WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

Feb 24, 2022 – 08:27 AM EST

PDB ID	:	1ZK6
Title	:	NMR solution structure of B. subtilis PrsA PPIase
Authors	:	Tossavainen, H.; Permi, P.; Purhonen, S.L.; Sarvas, M.; Kilpelainen, I.; Seppala, R.
Deposited on	:	2005-05-02

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

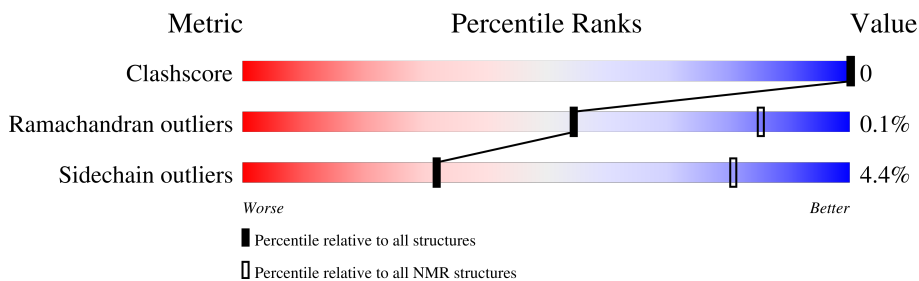|		
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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		0.1%
Sidechain outliers		4.4%

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	93	87% · 9% ·

2 Ensemble composition and analysis ⓘ

This entry contains 15 models. Model 13 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:117-A:167, A:175-A:206 (83)	0.23	13

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 6 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1434 atoms, of which 719 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Foldase protein prsA.

| Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	91	1434	452	719	118	144	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	114	GLY	-	cloning artifact	UNP P24327
A	115	SER	-	cloning artifact	UNP P24327

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.
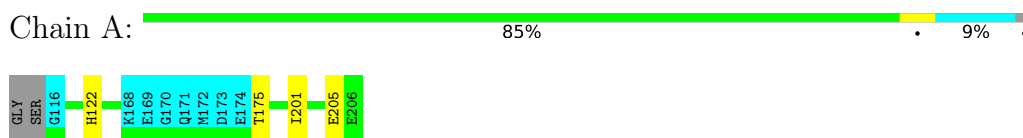
• Molecule 1: Foldase protein prsA

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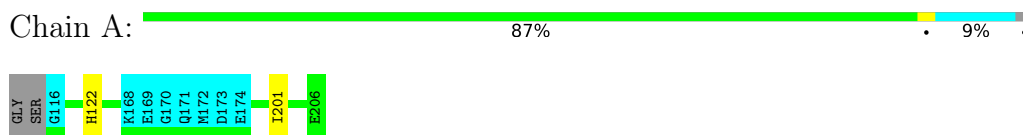
4.2 Scores per residue for each member of the ensemble
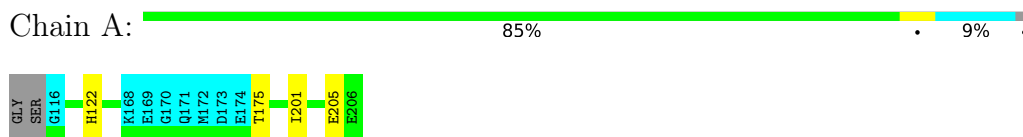
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Foldase protein prsA

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: Foldase protein prsA

Chain A:

4.2.3   Score per residue for model 3

• Molecule 1: Foldase protein prsA

Chain A: 86%   •   9%   •

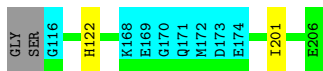### 4.2.4   Score per residue for model 4

• Molecule 1: Foldase protein prsA

Chain A: 87%   •   9%   •

4.2.5   Score per residue for model 5

• Molecule 1: Foldase protein prsA

Chain A: 87%   •   9%   •

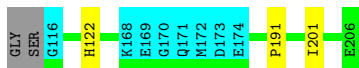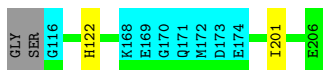### 4.2.6   Score per residue for model 6

• Molecule 1: Foldase protein prsA

Chain A: 86%   •   9%   •

4.2.7   Score per residue for model 7

• Molecule 1: Foldase protein prsA

Chain A: 87%   •   9%   •
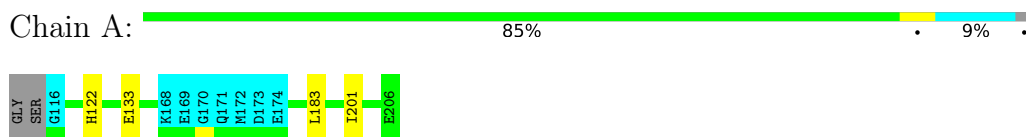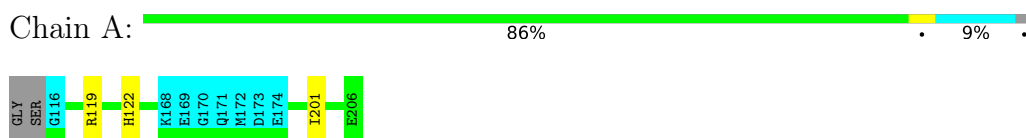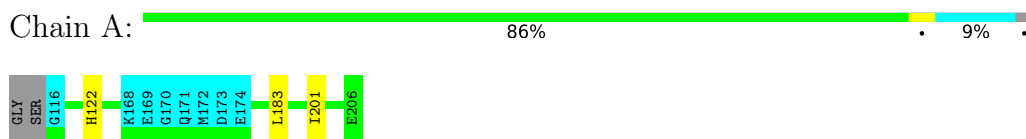
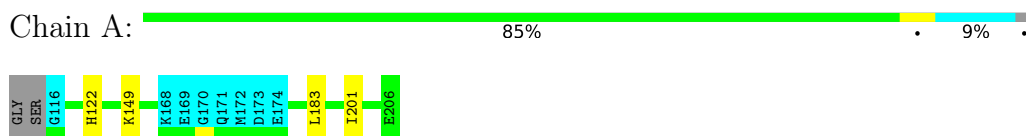#### 4.2.8 Score per residue for model 8

• Molecule 1: Foldase protein prsA

Chain A:

4.2.9 Score per residue for model 9

• Molecule 1: Foldase protein prsA

Chain A:

4.2.10 Score per residue for model 10

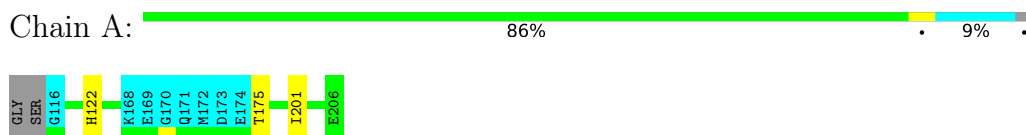• Molecule 1: Foldase protein prsA

Chain A:

4.2.11 Score per residue for model 11

• Molecule 1: Foldase protein prsA

Chain A:

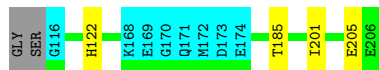#### 4.2.12 Score per residue for model 12

• Molecule 1: Foldase protein prsA

Chain A:

4.2.13 Score per residue for model 13 (medoid)

• Molecule 1: Foldase protein prsA

Chain A:

85% • 9% •

GLY SER G116 H122 K168 E169 G170 Q171 M172 D173 E174 T185 I201 E205 E206

4.2.14 Score per residue for model 14

• Molecule 1: Foldase protein prsA

Chain A:

85% • 9% •

GLY SER G116 H122 E133 K168 E169 G170 Q171 M172 D173 E174 L183 I201 E206
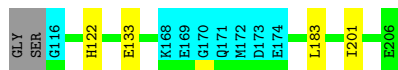
4.2.15 Score per residue for model 15

• Molecule 1: Foldase protein prsA

Chain A:

86% • 9% •

GLY SER G116 H122 K168 E169 G170 Q171 M172 D173 E174 L183 I201 E206
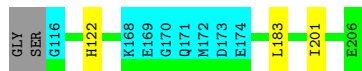
5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *simulated annealing.*

Of the 30 calculated structures, 15 were deposited, based on the following criterion: *structures with the least restraint violations, structures with the lowest energy.*

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

Software name	Classification	Version
CYANA	structure solution	1.0
Amber	refinement	8.0

No chemical shift data was provided.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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	0.72±0.01	0±0/666 (0.0± 0.0%)	0.85±0.01	0±0/886 (0.0± 0.0%)
All	All	0.72	0/9990 (0.0%)	0.85	1/13290 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	119	ARG	NE-CZ-NH1	7.00	123.80	120.30	9	1

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts ⓘ

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
All	All	9825	10005	10005	-

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 -.

There are no clashes.

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	82/93 (88%)	80±1 (97±1%)	2±1 (3±1%)	0±0 (0±0%)	54	85
All	All	1230/1395 (88%)	1198 (97%)	31 (3%)	1 (0%)	54	85

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	191	PRO	1

6.3.2 Protein sidechains ⓘ

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	69/76 (91%)	66±1 (96±1%)	3±1 (4±1%)	32	81
All	All	1035/1140 (91%)	989 (96%)	46 (4%)	32	81

All 8 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	122	HIS	15
1	A	201	ILE	15
1	A	183	LEU	5
1	A	175	THR	4
1	A	205	GLU	3
1	A	133	GLU	2
1	A	149	LYS	1
1	A	185	THR	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates ⓘ

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 ⓘ

No chemical shift data were provided