



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 12, 2017 – 11:31 pm GMT

PDB ID : 1AC8  
Title : VARIATION IN THE STRENGTH OF A CH TO O HYDROGEN BOND IN AN ARTIFICIAL PROTEIN CAVITY (3,4,5-TRIMETHYLTHIAZOLE)  
Authors : Musah, R.A.; Jensen, G.M.; Bunte, S.W.; Rosenfeld, R.; Mcree, D.E.; Goodin, D.B.  
Deposited on : 1997-02-14  
Resolution : 2.10 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 : 1.7.2 (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

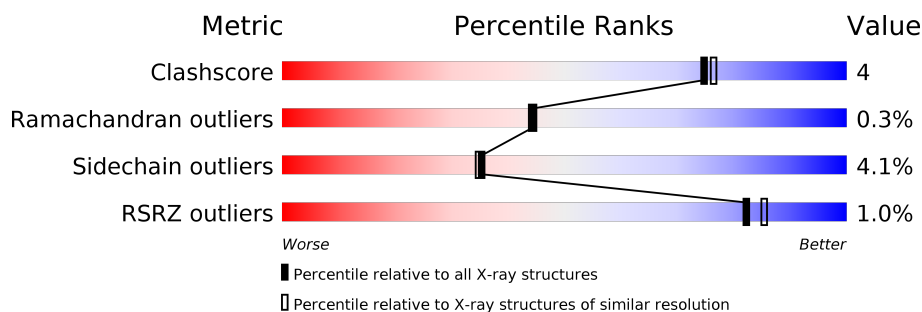
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (2.10-2.10)

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	294	<div> <div></div> <div>84%</div> <div>14%</div> <div>..</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
3	TMZ	A	296	-	-	X	-

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 2992 atoms, of which 503 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 CYTOCHROME C PEROXIDASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	291	Total	C	H	N	O	S	0	0	0
			2841	1492	503	388	452	6			

There are 4 discrepancies between the modelled and reference sequences:

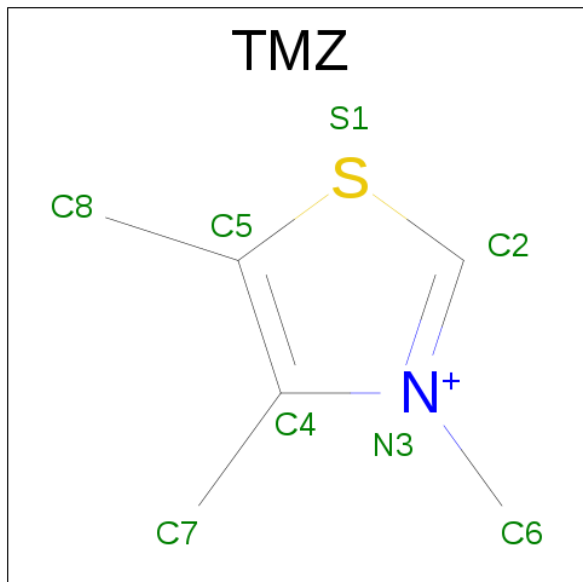
Chain	Residue	Modelled	Actual	Comment	Reference
A	53	ILE	THR	SEE REMARK 999	UNP P00431
A	152	GLY	ASP	SEE REMARK 999	UNP P00431
A	191	GLY	TRP	ENGINEERED	UNP P00431
A	272	ASP	ASN	SEE REMARK 999	UNP P00431

- Molecule 2 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula:  $C_{34}H_{32}FeN_4O_4$ ).



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

- Molecule 3 is 3,4,5-TRIMETHYL-1,3-THIAZOLE (three-letter code: TMZ) (formula:  $C_6H_{10}NS$ ).



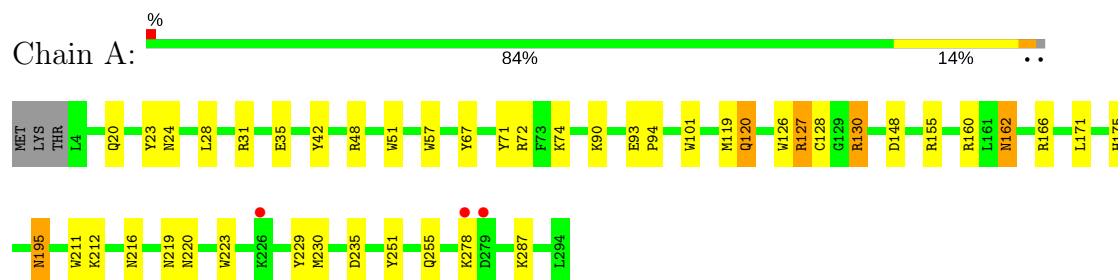
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	S	0	0
			8	6	1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	100	Total	O	0	0
			100	100		



● Molecule 1: CYTOCHROME C PEROXIDASE



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	108.00Å 77.30Å 51.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	7.00 – 2.10 21.10 – 1.95	Depositor EDS
% Data completeness (in resolution range)	79.0 (7.00-2.10) 66.3 (21.10-1.95)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.53 (at 1.95Å)	Xtriage
Refinement program	XTALVIEW	Depositor
R, $R_{free}$	(Not available) , (Not available) 0.208 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	15.8	Xtriage
Anisotropy	0.788	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 61.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	2992	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.54% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HEM, TMZ

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	0.83	1/2402 (0.0%)	1.58	46/3250 (1.4%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	42	TYR	CD2-CE2	-6.72	1.29	1.39

All (46) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	166	ARG	NE-CZ-NH2	-12.12	114.24	120.30
1	A	48	ARG	NE-CZ-NH1	9.75	125.17	120.30
1	A	31	ARG	NE-CZ-NH2	-9.32	115.64	120.30
1	A	127	ARG	NE-CZ-NH1	8.51	124.55	120.30
1	A	101	TRP	CD1-CG-CD2	8.10	112.78	106.30
1	A	126	TRP	CD1-CG-CD2	8.10	112.78	106.30
1	A	127	ARG	NE-CZ-NH2	-7.92	116.34	120.30
1	A	51	TRP	CE2-CD2-CG	-7.74	101.11	107.30
1	A	57	TRP	CD1-CG-CD2	7.70	112.46	106.30
1	A	126	TRP	CE2-CD2-CG	-7.65	101.18	107.30
1	A	223	TRP	CE2-CD2-CG	-7.58	101.23	107.30
1	A	71	TYR	CB-CG-CD2	-7.48	116.51	121.00
1	A	127	ARG	CB-CG-CD	-7.38	92.40	111.60
1	A	223	TRP	CD1-CG-CD2	7.35	112.18	106.30
1	A	51	TRP	CB-CG-CD1	-7.33	117.47	127.00
1	A	51	TRP	CD1-CG-CD2	7.30	112.14	106.30
1	A	229	TYR	CB-CG-CD2	-7.30	116.62	121.00
1	A	130	ARG	NE-CZ-NH2	-7.21	116.69	120.30
1	A	162	ASN	CA-C-N	-7.09	101.60	117.20
1	A	101	TRP	CE2-CD2-CG	-6.90	101.78	107.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	211	TRP	CE2-CD2-CG	-6.58	102.03	107.30
1	A	211	TRP	CD1-CG-CD2	6.58	111.56	106.30
1	A	160	ARG	NE-CZ-NH1	6.52	123.56	120.30
1	A	57	TRP	CE2-CD2-CG	-6.46	102.13	107.30
1	A	51	TRP	CG-CD2-CE3	6.42	139.68	133.90
1	A	162	ASN	N-CA-C	6.29	128.00	111.00
1	A	160	ARG	NE-CZ-NH2	-6.21	117.20	120.30
1	A	31	ARG	NE-CZ-NH1	6.03	123.31	120.30
1	A	42	TYR	CB-CG-CD1	-6.03	117.38	121.00
1	A	130	ARG	NE-CZ-NH1	6.02	123.31	120.30
1	A	130	ARG	CG-CD-NE	-6.01	99.18	111.80
1	A	251	TYR	CB-CG-CD2	-5.71	117.57	121.00
1	A	126	TRP	CG-CD2-CE3	5.66	138.99	133.90
1	A	148	ASP	CB-CG-OD1	5.64	123.38	118.30
1	A	101	TRP	CG-CD1-NE1	-5.56	104.54	110.10
1	A	57	TRP	CG-CD1-NE1	-5.54	104.56	110.10
1	A	255	GLN	CG-CD-NE2	5.24	129.27	116.70
1	A	48	ARG	NE-CZ-NH2	-5.23	117.68	120.30
1	A	155	ARG	NE-CZ-NH1	5.22	122.91	120.30
1	A	90	LYS	CA-CB-CG	-5.21	101.94	113.40
1	A	230	MET	CA-CB-CG	5.16	122.06	113.30
1	A	126	TRP	CG-CD1-NE1	-5.08	105.02	110.10
1	A	166	ARG	CG-CD-NE	-5.08	101.13	111.80
1	A	255	GLN	OE1-CD-NE2	-5.08	110.22	121.90
1	A	162	ASN	CA-C-O	5.07	130.74	120.10
1	A	223	TRP	CG-CD2-CE3	5.02	138.42	133.90

There are no chirality outliers.

There are no planarity outliers.

## 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	2338	503	2209	15	0
2	A	43	0	30	2	0
3	A	8	0	10	4	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	100	0	0	4	0
All	All	2489	503	2249	17	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 4.

All (17) 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:128:CYS:HA	4:A:326:HOH:O	1.62	0.99
1:A:235:ASP:OD2	3:A:296:TMZ:H2	1.94	0.67
1:A:175:HIS:O	3:A:296:TMZ:H72	1.96	0.66
1:A:130:ARG:NE	4:A:442:HOH:O	2.37	0.57
1:A:20:GLN:HE22	1:A:287:LYS:H	1.52	0.57
1:A:195:ASN:HD22	1:A:195:ASN:H	1.54	0.55
1:A:67:TYR:HA	1:A:130:ARG:HG2	1.89	0.53
1:A:130:ARG:CZ	4:A:442:HOH:O	2.57	0.52
1:A:119:MET:O	1:A:120:GLN:HG2	2.11	0.50
1:A:130:ARG:NH2	4:A:442:HOH:O	2.45	0.49
1:A:93:GLU:HB3	1:A:94:PRO:HD2	1.97	0.46
1:A:195:ASN:ND2	1:A:195:ASN:H	2.14	0.44
2:A:295:HEM:HAD2	3:A:296:TMZ:H83	2.01	0.43
1:A:23:TYR:HD1	1:A:24:ASN:HD22	1.66	0.42
1:A:216:ASN:HD21	1:A:220:ASN:HB2	1.84	0.42
2:A:295:HEM:HAD2	3:A:296:TMZ:C8	2.50	0.41
1:A:127:ARG:HD3	1:A:127:ARG:HH11	1.78	0.41

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	289/294 (98%)	283 (98%)	5 (2%)	1 (0%)	44	44

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	162	ASN

### 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	246/251 (98%)	236 (96%)	10 (4%)	35	35

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	28	LEU
1	A	35	GLU
1	A	72	ARG
1	A	74	LYS
1	A	120	GLN
1	A	171	LEU
1	A	195	ASN
1	A	212	LYS
1	A	219	ASN
1	A	278	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	20	GLN
1	A	24	ASN
1	A	78	ASN
1	A	195	ASN
1	A	220	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	240	GLN
1	A	292	GLN

### 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 ⓘ

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	$\# Z  > 2$
2	HEM	A	295	1,4	28,50,50	1.67	7 (25%)	17,82,82	1.67	3 (17%)
3	TMZ	A	296	-	3,8,8	3.52	2 (66%)	4,11,11	2.99	2 (50%)

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
2	HEM	A	295	1,4	-	0/6/54/54	0/0/8/8
3	TMZ	A	296	-	-	0/0/0/0	0/1/1/1

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	296	TMZ	C6-N3	-5.60	1.35	1.47
2	A	295	HEM	C3B-CAB	-3.95	1.40	1.47
2	A	295	HEM	C1B-NB	-3.52	1.32	1.36
2	A	295	HEM	C3C-CAC	-3.02	1.41	1.47
2	A	295	HEM	C4D-ND	-2.81	1.33	1.36
2	A	295	HEM	C4C-NC	-2.53	1.33	1.36
3	A	296	TMZ	C4-N3	2.35	1.42	1.39
2	A	295	HEM	CBB-CAB	2.45	1.46	1.28
2	A	295	HEM	CBC-CAC	2.46	1.46	1.28

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	296	TMZ	C7-C4-N3	-4.75	116.53	122.53
2	A	295	HEM	C1D-C2D-C3D	-2.41	105.32	107.00
2	A	295	HEM	CMC-C2C-C3C	3.24	130.91	124.89
3	A	296	TMZ	C5-C4-N3	3.58	114.96	107.38
2	A	295	HEM	CMB-C2B-C3B	4.13	132.55	124.89

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	295	HEM	2	0
3	A	296	TMZ	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	291/294 (98%)	-0.25	3 (1%) 82 85	5, 16, 34, 54	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	278	LYS	2.7
1	A	279	ASP	2.7
1	A	226	LYS	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
2	HEM	A	295	43/43	0.97	0.10	-0.50	3,9,12,16	0
3	TMZ	A	296	8/8	0.97	0.11	-0.61	15,15,15,15	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.