



## Full wwPDB X-ray Structure Validation Report ⓘ

May 15, 2020 – 07:45 am BST

PDB ID : 4DQV  
Title : Crystal structure of reductase (R) domain of non-ribosomal peptide synthetase from *Mycobacterium tuberculosis*  
Authors : Haque, A.S.; Panjikar, S.; Sankaranarayanan, R.  
Deposited on : 2012-02-16  
Resolution : 2.30 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 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 : 2.11  
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) : 2.11

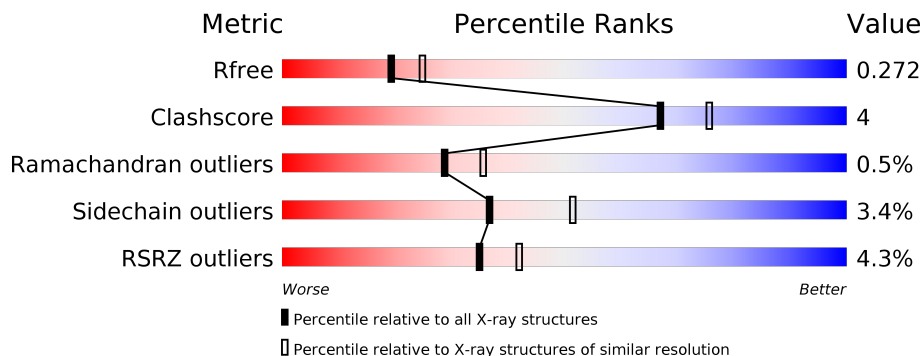
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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 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	478	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 3462 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 PROBABLE PEPTIDE SYNTHETASE NRP (PEPTIDE SYNTHASE).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	428	3301	2083	579	627	4	8	73	0	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MSE	-	EXPRESSION TAG	UNP Q10896
A	-19	GLY	-	EXPRESSION TAG	UNP Q10896
A	-18	SER	-	EXPRESSION TAG	UNP Q10896
A	-17	SER	-	EXPRESSION TAG	UNP Q10896
A	-16	HIS	-	EXPRESSION TAG	UNP Q10896
A	-15	HIS	-	EXPRESSION TAG	UNP Q10896
A	-14	HIS	-	EXPRESSION TAG	UNP Q10896
A	-13	HIS	-	EXPRESSION TAG	UNP Q10896
A	-12	HIS	-	EXPRESSION TAG	UNP Q10896
A	-11	HIS	-	EXPRESSION TAG	UNP Q10896
A	-10	SER	-	EXPRESSION TAG	UNP Q10896
A	-9	SER	-	EXPRESSION TAG	UNP Q10896
A	-8	GLY	-	EXPRESSION TAG	UNP Q10896
A	-7	LEU	-	EXPRESSION TAG	UNP Q10896
A	-6	VAL	-	EXPRESSION TAG	UNP Q10896
A	-5	PRO	-	EXPRESSION TAG	UNP Q10896
A	-4	ARG	-	EXPRESSION TAG	UNP Q10896
A	-3	GLY	-	EXPRESSION TAG	UNP Q10896
A	-2	SER	-	EXPRESSION TAG	UNP Q10896
A	-1	HIS	-	EXPRESSION TAG	UNP Q10896
A	0	MSE	-	EXPRESSION TAG	UNP Q10896

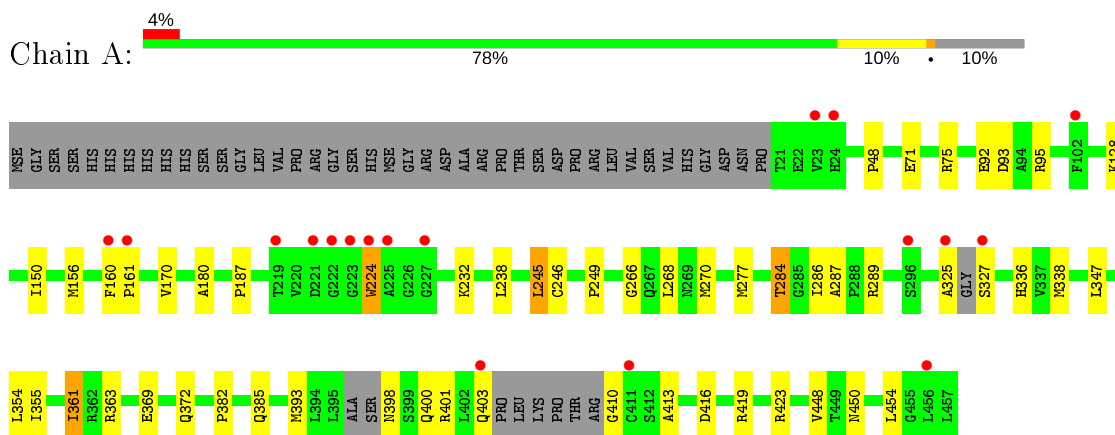
- Molecule 2 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
2	A	161	Total 161	O 161	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 the various outlier classes 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.

- Molecule 1: PROBABLE PEPTIDE SYNTHETASE NRP (PEPTIDE SYNTHASE)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	59.06 Å 59.06 Å 238.99 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	51.15 – 2.30 50.01 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.5 (51.15-2.30) 99.5 (50.01-2.30)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.70 (at 2.29 Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.229 , 0.278 0.227 , 0.272	Depositor DCC
$R_{free}$ test set	1127 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.8	Xtrriage
Anisotropy	0.669	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 38.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.043 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	3462	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.04% 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

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	0.67	6/3358 (0.2%)	0.65	6/4554 (0.1%)

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	224	TRP	CB-CG	-23.97	1.07	1.50
1	A	369	GLU	CB-CG	-11.18	1.30	1.52
1	A	95	ARG	CB-CG	-8.35	1.30	1.52
1	A	454	LEU	CB-CG	-6.27	1.34	1.52
1	A	403	GLN	CB-CG	-6.10	1.36	1.52
1	A	398	ASN	CB-CG	5.69	1.64	1.51

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	224	TRP	CA-CB-CG	11.62	135.79	113.70
1	A	372	GLN	CB-CG-CD	-6.83	93.84	111.60
1	A	400	GLN	CA-CB-CG	5.96	126.50	113.40
1	A	398	ASN	CB-CG-OD1	5.91	133.42	121.60
1	A	398	ASN	CB-CG-ND2	-5.53	103.44	116.70
1	A	454	LEU	CB-CG-CD2	-5.23	102.11	111.00

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	3301	0	3267	25	0
2	A	161	0	0	2	0
All	All	3462	0	3267	25	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 (25) 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:284:THR:HG23	1:A:286:ILE:H	1.47	0.80
1:A:410:GLY:N	2:A:542:HOH:O	2.28	0.67
1:A:325:ALA:C	1:A:327:SER:O	2.36	0.64
1:A:338:MSE:HE1	1:A:413:ALA:HB3	1.80	0.62
1:A:336:HIS:HB3	1:A:338:MSE:HE2	1.83	0.61
1:A:382:PRO:HD2	1:A:385:GLN:HE21	1.65	0.60
1:A:48:PRO:HG2	1:A:249:PRO:HG2	1.86	0.57
1:A:289:ARG:HB3	1:A:289:ARG:HH11	1.70	0.55
1:A:355:ILE:HG12	1:A:361:ILE:HG23	1.89	0.54
1:A:150:ILE:HD12	1:A:180:ALA:HB2	1.92	0.51
1:A:170:VAL:HG13	1:A:238:LEU:HD11	1.92	0.50
1:A:287:ALA:O	1:A:363:ARG:HA	2.13	0.49
1:A:268:LEU:HD23	1:A:270:MSE:HE1	1.94	0.49
1:A:266:GLY:H	1:A:450:ASN:ND2	2.11	0.48
1:A:382:PRO:HD2	1:A:385:GLN:NE2	2.28	0.48
1:A:48:PRO:HB3	1:A:187:PRO:HD3	1.95	0.48
1:A:71:GLU:O	1:A:75:ARG:HG2	2.16	0.46
1:A:419:ARG:O	1:A:423:ARG:HG3	2.17	0.45
1:A:289:ARG:HB3	1:A:289:ARG:NH1	2.32	0.44
1:A:268:LEU:HD22	1:A:450:ASN:HB3	2.00	0.43
1:A:245:LEU:HB3	1:A:246:CYS:SG	2.59	0.43
1:A:128:LYS:HE3	2:A:562:HOH:O	2.19	0.43
1:A:287:ALA:HB2	1:A:361:ILE:HD11	2.01	0.42
1:A:160:PHE:HB3	1:A:161:PRO:HD3	2.02	0.42
1:A:232:LYS:HA	1:A:232:LYS:HD3	1.91	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	420/478 (88%)	402 (96%)	16 (4%)	2 (0%)	29 35

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	401	ARG
1	A	393	MSE

### 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	348/380 (92%)	336 (97%)	12 (3%)	37 51

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	92	GLU
1	A	93	ASP
1	A	156	MSE
1	A	224	TRP
1	A	245	LEU
1	A	277	MSE
1	A	284	THR
1	A	347	LEU
1	A	354	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	361	ILE
1	A	416	ASP
1	A	448	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	269	ASN
1	A	385	GLN
1	A	438	HIS
1	A	450	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](#)

There are no ligands in this entry.

### 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	420/478 (87%)	0.15	18 (4%) 35 42	24, 43, 63, 82	15 (3%)

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	222	GLY	7.1
1	A	325	ALA	5.5
1	A	224	TRP	4.0
1	A	160	PHE	3.7
1	A	411	CYS	3.4
1	A	296	SER	3.3
1	A	161	PRO	2.9
1	A	219	THR	2.8
1	A	225	ALA	2.7
1	A	327	SER	2.6
1	A	227	GLY	2.5
1	A	223	GLY	2.3
1	A	23	VAL	2.2
1	A	221	ASP	2.2
1	A	24	HIS	2.1
1	A	102	PHE	2.1
1	A	403	GLN	2.0
1	A	456	LEU	2.0

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.