Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Bioinformatics. Author manuscript; available in PMC 2016 October 17.

Published in final edited form as:

Bioinformatics. 2007 October 1; 23(19): 2518–2521. doi:10.1093/bioinformatics/btm384.

FIST: a sensory domain for diverse signal transduction pathways in prokaryotes and ubiquitin signaling in eukaryotes

Kirill Borziak and **Igor B. Zhulin**

Joint Institute for Computational Sciences, The University of Tennessee – Oak Ridge National Laboratory, Oak Ridge, TN 37831-6173, USA

Igor B. Zhulin: joulineib@ornl.gov

Abstract

Motivation—Sensory domains that are conserved among Bacteria, Archaea and Eucarya are important detectors of common signals detected by living cells. Due to their high sequence divergence, sensory domains are difficult to identify. We systematically look for novel sensory domains using sensitive profile-based searches initiated with regions of signal transduction proteins where no known domains can be identified by current domain models.

Results—Using profile searches followed by multiple sequence alignment, structure prediction and domain architecture analysis, we have identified a novel sensory domain termed FIST, which is present in signal transduction proteins from Bacteria, Archaea and Eucarya. Chromosomal proximity of FIST-encoding genes to those coding for proteins involved in amino acid metabolism and transport suggest that FIST domains bind small ligands, such as amino acids.

1 INTRODUCTION

Signal transduction systems are information processing pathways that link environmental cues to adaptive responses in all living organisms. Major prokaryotic and eukaryotic signal transduction systems are different at the level of their principal components and complexity. Prokaryotic signal transduction pathways consist of simple one- and two-component systems (Ulrich *et al.*, 2005), whereas signal transduction in eukaryotes often involves multi-protein, branched cascades (Citri and Yarden, 2006). However, all living cells react to many similar signals that are present in the environment and inside cells (e.g. small molecules, such as amino acids and carbohydrates), therefore, some level of similarity is expected in their signal input elements. Indeed, several universal input domains have been described in sensory receptors from both prokaryotes and eukaryotes. The most abundant such domain is PAS (named after eukaryotic Per, ARNT and SIM proteins, where it was first described), which is found in diverse signal transduction pathways in organisms ranging from all major prokaryotic clades to humans (Ponting and Aravind, 1997; Zhulin *et al.*, 1997). PAS domains serve as detectors of small molecules, redox potential, oxygen, light and other important parameters (Taylor and Zhulin, 1999). Other input, sensory elements that are

Correspondence to: Igor B. Zhulin, joulineib@ornl.gov.

Conflict of Interest: none declared.

Supplementary information: http://genomics.ornl.gov/research/pubs/fist

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

found in both branches of life include small-ligand binding GAF (Aravind and Ponting, 1997), Cache (Anantharanam and Aravind, 2000) and CHASE (Anantharanam and Aravind, 2001; Mougel and Zhulin, 2001) domains. On the other hand, many other input domains are limited to bacterial signal transduction, e.g. MHYT (Galperin *et al.*, 2001), NIT (Shu *et al.*, 2003), CHASE2 through CHASE6 (Zhulin *et al.*, 2003) and 4HB_MCP (Ulrich and Zhulin, 2005), etc. Identification of novel sensory domains is a difficult task due to their extreme sequence variation, both in composition and in length. Sensitive similarity search methods, such as Position-Specific-Iterative (PSI) BLAST (Altschul *et al.*, 1997) must be employed to detect relationships between the domain family members, which must be further verified through a careful analysis of a multiple alignment of the domain family.

Recently developed MiST (Microbial Signal Transduction) database (Ulrich and Zhulin, 2007) enables rapid identification of sensory receptors where current domain models implemented in leading domain databases Pfam (Finn *et al.*, 2006) and SMART (Letunic *et al.*, 2006) do not detect any input domains. We carry out systematic similarity searches using protein regions of receptor proteins implicated in signal transduction that contain no identifiable domain. In many cases, such regions do contain known domains; however, current domain models are not sensitive enough to detect them. In other cases, such regions contain potentially new domain. In this report, we describe the identification of such novel domain, which we termed FIST, which is implicated in sensory reception in diverse signal transduction pathways in all three domains of life: Bacteria, Archaea and Eucarya.

2 DOMAIN IDENTIFICATION

During systematic analysis of microbial sensory proteins, we focused on a methyl-accepting chemotaxis protein (MCP) blr4191 from *Bradyrhizobium japonicum* (gi27352453), which was predicted to be a cytoplasmic chemotaxis receptor. All known chemotaxis receptors contain a sensory domain in their N-terminus followed by a conserved C-terminal signaling domain (Alexander and Zhulin, 2007), which can be identified by a Pfam domain model MCPsignal (accession PF00015). The blr4191 protein contained the full-length C-terminal MCPsignal domain, but lacked any detectable domain in its large (more than 400 amino acid residues) N-terminus. Exhaustive PSI-BLAST searches (*E*-value cutoff 0.01, composition-biased statistics on, filter on) against the NCBI nr database (1 March 2006) were initiated with residues 1–449 that include the entire N-terminal region up to the first residue of the MCPsignal domain and continued until no new homologs were identified. Duplicate sequences were excluded from analysis. Profile searches yielded the list of domain family, which we termed FIST (F-box and intracellular signal transduction proteins), containing 176 proteins. Of these, 155 were full length matches, and 15 matched the C-terminal half of the domain (FIST_C) only. This search also revealed an overlap of the newly proposed domain with the Pfam domain of unknown function DUF1745 (also found in the COG database as COG3287, ‘uncharacterized conserved protein’), which was detected in more than a half of proteins identified by PSI-BLAST searches.

Sequences that matched only the FIST_C subdomain all belong to F-box-containing eukaryotic proteins. When the region of these proteins that did not match the N-terminal half of the FIST domain (FIST_N) was subjected to a PSI-BLAST search, the only profile

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

matches returned were from the same subfamily of F-box proteins. However, searches initiated with FIST_C from these proteins returned most of full-length FIST domain proteins. Based on results of PSI-BLAST searches and predicted secondary structure, we built domain models from multiple alignments for FIST_N and FIST_C, as well as for the entire FIST domain, and carried out searches using the HMMER program to retrieve all homologs. Resulting sets contained 253 proteins with full-length FIST, 4 proteins with FIST_N only and 30 proteins with FIST_C only (nr database, 1 May 2007). Representative and complete multiple alignments of subdomains are provided as Figure 1 and Supplementary Material, correspondingly.

3 DOMAIN FEATURES AND ARCHITECTURE

Multiple alignment of full-length FIST domain sequences was built using ClustalX (Thompson *et al.*, 1997) with default parameters and edited with the VISSA technique (Ulrich and Zhulin, 2005), which utilizes PSIPRED (McGuffin *et al.*, 2000) to guide the editing process using predicted secondary structure. The final fold profile consists of 20 beta strands and 7 alpha helices. Multiple alignment of the FIST domain family revealed several highly conserved residues. The most conspicuous motif of the FIST domain family, which contains highly conserved cysteine and arginine, is found in the beta-strand 19 and the following loop (Fig. 1B). It may represent a site for ligand binding or protein–protein interactions. Using SMART (Letunic *et al.*, 2006) and Pfam (Finn *et al.*, 2006) domain architecture tools, we determined that in more than 70% of analyzed sequences the FIST domain comprises a single-domain protein; in the remaining sequences it is present in association with well-known signal transduction domains (Fig. 2) including the sensory PAS domain, GGDEF and EAL domains involved in the turnover of cyclic di-GMP, the MCP signaling domain, HisKA and HATPase_c domains of sensor histidine kinases, the DNA-binding helix-turn-helix TrmB domain and F-box domains of the ubiquitin signaling pathway. The search for transmembrane regions and signal peptides using Phobius (Kall *et al.*, 2004) showed that most FIST-containing proteins analyzed in this study are predicted to be intracellular; however, signal peptides were predicted for a few stand-alone FIST-domain proteins indicating that they are likely to be extracellular.

4 BIOLOGICAL FUNCTION

FIST represents a new sensory (input) domain in signal transduction. As many other sensory domains, it is found either as a single-domain protein or exclusively in a combination with other domains that transmit signals from the sensory domain down the regulatory pathway, namely transducers (MCPsignal, HATPase_c domains) and output domains (GGDEF, EAL, TrmB). The FIST domain was found in four major classes of prokaryotic signal transduction: methyl-accepting chemotaxis proteins, sensor histidine kinases, di-guanylate cyclases and diesterases, and transcriptional regulators.

Further evidence suggesting FIST involvement in signal transduction comes from the analysis of the genome context (Overbeek *et al.*, 1999) of FIST-encoding genes. Using the MIST database (Ulrich and Zhulin, 2007), we have determined that in more than 30% cases, adjacent (immediately upstream or downstream) to the FIST-encoding gene there is a gene

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

coding for a known signal transduction protein. Overall, signal transduction genes in 198 prokaryotic genomes that contain FIST domains comprise <7% of the total number of genes (http://genomics.ornl.gov/mist). Thus there is a significant enrichment in signal transduction genes next to FIST-encoding genes. As a sensory domain, FIST is predicted to bind small molecule ligands. In eukaryotes, FIST is found in F-box proteins that are involved in ubiquitin mediated degradation of regulatory proteins, a frequent means of controlling progression through signaling pathways (Hershko and Ciechanover, 1998). Interestingly, genes encoding FIST-containing proteins are also often found next to genes coding for enzymes involved in amino acid metabolism (peptidases, aspartamine synthase) and amino acid transporters. In plants, F-box proteins are known to bind a plant hormone auxin, which is a derivative of tryptophan (Parry and Estelle, 2006). This further suggests that small molecule ligands detected by FIST might be amino acids and their derivatives. Detecting amino acid concentration both inside and outside the cell is important for many physiological processes, and it is likely that FIST-containing sensors play a significant role in converting these signals into changes in transcription, metabolism, development and behavior.

FIST domains are found in 16 phyla (Supplementary Table 1) representing all three kingdoms of life. This broad phyletic distribution suggests the ancient origin of this domain and further underscores its importance as a ubiquitous sensor module in diverse signal transduction pathways.

Acknowledgments

We thank Luke Ulrich for expert technical assistance and two anonymous reviewers for excellent suggestions. This work was supported in part by a grant GM072285 from the National Institutes of Health to I.B.Z.

References

Alexander RP, Zhulin IB. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. Proc Natl Acad Sci USA. 2007; 104:2885–2890. [PubMed: 17299051]

Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]

Anantharaman V, Aravind L. Cache – a signaling domain common to animal Ca(2+)-channel subunits and a class of prokaryotic chemotaxis receptors. Trends Biochem Sci. 2000; 25:535–537. [PubMed: 11084361]

Anantharaman V, Aravind L. The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. Trends Biochem Sci. 2001; 26:579–582. [PubMed: 11590000]

Aravind L, Ponting CP. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci. 1997; 22:458–459. [PubMed: 9433123]

Citri A, Yarden Y. EGF-ERBB signalling: towards the systems level. Nat Rev Mol Cell Biol. 2006; 7:505–516. [PubMed: 16829981]

Finn RD, et al. Pfam: clans, web tools and services. Nucleic Acids Res. 2006; 34:D247–D251. [PubMed: 16381856]

Galperin MY, et al. MHYT, a new integral membrane sensor domain. FEMS Microbiol Lett. 2001; 205:17–23. [PubMed: 11728710]

Hershko A, Ciechanover A. The ubiquitin system. Annu Rev Bbiochem. 1998; 67:425–479.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Kall L, et al. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004; 338:1027–1036. [PubMed: 15111065]

Letunic I, et al. SMART 5:domains in the context of genomes and networks. Nucleic Acids Res. 2006; 34:D257–D260. [PubMed: 16381859]

McGuffin LJ, et al. The PSIPRED protein structure prediction server. Bioinformatics. 2000; 16:404–405. [PubMed: 10869041]

Mougel C, Zhulin IB. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends Biochem Sci. 2001; 26:582–584. [PubMed: 11590001]

Overbeek R, et al. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA. 1999; 96:2896–2901. [PubMed: 10077608]

Parry G, Estelle M. Auxin receptors: a new role for F-box proteins. Curr Opin Cell Biol. 2006; 18:152–156. [PubMed: 16488128]

Ponting CP, Aravind L. PAS: a multifunctional domain family comes to light. Curr Biol. 1997; 7:R674–R677. [PubMed: 9382818]

Shu CJ, et al. The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors. Trends Biochem Sci. 2003; 28:121–124. [PubMed: 12633990]

Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev. 1999; 63:479–506. [PubMed: 10357859]

Thompson JD, et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997; 25:4876–4882. [PubMed: 9396791]

Ulrich LE, et al. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. 2005; 13:52–56. [PubMed: 15680762]

Ulrich LE, Zhulin IB. Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. Bioinformatics. 2005; 21(Suppl 3):iii45–iii48. [PubMed: 16306392]

Ulrich LE, Zhulin IB. MiST: a microbial signal transduction database. Nucleic Acids Res. 2007; 35:D386–D390. [PubMed: 17135192]

Zhulin IB, et al. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. Trends Biochem Sci. 1997; 22:331–333. [PubMed: 9301332]

Zhulin IB, et al. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. J Bacteriol. 2003; 185:285–294. [PubMed: 12486065]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

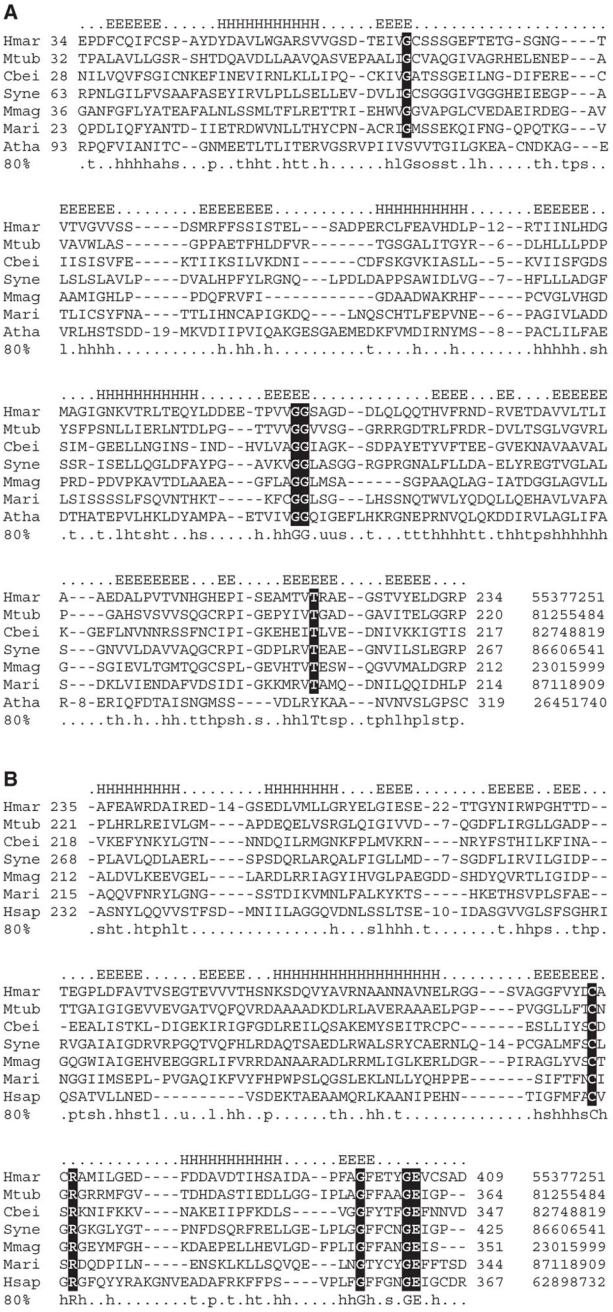

Fig. 1.

Representative multiple alignments of the FIST_N (**A**) and FIST_C (**B**) subdomains. Representatives of Archaea, Bacteria and Eucarya are included. Predicted secondary structure is shown above the alignment and consensus (80%) calculated for all members of the subdomain family is shown below. GI identification numbers are shown at the end of each sequence. Complete alignments are provided as Supplementary Material. Highly conserved residues are highlighted. Species abbreviations: Atha, *Arabidopsis thaliana*; Cbei, *Clostridium beijerincki*; Hmar, *Haloarcula marismortui*; Hsap, *Homo sapiens*; Mari,

Marinomonas species; Mmag, *Magnetospirillum magnetotacticum*; Mtub, *Mycobacterium tuberculosis* and Syne, *Synechococcus* species.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

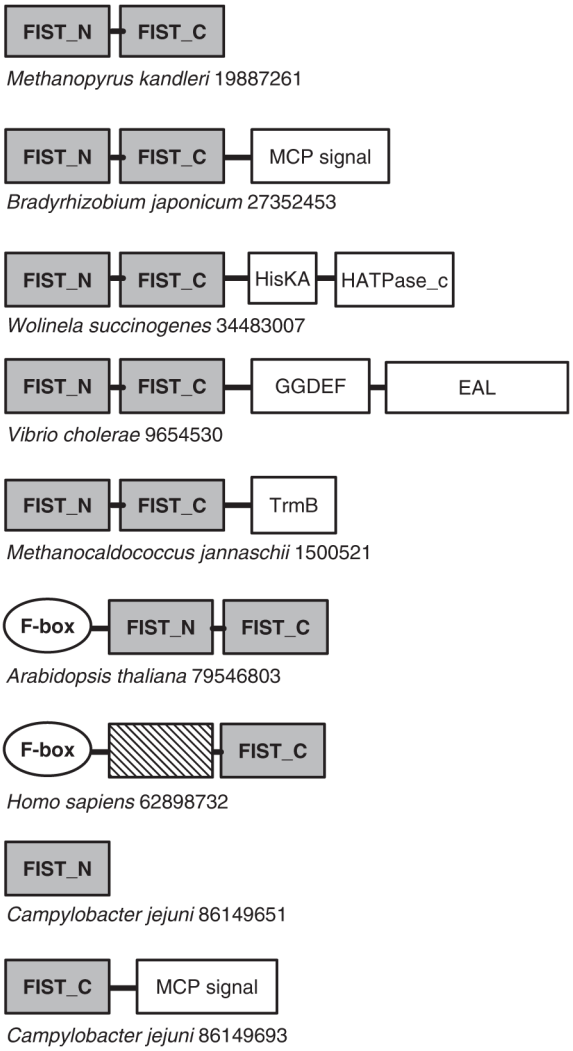

Fig. 2.

omain architecture of representative proteins containing the FIST domain. Species names and GenBank identification numbers are shown. Domain designations (Pfam accession numbers in parentheses): MCPsignal (PF00015), MCP C-terminal signaling domain; GGDEF (PF00990), di-guanylate cyclase; EAL (PF00563), di-guanylate esterase; HisKA (PF00512), histidine kinase dimerization domain; HATPase_c (PF02518), histidine kinase domain ATP-binding domain; TrmB (PF01978), helix-turn-helix DNA-binding domain; F-box (PF00646), protein-protein interaction domain and PAS (PF00989), the PAS sensory domain.