

Research article

Open Access

## Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes

A Krupa<sup>1,2</sup> and N Srinivasan\*<sup>1</sup>

Address: <sup>1</sup>Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India and <sup>2</sup>Cell Cycle Control Laboratory, London Research Institute, Cancer Research – UK, South Mimms, Hertfordshire, EN6 3LD UK

Email: A Krupa - [ananth.krupa@cancer.org.uk](mailto:ananth.krupa@cancer.org.uk); N Srinivasan\* - [ns@mbu.iisc.ernet.in](mailto:ns@mbu.iisc.ernet.in)

\* Corresponding author

Published: 19 September 2005

Received: 02 October 2004

*BMC Genomics* 2005, **6**:129 doi:10.1186/1471-2164-6-129

Accepted: 19 September 2005

This article is available from: <http://www.biomedcentral.com/1471-2164/6/129>

© 2005 Krupa and Srinivasan; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Ser/Thr/Tyr kinases (STYKs) commonly found in eukaryotes have been recently reported in many bacterial species. Recent studies elucidating their cellular functions have established their roles in bacterial growth and development. However functions of a large number of bacterial STYKs still remain elusive. The organisation of domains in a large dataset of bacterial STYKs has been investigated here in order to recognise variety in domain combinations which determine functions of bacterial STYKs.

**Results:** Using sensitive sequence and profile search methods, domain organisation of over 600 STYKs from 125 prokaryotic genomes have been examined. Kinase catalytic domains of STYKs tethered to a wide range of enzymatic domains such as phosphatases, HSP70, peptidyl prolyl isomerases, pectin esterases and glycoproteases have been identified. Such distinct preferences for domain combinations are not known to be present in either the Histidine kinase or the eukaryotic STYK families. Domain organisation of STYKs specific to certain groups of bacteria has also been noted in the current analysis. For example, Hydrophobin like domains in Mycobacterial STYK and penicillin binding domains in few STYKs of Gram-positive organisms and FHA domains in cyanobacterial STYKs. Homologues of characterised substrates of prokaryotic STYKs have also been identified.

**Conclusion:** The domains and domain architectures of most of the bacterial STYKs identified are very different from the known domain organisation in STYKs of eukaryotes. This observation highlights distinct biological roles of bacterial STYKs compared to eukaryotic STYKs. Bacterial STYKs reveal high diversity in domain organisation. Some of the modular organisations conserved across diverse bacterial species suggests their central role in bacterial physiology. Unique domain architectures of few other groups of STYKs reveal recruitment of functions specific to the species.

### Background

Extracellular signals perceived by the cells invoke an appropriate intracellular response triggering cascades of events that alter the activities of various signalling molecules. The variations in functional states of the proteins

are brought about by covalent and non-covalent molecular interactions. Intracellular signalling in prokaryotes and eukaryotes is often achieved by phosphorylation of target proteins. The phosphorylated proteins undergo changes in their three-dimensional (3-D) shapes resulting in

altered functional states. Phosphorylation events also lead to changes in the electrostatic interactions and ionisation states of residues at the catalytic site as exemplified by regulation of isocitrate dehydrogenases and HMG-CoA reductase [1,2]. Covalent attachment of the phosphoryl groups to various proteins is catalysed by protein kinases. These enzymes transfer a phosphate group from adenosine tri phosphate (ATP) onto an acceptor amino acid in a substrate protein. The acceptor groups on amino acid residues could be alcoholic as in Ser and Thr, phenolic group as in tyrosine or basic amino acid such as histidine ( $N^{\pi}$  or  $N^{\tau}$ ), arginine (guanidino group), and the  $\epsilon$ -NH<sub>2</sub> group of lysine, carboxyl groups of aspartate and glutamate or thiol group of cysteine [3,4]. In addition to nucleoside triphosphates as donors of the phosphate groups, certain class of phosphorylating enzymes makes use of either phosphoenzyme intermediates or low molecular weight metabolites such as phosphoenol pyruvate, acetyl phosphate or carbamoyl phosphates and polyphosphates as donors [5-7].

Despite the occurrences of various phosphorylating enzymes, cellular signalling mechanisms are dependent on a subset of the list of enzymes mentioned above, for intra-cellular communications. In eubacteria and archaea, two component signal transduction systems [8], also referred to as His-Asp phospho-relay systems, mediate adaptive responses to changes in environmental conditions. The two-component system is composed of a histidine kinase (HK) and a cognate regulatory protein referred to as a response regulator (RR). HKs consist of an ATP-binding kinase domain and the H-box domain carrying the histidine which is the site of phosphorylation. HKs have also been identified in eukaryotes, such as yeast, *Neurospora* and *Arabidopsis thaliana*, and their functional roles have been studied [9,10].

However, in eukaryotes, Ser/Thr and Tyr kinases play a key role in cellular signal transduction. The Ser/Thr and Tyr kinases differ completely from the HKs both in sequence and structure. The Ser/Thr and the Tyr kinase share a common catalytic core of about 270 residues made up of N-terminal lobe consisting of mainly  $\beta$ -strands and a C-terminal lobe composed predominantly of  $\alpha$ -helices [11]. The eukaryotes also possess two other distinct classes of protein kinases, the myosin heavy chain kinase/EF-2 kinase [12] and the mitochondrial protein kinases that include the pyruvate dehydrogenase kinase and branched-chain alpha-ketoacid dehydrogenase kinase [13] that differ from the Ser/Thr kinases in sequence and structure. Protein phosphorylation is hence considered to have originated independently in prokaryotes and eukaryotes in order to meet their diverse cellular and environmental requirements.

However studies on the eukaryotic protein-Ser/Thr/Tyr kinase like sequences (STYKs) in the prokaryotes emerged with the identification of proteins phosphorylated on Ser/Thr and Tyr residues in bacteria [14]. Phosphorylation of Ser/Thr residues in archaeal proteins has also been reported for *Sulfolobus acidocaldarius* by earlier studies [15]. Subsequent identification and characterisation of genes encoding STYKs in *Myxococcus xanthus* [16] and *Streptomyces* species [17] confirmed the presence of homologues of eukaryotic protein kinases in bacteria. Following these reports of STYKs, and the release of large number of bacterial genome sequences, a number of groups have surveyed for the occurrences of STYKs in them [18-22]. Their surveys suggested the occurrences of significant number of STYKs in both archaea and eubacterial genomes and they have been further classified based on their sequence similarity to various subfamilies [19]. A previous study [20] proposed that the Ser/Thr/Tyr protein kinases in bacteria have bacterial origin. Kinase activity in the STYKs has been demonstrated in a few species. These studies have suggested the involvement of STYKs in secondary metabolism [17], morphology of the aerial hyphae in *Streptomyces granticolor* [23], regulation of fruiting body formation in *M.xanthus* [24]. STYKs in *Pseudomonas aeruginosa* and *Yersinia pseudotuberculosis* [25,26] are shown to be critical for progression of infection and expression of virulence factors during certain stages of infection. The role of STYKs in the repression of proline utilisation operon and degradation of proline in *S.typhimurium* has been suggested [27].

The recently reported 3-D crystal structures of Mycobacterial PKN-B [28,29] has revealed an overall similarity of their structure to the eukaryotic Ser/Thr protein kinases with variations in the nucleotide binding regions. The prokaryotic homologues also include aminoglycoside kinases which phosphorylate antibiotics, and share remote sequence similarity with the STYKs. The 3-D fold of their catalytic domain is highly similar to the structures of eukaryotic STYKs [30]. Studies on the evolution of STYKs in *Synechocystis* sp. PCC 6803 have proposed the origin of STYKs before the divergence between prokaryotes and eukaryotes during evolution [20]. The conservation of these sequences in significant number of bacterial genomes despite limitation on their genomic size during evolution have therefore suggested their functions are essential in regulation of cellular activities in prokaryotes, as in their eukaryotic counterparts.

The present analysis is focussed on understanding the biological roles of the prokaryotic STYKs which have been shown to be encoded by significant number of genes in a large number of bacterial genomes by various groups. The present study is based on representative domain architectures of STYKs identified from the completely sequenced genomes of 125 prokaryotes which includes 35 genomes

compiled in our early version of KinG database [31]. The domain architectures of STYKs and their homologues in 90 additional genomes included in this paper are recently added to the KinG database <http://hodgkin.mbu.iisc.ernet.in/~king>.

About 600 STYKs and their homologues have been identified in the current study in 125 bacterial genomes, of complete genomic data, using various sequence and profile search methods. Manual analysis of each of these STYKs and homologues has been made in order to check for the presence or absence of residues known to be critical for the activity of eukaryotic protein kinases. Using profile based search method (See *Materials and Methods*) the extent of similarity of bacterial homologues with different subfamilies of eukaryotic protein kinases has been analysed. The analysis suggests that the bacterial homologues form a distinct class of protein kinases as indicated by the highest similarity (20–30% sequence identity) of their catalytic kinase domain with the generic kinase profiles rather than any individual subfamily profile of eukaryotic protein kinases. i.e., the percent sequence identities of bacterial kinases with specific sub-families of eukaryotic STYKs is lower.

During the course of current analysis we encountered lipopolysaccharide kinases (LPSKs) which are well known for their remote relationship with STYKs [32]. RIO1 class of STYKs have been previously identified in eukaryotes [33,34] and in archaea [19] and our earlier work reported the occurrence of RIO1 proteins in all the three major taxons namely, Archaea, Bacteria and Eukaryota [32].

The domain organisation of STYKs identified from various genomes has been studied in an attempt to understand their probable biological roles. Groups of STYKs conserved only among specific classes of bacteria have been identified. Diverse domain arrangements observed in bacterial STYKs are radically different from those observed in eukaryotic Ser/Thr or Tyr kinases and prokaryotic HKs. These domain arrangements unique to bacterial STYKs are therefore suggestive of their functional roles distinct from that of other well-known families of protein kinases. Consideration of known functions of domains tethered to catalytic kinase domain of bacterial STYKs enabled obtaining clues about gross functional roles of various bacterial STYKs. Involvement of STYKs in stress response, protein folding, extensive protein-protein interactions and sugar transport has been suggested by covalent tethering of catalytic kinase domains of STYKs with other enzyme and non-enzyme modules. Homologues of bacterial PKs that have been previously well characterised in terms of biochemical activity and function have also been identified in the study. Putative substrates for a small set of bacterial STYKs have also been identified based on

their similarity to bacterial kinases with well characterised substrates.

## Results

The STYK-like sequences have been identified in various completely sequenced bacterial genomes and have been analysed for the occurrence of other functional domains. The domains have been identified using different profile search methods with stringent search parameters as described in the section on *Materials and Methods* and have also been manually checked for the conservation of key motifs. It should be noted that we have used stringent e-value cut-off of  $10^{-8}$  and 0.01 respectively in the PSSM (Position Specific Scoring Matrix) matching and HMM (Hidden Markov Model) matching procedures. All the functional domain assignments discussed below are detected within the e-value limit of 0.01 by HMMER and are ensured, manually, that there is no obvious erroneous domain assignment.

### **Distribution of the STYKs and STYK-like sequences in bacterial genomes**

The data set of bacterial genomes considered in the current analysis includes predicted ORFs from the genomes of sixteen archaeal and 109 bacterial genomes representing major division of these prokaryotic phylogenetic domains. A list of these completed genomes and number of Protein Kinase-like sequences (PKLS) that includes STYKs, RIO1 and ABC1-like kinases is given in Table 1. Table 1 also includes information on number of protein kinase-like sequences lacking the catalytic base residue (aspartate) and hence are likely to be non-functional as kinase. All the archaeal genomes analysed have at least one STYK, which belongs to the RIO1 family [19]. The RIO1s share a significant sequence similarity to the eukaryotic protein kinases and have recently been shown to possess a Ser/Thr autophosphorylation activity [35]. These RIO1 sequences have been initially observed in yeast [33,34] and subsequently detected in other eukaryotic genomes as well [36]. We have previously identified RIO1 like sequences in two of the eubacterial species, *Deinococcus radiodurans* and *Pseudomonas aeruginosa* [32].

Occurrence of RIO1 in many other eubacterial genomes has been reported in this paper for the first time. RIO1 sequences have been identified in other bacterial species including *Bradyrhizobium japonicum*, *Pseudomonas putida*, *Pseudomonas syringae*, *Shewanella oneidensis* and *Yersinia pestis*. The ubiquitous nature of RIO1 across archaea, bacteria and eukaryota is thus evident. Examination for the occurrences of HKs among the genome analysed reveal their absence in significant number of archaeal genomes, as well as in mycoplasma, *Buchnera* and *Onion yellows phytoplasma* genomes.

**Table 2: Representative modular organisation of bacterial STYKs identified in the current analysis.**

Representative STYK	Domain organisation as identified in the current analysis	Bacterial species encoding STYKs of similar domain organisation
gij 9947763	Phosphatase-2C (PP2C) + Kinase	<i>Pseudomonas aeruginosa</i> , <i>Pseudomonas putida</i> , <i>Deinococcus radiodurans</i> .
gij 32474490	Kinase + HSP70	<i>Pirellula</i> sp.
gij 31793271	Kinase + Hydrophobin	<i>Mycobacterium bovis</i>
gij 3261694	Kinase + MalT-like ABC transporter domain	<i>Mycobacterium tuberculosis</i> H37Rv
gij 41407147	Kinase + trans-membrane segment + peptidyl-prolyl cis-trans isomerase domain	<i>Mycobacterium avium</i> sub sp. paratuberculosis
gij 23020879	Kinase + 6 TPR repeats + GGDEF	<i>Clostridium thermocellum</i>
gij 17229994	Kinase + pectinesterase	<i>Nostoc</i> sp. PCC 7120
gij 17134087	Kinase + GUN4	<i>Nostoc</i> sp. PCC 7120
gij 9945898	Kinase + von Willebrand Type A domain	<i>Pseudomonas aeruginosa</i>
gij 17129689	Kinase + ANF Receptor	<i>Nostoc</i> sp. PCC 7120
gij 21223285	Kinase + bacterial extracellular solute binding protein	<i>Streptomyces coelicolor</i>
gij 1006577	Kinase + SH3b	<i>Synechocystis</i> species
gij 17129893	FHA + Kinase	<i>Nostoc</i> sp. PCC 7120, <i>Thermodesmium erythraeum</i> , <i>Thermosynechococcus elongatus</i> , <i>Chloroflexus aurantia</i> , <i>Nostoc punctiforme</i> .
gij 1709642 (PKN2)	Kinase + Guanylate cyclase + Trans-membrane segment	<i>Myxococcus xanthus</i>

### Domain arrangements in bacterial homologues of eukaryotic protein kinases

Interactions of multi-modular signalling proteins with cognate ligands are often mediated by the constituent modules. The domain organisation of the STYKs has been analysed extensively by various profile-search and fold recognition methods described in a later section. The details of the domain organisation of the bacterial kinases analysed in this paper have been included in the KinG database <http://hodgkin.mbu.iisc.ernet.in/~king>[31,37]. Incorporation of information about STYKs in 125 prokaryotic genomes in the KinG database represents a major update in the database compared to the previous version with STYKs only from 35 prokaryotic genomes. A comprehensive study on the domain organisation of the previously compiled repertoires of STYKs and those encoded in the additional genomes has been described here. Modular organisation of STYKs, newly identified in the current analysis have been listed in Table 2

Influence of the diverse domain composition of STYKs and their probable biological functions are discussed in the following sections.

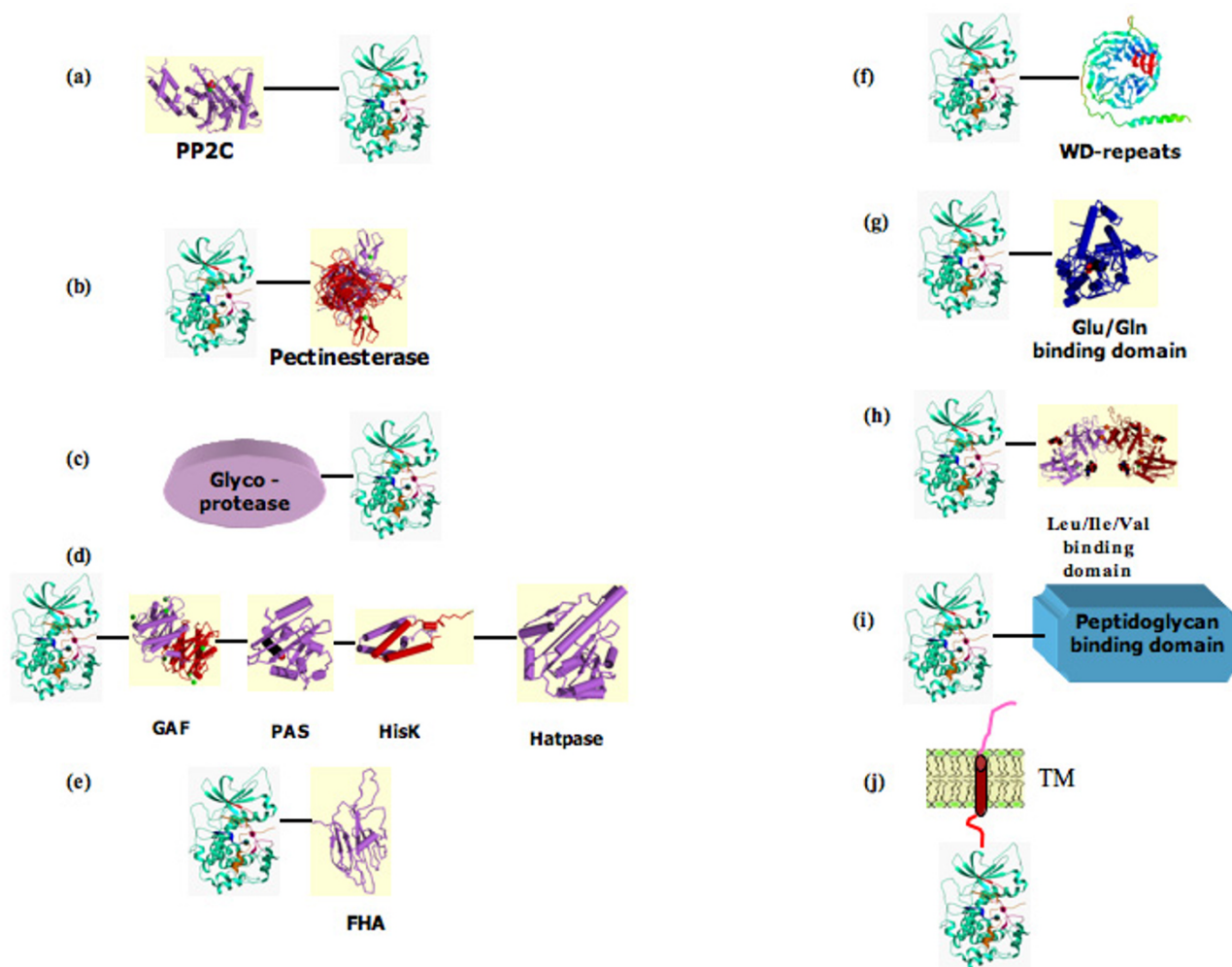
### Transmembrane bacterial Ser/Thr kinases

The amino acid sequences of kinase catalytic domains of bacterial STYKs are generally more similar to eukaryotic Ser/Thr kinases than the eukaryotic tyrosine kinases. The profiles of Ser/Thr kinase subfamilies match the bacterial kinases with better similarity ( $<1e-32$ ) in contrast to the tyrosine kinases ( $>1e-17$ ). However a significant number of bacterial STYKs contain trans-membrane spanning seg-

ments unlike the eukaryotic Ser/Thr kinases. The known rare exceptions of eukaryotic transmembrane protein with Ser/Thr kinase domain are TGF-beta receptor and the IRE1. Trans-membrane STYKs are absent in certain Gram-negative species *Haemophilus influenzae*, *Rickettsia*, *Helicobacter pylori*, *Chlamydiae* and in the hyperthermophilic bacteria, *Aquifex aeolicus*. The Gram-positive species analysed in the data set have at least one trans-membrane STYKs in them. The presence of significant number of trans-membrane kinases (TM-kinases) in most bacterial species suggests many of them could play as receptor like kinases and are probably involved in direct interaction with the extra-cellular ligands. Further the TM kinases with extracellular PASTA (Fig 1i) domain such as PKNB [28,29] of *Mycobacterium tuberculosis* are encoded in diverse groups of Gram-positive bacteria analysed in the current study. The selective conservation of PASTA domains in STYKs encoded by Gram-positive bacteria implies their role in signalling pathways conserved across diverse Gram positive species. Currently the nature of ligands stimulating the trans-membrane STYKs is not known.

### Archaeal multi-domain STYKs

The multidomain STYKs of archaea include the kinases with O-sialoglycoprotein endopeptidases (OSGP) domain (Fig: 1c) present at the N-terminal of the kinase domain. The OSGP domains belong to the glycoprotease family (PFAM00814). The members of this family are known to cleave proteins that are heavily sialylated on serine or threonine residues. The OSGP containing STYKs described in earlier studies [19] are also encoded only in



**Figure 1**

The representative domain combinations of the bacterial homologues of eukaryotic protein kinases are shown. The accession numbers of the protein domain families classified by PFAM are given below. (a) PP2c: phosphatase 2C [PF00481]; (b) Pectinesterase [PF01095]; (c) Glycoprotease [Peptidase\_M22; PF00814]; (d) GAF[PF01590], PAS[PF00989], HisK[PF00512], HAT-Pase[PF02518]. (e) FHA: Fork head associated domain [PF00498]; (f) VVD-repeats [PF00040]; (g) and (h) amino acid binding domain [ANF\_receptor; PF01094]; (i) peptidoglycan binding domain [PASTA, PF03973] k: trans-membrane segment [TM] shown in red.

archaeal genomes. The archaeal species with the OSGP domains in their STYKs include *Methanococcus janaschii*, gi|2826367), *Methanothermobacter thermoautotrophicus* (gi|2622538), *Halobacterium sp* (gi|1058146), and *Thermoplasma acidophilum* (gi|10639497). OSGP domains of archaeal STYKs bear highest sequence similarity with the OSGP domains of the yeast gene product YK18 which is a putative glyco protease.

**Phosphatase associated protein kinases**

We report for the first time the tethering of the PP2C domains with STYK domain in bacteria. This unique class

of STYKs has not been characterised so far. The PP2C like domains (Fig 1a) are tethered to the N-terminal of kinase domains in few STYKs encoded in *Pseudomonas putida* and *Pseudomonas syringae*. In addition to the well-characterised STYK of *P.aeruginosa* [38] a protein kinase (gi|9947763) with phosphatase-like domain described above has been identified in the current study. Earlier studies have biochemically characterised Stp1 of *Pseudomonas aeruginosa*, a single domain protein homologous to PP2C [38]. The regulation of Stp1 by its neighbouring gene encoding the protein kinase Stk1 has been suggested to be similar to the regulation of phosphatase and Ser/Thr kinases, YopH

and YpkA of *Y.pseudotuberculosis* that serve as critical virulence determinants. Phosphatase like domain of *Pseudomonas* species of the composite STYK shares 35% identity with Stp1 and about 30% identity with PP2C domains of Gram positive bacteria.

#### **Protein kinases associated with nitrate dissimilation**

A group of STYKs (gi|7190465, gi|4376414, gi|8978521, gi|7189538) specifically conserved in different species of *Chlamydiae* has a highly conserved C-terminal domain. A domain family previously referred to as NirV in Conserved Domain Database [39] has been assigned to this C-terminal region (E-58). This homologous domain family comprises of NirV like proteins, methylase involved in chemotaxis and sulphatase modifying enzymes and is referred to as DUF323 (Domain of unknown Function) in the PFAM database. The crystal structure of DUF323 has been recently solved [40] for a human homolog of Formylglycine generating enzyme (pFGE), which is a non-functional counter part of human FGE. FGE modifies the active site residue glycine of the sulphatases. The NirV domain of the Chlamydial kinases shares a sequence identity of 37% over a stretch of 100 residues with the NirV protein from *Rhodobacter spaeroides*. The NirV gene is situated in the operon containing two other genes namely, NirK and ppaZ whose gene products regulate nitrate dissimilation. The disruption of NirV gene in other organisms does not affect the nitrate reduction [41].

#### **Protein kinases with domains homologous to modules associated with chaperones**

We have identified a protein kinase (gi|32474490) with C-terminal domain homologous to HSP70 family of proteins in *Pirellula* species. The HSP70 family includes proteins that act as chaperones and assist the folding of a wide range of proteins.

Another class of STYKs that are also potentially involved in interactions with chaperonic proteins has been identified. STYKs with Tetratric peptide repeats (TPR) have been identified in a number of bacterial species including *M.tuberculosis*, *Clostridium acetobutylicum*, *Gloeobacter violaceus*, *Leptospira interrogans*, *Nostoc sp.*, *Pirellula*, *Streptomyces avermitilis*, *Sulfolobus solfataricus* and *Sulfolobus tokodaii*. TPR is known to form a series of anti parallel amphipathic  $\alpha$ -helices that bundle together through hydrophobic interactions to form a fold with deep concave groove. They are suggested to bind a number of proteins of regulatory importance such as HSP90 based chaperone system [42].

#### **Stress-responsive protein kinases**

In the current analysis we have identified STYKs (gi|13475495, gi|27353958 and gi|30180007) with universal stress induced protein associated (USPA) domain

encoded in the genomes of *Mesorhizobium loti*, *Bradyrhizobium japonicum* and *Nitrosomonas europaea*. Protein kinases with USPA domains have been recently identified in STYKs of *Arabidopsis thaliana* [43] and HisK of bacteria [44].

#### **Protein kinases with protein – protein interaction modules**

Protein kinases considered in the analysis have also revealed their association with domains made up of repeats that serve as an interface for the interaction of protein modules. An other class of STYKs with PQQ family repeats have been identified in *Deinococcus radiodurans*, *Streptomyces avermitilis* and *Streptomyces coelicolor* that have been described in the greater detail in later section with respect to their role in regulation of secondary metabolism. The PQQ repeats are known to form  $\beta$ -propeller like structures similar to WD – repeats that could serve as sites of interactions with target proteins.

STYK (gi|9948555) with leucine-rich repeats (LRR) have also been identified in *Pseudomonas aeruginosa*. LRRs are known to serve as protein interaction domains that form non-globular structures made up repeated by subunits [45].

NHL repeats are also identified in two STYKs of *M. tuberculosis* (gi|2078052) and *D.radiodurans* (gi|6457717) which have 6 and 3 repeats following the kinase domain. They are known to serve as protein interaction surfaces and some times possess enzymatic activity as exemplified by the bifunctional peptidyl glycine  $\alpha$ -amidylating monooxygenase protein.

#### **Ubiquinone biosynthesis associated protein kinases**

Significant number of archaeal and bacterial genomes encode STYKs with regions most similar to ubiB gene products flanking the catalytic kinase domain. Ubiquinone, is a critical component of the respiratory chain of most of organisms. The gene product of ubiB is involved in the production of an intermediate of Ubiquinone or the co-enzyme Q [46]. Based on the sequence comparisons, this group of kinases has been predicted to be involved in the regulation of ubiquinone biosynthesis [19]. This group of STYKs had been grouped into ABC1 family and was initially believed to have restricted occurrence in archaea [19]. The currently available genomic data reveals a wider representation of the ABC1 kinases in significant number of archaeal species.

They include *Halobacterium salinarium*, *Thermoplasma acidophilum*, *Methanosarcina acetovirans*, *Methanosarcina mazei*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Thermoplasma volcanium* All the ABC1 kinases encoded in archeal species analysed in the current study have an N-terminal ABC1 kinase domain and a putative trans-membrane

spanning segments at the C-terminal. *Thermoplasma acidophilum* (gi|10639567) also has an unique ABC1 kinase with N-terminal RI01 like domain and C-terminal ABC1 Kinase-like domain followed by a putative trans-membrane spanning segment. ABC1 family of STYKs has a wide representation in large number of bacterial species. As described before *Nitrosomonas europaea* encodes a STYK of ABC1 family that has a N-terminal Usp domain.

#### **Phosphorylation and structural components of bacterial cells**

A set of closely related protein kinases is conserved in *Mycobacterium bovis subsp. bovis AF2*, *Mycobacterium tuberculosis CDC1551* and *Mycobacterium tuberculosis H37Rv*. This new group of kinases identified in the current analysis has a N-terminal protein kinase domain and a C-terminal hydrophobin like regions. The hydrophobins are low molecular weight cysteine rich proteins that form a critical component of the hydrophobic sheaths covering the fungal spores.

STYKs conserved among all the Gram-positive bacteria considered in the current analysis, have a C-terminal region that is highly similar to the C-terminal region of high molecular weight penicillin binding proteins. This domain is believed to play a role in the recognition of D-alanyl D-alanine dipeptides used to build up the peptidoglycan layers and is classified into the family called PASTA [47] in Pfam database due to their occurrence in bacterial Ser/Thr protein kinases and penicillin binding proteins.

#### **Other domain combinations observed in bacterial protein kinases**

A STYK found in the *M.tuberculosis* (gi|3261694) is comprised of an N-terminal kinase domain followed by C-terminal domain that shares high sequence similarity to distinct group of ABC transporters, MalT (E-value:  $10^{-180}$ ). MalT group of transporters are unique in terms of their direct participation of the transporter in transcriptional regulation of *mal* regulon [48]. These observations suggest the possible role of the STYK in the regulation of uptake of sugars. This domain has been observed only in *M.tuberculosis* STYK and is not found in the closely related species, *M.leprae*. This protein kinase is likely to be critical for nutrient uptake.

*M.tuberculosis* also contains a PK (gi|2131007) with a domain that shares significant similarity (e-value  $10^{-4}$ ) with the periplasmic oxido reductase of thioredoxin family (DSBA). The thioredoxins are known for their interaction with a broad range of protein as a general protein disulphide oxido reductase. The potential role of this kinase for its possible involvement in disulphide isomerisation of various proteins during protein folding is there-

fore suggested. The occurrence of the DSBA domain in the extracellular region of the PK emphasises their role in reduction of periplasmic mycobacterial proteins. This is unique to *Mycobacterium tuberculosis* and is not found in *M.leprae*.

The genome of *Mycobacterium avium subsp. paratuberculosis* also encodes another STYK probably assisting the folding of proteins. The domain organisation of this novel kinase includes an N-terminal kinase domain, a central trans-membrane domain and a C-terminal peptidyl-prolyl cis-trans isomerase (Pro\_isomerase) domain. Among all available completely sequenced genomes of various species of Mycobacteria, this novel protein kinase is encoded in only in the *M.avium* genome.

*Clostridium thermocellum* ATCC 27405 encodes one of the largest STYK of 1826 amino acids encompassing a N-terminal kinase domain with six TPR repeats followed by the GGDEF domain at the C-terminal end. The domains of sensor kinases and response regulators of the His-Asp phospho relay systems have been studied previously [49]. The GGDEF domain is suggested and subsequently shown to be homologous to the catalytic domain of adenyl cyclases [50,51]. Recent studies have also suggested their role in the regulation of bacterial growth and development through novel response regulators [52] by turnover of cyclic diguanosine monophosphate. Studies have also underlined the importance of adjacent sensor domains in the regulation of the catalytic activity of GGDEF [53]. The occurrence of GGDEF with STYKs implies the sharing of down stream signalling components by the HisKs and STYKs and their role in production of cyclicdiguanylate.

The Nostoc sp. PCC 7120 encodes a protein kinase (gi|17229994) with a pectinesterase domain (Fig 1b). The pectin esterases catalyse the hydrolysis of pectin, a component of extra cellular capsules and cell walls. The association of two such enzymatic activities is unique to this organism and suggests the probable role of phosphorylation in maintenance of integrity of bacterial cellular structural components.

A novel protein kinase (gi|17134087) with a C-terminal GUN4 domain has been identified in the Nostoc sp. PCC 7120. The GUN4 proteins in *Arabidopsis thaliana* are known to bind to magnesium-protoporphyrin IX, an intermediate in chlorophyll biosynthesis and there by regulating the enzyme Mg-chelatase involved in the biosynthesis [54]. The novel protein kinase is therefore suggested to be involved in the regulation of chlorophyll production in Nostoc species.

*Pseudomonas aeruginosa*, known to be an opportunistic pathogen encodes a total of 11 STYK-like sequences.

Seven of these are STYKs and the rest are KDO kinases. We have identified the von Willebrand A domain located to the C-terminal of kinase domain in one of the STYKs that has been previously referred to as Ppka [55]. The glycoprotein binding vWA domain [56] might be playing an important role in its interaction with the host proteins as described in the later sections.

A STYK (gi|3328716) in *Chlamydia* species has a N-terminal kinase domain, and a C-terminal domain with similarity to TPR domain. Using remote homology detection methods as described in the 'Materials and Methods' section, the similarity of this region with the TPR domain of O-linked N-acetylglucosamine (O-GlcNAc) transferases [57] has been detected. The TPR domain of O-GlcNAc transferase (OGT) has been recently shown to be responsible for its targeting OGT and in determining its substrate specificity [58]. The domain homologous to TPR in the STYK is therefore likely to mediate its substrate interactions.

#### **Ser/Thr kinases of bacteria with periplasmic protein domains**

The periplasmic solute binding proteins are a part of the ATP binding cassette (ABC) transporter that transfers the amino acids and sugars into the cytoplasm. Ser/Thr kinases from *Streptomyces coelicolor* (gi|21223285) and *Nostoc* sp. 7120 (gi|17227840) contain a solute binding domain for Glu, Gln (Fig 1g) and ANF receptor like domain (Fig 1h), which is involved in the binding of branched chain amino acids respectively.

#### **Bacterial STYKs with non catalytic domains homologous to domains associated to eukaryotic PKs**

This section describes the possible functions of a few bacterial STYKs inferred based on their similarity in domain combination with the well characterised eukaryotic protein kinases.

*Synechocystis* sp. is a cyanobacteria that encodes 12 STYKs. One of the STYK (gi|1006577) has a SH3b (bacterial SH3 domain) following the N-terminal kinase domain. Occurrences of SH3 domains in bacteria have been previously well documented [59]. They have also been classified into a different family in SMART database [60].

Protein kinases with WD repeats (Fig 1f) were identified in many bacterial species including *Thermospora curvatum*, *Thermodesmium erythraeum*, *Chloroflexus aurantia*, *Nostoc punctiforme*, *Nostoc* sp. PCC 7120 and *Thermobifida fusca*. Most of these kinases contain 7 WD repeats and are identified in cyanobacteria. The number of repeats varied in few from 4 to 14 and all the species mentioned above contain more than one such protein kinase with the exception of *Thermobifida fusca*. The WD repeats are associated with

beta-propeller like folds and the most extensively studied are the G $\beta$ -subunit of the heterotrimeric G-proteins in eukaryotes.

The FHA domain containing PKs (Fig: 1e) have been identified in the current analysis in the cyanobacterial species, *Thermodesmium erythraeum* (gi|23042923, gi|23043171, gi|23041283), *Thermosynechococcus elongatus* BP-1 (gi|22298671, gi|22299030) *Chloroflexus aurantia* (gi|22970578), *Nostoc punctiforme* (gi|23127167), and *Nostoc* sp. PCC 7120 (gi|17232446, gi|17228044, gi|23043171). The phospho - threonine binding residues GR, SxxH and NG are well conserved in all these FHA domain containing protein kinases (Fig. 2). In eukaryotes the FHA containing protein kinases are conserved from yeast to mammals and are involved in the regulation of the various stages of cell cycle and development [61].

The *Myxococcus xanthus* protein kinase, PKN2 has an interesting domain organisation. The PK consists of an N-terminal kinase domain that extends as a stretch of Ala/Ser/Pro/Thr rich segments which further continues as a trans membrane segment as described in earlier studies [62,63]. We have identified a Adenylyl/ Guanylyl cyclase (AC/GC) domain in the region between the Ala/Ser/Pro/Thr rich regions and the trans membrane (TM)- region. Regions homologous to cyclases have been also observed in previous studies [64,65]. Cyclases of this nature also occur in the eukaryotes. A stretch of nearly 200 residues with no similarity to any extra cellular domains known so far lies on the external side of the TM segment. However the protein kinase like domain (PKLD) in the eukaryotic GCs lack the catalytic residues. In the case of PKN2 the PK domain has all the catalytic residues and is also known to phosphorylate on Ser/Thr residues [62]. The conservation of catalytic residues of cyclase domain and the topology of the gene product has been analysed in the current study. The key residues at the catalytic sites of the GC, Glu and Cys confer nucleotide specificity to the enzyme. Mutation of Glu and Cys to Lys and Asp, the corresponding residues of the adenylyl cyclases, changes the nucleotide specificity of GC to adenylylate [66]. Among the residues known to confer the nucleotide specificity of the GCs the glutamate residue is conserved (Fig. 3) while the other residue involved in specific nucleotide binding, Cys, is substituted by Thr which is found in equivalent positions in few the adenylylate cyclases. The nucleotide specificity of the GC-like domain of PKN2 sharing less than 15% identity with receptor GC is unclear. Trans membrane region of PKN2 follows the GC domain unlike eukaryotic transmembrane GCs.

A wide representation of eukaryotic signalling domains in the bacterial kingdom has been described in previous studies [67-69] and is also manifested in the bacterial Ser/



gi 17228044	(26)	IRI <b>GR</b> AADNHVILSDNLV----- <b>SRHH</b> LEIRQVSSGGGGS-WQVVSKG-TN	(69)
gi 23127167	(26)	IRI <b>GR</b> ATDNHVVLTDSLV----- <b>SRHH</b> LELRHVDSADNGSGWRLISQG-TN	(70)
gi 22299030	(26)	IRI <b>GR</b> AVDNHVVLYSAVV----- <b>SRHH</b> VELRR-----HGLQWEVVNLG-TN	(64)
gi 23041283	(26)	IKV <b>GR</b> STDNHIVLSDSLV----- <b>SRYH</b> LELHKISKSPSGNIWRLVSLG-TN	(70)
gi 22298671	(26)	IRI <b>GR</b> AADNDVILNDILV----- <b>SRYH</b> AELSCYRDPENLGRWFLKSLG-AM	(70)
gi 17232446	(27)	CIM <b>GR</b> SPEANIQLPDDAEHKT <b>IRYH</b> CLLD----IMPPN-IRIRDFGSKN	(71)
gi 23043171	(27)	CII <b>GR</b> HHCNCHIRLPNNEEHRT <b>IRYH</b> CILD----INYPY-VQIYDFGSKN	(71)
gi 23042923	(319)	LIL <b>GR</b> ---VDIAGADTTI----- <b>SRQH</b> LEFMR-----KENNYFVRDLGSKN	(356)
gi 17228044	(70)	GTFL <b>NG</b> VLVI-----QDALPNNALLQLA	(91)
gi 23127167	(71)	GTFL <b>NG</b> VLVI-----QSPLPDNSLLQLA	(92)
gi 22299030	(71)	GTYLD <b>G</b> KRIQ-----QATLTDGGILRLA	(92)
gi 23041283	(71)	GTFV <b>NG</b> VLMS-----QGIVSDGSLIQLA	(92)
gi 22298671	(72)	GTFV <b>DG</b> RLVN-----ESKLAGSGLIQLG	(93)
gi 17232446	(72)	GTYV <b>NG</b> QKIGQREAHQTPEEGAKINFPEYDLQSDDEIELG	(110)
gi 23043171	(72)	GTY <b>ING</b> KKIGKLQAKYKSEATAKMQFPEYDLKSGDKIQLG	(110)
gi 23042923	(357)	GTML <b>KG</b> KMIDN-----YPVEIKHGNMIKVG	(380)

**Figure 2**

Multiple sequence alignment of FHA domains seen in bacterial Ser/Thr kinases with the conserved motifs involved in phosphothreonine binding, 'GR', 'SXXH', 'NG' shaded in yellow. gi|17228044|ref|NP\_484592.1| (*Nostoc sp. PCC 7120*), gi|23127167|gb|ZP\_00109042.1| (*Nostoc punctiforme*), gi|22299030|ref|NP\_682277.1| (*Thermosynechococcus elongatus BP-1*), gi|23041283|gb|ZP\_00072750.1| (*Trichodesmium erythraeum IMS101*), gi|22298671|ref|NP\_681918.1| (*Thermosynechococcus elongatus BP-1*) gi|17232446|ref|NP\_488994.1| (*Nostoc sp. PCC 7120*) gi|23043171|gb|ZP\_00074493.1| (*Trichodesmium erythraeum IMS101*) gi|23042923|gb|ZP\_00074272.1| (*Trichodesmium erythraeum IMS101*).

Thr kinase family. Such domain recruitments across diverse signalling protein families imply the versatility in the nature of interactions of these domains.

**Dual protein kinases in bacteria**

It has been observed in the yeast, an unicellular eukaryote, that the osmo - regulation is brought about by both HisK and mitogen activated protein kinases (MAPKs) suggesting the signal transduction from HisK to the MAPKs to elicit the appropriate response [70,71]. The occurrences of such composite signalling pathways have been suggested in bacteria from other very early studies [72] and in studies on STYKs in *Anabaena sp. PCC 7120* [72]. The STYKs with HisK, HATPase (the catalytic domain of HisKs) and GAF (regulatory domains found in phosphodiesterases, adenylate cyclases and phytochromes) have been referred to as dual protein kinases (Fig 1d) as they contain catalytic domain of two distinct kinases fused to form a single gene product. Hybrid kinases such as those described above have so far been identified in cyanobacterial species [64,65,73,74]. The cyanobacterial species *Nostoc sp. PCC 7120* and *Nostoc punctiforme* have 12 and 11 dual protein kinases respectively. In the current study, the occurrences of such composite kinases have been noted in 5 different species. The other organisms that encode similar STYKs include *Trichodesmium erythraeum*, *Leptospira interrogans*,

*Ralstonia metallidurans* and *Rhodospseudomonas plustris* and *Pseudomonas putida*. The nature of signals these composite kinases perceive and the cellular processes they influence is intriguing. Some of these STYKs also have the PAS domains in the same polypeptide chain which is likely to serve as redox sensors and thereby suggesting the roles of such dual protein kinases in signalling pathways sensitive to redox potentials.

**Protein kinases involved in bacterial cell division**

Protein kinases encoded in the genomes of *Bacillus subtilis* and *Bacillus halodurans*, yabT (gi|2632333gnl|PIDe1181999) and BH0080 (gi|10172692dbj|BAB03799.1) respectively, are located in the same gene clusters as *spoIIE* and *FtsH/MesJ*. These STYKs are therefore suggested to be involved in the control of cell division like other genes in the same cluster. Among the many genes known so far to control the cell division in bacteria are the *spoIIE* and the *FtsH/MesJ* genes [75]. This group of STYK that are likely to be associated with cell division are analogous in their broader function to cyclin dependent kinases of the eukaryotes that control cell cycle.

P26769	(876)	LYHQSYDCVC	VMFASIPD--	-----	FKEFYTESDV	NKEGLECLRL	(913)
P25092	(815)	VEPELYEEVT	IYFSDIVG--	-----	FTTICKYSTP	----MEVVDM	(848)
P54736	(439)	LVQKRGDALL	VVFRSPTAGV	LCGMAMQDRL	WRHNQTVPEV	DRLNVRVCLH	(488)
P26769	(914)	LNEIIADFDD	LLSKPKFSGV	EKIKTIGST-	---YMAATGL	SAIP-SQQYM	(967)
P25092	(849)	LNDIYKSFDD	IVDH---HDV	YKVEITIGDA-	---YMVASGL	PKRNGNRHAI	(891)
P54736	(489)	AGEVLATPDS	VLGEP-MEVI	EAVEHVASAG	EVTFTTEAVNL	ARNRAEATAE	(537)
P26769	(968)	HIGTMVE---	-FAYALVGKL	DAINKHSFND	FKLRVGINH-	--GPVIAGVI	(1011)
P25092	(892)	DIAKMAL---	-EILSFMGTF	ELEHLPLPLP-	IWIRIGVHS-	--GPCAAGVV	(1033)
P54736	(538)	PCGAITLPGR	NEQLQLYRCQ	RAAEGPPFGD	RFASQGSRGV	ALAPLLAKLQ	(587)
P26769	(1012)	GAQKPQYDIW	GNTVNVASRM	DSTGVLDKIQ	VTEETSLLIQ	TLGYTCTCFV	(1061)
P25092	(1034)	GIKMPRYCLF	GDTVNTASRM	ESTGLPLRIH	VSGSTIAILK	RTECQFLYEV	(1083)
P54736	(588)	AVKLP--TGL	GELLRQRRE	AALVAGAVVL	LGAGAAWLSQ	RNDAGTRAFA	(635)
P26769	(1062)	N-----	-----	----			
P25092	(1084)	RGETYLKGRG	NETTYWLTG-	----	(1102)		
P54736	(636)	LLE---DGKL	NEALALMDAA	TDEE	(656)		

**Figure 3**  
Multiple sequence alignment of the human guanylate cyclase, rat adenylate cyclase and GC-like domain of PKN2 of *Myxococcus xanthus* is shown. The residues conferring nucleotide specificity are shaded yellow. Conserved hydrophobic and charged residues are shaded green and grey respectively.

**Bacterial protein kinases and their substrates**  
Despite the identification of bacterial STYKs by genome-wide analysis and demonstration of *in vitro* phosphorylation activity on exogenous substrates of significant number of STYKs, very limited information is available regarding their natural substrates and their specificity.

Putative homologues of AfsK, a well characterised Ser/Thr kinase *S.coelicolor*, has been identified in *Deinococcus radiodurans* and *Streptomyces avermitilis*. In *S. coelicolor*, AfsK phosphorylates a global regulator of secondary metabolism AfsR that also controls the production of Actinorhodin [17,76,77]. This is brought about by the Ser/Thr kinase AfsK. AfsK, encoded in the genomes of *Streptomyces coelicolor* and *S. lividans* is a membrane associated phosphokinase encoded by the *afsK* gene. The analysis of the domain composition of AfsK revealed the presence of nine bac\_PQQ repeats located at the C-terminal of the protein.

A search for the AfsR homologues in *Deinococcus radiodurans* that are likely to be the putative targets of AfsK revealed a gene product (gi|15807540) with N-terminal transcription regulator domain (*trans\_reg*) and Bacterial trans-activation domain (BTAD) similar to the domains observed in AfsR. This observation suggests that the gene product could probably be the target of the AfsK like

kinase (gi|6460339) of *D.radiodurans*. *Trans\_reg* or the BTAD domain conserved among all the transcription regulators of AFSR family are therefore likely to harbour sites for phosphorylation by AfsK.

The STYK sharing highest sequence similarity with AfsK in the catalytic domain (38% identity) in the *M.tuberculosis* genome is PKNB. A search for AfsR like gene products in the *M.tuberculosis* genome revealed the gene product of *embR* as its closest homologue. A further sequence analysis of *embR* gene product is known to be involved in the resistance to etambutol revealed the presence of BTAD, transactivation domain. This common domain shared by the AfsR and its homologue in *M.tuberculosis* and *D.radiodurans* further suggests that the BTAD, transactivation domain is critical for regulation of AfsR like gene products and as the most likely region of the phosphorylation mediated control of its activity. Recently *embR* gene product has been demonstrated to be phosphorylated by PKNH [78].

Substrate identification studies on PKN2 of *Myxococcus xanthus* have revealed histone like proteins (HU $\alpha$  and HU $\beta$ )[63] and beta-lactamase [79] as the target proteins phosphorylated by the recombinant PKN2 expressed in *E.coli*. The phosphorylation of beta-lactamase inhibits its secretion [79] while the phosphorylation of HU $\alpha$  and

HU $\beta$ + [63] has been shown to prevent their binding to DNA resulting in toxicity to cells. PKN4 has been further shown to phosphorylate phosphofructo kinase [80,81].

#### **Phosphorylation as a mode of regulation of bacterial STYKs**

Among the completed genomes analysed in this study, a significant number of STYKs are 'RD' protein kinases. The protein kinases in the eukaryotes are in turn regulated in a number of ways. Phosphorylation of Ser/Thr residues in the activation segment of the catalytic domains of these kinases is a common mode of regulation [82]. The PKs that undergo phosphorylation in the activation segment have a characteristic 'RD' doublet sequence motif in the catalytic loop wherein the Arg helps in neutralisation of the negative charge on the phosphorylated residue to suitably poise the catalytic aspartate for phosphorylation. These protein kinases are referred to as 'RD' kinases. A multiple sequence alignment of the putative activation segment in the catalytic domains of the bacterial 'RD' kinases have revealed the strong conservation of threonines in their activation segments (Fig: 4) These threonines are therefore suggested to be the sites of phosphorylation. Most of these kinases have two threonines that are well conserved and hence multiple phosphorylation for activation of these kinases as exemplified by the PKN-B of *M.tuberculosis* [83] is suggested.

#### **Discussion**

The genome-wide analyses reveal the occurrences of a large number of STYKs in bacteria. However the available knowledge about their biological functions is limited.

The STYKs show a lower representation compared to HKs in most of the genomes analysed (Data not shown). HKs are encoded by higher number of genes in comparison with STYK encoding genes in the prokaryotes. Nearly two percent of the genomes of eukaryotic organisms encodes STYKs as suggested from the genome-wide analysis by various groups [34,84-87]. Difference in the relative abundance of kinases belonging to the two classes, namely HKs and STYKs in the eukaryotes and the prokaryotes probably arises from the selection of specific class of PKs during the course of evolution, as an adaptation to distinct cellular environment. Low occurrence of STYKs compared to HKs in most bacterial species with the exception of *Pirellula* (*Rhodopirellula baltica* SH1) suggests, in general, the predominant role of HKs in prokaryotes.

HisK in archaeal genomes have been suggested to have arisen from lateral gene transfer events explaining their limited representation [88,89]. With the exception of *Buchnera*, *Onion phytoplasma* and *Thermococcus kodakaraensis*, all the above mentioned genomes have at least one of

the two kinds of protein kinases, namely HisK and STYK. STYKs have not been identified in genomes of the two spirochaetes, *Treponema pallidum* and *Borrelia burgdorferi*, *Chlorobium tepidum*, *Helicobacter hepaticus*, *Thermotoga maritima* and *Xanthomonas species*. The absence of either classes of protein kinases in *Buchnera* species, an obligate symbiont of the Aphids and *Onion phytoplasma* (intra-cellular plant pathogen) suggests the loss of such enzymes in the above species is likely to be an adaptation to the host-intracellular environment. Our analysis also identifies an anaerobic obligate heterotrophic thermophile, *T.kodakaraensis* as the only species with a non-symbiotic or non-parasitic habitat that lacks HisK and STYKs. Further studies on influence of phosphorylation on response to external stimuli will therefore be able to provide insight into the nature of enzymes and significance of phosphorylation in these organisms.

The number of STYKs in phylogenetic domains of archaea and bacteria increase with the growing size of bacterial genomic data. The assignment of probable biochemical functions and biological roles to large number of sequences are often dependent on the occurrence of homologous gene products in other phylogenetic domains despite the vast divergence between the species. Earlier studies have also indicated that the gross biological function and interaction network of multi-domain proteins are conserved by the preservation of domain composition and arrangement [90-92]. Therefore domain organisation of the bacterial STYKs have been studied here to get an insight into their putative biological roles in comparison with other proteins of similar domain organisation.

Enzymatic and non-enzymatic domains described in the earlier sections have been found covalently tethered to the kinase catalytic domain. Among the archaeal multi-domain STYKs, the OSGP-containing STYKs are widely represented. Knowledge of substrates of the OSGP-containing STYKs and their sites of phosphorylation is required to understand their role in regulation of functions of glycoproteins. The Ser/Thr residues serve as sites for both glycosylation and the phosphorylation [93,94]. In many instances the phosphorylation and glycosylation occurs at common sites in a competitive manner. The phosphorylation would therefore prevent the glycosylation of the proteins. OSGP domains are involved in the cleavage of those proteins at the sites containing glycosylated Ser/Thr residues. The phosphorylation of the target proteins by the associated kinase domains of the OSGP containing STYKs might therefore serve as a regulatory mechanism to inhibit the proteolysis of glycoproteins.

gi 3261694	(167)	DFGIARIAG-----GFETATGVIAGSPAFTAPE	(194)
gi 8978468	(142)	DWGAAVACGEEEDLLDIDVSKEEVLSSRMTIPGRIVGTPDYMAPE	(186)
PKNB	(156)	DFGIARAIADSGN-----SVTQTAAVIGTAQYLSPE	(187)
gi 13092426	(156)	DFGIARAIADS-T-----SVTQTAAVIGTAQYLSPE	(185)
gi 2633949	(152)	DFGIATALSST-----TITHTNSVLGSHVHLSPE	(180)
gi 10175124	(151)	DFGIARAMSSA-----TITHTNSVMGSHVHLSPE	(179)
gi 12724921	(154)	DFGIAKALSET-----SLTQTNTMFGSVHLSPE	(182)
gi 13622698	(154)	DFGIAVAFSAET-----SLTQTNSMLGSHVHLSPE	(182)
gi 13701020	(151)	DFGIAKALSET-----SLTQTNHVLGTVQYFSPE	(179)
gi 13092427	(159)	DFGIAKAVDAA-----PVTQTGMVMGTAQYIAPPE	(187)
gi 1552570	(159)	DFGIAKAVDAA-----PVTQTGMVMGTAQYIAPPE	(187)
gi 13092968	(147)	DFGLVRAVAAA-----GITSTSVILGTVAYLSPE	(175)
gi 2911096	(160)	DFGLVRAVAAA-----SITSTGVILGTAAYLSPE	(188)
gi 1370255	(158)	DFGIARALGD--T-----GLTSTGSLVLAATLAAAPPE	(186)
gi 2131011	(159)	DFGIAGWVDDP--S-----GLTATNMTVGTVSYAAPPE	(188)
gi 3261596	(157)	DFGIASATTD--E-----KLTQLGTAVGTWKYMAPE	(185)
gi 2131007	(157)	DFGIASATTD--E-----KLTQLGNTVGTLYYMAPE	(185)
gi 2078052	(156)	DFGIARAASD--P-----GLTQTGTAVGTYNYMAPE	(184)
pkn2	(156)	DFGIARLVEPDAAS-----SVSQIGVVLGTPPEYLSPE	(188)
gi 6460339	(154)	DFGLVALSEQTRH-----LTRSGVTLGTPAYMAPE	(184)
gi 6457717	(159)	DFGIAVGKVLTRL-----TMPGARVGTPIYMAPE	(188)
gi 13472161	(170)	DFGIARSAAVG--G-----ETLIGGKFAFKYNYVSPPE	(199)
gi 1006577	(155)	DFGAVKETMGAVTLGSG-----STVSS---VVIGTRGFMAPPE	(188)
gi 1653955	(161)	DFGIKETMGTLVNPDG-----RSAYS---VALGTAGYMASE	(194)
gi 1653478	(182)	DFGAVKNQIDSVLSSNT-----SAQTALTAFAVGTGAFAPPE	(218)
gi 1652588	(207)	DFGAVKQATAGVGASNE-----GST-----GIYSMGFAPPE	(235)
gi 3844698	(149)	DFGIASTVVKVAEKT-----EVLTDENSLFGTVSYMIPD	(191)
gi 1674285	(183)	DFGIASAVIRNAEKT-----EVLTDENSLFGTVSYMTPPE	(216)
gi 6460744	(207)	DFGSVRAFDSD-----KTVAHTRLVTPGYAPPE	(234)
gi 9947763	(376)	DFGLAWCPGLS-----REDPHLLPGTPSYLAPE	(403)
gi 6459630	(148)	DYDLSGPLGED-----PGGAVRIGTPAFQSPPE	(174)
gi 13701020	(151)	DLGLGKNLDEIHS-----HQTFTDNGVGOYKYCAPPE	(179)
gi 6899183	(169)	DYGAASVFFYNRED-----LTKDQEIKCTIIYASPK	(198)

**Figure 4**

Multiple sequence alignment of the activation segment of 'RD' protein kinases in bacteria with the canonical 'DFG', 'APE' shaded in blue and conserved threonines, the potential autophosphorylation sites shaded yellow. gi2911096, gi1370255, gi2131011, gi3261596, gi2131007, gi2078052, (M.tuberculosis); gi13092426 gi13092427 gi13092968 (M.leprae); gi8978468 (C.pneumoniae); gi2633949 (B.subtilis); gi10175124 (B.halodurans); gi12724921 (Lactococcus lactis); gi13622698 (S.pyogenes); gi13701020 gi13701020 (S.aureus); gi6460339 gi6457717 gi6460744 gi6459630 (D.radiodurans); gi13472161 (M.meliloti); gi1652313, gi1006577 gi1653955 gi1653478 gi1652588 (Synechocystis.sp); gi3844698 (M.genitalium); gi1674285 (M.pneumoniae); gi9947763 (P.aeruginosa); gi6899183 (U.ureolyticum).

Eubacterial STYKs are associated with a large variety of signalling domains. Chase 2 domains containing STYKs identified by recent studies [64,65,73] are also found in other classes of trans membrane receptors including histidine kinases, diguanylate cyclases and methyl accepting proteins. Presence of widely occurring sensory domain in Ser/Thr kinases of Nostoc further emphasises on its role in the perception of some critical signals that are likely to trigger cellular pathways synergistically.

STYK of P.aeruginosa with phosphatase like domain has significant similarity with spoIIE protein family in the PFAM database. Recent studies [95,96] have shown the involvement of Ser/Thr kinase-Phosphatase pair, PrkC and PrpC in the sporulation of B.subtilis. Psuedomonas aeruginosa however encodes a composite kinase-phosphatase in the same gene. The phosphatase-2C (PP2C) domains are known to selectively dephosphorylate protein phosphorylated on Ser/Thr residues and their homologues have been previously identified and characterised in various bacteria [97-101]. The PP2C domains have been pre-

dicted to be functionally active in SpoIIE [102,103] of *Bacillus subtilis*, a Gram positive bacterial species. SpoIIE controls the activation of sporulation transcription factors and critical for the development of the bacteria.

The bacterial STYKs with TPR repeats identified here are suggested to be influenced by the chaperone like protein during their trafficking, folding or interactions with substrate proteins. The involvement of HSP90 in hormonal and growth factor signal transduction mediated by protein kinases through proteins containing TPR repeats is well characterised in eukaryotic cells [42]. Further phosphatase pp5 of *Plasmodium falciparum* with TPR domain is known to act as co-chaperone for hsp90 and it has been therefore suggested that protein dephosphorylation might play a regulatory role in protein folding [104,105]. Despite their high sequence divergence the TPR repeats and their constituting domain form a general interface of interactions with HSP90 like chaperones [104].

STYKs with stress response domains such as USPA have been identified. The USPA proteins are suggested to be autophosphorylated to switch on to active state. Subsequent to the induction of USPA proteins, many phosphorylated proteins have been detected [106]. The two-component system involved in nitrate assimilation is known to induce bacterial universal stress proteins of the nucleotide binding class [44]. Majority of these proteins are tightly regulated and expressed under various conditions of stress in the bacteria.

The USPA containing kinases of plants have been implicated in disease resistance signalling pathway based on their similarity to other protein kinases like Pto and Pto interacting proteins of tomatoes that are involved in disease resistance [43,107]. Based on the role of USPA containing proteins in stress response and their regulation by the some two component system, we suggest the bacterial STYKs with USPA domains are likely to be induced during stress response. In the eukaryotes, MAP kinases are also known to trigger stress induced signalling pathways. This group of bacterial kinases could therefore be functionally analogous to their eukaryotic counterparts that are induced during stress.

Protein kinase containing gene products with C-terminal pentapeptide repeats have been identified initially by an early study [73] in *Nostoc sp. PCC 7120* (gi|17132363) and *Synechocystis sp. PCC 7803* (gi|1653478). Nitrate assimilation and nitrogen fixation are well characterised in various species of Cyanobacteria. The key enzyme involved in this pathway, the nitrogenase, is highly sensitive to oxygen. Some multi-cellular cyanobacterial species like *Nostoc sp.* form specialised cells called heterocysts wherein the glycolipid layer prevents the diffusion of oxygen [77]. The

accumulation of glycolipids into the heterocysts is brought about by a number of proteins. Some of these proteins have the pentapeptide repeats, which are known to be critical for the localisation of glycolipids [108].

PASTA domains conserved in a group of STYKs encoded by Gram positive bacteria are found in proteins comprising the divisome complex. The representation of this group of STYKs in all gram positive bacteria, including the *Lactococcus lactis* (gi|12724921), and *Streptococcus pyogenes* (gi|13622698), which contain only one STYK suggest their role to be critical in their cell division. The role of these protein kinases in cell division is reiterated with their genes clustering with *rodA* involved in the regulation of cell morphology during cell division. The phosphatase 2C encoding genes also lie in the same cluster supporting the role of reversible phosphorylation during cell division.

Lipopolysaccharide kinases (LPSK) engaged in the phosphorylation of the lipopolysaccharides in the outer membrane of Gram-negative bacteria share remote similarity with eukaryotic STYKs [32]. They could be grouped into two classes based on their sites of phosphorylation into *waaP* gene products and Kdo kinases (KdoK) [109,110]. A remote similarity between STYKs and the two family of LPSKs has been recently reported by our group [32]. The similarity in the potential 3-D structure and catalytic residues of LPSKs with the eukaryotic protein kinases suggests analogous catalytic functions to the residues that are well conserved across the two distinct families of enzymes. The occurrence of Kdo kinase is so far restricted to pathogenic Gram-negative species and is also known to be influencing the virulence in *H.influenzae* [109]. Inferences drawn on the structure and the catalytic mechanism of LPSK based on their similarity to eukaryotic protein kinases can therefore serve as a guide to the design of inhibitors to the LPSKs in virulent organisms.

The genomic organisation of all STYKs encoded in *M.tuberculosis* H37Rv genome has been analysed in an earlier study [111]. The gene encoding the protein kinase PKN-G of *M.tuberculosis* is clustered with the genes, *glnH* for periplasmic Gln binding proteins. Therefore PKN-G is likely to be the functional homologue of the other Gln binding periplasmic Ser/Thr protein kinase of *Streptomyces coelicolor* (gi|21223285). These protein kinases therefore suggests the intake of amino acids in to the cell is also likely to be influenced by phosphorylation, in addition to the sugar transport in *M.tuberculosis* as described previously. Recent studies have established the involvement of PKN-G in the regulation of cellular levels of Gln/Glu [112,113] and in evading phagocytosis by the host cells.

Eukaryotic signalling domains have been observed in a large number of bacterial proteins including in the STYKs

[67-69]. The cyanobacterial STYK with a SH3b domain has a transmembrane segment between the kinase domain and the SH3 domain. The nature of signals perceived by this trans-membrane STYK is not clear. These SH3 domains are distantly related to their eukaryotic counterparts. Eukaryotic Ser/Thr kinases with SH3 domains have been well studied which is exemplified by mitogen activated protein kinase kinase [114,115].

WD-repeats, a versatile protein-protein interaction module has also been observed bacterial protein kinases [64,73,116,117] and in human protein kinases [46]. Pkwa of *Themospora curvata*, a member of WD-repeat kinase family has been shown to be phosphorylated by exogenous protein kinase from *S. granticolor* [100].

The fork head associated (FHA) domain was initially found in transcription factors and has been recently identified as protein – protein interaction modules known to bind to phospho-threonine residues on target proteins. Hence the interactions of the FHA domains with their cognate proteins are largely influenced by phosphorylation. Many bacterial proteins have been recently identified with FHA domains [117]. Previously it has been observed that the genes encoding FHA containing proteins cluster along with Ser/Thr protein kinases and phosphatases in the *M. tuberculosis* genome [111]. Thus the identification of the PKs with FHA domains in cyanobacteria suggests synergistic actions of FHA domains containing proteins and enzymes involved in reversible phosphorylation as observed in eukaryotes.

Very few studies have shed light on the nature of substrates of bacterial kinase [17]. The AfsR, the substrate of AfsK has an N-terminal transcription regulator domain (trans\_reg domain), trans-activation domain (bacterial trans-activation domain, BTAD) followed by a signalling domain NB-ARC and a C-terminal tetratricopeptide repeat region which is known to be involved in protein-protein interactions. The BTAD domain is shared by all the members DNRI/REDD/AFSR family of transcription regulators involved in secondary metabolism. Other proteins in *Streptomyces coelicolor* and related species that control actinorhodin production share the trans\_reg domain and the BTAD domain. A similar domain organisation among the homologues of AFSR is intriguing and is likely to contain sites of phosphorylation.

The phosphorylation site of HU $\alpha$  has been identified as a threonine residue which is well conserved across HU $\alpha$  of various species. RXXRTGR is the sequence pattern conserved at the phosphorylation site (P-site). The occurrence of R at the P-1 site suggests PKN2 to be a 'basic directed' protein kinase like Protein Kinase A (PKA) that specifically phosphorylates substrates with consensus motif

RRXS/TZ' (X is any amino acid and Z is any hydrophobic amino acid). Based on the residue preferences of PKN2, at known phosphorylation sites on substrates, bacterial STYKs sharing significant similarity with PKN2 have been examined, that are likely to seek similar residue determinants on their substrates. The catalytic domain of PKN2 shares highest similarity (40% identity) with PKNB of *M.tuberculosis* and its homologue in *M.leprae*. The high sequence similarity between PKN2 and PKNB is likely to reflect in similar substrate binding modes such as preferences for N-terminal basic amino acid residues as also observed for PKA.

PKN2 of *Myxococcus xanthus* is known to autophosphorylate in the region following the kinase domain. The GC domain that lies C-terminal to the kinase domain contains 10 threonines. Further experimental studies are therefore necessary to ascertain if these sites could be the natural phosphorylation sites targeted by their respective enzymes.

A significant number of bacterial STYKs have been biochemically characterised previously to elucidate their biological roles. STYKs of the PKN subfamily have been identified and well characterised in *Myxococcus xanthus* [79-81], *Streptomyces coelicolor* [17], *Streptomyces granticolor* [86], *Synechocystis sp* [118-120], *Yersinia tuberculosis* [33], *Mycobacterium tuberculosis* [111,122] and *Pseudomonas aeruginosa* [55]. The PKN representatives of *Myxococcus xanthus* and *Streptomyces* species are involved in the specialised developmental cycles characteristic to these species. They are also shown to be capable of autophosphorylation although it has been suggested to be unimportant for their kinase activity *in vitro* [17,79-82].

The YpkA of *Yersinia pseudotuberculosis* is a secreted STYK that has been previously shown to interact with small GTPases like RhoA and Rac1 of the host [26]. We have identified the homologue of YpkA, the only STYK in the completely sequenced genome of *Yersinia pestis* CO92 sharing 99% sequence identity with YpkA. However the genome of *Yersinia pestis* KIM does not encode any YpkA homologues. In addition to ABC1 and RIO1 class of STYKs, a single gene product containing only the kinase domain is identified.

PpkA of *Pseudomonas aeruginosa* have been implicated in their virulence and are suggested to be induced during the process of infection [55]. An unusual domain arrangement of PpkA identified in the current analysis comprises of an N-terminal kinase domain followed by a stretch of 300 residues of unknown function and a C-terminal von Willebrand factor Type A (vWA) domain. The vWA domain, known to bind to glycoproteins and growth fac-

**Table 1: Prokaryotes of complete genomic data analysed and the distribution of encoded STYKs and homologues. The list includes ABCI, RIO1 and kinases that share significant sequence similarity with STYK family. Number in parenthesis indicates the sequences lacking catalytic Asp and hence are unlikely to function as a kinase.**

Organism	STYK	ABCI	RIO1
<b>ARCHAEA</b>			
<i>Aeropyrum pernix</i>	7	-	2
<i>Archaeoglobus fulgidus</i>	2	-	2
<i>Halobacterium salinarium</i>	5 (1)	1	2
<i>Methanococcus janaschii</i>	4	-	2
<i>Methanopyrus kandleri</i>	1	-	1
<i>Methanosarcina acetivorans</i>	5	2	2
<i>Methanosarcina mazei Goel</i>	4	1	2
<i>Methanothermobacter thermoautotrophius</i>	3	1	1
<i>Nanoarchaeum equitans</i>	3	-	2
<i>Pyrobaculum aerophilum</i>	4	-	2
<i>Pyrococcus abyssi</i>	4	-	2
<i>Pyrococcus horikoshii</i>	4	-	2
<i>Sulfolobus solfataricus</i>	7	1	2
<i>Sulfolobus tokodaii</i>	11	1	2
<i>Thermoplasma acidophilum</i>	3	1	1
<i>Thermoplasma volcanium</i>	4	1	2
<b>BACTERIA</b>			
<i>Agrobacterium tumefaciens</i> str. C58	2	1	-
<i>Aquifex aeolicus</i>	2	-	1
<i>Bacillus anthracis</i> str. Ames	3	-	-
<i>Bacillus cereus</i> ATCC 14579	6	2	-
<i>Bacillus halodurans</i>	4	1	-
<i>Bacillus subtilis</i>	3	-	-
<i>Bacteroides thetaiotaomicron</i> VPI-5482	2	-	-
<i>Bifidobacterium longum</i> NCC2705	7	1	-
<i>Bordetella bronchiseptica</i>	2	1	-
<i>Bordetella parapertussis</i>	2	1	-
<i>Bordetella pertussis</i>	2	1	-
<i>Borrelia burgdorferi</i>	-	-	-
<i>Bradyrhizobium japonicum</i>	5	1	1
<i>Brucella melitensis</i>	-	-	-
<i>Brucella suis</i>	2	2	-
<i>Buchnera aphidicola</i>	-	-	-
<i>Buchnera</i> sp. APS	-	-	-
<i>Candidatus Blochmannia floridanus</i>	1	1	-
<i>Caulobacter crescentus</i>	1	1	-
<i>Chlamydia muridarum</i>	1 (1)	-	-
<i>Chlamydia pneumoniae</i> J138	3 (1)	-	-
<i>Chlamydia trachomatis</i>	3 (1)	-	-
<i>Chlamydia caviae</i> GPIC	3 (1)	-	-
<i>Chlamydia pneumoniae</i> AR39	3 (1)	-	-
<i>Chlorobium tepidum</i> TLS	-	-	-
<i>Chromobacterium violaceum</i> ATCC 12472	3	2	-
<i>Clostridium acetobutylicum</i>	3	1	-
<i>Clostridium perfringens</i>	2	1	-
<i>Clostridium tetani</i> E88	1	-	-
<i>Corynebacterium diphtheriae</i>	4	-	-
<i>Corynebacterium efficiens</i> YS-314	4	-	-
<i>Corynebacterium glutamicum</i> ATCC 13032	4	-	-
<i>Coxiella burnetii</i> RSA 493	4	1	-
<i>Deinococcus radiodurans</i>	10	-	1
<i>Enterococcus faecalis</i> V583	2	-	-
<i>Escherichia coli</i> K12	4	1	-
<i>Escherichia coli</i> O157:H7 EDL933	4	1	-

**Table 1: Prokaryotes of complete genomic data analysed and the distribution of encoded STYKs and homologues. The list includes ABCI, RIO1 and kinases that share significant sequence similarity with STYK family. Number in parenthesis indicates the sequences lacking catalytic Asp and hence are unlikely to function as a kinase. (Continued)**

<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i> ATCC 25586	1	-	-
<i>Geobacter sulfurreducens</i> PCA	1	-	-
<i>Gloeobacter violaceus</i> [B]4.7NC_005125	19 (1)	4	-
<i>Haemophilus ducreyi</i>	2	1	-
<i>Haemophilus influenzae</i>	1	-	-
<i>Helicobacter hepaticus</i> ATCC 51449	-	-	-
<i>Helicobacter pylori</i> J99	1	-	-
<i>Lactobacillus plantarum</i> WCFS1	3	1	-
<i>Lactococcus lactis</i>	1	-	-
<i>Leptospira interrogans</i>	5 (2)	2	-
<i>Mesorhizobium meliloti</i>	3	1	-
<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i>	12	2	-
<i>Mycobacterium bovis</i>	10	2	-
<i>Mycobacterium leprae</i>	6	2	-
<i>Mycobacterium tuberculosis</i>	13	2	-
<i>Mycoplasma gallisepticum</i>	1	-	-
<i>Mycoplasma genitalium</i>	1	-	-
<i>Mycoplasma penetrans</i>	1	-	-
<i>Mycoplasma pneumoniae</i>	1	-	-
<i>Mycoplasma pulmonis</i>	1	-	-
<i>Neisseria meningitidis</i>	1	1	-
<i>Nitrosomonas europaea</i>	2	2	-
<i>Nostoc</i> sp. PCC 7120	53	4	-
<i>Oceanobacillus iheyensis</i>	1	-	-
<i>Onion yellows phytoplasma</i>	-	-	-
<i>Pasteurella multocida</i>	2	1	-
<i>Photobacterium luminescens</i> subsp. <i>laumondii</i>	1	1	-
<i>Pirellula</i> sp.	59 (1)	2	-
<i>Porphyromonas gingivalis</i> W83	1	-	-
<i>Prochlorococcus marinus</i> MED4	3	3	-
<i>Prochlorococcus marinus</i> MIT 9313	4	3	-
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str.	3	3	-
<i>Pseudomonas aeruginosa</i>	11	1	1
<i>Pseudomonas putida</i> KT2440	14 (2)	-	2
<i>Pseudomonas syringae</i>	22	2	2
<i>Ralstonia solanacearum</i>	3	1	-
<i>Rhodospseudomonas palustris</i>	5 (3)	1	-
<i>Rickettsia conorii</i>	1	1	-
<i>Rickettsia prowazekii</i>	1	1	-
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi Ty2	4	1	-
<i>Salmonella typhimurium</i> LT2	2	1	-
<i>Shewanella oneidensis</i> MR-1	8 (2)	1	1
<i>Shigella flexneri</i> 2a str. 2457T	3	1	-
<i>Sinorhizobium meliloti</i>	1	1	-
<i>Staphylococcus aureus</i>	2	-	-
<i>Staphylococcus epidermidis</i> ATCC 12228	1	-	-
<i>Streptococcus mutans</i> UA159	1	-	-
<i>Streptococcus pyogenes</i> M1 GAS	1	-	-
<i>Streptococcus pyogenes</i> SSI-1	1	-	-
<i>Streptomyces avermitilis</i>	34 (2)	1	-
<i>Streptomyces coelicolor</i> A3	18	-	-
<i>Synechococcus</i> sp. WH 8102	3	3	-
<i>Synechocystis</i>	12 (1)	5	-
<i>Thermoanaerobacter tengcongensis</i>	1	-	-
<i>Thermosynechococcus elongatus</i>	14	-	-

**Table 1: Prokaryotes of complete genomic data analysed and the distribution of encoded STYKs and homologues. The list includes ABCI, RIO1 and kinases that share significant sequence similarity with STYK family. Number in parenthesis indicates the sequences lacking catalytic Asp and hence are unlikely to function as a kinase. (Continued)**

<i>Thermotoga maritima</i>	1 (1)		-
<i>Treponema pallidum</i>	-	-	-
<i>Tropheryma whippelii</i> str Twist	4 (2)	-	-
<i>Ureaplasma ureolyticum</i>	1	-	-
<i>Vibrio cholerae</i>	2		-
<i>Vibrio parahaemolyticus</i> RIMD 2210633	6		-
<i>Vibrio vulnificus</i> CMCP6	5		-
<i>Wigglesworthia brevipalpis</i>	1		-
<i>Wolinella succinogenes</i>	1		-
<i>Xanthomonas axonopodis</i> pv. citri str. 306	5	-	-
<i>Xanthomonas campestris</i> pv. campestris str. ATCC 33913	1	-	-
<i>Xylella fastidiosa</i>	3		
<i>Xylella fastidiosa</i> Temecula1	3		
<i>Yersinia pestis</i>	4		

tor receptors, in PpkA is therefore suggested to aid in interactions at the surface of the host cell.

Mbk, is a Ser/Thr kinase from *Mycobacterium tuberculosis* encoded by the gene mbk [121] that lies in pst operon that controls phosphate transport. MbK corresponds to STYK gi|2078052 in *M. tuberculosis* genome, which has an N-terminal kinase domain, a transmembrane region and a C-terminal NHL-repeats. The occurrence of membrane spanning region and extra-cellular NHL-repeats that forms an interaction surface suggests the MbK to serve as receptors for unknown ligands regulating the phosphate transport.

The STYK, referred to as SpkA in *Synechocystis* is previously shown to be required for the motility of the cyanobacteria. SpkA corresponds to STYK (gi|1652312), in the genome of *Synechocystis* sp that has an N-terminal kinase domain and a C-terminal extension of 200 residues to which no known functional domain could be associated. The role of the C-terminal extension in the regulation or interaction with other proteins controlling the motility of the organism remains to be elucidated.

The attempt made in the current analysis to study the bacterial STYKs by *In silico* analysis is hoped to aid in the further understanding of STYKs encoded in various genomes. The diversity of the functional domains occurring in these protein kinase like gene products provide clues to their biological functions that are yet to be explored experimentally. The involvement of bacterial STYKs in various cellular processes described in the previous sections suggests a very little overlap with the Histidine kinases and the protein kinases of the eukaryotes.

Large-scale genomic analysis are fraught with uncertainties related to the inheritance of function among remotely homologous proteins. Nevertheless search procedures employed in the current study for the study of domain organisation of gene products has been extensively benchmarked [122] An accuracy level of 96% is achieved by use of stringent search parameters described in 'Materials and Methods' section. The functional roles of STYKs and related systems encoded in various prokaryotic organisms deduced using computational approaches in the present analysis and in a recent independent analysis [123] is hoped to provide useful leads for further studies on their regulation.

Experimental studies [eg., [124]] elucidating the mechanisms of action STYKs and their homologues in various cellular processes as suggested from the analysis would therefore have implications in understanding their influence on the growth and development of the bacteria.

## Methods

The complete set of predicted protein sequences from the ORFs of the bacterial genomes has been obtained from NCBI [125] Using sensitive sequence profile matching algorithms, STYK like sequences have been identified in the genomes.

We have employed multiple sensitive sequence search and analysis methods PSI-BLAST [126], IMPALA [127] and HMMer which matches Hidden Markov Models (HMMs) [128]. These programs have been previously benchmarked [122,129,130] and we have used a stringent cutoff for e-values (0.0005 in PSI-BLAST,  $10^{-5}$  in IMPALA and 0.01 for HMMer) for identifying homologues of kinases. The list of predicted STYKs used for further analysis has been arrived at after careful cross-referencing between the results of these methods as well as manual scrutiny for a variety of factors such as length of the kinase domains and presence of critical functional residues. Kinase catalytic domains of the PKLA were aligned using CLUSTALW. Manual decisions are taken considering the presence/absence of functional motifs and the lengths of putative kinase domains. Those kinase-like sequences without functional residues, such as an aspartate in the catalytic base position, are segregated separately and are not considered for the detailed analysis. Number of such sequences in various genomes is provided in Table 2.

Domain assignment to the non-catalytic regions of the kinase containing genes has been carried out using the HMM search methods by querying each of the kinase containing sequences against the 6190 protein family HMMs available in the Pfam database [131]. Trans-membrane segments were detected using TMHMM [132].



The 3-dimensional structures were superimposed using STAMP [133].

### Authors' contributions

AK performed the computational sequence analysis and modelling. AK and NS conceived the study and participated in its design and coordination. Both the authors read and approved the final manuscript.

### Additional material

#### Additional File 1

Data files comprising of the description of protein kinases and homologues encoded in genomes of organisms considered in the current analysis are provided as supplementary information accompanying this article. Each additional data file lists the gene identifiers, length, and domain arrangement of protein kinases and homologues identified in the current analysis.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-129-S1.tar>]

### Acknowledgements

Authors thank Dr. Galperin and anonymous referees for their critical comments and suggestions. A.K. is supported by a fellowship from the Council of Scientific and Industrial Research, India and Wellcome Trust, UK. This research is supported by International Senior Fellowship in biomedical sciences to N.S. by the Wellcome Trust, UK, National Bioscience award by the Department of Biotechnology, India and by the computational genomics project supported by the Department of Biotechnology, India.

### References

- Hurley JH, Dean AM, Thorsness PE, Koshland DE Jr, Stroud RM: **Regulation of isocitrate dehydrogenase by phosphorylation involves no long-range conformational change in the free enzyme.** *J Biol Chem* 1990, **265**:3599-3602.
- Kim DY, Stauffacher CV, Rodwell VW: **Engineering of Sulfolobus solfataricus HMG-CoA reductase to a form whose activity is regulated by phosphorylation and dephosphorylation.** *Biochemistry* 2000, **39**:2269-2275.
- Hunter T: **Protein kinase classification.** *Methods Enzymol* 1991, **200**:3-37.
- Duclos B, Marcandier S, Cozzzone AJ: **Chemical properties and separation of phosphoamino acids by thin-layer chromatography and/or electrophoresis.** *Methods Enzymol* 1991, **201**:10-21.
- Saier MH Jr, Wu LF, Reizer J: **Regulation of bacterial physiological processes by three types of protein phosphorylating systems.** *Trends Biochem Sci* 1990, **15**:391-395.
- Lukat GS, McCleary WVR, Stock AM, Stock JB: **Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors.** *Proc Natl Acad Sci U S A* 1992, **89**:718-722.
- Skorko R: **Polyphosphate as a source of phosphoryl group in protein modification in the archaeobacterium Sulfolobus acidocaldarius.** *Biochimie* 1989, **71**:1089-1093.
- Dutta R, Qin L, Inouye M: **Histidine kinases: diversity of domain organization.** *Mol Microbiol* 1999, **34**:633-640.
- Thomason P, Kay R: **Eukaryotic signal transduction via histidine-aspartate phosphorylation.** *J Cell Sci* 2000, **113**:3141-150.
- Wolanin PM, Thomason PA, Stock JB: **Histidine protein kinases: key signal transducers outside the animal kingdom.** *Genome Biol* 2002, **3**(10):3013.
- Zheng J, Knighton DR, ten Eyck LF, Karlsson R, Xuong N, Taylor SS, Sowadski JM: **Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor.** *Biochemistry* 1993, **32**:2154-2161.
- Ryazanov AG, Ward MD, Mendola CE, Pavur KS, Dorovkov MV, Wiedmann M, Erdjument-Bromage H, Tempst P, Parmer TG, Probst CR, Germino FJ, Hait WN: **Identification of a new class of protein kinases represented by eukaryotic elongation factor-2 kinase.** *Proc Natl Acad Sci U S A* 1997, **94**:4884-4889.
- Thomson M: **Evidence of undiscovered cell regulatory mechanisms: phosphoproteins and protein kinases in mitochondria.** *Cell Mol Life Sci* 2002, **59**:213-219.
- Cozzzone AJ: **Protein phosphorylation in prokaryotes.** *Annu Rev Microbiol* 1988, **42**:97-125.
- Skorko R: **Protein phosphorylation in the archaeobacterium Sulfolobus acidocaldarius.** *Eur J Biochem* 1984, **145**:617-622.
- Munoz-Dorado J, Inouye S, Inouye M: **A gene encoding a protein serine/threonine kinase is required for normal development of M. xanthus, a gram-negative bacterium.** *Cell* 1991, **67**:995-1006.
- Matsumoto A, Hong SK, Ishizuka H, Horinouchi S, Beppu T: **Phosphorylation of the AfsR protein involved in secondary metabolism in Streptomyces species by a eukaryotic-type protein kinase.** *Gene* 1994, **146**:47-56.
- Shi L, Potts M, Kennelly PJ: **The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait.** *FEMS Microbiol Rev* 1998, **22**:229-253.
- Leonard CJ, Aravind L, Koonin EV: **Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily.** *Genome Res* 1998, **8**:1038-1047.
- Han G, Zheng CC: **On the origin of Ser/Thr kinases in a prokaryote.** *FEMS Microbiol Lett* 2001, **200**:79-84.
- Kennelly PJ: **Protein kinases and protein phosphatases in prokaryotes: a genomic perspective.** *FEMS Microbiol Lett* 2002, **206**:1-8.
- Kennelly PJ: **Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry.** *Biochem J* 2003, **370**:373-389.
- Nadvornik R, Vomastek T, Janecek J, Technikova Z, Branny P: **Pkg2, a novel transmembrane protein Ser/Thr kinase of Streptomyces graticolor.** *J Bacteriol* 1999, **181**:15-23.
- Udo H, Inouye M, Inouye S: **Effects of overexpression of Pkn2, a transmembrane protein serine/threonine kinase, on development of Myxococcus xanthus.** *J Bacteriol* 1996, **178**:6647-6649.
- Barz C, Abahji TN, Trulzsch K, Heesemann J: **The Yersinia Ser/Thr protein kinase YpkA/YopO directly interacts with the small GTPases RhoA and Rac-1.** *FEBS Lett* 2000, **482**:139-143.
- Hakansson EE, Forsberg SA, Wolf-Watz H: **A secreted protein kinase of Yersinia pseudotuberculosis is an indispensable virulence determinant.** *Nature* 1993, **361**:730-732.
- Ostrovsky PC, Maloy S: **Protein phosphorylation on serine, threonine, and tyrosine residues modulates membrane-protein interactions and transcriptional regulation in Salmonella typhimurium.** *Genes Dev* 1995, **9**:2034-2041.
- Ortiz-Lombardia M, Pompeo F, Boitel B, Alzari PM: **Crystal structure of the catalytic domain of the PknB serine/threonine kinase from Mycobacterium tuberculosis.** *J Biol Chem* 2003, **278**:13094-13100.
- Young TA, Delagoutte B, Endrizzi JA, Falick AM, Alber T: **Structure of Mycobacterium tuberculosis PknB supports a universal activation mechanism for Ser/Thr protein kinases.** *Nat Struct Biol* 2003, **10**:168-174.
- Hon WC, McKay GA, Thompson PR, Sweet RM, Yang DS, Wright GD, Berghuis AM: **Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases.** *Cell* 1997, **89**:887-895.
- Krupa A, Abhinandan KR, Srinivasan N: **Kin G: A database of protein kinases in genomes.** *Nucl Acids Res* 2004, **32**:D153-155.
- Krupa A, Srinivasan N: **Lipopolysaccharide phosphorylating enzymes encoded in the genomes of Gram-negative bacteria are related to the eukaryotic protein kinases.** *Protein Sci* 2002, **11**:1580-1584.
- Angermayr M, Roidl A, Bandlow W: **Yeast Rio1p is the founding member of a novel subfamily of protein serine kinases**

- involved in the control of cell cycle progression. *Mol Microbiol* 2002, **44**:309-324.
34. Hunter T, Plowman GD: **The protein kinases of budding yeast: six score and more.** *Trends Biochem Sci* 1997, **22**:18-22.
  35. Angermayr M, Bandlow W: **RIO1, an extraordinary novel protein kinase.** *FEBS Lett* 2002, **524**:31-36.
  36. Plowman GD, Sudarsana S, Bingham J, Whyte D, Hunter T: **The protein kinases of *Caenorhabditis elegans*: a model for signal transduction in multicellular organisms.** *Proc Natl Acad Sci U S A* 1999, **96**:13603-13610.
  37. **KinG: Kinases in genomes** [<http://hodgkin.mbu.iisc.ernet.in/~king>]
  38. Mukhopadhyay S, Kapratil V, Xu W, Chakrabarty AM: **Characterisation of a Hank's type serine/threonine kinase and ser/threonine phosphoprotein phosphatase in *Pseudomonas aeruginosa*.** *J Bacteriol* 1999, **181**:6615-6622.
  39. Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR, Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH: **CDD: a curated Entrez database of conserved domain alignments.** *Nucleic Acids Res* 2003, **31**:383-387.
  40. Dickmanns A, Schmidt B, Rudolph MG, Mariappan M, Dierks T, von Figura K, Ficner R: **Crystal structure of human pFGE, the paralog of the Calphaformylglycine-generating enzyme.** *J Biol Chem* 2005, **280**:15180-15187.
  41. Jain R, Shapleigh JP: **Characterization of nirV and a gene encoding a novel pseudoazurin in *Rhodobacter sphaeroides* 2.4.3.** *Microbiology* 2001, **147**:2505-2515.
  42. Pratt WB: **The hsp90-based chaperone system: involvement in signal transduction from a variety of hormone and growth factor receptors.** *Proc Soc Exp Biol Med* 1998, **217**:420-434.
  43. Kerk D, Bulgrien J, Smith DW, Gribskov M: **Arabidopsis proteins containing similarity to the universal stress protein domain of bacteria.** *Plant Physiol* 2003, **131**:1209-1219.
  44. Kvint K, Nachin L, Diez A, Nystrom T: **The bacterial universal stress protein: function and regulation.** *Curr Opin Microbiol* 2003, **6**:140-145.
  45. Kobe B, Deisenhofer J: **The leucine-rich repeat: a versatile binding motif.** *Trends Biochem Sci* 1994, **19**:415-421.
  46. Poon WW, Davis DE, Ha HT, Jonassen T, Rather PN, Clarke CF: **Identification of *Escherichia coli* ubiB, a gene required for the first monoxygenase step in ubiquinone biosynthesis.** *J Bacteriol* 2000, **182**:5139-5146.
  47. Yeats C, Finn RD, Bateman A: **The PASTA domain: a beta-lactam-binding domain.** *Trends Biochem Sci* 2002, **27**:438.
  48. Boos W, Shuman H: **Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation.** *Microbiol Mol Biol Rev* 1998, **62**:204-229.
  49. Galperin MY, Nikolskaya AN, Koonin EV: **Novel domains of the prokaryotic two-component signal transduction systems.** *FEMS Microbiol Lett* 2001, **203**:11-21.
  50. Pei J, Grishin NV: **GGDEF domain is homologous to adenylyl cyclase.** *Proteins* 2001, **42**:210-216.
  51. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T: **Structural basis of activity and allosteric control of diguanylate cyclase.** *Proc Natl Acad Sci U S A* 2004, **101**:17084-17089.
  52. Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, Jenal U: **Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain.** *Genes Dev* 2004, **18**:715-727.
  53. Ryjenkov DA, Tarutina M, Moskvina OV, Gomelsky M: **Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain.** *J Bacteriol* 2005, **187**:1792-1798.
  54. Larkin RM, Alonso JM, Ecker JR, Chory J: **GUN4, a regulator of chlorophyll synthesis and intracellular signaling.** *Science* 2003, **299**:902-906.
  55. Motley ST, Lory S: **Functional characterization of a serine/threonine protein kinase of *Pseudomonas aeruginosa*.** *Infect Immun* 1999, **67**:5386-5394.
  56. Whittaker CA, Hynes RO: **Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere.** *Mol Biol Cell* 2002, **13**:3369-3387.
  57. Jinek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E: **The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha.** *Nat Struct Mol Biol* 2004, **11**:1001-1007.
  58. Iyer SP, Hart GW: **Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity.** *J Biol Chem* 2003, **278**:24608-24616.
  59. Whisstock JC, Lesk AM: **SH3 domains in prokaryotes.** *Trends Biochem Sci* 1999, **24**:132-133.
  60. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P: **SMART: a web-based tool for the study of genetically mobile domains.** *Nucleic Acids Res* 2000, **28**:231-234.
  61. Li J, Williams BL, Haire LF, Goldberg M, Wilker E, Durocher D, Yaffe MB, Jackson SP, Smerdon SJ: **Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2.** *Mol Cell* 2002, **9**:1045-1054.
  62. Udo H, Inouye M, Inouye S: **Biochemical characterization of Pkn2, a protein Ser/Thr kinase from *Myxococcus xanthus*, a Gram-negative developmental bacterium.** *FEBS Lett* 1997, **400**:188-192.
  63. Udo H, Lam CK, Mori S, Inouye M, Inouye S: **Identification of a substrate for Pkn2, a protein Ser/Thr kinase from *Myxococcus xanthus* by a novel method for substrate identification.** *J Mol Microbiol Biotechnol* 2000, **2**:557-563.
  64. Ohmori M, Ikeuchi M, Sato N, Wolk P, Kaneko T, Ogawa T, Kanehisa M, Goto S, Kawashima S, Okamoto S, Yoshimura H, Katoh H, Fujisawa T, Ehira S, Kamei A, Yoshihara S, Narikawa R, Tabata Satoshi: **Characterisation of genes encoding Multi-domain proteins in the genome of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. Strain PCC 7120.** *DNA Res* 2001, **8**:271-284.
  65. Wang L, Sun YP, Chen WL, Li JH, Zhang CC: **Genomic analysis of protein phosphatases and two-component regulatory systems of the cyanobacterium *Anabaena* sp. Strain PCC 7120.** *FEMS Microbiol Lett* 2002, **217**:155-165.
  66. Tucker CL, Woodcock SC, Kelsell RE, Ramamurthy V, Hunt DM, Hurley JB: **Biochemical analysis of a dimerization domain mutation in RetGC-I associated with dominant cone-rod dystrophy.** *Proc Natl Acad Sci U S A* 1999, **96**:9039-9044.
  67. Bakal CJ, Davies JE: **No longer an exclusive club: eukaryotic signalling domains in bacteria.** *Trends Cell Biol* 2000, **10**:32-38.
  68. Zhulin IB, Nikolskaya AN, Galperin MY: **Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea.** *J Bacteriol* 2003, **185**:285-294.
  69. Koonin EV, Makarova KS, Aravind L: **Horizontal gene transfer in prokaryotes: quantification and classification.** *Annu Rev Microbiol* 2002, **55**:709-742.
  70. Hughes DA: **Signal transduction. Histidine kinases hog the limelight.** *Nature* 1994, **369**:187-188.
  71. Maeda T, Wurgler-Murphy SM, Saito H: **A two-component system that regulates an osmosensing MAP kinase cascade in yeast.** *Nature* 1994, **369**:242-245.
  72. Phalip V, Li JH, Zhang CC: **HstK, a cyanobacterial protein with both a serine/threonine kinase domain and a histidine kinase domain: implication for the mechanism of signal transduction.** *Biochem J* 2001, **360**:639-644.
  73. Zhang CC, Gonzalez L, Phalip V: **Survey, analysis and genetic organisation of genes encoding eukaryotic proteins on a cyanobacterial genome.** *Nucl Acids Res* 1998, **26**:3619-3625.
  74. Zhang CC: **Bacterial signalling involving eukaryotic-type protein kinases.** *Mol Microbiol* 1996, **20**:9-15.
  75. Wehr W, Niederweis M, Schumann W: **The FtsH protein accumulates at the septum of *Bacillus subtilis* during cell division and sporulation.** *J Bacteriol* 2000, **182**:3870-3873.
  76. Umeyama T, Horinouchi S: **Autophosphorylation of a bacterial serine/threonine kinase, AfsK, is inhibited by KbpA, an AfsK-binding protein.** *J Bacteriol* 2001, **183**:5506-5512.
  77. Hong SK, Matsumoto A, Horinouchi S, Beppu T: **Effects of protein kinase inhibitors on *in vitro* protein phosphorylation and cellular differentiation of *Streptomyces griseus*.** *Mol Gen Genet* 1993, **236**:347-354.
  78. Molle V, Kremer L, Girard-Blanc C, Besra GS, Cozzzone AJ, Prost JF: **An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*.** *Biochemistry* 2003, **42**:15300-15309.

79. Udo H, Munoz-Dorado J, Inouye M, Inouye S: **Myxococcus xanthus**, a gram-negative bacterium, contains a transmembrane protein serine/threonine kinase that blocks the secretion of beta-lactamase by phosphorylation. *Genes Dev* 1995, **9**:972-983.
80. Nariya H, Inouye S: **Modulating factors for the Pkn4 kinase cascade in regulating 6-phosphofructokinase in Myxococcus xanthus**. *Mol Microbiol* 2005, **56**:1314-1328.
81. Nariya H, Inouye S: **Activation of 6-phosphofructokinase via phosphorylation by Pkn4, a proteinSer/Thr kinase of Myxococcus xanthus**. *Mol Microbiol* 2002, **46**:1353-1366.
82. Johnson LN, Noble ME, Owen DJ: **Active and inactive protein kinases: structural basis for regulation**. *Cell* 1996, **85**:149-158.
83. Boitel B, Ortiz-Lombardia M, Duran R, Pompeo F, Cole ST, Cervensky C, Alzari PM: **PknB kinase activity is regulated by phosphorylation in two Thr residues and dephosphorylation by PstP, the cognate phospho-Ser/Thr phosphatase, in Mycobacterium tuberculosis**. *Mol Microbiol* 2003, **49**:1493-508.
84. Morrison DK, Murakami MS, Cleghon V: **Protein kinases and phosphatases in the Drosophila genome**. *J Cell Biol* 2000, **150**:F57-62.
85. Kostich M, English J, Madison V, Gheyas F, Wang L, Qiu P, Greene J, Laz TM: **Human members of the eukaryotic protein kinase family**. *Genome Biol* 2002, **3**:RESEARCH0043.
86. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S: **The protein kinase complement of the human genome**. *Science* 2002, **298**:1912-1934.
87. Krupa A, Srinivasan N: **The repertoire of protein kinases encoded in the draft version of the human genome: atypical variations and uncommon domain combinations**. *Genome Biol* 2002, **3**:RESEARCH0066.
88. Kim D, Forst S: **Genomic analysis of the histidine kinase family in bacteria and archaea**. *Microbiology* 2001, **147**:1197-1212.
89. Koretke KK, Lupas AN, Warren PV, Rosenberg M, Brown JR: **Evolution of two-component signal transduction**. *Mol Biol Evol* 2000, **17**:1956-1970.
90. Bashton M, Chothia C: **The geometry of domain combination in proteins**. *J Mol Biol* 2002, **315**:927-939.
91. Teichmann SA, Murzin AG, Chothia C: **Determination of protein function, evolution and interactions by structural genomics**. *Curr Opin Struct Biol* 2001, **11**:354-363.
92. Park J, Lappe M, Teichmann SA: **Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast**. *J Mol Biol* 2001, **307**:929-938.
93. Kamemura K, Hart GW: **Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: a new paradigm for metabolic control of signal transduction and transcription**. *Prog Nucleic Acid Res Mol Biol* 2003, **73**:107-136.
94. Wells L, Vosseller K, Hart GW: **Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc**. *Science* 2001, **291**:2376-2378.
95. Madec E, Stensballe A, Kjellstrom S, Cladiere L, Obuchowski M, Jensen ON, Seror SJ: **Mass spectrometry and site-directed mutagenesis identify several autophosphorylated residues required for the activity of PrkC, a Ser/Thr kinase from Bacillus subtilis**. *J Mol Biol* 2003, **330**:459-472.
96. Madec E, Laszkiewicz A, Iwanicki A, Obuchowski M, Seror S: **Characterization of a membrane-linked Ser/Thr protein kinase in Bacillus subtilis, implicated in developmental processes**. *Mol Microbiol* 2002, **46**:571-586.
97. Bork P, Brown NP, Hegyi H, Schultz J: **The protein phosphatase 2C (PP2C) superfamily: detection of bacterial homologues**. *Protein Sci* 1996, **5**:1421-1425.
98. Irmiler A, Forchhammer K: **A PP2C-type phosphatase dephosphorylates the PII signaling protein in the cyanobacterium Synechocystis PCC 6803**. *Proc Natl Acad Sci U S A* 2001, **98**:12978-12983.
99. Dutta S, Lewis RJ: **Crystallization and preliminary crystallographic analysis of the kinase-recruitment domain of the PP2C-type phosphatase RsbU**. *Acta Crystallogr D Biol Crystallogr* 2003, **59**:191-193.
100. Delumeau O, Dutta S, Brigulla M, Kuhnke G, Hardwick SW, Volker U, Yudkin MD, Lewis RJ: **Functional and structural characterization of RsbU, a stress signaling protein phosphatase 2C**. *J Biol Chem* 2004, **279**:40927-40937.
101. Lai SM, Le Moual H: **PrpZ, a Salmonella enterica serovar Typhi serine/threonine protein phosphatase 2C with dual substrate specificity**. *Microbiology* 2005, **151**:1159-1167.
102. Duncan L, Alper S, Arigoni F, Losick R, Stragier P: **Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division**. *Science* 1995, **270**:641-644.
103. Feucht A, Magnin T, Yudkin MD, Errington J: **Bifunctional protein required for asymmetric cell division and cell-specific transcription in Bacillus subtilis**. *Genes Dev* 1996, **10**:794-803.
104. Russell LC, Whitt SR, Chen MS, Chinkers M: **Identification of conserved residues required for the binding of a tetratricopeptide repeat domain to heat shock protein90**. *J Biol Chem* 1999, **274**:20060-20063.
105. Chen MS, Silverstein AM, Pratt WB, Chinkers M: **The tetratricopeptide repeat domain of protein phosphatase 5 mediates binding to glucocorticoid receptor heterocomplexes and acts as a dominant negative mutant**. *J Biol Chem* 1996, **271**:32315-32320.
106. Freestone P, Nystrom T, Trinei M, Norris V: **The universal stress protein, UspA, of Escherichia coli is phosphorylated in response to stasis**. *J Mol Biol* 1997, **274**:318-324.
107. Sessa G, Martin GB: **Signal recognition and transduction mediated by the tomato Pto kinase: a paradigm of innate immunity in plants**. *Microbes Infect* 2000, **2**:1591-1597.
108. Black K, Buikema WJ, Haselkorn R: **The hglK gene is required for localization of heterocyst-specific glycolipids in the cyanobacterium Anabaena sp. strain PCC 7120**. *J Bacteriol* 1995, **177**:6440-6448.
109. White KA, Lin S, Cotter RJ, Raetz CR: **A Haemophilus influenzae gene that encodes a membrane bound 3-deoxy-D-mannooctulosonic acid (Kdo) kinase. Possible involvement of kdo phosphorylation in bacterial virulence**. *J Biol Chem* 1999, **274**:31391-31400.
110. Yethon JA, Whitfield C: **Purification and characterization of WaaP from Escherichia coli, a lipopolysaccharide kinase essential for outer membrane stability**. *J Biol Chem* 2001, **276**:5498-5504.
111. Av-Gay Y, Everett M: **The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis**. *Trends Microbiol* 2000, **8**:238-244.
112. Walburger A, Koul A, Ferrari G, Nguyen L, Prescianotto-Baschong C, Huygen K, Klebl B, Thompson C, Bacher G, Pieters J: **Protein kinase G from pathogenic mycobacteria promotes survival within macrophages**. *Science* 2004, **304**:1800-1804.
113. Cowley S, Ko M, Pick N, Chow R, Downing KJ, Gordhan BG, Betts JC, Mizrahi V, Smith DA, Stokes RW, Av-Gay Y: **The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo**. *Mol Microbiol* 2004, **52**:1691-702.
114. Dorow DS, Devereux L, Tu GF, Price G, Nicholl JK, Sutherland GR, Simpson RJ: **Complete nucleotide sequence, expression, and chromosomal localisation of human mixed-lineage kinase 2**. *Eur J Biochem* 1995, **234**:492-500.
115. Katoh M, Hirai M, Sugimura T, Terada M: **Cloning and characterization of MST, a novel (putative) serine/threonine kinase with SH3 domain**. *Oncogene* 1995, **10**:1447-1451.
116. Joshi B, Janda L, Stoytcheva Z, Tichy P: **PkwA, a WD-repeat protein, is expressed in spore-derived mycelium of Thermomonospora curvata and phosphorylation of its WD domain could act as a molecular switch**. *Microbiology* 2000, **146**:3259-3267.
117. Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV: **Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer**. *J Mol Biol* 1999, **289**:729-745.
118. Kamei A, Yoshihara S, Yuasa T, Geng X, Ikeuchi M: **Biochemical and functional characterization of a eukaryotic-type protein kinase, SpkB, in the cyanobacterium, Synechocystis sp. PCC 6803**. *Curr Microbiol* 2003, **46**:296-301.
119. Kamei A, Yuasa T, Geng X, Ikeuchi M: **Biochemical examination of the potential eukaryotic-type protein kinase genes in the complete genome of the unicellular Cyanobacterium synechocystis sp. PCC 6803**. *DNA Res* 2002, **9**:71-78.

120. Kamei A, Yuasa T, Orikawa K, Geng XX, Ikeuchi M: **A eukaryotic-type protein kinase, SpkA, is required for normal motility of the unicellular Cyanobacterium synechocystis sp. strain PCC 6803.** *J Bacteriol* 2001, **183**:1505-1510.
121. Peirs P, De Wit L, Braibant M, Huygen K, Content J: **A serine/threonine protein kinase from Mycobacterium tuberculosis.** *Eur J Biochem* 1997, **244**:604-612.
122. Namboori S, Srinivasan N, Pandit SB: **Recognition of remotely related structural homologues using sequence profiles of aligned homologous protein structures.** *In Silico Biol* 2004, **4**:0037.
123. Galperin MY: **A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts.** *BMC Micro Biol* 2005, **5**:35.
124. Grundner C, Gay LM, Alber T: **Mycobacterium tuberculosis serine/threonine kinases PknB, PknD, PknE, and PknF phosphorylate multiple FHA domains.** *Protein Sci* 2005, **14**:1918-1921.
125. **FTP site at NCBI** [<ftp://ftp.ncbi.nih.gov/genomes/Bacteria>]
126. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller , Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
127. Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L, Altschul SF: **IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position specific score matrices.** *Bioinformatics* 1999, **15**:1000-1011.
128. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14**:755-763.
129. Muller A, MacCallum RM, Sternberg MJ: **Benchmarking PSI-BLAST in genome annotation.** *J Mol Biol* 1999, **293**:1257-1271.
130. Lindahl E, Elofsson A: **Identification of related proteins on family, superfamily and fold level.** *J Mol Biol* 2000, **295**:613-625.
131. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL: **The Pfam protein families database.** *Nucleic Acids Res* 2002, **30**:276-280.
132. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**:567-580.
133. Russell RB, Barton GJ: **Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels.** *Proteins* 1992, **2**:309-323.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

