CHAPTER 6

Diversity of Ser/Thr Kinases in the Genomes of Various
Mycobacterium Species
6.1 Introduction

Ser/Thr kinases (STKs) play a key role in cellular signal transduction and their biological functions have established their roles in cell growth and differentiation. These kinases were originally identified in eukaryotes and later identified in several prokaryotes. Kinases can be classified into two types based on their cellular location, as receptor or non-receptor kinases. Receptor kinases have a membrane spanning region and hence bound to the membrane, and the non-receptor kinases are cytosolic. The other types of kinases are tyrosine and histidine kinases. These proteins function by phosphorylating the cellular target proteins, thereby bringing about a conformational change in them, such that, they either covalently or non-covalently interact with signaling partners to decide the fate of the cell.

The kinase domain of STKs has similar 3-D structures, and their catalytic domain comprises 270 amino acid residues. The 3-D structure comprises two domains, the N-terminal domain comprises mainly β-strands and the C-terminal domain comprises α-helices (Zheng et al., 1993). Previous studies have shown that protein phosphorylation and regulation of kinase activity is sufficiently, divergently evolved in eukaryotes and prokaryotes (Han & Zheng, 2001). Several studies on the evolution of eukaryotic and prokaryotic STKs have been addressed in order to assess the essential features in the regulation of their cellular activities (Hanks & Hunter, 1995; Av-Gay & Everett, 2002; Petrickova & Petricek, 2003). It has been observed that STKs occur as multi-domain proteins viz, a kinase domain is present as combination with other protein domains. Such coexistence of domains in proteins is an essential feature observed in some proteins since: 1) the different domains in proteins are involved in cascade of related events, 2) the activation of one domain by some cofactors brings about an allosteric change, thus leading to the activity of other domains and 3) these domains might be required for the proper positioning of the protein and then translocate it to the site of activity.
Chapter 6

The organization of domains in a large dataset of bacterial STKs has been investigated in order to recognize variety in domain combinations which determine the functions of bacterial STKs. Previous studies (Krupa & Srinivasan, 2005) have shown that STKs in prokaryotic genomes have diverse domain arrangements which are different from that of eukaryotes. Similar results were observed by Zhang et al., 2007 in the genome wide survey of cyanobacteria.

The Mycobacterium genus comprises a number of Gram-positive, acid-fast, rod-shaped aerobic bacteria and is the only member of the family Mycobacteriaceae within the order Actinomycetales. Like other closely related Actinomycetales, such as Nocardia and Corynebacterium, Mycobacteria also have unusually high genomic DNA GC content and are capable of producing mycolic acids as major components of their cell wall.

*M. tuberculosis* is the causative agent of tuberculosis, a chronic infectious disease with a growing incidence worldwide. This species is responsible for more morbidity in humans than any other bacterial disease. It infects 1.7 billion people a year (~33% of the entire world population) and causes over 3 million deaths per year. The *M. tuberculosis* H37Rv genome comprises 3989 proteins (Cole et al., 1998). *M. avium* 104 was earlier thought to cause tuberculosis in birds, but it now proved that it causes infections in immuno-compromized humans, such as the elderly, children and especially patients with AIDS. This genome codes for 5120 proteins (genome project; txid: 243243). *M. bovis* is the causative agent of classic bovine tuberculosis, but it can also cause the disease in humans, especially if contaminated milk is consumed without prior pasteurization. This genome comprises 3920 proteins (Garnier et al., 2003). *M. bovis* strain BCG was used to produce BCG (Bacille de Calmette et Guan) vaccine, a well-known tuberculosis vaccine, originally developed by Calmette and Guan in the 1920's by multiple subculturings that resulted in attenuation or weakened virulence of the strain. This genome comprises 3952 proteins (Brosch et al., 2007). *M. leprae* is the causative agent of leprosy or Hanson's disease. The infection is thought to be spread by the
respiratory route because lepromatous patients harbour bacilli in their nasal passages. This genome consists of 1605 proteins (Geluk et al., 2005). The genome of *M. leprae* is smaller due to reductive genome evolution, with many important metabolic activities and their regulatory circuits eliminated due to extensive recombination events between dispersed repetitive sequences. The bacterium *M. smegmatis* was initially isolated from human smegma. It is associated with soft tissue lesions following trauma or surgery and is also reported as a possible factor in penile carcinogenesis. This genome comprises 6716 proteins (genome project; txid: 246196). *M. ulcerans* Agy99 causes Buruli ulcer, the third most common mycobacterial pathogen after *M. tuberculosis* and *M. leprae*. This genome comprises 4160 genes (Stinear et al., 2007). *M. gilvum* PYR-GCK was isolated from river sediment and is able to degrade pyrene and other aromatic hydrocarbons. This genome consists of 5579 proteins (genome project; txid: 350054). *M. vanbaalenii* PYR-1 strain was isolated from contaminated sites exposed to petrogenic chemicals in the watershed of Redfish Bay, Texas, in 1986. It can degrade polycyclic aromatic hydrocarbons such as fluoranthene, pyrene, phenanthrene. This genome comprises 5979 proteins (genome project; txid: 350058). *Mycobacterium* sp. JLS. was isolated from creosote-contaminated soil and this microbe, along with some others collected at this site, is able to rapidly mineralize 

14C-labeled pyrene. This genome comprises 5739 genes (Miller et al., 2004).

The original genus mycobacterium has been speciated with several variations within the bacterium such that the resultant genomes have diverse functions and variable genome size. While some species are pathogenic to humans under normal conditions, other species are disease causing under immunocompromized conditions and some bacteria have a hydrocarbon processing ability. Due to this diversity in the chemical functions of these species and the recently availability of their complete proteome information has prompted
Chapter 6

us to carry out \textit{in silico} analysis of the Ser/Thr kinases and study their domain architecture in these genomes.

In the current study, we have identified 100 STKs and homologs in 10 completed representative mycobacterial genomes using profile based search methods adapted in PSI-BLAST searches available at NCBI. The domain organization of these proteins has been studied in order to identify other coexisting domains. This analysis will aid decipher the various biological functions of these proteins alongside the kinase activity.
6.2 Methods

Initially sequences encoded by the Ser/Thr kinases were obtained from protein database at NCBI (www.ncbi.nlm.nih.gov/) using keyword searches. These proteins were used by PSI-BLAST searches as queries in order to search ten mycobacterium genomes namely; *M. tuberculosis* H37Rv, *M. avium* 104, *M. bovis*, *M. bovis* strain BCG, *M. leprae*, *M. smegmatis*, *M. ulcerans* Agy99, *M. gilvum* PYR-GCK, *M. vanbaalenii* PYR-1, *Mycobacterium* sp. JLS. The identified proteins were validated using reciprocal BLAST searches, till no new proteins were identified. BLOSUM62 matrix, with existence 11 and extension 1 as gap penalties, Expect threshold 10 and PSI-BLAST threshold 0.005 was chosen for all the PSI-BLAST searches. Multiple sequence alignment was generated to confirm that these proteins do belong to the STK family.

These proteins confirmed as Ser/Thr kinases were submitted to SMART (Schultz *et al.*, 1998) using the batch mode, in order to analyse the domains present in the proteins. The regions that are not identified by SMART as already known domains were analysed manually using PSI-BLAST against nr database at NCBI.
6.3 Results and Discussion

From the analysis of ten complete mycobacterial genomes, we identified 120 STKs and their homologs using profile based database searching methods incorporated in PSI-BLAST. Some of these proteins share 100% sequence identity and hence removed as being redundant. The final database comprises 100 proteins. Multiple sequence alignment of these proteins indicated that the similarity among the kinase domain regions of these proteins is high (>40% sequence homology), compared to the non-kinase domain regions within the proteins (data not shown). The domain organization of these proteins was identified using SMART analysis and we observed that all the proteins comprise a STK domain. Further, we also identified that nearly 75% of the STKs have a membrane spanning region positioned C-terminus to the kinase domain indicating that these are receptor STKs.

The PSI-BLAST searches of the regions, that were not identified by SMART as already known domains revealed the extent of similarity or variation among the STK family members of mycobacterial species. Further, based on the extent of homology in the nonkinase domain regions, these proteins have been classified into subclasses and proteins from each subclass were examined. Proteins in the subclass 1 comprise representative members from all the mycobacterial genomes analysed in this work. The STK domain is followed by 4 tandem PASTA domains at the C-terminus. These PASTA domains were earlier reported by Krupa and Srinivasan, 2005 to be present in bacterial STKs. The presence of PASTA domains strongly suggests that it is a signal-binding sensor domain.

Members of this subclass 2 are present in *M. smegmatis*, *Mycobacterium* sp. *JLS*, *M. vanbaalenii* and *M. gilvum*. These proteins comprise a bacterial periplasmic substrate-binding protein domain (PBPb), C-terminus to the STKs. It is known that bacterial high affinity transport systems are involved in active transport of solutes across the cytoplasmic membrane. The protein components of these traffic systems include one or two transmembrane protein components, one or two membrane-
associated ATP-binding proteins and a high affinity periplasmic solute-binding protein. The latter are thought to bind the substrate in the vicinity of the inner membrane and to transfer it to a complex of inner membrane proteins for concentration into the cytoplasm. It is known that some solute-binding proteins function in the initiation of sensory transduction pathways. Proteins of the subclass 3 comprise NHL repeats, positioned C-terminus to the STK domain. The members of this subclass are proteins from *M. avium* 104, *M. tuberculosis* H37Rv, *M. bovis* BCG. NHL repeats are known to be present in cell surface proteins and these tandem repeats fold into beta-propeller architecture and it is possible that these proteins function by ligand binding to the extracellular beta-propeller structure. The members of subclasses 2 and 3 were earlier also reported by Krupa and Srinivasan, 2005 but, we have identified these proteins in more genomes due to the larger dataset understudy.

Our analysis has further identified novel domain architectures in mycobacterial STKs that have not been reported earlier. Proteins in the subclass 4 comprise two distinct domains, one each at the N- and C-termini, that sandwich the central STK domain. The sequence analysis of N- and C-terminal regions using PSI-BLAST identified STKs from actinobacterial genomes such as *Rhodococcus*, *Janibacter*, *Streptomyces*, *Corynebacterium* and *Nocardia*. This indicates that these two domains are responsible for a function, conserved in all species from actinobacteria. The multiple sequence alignments of N and C-terminal regions are indicated in Figures 6.1a and 6.1b. We do not know the structure or function of these domains, but the presence of conservatively varied residues and cysteines at conserved positions in the multiple sequence alignment (Figure 6.1a) and several conserved sequence motifs in the multiple sequence alignment (Figure 6.1b) implies a common function mediated by these novel domains.

Members of the subclass 5 are present in *Mycobacterium* sp. *JLS* (YP_001072984) and *M. vanbaalenii* (YP_951920) genomes. They comprise Kelch repeats, positioned C-terminus to the STK domain. Kelch is a 50 amino acid residue
sequence motif that represents one beta-sheet blade and several of these repeats can associate to form a beta-propeller. The funnel of the beta propeller structure is often important for ligand recognition and this might be important for the activation of these kinases.

Members of the subclass 6 are present in all mycobacterial species analysed. They comprise a conserved C-terminal domain as indicated in Figure 6.2. Also from this multiple sequence alignment it can be seen that several sequence motifs are conserved, indicating a common function mediated by these proteins. One protein from *M. avium* (YP_025533), comprises Peptidyl-prolyl cis-trans isomerase domain. One protein from *M. gilvum* (YP_001134469) comprises NERD domain positioned N-terminus to the kinase domain. The NERD (nuclease-related domain) is found in a range of bacterial as well as archaeal and plant proteins. NERD domain has distant similarity to endonucleases. The representation of STKs in these subclasses is shown in Figure 6.3.

From the analysis of STKs in ten representative mycobacterial genomes, we observed that the kinase domain is highly conserved among all members of mycobacterial species. Kinase domain coexists with PASTA, PBPb, peptidyl-prolyl cis-trans isomerase and NERD domains, as well as NHL and Kelch repeats. Certain STKs (from subclass 6) consists of a conserved domain specifically present in mycobacterial species while STKs (from subclass 4) consists of domains at N- and C-terminus that are also present in several actinomycetales indicating the restriction in the divergent evolution of STKs.
Figure 6.1a. The Multiple Sequence Alignment of N-terminal Regions of Some STKs. (These proteins also occur in Rhodococcus sp, Janibacter sp, Streptomyces coelicolor, Streptomyces avermitilis, Frankia alni and Nocardia farcinica Apart from mycobacterium species. * indicates a conserved residue).
Chapter 6

**Figure 6.1b.** The Multiple Sequence Alignment of C-terminal Regions of Some STKs. (These proteins also occur in *Rhodococcus* sp, *Janibacter* sp, *Streptomyces coelicolor*, *Streptomyces avermitilis*, *Frankia alni*, *Corynebacterium glutamicum*, *Corynebacterium efficiens*, *Corynebacterium diphtheriae*, *Corynebacterium jeikeium*, *Streptomyces avermitilis*, *Janibacter* sp and *Nocardioides farcinica* apart from mycobacterium species. * indicates a conserved residue).

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Sequence</th>
<th>Protein ID</th>
<th>Sequence</th>
<th>Protein ID</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP_826552</td>
<td>AQARTEAEIVASLV1</td>
<td>NP_301338</td>
<td>AQARTEAEIVASLV1</td>
<td>2PZI_1</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>NP_739199</td>
<td>AQARTEAEIVASLV1</td>
<td>NP_940377</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_250022</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>YF_00113951</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_736553</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_702160</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>YF_121560</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_915147</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_885191</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>YF_001131485</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_8R6576</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_883788</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>NF_334633</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_129492</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_301138</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>2PZI_1</td>
<td>AQARTEAEIVASLV1</td>
<td>NF_862827</td>
<td>AQARTEAEIVASLV1</td>
<td>NF_326552</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>YF_883788</td>
<td>AQARTEAEIVASLV1</td>
<td>NF_419785</td>
<td>AQARTEAEIVASLV1</td>
<td>NF_862901</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>YF_714096</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_708503</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_125749</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>NF_418150</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_714532</td>
<td>AQARTEAEIVASLV1</td>
<td>NF_627549</td>
<td>AQARTEAEIVASLV1</td>
</tr>
<tr>
<td>YF_720616</td>
<td>AQARTEAEIVASLV1</td>
<td>YF_121560</td>
<td>AQARTEAEIVASLV1</td>
<td>ZF_00995933</td>
<td>AQARTEAEIVASLV1</td>
</tr>
</tbody>
</table>

---

..... continued
Diversity of Ser/Thr kinases...

---

...... continued
Chapter 6

184
Figure 6.2. The Multiple Sequence Alignment of C-terminal Regions of Some STKs from Mycobacterium Species. (* indicates a conserved residue).
Figure 6.3. A Representative Domain Architecture Diagram of STKs from Ten Mycobacterial Genomes Analysed.
6.4 Conclusions

1. The kinase domain in STKs is highly conserved compared to the other regions in the proteins.

2. In mycobacterial STKs, the kinase domain coexists with PASTA, PBPb, peptidyl-prolyl cis-trans isomerase and NERD domains. It also coexists with NHL and Kelch repeats.

3. Diversity in the STKs is evident from the fact that some STKs are present only in mycobacterial species while some STKs are common to other members of Actinobacteria.
6.5 References


List of Publications


2) Chemical Function Based Virtual Screening: Discovery of potent lead molecules for the \textit{Bcr-Abl} tyrosine kinase using VX-680. \textbf{Krishna Kishore Inampudi} and Lalitha Guruprasad (being communicated).

3) The Identification of New Aurora A Kinase Inhibitors by Pharmacophore Modeling, Virtual Screening and Molecular Docking. \textbf{Krishna Kishore Inampudi} and Lalitha Guruprasad (being communicated).


5) Diversity of Ser/Thr kinases in the genomes of various \textit{Mycobacterium} species. \textbf{Krishna Kishore Inampudi}, N. Srinivas and Lalitha Guruprasad (being communicated).

