2:L:137:VAL:HG21	1.17	1.01

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	137/141 (97%)	110 (80%)	18 (13%)	9 (7%)	1	2
1	C	136/141 (96%)	107 (79%)	20 (15%)	9 (7%)	1	2
1	I	133/141 (94%)	112 (84%)	14 (10%)	7 (5%)	2	4
1	K	136/141 (96%)	99 (73%)	26 (19%)	11 (8%)	1	1
2	B	140/145 (97%)	116 (83%)	17 (12%)	7 (5%)	2	4
2	D	142/145 (98%)	118 (83%)	14 (10%)	10 (7%)	1	2
2	J	142/145 (98%)	108 (76%)	25 (18%)	9 (6%)	1	2
2	L	143/145 (99%)	116 (81%)	17 (12%)	10 (7%)	1	2
All	All	1109/1144 (97%)	886 (80%)	151 (14%)	72 (6%)	1	2

5 of 72 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	56	LYS
1	A	85	ASP
2	B	5	ALA
2	B	129	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	43	PHE

### 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	111/113 (98%)	91 (82%)	20 (18%)	2 5
1	C	110/113 (97%)	95 (86%)	15 (14%)	4 10
1	I	109/113 (96%)	96 (88%)	13 (12%)	6 14
1	K	110/113 (97%)	92 (84%)	18 (16%)	2 7
2	B	115/118 (98%)	90 (78%)	25 (22%)	1 3
2	D	117/118 (99%)	101 (86%)	16 (14%)	4 10
2	J	117/118 (99%)	96 (82%)	21 (18%)	2 5
2	L	118/118 (100%)	99 (84%)	19 (16%)	3 7
All	All	907/924 (98%)	760 (84%)	147 (16%)	3 7

5 of 147 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	40	ARG
1	I	74	ASP
2	L	71	PHE
2	D	48	LEU
2	D	110	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	I	45	HIS
2	J	87	GLN
2	L	77	HIS
1	I	97	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	J	108	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](#)

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	142	1	28,50,50	2.27	7 (25%)	17,82,82	2.38	5 (29%)
3	HEM	B	147	2	28,50,50	2.20	6 (21%)	17,82,82	1.66	3 (17%)
3	HEM	C	142	1,4	28,50,50	2.31	7 (25%)	17,82,82	1.79	2 (11%)
3	HEM	D	147	2,4	28,50,50	2.44	10 (35%)	17,82,82	1.35	3 (17%)
3	HEM	I	142	1,4	28,50,50	2.31	9 (32%)	17,82,82	1.91	4 (23%)
3	HEM	J	147	2,4	28,50,50	2.39	8 (28%)	17,82,82	1.49	4 (23%)
3	HEM	K	142	1	28,50,50	2.28	7 (25%)	17,82,82	2.60	7 (41%)
3	HEM	L	147	2	28,50,50	2.18	8 (28%)	17,82,82	1.79	3 (17%)

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	-	0/6/54/54	0/0/8/8
3	HEM	B	147	2	-	0/6/54/54	0/0/8/8
3	HEM	C	142	1,4	-	0/6/54/54	0/0/8/8
3	HEM	D	147	2,4	-	2/6/54/54	0/0/8/8
3	HEM	I	142	1,4	-	0/6/54/54	0/0/8/8
3	HEM	J	147	2,4	-	0/6/54/54	0/0/8/8
3	HEM	K	142	1	-	0/6/54/54	0/0/8/8
3	HEM	L	147	2	-	0/6/54/54	0/0/8/8

The worst 5 of 62 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	142	HEM	C3B-C2B	-5.83	1.32	1.40
3	A	142	HEM	C3C-C2C	-5.26	1.33	1.40
3	I	142	HEM	C3C-C2C	-5.05	1.33	1.40
3	C	142	HEM	C3C-C2C	-4.95	1.33	1.40
3	K	142	HEM	C3C-C2C	-4.79	1.34	1.40

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	K	142	HEM	CAA-CBA-CGA	-5.68	102.96	112.66
3	A	142	HEM	CAA-CBA-CGA	-5.46	103.33	112.66
3	A	142	HEM	CAD-CBD-CGD	-5.12	103.91	112.66
3	I	142	HEM	CMA-C3A-C4A	-4.81	121.07	128.46
3	C	142	HEM	CAA-CBA-CGA	-4.72	104.59	112.66

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	D	147	HEM	C4D-C3D-CAD-CBD
3	D	147	HEM	C2D-C3D-CAD-CBD

There are no ring outliers.

8 monomers are involved in 40 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	142	HEM	6	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	147	HEM	10	0
3	C	142	HEM	2	0
3	D	147	HEM	11	0
3	I	142	HEM	2	0
3	J	147	HEM	4	0
3	K	142	HEM	1	0
3	L	147	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	139/141 (98%)	-0.70	0 100 100	29, 42, 55, 60	0
1	C	138/141 (97%)	-0.81	0 100 100	21, 31, 47, 55	0
1	I	137/141 (97%)	-0.85	0 100 100	22, 33, 48, 53	0
1	K	138/141 (97%)	-0.67	0 100 100	28, 47, 60, 67	0
2	B	142/145 (97%)	-0.61	0 100 100	26, 45, 69, 70	0
2	D	144/145 (99%)	-0.75	0 100 100	22, 38, 60, 67	0
2	J	144/145 (99%)	-0.67	0 100 100	24, 45, 63, 71	0
2	L	145/145 (100%)	-0.69	0 100 100	22, 43, 64, 76	0
All	All	1127/1144 (98%)	-0.72	0 100 100	21, 41, 62, 76	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( $\text{\AA}^2$ )	Q<0.9
3	HEM	J	147	43/43	0.96	0.16	1.84	50,55,58,59	0
3	HEM	C	142	43/43	0.98	0.12	0.08	25,33,41,46	0
3	HEM	A	142	43/43	0.97	0.12	-0.15	25,32,42,45	0
3	HEM	L	147	43/43	0.96	0.13	-0.24	34,40,48,53	0
3	HEM	B	147	43/43	0.96	0.12	-0.32	48,53,56,57	0
3	HEM	D	147	43/43	0.97	0.11	-0.46	28,37,39,39	0
3	HEM	I	142	43/43	0.98	0.11	-0.56	23,29,37,41	0
3	HEM	K	142	43/43	0.97	0.12	-0.59	41,44,46,50	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.