Full wwPDB NMR Structure Validation Report (i)

Jun 3, 2023 – 11:58 AM EDT

PDB ID	:	6NVZ
BMRB ID	:	30567
Title	:	NMR Structure of the DNA binding domain of EhMybS3
Authors	:	Titaux-Delgado, G.A.; del Rio-Portilla, f.
Deposited on	:	2019-02-05

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 (1)) were used in the production of this report:

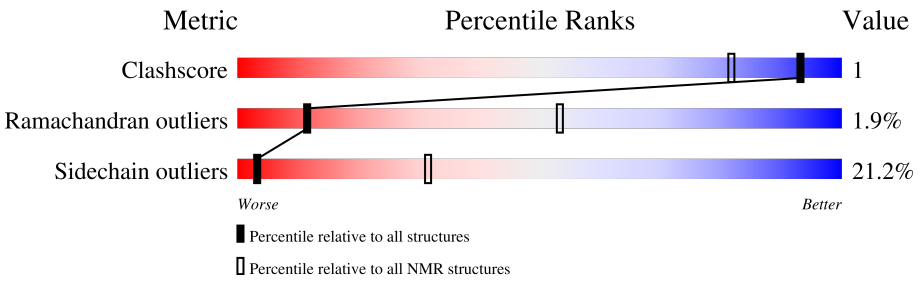
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

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 47%.

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	60	53% 10% • 35%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 16 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:12-A:38, A:42-A:53 (39)	0.38	16

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 3 clusters and 2 single-model clusters were found.

Cluster number	Models
1	2, 5, 7, 9, 12, 13, 14, 15, 16, 17, 19
2	3, 4, 6, 11, 20
3	8, 10
Single-model clusters	1; 18

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1012 atoms, of which 517 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Myb family DNA-binding protein shaqkyf family.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	60	1012	323	517	88	84	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP A0A175JNU7
A	2	PRO	-	expression tag	UNP A0A175JNU7
A	3	LEU	-	expression tag	UNP A0A175JNU7
A	4	GLY	-	expression tag	UNP A0A175JNU7
A	5	SER	-	expression tag	UNP A0A175JNU7

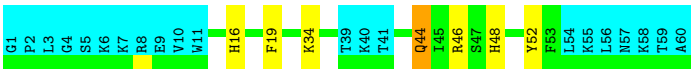
4 Residue-property plots (i)

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: Myb family DNA-binding protein shaqkyf family

Chain A:



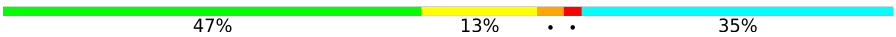
4.2 Scores per residue for each member of the ensemble

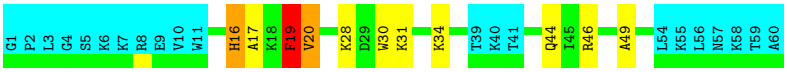
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A:

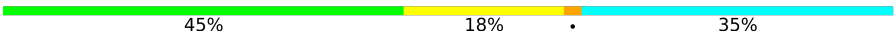


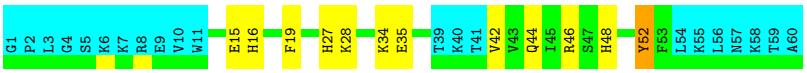


4.2.2 Score per residue for model 2

• Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A:





4.2.3 Score per residue for model 3

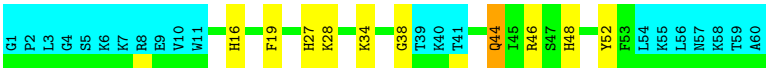
- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 50% 15% 35%

4.2.4 Score per residue for model 4

- Molecule 1: Myb family DNA-binding protein shaqkyf family

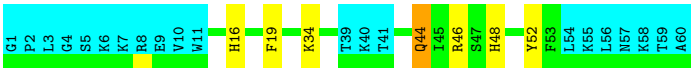
Chain A: 48% 15% • 35%



4.2.5 Score per residue for model 5

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 53% 10% • 35%



4.2.6 Score per residue for model 6

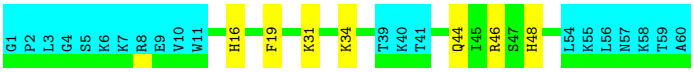
- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 53% 10% • 35%

4.2.7 Score per residue for model 7

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 53% 12% 35%



4.2.8 Score per residue for model 8

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 53% 8% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 H16 F19 K34 T39 K40 T41 Q44 I45 R46 S47 H48 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.9 Score per residue for model 9

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 52% 10% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 H16 F19 H27 K34 T39 K40 T41 Q44 I45 R46 S47 H48 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.10 Score per residue for model 10

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 48% 13% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 T12 H16 F19 K28 K31 K34 T39 K40 T41 Q44 I45 R46 S47 H48 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.11 Score per residue for model 11

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 53% 10% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 E15 F19 K34 T39 K40 T41 Q44 I45 R46 S47 H48 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.12 Score per residue for model 12

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 45% 17% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 T12 E15 H16 F19 H27 K28 K34 G38 T39 K40 T41 Q44 I45 R46 S47 H48 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.13 Score per residue for model 13

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 53% 8% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 H16 F19 K34 T39 K40 T41 Q44 I45 R46 S47 H48 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.14 Score per residue for model 14

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 52% 8% 5% 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 H16 F19 H27 K34 T39 K40 T41 Q44 I45 R46 S47 H48 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.15 Score per residue for model 15

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 50% 12% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 H16 F19 K31 K34 G38 T39 K40 T41 Q44 I45 R46 Q50 K51 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.16 Score per residue for model 16 (medoid)

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 53% 8% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 H16 F19 H27 K34 T39 K40 T41 Q44 H48 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.17 Score per residue for model 17

- Molecule 1: Myb family DNA-binding protein shaqkyf family

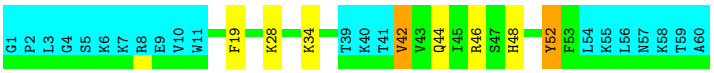
Chain A: 48% 13% • 35%

G1 P2 L3 G4 S5 K6 K7 R8 E9 V10 W11 E15 H16 F19 K34 G38 T39 K40 T41 Q44 I45 R46 S47 H48 A49 Q50 K51 Y52 F53 L54 K55 L56 N57 K58 T59 A60

4.2.18 Score per residue for model 18

- Molecule 1: Myb family DNA-binding protein shaqkyf family

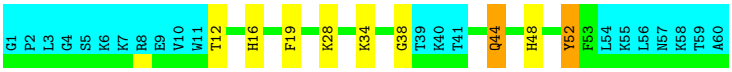
Chain A: 52% 10% • 35%



4.2.19 Score per residue for model 19

- Molecule 1: Myb family DNA-binding protein shaqkyf family

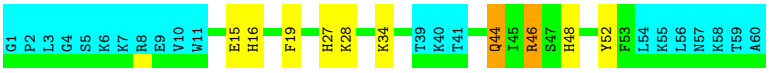
Chain A: 50% 12% • 35%



4.2.20 Score per residue for model 20

- Molecule 1: Myb family DNA-binding protein shaqkyf family

Chain A: 48% 13% • 35%



5 Refinement protocol and experimental data overview (i)

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

Of the 500 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	**Classification**	**Version**
CYANA	structure calculation	
Amber	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	355
Number of shifts mapped to atoms	355
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	47%

6 Model quality (i)

6.1 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 (average) 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	0.73±0.01	0±0/339 (0.0± 0.0%)	1.08±0.05	1±0/455 (0.2± 0.1%)
All	All	0.73	0/6780 (0.0%)	1.08	18/9100 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.8±0.6
All	All	0	16

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	46	ARG	NE-CZ-NH1	8.54	124.57	120.30	18	16
1	A	46	ARG	NE-CZ-NH2	5.53	123.06	120.30	8	1
1	A	17	ALA	N-CA-CB	-5.16	102.87	110.10	1	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	52	TYR	Sidechain	13
1	A	16	HIS	Sidechain	1
1	A	19	PHE	Sidechain	1
1	A	46	ARG	Sidechain	1

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	329	329	329	1±1
All	All	6580	6580	6580	17

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:44:GLN:H	1:A:44:GLN:CD	0.52	2.08	20	14
1:A:30:TRP:CH2	1:A:49:ALA:HB1	0.48	2.43	1	1
1:A:19:PHE:CD1	1:A:20:VAL:N	0.47	2.82	1	1
1:A:16:HIS:CE1	1:A:19:PHE:CZ	0.43	3.06	1	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	39/60 (65%)	34±1 (87±3%)	4±1 (11±3%)	1±1 (2±2%)	11 53
All	All	780/1200 (65%)	679 (87%)	86 (11%)	15 (2%)	11 53

All 4 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	27	HIS	7
1	A	38	GLY	5
1	A	42	VAL	2
1	A	31	LYS	1

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	33/51 (65%)	26±1 (79±4%)	7±1 (21±4%)	3	31
All	All	660/1020 (65%)	520 (79%)	140 (21%)	3	31

All 15 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	19	PHE	20
1	A	34	LYS	20
1	A	44	GLN	20
1	A	16	HIS	17
1	A	48	HIS	17
1	A	52	TYR	16
1	A	28	LYS	8
1	A	15	GLU	6
1	A	31	LYS	5
1	A	12	THR	5
1	A	50	GLN	2
1	A	20	VAL	1
1	A	35	GLU	1
1	A	42	VAL	1
1	A	46	ARG	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 47% for the well-defined parts and 41% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *myb.str*

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	355
Number of shifts mapped to atoms	355
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)

No chemical shift referencing corrections were calculated (not enough data).

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 47%, i.e. 268 atoms were assigned a chemical shift out of a possible 570. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	77/197 (39%)	77/80 (96%)	0/78 (0%)	0/39 (0%)
Sidechain	165/292 (57%)	165/189 (87%)	0/92 (0%)	0/11 (0%)
Aromatic	26/81 (32%)	26/41 (63%)	0/36 (0%)	0/4 (0%)
Overall	268/570 (47%)	268/310 (86%)	0/206 (0%)	0/54 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 41%, i.e. 355 atoms were assigned a chemical shift out of a possible 870. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	100/302 (33%)	100/123 (81%)	0/120 (0%)	0/59 (0%)
Sidechain	224/475 (47%)	224/307 (73%)	0/148 (0%)	0/20 (0%)
Aromatic	31/93 (33%)	31/47 (66%)	0/41 (0%)	0/5 (0%)
Overall	355/870 (41%)	355/477 (74%)	0/309 (0%)	0/84 (0%)

7.1.4 Statistically unusual chemical shifts (i)

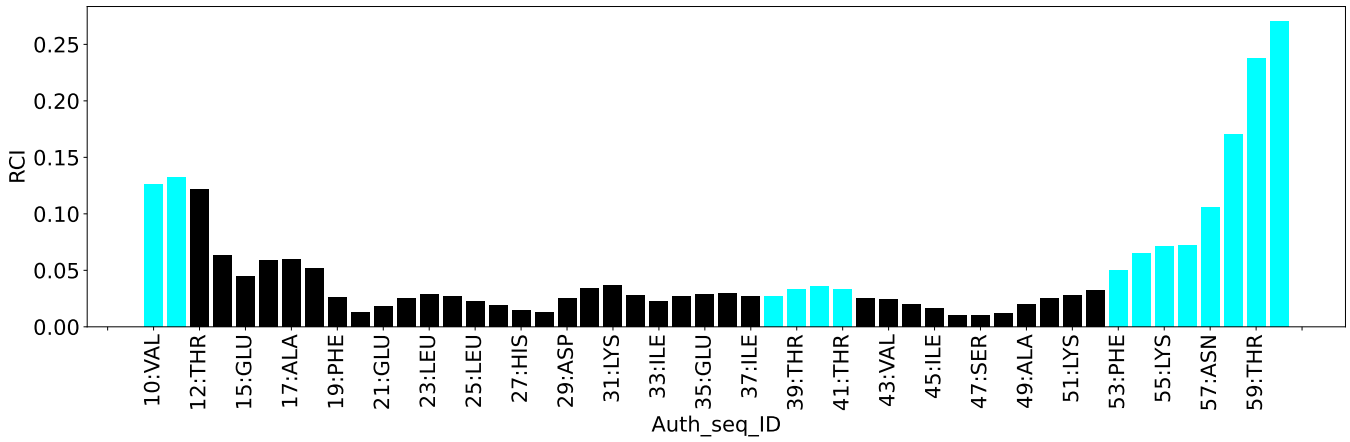
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	46	ARG	HD2	1.24	1.97 – 4.26	-8.2
1	A	46	ARG	HD3	1.24	1.81 – 4.39	-7.2

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:



8 NMR restraints analysis (i)

No restraints data found

9 Distance violation analysis (i)

No distance restraints data found

10 Dihedral-angle violation analysis (i)

Dihedral angle analysis failed due to data error in the dihedral angle restraints, possibly missing target value