Full wwPDB X-ray Structure Validation Report (i)

Jan 31, 2016 – 07:53 PM GMT

PDB ID	:	1HSY
Title	:	ORIGIN OF THE PH-DEPENDENT SPECTROSCOPIC PROPERTIES OF PENTACOORDINATE METMYOGLOBIN VARIANTS
Authors	:	Maurus, R.; Brayer, G.D.
Deposited on	:	1994-12-12
Resolution	:	1.90 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
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

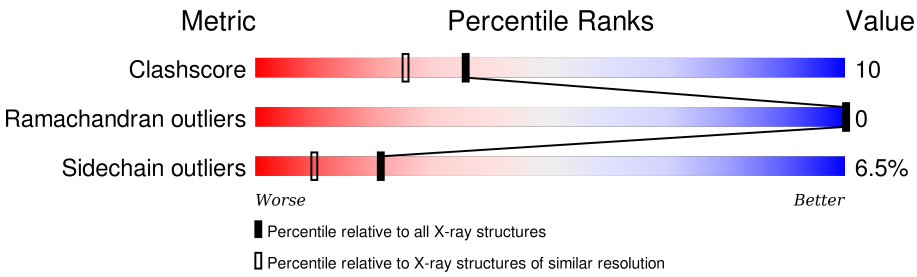
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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))
Clashscore	102246	5398 (1.90-1.90)
Ramachandran outliers	100387	5338 (1.90-1.90)
Sidechain outliers	100360	5339 (1.90-1.90)

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 >=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 <=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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	153	71% 25% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1297 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.

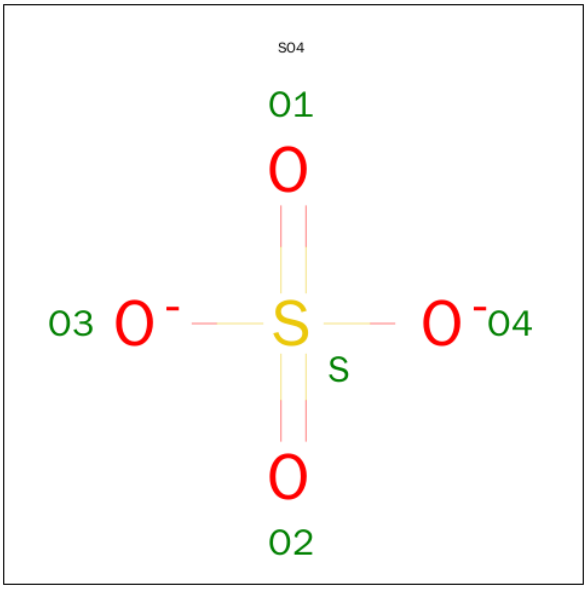
- Molecule 1 is a protein called MYOGLOBIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	153	1196	767	208	219	2	0	0	0

There is a discrepancy between the modelled and reference sequences:

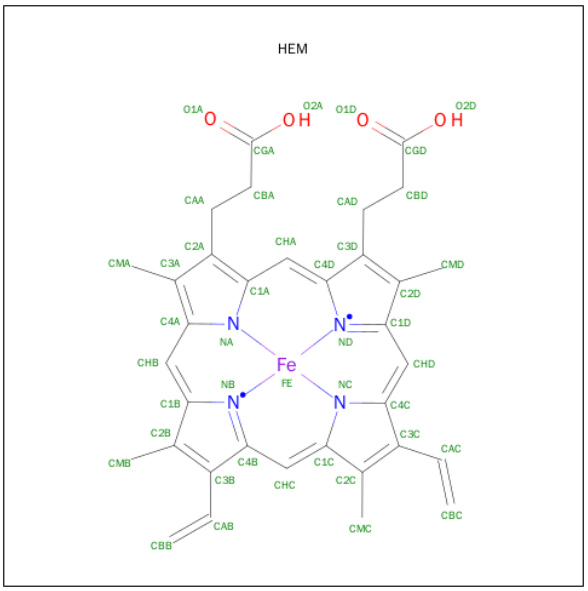
Chain	Residue	Modelled	Actual	Comment	Reference
A	64	THR	HIS	CONFLICT	UNP P68082

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
2	A	1	5	4	1	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	43	34	1	4	4	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	53	53	53	0	0

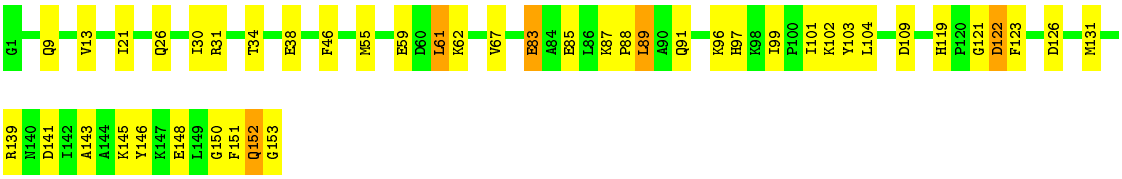
3 Residue-property plots (i)

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.

Note EDS was not executed.

- Molecule 1: MYOGLOBIN

Chain A:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	64.40Å 28.86Å 35.79Å 90.00° 107.20° 90.00°	Depositor
Resolution (Å)	6.00 – 1.90	Depositor
% Data completeness (in resolution range)	71.5 (6.00-1.90)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, R_{free}	0.163 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1297	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

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, 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	0.89	0/1222	1.76	25/1640 (1.5%)

There are no bond length outliers.

All (25) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	31	ARG	NE-CZ-NH1	13.27	126.93	120.30
1	A	126	ASP	CB-CG-OD1	9.63	126.97	118.30
1	A	139	ARG	NE-CZ-NH1	8.21	124.40	120.30
1	A	109	ASP	CB-CG-OD1	7.90	125.41	118.30
1	A	123	PHE	CB-CG-CD2	-7.75	115.37	120.80
1	A	122	ASP	CB-CG-OD1	-7.26	111.77	118.30
1	A	146	TYR	CB-CG-CD1	7.21	125.33	121.00
1	A	31	ARG	NE-CZ-NH2	-7.15	116.73	120.30
1	A	139	ARG	CD-NE-CZ	6.75	133.06	123.60
1	A	67	VAL	CA-CB-CG1	6.69	120.93	110.90
1	A	61	LEU	CA-CB-CG	6.62	130.52	115.30
1	A	103	TYR	CB-CG-CD2	-6.62	117.03	121.00
1	A	148	GLU	O-C-N	6.26	132.72	122.70
1	A	146	TYR	CB-CG-CD2	-6.26	117.25	121.00
1	A	103	TYR	CB-CG-CD1	6.10	124.66	121.00
1	A	38	GLU	OE1-CD-OE2	-5.88	116.24	123.30
1	A	141	ASP	CB-CG-OD1	5.84	123.56	118.30
1	A	83	GLU	CA-CB-CG	5.79	126.13	113.40
1	A	85	GLU	OE1-CD-OE2	-5.76	116.39	123.30
1	A	139	ARG	NE-CZ-NH2	-5.70	117.45	120.30
1	A	89	LEU	CA-CB-CG	5.67	128.35	115.30
1	A	121	GLY	C-N-CA	-5.41	108.17	121.70
1	A	46	PHE	CB-CG-CD1	5.40	124.58	120.80
1	A	85	GLU	CG-CD-OE1	5.38	129.06	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	151	PHE	C-N-CA	5.06	134.35	121.70

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	1196	0	1212	24	0
2	A	5	0	0	0	0
3	A	43	0	30	1	0
4	A	53	0	0	1	0
All	All	1297	0	1242	24	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 10.

All (24) 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:30:ILE:HG23	1:A:55:MET:HE2	1.68	0.74
1:A:101:ILE:HD13	1:A:143:ALA:HB2	1.70	0.74
1:A:87:LYS:HG2	1:A:91:GLN:OE1	2.02	0.60
1:A:30:ILE:HA	1:A:55:MET:CE	2.33	0.58
1:A:9:GLN:O	1:A:13:VAL:HG23	2.07	0.54
1:A:87:LYS:HB3	1:A:88:PRO:CD	2.38	0.54
1:A:34:THR:HG23	1:A:55:MET:HE1	1.91	0.52
1:A:26:GLN:O	1:A:30:ILE:HG13	2.11	0.51
1:A:101:ILE:CD1	1:A:143:ALA:HB2	2.41	0.51
1:A:119:HIS:HB3	1:A:122:ASP:OD1	2.11	0.50
1:A:30:ILE:HG12	1:A:55:MET:HE3	1.94	0.49
1:A:87:LYS:HB3	1:A:88:PRO:HD3	1.93	0.49
1:A:87:LYS:HE2	1:A:91:GLN:OE1	2.12	0.48
1:A:152:GLN:HG3	1:A:153:GLY:H	1.79	0.48
1:A:30:ILE:HA	1:A:55:MET:HE3	1.96	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:30:ILE:HG23	1:A:55:MET:CE	2.43	0.44
1:A:59:GLU:OE1	1:A:62:LYS:HE3	2.17	0.44
1:A:150:GLY:O	1:A:152:GLN:N	2.42	0.44
1:A:99:ILE:HD12	3:A:154:HEM:HAC	1.99	0.43
1:A:97:HIS:HB2	1:A:99:ILE:HG12	2.01	0.43
1:A:26:GLN:HG3	4:A:217:HOH:O	2.18	0.43
1:A:13:VAL:HB	1:A:131:MET:HE3	2.02	0.41
1:A:101:ILE:HA	1:A:104:LEU:HD12	2.03	0.41
1:A:87:LYS:N	1:A:88:PRO:HD2	2.35	0.41

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	151/153 (99%)	145 (96%)	6 (4%)	0	100 100

There are no Ramachandran outliers to report.

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	123/123 (100%)	115 (94%)	8 (6%)	21 10

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	21	ILE
1	A	61	LEU
1	A	83	GLU
1	A	89	LEU
1	A	96	LYS
1	A	102	LYS
1	A	145	LYS
1	A	152	GLN

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

Mol	Chain	Res	Type
1	A	128	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)

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	2.59	8 (26%)	24,82,82	2.68	9 (37%)
2	SO4	A	155	-	4,4,4	1.05	0	6,6,6	0.55	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	155	-	-	0/0/0/0	0/0/0/0

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	154	HEM	C3B-C4B	-7.88	1.44	1.51
3	A	154	HEM	C2D-C3D	-6.52	1.35	1.54
3	A	154	HEM	C3D-C4D	-5.89	1.44	1.51
3	A	154	HEM	C2C-C1C	-3.94	1.45	1.52
3	A	154	HEM	C2B-C1B	-2.14	1.44	1.51
3	A	154	HEM	FE-NC	2.16	2.04	1.95
3	A	154	HEM	CAA-C2A	2.61	1.56	1.52
3	A	154	HEM	C1C-NC	2.96	1.39	1.36

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	154	HEM	CAA-C2A-C1A	-4.60	122.01	127.01
3	A	154	HEM	CBD-CAD-C3D	-3.23	104.16	113.55
3	A	154	HEM	C1D-CHD-C4C	-2.99	120.83	125.82
3	A	154	HEM	CMD-C2D-C3D	2.91	127.23	114.35
3	A	154	HEM	CAD-C3D-C4D	3.38	124.38	112.47
3	A	154	HEM	CAD-C3D-C2D	4.37	125.78	113.22
3	A	154	HEM	C2D-C3D-C4D	4.70	109.47	101.50
3	A	154	HEM	CMC-C2C-C3C	4.71	128.30	116.53
3	A	154	HEM	CMB-C2B-C3B	5.72	130.82	116.53

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	154	HEM	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)

EDS was not executed - this section will therefore be empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.