J Biomol NMR (2013) 57:363–368
DOI 10.1007/s10858-013-9794-y

NMR STRUCTURE NOTE

A PilT N-terminus domain protein SSO1118 from hyperthemophilic archaeon *Sulfolobus solfataricus* P2

Jinsong Xuan · Xiaxia Song · Chao Chen · Jinfeng Wang · Yingang Feng

Received: 17 September 2013 / Accepted: 23 October 2013 / Published online: 30 October 2013

Biological context

The *Pil*T *N* terminus (PIN) domains with about 130 amino acids in length comprise a very large protein family present in all three kingdoms of life (Arcus et al. 2011). In the Pfam database, the PIN-domain family (PF01850) currently contains 8,807 members from bacteria, archaea, and eukaryotes. The biological functions of PIN-domains are diverse in various species. Most PIN domains have ribonuclease activity involved in different biological process. In eukaryotes, PIN domains are involved in nonsense mediated mRNA decay (NMD), RNA interference (RNAi), ribosomal RNA processing, and RNA degradation in immune response regulation (Bleichert et al. 2006; Xu et al. 2012). In prokaryotes, the majority of PIN domain proteins are the toxic components of VapBC-type toxin-antitoxin systems for stress response, while the toxic activity comes from their nuclease activity (Arcus et al. 2011; Blower et al. 2011). Recently, PIN domains were found in Chp1 of RITS (the RNA-induced initiation of transcriptional gene silencing) complex (Schalch et al. 2011) and Rrp44 in the yeast exosome (Makino et al. 2013).

PIN-domains have poor sequence conservation but a conserved three-dimensional structure (Arcus et al. 2011). As is shown by the determined structures of many PIN-domain proteins using X-ray crystallography, PIN-domains from various organisms have a 3-layer α/β/α sandwich structure which contains a 5-stranded parallel β-sheet with the order 32145 (Arcus et al. 2011). The revealed structural fold of PIN-domains has significant similarity with the Rossmann fold, a nucleotide-binding module existing in many dehydrogenases, kinases, and flavodoxins (Rossmann et al. 1974). Besides the conserved core structure, PIN-domain proteins often contain structural decorations and variations of loop and secondary structure elements, such as different α-helix orientation, different length of β-stand, and additional α-helix or β-stand (Takeshita et al. 2007; Bunker et al. 2008). Despite the poor sequence conservation, PIN domains contain a highly conserved active site constituted by several acidic residues for metal binding and ribonuclease activity (Arcus et al. 2011).

Protein SSO1118 with the full length of 111 residues from hyperthermophilic archaeon *Sulfolobus solfataricus* P2 was annotated as a hypothetical protein conserved in *Sulfolobale* (Fig. 1a). Our previous sequence analysis and NMR chemical shift assignment studies suggested that SSO1118 is a novel putative PIN domain protein (Xuan et al. 2011). Most archaeal PIN domains are from VapC gene of VapBC toxin-antitoxin pair whose genes are in an operon in genome, but the gene of SSO1118 is alone in the genome of *S. solfataricus* P2. Blast search in PDB does not give significant hit. This indicates that SSO1118 has no significant homology with structure-known proteins. In the present study, the solution structure of SSO1118 was determined by NMR

J. Xuan (✉) · X. Song
Department of Biological Science and Engineering, School of Chemical and Biological Engineering, University of Science and Technology Beijing, 30 Xueyuan Road, Beijing 100083, China
e-mail: jsxuan@sas.ustb.edu.cn

C. Chen · Y. Feng (✉)
Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, 189 Songling Road, Qingdao 266101, China
e-mail: fengyg@qibebt.ac.cn

J. Wang
National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China

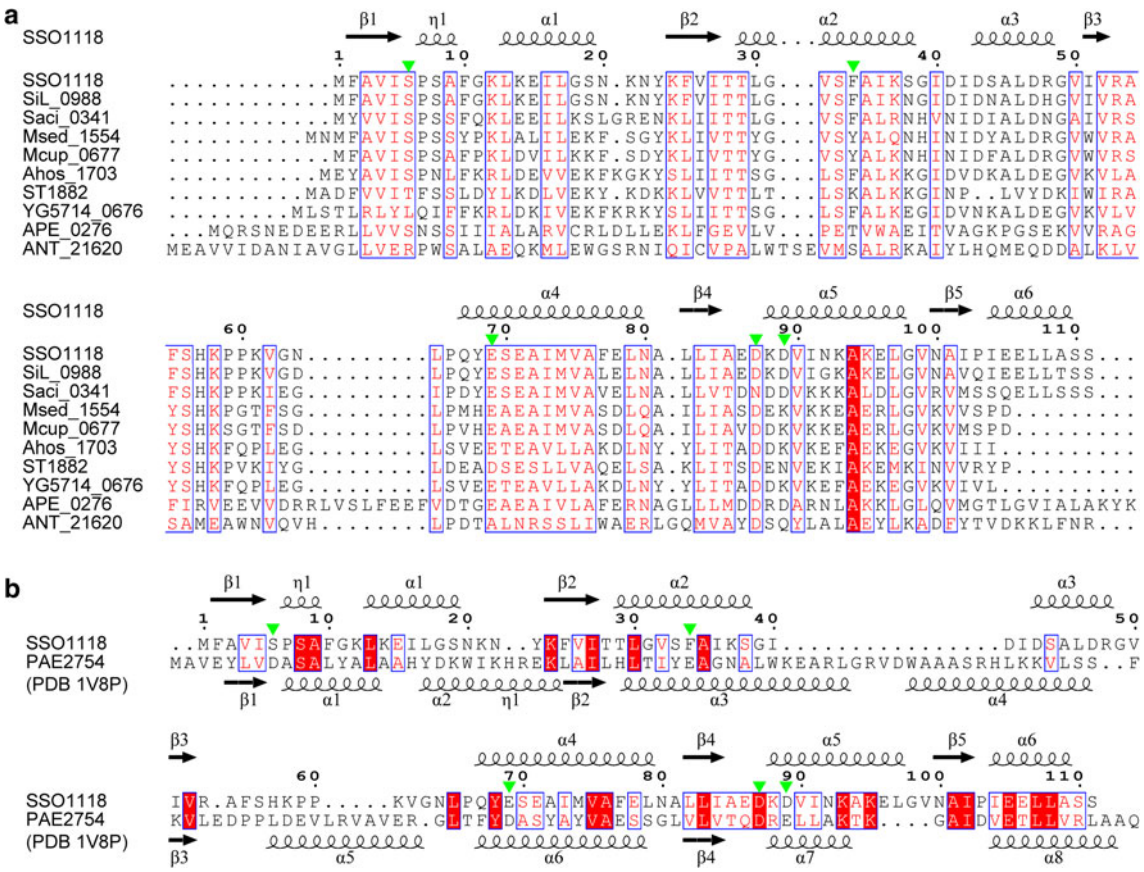

Fig. 1 Sequence alignments of SSO1118 and homologues. **a** The sequence alignment of SSO1118 and homologues identified by Blast search. **b** The structure-based sequence alignment of SSO1118 and PAE2754 (PDB 1V8P). Identical residues in the alignments are shown with *white characters in red box*, while similar residues are shown with *red characters*. The identical and similar residues are shown in *blue frame*. The secondary structures of SSO1118 and PAE2754 are shown on the *top and bottom*, respectively, of the alignment. The residues for putative active site are indicated by *green triangles*

spectroscopy. The structure determination reveals that SSO1118 is a divergent PIN domain protein with distinct structure features.

Methods and results

Sample preparation

SSO1118 protein was expressed and purified as described previously (Xuan et al. 2011). The NMR samples consisted of ~0.8 mM $^{15}N/^{13}C$-labeled SSO1118 in 90 % H_2O/10 % D_2O containing 50 mM potassium phosphate buffer (pH 7.0), 0.02 % (w/v) NaN_3, 0.02 % (w/v) sodium 2,2-dimethylsilapentane-5-sulfonate (DSS).

NMR spectroscopy and structure calculation

All NMR experiments were performed at 298 K on a Bruker DMX 600 spectrometer equipped with a z-gradient triple-resonance cryoprobe. The backbone and side chain resonance assignments for SSO1118 have been reported in previous work (Xuan et al. 2011). The distance restraints used for structure calculation were derived from NOE peaks in three-dimensional $^{1}H–^{15}N$ NOESY-HSQC and $^{1}H–^{13}C$ NOESY-HSQC spectra with a mixing time of 150 ms. Heteronuclear steady-state $^{1}H–^{15}N$ NOE, ^{15}N-transverse relaxation rate (R_2), and ^{15}N-longitudinal relaxation rate (R_1) were measured using standard pulse sequences. The rotational correlation time was estimated using R_2/R_1 ratio in secondary structure regions by the program r2r1_tm (http://www.palmer.hs.columbia.edu/software/quadric.html). All NMR data were processed with FELIX software (Accelrys Inc.) and analyzed with NMRViewJ (One Moon Scientific Inc.). Proton chemical shifts were referenced to the internal DSS, and ^{15}N and ^{13}C chemical shifts were referenced indirectly.

The solution structures of SSO1118 were determined with NOE-derived distance, backbone dihedral angle and hydrogen bond restraints (Table 1). Distance restraints were obtained from the assigned NOEs. Dihedral angle restraints were determined from chemical shifts using

Table 1 The experimental restraints and structural statistics for the 20 lowest energy structures of SSO1118

Distance restraints	
Intra-residue	857
Sequential	526
Medium	330
Long-range	487
Ambiguous	1,414
Total	3,614
Hydrogen bond restraints	104
Dihedral angle restraints	
ϕ	102
ψ	102
Total	204
Violations	
Max. NOE violation (Å)	0.163
Max. torsion angle violation (°)	3.76
PROCHECK statistics (%)	
Most favored regions	89.5
Additional allowed regions	7.8
Generously allowed regions	2.0
Disallowed regions	0.7
RMSD from mean structure (Å)	
Backbone heavy atoms	
All residue[a]	0.61 ± 0.09
Regular secondary structure[b]	0.34 ± 0.05
All heavy atoms	
All residue[a]	1.12 ± 0.06
Regular secondary structure[b]	0.78 ± 0.07

[a] The C-terminal 8 residue His-tag (LEHHHHHH) is excluded

[b] Regular secondary structure regions include residues 2–5, 7–9, 13–20, 24–38, 43–48, 51–53, 67–79, 82–84, 88–97, 100–102, and 104–109

TALOS+ (Shen et al. 2009). Hydrogen bond restraints according to the regular secondary structure patterns were incorporated into the structure calculation during the late stage of the structure refinement. Initial structures of SSO1118 were generated using the CANDID module of the CYANA software (Herrmann et al. 2002). Then, 100 structures were calculated by CNS (Brunger et al. 1998), and the 50 lowest energy structures were selected to be refined in explicit water by RECOORDScript (Nederveen et al. 2005) and CNS. The 20 lowest-energy structures in the refinement were selected to represent the final ensemble of structures for SSO1118. Secondary structures were determined by MOLMOL (Koradi et al. 1996). Because the secondary structural elements in the 20 selected structures were slightly different, a consensus secondary structure boundary was deduced from those secondary structures which were observed in more than 10 selected SSO1118 structures. All figures depicting structures were generated using PyMol (http://www.pymol.org/). The Dali server (Holm and Rosenstrom 2010) and SSM server (Krissinel and Henrick 2004) were utilized for searching the structural similarity.

Structure of SSO1118

The determined SSO1118 structure represents a 3-layer α/β/α sandwich structure containing six α-helices (α1: 13–20, α2: 29–38, α3: 43–48, α4: 67–79, α5: 88–97, α6: 104–109), a 3–10 helix (η1: 7–9), and five β-strands (β1: 2–5, β2: 24–28, β3: 51–53, β4: 82–84, β5: 100–102) with an order β1–η1–α1–β2–α2–α3–β3–α4–β4–α5–β5–α6 (Fig. 2a, b). Helices α1, α2, α3, α6, and η1 are flanked on one side of the β sheet, while helices α4 and α5 are on the other side. In this sandwich structure, the hydrophobic residues from both sides of the β sheet form a major hydrophobic core with the hydrophobic residues from flanked α helices. The 13-residues loop linking β3 and α4 ($L_{\beta3\alpha4}$) is poorly converged, suggesting the flexibility of this loop. The backbone relaxation measurements showed that NOE values of loops $L_{\beta3\alpha4}$ and $L_{\alpha1\beta2}$ were smaller than that of most other secondary structure regions and loops (Fig. 2d). Therefore, loops $L_{\beta3\alpha4}$ and $L_{\alpha1\beta2}$ are more flexible than other structural regions except the disordered C-terminal His-tag. SSO1118 is a slightly acidic protein with a predicted pI 5.83, and the electrostatic surface on the protein indicates that one side of SSO1118 is mainly negatively charged whereas the other side is positively and neutral (Fig. 2c).

Deposition of structure coordinates

The atomic coordinates and NMR-derived restraints of SSO1118 have been deposited in the Protein Data Bank (http://www.rcsb.org/) with accession code 2MDT.

Discussion and conclusions

Structure comparison

Similar structures to SSO1118 were searched using Dali and SSM servers in PDB (Fig. 3a). The results indicate that the most similar structures to SSO1118 are PIN domain proteins from archaeal VapBC toxin-antitoxin, whereas the other PIN domain proteins, as well as Rossmann fold proteins, show more or less similarity to SSO1118 (Fig. 3a–c). All of these protein structures show similar α/β/α sandwich fold with a parallel β-sheet in the center of the core structure. The largest structural differences can be found in the number, length, and orientations of the flanked helices. PAE2754 (PDB 1V8P; Arcus et al. 2004), a VapC toxin from the hyperthermophilic

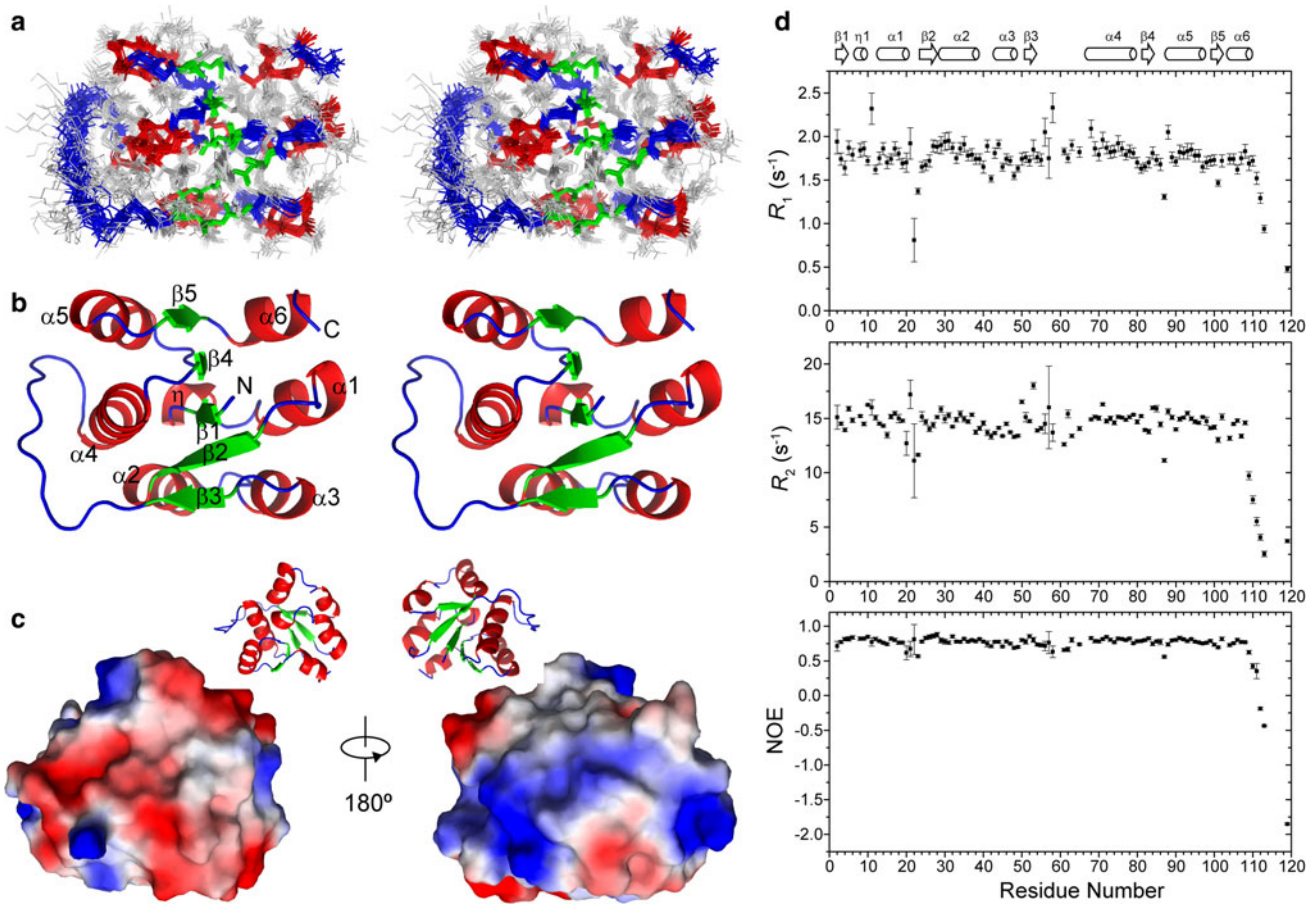

Fig. 2 Solution structure and backbone dynamics of SSO1118. **a** Stereo view of 20 structure ensemble of SSO1118. The backbone of helices, β-strands, and loops were in *red*, *green*, and *blue*, respectively. The side chains are in *light grey*. **b** Stereo view of ribbon representation of SSO1118. **c** The electrostatic surface of SSO1118. The protein orientation is indicated by the corresponding ribbon representation on the *top*. For clarification, the C-terminal 8-residue His-tag (LEHHHHHH) is not shown in this figure and Fig. 3. **d** Backbone relaxation parameters (R_1, R_2 and ^{1}H–^{15}N heteronuclear NOE) of SSO1118. The secondary structures are marked at the *top* of the figure

crenarchaeon *Pyrobaculum aerophilum*, is the protein with the highest Z-score in Dali search. PAE2754 has a very similar β-sheet to SSO1118, but the number and length of helices of PAE2754 are very different from SSO1118 (Figs. 1b, 3b). For other proteins, the structural differences can be found also in the length of the β-strands although the topology of the sheet is same (Fig. 3a). Like other PIN domain proteins, SSO1118 also shows topological similarity to Rossmann fold proteins. However, Rossmann fold proteins, compared to other PIN domain proteins, shows more structural differences from SSO1118 in term of β-stand length, helix number, length, and orientation, as well as in loop length and positions (Fig. 3c).

Some archaeal PIN domain proteins form oligomer in solution (Arcus et al. 2004; Bunker et al. 2008), while the PIN domain of human EST1A is monomer in solution because the oligomer interface of archaeal PIN domains is not conserved in EST1A (Takeshita et al. 2007). SSO1118 is a monomer in solution according to both the gel filtration during the protein purification and the overall rotational correlation time of 8.6 ± 0.3 ns estimated from the backbone relaxation R_2/R_1 data. This could be explained by the structural specificity of SSO1118, that the α5 in the dimer interface of other archaeal PIN domains is replaced by the long flexible loop $L_{\beta3\alpha4}$ (Fig. 3a, b).

Although the secondary structures show great variance in various PIN domain proteins, most PIN domains possess a conserved active site for metal binding and nuclease activity (Min et al. 2012). The SSO1118 structure has a conserved negatively charged patch forming by several acidic residues which can function as the active site like in other PIN domains (Fig. 3d, e). Particularly, the metal binding site formed by three acidic residues (E69, D87, and D89) is conserved in SSO1118, while the two other putative catalytic residues become S6 and F34. Therefore, unlike other archaeal PIN domain proteins, SSO1118 may not possess the

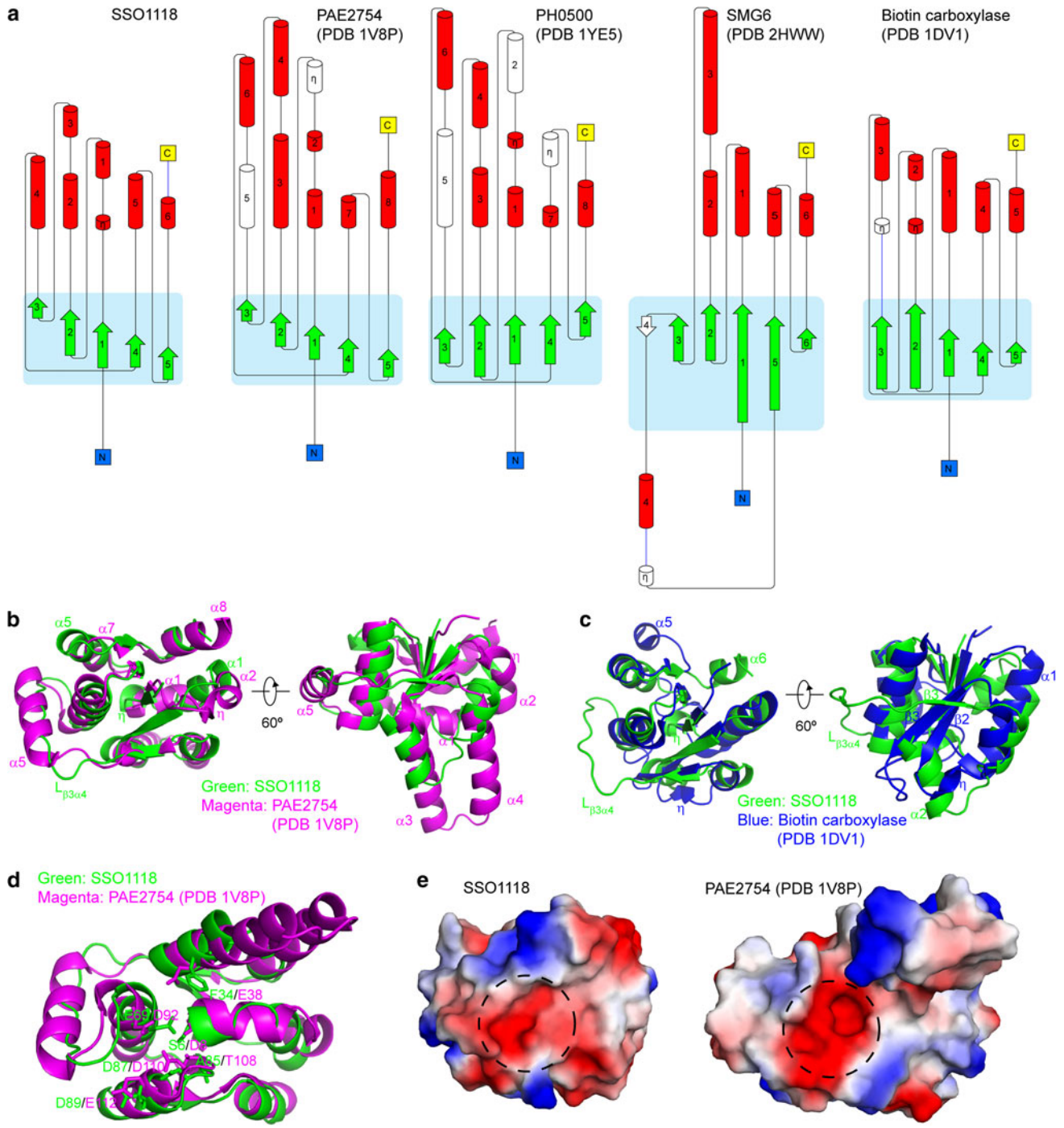

Fig. 3 Structure comparison. **a** Topology of SSO1118 and selected similar proteins identified by Dali search. The additional secondary structural elements in the similar proteins are shown in *white*. The figure was produced by Pro-origami (Stivala et al. 2011) with manual edit. **b** Structural superimposition of SSO1118 and an archaeal PIN domain protein PAE2754 (PDB 1V8P). The secondary structural elements with significant difference are labeled. **c** Structural superimposition of SSO1118 and a Rossmann fold protein (PDB 1DV1). The secondary structural elements with significant difference are labeled. **d** Putative active sites of SSO1118 and PAE2754. Putative active site residues are shown as *sticks with label*. **e** Electrostatic surfaces of SSO1118 and PAE2754. The putative active site regions are indicated by *dashed circles*

nuclease activity but still maintain a metal binding site. However, since the key active site residues in other archaeal PIN domain proteins were speculated from structure without experimental evidence, whether SSO1118 have nuclease activity and what is the physiological function of SSO1118 need further functional study in future.

Conclusions

SSO1118 has an α/β/α sandwich structure with a central parallel β-sheet. The structural analysis indicates that SSO1118 is a novel PIN domain protein with distinct structure features. The putative active site of SSO1118 is different from other archaeal PIN domain proteins, suggesting that SSO1118 may have different function.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant No. 31300635 to J.X., Grant No. 31170701 to Y.F.).

References

Arcus VL, Backbro K, Roos A, Daniel EL, Baker EN (2004) Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease. J Biol Chem 279(16): 16471–16478. doi:10.1074/jbc.M313833200

Arcus VL, McKenzie JL, Robson J, Cook GM (2011) The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array. Protein Eng Des Sel 24(1–2):33–40. doi:10.1093/protein/gzq081

Bleichert F, Grannemant S, Osheim YN, Beyer AL, Baserga SJ (2006) The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. Proc Natl Acad Sci USA 103(25):9464–9469. doi:10.1073/pnas.0603673103

Blower TR, Salmond GPC, Luisi B (2011) Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin-antitoxin partners. Curr Opin Struc Biol 21(1):109–118. doi:10.1016/j.sbi.2010.10.009

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 54:905–921. doi:10.1107/S0907444998003254

Bunker RD, McKenzie JL, Baker EN, Arcus VL (2008) Crystal structure of PAE0151 from *Pyrobaculum aerophilum*, a PIN-domain (VapC) protein from a toxin-antitoxin operon. Proteins 72(1):510–518. doi:10.1002/Prot.22048

Herrmann T, Guntert P, Wuthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319(1):209–227. doi:10.1016/S0022-2836(02)00241-3

Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38:W545–W549. doi:10.1093/Nar/Gkq366

Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14(1):51–&. doi:10.1016/0263-7855(96)00009-4

Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D 60:2256–2268. doi:10.1107/S0907444904026460

Makino DL, Baumgartner M, Conti E (2013) Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. Nature 495(7439):70–75. doi:10.1038/Nature11870

Min AB, Miallau L, Sawaya MR, Habel J, Cascio D, Eisenberg D (2012) The crystal structure of the Rv0301-Rv0300 VapBC-3 toxin-antitoxin complex from *M. tuberculosis* reveals a Mg^{2+} ion in the active site and a putative RNA-binding site. Protein Sci 21(11):1754–1767. doi:10.1002/Pro.2161

Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CAEM, Nabuurs SB, Guntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AMJJ (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins 59(4):662–672. doi:10.1002/Prot.20408

Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of a nucleotide-binding protein. Nature 250(5463):194–199. doi:10.1038/250194a0

Schalch T, Job G, Shanker S, Partridge JF, Joshua-Tor L (2011) The Chp1-Tas3 core is a multifunctional platform critical for gene silencing by RITS. Nat Struct Mol Biol 18(12):U1351–U1363. doi:10.1038/Nsmb.2151

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44(4):213–223. doi:10.1007/s10858-009-9333-z

Stivala A, Wybrow M, Wirth A, Whisstock JC, Stuckey PJ (2011) Automatic generation of protein structure cartoons with Pro-origami. Bioinformatics 27(23):3315–3316. doi:10.1093/bioinformatics/btr575

Takeshita D, Zenno S, Lee WC, Saigo K, Tanokura M (2007) Crystal structure of the PIN domain of human telomerase-associated protein EST1A. Proteins 68(4):980–989. doi:10.1002/Prot.21351

Xu JW, Peng W, Sun Y, Wang XX, Xu YH, Li XM, Gao GX, Rao ZH (2012) Structural study of MCPIP1 N-terminal conserved domain reveals a PIN-like RNase. Nucleic Acids Res 40(14): 6957–6965. doi:10.1093/Nar/Gks359

Xuan J, Song X, Wang J, Feng Y (2011) Resonance assignments of a putative PilT N-terminus domain protein SSO1118 from hyperthermophilic archaeon *Sulfolobus solfataricus* P2. Biomol NMR Assign 5(2):161–164. doi:10.1007/s12104-010-9291-0