wwPDB NMR Structure Validation Summary Report

Jun 24, 2024 – 12:35 PM EDT

PDB ID	:	6QFP
BMRB ID	:	25277
Title	:	Solution NMR ensemble for MlbQ at 298K compiled using the CoMAND method
Authors	:	ElGamacy, M.; Truffault, V.; Zhu, H.; Coles, M.
Deposited on	:	2019-01-10

This is a wwPDB NMR Structure Validation Summary 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:

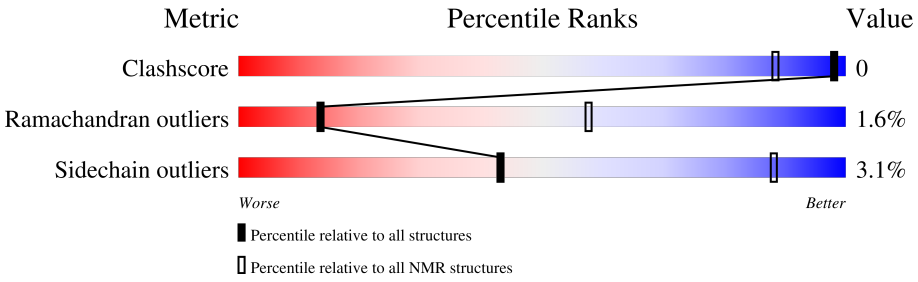
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance

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

The overall completeness of chemical shifts assignment is 86%.

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	147	52% · 8% 37%

2 Ensemble composition and analysis

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:60-A:139 (80)	0.85	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 1 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1328 atoms, of which 659 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative lipoprotein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	92	1328	422	659	112	133	2	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	expression tag	UNP W2EQT0
A	-4	HIS	-	expression tag	UNP W2EQT0
A	-3	HIS	-	expression tag	UNP W2EQT0
A	-2	HIS	-	expression tag	UNP W2EQT0
A	-1	HIS	-	expression tag	UNP W2EQT0
A	0	HIS	-	expression tag	UNP W2EQT0
A	1	SER	-	expression tag	UNP W2EQT0
A	2	SER	-	expression tag	UNP W2EQT0
A	3	GLY	-	expression tag	UNP W2EQT0
A	4	LEU	-	expression tag	UNP W2EQT0
A	5	VAL	-	expression tag	UNP W2EQT0
A	6	PRO	-	expression tag	UNP W2EQT0
A	7	ARG	-	expression tag	UNP W2EQT0
A	8	GLY	-	expression tag	UNP W2EQT0
A	9	SER	-	expression tag	UNP W2EQT0
A	10	GLY	-	expression tag	UNP W2EQT0
A	11	MET	-	expression tag	UNP W2EQT0
A	12	LYS	-	expression tag	UNP W2EQT0
A	13	GLU	-	expression tag	UNP W2EQT0
A	14	THR	-	expression tag	UNP W2EQT0
A	15	ALA	-	expression tag	UNP W2EQT0
A	16	ALA	-	expression tag	UNP W2EQT0
A	17	ALA	-	expression tag	UNP W2EQT0
A	18	LYS	-	expression tag	UNP W2EQT0
A	19	PHE	-	expression tag	UNP W2EQT0
A	20	GLU	-	expression tag	UNP W2EQT0
A	21	ARG	-	expression tag	UNP W2EQT0
A	22	GLN	-	expression tag	UNP W2EQT0
A	23	HIS	-	expression tag	UNP W2EQT0
A	24	MET	-	expression tag	UNP W2EQT0
A	25	ASP	-	expression tag	UNP W2EQT0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	SER	-	expression tag	UNP W2EQT0
A	27	PRO	-	expression tag	UNP W2EQT0
A	28	ASP	-	expression tag	UNP W2EQT0
A	29	LEU	-	expression tag	UNP W2EQT0
A	30	GLY	-	expression tag	UNP W2EQT0
A	31	THR	-	expression tag	UNP W2EQT0
A	32	ASP	-	expression tag	UNP W2EQT0
A	33	ASP	-	expression tag	UNP W2EQT0
A	34	ASP	-	expression tag	UNP W2EQT0
A	35	ASP	-	expression tag	UNP W2EQT0
A	36	LYS	-	expression tag	UNP W2EQT0

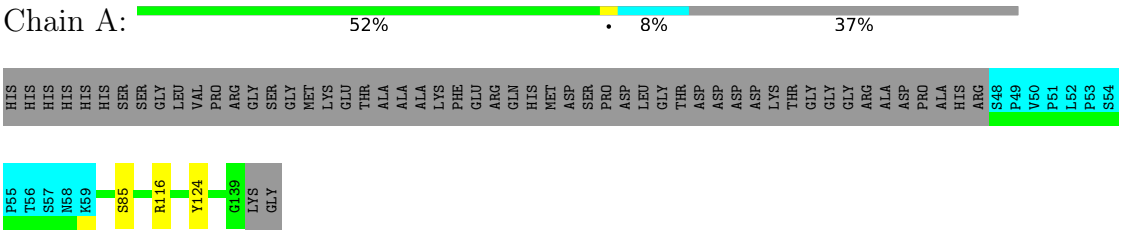
4 Residue-property plots

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: Putative lipoprotein

Chain A:

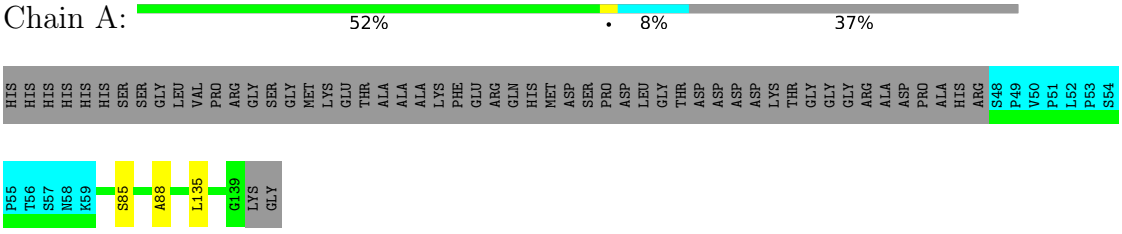


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: Putative lipoprotein

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *R-factor based MD frame picking.*

Of the 10000 calculated structures, 10 were deposited, based on the following criterion: *CNH-NOESY based R-factor.*

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

Software name	Classification	Version
Shine	structure calculation	
NAMD	structure calculation	2.12
NAMD	refinement	2.12
CoMAND	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	1002
Number of shifts mapped to atoms	961
Number of unparsed shifts	0
Number of shifts with mapping errors	41
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	86%

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	1.25±0.05	0±0/597 (0.1± 0.1%)	1.60±0.13	6±4/818 (0.7± 0.5%)
All	All	1.25	4/5970 (0.1%)	1.60	58/8180 (0.7%)

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	0.6±0.8
All	All	0	6

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	105	GLU	CB-CG	5.51	1.62	1.52	10	1
1	A	126	SER	CA-CB	5.17	1.60	1.52	3	1
1	A	116	ARG	CD-NE	5.14	1.55	1.46	4	1
1	A	111	PRO	N-CD	5.13	1.55	1.47	7	1

5 of 37 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(°)	Ideal(°)	Models	
								Worst	Total
1	A	116	ARG	NE-CZ-NH1	14.15	127.37	120.30	6	6
1	A	69	TYR	CB-CG-CD2	-11.22	114.27	121.00	2	2
1	A	98	TYR	CB-CG-CD1	-8.22	116.07	121.00	6	3
1	A	90	PHE	CB-CG-CD2	-7.89	115.28	120.80	7	2
1	A	116	ARG	NE-CZ-NH2	-7.85	116.38	120.30	6	4

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	124	TYR	Sidechain	4
1	A	98	TYR	Sidechain	1
1	A	69	TYR	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	584	570	570	0±0
All	All	5840	5700	5700	4

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:88:ALA:HB2	1:A:135:LEU:HD21	0.48	1.86	3	3
1:A:63:SER:HB2	1:A:77:HIS:CE1	0.41	2.50	9	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	79/147 (54%)	74±2 (94±2%)	4±1 (5±1%)	1±1 (2±1%)	13	57
All	All	790/1470 (54%)	739 (94%)	38 (5%)	13 (2%)	13	57

5 of 7 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	85	SER	5
1	A	77	HIS	2
1	A	70	LEU	2
1	A	68	ALA	1
1	A	111	PRO	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	62/116 (53%)	60±2 (97±2%)	2±2 (3±2%)	43	88
All	All	620/1160 (53%)	601 (97%)	19 (3%)	43	88

5 of 13 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	110	PRO	3
1	A	113	THR	3
1	A	72	PRO	2
1	A	84	PRO	2
1	A	111	PRO	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

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 86% for the well-defined parts and 84% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *Mlbq_ bmrb.str*

7.1.1 Bookkeeping

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	1002
Number of shifts mapped to atoms	961
Number of unparsed shifts	0
Number of shifts with mapping errors	41
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 41) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	27	PRO	CA	63.23	0.05	1
1	A	27	PRO	CB	32.05	0.05	1
1	A	27	PRO	CG	27.11	0.05	1
1	A	27	PRO	CD	50.47	0.05	1
1	A	28	ASP	H	8.26	0.02	1
1	A	28	ASP	C	176.13	0.05	1
1	A	28	ASP	CA	54.17	0.05	1
1	A	28	ASP	CB	40.86	0.05	1
1	A	28	ASP	N	120.03	0.05	1
1	A	29	LEU	H	8.12	0.02	1
1	A	29	LEU	C	177.87	0.05	1
1	A	29	LEU	CA	55.16	0.05	1
1	A	29	LEU	CB	42.27	0.05	1
1	A	29	LEU	CG	26.82	0.05	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	29	LEU	N	123.03	0.05	1
1	A	30	GLY	H	8.41	0.02	1
1	A	30	GLY	CA	45.34	0.05	1
1	A	30	GLY	N	109.35	0.05	1
1	A	44	PRO	CA	63.59	0.05	1
1	A	44	PRO	CB	31.99	0.05	1
1	A	44	PRO	CG	27.11	0.05	1
1	A	44	PRO	CD	50.72	0.05	1
1	A	45	ALA	H	8.35	0.02	1
1	A	45	ALA	CA	52.67	0.05	1
1	A	45	ALA	N	122.14	0.05	1
1	A	140	LYS	H	7.76	0.02	1
1	A	140	LYS	HA	4.32	0.02	1
1	A	140	LYS	HB2	1.67	0.02	1
1	A	140	LYS	HB3	1.67	0.02	1
1	A	140	LYS	C	175.82	0.05	1
1	A	140	LYS	CA	55.68	0.05	1
1	A	140	LYS	CB	32.8	0.05	1
1	A	140	LYS	CG	24.96	0.05	1
1	A	140	LYS	CD	28.45	0.05	1
1	A	140	LYS	CE	41.36	0.05	1
1	A	140	LYS	N	119.83	0.05	1
1	A	141	GLY	H	7.96	0.02	1
1	A	141	GLY	HA2	3.72	0.02	1
1	A	141	GLY	HA3	3.72	0.02	1
1	A	141	GLY	CA	46.1	0.05	1
1	A	141	GLY	N	114.86	0.05	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	93	0.10 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	84	0.14 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	63	0.19 ± 0.09	None needed (< 0.5 ppm)
^{15}N	74	0.51 ± 0.52	None needed (imprecise)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 86%, i.e. 858 atoms were assigned a chemical shift out of a possible 1000. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	340/397 (86%)	143/163 (88%)	133/160 (83%)	64/74 (86%)
Sidechain	492/540 (91%)	334/357 (94%)	152/170 (89%)	6/13 (46%)
Aromatic	26/63 (41%)	25/31 (81%)	0/29 (0%)	1/3 (33%)
Overall	858/1000 (86%)	502/551 (91%)	285/359 (79%)	71/90 (79%)

7.1.4 Statistically unusual chemical shifts

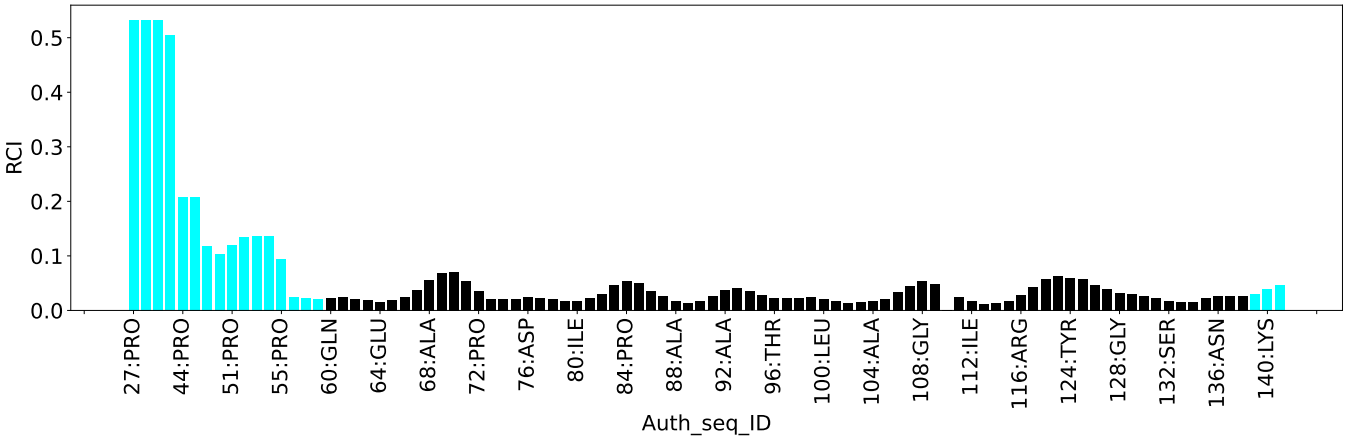
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	73	LEU	HD11	-0.67	-0.61 – 2.12	-5.2
1	A	73	LEU	HD12	-0.67	-0.61 – 2.12	-5.2
1	A	73	LEU	HD13	-0.67	-0.61 – 2.12	-5.2

7.1.5 Random Coil Index (RCI) plots

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
0.0
0.1
0.2
0.3
0.4
0.5
27:PRO
44:PRO
51:PRO
55:PRO
60:GLN
64:GLU
68:ALA
72:PRO
76:ASP
80:ILE
84:PRO
88:ALA
92:ALA
96:THR
100:LEU
104:ALA
108:GLY
112:ILE
116:ARG
124:TYR
128:GLY
132:SER
136:ASN
140:LYS
Auth_seq_ID