SUMMARY

Streptokinase, an extracellular protein produced by several pathogenic streptococci, is one of the most frequently utilized thrombolytic agent for the treatment of acute myocardial infarction. Pathogenicity of the native host prompted many research groups to explore heterologous non-pathogenic hosts that can generate large amount of streptokinase (SK) of high quality. Great degree of complexity, like degradation of protein and instability of cloned gene, has been encountered during streptokinase production through recombinant route. Thus, to develop an ideal system for its efficient production and to unravel the molecular basis of these problems, associated with the production of genetically engineered streptokinase, detailed study targeted towards these aspects are highly essential. With this objective in mind, a study was initiated to design and develop a highly efficient expression system for the extracytoplasmic production of SK in *E. coli* which can also overcome the problems of its tedious purification procedure utilized to harvest this protein. Accordingly, gene encoding SK was cloned in *E. coli* along with its native regulatory signals which resulted in low expression of *skc* gene and partial transport of protein into the extrcellular milieu. In order to enhance the level of protein production and to overcome the problem of incompatibility of SK transport signal with *E. coli* secretion machinery, structural gene of SK was fused in frame with various signal sequences of *E. coli*, viz. *pel*B, *omp*A, *lam*B and *mal*E and cloned under the strong regulatable *tac* promoter. It was evident from the present study that *omp*A-*skc* fusion resulted in the maximum transport of SK in the culture supernatant of *E. coli*. To further optimize the level of secreted SK, *E. coli* strains of various genetic background were screened and a combination of ideal host vector system was established which gave approx. 10 times higher level of secreted SK as compared to that obtained using its natural regulatory signals. Furthermore, expression plasmid, pJKD56, was found highly stable even in the absence of any selection pressure. Using this combination of host-vector system, critical

parameters affecting the protein production, e.g., induction condition, age and percentage of seed inoculum, harvest time of recombinant SK (rSK) etc., were studied and optimal condition for its extracellular production in *E. coli* were established.

The complete nucleotide sequence of *skc* gene and *ompA-skc* fusion was determined which confirmed the authenticity of *skc* gene and its correct alignment with the *ompA* signal sequence. Recombinant SK (rSK) was further characterized by SDS-PAGE and immunoblot analysis which indicated a major 47kDa protein product corresponding to the size of the natural SK along with a minor 44kDa protein product. Degradation of 47kDa SK into 44kDa form has been reported due to proteolytic degradation of C-terminal residues of SK in *Streptococcus sanguis*. To understand the problem of this proteolytic degradation and to check that it is only associated with some specific hosts, secretion of SK was studied in the eucaryotic host, *Pichia pastoris*, by integrating the *skc* gene into its genome. Processing of 47kDa into 44kDa product was found to be substantially reduced in this host and only one major SK product was detectable by SDS-PAGE and western blot analysis. Secreted SK was found glycosylated but fully functional, indicating that glycosylation has not occurred within the functional domain of SK. Glycosylated SK exhibited great degree of stability after prolonged storage. Enhanced stability of SK and significant reduction in its degradation was attributed to the masking of protease sensitive sites due to glycosylation of SK.

Amino-terminal sequence analysis of 47kDa and 44kDa products confirmed that both species are having identical amino-terminus, and 44kDa form of SK has appeared due to the processing of carboxy-terminal residues. Interestingly, both products were found to lack their first two amino terminal residues. In contrast, SK secreted in *E. coli* through its native transport signal retained its correct amino terminus. This was totally unexpected as OmpA signal peptide has been shown to be processed correctly in *E. coli*. Therefore, to check that this is the result of erroneous processing of OmpA-SK fusion product due to the cleavage at alternate site during protein transport, detailed investigations were carried out using various hybrids of SK carrying different *E. coli*

derived transport signal and site directed mutagenesis of amino-terminal segment of Sl PelB-SK and LamB-SK fusions which showed high probability of turn structure ne the pre-protein processing site as present in the natural pre-SK, resulted in the nativ amino-terminus; whereas in OmpA-SK and MalE-SK fusions reduction of turn structur in the vicinity of cleavage junction, resulted in the recognition of a nearby alternat cleavage site, created due to the presence of an alanine residue at second position of amino terminus of SK. These results unveiled the unique phenomenon of SK processing in *E. coli* and led to the conclusion that the compatibility of leader peptide with the mature sequence of SK, which can satisfy the secondary structure requirement within the cleavage region of preprotein, is essential for maintaining the correct processing. Furthermore, site directed mutagenesis of *ompA-skc* fusion junction confirmed that alteration in the secondary structure in the OmpA-SK fusion near the processing site, has resulted in the recognition of second alanine of the mature SK and is responsible for the deletion of two amino acids from the amino terminus of secreted rSK. These observations also provide the evidence that signal peptidase stutters when it encounters two or more juxtaposed target sequences and its proper alignment with respect to the cleavage site is modulated by the secondary structure attained near the processing region.

Studies conducted on SK in the past mainly focussed on identification and characterization of core segment of SK involved in binding and activation of plasminogen molecule while the role(s) of other accessory domains in protein function is still largely unknown. During the present study efforts were made to understand the role(s) of terminal regions of SK by characterizing the genetically engineered forms of SK altered in amino and carboxy terminal regions. Characterization of various amino-terminal deleted forms of rSK revealed that its functional properties remained unchanged even when the first 13 N -terminal amino acid residues were removed. Role of this small segment was explored by characterizing the deleted forms of SK. It was found that full length SK, devoid of any leader sequence, was translocated slowly

across the cytoplasmic and outer membrane, where as deletion mutant, devoid of first 13 N-terminal residues remained inside the cell. Cell fractionation studies as well as genetic evidence utilizing alkaline phosphate fusion pointed towards the existence of additional information for protein transport within the N-terminal domain of SK. This was further substantiated by creating in frame genetic fusion of ompA transport signal with full length and 13 aa residues deleted SK. Pattern of protein export of these fusion constructs in *E. coli* revealed that translocation of proteins is 3-4 times faster when the first 13 N-terminal amino acid residues of SK were intact. Based on these observations it was proposed that N-terminus of mature SK maintains the export competent status of protein and thus confers speed and efficiency upon the translocation process of streptokinase. Additionally, it provided the evidence that amino-terminal region of SK modulates the folding characteristics of full length molecule.

Secretion of SK in *E. coli* was found to be associated with C-terminal cleavage resulting in appearance of degradation product of mature form and thus affecting the overall yield of mature protein. To analyse the role of *E. coli* protease system in the C-terminal processing of SK, *E. coli* mutants deficient in the combination of different intracellular or membrane bound proteases were utilized. Only *E. coli* strain, deficient in seven major proteases, was able to secrete SK in the complete full length form whereas, other protease deficient strains blocked in various one or two proteases, resulted in C-terminal processing of SK with varying efficiency. These results suggested that C-terminal degradation of SK is the result of action of two or more proteases and more than one protease susceptible sites may be present in this region. Structural analysis of C-terminal domain of SK showed that it has the tendency to form beta-turn structure and, thus, may be surface exposed, making it prone to *E. coli* protease attack. Alterations in the last 30-50 carboxy-terminal residues of SK, significantly affected the functional characteristics of SK. How this effect is manifested via carboxy terminal region of SK is not known. Role of this segment in protein function was studied by characterizing the five site directed mutants in which charge density and hydrophobicity

of protease sensitive site of C-terminal region was altered. Transport pattern of C-terminally modified forms of SK was identical to the wild type indicating that the C-terminal region of SK does not carry any determinant for the protein export. Additionally, conformational modification as well as blockage of putative protease susceptible site did not completely abolish the formation of 44kDa SK. Also, an additional minor 45 kDa form of SK was also visible in some cases. These results suggested that more than one protease sensitive sites are present in this region and suppression of one site exposes another protease susceptible site in this region. Functional characterization of C-terminally altered forms of SK revealed that this region directly or indirectly modulates overall activator activity of SK. Based on the results obtained during these studies it has been proposed that C-terminal domain of SK is involved in potentiating the function of SK, probably by stabilizing the SK-PG activator complex.

Overall, this study has resulted in the establishment of a novel hyperproduction system for the extracellular streptokinase. This system, if upscaled at higher scale, may provide a facile system for large scale preparation of recombinant SK having functional characteristics identical to the natural SK. Molecular studies of genetically engineered SK, carrying modifications in the amino and carboxy-terminal domain of SK has provided the evidences for the role(s) of these regions in protein function. Simultaneously, it revealed the role of structural features of secreted proteins at the preprotein cleavage junction which may be helpful in designing the fusion of heterologous proteins with *E. coli* transport signals. Characterization of terminal regions of SK may be helpful in developing a smaller version of fully functional or better molecule lacking some of the highly immunogenic region of SK, and, thus, may come out as a better substitute of SK for the thrombolytic therapy.