

# wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 06:37 am BST

PDB ID : 3R0J

Title: Structure of PhoP from Mycobacterium tuberculosis

Authors : Menon, S.; Wang, S.

Deposited on : 2011-03-08

Resolution : 2.50 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

A user guide is available at

https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) 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

 $\begin{array}{ccc} \mathrm{CCP4} & : & 7.0.044 \; \mathrm{(Gargrove)} \\ \mathrm{roteins)} & : & \mathrm{Engh} \; \& \; \mathrm{Huber} \; \mathrm{(2001)} \end{array}$ 

Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

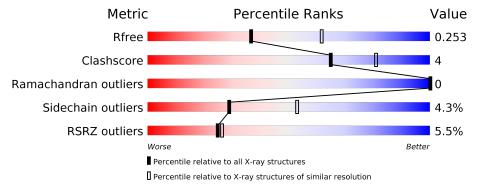
Validation Pipeline (wwPDB-VP) : 2.11

# 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 2.50 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	$\begin{array}{c} \text{Whole archive} \\ (\#\text{Entries}) \end{array}$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries,\ resolution\ range(\AA)}) \end{array}$
$R_{free}$	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	250	78%	10%	·	10%		
1	В	250	80%	7%		12%		



# 2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3622 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.

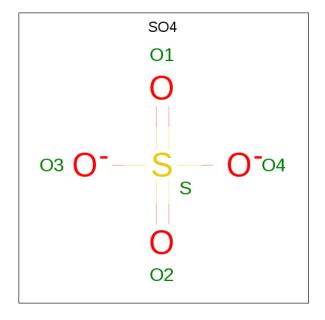
• Molecule 1 is a protein called POSSIBLE TWO COMPONENT SYSTEM RESPONSE TRANSCRIPTIONAL POSITIVE REGULATOR PHOP.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	A	224	Total 1768	C 1124	• '		Se 3	0	0	0
1	В	221	Total 1746	C 1111	• '	O 326	Se 3	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP P71814
A	-1	THR	-	EXPRESSION TAG	UNP P71814
A	0	HIS	-	EXPRESSION TAG	UNP P71814
В	-2	GLY	-	EXPRESSION TAG	UNP P71814
В	-1	THR	=	EXPRESSION TAG	UNP P71814
В	0	HIS	-	EXPRESSION TAG	UNP P71814

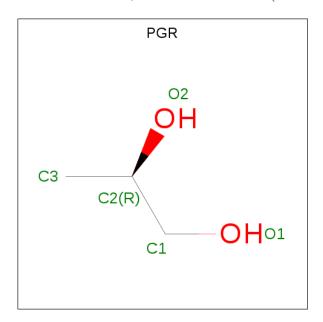
• Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	В	1	Total O S 5 4 1	0	0
2	В	1	Total O S 5 4 1	0	0
2	В	1	Total O S 5 4 1	0	0

 $\bullet$  Molecule 3 is R-1,2-PROPANEDIOL (three-letter code: PGR) (formula:  $\mathrm{C_3H_8O_2}).$ 



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	В	1	Total C O 5 3 2	0	0

• Molecule 4 is water.

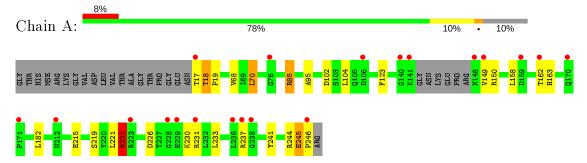
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	43	Total O 43 43	0	0
4	В	35	Total O 35 35	0	0



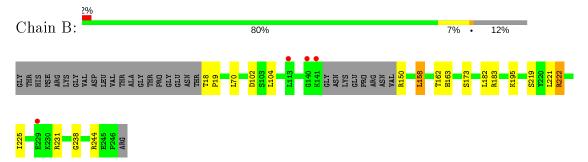
# 3 Residue-property plots (i)

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: POSSIBLE TWO COMPONENT SYSTEM RESPONSE TRANSCRIPTIONAL POSITIVE REGULATOR PHOP



• Molecule 1: POSSIBLE TWO COMPONENT SYSTEM RESPONSE TRANSCRIPTIONAL POSITIVE REGULATOR PHOP





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants	79.23Å 138.24Å 58.91Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.81 - 2.50	Depositor
Resolution (A)	29.81 - 2.50	EDS
% Data completeness	98.6 (29.81-2.50)	Depositor
(in resolution range)	98.6 (29.81-2.50)	EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	3.18 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
D D.	0.212 , 0.245	Depositor
$R, R_{free}$	0.220 , $0.253$	DCC
$R_{free}$ test set	1162 reflections $(5.08\%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.0	Xtriage
Anisotropy	0.434	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.34 , 49.6	EDS
L-test for twinning <sup>2</sup>	$ < L >=0.46, < L^2>=0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	3622	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The analyses of the Patterson function reveals a significant off-origin peak that is 41.64 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.2908e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.

<sup>&</sup>lt;sup>2</sup>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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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: MSO, PGR, SO4

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	Boı	nd lengths	Bond angles		
IVIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z >5	
1	A	0.86	3/1787 (0.2%)	0.86	5/2420 (0.2%)	
1	В	0.76	0/1765	0.83	4/2389 (0.2%)	
All	All	0.81	$3/3552 \ (0.1\%)$	0.84	9/4809 (0.2%)	

#### All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	${f Observed(\AA)}$	$\operatorname{Ideal}( ext{\AA})$
1	A	237	ARG	CZ-NH1	10.64	1.46	1.33
1	A	245	GLU	CG-CD	-8.29	1.39	1.51
1	A	237	ARG	NE-CZ	5.86	1.40	1.33

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

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$Observed(^o)$	$\operatorname{Ideal}({}^o)$
1	A	222	ARG	NE-CZ-NH2	9.58	125.09	120.30
1	В	222	ARG	NE-CZ-NH2	-8.14	116.23	120.30
1	A	222	ARG	NE-CZ-NH1	-7.76	116.42	120.30
1	В	222	ARG	NE-CZ-NH1	6.94	123.77	120.30
1	A	244	ARG	NE-CZ-NH1	-5.91	117.34	120.30

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	Α	1768	0	1787	16	0
1	В	1746	0	1765	13	0
2	A	10	0	0	0	0
2	В	15	0	0	1	0
3	В	5	0	8	3	0
4	A	43	0	0	1	0
4	В	35	0	0	1	0
All	All	3622	0	3560	29	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.

The worst 5 of 29 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{array}{c}  ext{Interatomic} \  ext{distance } ( ext{Å}) \end{array}$	$egin{array}{c}  ext{Clash} \  ext{overlap } ( ext{Å}) \end{array}$
1:B:183:ARG:HD3	2:B:301:SO4:O3	1.65	0.97
1:B:195:LYS:NZ	3:B:248:PGR:H2	2.18	0.59
1:A:231:ARG:O	1:A:246:PRO:HD3	2.07	0.54
1:B:18:THR:HG22	1:B:19:PRO:O	2.07	0.54
1:B:238:GLY:H	3:B:248:PGR:H12	1.77	0.50

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	$\mathbf{Outliers}$	Perce	$\mathbf{ntiles}$
1	A	219/250~(88%)	214 (98%)	5 (2%)	0	100	100
1	В	$216/250 \; (86\%)$	208 (96%)	8 (4%)	0	100	100
All	All	435/500 (87%)	422 (97%)	13 (3%)	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	189/206 (92%)	178 (94%)	11 (6%)	20	38	
1	В	186/206 (90%)	181 (97%)	5 (3%)	44	71	
All	All	375/412 (91%)	359 (96%)	16 (4%)	29	53	

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	158	LEU
1	A	219	SER
1	В	158	LEU
1	A	150	ARG
1	В	173	SER

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)

2 non-standard protein/DNA/RNA residues 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	Т	Chain	Res Link		В	ond leng	${ m gths}$	В	ond ang	gles
MIOI	Type	Chain	Res	Link	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
1	MSO	A	73	1	4,8,9	0.76	0	1,9,11	0.03	0
1	MSO	В	73	1	4,8,9	0.50	0	1,9,11	0.16	0

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	${f Res}$	Link	Chirals	Torsions	Rings
1	MSO	A	73	1	-	0/4/7/9	-
1	MSO	В	73	1	-	0/4/7/9	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry (i)

6 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	Т	Chain	Res	Link	В	ond leng	${ m gths}$	В	ond ang	gles
MIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	SO4	В	303	-	4,4,4	0.18	0	6,6,6	0.24	0
2	SO4	В	304	_	4,4,4	0.09	0	6,6,6	0.54	0
3	PGR	В	248	_	3,4,4	0.55	0	1,4,4	0.38	0



Mol	Mol Type Chain Res L		Link	B	Bond lengths			Bond angles		
MIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	SO4	A	305	_	4,4,4	0.13	0	6,6,6	0.16	0
2	SO4	В	301	_	4,4,4	0.20	0	6,6,6	0.62	0
2	SO4	A	302	-	4,4,4	0.20	0	6,6,6	0.33	0

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	PGR	В	248	-	-	0/2/2/2	-

There are no bond length outliers.

There are no bond angle outliers.

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
3	В	248	PGR	3	0
2	В	301	SO4	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,  $95^{th}$  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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(\AA^2)$	Q < 0.9
1	A	221/250 (88%)	0.47	20 (9%) 9 9	34, 61, 110, 145	0
1	В	$218/250 \ (87\%)$	0.13	4 (1%) 68 71	34, 52, 82, 100	0
All	All	439/500 (87%)	0.30	24 (5%) 25 26	34, 57, 96, 145	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	148	ASN	6.8
1	В	141	LYS	6.2
1	A	149	VAL	6.1
1	A	17	THR	5.3
1	A	140	GLY	4.8

### 6.2 Non-standard residues in protein, DNA, RNA chains (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,  $95^{th}$  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	${f Res}$	Atoms	RSCC	RSR	${f B-factors({ m \AA}^2)}$	Q<0.9
1	MSO	В	73	9/10	0.86	0.24	53,58,66,75	0
1	MSO	A	73	9/10	0.87	0.24	53,58,65,74	0

## 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. The B-factors column lists the minimum, median,  $95^{th}$  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	$ extbf{B-factors}( extbf{A}^2)$	Q<0.9
3	PGR	В	248	5/5	0.60	0.21	76,76,81,82	0
2	SO4	A	305	5/5	0.77	0.32	$130,\!130,\!131,\!131$	0
2	SO4	В	304	5/5	0.91	0.26	77,83,84,84	0
2	SO4	В	301	5/5	0.91	0.18	80,81,83,84	0
2	SO4	В	303	5/5	0.96	0.28	59,62,63,63	0
2	SO4	A	302	5/5	0.96	0.29	67,67,69,70	0

### 6.5 Other polymers (i)

There are no such residues in this entry.

