Introduction
One

Introduction

*Streptococcus pyogenes* is a gram positive human bacterial pathogen causing spectrum of diseases ranging from simple and uncomplicated pharyngitis to severe invasive infections and the post streptococcal nonsuppurative sequelae of acute rheumatic fever (RF) and acute glomerulonephritis. Streptococcus is classified on the basis of the presence of different group specific antigens. *S. pyogenes* has group A specific antigens in their cell wall therefore also known as group A streptococcus (GAS) (Cunningham, 2000). GAS mediated diseases continue to be a major problem world-wide. From the mid-1980s to the 1990s, eight RF outbreaks were documented in the United States, with the largest in Salt Lake City, Utah (Ayoub, 1992). In the late 1980s, streptococcal toxic shock syndrome (STSS), bacteremia, and severe, invasive GAS skin and soft tissue infections were reported in the United States and Europe (Greenberg et al., 1983; Cone et al., 1987; Stevens et al., 1989; Holm, 1996). Increased bacteremic infections were reported in Colorado, Sweden, and in United Kingdom (Stromberg et al., 1991).

In developing countries, RF remains an endemic disease with annual incidences ranging from 100 to 200 per 100,000 school-aged children and is a major cause of cardiovascular mortality (Olivier, 2000). GAS, *S. pneumoniae* and *Staphylococcus aureus* are important causes of severe infection in young children in the Papua New Guinea highlands (Lehmann et al., 1999). It has been estimated that there may be 30 million children with rheumatic heart disease (RHD) in developing countries, compared to only 1.5 million in developed countries (Nandi et al., 2001). Although GAS associated diseases have decreased in many parts of the world but in India still they continue to be a major cause of cardiovascular morbidity and mortality. It is estimated that there are at least 517,000 deaths each year due to severe GAS diseases. The greatest burden is due to RHD, with a prevalence of at least 15.6 million cases, with 282,000 new cases and 233,000 deaths each year. Incidence of such infections is high in developing countries like India (Carapetis et al., 2005; Sagar et al., 2008).

In spite of the high mortality and substantial economic losses caused by these diseases, there is currently no licensed vaccine to prevent GAS infections in human. The ideal universal vaccine should be effective against all diseases caused by GAS serotypes. Therefore the distribution pattern of different serotypes of GAS causing
pathogenesis/diseases in the world population including developed and developing countries is essential to be explored to prepare a universal vaccine (WHO, 2005). Epidemiological studies unfold the facts related to the distribution pattern of GAS in the world; and the dynamics of variations in serotypes distribution of GAS with respect to time, region and climatic factors. The epidemiological data are helpful to decide the potential target serotypes, subtypes and strains of any pathogen like GAS.

M protein is a major virulence factor of GAS which is expressed on surface. M protein based serotyping method is used for over the past 60 years. High diversity of GAS serotype and strain is due to highly variable amino acid sequences of M protein in different serotypes. More than 150 serotypes have been reported till date (Enright et al., 2001). Conventionally, simple antigen-antibody agglutination reaction is followed for serotyping of GAS. But it is expensive, time consuming and difficult to prepare specific antibodies for each of 150 M serotypes of GAS. Further, there is always a chance of erroneous handling when conventional serotyping method is opted. Due to these limitations, this approach is outdated at present. That is why an advanced method is required for serotyping of GAS. Molecular method of serotyping has been developed which is better, highly acceptable, accurate, cheaper and advanced option at present.

Accumulated evidences indicate that the \textit{emm} gene encodes the cell surface M protein that is responsible for at least more than 150 known M serotypes of GAS. By using a system based on sequence analysis of the portion of the \textit{emm} gene that encodes M serospecificity, the problems associated with conventional serotyping approach can be avoided. This system is also called \textit{emm} typing that relies upon the use of the two highly conserved primers to amplify a large portion of the \textit{emm} gene. The hyper variable sequence encoding M serospecificity lies adjacent to one of the amplifying primer sequences, allowing for direct sequencing.

Several epidemiological reports have been published related to the distribution pattern of GAS serotypes in the developed countries (Alberti et al., 2003; Brandt et al., 2001; Dicuonzo et al., 2001; Espinosa et al., 2003; Ho et al., 2003; Kao et al., 2005; Moses et al., 2003) but little is known regarding the \textit{emm} type distribution of GAS from India as only few reports are available on \textit{emm} typing and GAS serotyping from India (Menon et al., 2001; Navaneeth et al., 2001; Sagar et al., 2004, 2008). The differences in
serotype distribution among various populations also may reflect differences in pathogenesis among the serotypes. Ongoing monitoring of the distribution of GAS serotypes is important for charting changes in serotype prevalence and for determining the components of an effective GAS vaccine because the current multivalent vaccine preparation do not include the vaccine candidate against the majority of the circulating serotypes from India (Sagar et al., 2008). In addition, there is a lack of information regarding the distribution of emm types over the last five years (2004-2008) in India. Therefore present study was conducted to explore the epidemiology of GAS to know the prevalent serotypes in North Indian population from the samples collected during 2004-2008.

Successful pathogenicity of GAS is dependent on efficient adherence, colonization, invasion, evasion from host immune system and systemic persistence inside the host. Host-pathogen interactions occur due to binding of surface streptococcal ligands to specific receptors on host cells. Attachment of GAS to pharyngeal or dermal epithelial cells is the most important initial step in colonization of the host. Colonization is followed by attachment of GAS as a result of rapid division. It not only adheres to epithelial cells but also invades them. The fact that GAS evolved multiple routes to the interior of epithelial cells is a strong indication that intracellular invasion plays an important role in their pathogenesis.

As GAS attaches, colonizes and invades the host tissues, it is resisted by host immune system. Invasive GAS occasionally establishes a deeper infection. Depending on both the virulence of the strain and the host susceptibility (Norrby-Teglund et al., 2000), severe invasive infections may lead to toxic shock and necrotizing fasciitis due to the induction of high levels of inflammatory cytokines like IL-1β, IL-6, IL-8, IL-10, IL-12 and TNF. Non-specific or misdirected release of cytotoxic agents is often the underlying cause of inflammatory disorders (Kotb, 1995).

Invasiveness of GAS depends on the expression of different virulence factors [M protein, C5a peptidase, Streptococcal fibronectin binding protein (sfb1), exotoxins and many other proteins]. Expression of these factors is regulated by several two-component systems in GAS (Musser and Sitkiewicz, 2006). When GAS invades human body, it encounters with different physiological conditions (temperature, salt concentration and
pH etc). Therefore we hypothesize whether physiological factors play any role to affect the invasion efficiency of GAS. Additionally, the proinflammatory immune response of host against invading GAS was also checked and the correlation of invasion efficiency and host immune response against GAS was developed for the first time.

It is desirable to develop vaccine against GAS because there is no licensed vaccine to prevent GAS infections in world human population even after intensive research for last two decades in this direction (WHO, 2005). This is due to several reasons like high diversity and heterogeneity of GAS at serotype and antigen level. After genome sequencing of several GAS strains, some research groups followed bioinformatics coupled with proteomics approaches for high throughput genome screening to find out most eligible and universal vaccine candidates against GAS (Rodriguez-Ortega et al., 2006; Severin et al., 2007). But bioinformatics screening was not done by considering all contemporary sequenced GAS strains. Therefore vaccine candidates selected by them may not be considered as universal. In the present study, sequences of already sequenced eight serotypes like M1, M2, M3, M4, M6, M12, M18 and M28 were used for the first time to mine the best vaccine candidates. By using different bioinformatics tools like SignalP3.0, Blast, HMMTOP, Pfam, Prosite and Multialignment, we could predict the most effective vaccine candidates against GAS.

Invasiveness of GAS depends on type and site of isolation of GAS (Molinari and Chhatwal, 1998). By comparing the proteome of highly invasive and less invasive GAS serotypes, good vaccine candidates may be screened. Absolute Quantification of protein (AQUA), Isotope Coded Affinity Tag (ICAT), Mass Assisted Laser Desorption and Ionization (MALDI) and Electrosoray Ionization-Liquid Chromatography-Mass Spectrometry (ESI-LC-MS/MS) are such powerful proteomic tools which can be exploited to explore the possibility in identification of proteins vaccine candidates (Johri et al., 2006; 2007). In present study, proteomic approach (1D-SDS-PAGE and ESI-LC-MS/MS) was used to compare the proteomes of highly invasive M49 and less invasive M1 GAS serotypes to find out differentially expressed proteins in both the serotypes and to screen out proteins as good vaccine candidates expressed exclusively in invasive M49 serotype. The vaccine candidates found common in proteomics and bioinformatics study
were considered as the best candidates against GAS which may be targeted in future for vaccine development.

Therefore, for the first time, following objectives were undertaken for the present Ph.D. work,

- **Molecular serotyping of GAS isolates of Indian origin.**
- **Interaction study of most prevalent vs. less prevalent GAS serotypes with human lung alveolar carcinoma epithelial cell line type II pneumocytes (A549) and human hypopharyngeal carcinoma cell line (HEp-2) under different physiological conditions.**
- To analyze the production of proinflammatory cytokines (IL-1β, IL-6, IL-8, IL-10, IL-12P70 and TNF-α) by A549 and HEp-2 cells when interacted with most prevalent (M1) and second most prevalent serotype (M49) of GAS under different physiological conditions.
- **In silico** prediction of universal vaccine candidates.
- **Application of the proteomic tools to identify protein vaccine candidates that are expressed differentially in the most prevalent and invasive GAS serotype.**

Our results on *emm* typing revealed M1 (21.73 %) and M49 (8.7 %) as the most and the second most prevalent serotypes in all 92 identified GAS isolates out of total 160 clinical samples collected from North India. We have found that M1 was less invasive than M49 serotype. In case of M1 serotype maximum invasion i.e., 3.61 % was found at 0.6M NaCl and in case of M49 it was 12.2 % at 25°C with A549 cells. We also checked the *in vitro* proinflammatory immune response of cytokines like TNF-α, IL-1β, IL-10, IL-6, IL-8, and IL-12P70 by A549 and HEp-2 cells during the 2 h interaction with M1 and M49 serotypes. HEp-2 cells produced significantly (*P* ≤ 0.001) higher amount of IL-6 i.e., 317.4±12.7 and 305.45±31 pg/ml when stimulated by M49 serotype grown at 25°C and 0.3 M NaCl conditions respectively as compared to normal conditions (0.1M NaCl, 37°C) (production was 184.05±24.11 pg/ml). IL-6 and IL-8 production was higher in comparison to other cytokines suggesting the activation of humoral and cellular immune response by A549 and HEp-2 cells. We found that *in vitro* proinflammatory response by
the host depends on cell type and serotype of GAS; and is inversely proportional to the invasiveness of GAS under different physiological conditions.

By using *in silico* analysis, we have selected 147 most eligible vaccine candidates in GAS which were predicted to be expressed on GAS surface and having externally exposed conserved portion. These proteins mainly include transporters, proteases/peptidases, metabolic proteins, permeases, toxins and kinases etc. One highest invasive isolate of M49 and least invasive isolate of M1 GAS serotype were selected after conducting invasion study with A549 and HEp-2 cells. To identify the best vaccine candidates, we have coupled bioinformatic tools with proteomics approach and focused on surface proteins. Among the 919 proteins identified in M49 serotype (whole proteome), 582 were detected exclusively in M49 isolate and 337 were detected in both M1 and M49. Only 70 out of total 407 proteins were detected exclusively in M1. We found that most of these proteins were related to metabolic enzymes/proteins, ribosomal proteins and transporter/solute binding proteins indicating the active stage of GAS at log phase. 16 proteins were found common using bioinformatics and proteomics in cell wall and membrane fraction of both M1 and M49 while 23 and 1 predicted proteins were detected exclusively in cell wall and membrane fraction of highly invasive M49 and M1 GAS isolate respectively. We propose that these proteins (n=40) selected using both bioinformatics and proteomics approach may be targeted as potential universal candidates for vaccine development against GAS in future.