



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:46 AM GMT

PDB ID : 3C4V  
Title : Structure of the retaining glycosyltransferase MshA: The first step in mycothiol biosynthesis. Organism: Corynebacterium glutamicum : Complex with UDP and 1L-INS-1-P.  
Authors : Vetting, M.W.; Frantom, P.A.; Blanchard, J.S.  
Deposited on : 2008-01-30  
Resolution : 2.60 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 (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

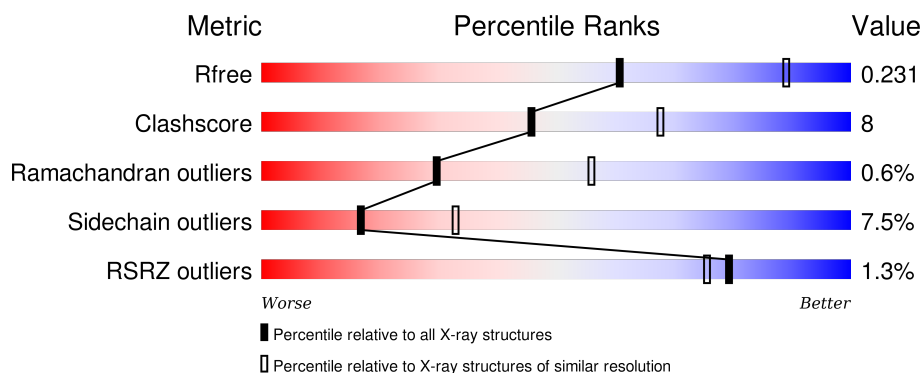
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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))
$R_{free}$	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	426	
1	B	426	

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
4	LIP	B	601	X	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6189 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 Predicted glycosyltransferases.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	393	Total	C	N	O	S	0	0	0
			3026	1896	540	578	12			
1	B	393	Total	C	N	O	S	0	0	0
			3025	1895	540	578	12			

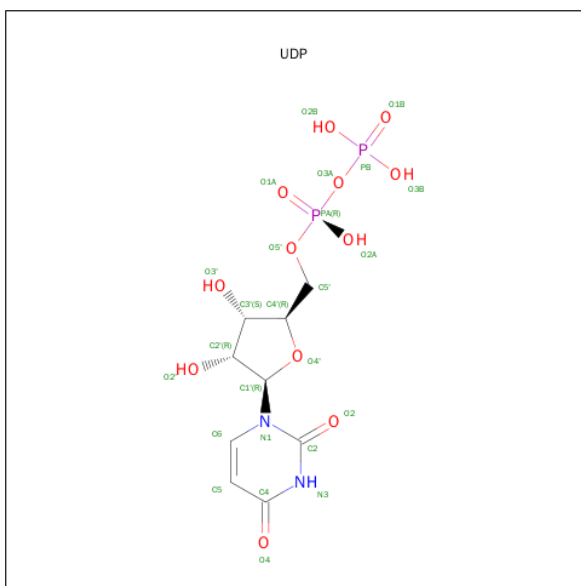
There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	419	LEU	-	EXPRESSION TAG	UNP Q8NTA6
A	420	GLU	-	EXPRESSION TAG	UNP Q8NTA6
A	421	HIS	-	EXPRESSION TAG	UNP Q8NTA6
A	422	HIS	-	EXPRESSION TAG	UNP Q8NTA6
A	423	HIS	-	EXPRESSION TAG	UNP Q8NTA6
A	424	HIS	-	EXPRESSION TAG	UNP Q8NTA6
A	425	HIS	-	EXPRESSION TAG	UNP Q8NTA6
A	426	HIS	-	EXPRESSION TAG	UNP Q8NTA6
B	419	LEU	-	EXPRESSION TAG	UNP Q8NTA6
B	420	GLU	-	EXPRESSION TAG	UNP Q8NTA6
B	421	HIS	-	EXPRESSION TAG	UNP Q8NTA6
B	422	HIS	-	EXPRESSION TAG	UNP Q8NTA6
B	423	HIS	-	EXPRESSION TAG	UNP Q8NTA6
B	424	HIS	-	EXPRESSION TAG	UNP Q8NTA6
B	425	HIS	-	EXPRESSION TAG	UNP Q8NTA6
B	426	HIS	-	EXPRESSION TAG	UNP Q8NTA6

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

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

- Molecule 3 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula:  $C_9H_{14}N_2O_{12}P_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total	C	O	P	0	0
			16	6	9	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	38	Total	O	0	0
			38	38		
5	B	32	Total	O	0	0
			32	32		



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	223.94Å 223.94Å 125.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	79.18 – 2.60 79.17 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.4 (79.18-2.60) 99.4 (79.17-2.60)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.32 (at 2.62Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.190 , 0.213 0.210 , 0.231	Depositor DCC
$R_{free}$ test set	2463 reflections (5.34%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.5	Xtriage
Anisotropy	0.352	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 30.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 48552 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6189	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.79% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: MG, LIP, UDP

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	0.72	3/3086 (0.1%)	0.78	6/4199 (0.1%)
1	B	0.84	11/3085 (0.4%)	0.74	7/4197 (0.2%)
All	All	0.78	14/6171 (0.2%)	0.76	13/8396 (0.2%)

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	8
1	B	0	4
All	All	0	12

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	319	GLY	N-CA	28.58	1.89	1.46
1	A	317	SER	CB-OG	20.22	1.68	1.42
1	B	318	PHE	C-N	9.18	1.49	1.33
1	B	318	PHE	N-CA	8.23	1.62	1.46
1	B	319	GLY	C-O	7.46	1.35	1.23

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	319	GLY	N-CA-C	-10.58	86.66	113.10
1	A	17	GLY	N-CA-C	-9.28	89.91	113.10
1	B	318	PHE	N-CA-CB	7.96	124.94	110.60
1	B	318	PHE	C-N-CA	-7.73	106.06	122.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	205	GLY	N-CA-C	-7.61	94.08	113.10

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	16	PRO	Peptide
1	A	17	GLY	Peptide
1	A	204	PRO	Peptide
1	A	262	CYS	Peptide
1	A	316	GLU	Peptide

## 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	3026	0	2989	52	0
1	B	3025	0	2987	42	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	25	0	11	1	0
3	B	25	0	11	1	0
4	B	16	0	11	0	0
5	A	38	0	0	0	0
5	B	32	0	0	0	0
All	All	6189	0	6009	93	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 8.

The worst 5 of 93 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:317:SER:CB	1:A:317:SER:OG	1.68	1.38
1:B:319:GLY:CA	1:B:319:GLY:N	1.88	1.35

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:316:GLU:HG3	1:A:317:SER:HB2	1.26	1.18
1:A:18:THR:CG2	1:A:264:GLY:HA3	2.01	0.91
1:A:18:THR:HG23	1:A:264:GLY:HA3	1.52	0.90

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	387/426 (91%)	369 (95%)	16 (4%)	2 (0%)	34	60
1	B	387/426 (91%)	373 (96%)	11 (3%)	3 (1%)	24	46
All	All	774/852 (91%)	742 (96%)	27 (4%)	5 (1%)	30	56

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	206	ASN
1	B	206	ASN
1	B	317	SER
1	A	316	GLU
1	B	318	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	319/348 (92%)	291 (91%)	28 (9%)	12	24
1	B	319/348 (92%)	299 (94%)	20 (6%)	22	44
All	All	638/696 (92%)	590 (92%)	48 (8%)	17	33

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	311	VAL
1	A	367	THR
1	B	311	VAL
1	A	318	PHE
1	A	326	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	326	GLN
1	B	62	ASN
1	A	385	HIS
1	A	14	GLN
1	B	38	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 ⓘ

Of 5 ligands modelled in this entry, 2 are monoatomic - leaving 3 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	UDP	A	600	-	18,26,26	1.19	3 (16%)	26,40,40	1.83	3 (11%)
3	UDP	B	600	-	18,26,26	1.15	1 (5%)	26,40,40	1.84	5 (19%)
4	LIP	B	601	-	16,16,16	2.60	2 (12%)	24,25,25	1.09	1 (4%)

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	UDP	A	600	-	-	0/12/32/32	0/2/2/2
3	UDP	B	600	-	-	0/12/32/32	0/2/2/2
4	LIP	B	601	-	2/2/7/7	0/5/29/29	0/1/1/1

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	600	UDP	PB-O2B	-2.35	1.46	1.54
3	A	600	UDP	PB-O3B	-2.03	1.47	1.54
3	A	600	UDP	C4-N3	2.49	1.37	1.33
3	B	600	UDP	C4-N3	3.21	1.39	1.33
4	B	601	LIP	P1-O7	6.39	1.64	1.51

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	600	UDP	PA-O3A-PB	-4.84	116.44	132.67
3	B	600	UDP	PA-O3A-PB	-4.14	118.79	132.67
3	B	600	UDP	O3'-C3'-C4'	-3.24	101.33	111.05
4	B	601	LIP	O9-P1-O7	-2.66	106.59	112.76
3	B	600	UDP	O3A-PA-O5'	-2.51	96.27	102.94

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	B	601	LIP	C5
4	B	601	LIP	C3

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	600	UDP	1	0
3	B	600	UDP	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	393/426 (92%)	0.18	2 (0%) 91 90	25, 42, 56, 64	0
1	B	393/426 (92%)	0.31	8 (2%) 68 63	35, 46, 58, 71	0
All	All	786/852 (92%)	0.24	10 (1%) 79 75	25, 45, 57, 71	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	207	ASP	3.5
1	B	318	PHE	3.3
1	B	281	GLU	2.7
1	B	409	GLU	2.4
1	B	315	ASN	2.3

### 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
4	LIP	B	601	16/16	0.86	0.27	2.97	80,84,85,85	0
3	UDP	A	600	25/25	0.94	0.17	-0.79	29,34,55,57	0
3	UDP	B	600	25/25	0.97	0.14	-1.85	24,33,48,50	0
2	MG	A	602	1/1	0.87	0.13	-1.98	41,41,41,41	0
2	MG	B	602	1/1	0.93	0.15	-2.60	46,46,46,46	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.