



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 18, 2023 – 09:09 PM EDT

PDB ID : 5CBB  
Title : Crystal structure of Mycobacterium tuberculosis malate synthase in complex with 5-(3H-indol-3-ylidene)-2,5-dihydro-1H-pyrazole-3-carboxylate  
Authors : Huang, H.-L.; Sacchettini, J.C.  
Deposited on : 2015-06-30  
Resolution : 2.01 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35.1  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.35.1

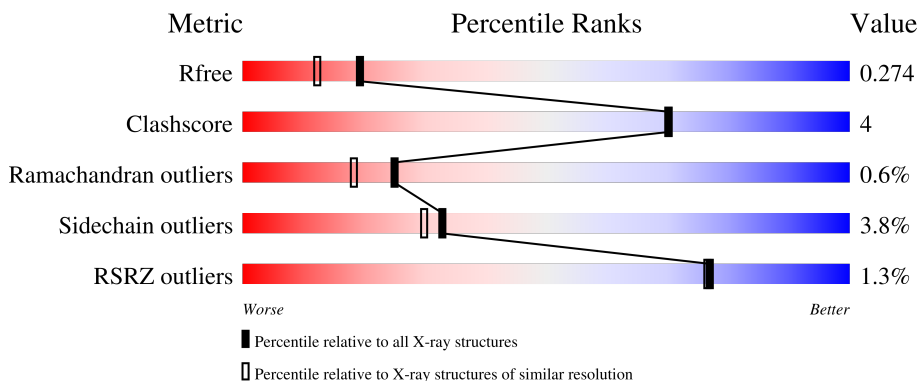
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.01 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))
$R_{free}$	130704	10434 (2.04-2.00)
Clashscore	141614	11643 (2.04-2.00)
Ramachandran outliers	138981	11493 (2.04-2.00)
Sidechain outliers	138945	11492 (2.04-2.00)
RSRZ outliers	127900	10220 (2.04-2.00)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	741	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 5994 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 Malate synthase G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	713	5454	3422	962	1048	22	0	1	0

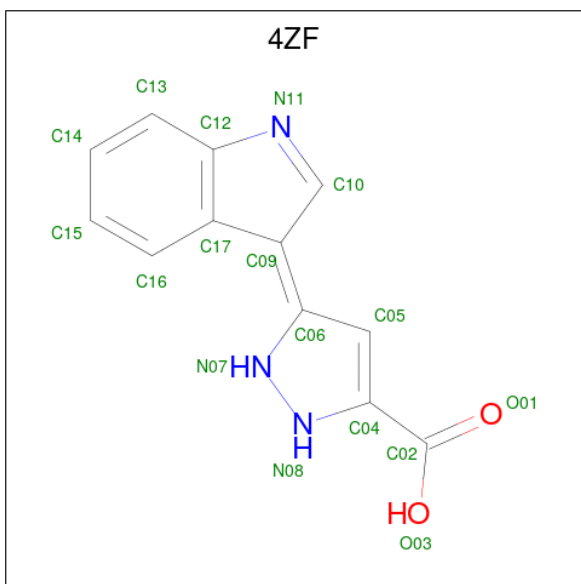
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	619	ALA	CYS	engineered mutation	UNP P9WK17

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Mg	0	0
			1	1		

- Molecule 3 is (5E)-5-(3H-indol-3-ylidene)-2,5-dihydro-1H-pyrazole-3-carboxylic acid (three-letter code: 4ZF) (formula: C<sub>12</sub>H<sub>9</sub>N<sub>3</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	17	12	3	2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	522	522	522	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	80.82Å 80.82Å 225.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.53 – 2.01 45.53 – 2.01	Depositor EDS
% Data completeness (in resolution range)	99.7 (45.53-2.01) 99.8 (45.53-2.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.03	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.03 (at 2.01Å)	Xtrriage
Refinement program	PHENIX 1.8.2_1309	Depositor
R, $R_{free}$	0.214 , 0.275 0.215 , 0.274	Depositor DCC
$R_{free}$ test set	2568 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	34.5	Xtrriage
Anisotropy	0.597	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 36.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	5994	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.45% 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: 4ZF, MG

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	0.36	0/5559	0.53	1/7558 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	151	THR	N-CA-C	5.08	124.71	111.00

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	5454	0	5398	39	0
2	A	1	0	0	0	0
3	A	17	0	0	0	0
4	A	522	0	0	8	2
All	All	5994	0	5398	39	2

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 (39) 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:506:SER:OG	4:A:901:HOH:O	1.76	0.88
1:A:717:ARG:NH2	4:A:905:HOH:O	2.30	0.63
1:A:235:HIS:NE2	4:A:908:HOH:O	2.32	0.59
1:A:524:ASP:O	1:A:528:THR:HG23	2.03	0.59
1:A:296:LYS:HA	1:A:382:ASN:O	2.03	0.58
1:A:376:LEU:HD22	1:A:391:GLY:HA2	1.85	0.58
1:A:198:GLY:HA2	1:A:323:PRO:O	2.03	0.58
1:A:461:LEU:HD12	1:A:639:ILE:HG21	1.87	0.57
1:A:195:VAL:HB	1:A:224:ALA:HB1	1.89	0.55
1:A:713:ILE:HD13	1:A:716:ARG:NH1	2.22	0.55
1:A:109:ILE:HG23	1:A:510:GLN:HG3	1.89	0.55
1:A:62:ALA:O	1:A:66:LYS:HG3	2.07	0.54
1:A:226:SER:O	1:A:226:SER:OG	2.24	0.52
1:A:333:ARG:NH2	4:A:913:HOH:O	2.35	0.51
1:A:41:LYS:NZ	1:A:45:ASP:OD2	2.44	0.51
1:A:342:GLY:HA2	1:A:397:LYS:HE3	1.92	0.50
1:A:379:SER:OG	1:A:380:ASP:N	2.44	0.50
1:A:119:VAL:HG21	1:A:130:ALA:HB2	1.95	0.48
1:A:297:GLY:HA3	1:A:385:LEU:HB2	1.95	0.47
1:A:697:GLU:HG2	1:A:713:ILE:HD12	1.98	0.46
1:A:5:VAL:HG21	1:A:17:TYR:CD2	2.51	0.46
1:A:20:VAL:HA	1:A:24:ALA:HB3	1.98	0.46
1:A:510:GLN:HG2	4:A:1055:HOH:O	2.16	0.46
1:A:392:SER:C	1:A:393:ILE:HD12	2.37	0.45
1:A:719:ARG:HG2	4:A:1064:HOH:O	2.15	0.45
1:A:151:THR:O	1:A:153:GLY:N	2.46	0.45
1:A:713:ILE:HG12	4:A:1100:HOH:O	2.16	0.44
1:A:259:ASP:OD1	1:A:260:VAL:N	2.49	0.44
1:A:231:LEU:HD21	1:A:330:LEU:HD11	1.99	0.44
1:A:273:GLU:OE2	1:A:399:LYS:HE3	2.19	0.43
1:A:686:PHE:HB3	1:A:692:PHE:CD1	2.53	0.43
1:A:128:LEU:HD11	1:A:300:ALA:HA	2.01	0.43
1:A:538:SER:HA	1:A:560:VAL:HG11	2.01	0.42
1:A:637:LEU:HB3	1:A:710:THR:HG21	2.01	0.42
1:A:553:LEU:O	1:A:557:GLN:HG3	2.20	0.42
1:A:47:THR:N	1:A:48:PRO:HD2	2.35	0.42
1:A:512:GLY:HA2	1:A:539:THR:O	2.21	0.41
1:A:653:ILE:HD12	1:A:658:VAL:HG22	2.03	0.41
1:A:49:GLN:HG3	4:A:910:HOH:O	2.21	0.40

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-



metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:A:1075:HOH:O	4:A:1224:HOH:O[3_654]	2.12	0.08
4:A:1292:HOH:O	4:A:1305:HOH:O[8_665]	2.18	0.02

## 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	708/741 (96%)	689 (97%)	15 (2%)	4 (1%)	25 18

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	152	ASP
1	A	206	ASP
1	A	3	ASP
1	A	151	THR

### 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	573/593 (97%)	551 (96%)	22 (4%)	33 30

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	THR
1	A	9	ASN
1	A	83	GLN
1	A	125	ARG
1	A	128	LEU
1	A	184	SER
1	A	186	SER
1	A	199	GLN
1	A	207	LYS
1	A	312	ARG
1	A	316	ARG
1	A	335	LEU
1	A	348	ASP
1	A	389	ARG
1	A	411	GLU
1	A	528	THR
1	A	583	LEU
1	A	595	ARG
1	A	603	GLN
1	A	637	LEU
1	A	648	LEU
1	A	649	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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	4ZF	A	802	-	18,19,19	3.10	6 (33%)	21,27,27	1.26	2 (9%)

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	4ZF	A	802	-	-	0/6/17/17	0/3/3/3

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	802	4ZF	C06-N07	10.16	1.43	1.35
3	A	802	4ZF	C04-N08	5.41	1.39	1.35
3	A	802	4ZF	C04-C02	3.43	1.57	1.48
3	A	802	4ZF	C10-C09	3.00	1.44	1.40
3	A	802	4ZF	O03-C02	-2.40	1.23	1.30
3	A	802	4ZF	C17-C12	2.23	1.43	1.41

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	802	4ZF	O03-C02-O01	2.49	129.31	123.61
3	A	802	4ZF	O01-C02-C04	-2.25	116.46	121.09

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	713/741 (96%)	0.20	9 (1%) 77 76	15, 28, 48, 69	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	324	GLY	5.2
1	A	381	VAL	3.6
1	A	206	ASP	3.6
1	A	677	VAL	3.0
1	A	203	ALA	2.3
1	A	125	ARG	2.3
1	A	269	ILE	2.3
1	A	541	TRP	2.1
1	A	207	LYS	2.1

### 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 monosaccharides 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
2	MG	A	801	1/1	0.92	0.15	24,24,24,24	0
3	4ZF	A	802	17/17	0.92	0.15	22,26,29,29	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.