Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2017 – 05:04 PM EST

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 Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
http://wwpdb.org/validation/2016/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:

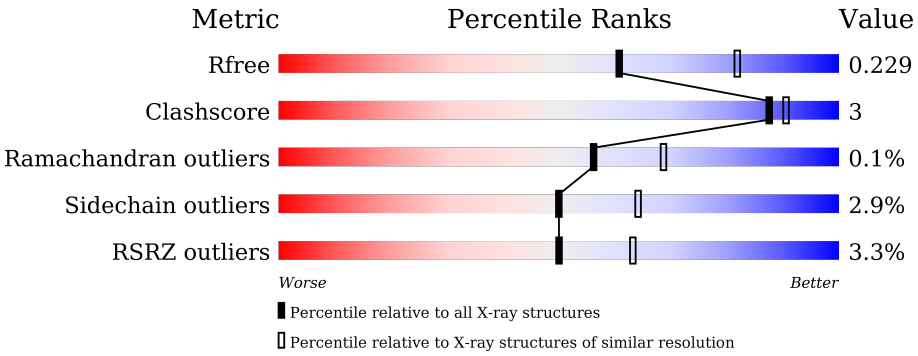
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

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 2.35 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}	91344	1352 (2.38-2.34)
Clashscore	102246	1456 (2.38-2.34)
Ramachandran outliers	100387	1435 (2.38-2.34)
Sidechain outliers	100360	1436 (2.38-2.34)
RSRZ outliers	91569	1358 (2.38-2.34)

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. 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	360	6% 72% 6% 22%
2	B	258	% 79% 6% • 14%
3	C	238	81% 10% 9%

2 Entry composition

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
1	A	281	Total	C	N	O	S	0	0	0
			2402	1556	402	433	11			

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
2	B	221	Total	C	N	O	S	0	0	0
			1668	1060	272	327	9			

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

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

- Molecule 4 is water.

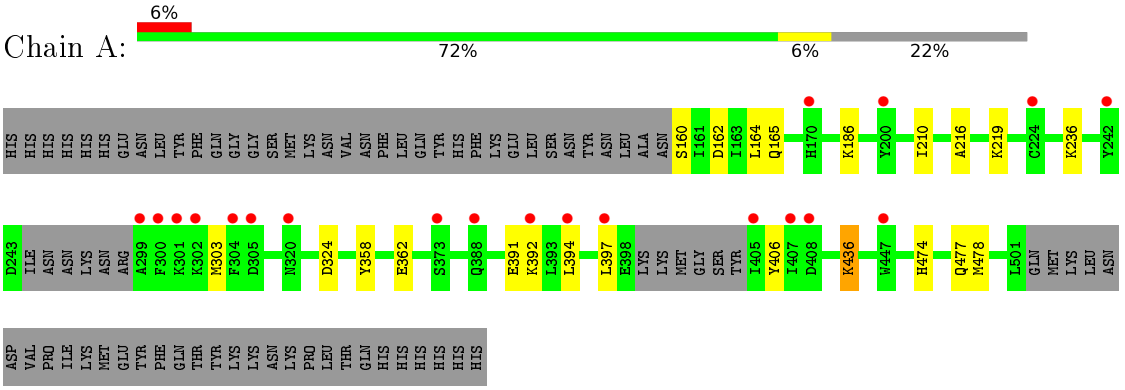
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	32	Total	O	0	0
			32	32		
4	B	116	Total	O	0	0
			116	116		
4	C	113	Total	O	0	0
			113	113		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

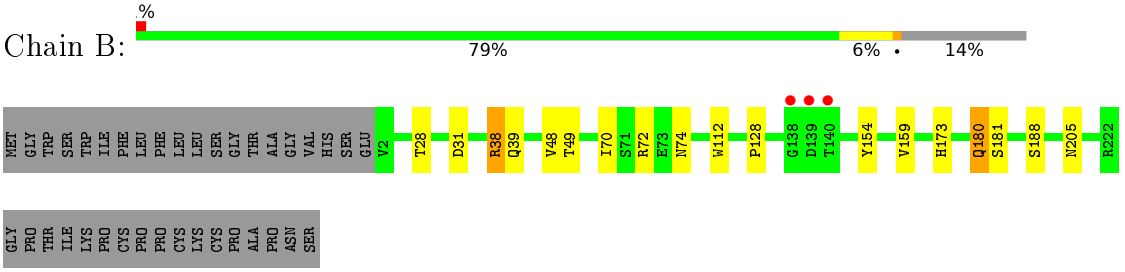
- Molecule 1: Reticulocyte binding-like protein 5,Reticulocyte binding protein 5

Chain A:



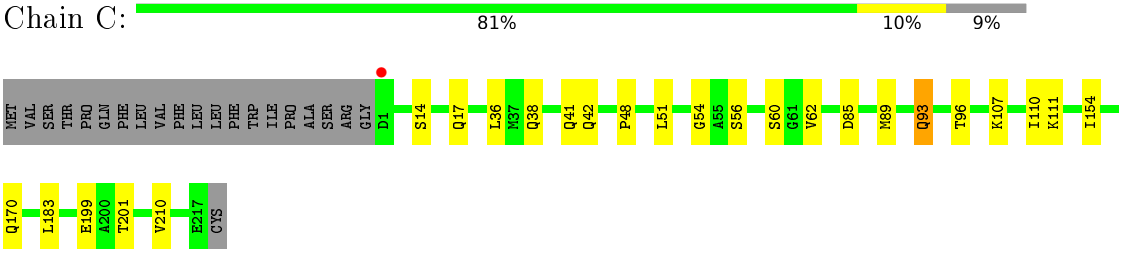
- Molecule 2: MONOCLONAL ANTIBODY 9AD4

Chain B:



- Molecule 3: MONOCLONAL ANTIBODY 9AD4

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	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
$< I/\sigma(I) >$ [1]	2.72 (at 2.34Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.176 , 0.222 0.178 , 0.229	Depositor DCC
R_{free} test set	1798 reflections (5.18%)	DCC
Wilson B-factor (Å^2)	38.2	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 52.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.041 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	6008	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage'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.*

[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

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	0.49	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

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.

All (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
1:A:164:LEU:HD22	1:A:478:MET:HB3	1.84	0.58
3:C:54:GLY:C	3:C:56:SER:H	2.07	0.57
1:A:474:HIS:HA	1:A:477:GLN:HE21	1.71	0.56
2:B:72:ARG:NE	2:B:74:ASN:HD21	2.04	0.53
2:B:180:GLN:O	2:B:180:GLN:HG2	2.10	0.51
2:B:49:THR:HG23	2:B:70:ILE:HD12	1.92	0.51
2:B:39:GLN:HE22	3:C:42:GLN:NE2	2.01	0.50
2:B:72:ARG:HE	2:B:74:ASN:ND2	2.03	0.50
3:C:89:MET:HG2	3:C:107:LYS:HD3	1.94	0.49
3:C:199:GLU:HG2	3:C:210:VAL:HG22	1.98	0.45
3:C:38:GLN:HB2	3:C:93:GLN:HE21	1.83	0.44
1:A:358:TYR:CZ	1:A:362:GLU:HG3	2.53	0.44
3:C:36:LEU:HD12	3:C:96:THR:HA	1.99	0.44
1:A:216:ALA:HA	1:A:219:LYS:HE3	1.99	0.44
2:B:38:ARG:HD2	2:B:48:VAL:HG22	2.00	0.43
1:A:216:ALA:HA	1:A:219:LYS:CE	2.47	0.43
3:C:41:GLN:HB2	3:C:51:LEU:HD11	2.00	0.43
2:B:112:TRP:CE3	3:C:48:PRO:HD2	2.55	0.42
1:A:186:LYS:HG3	1:A:210:ILE:HD13	2.00	0.42
3:C:51:LEU:HA	3:C:62:VAL:HG21	2.01	0.42
2:B:173:HIS:O	2:B:188:SER:HA	2.20	0.41
3:C:201:THR:HG23	4:C:377:HOH:O	2.20	0.40
2:B:128:PRO:HB3	2:B:154:TYR:HB3	2.04	0.40
2:B:28:THR:O	2:B:31:ASP:HB2	2.21	0.40

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%)	34 39
3	C	215/238 (90%)	207 (96%)	8 (4%)	0	100 100
All	All	709/856 (83%)	690 (97%)	18 (2%)	1 (0%)	56 69

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%)	37 48
2	B	185/216 (86%)	181 (98%)	4 (2%)	60 75
3	C	190/209 (91%)	186 (98%)	4 (2%)	61 76
All	All	646/771 (84%)	627 (97%)	19 (3%)	50 64

All (19) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	160	SER
1	A	162	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	165	GLN
1	A	236	LYS
1	A	303	MET
1	A	324	ASP
1	A	391	GLU
1	A	392	LYS
1	A	394	LEU
1	A	397	LEU
1	A	436	LYS
2	B	38	ARG
2	B	159	VAL
2	B	180	GLN
2	B	205	ASN
3	C	60	SER
3	C	85	ASP
3	C	93	GLN
3	C	111	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	312	ASN
1	A	316	ASN
1	A	477	GLN
2	B	39	GLN
2	B	74	ASN
3	C	17	GLN
3	C	38	GLN
3	C	93	GLN
3	C	141	ASN
3	C	170	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

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

6.1 Protein, DNA and RNA chains

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>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	281/360 (78%)	0.33	20 (7%) 19 28	36, 62, 97, 130	0
2	B	221/258 (85%)	-0.06	3 (1%) 78 87	22, 36, 62, 87	0
3	C	217/238 (91%)	-0.23	1 (0%) 91 96	22, 42, 65, 81	0
All	All	719/856 (83%)	0.04	24 (3%) 50 63	22, 46, 87, 130	0

All (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
1	A	300	PHE	3.8
1	A	394	LEU	3.6
1	A	242	TYR	3.5
1	A	200	TYR	3.4
1	A	320	ASN	3.1
1	A	405	ILE	3.1
2	B	139	ASP	3.1
1	A	304	PHE	2.9
1	A	388	GLN	2.7
1	A	407	ILE	2.6
1	A	392	LYS	2.5
1	A	373	SER	2.4
1	A	301	LYS	2.3
1	A	305	ASP	2.3
1	A	170	HIS	2.3
1	A	224	CYS	2.2
3	C	1	ASP	2.1
2	B	138	GLY	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	408	ASP	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.