Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 10:41 PM BST

PDB ID	:	2JZV
Title	:	Solution structure of S. aureus PrsA-PPIase
Authors	:	Seppala, R.; Tossavainen, H.; Heikkinen, S.; Koskela, H.; Kontinen, V.; Permi, P.
Deposited on	:	2008-01-21

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
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:

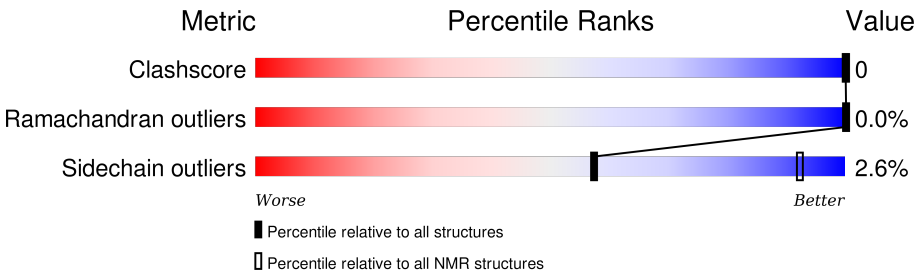
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 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	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	111	86% • 13%

2 Ensemble composition and analysis (i)

This entry contains 25 models. Model 8 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:141-A:151, A:160-A:245 (97)	0.21	8

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 4 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1745 atoms, of which 888 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	111	1745	538	888	143	175	1	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	GLY	-	EXPRESSION TAG	UNP P60747
A	-4	PRO	-	EXPRESSION TAG	UNP P60747
A	-3	LEU	-	EXPRESSION TAG	UNP P60747
A	-2	GLY	-	EXPRESSION TAG	UNP P60747
A	-1	SER	-	EXPRESSION TAG	UNP P60747

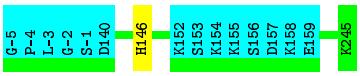
4 Residue-property plots (i)

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: Foldase protein prsA

Chain A:



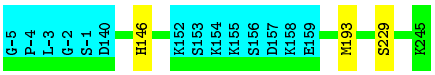
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Foldase protein prsA

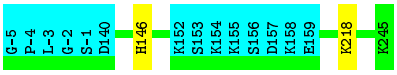
Chain A:



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: 85% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 Q174 L220 K245

4.2.4 Score per residue for model 4

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 Q174 K245

4.2.5 Score per residue for model 5

- Molecule 1: Foldase protein prsA

Chain A: 85% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K164 D213 K245

4.2.6 Score per residue for model 6

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K218 K245

4.2.7 Score per residue for model 7

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K245

4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K245

4.2.9 Score per residue for model 9

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 Q174 K245

4.2.10 Score per residue for model 10

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 Q174 K245

4.2.11 Score per residue for model 11

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K245

4.2.12 Score per residue for model 12

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K214 K245

4.2.13 Score per residue for model 13

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K218 K245

4.2.14 Score per residue for model 14

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K218 K245

4.2.15 Score per residue for model 15

- Molecule 1: Foldase protein prsA

Chain A: 83% 5% 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 Q174 D213 K214 D215 K224 K245

4.2.16 Score per residue for model 16

- Molecule 1: Foldase protein prsA

Chain A: 84% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 D213 K218 D244 K245

4.2.17 Score per residue for model 17

- Molecule 1: Foldase protein prsA

Chain A: 85% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 D163 Q174 K245

4.2.18 Score per residue for model 18

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K245

4.2.19 Score per residue for model 19

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K245

4.2.20 Score per residue for model 20

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K245

4.2.21 Score per residue for model 21

- Molecule 1: Foldase protein prsA

Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K175 K245

4.2.22 Score per residue for model 22

- Molecule 1: Foldase protein prsA

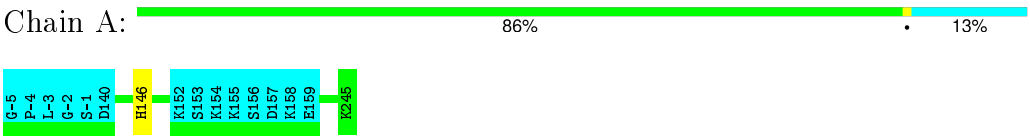
Chain A: 86% • 13%

G-5 P-4 L-3 G-2 S-1 D140 H146 K152 S153 K154 K155 S156 D157 K158 E159 K190 K245

4.2.23 Score per residue for model 23

- Molecule 1: Foldase protein prsA

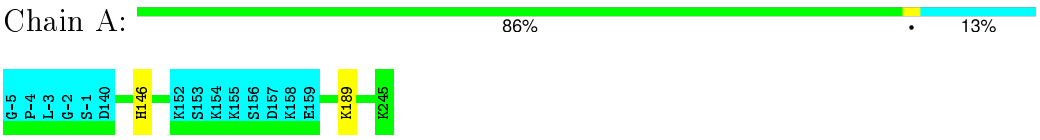
Chain A:



4.2.24 Score per residue for model 24

- Molecule 1: Foldase protein prsA

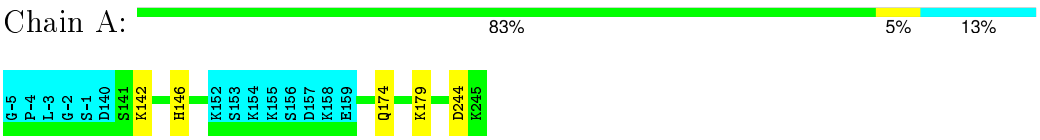
Chain A:



4.2.25 Score per residue for model 25

- Molecule 1: Foldase protein prsA

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 400 calculated structures, 25 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 solution	2.1
AMBER	refinement	8.0

No chemical shift data was provided. 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
All	All	18875	19525	19525	-

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 (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	96/111 (86%)	95±1 (99±1%)	1±1 (1±1%)	0±0 (0±0%)	100	100
All	All	2400/2775 (86%)	2381 (99%)	18 (1%)	1 (0%)	100	100

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	229	SER	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	82/94 (87%)	80±1 (97±1%)	2±1 (3±1%)	57 92
All	All	2050/2350 (87%)	1996 (97%)	54 (3%)	57 92

All 17 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	146	HIS	25
1	A	174	GLN	7
1	A	218	LYS	5
1	A	213	ASP	3
1	A	244	ASP	2
1	A	215	ASP	1
1	A	193	MET	1
1	A	163	ASP	1
1	A	190	LYS	1
1	A	175	LYS	1
1	A	179	LYS	1
1	A	142	LYS	1
1	A	189	LYS	1
1	A	214	LYS	1
1	A	220	LEU	1
1	A	224	LYS	1
1	A	164	LYS	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 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)

No chemical shift data were provided