Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 02:48 PM EDT

PDB ID	:	2MRC
BMRB ID	:	25068
Title	:	NMR Structure and 1H, 13C and 15N Chemical Shift Assignments for High mobility group protein from Plasmodium falciparum 3D7.
Authors	:	Tang, C.; Barnwal, R.; Varani, G.; Seattle Structural Genomics Center for Infectious Disease (SSGCID)
Deposited on	:	2014-07-02

This is a Full wwPDB NMR Structure Validation 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

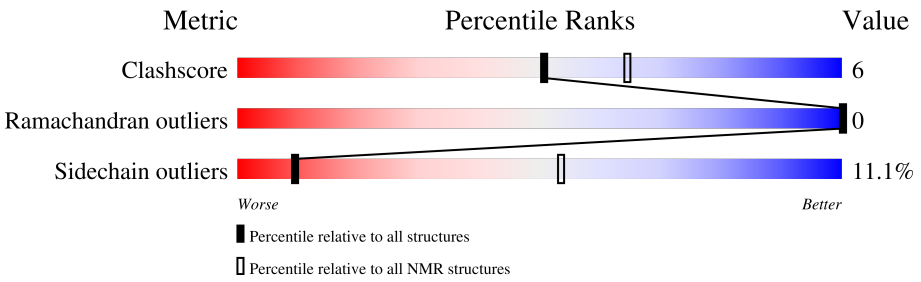
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 81%.

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)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	A	92	72% 10% 18%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 1 is the overall representative, medoid model (most similar to other models).

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:16-A:90 (75)	1.06	1

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters and 2 single-model clusters were found.

Cluster number	Models
1	1, 3, 5, 8, 9
2	2, 6, 7
Single-model clusters	4; 10

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1551 atoms, of which 788 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called High mobility group protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	92	1551	488	788	140	132	3	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP Q8IB14
A	2	ALA	-	expression tag	UNP Q8IB14
A	3	HIS	-	expression tag	UNP Q8IB14
A	4	HIS	-	expression tag	UNP Q8IB14
A	5	HIS	-	expression tag	UNP Q8IB14
A	6	HIS	-	expression tag	UNP Q8IB14
A	7	HIS	-	expression tag	UNP Q8IB14
A	8	HIS	-	expression tag	UNP Q8IB14
A	9	MET	-	expression tag	UNP Q8IB14

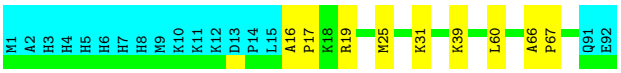
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: High mobility group protein

Chain A: 72% 10% 18%



4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

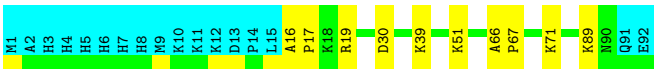
• Molecule 1: High mobility group protein

Chain A: 66% 15% 18%

4.2.2 Score per residue for model 2

• Molecule 1: High mobility group protein

Chain A: 71% 11% 18%



4.2.3 Score per residue for model 3

- Molecule 1: High mobility group protein

Chain A: 63% 18% 18%

M1 A2 H3 H4 H5 H6 H7 H8 M9 K10 K11 K12 D13 P14 L15 A16 P17 A23 Y24 M25 K31 K39 K44 D45 V49 G50 K51 S61 K65 A66 P67 Y68 E84 Q91 E92

4.2.4 Score per residue for model 4

- Molecule 1: High mobility group protein

Chain A: 58% 24% 18%

M1 A2 H3 H4 H5 H6 H7 H8 M9 K10 K11 K12 D13 P14 L15 A16 P17 K18 R19 A20 L21 S22 A23 Y24 M25 F26 Y27 V28 K29 D30 K31 R32 I35 I36 K37 V49 L52 W57 L60 A66 P67 K70 K71 A72 K88 Q91 E92

4.2.5 Score per residue for model 5

- Molecule 1: High mobility group protein

Chain A: 65% 15% • 18%

M1 A2 H3 H4 H5 H6 H7 H8 M9 K10 K11 K12 D13 P14 L15 A16 P17 K18 R19 A20 L21 M25 K31 K39 D45 K51 L60 S61 K65 A66 P67 K89 N90 Q91 E92

4.2.6 Score per residue for model 6

- Molecule 1: High mobility group protein

Chain A: 76% 5% 18%

M1 A2 H3 H4 H5 H6 H7 H8 M9 K10 K11 K12 D13 P14 L15 A16 P17 K39 A66 P67 Q91 E92

4.2.7 Score per residue for model 7

- Molecule 1: High mobility group protein

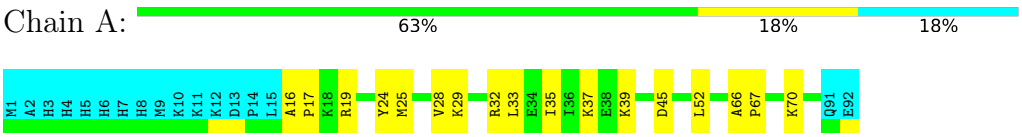
Chain A: 65% 14% • 18%

M1 A2 H3 H4 H5 H6 H7 H8 M9 K10 K11 K12 D13 P14 L15 A16 P17 K18 R19 A20 L21 M25 F26 Y27 K31 I35 L60 S61 A66 P67 K71 K81 K89 N90 Q91 E92

4.2.8 Score per residue for model 8

- Molecule 1: High mobility group protein

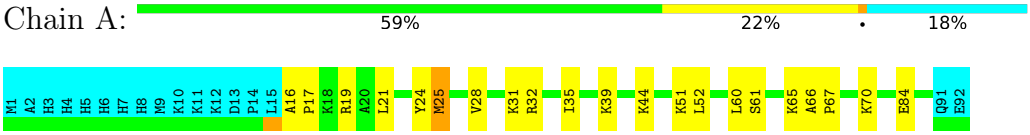
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: High mobility group protein

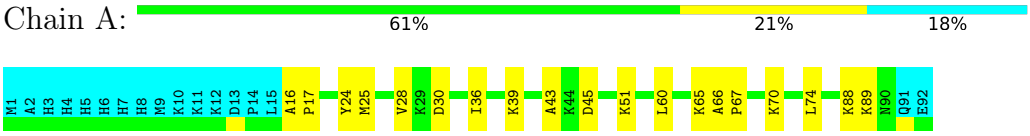
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: High mobility group protein

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics*.

Of the 200 calculated structures, 10 were deposited, based on the following criterion: *target function*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
CYANA	structure solution	
CYANA	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	928
Number of shifts mapped to atoms	928
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	81%

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	614	648	648	8±2
All	All	6140	6480	6480	75

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 6.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:66:ALA:HB3	1:A:67:PRO:HD3	0.68	1.66	5	10
1:A:66:ALA:HB3	1:A:67:PRO:CD	0.62	2.25	5	10
1:A:23:ALA:HB3	1:A:72:ALA:HB2	0.62	1.70	4	1
1:A:24:TYR:O	1:A:28:VAL:HG23	0.61	1.95	8	5
1:A:21:LEU:HD23	1:A:25:MET:CB	0.59	2.27	4	1
1:A:36:ILE:HG23	1:A:43:ALA:HB2	0.54	1.78	10	1
1:A:60:LEU:HD23	1:A:65:LYS:HG2	0.53	1.80	10	1
1:A:21:LEU:HD23	1:A:25:MET:HB3	0.52	1.79	4	2
1:A:35:ILE:HD13	1:A:52:LEU:HD12	0.52	1.81	9	4
1:A:21:LEU:HD22	1:A:25:MET:SD	0.52	2.45	9	1
1:A:70:LYS:O	1:A:74:LEU:HD12	0.51	2.06	10	1
1:A:45:ASP:O	1:A:49:VAL:HG22	0.49	2.06	3	1
1:A:60:LEU:HD22	1:A:65:LYS:HG2	0.49	1.83	5	1
1:A:73:GLN:O	1:A:77:VAL:HG23	0.49	2.08	1	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:35:ILE:CD1	1:A:52:LEU:HD12	0.48	2.38	8	1
1:A:36:ILE:HG23	1:A:43:ALA:CB	0.47	2.39	10	1
1:A:66:ALA:CB	1:A:67:PRO:CD	0.46	2.93	1	10
1:A:60:LEU:HD23	1:A:65:LYS:CG	0.44	2.43	10	1
1:A:16:ALA:HB1	1:A:17:PRO:HD2	0.44	1.90	2	10
1:A:32:ARG:CD	1:A:49:VAL:HG11	0.43	2.43	4	1
1:A:21:LEU:HD22	1:A:25:MET:CB	0.42	2.45	5	1
1:A:66:ALA:N	1:A:67:PRO:HD2	0.41	2.30	3	5
1:A:23:ALA:HB1	1:A:68:TYR:CD2	0.41	2.51	3	1
1:A:31:LYS:HD2	1:A:35:ILE:HD11	0.41	1.92	7	1
1:A:49:VAL:HG23	1:A:50:GLY:N	0.41	2.31	3	1
1:A:29:LYS:O	1:A:33:LEU:HD12	0.40	2.17	8	1
1:A:32:ARG:HD2	1:A:49:VAL:HG11	0.40	1.94	4	1

6.3 Torsion angles (i)

6.3.1 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 PDB entries followed by that with respect to all NMR entries. 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	75/92 (82%)	69±1 (93±2%)	6±1 (7±2%)	0±0 (0±0%)	100	100
All	All	750/920 (82%)	694 (93%)	56 (7%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	63/79 (80%)	56±3 (89±4%)	7±3 (11±4%)	9	53
All	All	630/790 (80%)	560 (89%)	70 (11%)	9	53

All 21 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	39	LYS	7
1	A	19	ARG	6
1	A	31	LYS	5
1	A	60	LEU	5
1	A	61	SER	5
1	A	51	LYS	5
1	A	25	MET	5
1	A	45	ASP	4
1	A	89	LYS	4
1	A	30	ASP	3
1	A	70	LYS	3
1	A	71	LYS	2
1	A	44	LYS	2
1	A	65	LYS	2
1	A	84	GLU	2
1	A	27	TYR	2
1	A	37	LYS	2
1	A	88	LYS	2
1	A	32	ARG	2
1	A	57	TRP	1
1	A	81	LYS	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 81% for the well-defined parts and 69% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	928
Number of shifts mapped to atoms	928
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	78	-1.18 ± 0.14	Should be checked
$^{13}C_{\beta}$	75	-0.74 ± 0.11	Should be checked
$^{13}C'$	78	-0.45 ± 0.15	None needed (< 0.5 ppm)
^{15}N	73	0.14 ± 0.24	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 81%, i.e. 876 atoms were assigned a chemical shift out of a possible 1088. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	364/370 (98%)	146/149 (98%)	148/150 (99%)	70/71 (99%)
Sidechain	510/651 (78%)	322/418 (77%)	183/202 (91%)	5/31 (16%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	2/67 (3%)	1/31 (3%)	0/35 (0%)	1/1 (100%)
Overall	876/1088 (81%)	469/598 (78%)	331/387 (86%)	76/103 (74%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 69%, i.e. 928 atoms were assigned a chemical shift out of a possible 1337. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	382/453 (84%)	153/182 (84%)	156/184 (85%)	73/87 (84%)
Sidechain	544/775 (70%)	343/498 (69%)	195/242 (81%)	6/35 (17%)
Aromatic	2/109 (2%)	1/55 (2%)	0/47 (0%)	1/7 (14%)
Overall	928/1337 (69%)	497/735 (68%)	351/473 (74%)	80/129 (62%)

7.1.4 Statistically unusual chemical shifts (i)

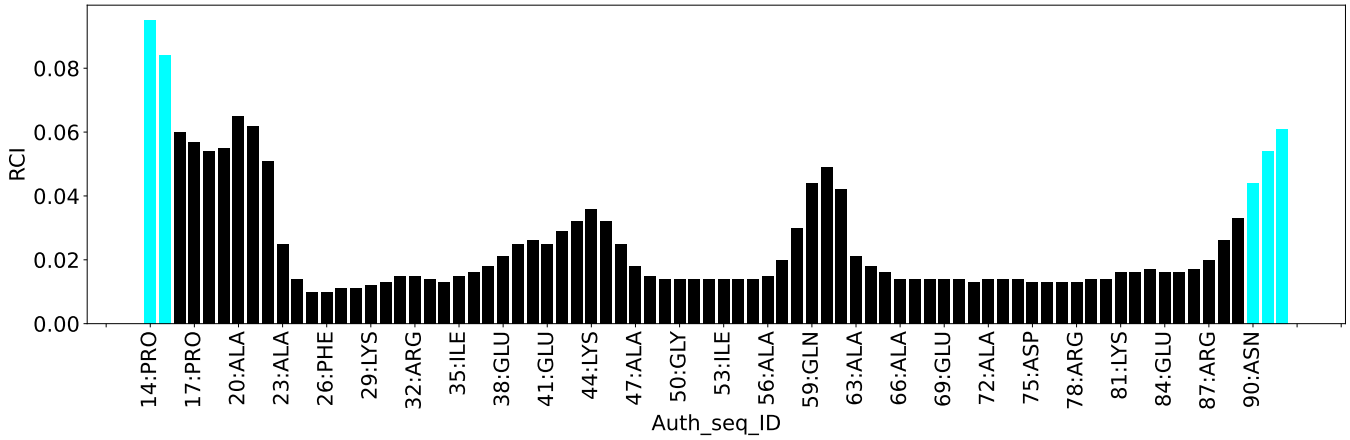
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	71	LYS	HB3	0.20	0.46 – 3.04	-6.0
1	A	65	LYS	HG2	0.06	0.13 – 2.61	-5.3

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:


RCI
0.00
0.02
0.04
0.06
0.08
14:PRO
17:PRO
20:ALA
23:ALA
26:PHE
29:LYS
32:ARG
35:ILE
38:GLU
41:GLU
44:LYS
47:ALA
50:GLY
53:ILE
56:ALA
59:GLN
63:ALA
66:ALA
69:GLU
72:ALA
75:ASP
78:ARG
81:LYS
84:GLU
87:ARG
90:ASN
Auth_seq_ID