Journal of Proteomics & Bioinformatics - Open Access

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jpb.1000148
JPB/Vol.3 Issue 8

In silico Discovery of Novel Protein Domains in *Streptococcus mutans*

Anand Ramesh Varsale[1], Amol Shriram Wadnerkar[1], Rajendra Haribhau Mandage[1]* and Priyanka Khasherao Jadhavrao[2]

[1]Dept of Bioinformatics, Centre for Advanced Life Sciences, Deogiri College, Aurangabad 431005, M.S. India
[2]Dept of Microbiology, Abeda Inamdar College, Pune 411001, M.S. India

Abstract

Streptococcus mutans is the principal etiological agent of dental caries worldwide and is considered to be the most cariogenic of all of the oral streptococci. We sought to expand our understanding of this organism at the molecular level through identification and function prediction of novel protein domains by two automated approaches using HMM and BLAST to develop a complete set of domains, which were then subsequently manually analysed. A final set of 19 novel domains and families was identified. This enhances our understanding of both *S mutans* and also general bacterial molecular mechanisms, including protein synthesis to catalytic action. Furthermore we demonstrate that using this type of *in silico* method it is possible to fairly rapid generation of new biological information from previously uncorrelated data to increase the rate of discoveries in the laboratory.

Keywords: Oral cariogen; Family; Domain hunting; All-against-all BLAST; HMM; Prosite

Introduction

Streptococcus mutans is Gram-positive, nonmotile, facultatively anaerobic, oral pathogen and is considered to be the most cariogenic of all of the oral streptococci (Loesche et al., 1986). *S. mutans* adheres to the tooth surface and produces a sticky polysaccharide called dextran that enables the further colonisation of other microorganisms, forming dental plaques which serve as a biofilm (Nyvad et al., 1990). These microorganisms have to withstand changes in temperature, nutrition and osmotic pressure, pH variations (Carlsson et al., 1997) as well as exposure to the mucosal immune system, natural virulence factors, antibiotics ,competitors as well as pathogenity (Kado et al., 2009; Ochman et al., 2000). Lateral gene transfer (LGT) or horizontal gene transfer (HGT) (Lawrence, 1997; Ochman et al., 2005) is a major way by which organisms acquires novel genes, and it has played an important role in how *S. mutans* has adapted to sustain the oral environment through resource acquisition, defense against host factors, and use of gene products that maintain its niche against microbial competitors. Thus LGT creates unusually high similarities among organisms, particularly those that are closely related or share the same habitat; it can also be helpful for the understanding of gene evolution and species diversification, but also for the development of drugs that inhibit the transfer of resistance (Ochman et al., 2000). *S. mutans* genome is composed of the 1963 open reading frames out of which 65% were initially assigned putative functions through developed Bioinformatics methodologies and more ORFs, or their orthologs, have been identified by microarray techniques, phenotype studies and so on (Niu et al., 2008; Deng et al., 2009; Ajdic and Pham, 2007). There are very small numbers of related proteins in the modern databases and these homologs are all hypothetical proteins having no function. The plethora of proteins reflects both a diversity of novel protein families and an expansion within identified families when compared to other organisms and that serves as good platform in the search for novel protein domains.These methodologies clearly showed how completely sequenced genomes can be exploited to further understand the biology of an organism by predicting relationships between molecular structure, function, and evolution.

Protein domains

After protein discovery, there are many questions that are associated with protein's overall identity, putative function and biologically significant sites identification. To answer these questions, a number of databases and tools have been customised. Structural and functional features of proteins are determined by using different methods to exploit characteristic sequence patterns and amino acid frequency and other properties (Bateman and Birney, 2000). Sometimes newly discovered protein sequence may lack considerable identity with known sequences of known functions over their entire length, which makes functional prediction difficult. So important aspect of protein sequence characterisation is identification of domains and families based on primary sequence data (Gribskov et al., 1996). These functional domains and family databases are useful to deal with the question, "What types of functional domains are present within this sequence? Or what family does this protein belong to?" Even though, some family and domain databases were developed for genomic sequences annotation purpose, these tools are better option to characterise proteins with unknown function (Bateman and Birney, 2000). A protein domain is a discrete portion of a protein sequence and structure that can evolve, function, and assume to fold independently of the rest of the protein and possessing its own function. The independent evolutionary theories of domains found within the same protein lead to the hypothesis that the domain is the basic unit of protein structure and function (Doolittle, 1995). The direct functional and structural determination of all the proteins in an organism is prohibitively costly and time consuming because of the relative scarcity of 3D structural information therefore primary sequence analysis is preferred to identify majority of protein domain families (Sonnhammer and Kahn, 1994). Large number of protein domain families recognised from sequence has been escalating progressively over the years which have led to the development of online domain and families databases such as SMART (Schultz et al.,

***Corresponding author:** Rajendra Haribhau Mandage, Dept of Bioinformatics, Centre for Advanced Life Sciences, Deogiri college, station road, Aurangabad, 431005, M.S. India. Tel: +91 9561585950; E-mail: rajendra.mandage@gmail.com

Received August 07, 2010; **Accepted** August 29, 2010; **Published** August 29, 2010

Citation: Varsale AR, Wadnerkar AS, Mandage RH, Jadhavrao PK (2010) Cheminformatics. J Proteomics Bioinform 3: 253-259. doi:10.4172/jpb.1000148

1998) and Pfam (Schultz et al., 1998; Bateman and Birney, 2000). Organism's metabolic potential as well as other molecular systems can be explored using sophisticated genomic tools that have been directed toward understanding the key function of a particular protein in a fundamental biological process using a primary sequence only. A classier means of analysing proteins is through the detection of their domain structures which is distinct steady amino acids piece of sequence, typically ranges between 40 and 400 residues. A numbers of evolutionarily related proteins may contain same domain as structural or functional unit (Bateman and Birney, 2000; Galperin and Koonin, 1998). Here we try to present some of available tools and techniques to detect possible functional domains and novel family of a protein sequences in *Streptococcus mutans*.

The domain hunting approach

One of the methods to correctly predict novel domains is through inspection of high resolution protein 3D structures however structural databases contain limited numbers of sequences that have representative structures (Amer et al., 2008). Novel domains are usually detected by employing automated methods to rapidly generate an optimised set of targets, which were subsequently analysed manually. At one extreme a researcher will start a solo protein sequence and hunt for partial matches against other sequences. These short matches can serves as template to build new families. At the other extreme fully automated methods that work on large protein sets to detect novel families are available (Yeats et al., 2003). We explored these two approaches to investigate the *S. mutans* genome by means of amalgamation of rapid automatic detection of potential novel domains followed by careful manual analyses to assist elucidation putative biological mechanisms and thorough understanding of described systems within the oral dwelling prokaryote *Streptococcus mutans*. Firstly a set of novel domains is detected using the recently completed genome sequence and explanatory information was obtained through literature searching and other analytical tools. These predictions were then viewed within the framework of the *Streptococcus mutans*. These outcomes provide functions for many proteins leading to a number of testable hypotheses.

Materials and Methods

Protein sequences retrieval

Uniprot database (Copleya et al., 2009) was used to retrieve all sequences provided in domain hunting approach. SMART (Letunic et al., 2004), Prosite (Falquet et al., 2002), Pfam (Finn et al., 2006) and InterPro (Apweiler et al., 2001) were used to identify novel domain on the basis of sequence similarity.

Approach I

A set of 3000 potential and known protein sequences from *Streptococcus mutans* was used as the preliminary data. An initial alignment generated by CLUSTAL W (Thompson et al., 1994) was used to create profile-HMMs using the HMMER3 tool (Eddy, 2001).The resultant profile-HMMs were searched against the Uniprot protein resource. A threshold of 0.01 was selected to detect homologs and this alignment was built by means of the hmmalign tool from the HMMER package. This alignment was then queried against the Pfam and Prosite database to identify any similarities with the known domain and families. The last step is a manual examination of the domain to widen its relationship as well as to develop better multiple sequence alignment and with anticipation of the domain function prediction. This analysis uses a wide variety of tools and methodologies.

Approach II

A complementary approach was also tried here for detection of novel domains that may be of significance to the biology of *S. mutans*. All-against-all BLAST (Altschul et al., 1990) was done by means of single-linkage clustering methodology. The proteins were clustered with a cutoff threshold of 50 bits, which helped to avoid clustering of unrelated proteins. Single proteins and all other clusters that corresponded to Pfam database were then removed from the primary dataset. T-Coffee (Notredame et al., 2000) was used to align the clustered sequences. The aligned sequences (clusters) were subsequently used as template for an iteration using HMMER 3, same as in approach I. The sequences were iterated until convergence. Afterward they were again realigned with T-Coffee and a single round of iteration was done. Then the iterative search process was repeated until new family members were identified.

Predictions of function

On the basis of information in the literature and/or co-occurrence with formerly well-known domains, some functional characteristics can be predicted for newly discovered domain and families. The predicted functions such as protein synthesis to drug resistance represent a range of cellular and molecular functions.

Pfam/Prosite Acc No	Family/Domain Name	Pfam Type	Function
PF00472	RF-1	Domain	peptidyl-tRNA hydrolase activity
PF00702	Hydrolase	Family	catalytic activity
PF01368	DHH	Family	phosphoesterase function
PF03462	PCRF	Domain	protein synthesis
PF04327	DUF464	Family	unknown function
PF00480	ROK	family	unknown function
PF00005	ABC transporter	family	Translocation of compounds across membranes.
PF00013	KH	domain	RNA binding
PF00293	NUDIX	domain	removing an oxidatively damaged form of guanine
PF00308	Bacterial dnaA	family	initiating and regulating chromosomal replication
PF00344	eubacterial secY	family	protein transport
PF00391	PEP-utilising enzyme	motif	transferase activity
PF00467	KOW	motif	rRNA tertiary structure
PF00595	PDZ	domain	targeting signalling molecules to sub-membranous sites
PF00627	UBA/TS-N	domain	ubiquitination pathway
PS51353	ArsC	Family	converts arsenate to arsenite
PS01125	ROK	family	transcriptional repressors
PF00254	FKBP-type peptidyl-prolyl cis-trans isomerase	domain	peptidyl-prolyl cis-trans isomerase activity
PS50847	LPxTG	motif	catalytic action

Table: 1 List of all domains identified by Approach I and II, as well as their probable function.

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jpb.1000148
JPB/Vol.3 Issue 8

Results and Discussion

From an initial set of 150 potential domain targets, 25 targets were removed by the step single-linkage clustering methodology that lay within Pfam families, Prosite domain database and most related to the same set of overlapping families. A final set of 19 targets were discovered as novel domains to *S. mutans*. Table 1 lists and briefly describes all novel domains identified in the domain hunting approaches.

Description of some significant domains

LPxTG motif (PS50847): LPXTG motif in a protein serves as a platform for the catalytic action of proteolytic enzyme sortase, resulting in a transpeptidation reaction. The targeted bond between threonine and glycine is cleaved by the enzyme exposing the carboxy terminal of threonine residue that in turn binds to amino terminus of Pentaglycine Bridge in the peptidoglycan, causing crosslinking through covalent interactions. The hydrophobic LPXTG motif is present as conserved sequence in a 35-residue sorting signal along with a tail of positively charged residues (Mazmanian et al., 1999). Figure 1 reveals the structure of rep sorting signal comprising a hydrophobic LPXTG motif and its positively charged residual tail.

Such motif is generally found in the surface proteins of gram positive cocci, possessing N-terminal signal peptide and a C-terminal sorting signal, the specific substrate for sortase, resulting in cleavage of LPXTG motif and attachment of the protein to the peptidoglycan as a consequence of transpeptidation reaction (Marraffini et al., 2006; Navarre and Schneewind, 1999). This particular activity of sortase enzyme and encoding and accessibility of such motif by pathogens is crucial for the establishment of an infectious disease. For example, *S. aureus* anchors on the host cell by the transpeptidation reaction processed by sortase enzyme. According to a recent study, mutations in sortase A and sortase B genes of *S. aureus* resulted in abortive infections due to failure in cell wall anchorage and projection of surface proteins on cells bearing sorting signal with LPXTG or NPQTN motifs (Mazmanian et al., 2001).

Structural analysis of LPXTG family proteins disclosed their modular architecture (Figure 2) and their evolutionary mode as the acquisition of distinct domain sized polypeptides may be evolved through duplication and homologous recombination. Such domains are also explored from various other species such as B repeats of sasA and sasG, SD and SX repeats of Sdr proteins, the conserved 212 residual domain of sasG and signal peptide consisting (Y/F)SIRK motif, as the evidence of horizontal transfer (Fiona et al., 2003). Proteins from LPXTG family show presence of N-terminal secretory signal sequence as a peculiar feature. These were found using the SIGNALP prediction algorithm. These when aligned with the signal sequence of S. aureus, resulted in identification of 15 sequences containing (Y/F) SIRK motif as a conserved sequence. Local alignment tools affirmed this motif as common in Sortase substrates of gram positive cocci (Tettelin et al. 2001).

ArsC family (PS51353): Detoxification of arsenate, arsenite and antimonite is observed as a chromosomal encoded resistance mechanism in many bacterial species (Carlin et al., 1995). The resistance is through efflux mechanism. Reduced Glutathione (GSH) acts as a cofactor for ArsC (~150-residue), an arsenate reductase in the conversion of arsenate to arsenite. Redox active cysteine is an active site conserved amino acid in ArsC (Figure 3) (Liu et al., 1997). Arsenate reductase and low molecular weight bovine protein tyrosine

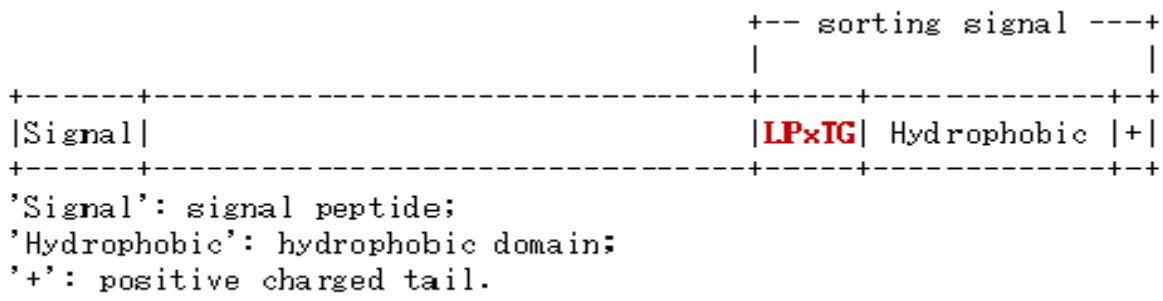

Figure 1: Schematic representation of surface proteins sorting signal composed of a conserved LPxTG motif, a hydrophobic domain, and a tail of positively charged residues.

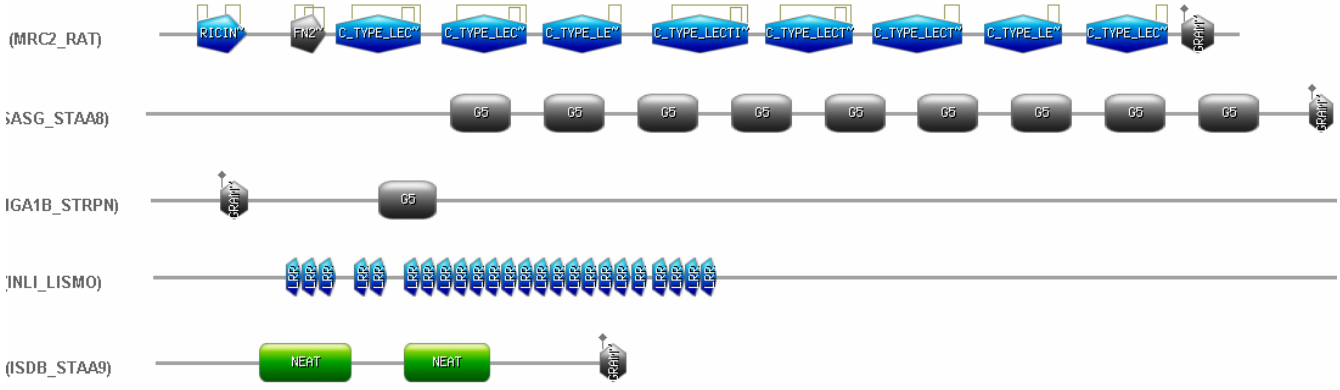

Figure 2: Domain architecture of LPxTG motif is shown in blue color.

Figure 3: Sequence logo of family proteins containing ARSC domain. Sequences from *O. tritici* were aligned with ArsC homologues from *E. coli* pR773 (AAA21096) and *Staphylococcus aureus* pI258 (AAA25638).

Figure 4: Multiple sequence alignment of gram-positive bacterial arsenate reductases, bacterial PTPase homologues, and mammalian LMW PTPases by CLUSTAL X. The protein sequences are obtained from the SWISS-PROT database.

phosphatase show significant structural similarity in spite of the low sequence identity. Similarity is significantly high in their active sites. In vitro analysis affirmed this structural homology functionally relevant by displaying phosphatase activity by arsenate reductase (Figure 4).

The arsC family proteins are also expressed in gram positive bacteria such as Spx proteins that act as transcription factors, regulating transcription of multiple genes under disulfide stress (Zuber, 2004).The structure of ArsC protein is found to be belonging to the thioredoxin superfamily fold characterized by α-helices wrapped around a β-sheet core. The loop between the first β-strand and the first helix encloses the active site cysteine residue. Such structure is found to be conserved in Spx proteins and other homologs (Martin et al., 2001). This suggests the horizontal transfer of this conserved domain.

ROK family (PS01125): It is a family of bacterial proteins which groups transcriptional repressors, uncharacterised ORFs and sugar kinases, for this reason known as ROK (Repressor, ORF, Kinase). At present, consist of Xylose operon repressor (gene xylR) in *Bacillus subtilis, Lactobacillus pentosus* and *Staphylococcus xylosus*, N-acetylglucosamine repressor (gene nagC) from *Escherichia coli* and Glucokinase (gene glk) from *Streptomyces coelicolor*.

The repressor proteins (xylR and nagC) from this family possess an N-terminal region contains a helix-turn-helix DNA-binding motif. The domain common to all these proteins consists of about 300 residues (Titgemeyer et al., 1994). Sequence logo demonstrates conservation of glycine residues in many positions (Figure 5).The presence of ROK (Repressor, ORF, Kinase) domain in the wide varieties from bacteria to humans designates its conservation (Figure 6).

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jpb.1000148
JPB/Vol.3 Issue 8

DUF464 Family (PF04327): This family is an interesting case, and has been previously mentioned as a family of uncharacterised archaeal proteins with 38 sequences in pfam (Shin et al., 2005). A protein BLAST search (Altschul et al., 1997) was performed on the NCBI site using the established default parameters in order to search whether selected DUF464 family member protein have sequence similarity with other proteins. Blast results showed 27% identity with Ribosomal protein of *Leptotrichia hofstadii* F0254. In Gene ontology it showed molecular function as ribonucleoprotein and cellular component as ribosome which is inferred from electronic annotation. This relevance with ribosome suggests that it may be used as a potential target for antibiotics in order to resist cariogenic pathogens. GOR server located at Expasy server was used to predict secondary structure of DUF464 Family member protein. It showed unusual number of extended coils and and random coil (Figure 7).

FKBP-type peptidyl-prolyl cis-trans isomerase (PF00254): FKBP (Tropschug et al., 1990) is a domain that typically occurs in the major high-affinity binding protein mostly found in vertebrates, for the immunosuppressive drug FK506. FKBP12 is notable in humans for binding the immunosuppressant molecule tacrolimus (originally designated FK506), which is used in treating patients after organ transplant and patients suffering from autoimmune disorders (Wang et al.,1994). Both the FKBP-tacrolimus complex and the ciclosporin-cyclophilin complex inhibit a phosphatase called calcineurin, thus blocking signal transduction in the T-lymphocyte transduction pathway. Slow protein-folding reactions are accelerated by a prolyl cis/trans isomerase isolated from porcine kidney which is identical to cyclophilin, a protein that is probably the cellular receptor for the immunosuppressant cyclosporin A. It exhibits peptidyl-prolyl cis-trans isomerase activity (PPIase or rotamase) (Stein, 1991). An

Figure 5: Sequence logo from multiple sequence alignment of ROK family proteins.

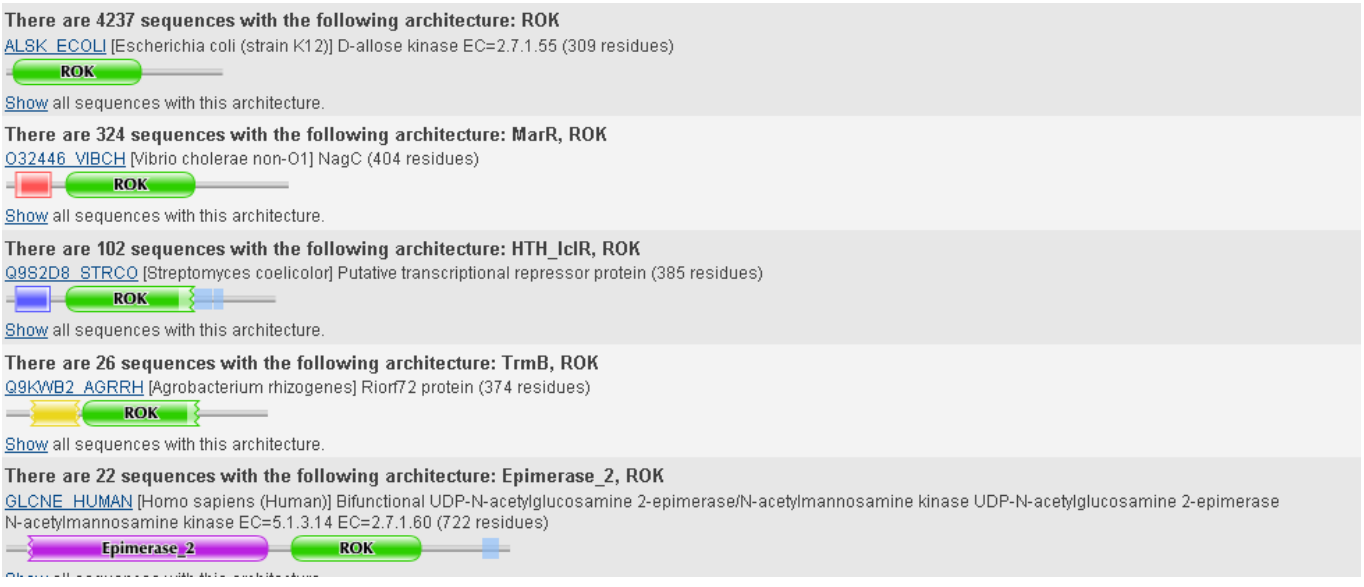

Figure 6: Domain architecture of proteins containing the ROK domain in various species.

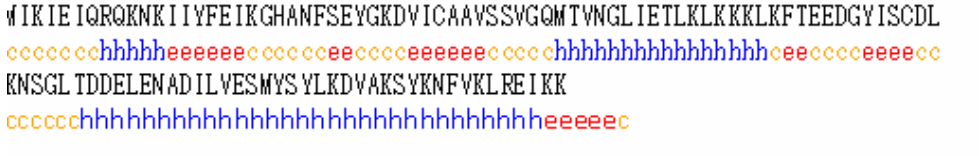

Figure 7: Secondary structure prediction of DUF464 Family member protein using GOR server.

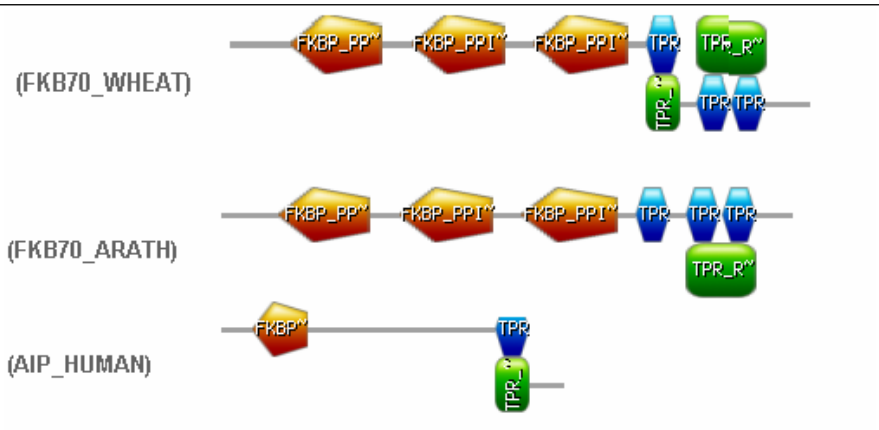

Figure 8: Domain architectures examples of the FKBP-type peptidyl-prolyl cis-trans isomerase -containing proteins.

FKBP was found to act as a modulator of an intracellular calcium release channel along with the cyclophilins. Presence of such immunosuppressive binder motif in *S. mutans* makes a proposition that it actually helps the pathogen in dealing with the therapeutic agents used against it. The FKBP domain is found in discrete varieties of organisms from prokaryotes to eukaryotes, showing its significant conservation (Figure 8).

Conclusion

Comparative genomics is still a growing field and it is hoped that through these methods we can get better categorisation of domains and families of protein sequences toward an understanding of the biology of *S. mutans*. Manual investigation of every single protein is an incalculably time-consuming activity therefore it would not be feasible to annotate protein families. We presented here a combination domain hunting approach in order to concentrate on potentially the most interesting domain families. Our approach discovered common domains e.g., the ROK domain that is observed in a wide variety of species. Majority of domains identified indicate that they have essential biological activities; they are, on average, present in smaller number of proteins than previously described domains. FKBP-type peptidyl-prolyl cis-trans isomerase domain was also found in *S. mutans* that is also present in various other pathogens is immunosuppressive in nature. So a hypothesis can be made that this domain allows development of *S. mutans* as a biofilm over tooth surfaces by suppressing mucosal immune barrier. The domains that are found in the study suggest that the highly conserved domains are not acquired through the lateral gene transfer but they are of ancient origin. Such investigations are helpful for phylogenetic analysis that will lead to demonstrate the single origin of functional domains and also be a significant contributor in the evolutionary aspects of life. These discoveries provide a basis for future drug development and new approach in prevention and treatment of dental caries.

Reference

1. Ajdic D, Pham VT (2007) Global transcriptional analysis of Streptococcus mutans sugar transporters using microarrays. J Bacteriol 199: 5049-5059.
2. Al-Badarneh AF, Khalil MA (2008) Improving Protein 3D Structure Prediction Accuracy using Dense Regions Areas of Secondary Structures in the Contact Map. American J Biochem Biotechnol 4: 375-384.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410.
4. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.
5. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, et al. (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res 29: 37-40.
6. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, et al. (2004) UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 32: D115-9.
7. Bateman A, Birney E (2000) Searching databases to find protein domain organization. Adv Protein Chem 54: 137-157.
8. Carlin A, Shi W, Dey S, Rosen BP (1995) The ars operon of Escherichia coli confers arsenical and antimonial resistance. J Bacteriol 177: 981-986.
9. Carlsson J (1997) Bacterial metabolism in dental biofilms. Adv Dent Res 11: 75-80.
10. Copleya RR, Doerksa TB, Letunica I, Bork P (2002) Protein domain analysis in the era of complete genomes. FEBS Lett 513: 129-134.
11. Deng DM, Urch JE, ten Cate JM, Rao VA, van Aalten DM (2009) Streptococcus mutans SMU.623c codes for a functional, metal-dependent polysaccharide deacetylase that modulates interactions with salivary agglutinin. J Bacteriol 191: 394-402.
12. Eddy SR (2001) Profile hidden Markov models. Bioinformatics 14: 755-763.
13. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJA, et al. (2002) The PROSITE database, its status in 2002. Nucleic Acids Res 30: 235-238.
14. Finn RD, Mistry J, Schuster-Bockler B (2006) Pfam: clans, web tools and services. Nucleic Acids Res 34: 247–251.
15. Fiona MR, Ruth M, Sharon JP (2003) Characterization of novel LPXTG-containing proteins of Staphylococcus aureus identified from genome sequences. Microbiol 149: 643-654.
16. Galperin MY, Koonin EV (1998) Sources of systematic error in functional annotation of genomes: domain rearrangement, nonorthologous gene displacement and operon disruption. In Silico Biol 1:55-67.
17. Kado CI (2009) Horizontal gene transfer: sustaining pathogenicity and optimizing host-pathogen interactions. Mol Plant Pathol 10: 143-150.
18. Krause A, Stoye J, Vingron M (2005) Large scale hierarchical clustering of protein sequences. BMC Bioinformatics 6: 15.
19. Lawrence JG (1997) Selfish operons and speciation by gene transfer. Trends Microbiol 5: 355–359.
20. Letunic I, Copley RR, Schmidt S (2004) SMART 4.0: towards genomic data integration. Nucleic Acids Res 32: 142-144.

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jpb.1000148
JPB/Vol.3 Issue 8

21. Liu J, Rosen BP (1997) Ligand interactions of the ArsC arsenate reductase. J Biol Chem 272: 21084-21089.

22. Loesche WJ (1986) Role of Streptococcus mutans in human dental decay. Microbiol Rev 50: 353-380.

23. Marraffini LA, DeDent AC, Schneewind O (2006) Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria. Microbiol. Mol. Biol. Rev70: 192–221.

24. Martin P, DeMel S, Shi J, Gladysheva T, Gatti DL (2001) Insights into the structure, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. Structure 9: 1071-1081.

25. Mazmanian SK, Liu G, Ton-That H, Schneewind O (1999) Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. Science 285: 760-763.

26. Navarre WW, Schneewind O (1999) Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol. Mol Biol Rev 63: 174-229.

27. Niu G, Okinaga T, Zhu L, Banas J, Qi F (2008) Characterization of irvR, a novel regulator of the irvA-dependent pathway required for genetic competence and dextran-dependent aggregation in Streptococcus mutans. J Bacteriol 190: 7268-7274.

28. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302: 205-217.

29. Nyvad B, Kilian M (1990) Comparison of the initial streptococcal microflora on dental enamel in caries-active and in caries-inactive individuals. Caries Res 24: 267–272.

30. Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405: 299–304.

31. Ochman H, Lerat E, Daubin V (2005) Examining bacterial species under the specter of gene transfer and exchange. Proc Natl Acad Sci 102: 16595-6599.

32. Rao TS, Basu N, Siddique HH (1982) Anti-Infl ammatory Activity of Curcumin Analogues. Indian J Med Res 75: 574-578.

33. Schultz J, Milpetz F, Bork P (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA 95: 5857-5864.

34. Shin DH, Lou Y, Jancarik J, Yokota H, Kim R, et al. (2005) Crystal structure of TM1457 from Thermotoga maritima. J Struct Biol 152: 113-117.

35. Sonnhammer ELL, Kahn D (1994) Modular Arrangement of Proteins as Inferred from Analysis of Homology. Protein Sci 3: 482-492.

36. Stein RL (1991) Exploring the catalytic activity of immunophilins. Curr Biol 1: 234-236.

37. Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W-Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. Nucleic Acids Res 22: 4673-4680.

38. Titgemeyer F, Reizer J, Reizer A, Saier MH (1994) Jr Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria. Microbiology 140: 2349-2354.

39. Tropschug M, Wachter E, Mayer S, Schoenbrunner ER, Schmid FX (1990) Isolation and sequence of an FK506-binding protein from N. crassa which catalyses protein folding. Nature 346: 674-677.

40. Wang T, Donahoe PK, Zervos AS (1994) Specific interaction of type I receptors of the TGF-beta family with the immunophilin FKBP-12. Science 265: 674-676.

41. Yeats C, Bentley S, Bateman A (2003) New Knowledge from Old: In silico discovery of novel protein domains in Streptomyces coelicolor. BMC Microbiol 3: 3.

42. Zhang RG, Grembecka J, Vinokour E, Collart F (2002) Structure of Bacillus subtilis YXKO--a member of the UPF0031 family and a putative kinase. J Struct Biol 139: 161-170.

43. Zhou YF, Li LF, Yang C, Liang YH (2008) Preliminary X-ray crystallographic analysis of SMU.573, a putative sugar kinase from Streptococcus mutans. Acta Crystallogr F 64: 47-49.

44. Zuber P (2004) Spx-RNA polymerase interaction and global transcriptional control during oxidative stress. J Bacteriol 196: 1911-1919.