wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 05:17 PM BST

PDB ID	:	1TM9
Title	:	NMR Structure of gene target number gi3844938 from Mycoplasma genitalium: Berkeley Structural Genomics Center
Authors	:	Pelton, J.G.; Shi, J.; Yokota, H.; Kim, R.; Wemmer, D.E.; Berkeley Structural Genomics Center (BSGC)
Deposited on	:	2004-06-10

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
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

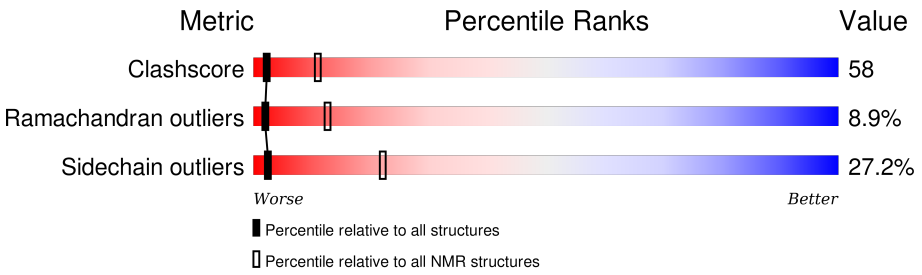
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 85%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	137	7% 70% 10% 13%

2 Ensemble composition and analysis (i)

This entry contains 26 models. Model 20 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:6-A:77, A:83-A:111, A:118-A:135 (119)	0.60	20

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 5 single-model clusters were found.

Cluster number	Models
1	3, 4, 5, 6, 9, 10, 12, 14, 15, 17, 19, 20, 21, 23, 24, 25
2	1, 7, 8
3	2, 26
Single-model clusters	11; 13; 16; 18; 22

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2186 atoms, of which 1082 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hypothetical protein MG354.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	137	2186	702	1082	178	219	5	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: Hypothetical protein MG354

Chain A: 7% 70% 10% 13%

M1 E2 Q3 N4 N5 I6 K7 E8 Q9 L10 I11 S12 F13 F14 N15 Q16 A17 C18 S19 T20 H21 Q22 E23 R24 L25 D26 F27 I28 C29 S30 T31 R32 E33 S34 D35 T36 F37 S38 S39 V40 D41 V42 P43 L44 E45 P46 I47 K48 N49 I50 I51 E52 I53 T54 K55 D56 E57 N58 Q59 D60

I61 E62 I63 T64 K65 I66 A67 V68 N69 N70 I71 K72 T73 L74 S75 S76 V77 G78 A79 T80 G81 Q82 Y83 M84 A85 S86 F87 F88 S89 T90 N91 S92 E93 P94 A95 I96 I97 F98 C99 V100 I101 Y102 F103 L104 Y105 H106 F107 G108 F109 L110 K111 D112 N113 N114 K115 K116 Q117 I118 I119 K120

K121 A122 Y123 E124 T125 I126 A127 D128 N129 I130 A131 D132 Y133 L134 N135 E136 N137

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

The representative model is number 20. Colouring as in section 4.1 above.

- Molecule 1: Hypothetical protein MG354

Chain A: 13% 47% 26% 13%

M1 E2 Q3 N4 N5 I6 K7 E8 Q9 L10 I11 S12 F13 F14 N15 Q16 A17 C18 S19 T20 H21 Q22 E23 D26 F27 I28 C29 S30 T31 R32 E33 S34 D35 T36 F37 S38 S39 V40 D41 V42 P43 L44 E45 P46 I47 K48 N49 I50 I51 E52 I53 T54 K55 Q59 D60 I61 E62 I63

T64 K65 I66 A67 V68 N69 N70 I71 K72 T73 L74 S75 S76 V77 G78 A79 T80 G81 Q82 Y83 M84 A85 S86 F87 F88 S89 T90 N91 S92 E93 P94 A95 I96 I97 F98 C99 V100 I101 Y102 F103 L104 Y105 H106 F107 G108 F109 L110 K111 D112 N113 N114 K115 K116 Q117 I118 I119 K120 K121 A122 Y123

E124 T125 I126 A127 D128 N129 I130 A131 D132 Y133 L134 N135 E136 N137

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 26 were deposited, based on the following criterion: *structures with favorable non-bond energy, structures with the least restraint violations, target function*.

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

Software name	Classification	Version
DYANA	structure solution	1.5
DYANA	refinement	1.5

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)	BMRB entry 6244
Number of chemical shift lists	1
Total number of shifts	1501
Number of shifts mapped to atoms	1501
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

No validations of the models with respect to experimental NMR restraints is performed at this time.

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	965	959	959	111±12
All	All	25090	24934	24934	2888

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 58.

5 of 959 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:77:VAL:HG11	1:A:87:PHE:CD2	1.06	1.84	15	1
1:A:85:ALA:HB2	1:A:130:ILE:HD11	1.04	1.16	8	26
1:A:13:PHE:O	1:A:15:ASN:N	1.03	1.91	26	1
1:A:66:ILE:HD11	1:A:101:ILE:HD11	0.96	1.36	17	2
1:A:14:PHE:CE1	1:A:100:VAL:HG21	0.94	1.96	13	3

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	119/137 (87%)	72±4 (61±4%)	36±4 (31±3%)	11±2 (9±2%)	2 13
All	All	3094/3562 (87%)	1876 (61%)	944 (31%)	274 (9%)	2 13

5 of 44 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	76	SER	21
1	A	134	LEU	20
1	A	35	ASP	20
1	A	6	ILE	18
1	A	20	THR	18

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	111/126 (88%)	81±5 (73±4%)	30±5 (27±4%)	2 22
All	All	2886/3276 (88%)	2101 (73%)	785 (27%)	2 22

5 of 84 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	64	THR	26
1	A	71	ILE	24
1	A	7	LYS	23
1	A	32	ARG	22
1	A	72	LYS	22

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 carbohydrates 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 85% for the well-defined parts and 84% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 6244

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	1501
Number of shifts mapped to atoms	1501
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	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
$^{13}C_{\alpha}$	137	-0.78 ± 0.20	Should be applied
$^{13}C_{\beta}$	134	0.07 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	132	-0.36 ± 0.17	None needed (< 0.5 ppm)
^{15}N	132	0.16 ± 0.26	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 85%, i.e. 1268 atoms were assigned a chemical shift out of a possible 1495. 12 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	586/589 (99%)	235/235 (100%)	235/238 (99%)	116/116 (100%)
Sidechain	628/762 (82%)	385/444 (87%)	243/291 (84%)	0/27 (0%)

Continued on next page...

Continued from previous page...

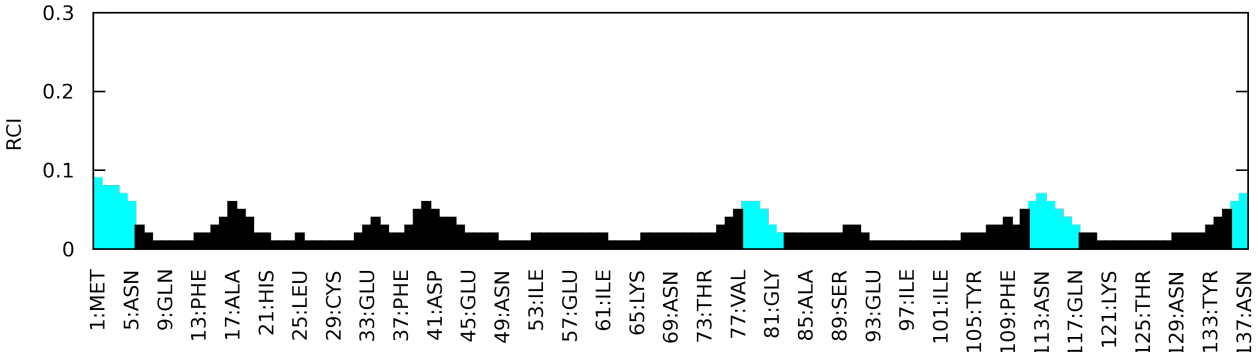
	Total	^{1}H	^{13}C	^{15}N
Aromatic	54/144 (38%)	38/78 (49%)	16/64 (25%)	0/2 (0%)
Overall	1268/1495 (85%)	658/757 (87%)	494/593 (83%)	116/145 (80%)

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.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	49	ASN	HB3	0.99	4.41 – 1.11	-5.4

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.

Random coil index (RCI) for chain A: