W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

May 16, 2020 – 04:39 pm BST

PDB ID	:	2X9Q
Title	:	Structure of the Mycobacterium tuberculosis protein, Rv2275, demonstrates that cyclodipeptide synthetases are related to type I tRNA-Synthetases.
Authors	:	Vetting, M.W.; Hegde, S.S.; Blanchard, J.S.
Deposited on	:	2010-03-23
Resolution	:	2.02 Å(reported)

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

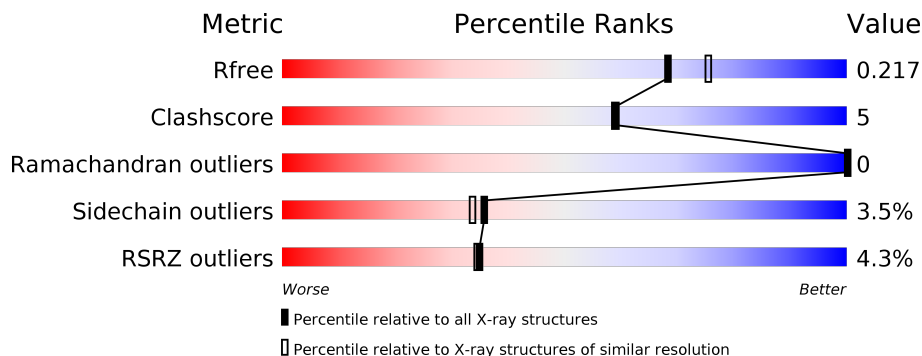|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.02 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	Percentile Ranks	Value
Rfree		0.217
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		3.5%
RSRZ outliers		4.3%

Worse → *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

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 on the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	289	2% 71% 8% · 20%
1	B	289	5% 66% 9% · 23%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3804 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 CYCLODIPEPTIDE SYNTHETASE.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	A	231	Total	C	N	O	S	Se		0	0	0
			1794	1114	330	340	6	4				
1	B	222	Total	C	N	O	S	Se		0	0	0
			1746	1090	320	326	6	4				

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	196	MSE	LEU	conflict	UNP Q50688
B	196	MSE	LEU	conflict	UNP Q50688

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	146	Total	O	0	0
			146	146		
2	B	118	Total	O	0	0
			118	118		

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: CYCLODIPEPTIDE SYNTHETASE
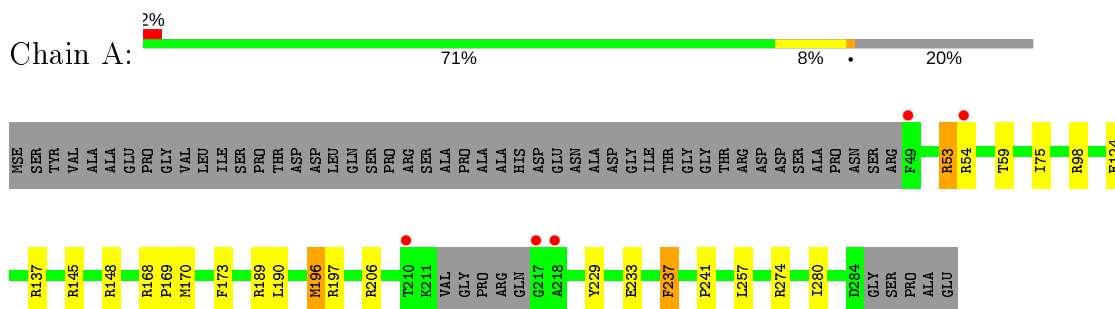
- Molecule 1: CYCLODIPEPTIDE SYNTHETASE
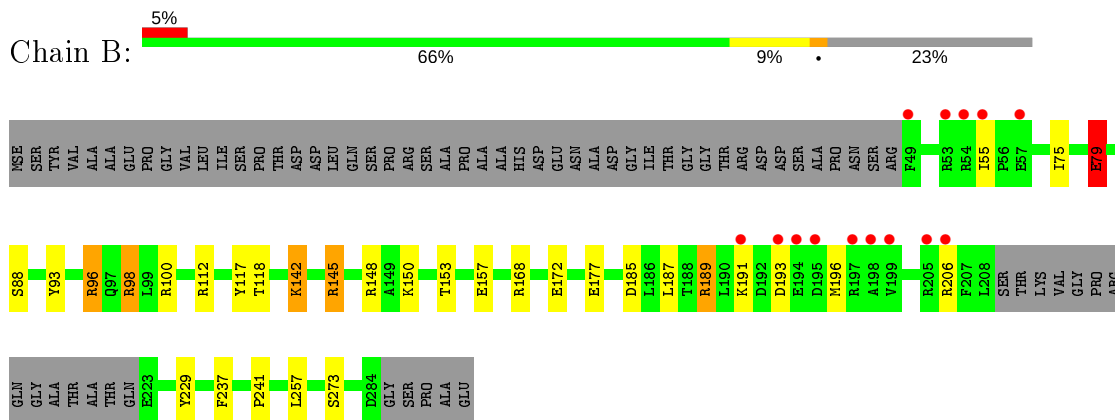
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	71.09Å 75.06Å 140.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	70.18 – 2.02	Depositor
	39.08 – 2.02	EDS
% Data completeness (in resolution range)	98.8 (70.18-2.02)	Depositor
	98.8 (39.08-2.02)	EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.80 (at 2.01Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.180 , 0.210	Depositor
	0.186 , 0.217	DCC
R_{free} test set	2525 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	28.7	Xtriage
Anisotropy	0.413	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 48.9	EDS
L-test for twinning[2]	$<|L|> = 0.48$, $<L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3804	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.84% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $<|L|>$, $<L^2>$ 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)

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	1.34	4/1821 (0.2%)	1.17	12/2460 (0.5%)
1	B	1.23	3/1774 (0.2%)	1.01	5/2395 (0.2%)
All	All	1.29	7/3595 (0.2%)	1.09	17/4855 (0.4%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	124	GLU	CG-CD	7.79	1.63	1.51
1	A	229	TYR	CD1-CE1	6.25	1.48	1.39
1	A	237	PHE	CE1-CZ	5.92	1.48	1.37
1	B	177	GLU	CG-CD	5.40	1.60	1.51
1	A	145	ARG	CG-CD	5.35	1.65	1.51

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	197	ARG	NE-CZ-NH2	-11.96	114.32	120.30
1	A	197	ARG	NE-CZ-NH1	10.42	125.51	120.30
1	A	148	ARG	NE-CZ-NH2	8.93	124.77	120.30
1	A	137	ARG	NE-CZ-NH1	8.12	124.36	120.30
1	A	196	MSE	CG-SE-CE	-7.71	81.93	98.90

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	1794	0	1759	14	0
1	B	1746	0	1718	24	0
2	A	146	0	0	0	0
2	B	118	0	0	2	0
All	All	3804	0	3477	33	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 5.

The worst 5 of 33 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:B:100:ARG:NH2	1:B:157:GLU:OE1	1.85	1.08
1:A:190:LEU:HD23	1:A:196:MSE:HE1	1.68	0.72
1:B:96:ARG:HD2	1:B:150:LYS:HG2	1.73	0.70
1:A:190:LEU:HD23	1:A:196:MSE:CE	2.26	0.65
1:B:191:LYS:NZ	1:B:191:LYS:HB3	2.13	0.63

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	227/289 (78%)	223 (98%)	4 (2%)	0	100	100
1	B	218/289 (75%)	212 (97%)	6 (3%)	0	100	100
All	All	445/578 (77%)	435 (98%)	10 (2%)	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	190/232 (82%)	186 (98%)	4 (2%)	53	55
1	B	186/232 (80%)	177 (95%)	9 (5%)	25	21
All	All	376/464 (81%)	363 (96%)	13 (4%)	36	34

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	79	GLU
1	B	88	SER
1	B	187	LEU
1	B	55	ILE
1	B	142	LYS

Some 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 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, 95th 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^2)	Q<0.9
1	A	227/289 (78%)	-0.05	5 (2%)	62	61	17, 27, 47, 67	0
1	B	218/289 (75%)	0.05	14 (6%)	19	18	17, 30, 65, 86	0
All	All	445/578 (76%)	-0.00	19 (4%)	35	34	17, 28, 56, 86	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	217	GLY	7.6
1	A	218	ALA	4.7
1	B	55	ILE	4.2
1	B	206	ARG	4.1
1	B	53	ARG	3.9

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.