



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 24, 2017 – 11:01 PM EDT

PDB ID : 3HGF  
Title : Expression, purification, spectroscopical and crystallographical studies of segments of the nucleotide binding domain of the reticulocyte binding protein Py235 of Plasmodium yoelii  
Authors : Gruber, A.; Manimekalai, M.S.S.; Balakrishna, A.M.; Hunke, C.; Jeyakanthan, J.; Preiser, P.R.; Gruber, G.  
Deposited on : unknown  
Resolution : 4.00 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030345

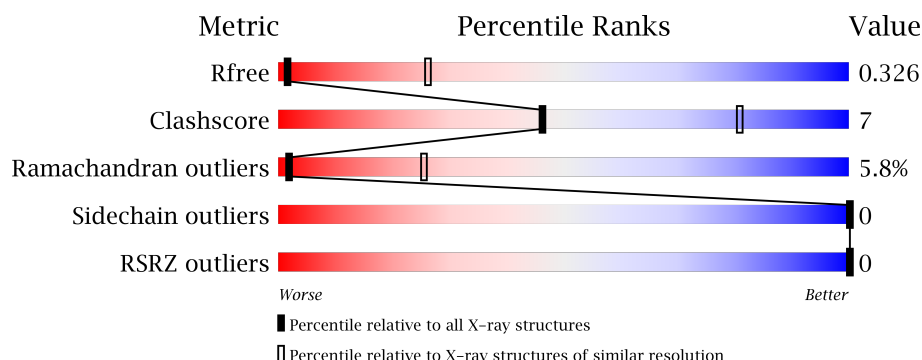
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.00 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))
$R_{free}$	100719	1088 (4.40-3.60)
Clashscore	112137	1187 (4.40-3.60)
Ramachandran outliers	110173	1139 (4.40-3.60)
Sidechain outliers	110143	1126 (4.40-3.60)
RSRZ outliers	101464	1099 (4.40-3.60)

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	107	
1	B	107	
1	C	107	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1375 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 Rhoptyr protein fragment.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	98	Total	C	N	O	0	0	0
			493	297	98	98			
1	B	89	Total	C	N	O	0	0	0
			446	268	89	89			
1	C	87	Total	C	N	O	0	0	0
			436	262	87	87			

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	HIS	-	EXPRESSION TAG	UNP Q7RPU0
A	2	HIS	-	EXPRESSION TAG	UNP Q7RPU0
A	3	HIS	-	EXPRESSION TAG	UNP Q7RPU0
A	4	HIS	-	EXPRESSION TAG	UNP Q7RPU0
A	5	HIS	-	EXPRESSION TAG	UNP Q7RPU0
A	6	HIS	-	EXPRESSION TAG	UNP Q7RPU0
A	7	PRO	-	EXPRESSION TAG	UNP Q7RPU0
A	8	MET	-	EXPRESSION TAG	UNP Q7RPU0
A	9	VAL	-	EXPRESSION TAG	UNP Q7RPU0
B	1	HIS	-	EXPRESSION TAG	UNP Q7RPU0
B	2	HIS	-	EXPRESSION TAG	UNP Q7RPU0
B	3	HIS	-	EXPRESSION TAG	UNP Q7RPU0
B	4	HIS	-	EXPRESSION TAG	UNP Q7RPU0
B	5	HIS	-	EXPRESSION TAG	UNP Q7RPU0
B	6	HIS	-	EXPRESSION TAG	UNP Q7RPU0
B	7	PRO	-	EXPRESSION TAG	UNP Q7RPU0
B	8	MET	-	EXPRESSION TAG	UNP Q7RPU0
B	9	VAL	-	EXPRESSION TAG	UNP Q7RPU0
C	1	HIS	-	EXPRESSION TAG	UNP Q7RPU0
C	2	HIS	-	EXPRESSION TAG	UNP Q7RPU0
C	3	HIS	-	EXPRESSION TAG	UNP Q7RPU0
C	4	HIS	-	EXPRESSION TAG	UNP Q7RPU0
C	5	HIS	-	EXPRESSION TAG	UNP Q7RPU0

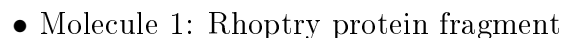
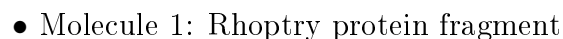
*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	6	HIS	-	EXPRESSION TAG	UNP Q7RPU0
C	7	PRO	-	EXPRESSION TAG	UNP Q7RPU0
C	8	MET	-	EXPRESSION TAG	UNP Q7RPU0
C	9	VAL	-	EXPRESSION TAG	UNP Q7RPU0

**i**

- Molecule 1: Rhoptry protein fragment



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.22Å 70.22Å 193.22Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	33.00 – 4.00 33.00 – 4.00	Depositor EDS
% Data completeness (in resolution range)	95.4 (33.00-4.00) 95.4 (33.00-4.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.09 (at 3.99Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, $R_{free}$	0.339 , 0.373 0.378 , 0.326	Depositor DCC
$R_{free}$ test set	117 reflections (4.48%)	DCC
Wilson B-factor (Å <sup>2</sup> )	133.2	Xtriage
Anisotropy	0.466	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	-0.11 , 107.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	1375	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.82% 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](#)

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	0.27	0/493	0.45	0/687
1	B	0.27	0/445	0.45	0/620
1	C	0.27	0/434	0.45	0/603
All	All	0.27	0/1372	0.45	0/1910

There are no bond length outliers.

There are no bond angle outliers.

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	493	0	212	8	0
1	B	446	0	188	3	0
1	C	436	0	183	3	0
All	All	1375	0	583	14	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 7.

The worst 5 of 14 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:C:99:GLN:HA	1:C:103:ALA:HB2	1.83	0.60

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:9:VAL:C	1:A:11:GLU:H	2.12	0.53
1:A:36:LEU:C	1:A:38:ASN:H	2.14	0.51
1:B:96:ILE:C	1:B:98:GLU:H	2.15	0.49
1:B:79:SER:C	1:B:81:ASN:H	2.18	0.47

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	94/107 (88%)	70 (74%)	14 (15%)	10 (11%)	0	10
1	B	85/107 (79%)	67 (79%)	17 (20%)	1 (1%)	15	58
1	C	81/107 (76%)	63 (78%)	14 (17%)	4 (5%)	2	29
All	All	260/321 (81%)	200 (77%)	45 (17%)	15 (6%)	2	25

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	20	VAL
1	A	43	ILE
1	A	62	LYS
1	A	84	LYS
1	C	101	PRO

### 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	2/104 (2%)	2 (100%)	0	100	100
1	B	1/104 (1%)	1 (100%)	0	100	100
1	C	1/104 (1%)	1 (100%)	0	100	100
All	All	4/312 (1%)	4 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

### 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 ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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	98/107 (91%)	-0.86	0 100 100	18, 42, 56, 57	0
1	B	89/107 (83%)	-0.70	0 100 100	55, 63, 73, 73	0
1	C	87/107 (81%)	-0.52	0 100 100	72, 101, 105, 105	0
All	All	274/321 (85%)	-0.70	0 100 100	18, 62, 105, 105	0

There are no RSRZ outliers to report.

### 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.