

## Isolation and Structure Analysis of a SH3 Domain Protein from *Monascus purpureus*

Haoming Li <sup>a</sup>, Lan Gao <sup>b</sup>

School of Biosciences and Biopharmaceutics, Guangdong Pharmaceutical University, Guangzhou Higher Education Mega Center, Guangzhou 510006, Peoples Republic of China

<sup>a</sup>Corresponding author, Lihaoming@gdpu.edu.cn, <sup>b</sup>glgdpu@aliyun.com

**Keywords:** SH3 domain; homology modeling; cloning; *Monascus purpureus*; cDNA library.

**Abstract.** SH3 (Src homology 3) domain is a small modules units (~60 amino acids) that are important for the generation of protein-protein interactions in cellular signal transduction pathways. SH3 domains commonly bind to proline-rich sequences. Here, we explore the SH3 domain proteins in *Monascus purpureus*. A *M. purpureus* cDNA library was constructed, and the MpSH3 cDNA was isolated from the cDNA library by sequencing method. The cDNA was 1239 bp in length, contains a predicted 744 bp ORF that encodes 247 amino acids, the gene was designated MpSH3. The deduced amino acid sequence showed homology with SH3 domain proteins of fungi. A three-dimensional model of MpSH3 was built using homology modeling method. The 3D structure model of MpSH3 indicated that the domain is 52 residues long and is composed of five  $\beta$ -strands, and has canonical RT, N-Src loops and 310-helix. The structure analysis implicated that the MpSH3 protein can mediate protein-proteins via recognition of proline-rich peptides.

### Introduction

Modular interaction domains are conserved regions in proteins that are specialized in mediating interactions of proteins with one another. Proteins composed of multiple protein-binding domains plays integral roles in signaling by serving as scaffolds for the assembly. And regulate almost all cellular functions, including cell growth, differentiation, motility and apoptosis [1-2].

Protein interaction domains are often identified through consensus sequence motifs. SH3 domains were first noted as regions of sequence similarity between divergent signaling proteins such as the Src family of tyrosine kinases, the Crk adaptor protein, and phospholipase C- $\gamma$  [3]. Src family kinases exhibit a highly conserved structural organization that includes a myristoylated N-terminal domain, a short unique region, modular SH3 and SH2 domains, an SH2 kinase-linker region, the catalytic domain, and a C-terminal tail. SH3 domains mediate protein-protein interactions via recognition of proline-rich sequences, such as peptides that bear a PxxP core motif (where x represents any amino acid) flanked by a positively charged residue [4]. SH3 is one of the most abundant PRDs (proline-recognition domains) in vertebrates, with an estimated 300, 63 and 27 copies in the human proteome, fruit fly and yeast respectively [5]. These domains are ~60 residues long and are composed of five  $\beta$ -strands, which are sequentially joined by the RT, N-Src, and distal loops, and a short 310-helix. The most common structure formed by two or more proline residues in a row is PPII (polyproline type II), a left-handed helix with three residues per turn. The SH3 surface could recognize the PPII helix were first identified by the determination of three dimensional structures of peptide complexes of the p85 SH3 domain by NMR [6] and the Abl and Fyn SH3 domains by X-ray crystallography [7]. Both the side chains and the backbone carbonyls in a PPII structure are projected outwards from the axis of the helix, they are poised to interact with another molecule. The PPII helix is only found in sequences that are rich in proline residues. Because of the inherent 2-fold rotational pseudosymmetry of the PPII helix, a PRD can engage a proline-rich sequence in one of the two opposing orientations [8].

SH3 possess intrinsic ability to recognize a diverse set of ligands [9-10]. While many SH3 domains broadly target two main binding motifs (Class I: tXXPXXP and Class II: PXXPXt, where P is proline, t is arginine or lysine, and X is any amino acid, they have differences in binding preference

at other positions surrounding the core PXXP motif, it is important for determining their interaction specificity [11-12]. In yeast, the binding specificities of 27 yeast SH3 domains was scanned, the (R/K)xx(K/R) motif is Class III consensus recognized by the SH3 domain [13-14]. More recently, a number of alternative SH3 domain binding motifs have been identified, highlighting a wider breadth of SH3 specificities [11, 15].

Here we identified the first SH3 domain protein gene from *Monascus purpureus*, and its structure was analyzed.

## Materials and Methods

**Strains, and Growth Conditions** *M. purpureus* strain CICC 5031 was purchased from china center of industrial culture collection. The stock culture was maintained on PDA (potato dextrose agar) slants. Spores were prepared by growth on these slants for 7 days at 30°C, washed and harvested with sterile water. A suspension of spores was used to inoculate a 250 ml flask containing 40 ml liquid fermentation medium PD (potato dextrose broth), for cDNA library construction, the *M. purpureus* was incubated at 30°C at 130 rpm for 5 days. The mycelium samples were collected and washed with sterile water, frozen in liquid nitrogen, and then were stored at -80°C until required.

**CDNA Library Construction, ESTs Sequencing and Cloning of MpSH3** Total RNA was extracted from frozen mycelium (40 ml culture which incubated at 30°C, 130 rpm for 5 days) using TRIzol® reagent (Invitrogen, USA), following the instructions of manufacturer. From total RNA, the cDNA library construction and amplification were performed following the users manual of the Creator™ SMART™ cDNA Library construction Kit (Clontech, USA). The SMART cDNAs were ligated into SfiI-digested pDNR-LIB vector and transformed into *Escherichia coli* strain DH5α. Colonies were randomly picked, inoculate each colony to separated PCR reaction solutions. The colony was lysed by heating the mixed solutions at 95°C in a PTC-200 Thermocycler (MJ Research, USA) for 5 min.

After then, went to PCR amplification procedure with M13 primers provided by the Creator™ SMART™ cDNA Library construction Kit. The amplified PCR products (ESTs, expressed sequence tags) were analyzed by 1.2% agarose gel electrophoresis. When the amplified cDNA fragments were longer than 1000bp, incubated the isolated colonies and sequenced the ESTs. After sequencing and analysis, the colony containing the predicted pDNR-LIB-MpSH3 was isolated.

## Results and Discussion

To identify gene encoding SH3 domain protein in *M. purpureus*, a cDNA library of *M. purpureus* mycelium was constructed and screened with primers M13 (see method). The inserts of about 200 clones were sequenced and analyzed. The full-length MpSH3 cDNA was then obtained (Genbank accession No. GU797552). The cDNA was 1239 bp in length, contains a predicted 744 bp ORF that encodes 271 amino acids, a 184 bp non-coding region at 5' end and a 311 bp of non-coding region flank at 3' end. The predicted amino acid sequence of MpSH3 specifies a polypeptide with a molecular mass of 26.6 kDa and a pI of 6.97.

Protein sequence alignment indicates that the deduced MpSH3 protein is similar to other known SH3 domain proteins in fungi. MpSH3 had high identity with species of Eurotiales, range from 66.80% (*Aspergillus oryzae*) to 60.31% (*Penicillium solitum*); and had 60.89% (*Coccidioides immitis*) to 55.16% (*Paracoccidioides* sp.) identity with species of Onygenales; and the sequence homology is lower (26.32%) with *Saccharomyces cerevisiae* (Fig.1). Larson et al [16] indicated that the sequence identity of any two randomly chosen SH3 domains is ~30%, relatively low sequence identity. Compare the predicted secondary structure of SH3 domain from MpSH3 and yeast, each domain has 5 β-strands, and there is the highly conserved tryptophan (W156 of MpSH3 and W98 of *Sce*) in the N-Src loop (the loop connected the second and the third β-strands) which will have interacts with positively charged residues such as arginine and lysine flanked the PxxP core motif

[17]. The complex characteristics of some interaction domains make it difficult to predict protein–protein interactions by homology-based sequence search.

Mpu	MASPNDFTVAMTNRSLRNIRSELDLFLADASVITPQQHSSILSQRDLDSMMPEVVSQPRSEKTTSSAPPAATPAPTQ--PSPGP----YTPPTSQF	91
Aor	MASQSNFMAAMTNRSLRTRVLELEFLADASVITPQQLSSIVSQLPQENTER-----SASVQSRHQPPVAAPTPTQ--PSPAP----YSPPTQQF	84
Cim	--MAPNDLSTAITNRSRLRTIKTELEFLCDSSVITPQQLSSILSOLPS-----QTQLHAPVQSTADTG-----FTPPSQGL	57
Pso	MAAPSEFASAMTTRSLRMVTELEFLADASVITPQQLSSILSOLPTDSEATRAVHASPPAQQQQAPPAQHVVPVQAQPPAPYTPSYSPFVSQM	95
Psp	--MAP--KTSVYEAKRSLRTIRSELEFLADTSLITPQSOLTLSSQLPS-----DAQLQASLHSTGDS-----PPPTNQL	66
Scs	-----MSASLNRSLTNIRTELEFLKESNVIITNDVFDQIN-----	42
PredS	HHHHHHHHHHHHHHHHHHHH	HHHHHHHH
PredM	HHHHHHHHHHHHHHHHHHHH	HHHHHHHHHH
Mpu	ANATLNEKSSLHNQFSTPPP---PAYQQGPPVLSVASALYAYTPTDAGDLALQPNDRIQVLEHMNDWWRGRNERTNLEGIFPRSYVRIIEKSV	183
Aor	ANTSINEKAAVQPPQHYTP--PPPAYQAPAVLSVANALYAYTPTDAGDLALQPHDRVQVLEHMNDWWRGRNERTNLEGIFPRSYVSVIENKA-	177
Cim	DKLALNEKSSQDPYSVDVPP--PPSYASAPVLCMASALYAYTPTDAGDLALQANDRIQVLEHMNDWWRGRNERTSMEGIFPRTYVTIIDEKPA	148
Pso	ANASVNEKAYAQNQYASQPPQAPPAYQVPPVVLGLASAMYAYNETDAGDLALLQPDRIQITEHMNDWWRGRNERTGQEGIFPRSYVNVVNEKAS	188
Psp	SSLRLNEKSPQAQHYHDEVA--PPAYASGPPVLAIASALYAYAPT DAGDLALQPDRIQVLEHMNDWWRGRNERTQLEGIFPRTYVSVIEEKSI	157
Scs	KSLPVKWDPSAPRNASTTS-----LEYYEALYQFDEQQDGLGLKPGDKVQLLEKLSPEWYKG--SCNGRTGIFRANYVKAASFSGSN	123
PredS	EEEE	EEEEEE
PredM	EEEE	EEEEEE
Mpu	PPAPVQPTN-YGNMPLVVSQSGSAPQAE-----KHSRFGEGGKFKGKLGNAALFAGATMGADLVNSIF	247
Aor	---AAPTS-YGNMPLVVSQSGSAENPDDK-----KTGKFEHGGKFKGKMGNAALFAGATIGSNIVNSIF	240
Cim	--ATPQPTN-YGNMPLVANSQSSQAG-----RKPSKFEENGKFKGKMGNAALFAGATIGSNIVNGIF	226
Pso	--VSSPPTN-YGNMPLAVSQSGQFENPEDP-----KKNKLEEGGKFKGKLGNAALFAGATVGSNIVNSIF	255
Psp	---IHSPPVVSQGNAPLEVAHLGSSGSGFGGHS-----SGRPSKFEENGKFKGKMGNAALFAGATVGSKIVNGIF	228
Scs	GPSSLPPPPQYKAQELQVPTQNSATSSVQQQFP PPSVTVYVQQPQQQQPQQQQSSSHSLKSGFKLGNAALFAGATIGSDIVNNIF	210
PredS	HHH	HHH HHH
PredM	HHH	HHH HHH

Fig. 1. Alignment of MpSH3 with 5 protein sequences from fungi. The abbreviation and Genbank accession numbers are: Mpu (Monascus purpureus AIX10963), Aor (Aspergillus oryzae, XP\_001818335), Cim (Coccidioides immitis, XP\_001240512), Pso (Penicillium solitum, KJJ13022), Psp (Paracoccidioides sp., XP\_002789274) and Sce (Saccharomyces cerevisiae, AJP42275). Residues in SH3 domains of Mpu and Sce are colored. PredS and PredM are predicted secondary structure of Mpu and Sce respectively by PSIPRED v 3.3. Alpha helices are marked with H. Beta sheets are marked with E. Identical residues are shaded.

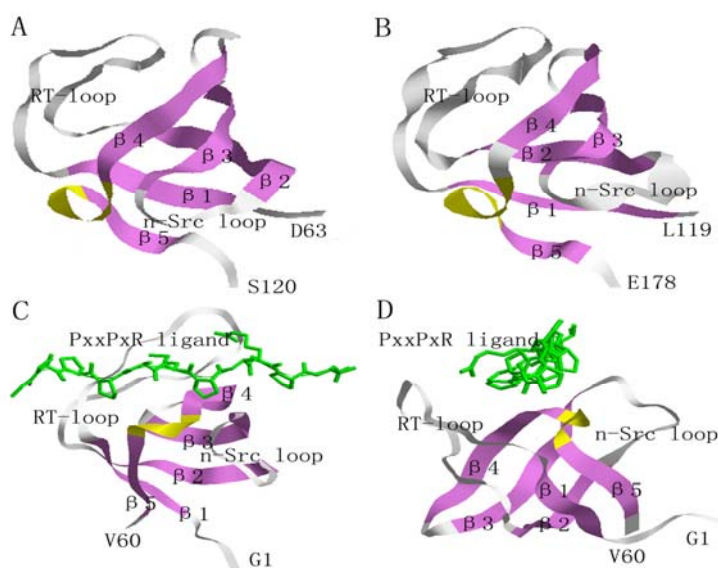


Fig. 2 3D structure of MpSH3. (A) Crystal structure of SH3 domain of human p56-lck (from 63-120 residue range; PDB code 1LCK), (B) Predicted 3D structure of MpSH3, (from 119-178 residue range, Genbank accession number: AIX10963) (C) and (D): Crystal Structure of a yeast SH3 domain protein (from 1-60 residue range) complex with a 12-mer peptide. Conserved  $\beta$ -strands and 310-helices are shown as violet and yellow ribbon diagram respectively. RT-loop and n-Src loop are shown as white ribbon diagram. In (C) and (D): The peptides are shown as green.

The three dimension model structure of MpSH3 was predicted using SWISS-PDB software. On the basis of the 2.5Å crystal structure of SH3-SH2 fragment of human p56-lck (PDB code 1LCK), p56lck is a member of the Src family of non-receptor tyrosine kinases, contains SH3 and SH2, and a tyrosine kinase domain. The structure of SH3 domain is shown in Fig.2; it's a small, independently folded module. The amino acid sequence of MpSH3 (from 119-178 residue range) has 23.81% amino acid identity with 1LCK (from 63-120 residue range). The structure of MpSH3 was successfully built as a single chain. The overall fold of the MpSH3 is very similar to that of p56-lck SH3 domain. The MpSH3 is composed of five  $\beta$ -strands, which are sequentially joined by the RT, N-Src, and distal loops, and a short  $3_{10}$ -helix. The region with structural difference is the N-Src loop connecting the  $\beta_2$  and  $\beta_3$ -strands. In Fig.2, the 1.4Å crystal structure of SH3 domain (from 1-60 residue range) from a *S. cerevisiae* hypothetical 40.4 kDa protein in complex with a peptide (EGPPPAMPARPT) was shown (PDB code 1SSH), the peptide located in the pocket formed by RT-loop and N-Src loop, the two loop

are involved in peptide recognition. Comparison with the crystal structures of the SH3 domains of 1SSH, the region with structural difference is the  $3_{10}$ -helix and the  $\beta 4$ ,  $\beta 5$ -strands.

Many SH3 domains have acidic residues localized at one end of the peptide binding surface of the SH3 domains, i.e. a glutamate or aspartate residue, which is observed to hydrogen bond to an arginine residue in the peptide [8]. In SH3 domain of MpSH3, there are a certain number of glutamate or aspartate residue located at  $\beta$ -strands and loops. The prolines of the core motif PXXP are accommodated in conserved hydrophobic grooves on one surface of the SH3 domain. In SH3 domain of MpSH3, there are hydrophobic residues such as proline, tryptophan, valine residues. The surface in the SH3 domain located between RT and N-Src loops displays little conservation across the SH3 domain family. So, to accurately predict the specificities of SH3 domains are complicated.

## Conclusion

The MpSH3 has similar 3D model structure to yeast SH3 domain protein and SH3 domain of human p56-lck. Suggest that the MpSH3 protein may have the function of protein interaction, and may combine with proline-rich sequences.

## Acknowledgements

This work was partially financially supported by the Guangdong Sci. & Tech. Program (2013B010404046).

## References

- [1] P. Reijntj, S. Jorde and J. Wendland: *Curr Genet*, Vol. 56(2010), p. 309-19.
- [2] E. Verschueren, M. Spiess, A. Gkourtsa *et al*: *PLoS One*, Vol. 10(2015), e0129229.
- [3] T. Pawson and J. Schlessinger: *Curr. Biol*, Vol. 3(1993), pp. 434-42.
- [4] B. Mayer: *J. Cell Sci*, Vol. 114(2001), p. 1253-1263.
- [5] L. Castagnoli, A. Costantini, C. Dall'Armi *et al*: *FEBS Lett*, Vol. 567(2004), p. 74-9.
- [6] H. Yu, J.K. Chen, S. Feng *et al*: *Cell*, Vol. 76(1994), p. 933-45.
- [7] A. Musacchio, M. Saraste and M. Wilmanns: *Nat. Struct. Biol*, Vol. 1(1994), p. 546-51.
- [8] W.A. Lim, F.M. Richards and R.O. Fox: *Nature*, Vol. 372(1994), p. 375-79.
- [9] C. Landgraf, S. Panni, L. Montecchi-Palazzi *et al*: *PLoS Biol*, Vol. 2(2004), e14.
- [10] X. Xin, D. Gfeller, J. Cheng *et al*: *Mol Syst Biol*, Vol. 9(2013), p. 652.
- [11] J. Kim, C.D. Lee, A. Rath *et al*: *Journal of Mol. Biol*, Vol. 377(2008), p. 889-901.
- [12] T. Hou, N.Li, Y.Li *et al*: *J Proteome Res*, Vol. 11(2012), p. 2982-95.
- [13] D.A. Liberles, M.D. Tisdell and J.A. Grahnen: *Proc Biol Sci*, Vol. 278(2011), p.1930-35.
- [14] M. Gorelik, K. Stanger and A.R. Davidson: *J Biol Chem*, Vol. 286(2011), p. 19470-77.
- [15] R.R. Bartelt, J. Light, A. Vacaflares *et al*: *Biochim Biophys Acta*, Vol. 1853(2015), p. 2560-69.
- [16] S.M. Larson and A.R. Davidson: *Protein Sci*, Vol. 9(2000), p. 2170-80.
- [17] S. Feng, C. Kasahara, R. Rickles *et al*: *Proc.Natl.Acad.Sci.U.S.A*, Vol. 92(1995), p. 12408-15.