WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:29 pm BST

PDB ID	:	2KGW
Title	:	Solution Structure of the carboxy-terminal domain of OmpATb, a pore forming protein from Mycobacterium tuberculosis
Authors	:	Yang, Y.; Auguin, D.; Delbecq, S.; Hoh, F.; Dumas, E.; Molle, V.; Saint, N.
Deposited on	:	2009-03-20

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

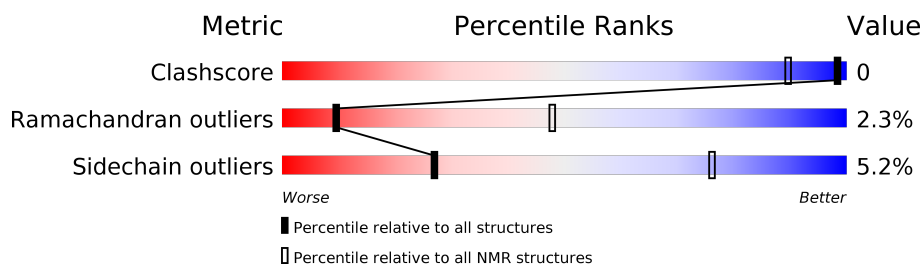|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 was not calculated.

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	Percentile Ranks	Value
Clashscore		0
Ramachandran outliers		2.3%
Sidechain outliers		5.2%

Worse   *Better*

■ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	129	84% 9% 6%

2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 4 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:206-A:326 (121)	0.36	4

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1841 atoms, of which 922 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Outer membrane protein A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	129	1841	567	922	168	182	2	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.

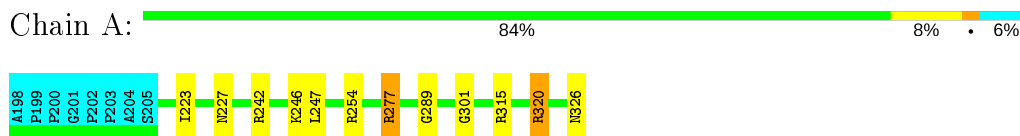• Molecule 1: Outer membrane protein A

Chain A:
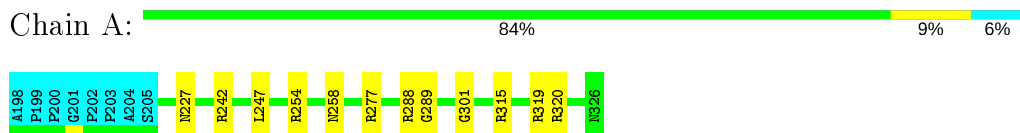
4.2 Scores per residue for each member of the ensemble
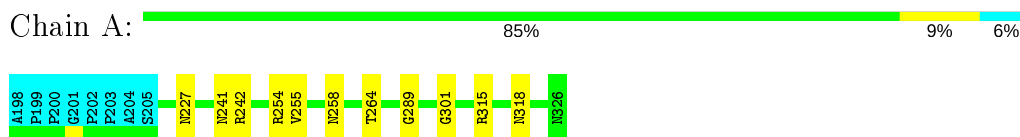
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
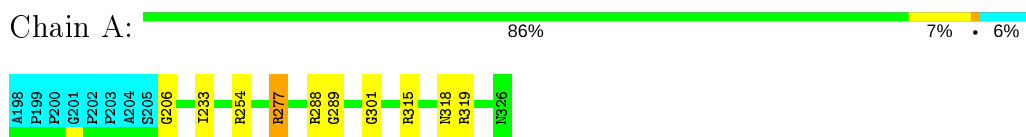
• Molecule 1: Outer membrane protein A

Chain A:

4.2.2 Score per residue for model 2

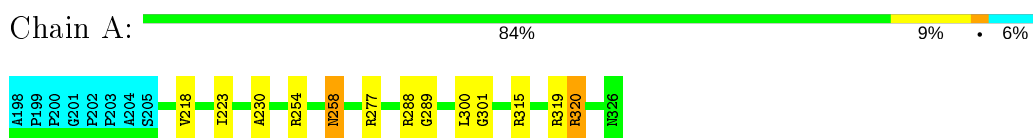• Molecule 1: Outer membrane protein A

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Outer membrane protein A

Chain A:

4.2.4 Score per residue for model 4 (medoid)

• Molecule 1: Outer membrane protein A

Chain A:

4.2.5 Score per residue for model 5

• Molecule 1: Outer membrane protein A

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: Outer membrane protein A

Chain A:

4.2.7 Score per residue for model 7

• Molecule 1: Outer membrane protein A

Chain A:

4.2.8 Score per residue for model 8
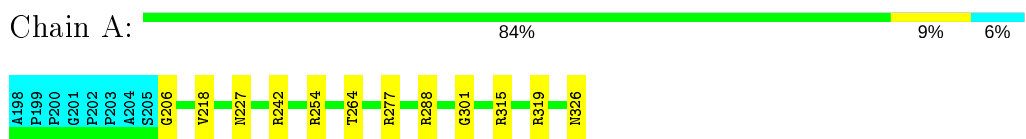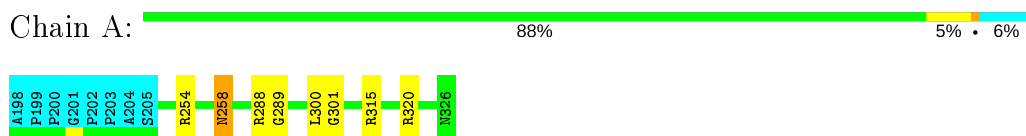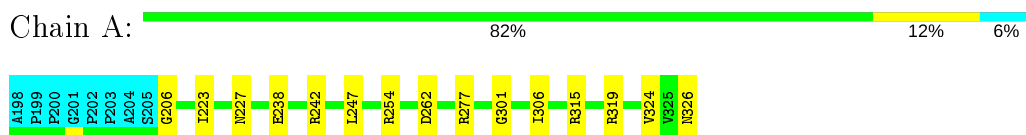
• Molecule 1: Outer membrane protein A

Chain A:

84%　　　　　　　　　　9%　　6%

A198 P199 P200 G201 P202 P203 A204 S205 A230 R242 R254 N258 R288 G289 L300 G301 R315 N318 R319 N326
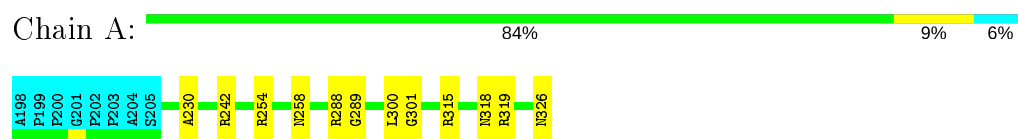
4.2.9 Score per residue for model 9

• Molecule 1: Outer membrane protein A

Chain A:

84%　　　　　　　　　　9%　　6%

A198 P199 P200 G201 P202 P203 A204 S205 V218 I223 N227 K246 L247 T256 R277 R288 G289 G301 N318 R319 N326
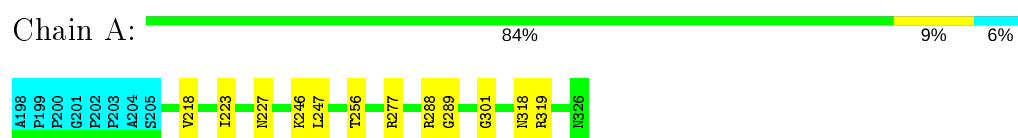
4.2.10 Score per residue for model 10

• Molecule 1: Outer membrane protein A

Chain A:

83%　　　　　　　　　9%　•　6%

A198 P199 P200 G201 P202 P203 A204 S205 V218 G226 D236 L247 V255 T256 N263 T264 S274 R288 G301 R315 R319 R320 N326
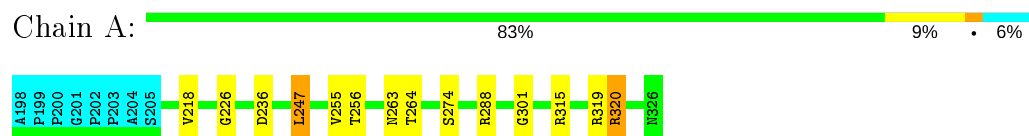
5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing, minimization.*

Of the 200 calculated structures, 10 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
AMBER	refinement	8
CYANA	structure solution	2.1

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	871	874	873	0±0
All	All	8710	8740	8730	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	Models Total
1:A:258:ASN:HB3	1:A:300:LEU:CD1	0.45	2.42	4	3
1:A:247:LEU:HD21	1:A:255:VAL:HG11	0.43	1.90	10	1

5.2 Torsion angles (i)

5.2.1 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	120/129 (93%)	111±2 (93±2%)	6±2 (5±2%)	3±0 (2±0%)	9	48
All	All	1200/1290 (93%)	1112 (93%)	60 (5%)	28 (2%)	9	48

All 6 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	301	GLY	10
1	A	289	GLY	7
1	A	227	ASN	5
1	A	206	GLY	3
1	A	230	ALA	2
1	A	226	GLY	1

5.2.2 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	90/95 (95%)	85±2 (95±2%)	5±2 (5±2%)	27	76
All	All	900/950 (95%)	853 (95%)	47 (5%)	27	76

All 20 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	247	LEU	4
1	A	218	VAL	4
1	A	288	ARG	4
1	A	318	ASN	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	223	ILE	4
1	A	326	ASN	4
1	A	277	ARG	3
1	A	258	ASN	3
1	A	320	ARG	3
1	A	264	THR	3
1	A	246	LYS	2
1	A	262	ASP	1
1	A	233	ILE	1
1	A	274	SER	1
1	A	236	ASP	1
1	A	241	ASN	1
1	A	238	GLU	1
1	A	306	ILE	1
1	A	263	ASN	1
1	A	324	VAL	1

5.2.3   RNA (i)

There are no RNA molecules in this entry.

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

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

5.4   Carbohydrates (i)

There are no carbohydrates in this entry.

5.5   Ligand geometry (i)

There are no ligands in this entry.

5.6   Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

No chemical shift data were provided