



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 01:31 am BST

PDB ID : 5MI0  
Title : A thermally stabilised version of Plasmodium falciparum RH5  
Authors : Campeotto, I.; Goldenzweig, A.; Davey, J.; Barfod, L.; Marshall, J.M.; Silk, S.E.; Wright, K.E.; Draper, S.J.; Higgins, M.K.; Fleishman, S.J.  
Deposited on : 2016-11-27  
Resolution : 2.35 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

<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  
Xtrriage (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

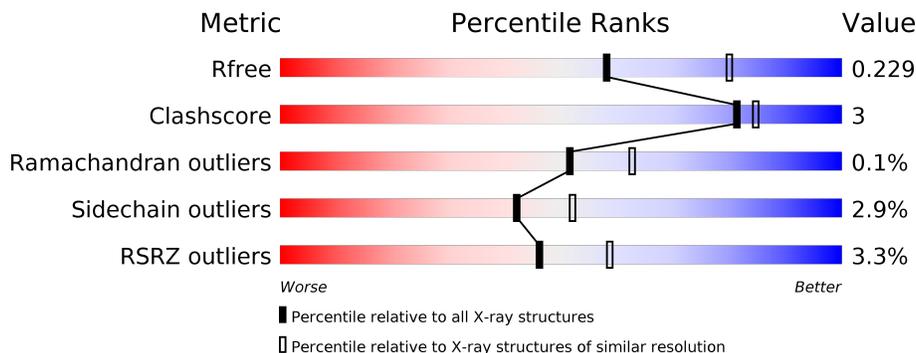
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	360	
2	B	258	
3	C	238	

## 2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 6008 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 Reticulocyte binding-like protein 5, Reticulocyte binding protein 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	281	2402	1556	402	433	11	0	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	124	HIS	-	expression tag	UNP A0A159SJI5
A	125	HIS	-	expression tag	UNP A0A159SJI5
A	126	HIS	-	expression tag	UNP A0A159SJI5
A	127	HIS	-	expression tag	UNP A0A159SJI5
A	128	HIS	-	expression tag	UNP A0A159SJI5
A	129	HIS	-	expression tag	UNP A0A159SJI5
A	130	GLU	-	expression tag	UNP A0A159SJI5
A	131	ASN	-	expression tag	UNP A0A159SJI5
A	132	LEU	-	expression tag	UNP A0A159SJI5
A	133	TYR	-	expression tag	UNP A0A159SJI5
A	134	PHE	-	expression tag	UNP A0A159SJI5
A	135	GLN	-	expression tag	UNP A0A159SJI5
A	136	GLY	-	expression tag	UNP A0A159SJI5
A	137	GLY	-	expression tag	UNP A0A159SJI5
A	138	SER	-	expression tag	UNP A0A159SJI5
A	139	MET	-	expression tag	UNP A0A159SJI5
A	157	LEU	ILE	conflict	UNP A0A159SJI5
A	183	GLU	ASP	conflict	UNP A0A159SJI5
A	216	ALA	THR	conflict	UNP A0A159SJI5
A	233	LYS	ALA	conflict	UNP A0A159SJI5
A	299	ALA	THR	conflict	UNP B2L3N7
A	304	PHE	MET	conflict	UNP B2L3N7
A	312	ASN	LYS	conflict	UNP B2L3N7
A	314	PHE	LEU	conflict	UNP B2L3N7
A	316	ASN	LYS	conflict	UNP B2L3N7
A	330	ASN	MET	conflict	UNP B2L3N7

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	370	ALA	SER	conflict	UNP B2L3N7
A	381	ASN	SER	conflict	UNP B2L3N7
A	384	LYS	THR	conflict	UNP B2L3N7
A	392	LYS	LEU	conflict	UNP B2L3N7
A	395	ASN	THR	conflict	UNP B2L3N7
A	398	GLU	ASN	conflict	UNP B2L3N7
A	458	LYS	ARG	conflict	UNP B2L3N7
A	464	LYS	ASN	conflict	UNP B2L3N7
A	467	ALA	SER	conflict	UNP B2L3N7
A	505	LEU	PHE	conflict	UNP B2L3N7
A	527	HIS	-	expression tag	UNP B2L3N7
A	528	HIS	-	expression tag	UNP B2L3N7
A	529	HIS	-	expression tag	UNP B2L3N7
A	530	HIS	-	expression tag	UNP B2L3N7
A	531	HIS	-	expression tag	UNP B2L3N7
A	532	HIS	-	expression tag	UNP B2L3N7

- Molecule 2 is a protein called MONOCLONAL ANTIBODY 9AD4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	221	1668	1060	272	327	9	0	0	0

- Molecule 3 is a protein called MONOCLONAL ANTIBODY 9AD4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	217	1677	1046	286	338	7	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	32	Total 32	O 32	0	0
4	B	116	Total 116	O 116	0	0
4	C	113	Total 113	O 113	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	39.66Å 85.55Å 132.87Å 90.00° 90.91° 90.00°	Depositor
Resolution (Å)	32.91 – 2.35 31.83 – 2.35	Depositor EDS
% Data completeness (in resolution range)	98.4 (32.91-2.35) 98.4 (31.83-2.35)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.72 (at 2.34Å)	Xtrriage
Refinement program	BUSTER 2.10.2, PHENIX 1.9	Depositor
R, $R_{free}$	0.176 , 0.222 0.178 , 0.229	Depositor DCC
$R_{free}$ test set	1798 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.2	Xtrriage
Anisotropy	0.047	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 52.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.041 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	6008	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.21% 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	0.49	0/2454	0.61	0/3291
2	B	0.52	0/1714	0.74	0/2343
3	C	0.50	0/1717	0.70	0/2332
All	All	0.50	0/5885	0.67	0/7966

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	2402	0	2409	6	1
2	B	1668	0	1610	12	0
3	C	1677	0	1603	14	0
4	A	32	0	0	0	0
4	B	116	0	0	0	0
4	C	113	0	0	1	0
All	All	6008	0	5622	29	1

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 3.

The worst 5 of 29 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)
2:B:39:GLN:HE22	3:C:42:GLN:HE22	1.21	0.88
2:B:72:ARG:HE	2:B:74:ASN:HD21	1.29	0.79
3:C:14:SER:H	3:C:17:GLN:HE21	1.45	0.63
3:C:154:ILE:HD11	3:C:183:LEU:HD21	1.83	0.60
3:C:110:ILE:H	3:C:170:GLN:HE22	1.51	0.59

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:406:TYR:OH	1:A:436:LYS:NZ[2_5412]	2.17	0.03

### 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	275/360 (76%)	271 (98%)	4 (2%)	0	100	100
2	B	219/258 (85%)	212 (97%)	6 (3%)	1 (0%)	29	32
3	C	215/238 (90%)	207 (96%)	8 (4%)	0	100	100
All	All	709/856 (83%)	690 (97%)	18 (2%)	1 (0%)	51	63

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	181	SER

#### 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	271/346 (78%)	260 (96%)	11 (4%)	30	37
2	B	185/216 (86%)	181 (98%)	4 (2%)	52	63
3	C	190/209 (91%)	186 (98%)	4 (2%)	53	65
All	All	646/771 (84%)	627 (97%)	19 (3%)	42	52

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	394	LEU
1	A	436	LYS
3	C	60	SER
1	A	392	LYS
3	C	85	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	74	ASN
3	C	17	GLN
3	C	93	GLN
2	B	39	GLN
3	C	38	GLN

### 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	281/360 (78%)	0.33	20 (7%) 16 23	36, 62, 97, 130	0
2	B	221/258 (85%)	-0.06	3 (1%) 75 83	22, 36, 62, 87	0
3	C	217/238 (91%)	-0.23	1 (0%) 91 95	22, 42, 65, 81	0
All	All	719/856 (83%)	0.04	24 (3%) 46 59	22, 46, 87, 130	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	140	THR	6.2
1	A	299	ALA	5.9
1	A	302	LYS	4.8
1	A	447	TRP	4.8
1	A	397	LEU	4.4

### 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

There are no such residues in this entry.