

Newly Identified Motifs within PAS Domains of Filamentous Cyanobacteria

Rei Narikawa

cc37728@mail.ecc.u-tokyo.ac.jp

Masahiko Ikeuchi

mikeuchi@bio.c.u-tokyo.ac.jp

Shinobu Okamoto

okamoto@bio.c.u-tokyo.ac.jp

Masayuki Ohmori

cohmori@mail.ecc.u-tokyo.ac.jp

Department of Life Sciences, University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo

Keywords: cyanobacteria, PAS domain, structure

1 Introduction

Many cyanobacterial genome projects have been finished or are ongoing. Now, complete genome sequence of *Anabaena* sp. PCC 7120 (*Anabaena*)[1] and draft genome sequence of *Nostoc punctiforme* (*Nostoc*)[3] are available. Phylogenetically, these species are closely related. These species have extremely abundant signal-transduction proteins, especially PAS domain-containing proteins (PAS-containing proteins)[5]. PAS domain is known to be multi-functional domain superfamily with highly conserved 3D structure [6]. We formerly detected 142 and 180 PAS domains in 60 and 84 PAS-containing proteins from *Anabaena* and *Nostoc*, respectively. Clustering analysis showed that they were classified into 64 distinct subclasses and 36 “orphans”. On the other hand, we detected 35 pairs of PAS-containing proteins that are orthologous to each other between the two species. PAS domains in these orthologous proteins were classified into two groups: orthologous and non-orthologous ones. We interpreted the non-orthologous PAS domains in the orthologous proteins as generated by domain shuffling after branching of the two related species. Finally, 60 pairs of the orthologous PAS domains were highlighted as candidates for functionality [4]. Here, we extended our trial to unravel the complex PAS superfamily by finding novel motifs, which may interact each other in the 3-dimensional configuration.

2 Methods and Results

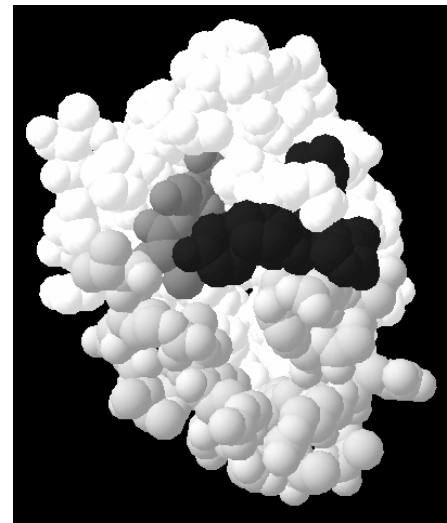
We predicted secondary structure of 60 pairs of the orthologous PAS domains by PSIPRED [2] and constructed multiple sequence alignment. Based on this alignment, we found several conserved Cys and His residues in certain pairs of PAS domains as possible coordination to ligands such as heme or flavin [4]. Here, we detected novel sequence motifs in 11 pairs of the orthologous PAS domains. They are HPDD and E[E,Y]R motifs with an intervening stretch of 19-21 amino acid residues (Figure 1). It is of note that the overall structure of HPDDX_{19–21}E[F,Y]R is shared by 8 distinct subclasses of the PAS superfamily. We also detected the same motifs in many other non-orthologous PAS domains. These motifs were always found together beyond the subclasses. In total, *Anabaena* has 40 PAS domains of these motifs and *Nostoc* has 45 domains. We performed mapping analysis of the motif based on the known 3D structure of other subclasses and found that HPDD and E[E,Y]R were closely located on the surface of the domain, although they were separated by X_{19–21} segment (Figure 2). The 3D structure of PAS domain shows a clam-shaped fold. HPDD and E[F,Y]R appear to interact with each other, leading to closure of a flexible pocket, where ligands such as heme or flavin are coordinated. Moreover, both contain acidic and basic amino acid residues, implying that they may be involved in electrostatic stabilization of the flexible structure or dimerization of the PAS-containing proteins.

Table 1: Number of PAS-containing proteins, PAS domains and PAS domains with HPDD and E[F,Y]R motifs.

	Number of		
	PAS proteins	PAS domains	PAS domains with HPDD and E[F,Y]R motifs
Anabaena	60	142	40
Nostoc	82	180	45
Ortholog	35 pairs	60 pairs	11 pairs

	Subclass Number
A110182_1/1 ALHLEDLERSYETWSKAVETG-ALYEN ^{EYR} FK	
Npun2175_1/1 ALHPDDIQNADQIWSNAVKNS-TTYNN ^{EYR} FK	
A110824_2/4 AVHPDDRGTGEANGIAVANR-SQYQI ^{EYR} IR	
Npun1852_2/5 AVHPDDRATGEVWSAAVANL-SMYQI ^{EYR} IR	#2
A110824_3/4 PIHPDDRTRSF EAWNASVVKR-SIYQI ^{EYR} LR	
Npun1852_3/5 PIHPDDCARSTEVWNAAMTNR-GIYQI ^{EYR} LR	
A111846_1/1 SVHSEDQERAIASWSTYIYQG-GDYSE ^{EYR} FQ	
Npun1355_1/1 SVHPEDRERAVASWSNYICEG-RDYSE ^{EYR} FQ	
A115210_4/4 ITHPDDVEADWVYAKQVLAREINGYSLE ^{EYR} RYL	#15
Npun3209_1/1 ITHPDDLAVDLENARRVLAKEINGYSLE ^{EYR} RYI	
A111145_1/1 FVHNDDVQLFTDYLSTVSNKNNKHQPLE ^{EYR} IK	#14
Npun4641_3/3 FVHPDDVPICADYFQRIATTEKQDAI ^{EYR} VK	
Alr3170_4/4 LIHPEDQQRVSQSIKPLLANGSH--EL ^{EYR} IRL	#32
Npun5680_4/4 VIHPEDQQQVHAAIQPLFLTEKQ--EL ^{EYR} IRV	
A111914_2/4 LIHPEDRENFSQSVKNSIEQGMP-FTV ^{EYR} TV	#33
Npun5547_4/5 LIHPEDQKAVAANIAQIIEEGQ--GS ^{EYR} IRTI	
A112379_3/5 IVHPEDREFLHQ ^S ALRAIEQQVP-HDI ^{EYR} FRFL	#21
Npun6860_2/4 IVHPEDRELLHQ ^S VLRAIDQQVP-HNL ^{EYR} FRFI	
Alr3442_1/3 CVHPEDRESVIQVIGVLENKTD-YYD ^{EYR} FRIV	#22
Npun4379_1/3 CVYPEDKQSVMOGIGQALAQKT-DYD ^{EYR} FRIV	
Alr1966_2/2 RVHPDDIDWVMQAVQDHLAKKTSHYVT ^{EYR} HRVL	#5
Npun4391_2/2 RIHPEDQDWLQAFQDHF ^{EYR} AKKTPFYVCEYRVO	

Figure 1: Multiple alignment of HPDD-E[F,Y]R motif. Subclass was defined in ref 4.

Figure 2: Mapping analysis. Black: HPDD motif, Gray: E[F,Y]R motif, Pale gray: intervening X₁₉₋₂₁ stretchend.

References

- [1] Kaneko, T., *et al.*, Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120, *DNA Res.*, 8:205–213; 227–253, 2001.
- [2] McGuffin, L.J., Bryson, K., and Jones, D.T., The PSIPRED protein structure prediction server, *Bioinformatics*, 16:404–405, 2000.
- [3] Meeks, J., *et al.*, An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium, *Photosyn Res.*, 70:85–106, 2001.
- [4] Narikawa, R., *et al.*, Domain shuffling occurs frequently in PAS domain-containing proteins of filamentous cyanobacteria, *Abstract of XIth International Symposium on Phototrophic Prokaryotes*, 67, 2003.
- [5] Ohmori, M., *et al.*, Characterization of genes encoding multi-domain proteins in the genome of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120, *DNA Res.*, 8:271–284, 2001.
- [6] Taylor, B.L., Zhulin, I.B., and Johnson, M.S., Aerotaxis and other energy-sensing behavior in bacteria, *Annu. Rev. Microbiol.*, 53:103–128, 1999.