wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 08:31 PM BST

PDB ID	:	2FYH
Title	:	Solution structure of the 2'-5' RNA ligase-like protein from Pyrococcus furiosus
Authors	:	Okada, K.; Matsuda, T.; Sakamoto, T.; Muto, Y.; Yokoyama, S.; Kanai, A.; Kawai, G.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2006-02-08

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:

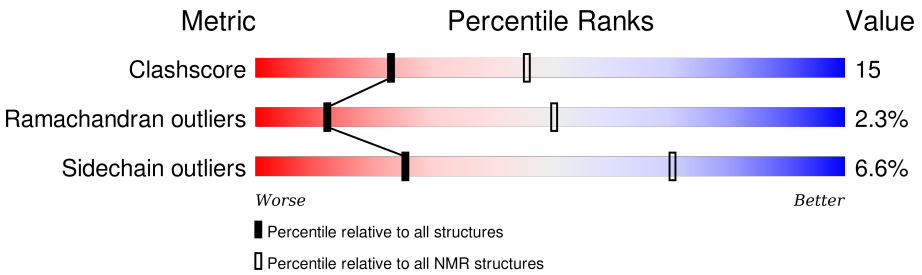
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 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	190	67% 28% • 5%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:1-A:181 (181)	0.30	11

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 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3141 atoms, of which 1599 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called putative integral membrane transport protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	190	3141	996	1599	267	277	2	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	185	HIS	-	EXPRESSION TAG	UNP Q8U4Q3
A	186	HIS	-	EXPRESSION TAG	UNP Q8U4Q3
A	187	HIS	-	EXPRESSION TAG	UNP Q8U4Q3
A	188	HIS	-	EXPRESSION TAG	UNP Q8U4Q3
A	189	HIS	-	EXPRESSION TAG	UNP Q8U4Q3
A	190	HIS	-	EXPRESSION TAG	UNP Q8U4Q3

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: putative integral membrane transport protein

Chain A: 67% 28% • 5%

M1 R2 A3 F4 I5 A6 I7 D8 V9 S10 R14 V18 Q21 K30 F33 V34 E35 F39 L43 K44 F45 L46 I49 Q53 E56 I57 K58 K59 I64 A65 K66 K67 Y68 H71 N74 V75 F81 P82 N83 W91 E95 N96 D97

I103 D108 D109 E110 L111 L114 F122 I126 F133 V134 K143 N149 E150 V157 E158 A159 K164 F180 E181 L182 S183 E184 H185 H186 H187 H188 H189 H190

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

The representative model is number 11. Colouring as in section 4.1 above.

- Molecule 1: putative integral membrane transport protein

Chain A: 63% 31% • 5%

M1 R2 A3 F4 I5 D8 V9 S10 R14 V18 Q21 D22 Y23 I24 K30 I31 K32 F33 V34 E35 R36 L43 K44 I49 I57 K58 K59 I64 A65 K66 K69 K70 H71 N74 V75 R76 F81 P82 N83 I90 W91 E95 N96 E106 I107

D108 D109 E110 L111 A112 K113 L114 G115 F116 F122 H125 I126 R130 V131 K132 F133 V134 K135 L140 L144 K145 E146 L147 A148 I156 V157 E158 A159 I160 A178 R179 L182 S183 E184 H185 H186 H187 H188 H189 H190

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 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	2.0.17
CYANA	refinement	2.0.17

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
1	A	1459	1529	1529	44±5
All	All	29180	30580	30580	876

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 15.

5 of 315 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:3:ALA:HB3	1:A:49:ILE:HD12	1.06	1.24	18	15
1:A:7:ILE:HG21	1:A:157:VAL:HG22	0.95	1.37	16	15
1:A:111:LEU:HD23	1:A:114:LEU:HD12	0.93	1.40	5	18
1:A:159:ALA:HB1	1:A:180:PHE:O	0.86	1.70	16	13
1:A:87:VAL:HG21	1:A:144:LEU:HD11	0.83	1.50	5	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	180/190 (95%)	145±3 (81±2%)	30±3 (17±2%)	4±1 (2±1%)	12 51
All	All	3600/3800 (95%)	2909 (81%)	608 (17%)	83 (2%)	12 51

5 of 17 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	10	SER	19
1	A	35	GLU	18
1	A	96	ASN	14
1	A	25	GLY	8
1	A	65	ALA	5

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	155/164 (95%)	145±3 (93±2%)	10±3 (7±2%)	25 71
All	All	3100/3280 (95%)	2896 (93%)	204 (7%)	25 71

5 of 59 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	83	ASN	18
1	A	8	ASP	9
1	A	67	LYS	8
1	A	44	LYS	8
1	A	96	ASN	8

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