Quantitative and Structural insights into the Streptococcal Antigen-ome

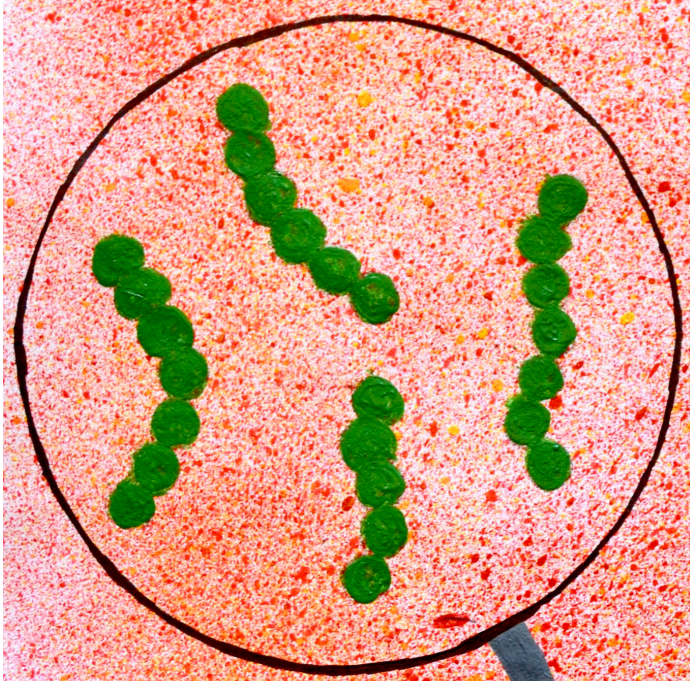
Chowdhury, Sounak

2022

Document Version:
Publisher's PDF, also known as Version of record

Quantitative and Structural insights into the Streptococcal Antigen-ome

SOUNAK CHOWDHURY
DEPARTMENT OF CLINICAL SCIENCES, LUND | FACULTY OF MEDICINE | LUND UNIVERSITY



Quantitative and Structural insights into the Streptococcal Antigen-ome

Quantitative and Structural insights into the Streptococcal Antigen-ome

Sounak Chowdhury

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on the 27th October 2022, Thursday at 9:30 AM in Belfragesalen, Biomedical Center, Lund, Sweden

Faculty opponent
Professor Shiranee Sriskandan

Department of Infectious Disease, Imperial College London

London, England

Organization
LUND UNIVERSITY

Faculty of Medicine
Department of Clinical Sciences
Division of Infection Medicine
BMC D13
22184, Lund, Sweden

Author(s): Sounak Chowdhury

Document name
Doctoral dissertation

Date of issue
2022-10-27

Sponsoring organization

Title and subtitle: Quantitative and Structural insights into the Streptococcal Antigen-ome

Abstract

Infections are a leading cause of human mortality and morbidity worldwide. The development and progression of an infection involves a complex interplay between pathogen and host factors, where pathogens try to evade the host immune system using their virulence factors and the host immune system tries to neutralize the pathogen to prevent the infection. This results in the formation of dynamic host-pathogen protein-protein interactions (HP-PPIs). Dissecting these protein networks using proteomics can provide critical insights into the underlying pathogenesis of a disease. This thesis is focused on using different proteomic tools to analyze such multimeric protein networks formed between *Streptococcus pyogenes* (*S. pyogenes*) and human proteins.

A proteomics based immuno-profiling strategy was developed that could quantitatively profile bacterial antigens based on host humoral immune responses. The method was then applied on one of the most common human pathogen, *S. pyogenes,* and we report a comprehensive snapshot of the streptococcal antigen-ome. The study revealed that only a small fraction of the *S. pyogenes* surface proteins elicited antibody responses in a host specific manner. Additionaly comparative analysis of the antibody responses during natural exposure and invasive infection confirmed that the composition of the antibody repertoires against the *S. pyogenes* antigens changes during infection as compared to the antibody repertoires during natural exposure. Moreover, the panel of 50 *S. pyogenes* antigens generated can further be interrogated for its capability as potential vaccine candidates. We further quantitatively and structurally analyzed complex protein interactions formed between human proteins and different *S. pyogenes* serotypes and one of its major antigen, the M protein. Using state-of-the art techniques, novel interactions were discovered and possible immune evasion strategies were suggested. The *S. pyogenes*-human protein interactome was found to be highly serotype specific and to a great extent mediated by the different M proteins. We also report that *S. pyogenes* forms local environment dependent protein complexes to evade human immune defenses. Furthermore, the antibody binding site was mapped on the M1 protein using integrative structural biology approaches thereby validating the role of M proteins in preventing phagocytosis of *S. pyogenes*. The work also demonstrates the immuno-modulatory role of fibrinogen binding to M proteins and platelet activation during *S. pyogenes* infections.

In conclusion, this study provided critical insights into the *S. pyogenes* antigen-ome, thereby paving ways for the development of vaccines and other therapeutic interventions. Furthermore, the work also provides a repository of different proteomics techniques that could be used to decipher protein interactions to great details for any other pathogen to advance our understanding of the pathogenesis.

Key words: Infection**,** host-pathogen protein-protein interactions, *Streptococcus pyogenes*, M protein, antibodies, vaccines, proteomics, mass spectrometer**.**

Classification system and/or index terms (if any)

Supplementary bibliographical information — **Language:** English

ISSN and key title: 1652-8220
Lund University, Faculty of Medicine Doctoral Dissertation Series 2022:134 — **ISBN:** 978-91-8021-296-0

Recipient's notes — **Number of pages: 97** — Price

Security classification

Signature Sounak Chowdhury. Date 2022-09-15

Quantitative and Structural insights into the Streptococcal Antigen-ome

Sounak Chowdhury

LUND
UNIVERSITY

Coverphoto: painted by Sounak Chowdhury, “Insights into the *Streptococcus pyogenes* infection”

Faculty of Medicine
Department of Clinical Sciences, Lund

Doctoral Dissertation Series 2022:134
ISBN 978-91-8021-296-0
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2022

To Ma & Baba…

“The important thing is not to stop questioning,
Curiosity has its own reason for existence.”

- Albert Einstein

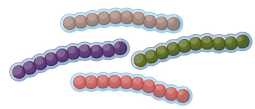


Table of Contents

Preface

PhD, a memorable journey.

I started my PhD in August 2018 to study host-pathogen interactions using proteomics. *Streptococcus pyogenes* and proteomics are the unifying elements of this thesis. These four years have taught me a lot about the different aspects of research and also trained me to become an independent researcher. During this period, I faced various challenges and every time I solved them, I learned new things. It is quite satisfying to see how different questions, ideas, problems and thoughts randomly accumulated during these years have resulted into this thesis.

The thesis consists of four main sections. The first section is a brief introduction to some key concepts, which would help the readers understand the papers appended in this thesis. In the next two sections the aim of the work and the major findings of the papers are highlighted. Finally, the thesis is concluded with a discussion on the different papers and future perspectives.

My sincere hope is that this work provides valuable information and that it stimulates further research.

Sounak Chowdhury

15th September 2022

Papers included in the thesis

Paper I

"Profiling bacterial antigenic proteins using humoral immune responses"

Chowdhury S, Toledo AG, Hjortswang E, Bläckberg A, Shannon O, Rasmussen M, Malmström L & Malmström J.

In manuscript

Paper II

"*Streptococcus pyogenes* Forms Serotype- and Local Environment-Dependent Interspecies Protein Complexes"

Chowdhury S, Khakzad H, Bergdahl GE, Lood R, Ekstrom S, Linke D, Malmström L, Happonen L & Malmström J.

mSystems, volume 6, issue 5, e00271-21. (2021)

Paper III

"Structural determination of *Streptococcus pyogenes* M1 protein interactions with human immunoglobulin G using integrative structural biology"

Khakzad H, Happonen L, Karami Y, **Chowdhury S**, Bergdahl GE, Nilges M, Nhieu GTV, Malmström J & Malmström L.

PLOS Computational Biology, volume 17, issue 1, e1008169. (2021)

Paper IV

"Distinct Serotypes of Streptococcal M Proteins Mediate Fibrinogen-Dependent Platelet Activation and Proinflammatory Effects"

Palm F, **Chowdhury S**, Wettemark S, Malmström J, Happonen L & Shannon O.

Infection and Immunity, volume 90, issue 2, e00462-21. (2022)

Excluded publications

Research articles

- "Long-Read Sequencing Reveals Genetic Adaptation of *Bartonella* Adhesin A Among Different *Bartonella henselae* Isolates"

  Thibau A, Hipp K, Vaca DJ, **Chowdhury S**, Malmström J, Saragliadis A, Ballhorn W, Linke D & Kempf VAJ.

  Frontiers in Microbiology, volume 13. (2022)

- "Invasive streptococcal infection can lead to the generation of cross-strain opsonic antibodies"

  Neergaard TD, Bläckberg A, Ivarsson H, Thomasson S, Ahnlide VK, **Chowdhury S**, Khakzad H, Malmström J, Rasmussen M & Nordenfelt P.

  Under revision

Review articles

- "Structural proteomics, electron cryo-microscopy and structural modeling approaches in bacteria-human protein interactions"

  Chowdhury S, Happonen L, Khakzad H, Malmström L & Malmström J

  Medical Microbiology and Immunology, volume 209, issue 3. (2020)

- "Host-Pathogen Adhesion as the Basis of Innovative Diagnostics for Emerging Pathogens"

  Belkum AV, Almedia C, Bardiaux B, Barrass SV, Butcher SJ, Caykara T, **Chowdhury S**, Datar R, Eastwood I, Goldman A, Goyal M, Happonen L, Pruneyre NI, Jacobsen T, Johnson PH, Kempf VAJ, Kiessling A, Bueno JL, Malik A, Malmström J, Meuskens I, Milner PA, Nilges M, Pamme N, Peyman SA, Rodrigues LR, Mateos PR, Sande MG, Silva CJ, Stasiak AC, Stehle T, Thibau A, Vaca DJ & Linke D.

  Diagnostics, volume 11, issue 7. (2021)

Abbreviations

aa- amino acid

ADCC- Antibody dependent cellular cytotoxicity

AMP- Antimicrobial peptide

AMR- Antimicrobial resistance

AP-MS- Affinity purification-mass spectrometry

APCs- Antigen presenting cells

APSGN- Acute poststreptococcal glomerulonephritis

ARF- Acute rheumatic fever

C domain- Constant domain

C4BP- C4b-binding protein

C5AP- C5a peptidase

C_H- Constant domain of heavy chain

CID- Collision induced dissociation

C_L- Constant domain of light chain

DC- Dendritic cells

DDA- Data dependent acquisition

DIA- Data independent acquisition

ECD- Electron capture dissociation

ELISA- Enzyme linked immunosorbent assay

EndoS- Endo-beta-N-acetylglucosaminidase of streptococci

ESI- Electron spray ionization

ETD- Electron transfer dissociation

Fab- Fragment antigen binding

Fc- Fragment crystallizable

FcRs- Fc receptors

FDR- False discovery rate

GAC- Group A carbohydrate

GAS- Group A *Streptococcus*

H chain- Heavy chain

HCD- High energy collision dissociation

HDX-MS- Hydrogen deuterium exchange-mass spectrometry

HP-PPIs- Host-pathogen protein-protein interactions

HVR- Hypervariable region

IdeS- IgG-degrading enzyme of *S. pyogenes*

Ig- Immunoglobulin

IgA- Immunoglobulin A

IgD- Immunoglobulin D

IgE- Immunoglobulin E

IgG- Immunoglobulin G

IgM- Immunoglobulin M

IL-8 – Interleukin 8

IVIG - Intravenous immunoglobulin

L chain- Light chain

LC- Liquid chromatography

LC-MS – Liquid chromatography-mass spectrometry

m/z- mass to charge ratio

MAC- Membrane attack complex

MBL- Mannose binding lectin

MD- Molecular dynamics

MHC- Major histocompatibility complex

MS- Mass spectrometry

MS/MS- Tandem mass spectrometry

NAD- Nicotinamide dehydrogenase

NETs- Neutrophil extracellular traps

NK- Natural killer cells

PAMPs- Pathogen associated microbial patterns

PPI- Protein-protein interaction

PROS-Vitamin K dependent protein S

PRRs- Pattern recognition receptors

RHD- Rheumatic heart disease

RP-LC- Reverse phase-liquid chromatography

S. pyogenes – Streptococcus pyogenes

SA-MS- Surface adsorption-mass spectrometry

SC- Secretory component

SIC- Streptococcal inhibitor of complement-mediated lysis

sIgA-Secretory IgA

SLO- Streptolysin O

SMEZ- Streptococcal mitogenic exotoxin Z

Spe- Streptococcal pyrogenic exotoxin

SPEB- Streptococcal pyrogenic exotoxin B

SPR- Surface plasmon resonance

SpyCEP- *S. pyogenes* cell envelope proteinase

SSA- Streptococcal superantigen A

STSS- Streptococcal toxic shock syndrome

SWATH- Sequential Window Acquisition of all Theoretical Mass Spectra

TCR- T cell receptor

TIG- Trigger factor

TX-MS- Targeted chemical cross-linking-mass spectrometry

V domain- Variable domain

V_H- Variable domain of heavy chain

V_L-Variable domain of light chain

WHO- World Health Organization

XL-MS – Cross-linking-mass spectrometry

Popular science summary

Proteins are the building blocks of all living organisms. There are thousands of different proteins organized into cells, tissues and organs performing different roles. Any change in the composition or structure of a protein can lead to the malfunctioning of the cells. Therefore, determining the amount and structure of proteins present in various cells are important for understanding how different organisms function. The study of proteins is proteomics and one of the instrument used to identify and measure proteins is known as the mass spectrometer.

Infections are a major cause of human sickness and death worldwide, and the World Health Organization (WHO) has identified infections as a major threat to humankind. Different infecting organisms (pathogens) use specialized proteins called virulence factors to adhere, colonize and invade the host to establish an infection. On the other side, the host immune system tries to prevent the infection resulting in a continuous battle between the host and the pathogen. To succeed over the host immune response, the pathogen uses its virulence factors to interact with multiple different host proteins leading to the formation of a multimeric protein interaction network. Analysis of such protein interaction networks would guide us in understanding the way pathogens initiate an infection and at the same time would help us understand the host defense strategies.

Streptococccus pyogenes (*S. pyogenes*) is a bacterium that only infects humans, and is also one of the leading causes of infection in humans. It causes both mild infections such as strep throat as well as serious life-threatening invasive infection. In this thesis, different mass spectrometry-based proteomics techniques were used to understand the protein interaction networks formed between *S. pyogenes* and human proteins. The human protein networks formed around different *S. pyogenes* and its main virulence factor M protein is demonstrated. The structure of the multimeric protein complexes formed around the bacteria and its role in immune evasion and adaptation in different host microenvironment is also reported. Moreover, the work also outlines the role of binding of host proteins like fibrinogen to M protein in activation of platelets during *S. pyogenes* infections.

The human immune system produces proteins called antibodies that target different pathogenic proteins. Antibodies bound to the pathogens lead to the killing of pathogens by different immune cells. The various *S. pyogenes* proteins targeted by

antibodies are described in this work. The mode of binding of antibodies to the M protein has also been elucidated.

In conclusion, this study helped us understand the different mechanisms involved in *S. pyogenes* infections. Additionally, the proteomic techniques described here can be used to characterize protein networks formed around any other pathogen to better understand the basis of different infections.

Populärvetenskapling sammanfattning

Proteiner är byggstenarna i alla levande organismer. Det finns tusentals olika proteiner organiserade i celler, vävnader och organ, som utför olika roller. Varje förändring i ett proteins sammansättning eller struktur kan leda till att cellerna inte fungerar som de ska. Därför är det viktigt att bestämma strukturen och mängden proteiner som finns i olika celler för att förstå hur olika organismer fungerar. Studier av proteiner kallas proteomik och ett av instrumenten som används för att identifiera och mäta proteiner är masspektrometern.

Infektioner är globalt en viktig orsak till människors sjukdomar och dödsfall, och Världshälsoorganisationen (WHO) har identifierat infektioner som ett stort hot mot mänskligheten. Olika infekterande organismer (patogener) använder specialiserade proteiner som kallas virulensfaktorer för att binda till, kolonisera och invadera värden för att etablera en infektion. Å andra sidan strävar värdens immunsystem till att förhindra infektionen, vilket resulterar i en kontinuerlig kamp mellan värden och patogenen. För att lyckas vinna över värdens immunsvar använder patogenen sina virulensfaktorer för att interagera med flera olika värdproteiner, vilket leder till uppkomsten av ett multimert proteininteraktionsnätverk. Analysen av sådana proteininteraktionsnätverk kunde vägleda oss i att förstå hur patogener initierar en infektion och skulle samtidigt hjälpa oss att förstå värdens försvarsstrategier.

Streptococcus pyogenes (*S. pyogenes*) är en bakterie som endast infekterar människor, och är en av de främsta orsakerna till infektion hos människor. Bakterien orsakar både milda infektioner som halsfluss och mer allvarliga livshotande invasiva infektioner. I denna avhandling har olika masspektrometribaserade proteomik tekniker använts för att förstå de proteininteraktionsnätverk som bildas mellan *S. pyogenes* bakterien och humana proteiner. De mänskliga proteinnätverken som bildas runt olika *S. pyogenes* bakterier och dess huvudsakliga virulensfaktor M proteinet demonstreras. Strukturen av de multimera proteinkomplex som bildas runt bakterierna och deras roll i undvikandet av värden immunförsvar, samt bakteriens anpassning i olika värdmikromiljöer rapporteras också. Dessutom beskriver detta arbete också rollen av inbindning av värdproteiner, såsom fibrinogen till M proteinet vid aktivering av blodplättar under *S. pyogenes*-infektioner.

Det mänskliga immunsystemet producerar proteiner som kallas antikroppar, som riktar sig mot olika patogena proteiner. Antikroppar bundna till patogenerna leder till att olika immunceller dödar dessa patogener. De olika *S. pyogenes*-proteinerna som antikropparna riktar sig mot beskrivs i detta arbete. Sättet hur inbindningen av antikroppar till M proteinet sker har också klarlagts.

Sammanfattningsvis hjälpte denna studie oss att förstå de olika mekanismerna som är involverade i *S. pyogenes*-infektioner. Dessutom kan de proteomik teknikerna som beskrivs här användas för att karakterisera proteinnätverk som bildas runt vilken annan patogen som helst för att bättre förstå grunden för olika infektioner.

Introduction

This section will provide the background information that is required to understand the aim and findings of this thesis. The various topics of these chapters are arranged in a progression that ranges from basic microbiology and immunology to advanced state-of-the art techniques that are used to solve different research questions. The goal of these chapters is to acquaint the readers to the different elements of the research work presented here in.

Chapter 1: Infection

"The deviation of man from the state in which he was originally placed by nature seems to have proved to him a prolific source of diseases"

-Edward Jenner[1]

In 1884, Robert Koch postulated the causative relationship between a microbe and disease[2]. These fundamental principles have guided the identification of different infectious disease causing agent[3]. The invasion and multiplication of these agents in the host might cause an infection if the immune system fails to prevent it. These causative agents of infections are termed as pathogens and the ability of such pathogens to cause infection is pathogenicity. There are four major types of pathogens; viruses, bacteria, fungi and parasites. Not only does the size of these pathogens vary but also the way they infect and cause damage to the host is different. Pathogens like most organisms are trying to survive and procreate, therefore the nutrient rich human body maintained at a uniform temperature provides a suitable environment for their survival[4]. Infectious diseases are a major cause of human morbidity and mortality and WHO has identified infections as a global concern. The problem of infections are further aggravated by the emergence of novel pathogens[5], quite obvious from the Covid-19 pandemic. The focus of this thesis is bacterial infection in the human body. This chapter provides information on how pathogens cause infection and what are the different ways of preventing and controlling an infection.

Setting up an infection

The three main steps for establishing an infection in the host are adhesion, colonization and dissemination. Adhesion is the initial and critical step where pathogens bind to host cells and interact with host molecules to initiate colonization. The pathogens then grow and multiply in number on different body surfaces followed by their invasion into different tissues to begin an infection. To achieve a successful infection the pathogen needs to overcome the immune surveillance of the host. The host tries to prevent an invasion while the pathogen counteracts to colonize and spread in the host, resulting in a continuous battle between the host and pathogen. Pathogens are capable of tilting the balance in their favor by expressing specific effector proteins also known as virulence factors. A virulence factor is a pathogenic component that allows it to cause damage in the host and at the same time keeps the pathogen viable[6]. During an infection, pathogens use these virulence factors to hijack and rewire multiple host processes that facilitates their efficient replication and spread[7]. The impact a pathogen shall have on a host is directly linked to the extent of rewiring[8]. As protein complexes and their interaction networks are the main functional modules of the cell[9], therefore disruption of these networks are a key manipulation strategy of the pathogens. This process results in the formation of host-pathogen protein-protein interactions (HP-PPIs)[7,10]. The general tendency of pathogens is to attack host proteins that have a high degree of connection with other proteins (hub) and are central to multiple interaction networks (bottlenecks) within the host protein-protein interaction (PPI) network[8,11,12]. Hub proteins are essential for an organism governed by a property called centrality-lethality rule[13,14], therefore interaction of pathogens with highly connected proteins of the host PPI network would enable the pathogen to disrupt critical processes of the host[15]. More than 5000 human proteins are targeted by different pathogens, revealing the multimeric nature of such networks formed[15]. The major host protein targets of pathogens include proteins involved in regulation of the immune system, cellular localization and those that mediate binding to host macro-molecules for efficient cell adhesion initiating efficient host-pathogen cross talks[16]. Moreover, evolutionarily conserved host proteins are preferential targets of the pathogens such that efficient HP-PPIs are sustained for a long evolutionary time-scale[16]. Thus, these HP-PPIs are dynamic and under continuous evolutionary pressure to overcome the host immune response[17-19]. Therefore, analysis of HP-PPIs would inform cellular processes altered by the pathogen, common human protein targets for different pathogens and host defense

mechanism[20,21]. Recently, there has been a tremendous increase in cataloguing of HP-PPIs but most of these studies are focused on viral infections leading to a huge scarcity of HP-PPI data for bacterial pathogens[16,22]. A quantitative characterization of HP-PPIs is critical for proper understanding of the molecular basis of bacterial infections. Not only the knowledge about the proteins involved in the interaction are important but structural insights into the pattern of these interactions are equally important to advance our understanding of bacterial infections and for the development of antimicrobial therapies. Structural characterization of the HP-PPIs on a global scale would aid the identification of therapeutic strategies that could precisely inhibit HP-PPI without disrupting the endogenous host PPI network. But, identifying HP-PPIs on a systems-wide scale is a formidable challenge and even more difficult is the structural characterization of these interaction networks[23]. Recent advancements in proteomics have enabled the quantitative and structural characterization of HP-PPIs, but so far HP-PPIs have been characterized for only a few bacterial species and the HP-PPIs for other major human bacterial pathogens yet remains to be determined. Moreover, bacterial pathogens are known to cause infections at different anatomical sites within the host thus the elucidation of HP-PPIs in different host microenvironment are of utmost importance, which also remains elusive. This thesis is focused on expanding the knowledge of HP-PPIs for a prominent human bacterial pathogen *S. pyogenes*.

The natural defense mechanism

The host immune system is the natural defense system to prevent an attack by a pathogen. The immune system is composed of different types of effector cells and molecules, which protects the body from various infectious pathogens, their toxins and effects[24]. The human immune system can be broadly classified into two types, the innate immune system and the adaptive immune system which are operational at different levels to prevent an infection. Even though there exist two different systems, they are quite interlinked with one another as many of the immune mechanisms are shared between the two systems. The different strategies adopted by the human immune system to avoid an invasion by a pathogen are discussed below.

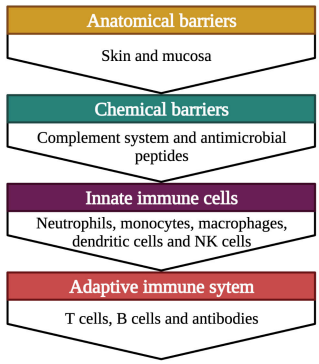


Figure 1: Different levels of the human immune system. First the skin and mucosa provides a anatomical barrier to the pathogen. Chemical barrier is the second barrier which includes the complement system and antimicrobial peptides that act on the pathogens close to the anatomical barrier. If these barriers are breached by the pathogen then the innate immune cells are rapidly activated and finally if the pathogen overcomes this then the slowly activating adaptive immune system is triggered. Figure created in Biorender

Innate immune system

The innate immune system is activated immediately *i.e.* within few minutes on exposure to an infectious agent. Anatomical barriers, chemical barriers and the innate immune cells function at different stages to activate an innate immune response.

The first line of defense comprises of skin and mucosa, which is the initial defense strategy of the humans. Skin and the mucosa are made up of highly organized epithelial cells which forms the anatomical barrier restricting the entry of pathogen from the external environment. Antimicrobial peptides (AMP) are present in the skin and mucosa which target the pathogen and facilitates their clearance in multiple different ways[25,26]. During an injury, damage of the skin might cause vascular injury leading to bleeding. Platelets are one kind of cells that are activated during the injury and plays an important role in preventing the loss of blood by the formation of blood clots[27]. The damaged skin might act as an entry point for the pathogens so along with its hemostatic role the platelets also have a role in immune response[28]. Platelets have enormous surface area and they express receptors, which allows them to bind pathogens. The immune role of platelets involves recruitment of immune cells, formation of complexes with the immune cells and complement activation[29-31]. Sometimes the pathogen can breach this physical barrier and gain access to the internal milieu, then the complement system comes in to play. The complement system consists of nearly 30 plasma and membrane associated proteins. This array

of proteins is organized into a hierarchy of proteolytic cascades that are initiated by the identification of different pathogenic components by three major pathways: the classical, alternative and lectin pathways. The classical pathway is initiated by the immune complexes formed by antibodies binding to pathogens or pathogenic components; specific carbohydrates, lipids or proteins on the pathogen surface triggers alternative pathway and the binding of either mannose binding lectin (MBL) or ficolin to the carbohydrate moieties on the pathogen surface activates lectin pathway. Once the complement system is activated this leads to the generation of proinflammatory mediators (anaphylatoxins), opsonization (coating) of the pathogen surface by different complement opsonins and finally lysis of the pathogen by the assembly of membrane-penetrating pores known as the membrane attack complex (MAC). Thus, the complement system links both the innate and adaptive immune system allowing an integrated host immune response against pathogenic challenges. Moreover, the complement system also exemplifies the complexity of HP-PPIs formed at the host-pathogen interface. As the complement system can exert a multitude of effects, the complement activation is tightly regulated by the presence of different host proteins like factor H and C4b-binding protein (C4BP)[32,33]. If the pathogen is able to overcome the anatomical and chemical barrier they then encounter the cellular defenses of the innate immune system. The innate immune cells mainly involve cells that are capable of engulfing the pathogen such as neutrophils, monocytes, macrophages and dendritic cells (DC). These immune cells express certain receptors known as pattern recognition receptors (PRRs), which recognize particular components of the invading pathogen know as pathogen associated microbial patterns (PAMPs). Moreover, the innate immune cells also express receptors that can recognize antibodies bound to pathogens leading to the opsonophagocytosis of pathogens. Natural killer (NK) cells are another type of immune cells that target and kill host cells infected with intracellular pathogens. Once these cells senses such a recognition, chemical mediators like cytokines are released which amplify the immune response by activating and attracting more immune cells[34].

Adaptive immune system

Many pathogens have evolved to circumvent the innate immune system by different strategies, primarily by preventing phagocytosis or by inhibiting complement activation. Such pathogens are then tackled by the adaptive immune system. The adaptive immune response is delayed and takes few days to activate, but is more

efficient in controlling an infection. The adaptive immune system is centered around two kinds of specialized cells: T cells and B cells[35]. The major component that triggers the activation of these cells is an antigen. An antigen is a molecular constituent of a pathogen that is capable of inducing a systemic response mediated by T cells and antibodies[36].

The antigen is presented by antigen presenting cells (APCs) like the DC and macrophages via major histocompatibility complexes (MHC) to the T cell receptors (TCR) resulting in the generation of CD8+ T cell and CD4+ T cell. These cells then either target and kill infected cells or activate other immune cells like the B cell[24,37,38].

B cells interact directly with antigens via specialized proteins known as immunoglobulins (Igs). Igs present on the B cell, called as B cell receptors (BCR) act as the cell's receptor for the antigen. The secreted form of Igs are the antibodies. The main function of B cells is to secrete such antibodies directed against pathogens. Antibodies have two major functions; one is to target the pathogens or antigens and the other is to bind to effector molecules or cells to destroy the antibody bound pathogen. Antibodies are large molecules composed of two distinct polypeptide chains; the heavy (H) chain and the light (L) chain. The two heavy chains are linked to each other by disulfide bonds and the heavy chain is linked to the light chain by a disulfide bond in a manner that it gives a Y shape to the antibody. The ends of the two arms of the Y recognize and bind antigens, and the stem of Y is involved in interaction with effector molecules or cells. The region of the antibody involved in the recognition of the antigen is referred to as paratope, and the antigenic region recognized by the antibody is the epitope. The two heavy and light chains are identical in an antibody, giving rise to two antigen-binding sites, thereby increasing the strength of interaction. There exist two types of light chains, lambda (λ) and kappa (κ); while there are five types of heavy chains denoted by α, δ, ε, γ and μ. The heavy chain type defines the class and the effector functions of the antibodies. Based on the heavy chain there are five major classes of antibodies namely IgA (α), IgD (δ), IgE (ε), IgG (γ) and IgM (μ). Antibodies contain discrete domains of around 110 amino acids; the light chain has two domains and the heavy chain has four domains. The amino acids of the first domain from the N-terminal end is highly variable giving rise to the variable domain (V domain) of the heavy and light chains (V_H and V_L), which determines the antigen-specificity of the antibodies. The remaining domains of a particular antibody class is constant and it forms the constant domain (C domain) of the heavy and light chain (C_H and C_L). The multiple constant domains of the heavy chain are numbered as C_H1, C_H2 and C_H3

respectively. Proteases such as papain can cleave the antibodies into variable fragment antigen binding (Fab) and constant fragment cystallizable (Fc) fragments. The Fab part comprising of the V_H, V_L, C_H1 and C_L domains possess the antigen-binding activity. The Fc part contains C_H2 and C_H3 domains that binds to the Fc receptor (FcRs) on immune cells, phagocytes and the complement components to induce antibody dependent cellular cytotoxicity (ADCC), opsonophagocytosis, and complement activation[24,39].

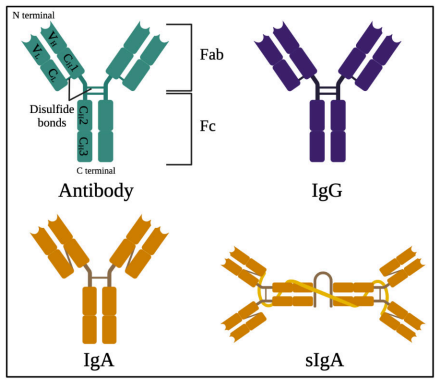


Figure 2: Antibody structure. Basic structure of an antibody depiciting different variable and constant domains, the disulfide linkage between the chains and the Fab and Fc part of the antibodies; and the strucutre of IgG, IgA and sIgA. Figure created in Biorender

IgM is the first antibody produced by the B cells, but a very small amount of IgM is found in the plasma. The production of IgD is very low, while IgE is also present in small amounts in plasma and is mainly responsible for allergic reactions. IgG and IgA are the two major contributors of the human antibody response. The predominant antibody in plasma is IgG. IgG is monomeric and there are four different subtypes of IgGs: IgG1, IgG2, IgG3 and IgG4 of which IgG1 is the most abundant. IgG's bind to pathogens and cause neutralization of the pathogen and toxins, opsonization and phagocytosis, they initiate ADCC to kill pathogens and also activate the complement system[24,39]. IgA exists both in monomeric and dimeric forms. IgA is monomeric in plasma and the dimeric form of IgA is the secretory IgA (sIgA) which is present in the mucous secretions like tears, sweat, saliva and colostrum. sIgA comprises of two IgA monomers connected by J chain and a secretory component (SC). Pathogens coated with IgA are eliminated via phagocytosis or ADCC. sIgA is the major driver of the mucosal immunity in the

respiratory tract, gastrointestinal tract and urogenital tract. sIgA plays an important role in the maintenance of commensal bacteria and also prevents systemic infection by blocking the invasion of pathogenic and commensal bacteria across the mucosal surfaces[40-42]. There are two IgA subtypes: IgA1 is prevalent in serum and IgA2 is mostly present in secretions[43]. According to the concept of "original antigenic sin", the first infection by a pathogen plays an important role in imprinting the antibody response against that pathogen and orienting the immunological memory[44]. However, antibodies generated by an initial infection are not always protective as they target non-protective domains of the antigens. Moreover, these non-protective antibodies interfere with protective antibodies generated by vaccines thereby dampening the protective antibody response, as evident for *Staphylococcus aureus* infection and its vaccine failures[45].

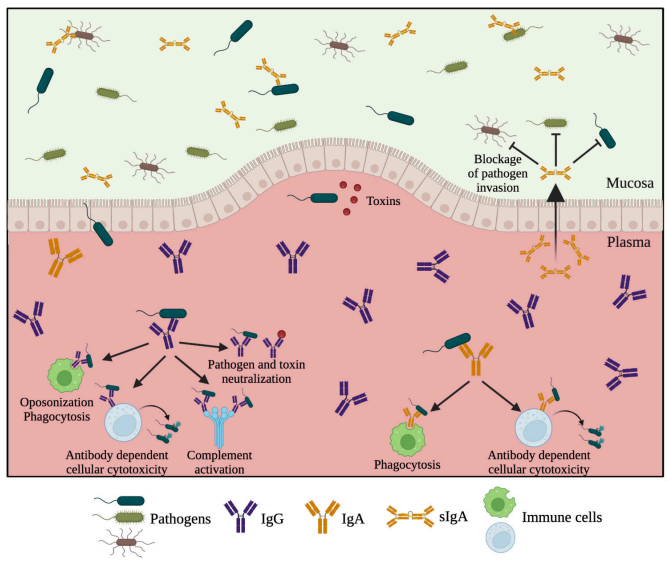


Figure 3: The role of IgA and IgG in different microenvironment. sIgA is the main antibody in the mucosa and it inhibits the entry of pathogens into the systemic circulation. IgG is the predominant antibody in the plasma. IgG causes neutralization of the pathogen and toxins, complement activation, antibody dependent cellular cytotoxicity, opsonization and phagocytosis of pathogens. IgA is also present in the plasma and it mainly causes phagocytosis and antibody dependent cellular cytotoxicity. Figure created in Biorender

Controlling infections: problems and solutions

Sometimes the immune system fails to control the invasion and spread of a pathogen which results in an infection. Then to control an infection different antimicrobial agents are administrated, such as antibiotics to treat bacterial infections[46]. However, most of the infectious agents are becoming less susceptible to these treatment strategies resulting in the emergence of antimicrobial resistance (AMR). AMR is a major threat to humans all around the world[47], therefore new measures are required to control the rapidly growing AMR. Vaccines have been successful in eradicating many viral and bacterial infections that were cause of millions of human deaths[48], thus proving vaccines to be an effective health intervention[49,50]. Vaccines are biological formulations containing attenuated, heat killed or genetically modified pathogens or pathogen derived components[51,52], that are administered to evoke host specific antibody and cellular immune response. The immune response triggered by the vaccine agent prepares the host to prevent itself from further infection of that particular pathogen. The prophylactic use of vaccines prevents pathogens from achieving large population size within the host, thereby reducing the chance of onward transmission[53-55]. The mode of administration and mode of action makes vaccines less susceptible to resistance. There has been rare instances of vaccine resistance, but still the pronounced health benefits of vaccine were achieved due to the preventive nature of vaccines, its durable protective effect and herd immunity[56]. Therefore the use of vaccines is one way to effectively tackle the problem of AMR [56]. More than 60 licensed vaccines have been successful in controlling different infections, still effective vaccines are urgently required for other serious pathogens[57]. Particularly vaccines are required against multiple bacterial pathogens because there has been a tremendous increase in bacterial AMR as evident by 4.95 million deaths globally in 2019 alone due to bacterial AMR[58]. This thesis is centered around one such bacteria, *S. pyogenes* for which no vaccine is yet available. Recently, new technologies are being developed to accelerate vaccine development. One such method is reverse vaccinology. Reverse vaccinology uses genomic information to predict potential vaccine candidates, which can then be tested in both *in vitro* and *in vivo* pre-clinical models[56]. Reverse vaccinology has been successful in the development of a vaccine against group B meningococcus[59]. High throughput screening of proteins using mass spectrometer has further expanded the scope of reverse vaccinology.

Chapter 2: *Streptococcus pyogenes*

"*not only were streptococci among the first organisms to be presumed to be the cause of contagious diseases, but also their existence forced the introduction of hygiene and asepsis into hospital wards*"

- M.Schwartz[60]

S. pyogenes or Group A *Streptococcus* (GAS) was first identified by Rebecca C. Lancefield. *S. pyogenes* results in complete hemolysis forming a clear zone when grown on blood agar plate and thus are categorized as beta-hemolytic bacteria. This Gram-positive bacterium only infects human. *S. pyogenes* are a prominent cause of death worldwide and is ranked ninth amongst other infection causing pathogens[61,62]. It causes a broad spectrum of human diseases and is one of the most versatile human pathogen[63]. The diseases caused by *S. pyogenes* ranges from mild throat infections (pharyngitis or commonly known as "strep throat") to severe invasive disease like the streptococcal toxic shock syndrome (STSS). There are more than 220 different serotypes of the bacteria. The primary reservoirs of *S. pyogenes* in humans are the nasopharyngeal mucosa and skin. *S. pyogenes* are known to asymptomatically colonize these sites and remain viable for long time. The bacteria are transmitted from asymptomatic carriers or infected patients generally by respiratory droplets or skin to skin contact. Small outbreaks of *S. pyogenes* infections were mainly caused by person-to-person spread in crowded settings, hospital-acquired infections or consumption of contaminated foods[64-66]. However, large *S. pyogenes* outbreaks were mainly caused by the emergence of new virulent clones that have undergone multiple genetic changes[67-69]. Large *S. pyogenes* outbreaks have been reported in Canada, North America, Hong Kong and China[70-73]. This chapter gives an overview of the different diseases caused by *S. pyogenes*, its virulence factors and vaccine strategies to control *S. pyogenes* infections.

Disease manifestations and burden

S. pyogenes is known to cause a wide array of mild and serious diseases in humans. Majority of the benign infections involve superficial infections of the throat and skin like pharyngitis, scarlet fever and impetigo[74,75]. *S. pyogenes* is the most common cause of bacterial pharyngitis[61]. The bacterium also possesses the ability to penetrate human epithelial surfaces and enter into systemic circulation thereby causing many invasive diseases like STSS, puerperal sepsis, meningitis and endocarditis. *S. pyogenes* are also known as the "flesh eating" bacterium, as they are able to invade skin and soft tissue and cause necrotizing fasciitis, which in worst cases can destroy the limb or tissue. *S. pyogenes* is also responsible for serious post streptococcal infection sequelae, such as the acute post streptococcal glomerulonephritis (APSGN), acute rheumatic fever (ARF) and rheumatic heart disease (RHD). Pharyngitis is very common in school-aged children and the infection triggers antibody repertoires capable of preventing further re-infection[76-79]. Despite the presence of antibody titers, there are incidences of serious *S. pyogenes* infections and the basis for immunity to *S. pyogenes* infections following a natural infection still remains poorly understood.

One of the major cause of morbidity and mortality globally in the twenty-first century is *S. pyogenes* infections. In the mid-twentieth century, there was a drop of *S. pyogenes* disease burden in industrialized countries, however in the last fifty years there has been an upsurge in *S. pyogenes* infections and outbreaks[65,80,81]. The estimated global incidence of pharyngitis is more than 600 million/year and around 111 million for impetigo. Globally *S. pyogenes* causes around 660,000 invasive infections each year with a death rate of 160,000. The prevalence of RHD is 15.6-19.6 million and that of ARF and APSGN is around 470,000/year. RHD and ARF are one of the major causes of children mortality and disability worldwide[61,74,82].

M protein

To colonize, disseminate and initiate infections in humans, *S. pyogenes* have evolved to express multiple secreted and surface attached virulence factors. Some

of these virulence factors are involved in multiple resistance mechanisms, one of them being the M protein. M protein, the major surface attached virulence factor of *S. pyogenes* is known to enable bacterial invasion, resist phagocytosis and promote bacterial survival in the host[83-86]. It is a fibrillar α-helically coiled-coiled dimeric protein covalently attached to the cell wall by the LPxTG motif and extending nearly 500Å from the bacterial surface[82,87]. M proteins can also be cleaved from the bacterial surface by streptococcal pyrogenic exotoxin B (SPEB)[88]. The protein consists of multiple domains and some of these domains are repeats. The first 50 amino acids (aa) from the N-terminal end of the protein is highly variable, and this part of the M protein is known as the hypervariable region (HVR)[89,90]. M proteins are classified based on the variation in the HVR and based on this classification more than 220 different serotypes of *S. pyogenes* have been reported[87]. After the HVR region the M protein consists of A domain, B repeats, C repeats and D domain. The 100-150 amino acid stretch after the HVR forms the semivariable region of the M-protein and encompasses A domain and B repeats. The conserved C-terminal part of the M protein consists of the C repeats and D domains.

M protein is encoded by the *emm*-gene. Based on the arrangement of different domains in the M proteins and occurrence of the *emm* genes in *S. pyogenes* genomes, M proteins are classified into three main types A-C, D and E. The A-C type consist of long M proteins comprising of all the four domains *i.e.* A, B, C and D. Strains in the A-C type are mainly responsible for throat infections. D type includes M proteins with B, C and D domains and the strains expressing D type M proteins are responsible for skin infections. M proteins belonging to the E type are shorter comprising of only the conserved C and D domains. Strains expressing the E type M protein are generalist strains infecting both throat and skin[91,92]. Based on comparative sequence analysis of the M proteins, M proteins have also been classified into two main clades X and Y. X consist of the E type and Y includes the A-C type while D type is distributed in both the clades. Common *S. pyogenes* strains such as M28, M49 and M89 belong to clade X and M1, M3 and M5 belong to clade Y[93].

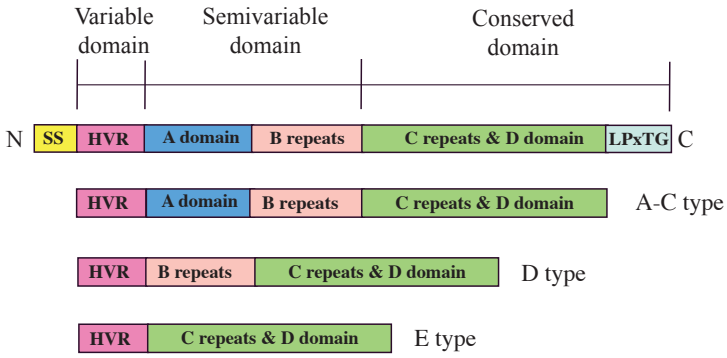


Figure 4: Structure of naïve and mature M protein and different types of M protein. Reproduced from *Chowdhury et.al., (2021)- "Streptococcus pyogenes Forms Serotype-and local Environment-Dependent Interspecies Protein Complexes", mSystems, vol 6, issue 5*. SS- signal sequence, cleavage of SS results in mature M protein and LPxTG domain anchors the M protein to the bacterial surface. SS and LPxTG is only shown for one M protein and has been omitted for the other M proteins. HVR- hypervariable region, A domain and B repeats are part of the semivariable domain and the C repeats and D domain are part of the conserved domain. A-C type M protein have A domain, B and C repeats and D domain; D type lacks A domains and contains B and C repeats and D domain; E type have only C repeats and D domain. M proteins are not drawn to scale

M protein interacts with a lot of host proteins, which promotes the survival and dissemination of *S. pyogenes* in the host[94]. Fibrinogen binds to the B repeats on the M protein, which prevents binding of opsonizing antibodies and inhibits phagocytosis[90,95]. Binding of the M protein to fibrinogen masks the antigenic sites of the M protein thereby making it difficult to opsonize the bacteria[96]. Albumin binding to the C repeats of the M protein helps the bacteria uptake fatty acids and survive in the stationary phase[97,98]. Additionally, M proteins can bind IgG[94] and IgA[99]. The orientation of IgG binding to *S. pyogenes* surface is dependent on the concentration of IgG *i.e.* in an IgG poor environment like saliva the interaction is Fc mediated, while in an IgG rich environment like plasma, IgG's bind via Fab[100]. The M protein also binds the C4BP at the N-terminal end therefore preventing bacterial clearance via complement pathways[101,102]. Factor H binding to M protein prevents complement deposition thereby leading to reduced phagocytosis[103]. M protein being of the key antigen of *S. pyogenes* has been the major candidate for vaccine attempts so far.

Other virulence factors

In addition to the M protein, *S. pyogenes* expresses a plethora of virulence factors which helps the bacteria evade host immune responses in multiple ways like preventing phagocytosis, altering the complement activation, degrading antibodies and inhibiting neutrophil killing. Some of the other important virulence factors of *S. pyogenes* and their functions are briefly described here.

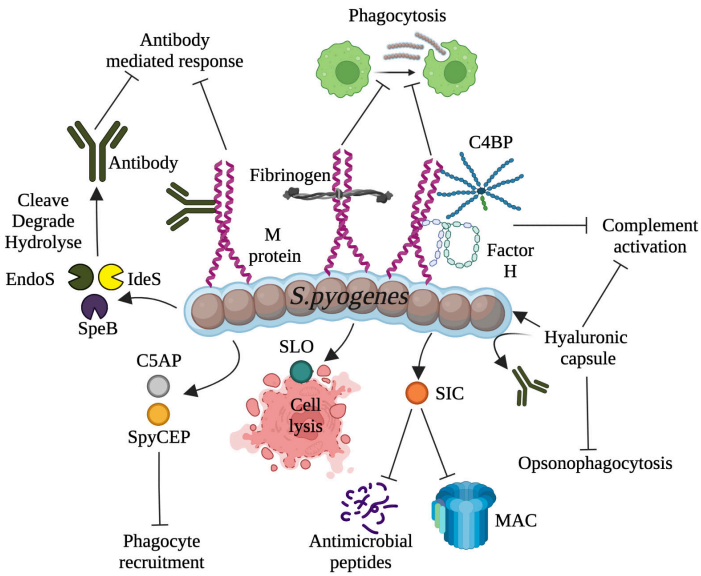


Figure 5: Different virulence factors expressed by *S. pyogenes* and their role in immune evasion. M protein binds C4BP, factor H, fibrinogen and Fc domains of antibodies to inhibit complement pathway, phagocytosis and antibody mediated immune response. The hyaluronic capsule blocks antibodies preventing complement activation and opsonophagocytosis. SIC inhibits MAC formation and antimicrobial peptide activity. SLO causes lysis of host cells. Secreted virulence factors C5AP and SpyCEP hinders phagocyte recruitment. EndoS, SpeB and IdeS alters antibody structure thus preventing antibody mediated immune response. Figure created in Biorender

- *Hyaluronic capsule*

Almost all of the *S. pyogenes* strains are encapsulated within a hyaluronic acid capsule. This capsule acts like a surface shield thereby blocking the access of antibody to the surface antigens, inhibiting complement activation, preventing opsonophagocytosis and leading to increased survival in the neutrophil extracellular

traps (NETs)[104-107]. Invasive strains of *S. pyogenes* were found to be hyperencapsulated[108].

- *Streptococcal inhibitor of complement-mediated lysis (SIC)*

This is a secreted virulence factor. SIC binds the C5b57 complement complex preventing its uptake on the cell membrane thus inhibiting the formation of MAC[109,110]. Other roles of SIC involve mucosal colonization, inhibition of secretory proteinase inhibitor, lysozyme and antimicrobial peptides[111-113].

- *Streptolysin O (SLO)*

SLO is a toxin which the bacteria secrete extracellularly to target different immune cells. It is a cholesterol dependent cytolysin and is lytic to almost all cholesterol containing cells. The beta-hemolysis caused on blood agar plate by *S. pyogenes* is due to SLO. It disrupts the host cell membranes by forming large pores on them leading to apoptosis of neutrophils, macrophages and epithelial cells[114-116]. SLO induced pores on the host cells act as entry point for another toxin nicotinamide dehydrogenase (NAD-glycohydrolase) which exhausts the host cell energy causing increased tissue damage[117].

- *Immunoglobulin modulating enzymes*

S. pyogenes produces two major enzymes, namely IgG-degrading enzyme of *S. pyogenes* (IdeS) and endo-beta-N-acetylglucosaminidase of streptococci (EndoS), which target the IgGs and alter their structure. IdeS possess cysteine protease activity and it cleaves the lower Fc region of the surface attached IgG. The cleaved IgG's are then not recognized by phagocytes leading to inhibition of opsonophagocytosis[118]. EndoS hydrolyzes the glycan in the heavy chain of IgG, preventing Fc receptor binding, complement activation and phagocytosis[119,120].

- *Streptococcal pyrogenic exotoxin B (SPEB)*

SPEB is a cysteine protease. This protease has a broad activity and can degrade multiple host proteins like immunoglobulins, cytokines, chemokines and complement proteins as well as AMPs[63,121]. SPEB can degrade IgA, IgD, IgE, IgG and IgM[119]. This virulence factor is also responsible for colonization and dissemination in the host. SPEB can effectively cleave the M protein from the streptococcal surface[88].

- *C5a peptidase (C5AP)*

It is a proteolytic enzyme expressed on the surface of *S. pyogenes*. C5AP is a highly specific serine peptidase. This peptidase cleaves the chemotaxin C5a[122]. C5a is a potent chemoattractant playing an important role in the recruitment and stimulation of neutrophils. C5AP inactivates C5a and its chemoattractant activity is lost, thereby impairing the recruitment of phagocytes at the infection site[123].

- *S. pyogenes cell envelope proteinase (SpyCEP)*

SpyCEP is a peptidase virulence factor. The expression of SpyCEP was found to be upregulated during invasive infections. SpyCEP degrades one of the potent chemokine IL-8. IL-8 is important for neutrophil recruitment at the site of infections. Therefore, the cleavage and inactivation of IL-8 inhibits neutrophil mediated killing of the bacteria[124,125].

- *Superantigens*

S. pyogenes are known to express numerous superantigens, which include the streptococcal pyrogenic exotoxin family (SpeA, SpeC, SpeG-SpeM), streptococcal superantigen A (SSA), streptococcal mitogenic exotoxin Z (SMEZ). These superantigens are responsible for STSS[126]. Superantigens interact simultaneously with MHC class II on antigen presenting cells and T cells leading to the release of cytokines. The elevated cytokine production causes increased inflammatory response, tissue necrosis and multi-organ failure[82,127].

S. pyogenes HP-PPIs

It is very evident that *S. pyogenes* interacts with multiple human proteins. So far, these interactions have mostly been studied as single entities. But within the host, all these interactions occur simultaneously resulting in a network of complex interactions formed around the bacterial surface and its different virulence factors. Therefore, it is vitally important to study all the *S. pyogenes*-human protein interaction networks together on a global scale to better understand the rewiring mechanisms adopted by *S. pyogenes*. As mass spectrometry is capable of characterizing and quantifying dynamic protein networks formed at the pathogen and host interface, therefore large-scale analysis of HP-PPIs for *S. pyogenes* has

been carried out to some extent using mass spectrometry[94,128,129]. The HP-PPIs around *S. pyogenes* was found to be highly dense and organized, and the absence of M proteins altered this network to a great extent[129]. Plasma proteins were highly enriched at the bacterial surface and the interaction profile differed between an invasive and non-invasive *S. pyogenes* strain[128]. Additionally, the protein interaction networks varied between saliva and plasma revealing the tendency of *S. pyogenes* to sequester plasma proteins even at low concentrations[94]. However, knowledge regarding the network complexity of human proteins interacting with different *S. pyogenes* serotypes is still lacking. Moreover, the dynamics of these protein interactions mediated by different virulence factors at the human - *S. pyogenes* surface needs to be addressed to better understand the role of HP-PPIs in facilitating *S. pyogenes* to cause both localized and invasive infections. Finally, atomic resolution of the structure of these protein networks is the unmet need to gain mechanistic understanding of *S. pyogenes* infections. In this work, different proteomics techniques have been used to quantitatively and structurally characterize the HP-PPIs for *S. pyogenes* to unravel the properties and functional classes of human proteins interacting with the bacterium in different ecological niches and determine its role during infection.

Streptococcal antigen-ome

Antibodies function as immune complexes targeting distinct antigens. High and low affinity antibodies target different antigens driving distinct effector functions such as complement activation, toxin neutralization, cellular toxicity and phagocytosis. Different levels of antibodies are also generated against various pathogenic proteins. Antibodies have been reported to be a predominant interacting partner of the *S. pyogenes* HP-PPIs[94]. Therefore, quantitative analysis of the streptococcal antigen-ome will aid in the ranking of antigens based on their ability to evoke antibody responses in the host and will also be critical in understanding the pathogenesis and for the development of therapeutic approaches. Here we define the antigen-ome as a set of pathogenic specific proteins targeted by host antibodies. Few attempts have been made to define the *S. pyogenes* protein targets of the host antibodies. Genomic library screening conducted with serum from patients recovered from common *S. pyogenes* infections and healthy individuals identified a total of 95 *S. pyogenes*

immunogenic proteins[130]. Using genomic libraries, prediction of the sub-cellular localisation of pre-dominant immunogenic proteins becomes difficult. Furthermore, bioinformatics tools were used to select proteins which were then probed with sera from children with *S. pyogenes* pharyngitis leading to the identification of 26 antigens[131]. Using such an approach the determination of the quantitative ability of different antigens to evoke antibody response in the host becomes difficult. Intravenous immunoglobulin (IVIG) has been used to determine *S. pyogenes* surface antigens recognised by pooled human immunoglobulins. The study resulted in the identification of 94 *S. pyogenes* antigens[132]. Many of the antigens identified in these studies overlapped, which included *S. pyogenes* virulence factors like M1, SLO and C5AP[130-132]. Therefore, based on these findings further studies should be performed to quantitatively compare the immune response of individuals during an *S. pyogenes* infection and post-infection to better understand the host immune response and define the antigenic repertoires. To suffice this purpose, we develop a quantitative mass spectrometry based strategy to identify streptococcal antigens evoking antibody response and denote the sub-cellular localization of those antigens. Moreover, the method is also capable of quantifying the antibody responses evoked by different antigens in a host specific and time resolved manner.

Vaccines

A safe and effective vaccine is yet to be developed against *S. pyogenes*[133]. The major challenges encountered in the development of a vaccine include serotype diversity, differences in geographical distribution of *S. pyogenes* serotypes and burden, antigenic variation and similarity of *S. pyogenes* antigens with host proteins causing increased possibility of autoimmune diseases[92,134-137].

M protein has been the major focus of *S. pyogenes* vaccinology. The M protein has been found to be immunologically cross reactive with cardiac myosin[138], therefore the HVR of M protein has been investigated as a vaccine candidate to avoid auto-immune diseases. To obtain a global coverage of the vaccine against multiple different serotypes of *S. pyogenes*, fragments from different M proteins have been incorporated to form multivalent M vaccine. The multivalent attempts include bivalent, tetravalent, hexavalent, octavalent, 26-valent and 30-valent vaccines[139-144]. The hexavalent and 26-valent M vaccines have reached human clinical trials[145,146]. These multivalent vaccines have been effective against few serotypes, whose M

protein fragments were not included in the vaccine but not against all globally circulating serotypes, particularly the endemic serotypes. To overcome the serotype specific concern, an alternative approach of using the conserved C-terminal end of the M protein as a vaccine candidate was adopted. Vaccine design with minimum peptide residues from the conserved C-terminal end (J8 and J14) has been shown to be effective in mice[147,148]. Additionally, a 55 amino acid sequence of the conserved C-terminal end of M protein named StrepInCor has been shown to encompass the human T and B cell epitope and was able to induce a robust protective immune response in mice[149]. Studies are underway for StrepInCor to verify its coverage for different *S. pyogenes* strains and to evaluate its ability in inducing a good mucosal immune response.

Some other *S. pyogenes* antigens that have been studied for vaccine targets involve group A carbohydrate (GAC), C5AP, SPEB and SpyCEP[150-153]. The major problem with such antigens are cross reactivity with host protein or toxic and enzymatic effects on host cells. Trigger factor (TIG), a surface attached protein required for the maturation and secretion of SPEB has been identified to confer protection in mice[154-156]. It is a highly conserved protein and lacks significant amino acid sequence similarity with human proteins; therefore, further investigation should be carried out on TIG as a potential vaccine candidate.

Reverse vaccinology supported by approaches like genomics, proteomics, bioinformatics and protein arrays are recently being used to identify potential novel *S. pyogenes* vaccine antigens. High throughput multigenome data mining coupled with proteomics identified 52 probable *S. pyogenes* vaccine candidates for further investigation[157]. To determine surface exposed vaccine candidate's, *S. pyogenes* were treated with trypsin to shave the surface proteins which were then identified by proteomics and then tested in mice models for protection[158]. Novel immunogenic proteins were identified by genomic surface display libraries and by antibodies from human serum exposed to or infected with *S. pyogenes*. Amongst these proteins, nine highly conserved protective antigens were identified which are currently being evaluated for their protective efficacies as individual candidates or as a combination[130]. A strategy including three high throughput technologies *i.e.* mass spectrometry based proteomics, protein array and flow cytometry was parallelly applied to determine *S. pyogenes* vaccine candidates. This resulted in the generation of a combination of three protective antigens in mouse models[131]. Such approaches combined with traditional methods, therefore could accelerate the development of a successful anti-*S. pyogenes* vaccine.

Chapter 3: Proteomics

"*The main strength of proteomics lies not in a massive cataloguing of protein spots, but in allowing proteins to be placed into functional classes by following changes in their expression and post-translational modification levels*"

- Peter James[159]

Proteins play an important role in catalysing and controlling all essential cellular processes. Proteins account about half of the total dry mass of a cell[160]. Proteome refers to the complete set of PROTEins encoded by the genOME. The term proteome was first coined by Marc Wilkins[161]. Proteins are coded by genes, but the number of proteins can be more than the genes because different forms of proteins are produced due to alternative splicing or post-translational modifications. Moreover, the proteome abundance and expression in an organism is cell or tissue specific and can vary in time under different conditions. Therefore, studying proteomes will help us determine and characterize protein expressions in different cells or tissues thereby providing important biological information. The study of proteins is termed as proteomics. One of the most commonly used technique to study, analyse and measure proteins is mass spectrometry (MS). For instance, MS was used to draft the human proteome map[162,163]. Recent advancements in proteomics has also resulted in the use of MS to elucidate the structure of proteins. This chapter is focussed on MS based proteomics and how it is being used to identify and measure proteins (quantitative proteomics) and define protein structures (structural proteomics).

Mass spectrometry

Mass spectrometry is a technique to measure the mass to charge ratio (m/z) of ionized molecules using a mass spectrometer. The result from a MS analysis is an intensity versus m/z chromatogram. There are two main approaches in MS-based proteomics to study proteins: top down proteomics and bottom up proteomics. In top down proteomics, intact proteins are measured. Even though top down proteomics provides better characterization of proteins, its use is limited because of several technical limitations such as difficulty in separation of intact protein mixture and unavailability of efficient methods for fragmenting large proteins[164]. Bottom up proteomics is the most commonly used proteomics method. In bottom up proteomics proteins are digested into peptides using different proteases and the resulting peptides are then analyzed in the mass spectrometer. The peptides generated in bottom up proteomics are further dissociated into fragments and this is known as tandem mass spectrometry (MS/MS)[164,165]. The peptide masses and sequence are then used to map the proteins. Bottom up MS/MS approach has been adopted for all protein analysis in this thesis.

Workflow of Bottom up proteomics

A typical bottom up proteomics pipeline consists of several sample preparation steps to ensure optimum digestion of proteins to peptides; and an online liquid chromatography (LC) system coupled to a mass spectrometer (LC-MS) for separation of peptides followed by their identification and quantification. To begin with, the proteins in the sample are denatured to linear conformation exposing the protease cleavage sites. Commonly in bottom up proteomics the protein is digested into peptides by trypsin[166]. Prior to MS analysis the peptides are separated on a solid phase to reduce the amount of salts in the peptide mixture. These peptides are then separated on the LC system operated in a reverse phase-liquid chromatography mode (RP-LC), where the peptides are separated based on hydrophobicity. A gradient of increasing percentage of hydrophobic/hydrophilic solvents are used to elute the peptides resulting in the elution of polar peptides first followed by the elution of non-polar peptides. When these peptides reach the tip of chromatographic column they are ionized into ions by the ion source. Electron spray ionization (ESI) is commonly used to ionize peptides in solution and is therefore easily coupled to the LC system[167]. The ions then enter the MS. In the MS, the mass analyzer measures

m/z of the ions and the detector records the number of ions with a particular m/z value[168], resulting in the generation of a MS chromatogram. For MS/MS there is an additional component in the MS for trapping the ions and breaking them into fragment ions by different energy sources such as electron capture dissociation (ECD), electron transfer dissociation (ETD), collision induced dissociation (CID) and high energy collision dissociation (HCD). The MS used for protein analysis in this thesis uses HCD[169]. One important consideration in bottom up proteomics is that the MS measures peptides and not proteins. Bioinformatics software's are used to annotate the spectra and search against a protein database to map the peptides. These peptides are then finally mapped to their parent proteins using different bioinformatics tools. Tracing back peptides to the protein is not straightforward as peptides originating from different proteins might be same *i.e.* non-proteotypic peptides, therefore determining their parent protein becomes difficult. Typically, unique peptides originating from a protein *i.e.* proteotypic peptides are considered in MS analysis. MS can be operated in different ways to address different research objectives.

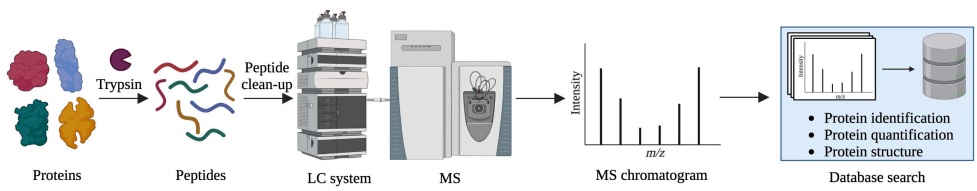


Figure 6: Schematic of bottom up proteomics. Proteins are digested into peptides using trypsin, then the peptide mixture is cleaned to reduce contaminants like salts. Peptides are then separated in the liquid chromatography (LC) system based on hydrophobicity before injecting into the mass spectrometer (MS). Analysis of peptides in the MS results in a MS chromatogram which is then searched against a database to identify proteins, quantify proteins or determine protein structure. Figure created in Biorender

Quantitative proteomics

Determining the presence of a protein in a sample is important but equally important is reporting the abundance of the protein. There are different approaches to quantify proteins using bottom up proteomics. Two of these approaches have been used to quantify proteins in this thesis and they are described below.

Data dependent acquisition (DDA)

DDA is a discovery or shotgun proteomics approach. The main aim with DDA is to identify maximum number of proteins. In DDA the m/z of all the eluting peptides (precursor ions) are measured by the mass analyzer at the full scan level (MS1). Then a predefined number of precursor ions, such as the top '*n*' precursor ions are selected and fragmented into fragment ions and the m/z is measured at MS2 level[170]. A dynamic exclusion list is generally created to avoid repeated measurements of the same precursor ions[171]. Different bioinformatics tools are then used to annotate the generated MS spectra (experimental spectra) to a peptide sequence. This is done by *in-silico* digestion of a query protein database by the same protease used for the experiment leading to the generation of a theoretical database containing predicted spectra. Then the experimental spectra and predicted spectra are matched to determine the presence of peptides in the sample of interest. In such matching, there might be chances of false identifications. To reduce this, a false discovery rate (FDR) cut off is set. For the statistical control a decoy database of proteins are created wherein the amino acids are either interchanged or reversed. Then the theoretical database and the decoy database is combined to generate a composite database. The experimental spectra are then searched against this composite database[172]. Match with the decoy sequences should be rare and random and a cut off score of generally 1% FDR is set, meaning only 1% false identifications would be considered. Quantification takes place at MS1 level by calculating the area under the curve of MS1 peak. As the identification of the peptides in DDA is based on predominant precursor ions, there is a chance of missing the peptides arising from low abundant proteins.

Data independent acquisition (DIA)

DIA also known as Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH), was introduced to overcome the limitations of DDA[173]. In DIA all precursor ions are fragmented, resulting in accurate and consistent in-depth proteome coverage. A predefined m/z range is set *i.e.* 400-1200 m/z and all the precursor ions in this range are scanned and MS1 spectrum is recorded. This dynamic range of 400-1200 m/z is mostly divided into 32 windows of uniform width *i.e.* 25 m/z with a slight overlapping width. The dynamic range can also be divided into windows with variable widths. All precursor ions from these windows are then scanned, isolated and fragmented to generate a large number of MS2 spectra[173,174]. Quantification in DIA takes place at MS2 level and the peptide intensities are then used to report the abundance of protein. The mass spectra generated from a DIA run

is highly complex as it contains fragment spectrum of all the precursor ions. This makes it very difficult to assign fragment ions to a particular precursor. One way of deconvolution of DIA spectrum is by using assay libraries; recently library free approaches are also being adopted for DIA data analysis. Assay libraries are generated from prior DDA runs which contains information about precursor and fragment ion m/z values, retention time and their intensities[175]. FDR cutoff is used to distinguish true hits from false hits. In DDA, the precursors are stochastically selected and thus not all peptides are detected in DDA. Peptides identified in DDA will only be included in the assay library. Therefore, the DIA data analysis is complex and is limited by the number of peptides included in the assay library.

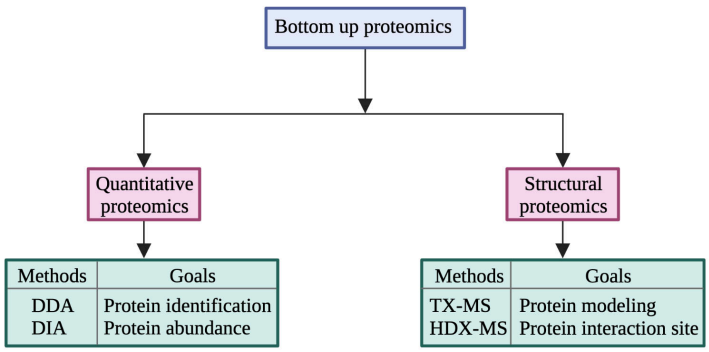


Figure 7: Different bottom up proteomics techniques used in this thesis. DDA and DIA were the two quantitative proteomics methods used to identify proteins and report their abundance. Strcutural proteomics techniques involved TX-MS and HDX-MS, used for protein modeling and mapping protein interaction sites. Figure created in Biorender.

Structural proteomics

Structural proteomics plays an important role in understanding protein-protein interactions, mapping protein binding interface and modeling protein structures. Two complementary structural proteomics strategies operating at the bottom up level used to map protein interactions in this thesis are discussed below.

Cross-linking-mass spectrometry (XL-MS)

In XL-MS a chemical cross-linker is added to the protein of interest to obtain structural information. The cross-linker is a chemical moiety with two reactive

groups separated by a fixed distance. These reactive groups covalently interact with certain amino acids of the proteins separated apart by the distance between the reactive groups of the cross-linker. The peptides containing the cross-linked amino acids can then be identified by MS. As cross-linking reactions can be carried in *in vitro* settings mimicking *in vivo* like conditions (such as live bacteria or human plasma), therefore the native conformation of protein or protein complexes can be preserved. Moreover, due to the covalent connection of interacting partners weak and transient interactions can also be captured using XL-MS[176]. XL-MS has been successfully used to investigate individual proteins as well as complex protein mixture[177-179]. The information from such cross-linking experiments provides valuable insights into the protein quaternary structure, protein domain and subunit composition, protein-protein interactions, direct and indirect binding and binding interfaces. XL-MS experiments also provide information regarding distance constraints which can be used for protein modeling. Recently, a new concept of targeted chemical cross-linking-mass spectrometry (TX-MS) has been proposed, where XL-MS has been combined with computational structural modeling approaches to build structural models of complex protein-protein interactions[180]. XL-MS are also used in an integrative fashion along with other structural techniques like HDX-MS and cryo-EM to solve complex structures[176,177,181].

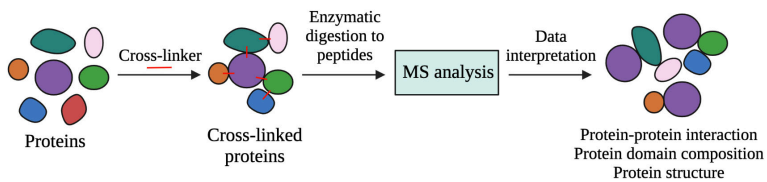


Figure 8: Overview of XL-MS workflow. Cross-linkers are added to the proteins. Proteins within the distance constraint of the corsslinker are cross linked. The proteins are digested to peptides followed by MS analysis. Data interpretation results in the identification of protein-protein interaction, knowledge about protein domain composition and protein structure. Figure created in Biorender.

Hydrogen deuterium exchange-mass spectrometry (HDX-MS)

HDX-MS is based on a chemical modification where the protein amide backbones are labelled with deuterium from D_20 depending on solvent accessibility[182] . HDX-MS is used to map protein interaction sites, conformational changes, ordered and disordered regions of the protein. To determine the binding site between proteins, a differential comparison of hydrogen deuterium exchange rate of peptides from the

protein alone versus the protein mixture is carried out. First, the individual proteins are dissolved in deuterated water; in doing so, the accessible hydrogens from the amide backbones are exchanged with deuterium, which is tracked by a MS. Such an exchange is only possible if the hydrogens are exposed, but if there is a protein-protein interaction then such exchange is not possible. Therefore, when the protein mixture is dissolved in deuterated water the number of deuterium uptake will reduce due to non-accessibility of sites because of protein interactions. A comparison of deuterium incorporation to the peptides in the unbound and bound state will facilitate the identification of protein binding sites[183,184]. HDX-MS has also been used for epitope mapping[182,183].

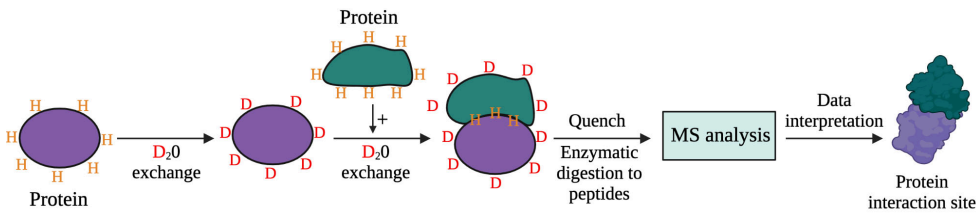


Figure 9: Overview of HDX-MS workflow. First the individual proteins are dissolved in D_20 and then the protein mixture is dissolved in D_20. The reaction is quenched and the proteins are digested to peptides for MS analysis. The data is then interpreted to compare the deuterium incorporation in the unbound and bound peptides resulting in the mapping of protein interaction site. Figure created in Biorender.

Methods to study HP-PPIs

The different strategies used to study HP-PPIs in this thesis are briefly explained below.

Surface adsorption-mass spectrometry (SA-MS)

SA-MS is a bacterial surface centred enrichment strategy. In SA-MS the intact live bacteria are incubated with biological samples (such as cell lysates, cell lines, plasma or saliva). During this incubation, the protein interactions established between the host proteins in the biological mixture and the bacterial surface proteins will be captured. The bacterial cells are then harvested followed by limited proteolysis using proteases like trypsin. The proteins arising from the proteolysis is

then analysed using MS to identify host proteins interacting with bacterial surface proteins. SA-MS has been used to study human plasma proteins interacting with different *S. pyogenes* serotypes[94,181].

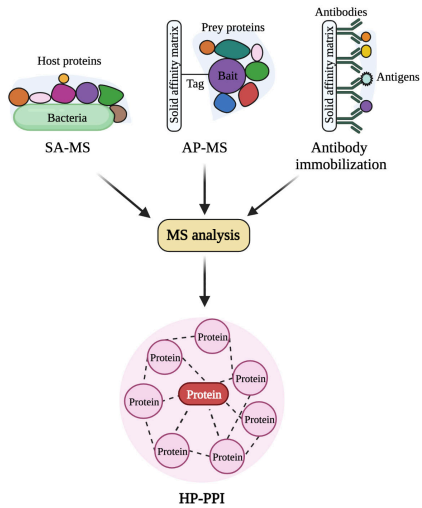


Figure 10: Different MS based strategies to study HP-PPI. SA-MS identifies the HP-PPIs formed on the bacterial surface, AP-MS determines the HP-PPIs formed around indvidual proteins and antibody immobilization is used to identify immune complexes formed between antigen and antibodies. Figure created in Biorender.

Affinity purification-mass spectrometry (AP-MS)

The underlying principle of AP-MS is the enrichment of proteins (prey) from a complex biological mixture (such as cell lysates, cell lines, plasma or saliva) using a bait protein coupled to a solid matrix via an affinity tag[177]. The recombinantly expressed affinity tagged bait protein and the biological sample is mixed to allow the prey proteins to interact and bind to the bait protein. The non-interacting unbound proteins are washed away and the resulting bait-prey complex is released from the solid matrix, enzymatically digested to peptides and analysed using MS[20,21,94]. AP-MS has been used to generate a quantitative interaction map between different M proteins and human plasma and saliva proteins[94,181].

Antibody immobilization

Different carrier material or solid matrix like sepharose, agarose, protein A or protein G columns are used to covalently or non-covalently immobilize antibodies[182]. Immobilization of antibodies using protein A or G, orients the

antibodies in a fashion such that their antigen binding sites are towards the bulk solution, allowing maximum interaction with antigens in native condition[182]. Individual antigens or biological samples such as plasma or saliva are then passed over the immobilized antibodies with a goal that the antibodies will form immune complexes with the antigens. Washing will remove the unbound proteins and the immune complex is then analysed by MS. Immobilization of IgG from IVIG has resulted in the identification of *S. pyogenes* surface antigens recognized by pooled human immunoglobulins[132].

Aim of the thesis

The human body is continuously challenged by different bacterial pathogens. The human immune system scouts and neutralize these pathogens, however, under certain conditions the bacteria is capable of adapting itself in the host leading to a successful infection. One main way it does so is by expressing a multitude of virulence factors which it uses to interact with multiple host proteins and as result a complex network of proteins are formed. *S. pyogenes* is one such bacterium that has been successful in infecting humans, accounting to its ability to form highly complex multimeric protein interaction networks. This thesis is focussed towards using different proteomics approaches to identify such *S. pyogenes* human protein networks and determine their role with an aim of advancing our understanding of the pathogenesis. Through, the research work presented in this thesis I try to answer three fundamental questions "*What, Why and How?*"

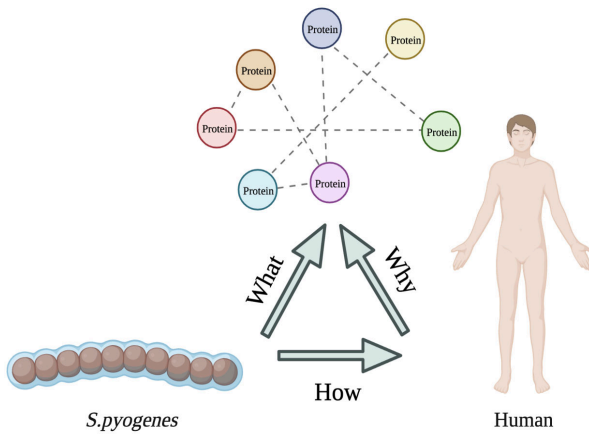


Figure 11: Schematics of the thesis aim. Three fundamental aspects of the *S. pyogenes* human protein interactions are studied, *What, Why and How*? Figure created in Biorender.

What proteins are part of this interactions?

A wide range of proteins are expressed by the bacteria, but the adaptive immune system evokes antibodies against only a limited set of proteins. **What** are the major *S. pyogenes* proteins that drive antibody response during and post-infection and **What** is their capacity to evoke the antibody response still remains partly unknown. Moreover, the virulence factors of the bacteria interact with various host proteins. **What** are the human proteins involved in the complex network of protein interactions formed around different *S. pyogenes* serotypes and its role in tissue tropism are yet to be explored. **Paper I** is focused towards identifying key *S. pyogenes* antigens and quantifying the immune response elicited by these antigens. **Paper II** is focused at mapping the human protein interactions on different *S. pyogenes* surface and around the key virulence factor M protein.

Why are such interactions important?

The next obvious question that arises is **Why** such protein interactions are important for *S. pyogenes*? The work explains the relevance of such interactions in the context of the adaptation of the bacteria in the host. **Paper II** reports the importance of protein interactions for the adaptation of *S. pyogenes* in different host microenvironment, whereas **Paper IV** demonstrates the role of M protein interaction with fibrinogen in immunomodulation.

How such protein networks are formed?

The final objective was to understand **How** *S. pyogenes* are capable of forming such multimeric protein complexes and its roles in escaping the host immune response. **Paper II** elucidates the nature of IgA and C4BP protein interactions with M protein and their role in immune evasion during a mucosal and systemic infection. **Paper III** reveals the molecular organization of IgG around M protein and its underlying mechanism in inhibition of phagocytosis.

Investigations and findings

In this section, a brief summary of the research carried out and the outcomes of the projects are presented. The background, aim, result and the conclusion of each of the papers are provided. For more details regarding each of the paper the reader can refer to the original form of papers included at the end of the thesis.

Paper I

Background

A large number of proteins are expressed by the bacteria, but only a few of them evoke an immune response in the host. The main aim of vaccine development is to identify key antigens that elicit a long-lasting antibody response, which can prevent further infection. However, what remains elusive for many bacterial pathogens including *S. pyogenes* is the complete spectrum of antigens that induce antibody response during and post-infection and whether the antigens evoke the same extent of antibody responses across different individuals. Therefore, quantifying the humoral immune response in humans could provide critical insights into the antibody repertoires and facilitate identification of correlates of protective immunity.

Aim

The specific aims of the paper were as follows:

- To develop a proteomics based immuno-profiling strategy that could quantitatively identify predominant bacterial antigens eliciting antibody response in the host and determine the sub-cellular localization of such antigens.
- Use the strategy to dissect humoral immune response during natural *S. pyogenes* exposure and under sever *S. pyogenes* infections.

Result

In this study, we developed a fully automated quantitative MS based method called "systems antigen-omics" to profile bacterial antigens. The method exploits the affinity of antigen-specific antibodies circulating in the plasma or serum to enrich antigens from a pool of fractionated bacterial proteins. We used this method to define the *S. pyogenes* antigen-ome. To determine the localization of the antigens an in-depth fractionation of the bacterium was performed. A total of 233 *S. pyogenes* proteins were identified from a pool of secreted, cell wall and surface fractions. To map antigens eliciting antibody response during natural exposure, IgG's were

enriched from a cohort of healthy donors, which were then used to affinity purify antigens from a pool of these proteins. The individuals had circulating antibodies against only a small fraction *i.e.* 17% of the total *S. pyogenes* surface attached proteins. These included some of the well characterized *S. pyogenes* antigens like M1, SLO and C5AP. The majority of the antigens were identified in the secreted fractions. Moreover, the response to these antigens were highly personalized. We then sought to understand the antibody response during severe *S. pyogenes* infections and we analyzed acute and convalescent sera samples from four *S. pyogenes* infected patients. The number of antigens eliciting antibody response during infections did not change to a great extent; however, the composition of the antigen repertoire varied a lot compared to a natural condition, revealing that mounting an antibody response against antigens is not always linked with protection. SPEB, one of the major *S. pyogenes* virulence factor was only identified in patients and was absent in healthy donors. The response was highly patient centric. Additionally, not much difference in the immune response was observed between the acute and convalescent phase.

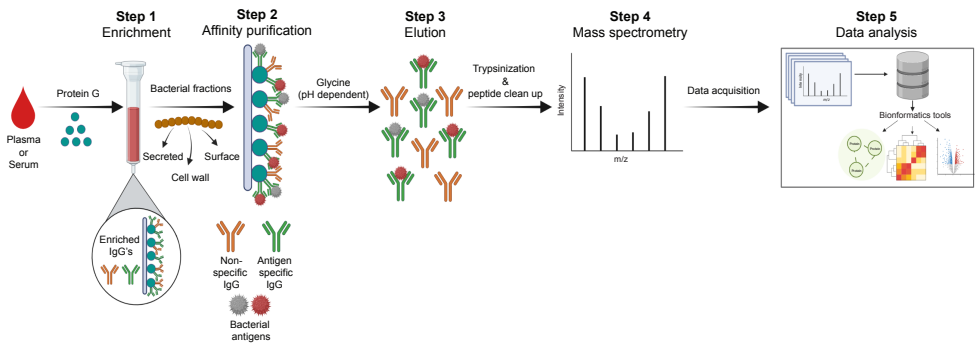


Figure 12: Schematics of the systems antigen-omics approach. Step 1- Enrichment, IgG's are enriched from plasma or sera using protein G column. Step 2- Affinity purification, IgG's are used to affinity purify antigens from a pool of bacterial fractions, antigen specific IgG's shall bind to antigens while the non-specific IgG's wont bind any antigens. Step 3- Elution, antigens and antibodies are eluted. Step 4- Mass spectrometry, antigens and antibodies are trypsinised into peptides followed by peptide clean up and then the peptides are measured using mass spectrometer. Step 5- Data analysis, the proteomics data are analysed resulting in the identifiaction of the bacterial antigens.

Conclusion

The systems antigen-omics approach can quantitatively profile antigens evoking antibody response in a species independent fashion and also map the sub-cellular

localization of the antigens. Moreover, the MS intensities of antigens measured using this method can be used to estimate titers enabling the comparative analysis of antigenic response across different host. The compressed panel of 50 *S. pyogenes* antigens generated using this approach can facilitate the selection of potential *S. pyogenes* vaccine candidates. In summary, the systems antigen-omics approach has the potential of identifying antigens for any pathogen, thereby accelerating the development of vaccine or other therapeutic interventions.

Paper II

Background

S. pyogenes is known to express a wide range of virulence factors which it uses to circumvent the human immune response. Among these virulence factors M protein is abundantly present on the surface of the bacteria and is one of the prominent targets of the antibodies from the host as identified in **Paper I**. Based on the domain arrangement of M proteins they are classified into A-C type, D type and E type. The combination of the domains of the M protein allows it to engage with numerous human proteins simultaneously. The protein interactions formed around A-C type is well described; however, the HP-PPIs formed at the surface of the different E type *S. pyogenes* serotypes and the extent they are mediated by the M protein still remain largely unexplored.

Aim

The aims of this paper were to:

- Quantify the different human plasma proteins interacting with different serotypes of *S. pyogenes*.
- Identify the extent of interactions mediated by M protein.
- Characterize HP-PPI in different host microenvironments.

Result

A combination of quantitative and structural mass spectrometry techniques was used to determine the structure and composition of the protein interaction network formed between human plasma or saliva proteins and the surface of *S. pyogenes* serotypes and the M protein. To determine the different human plasma proteins interacting with *S. pyogenes* surface, SA-MS was performed on six different *S. pyogenes* serotypes *i.e.* A-C type (M1, M3, M5,) and E type (M28, M49, M89). DIA-MS analysis of the protein interactions revealed a total of 92 surface bound human plasma proteins. A marked difference was observed in the HP-PPI around the A-C type and E type, particularly the binding of fibrinogen was specific to A-C type and C4BP binding was specific to E type. To understand the role of the M

proteins in dictating the formation of such HP-PPIs on the *S. pyogenes* surface, AP-MS was performed. AP-MS analysis on the recombinantly expressed M proteins corresponding to the same bacterial serotype identified a list of 32 human plasma proteins. Patterns similar to SA-MS were observed, such that the A-C type M proteins bound fibrinogen while the E type M proteins bound C4BP. One interesting observation was significantly high-level enrichment of IgA and C4BP by M28. As IgA is a predominant antibody in the mucosal environment and C4BP is present mostly in the plasma, we quantified the protein interactions of M28 in a mixture of saliva and plasma. The results showed that M28 binds sIgA in saliva and IgA and C4BP in plasma. To understand the concomitant binding of IgA and C4BP on M28 we used two orthogonal structural mass spectrometry techniques. Both TX-MS and HDX-MS identified a single C4BP binding site and two distinct IgA binding sites on the M28 protein. Based on the findings we propose three models of IgA binding to M28 in the mucosal niche: i) one dimeric IgA binding, ii) two dimeric IgA binding and iii) one monomeric and one dimeric IgA binding. However, in the plasma upsurge of C4BP drives the binding of M28 to C4BP and IgA.

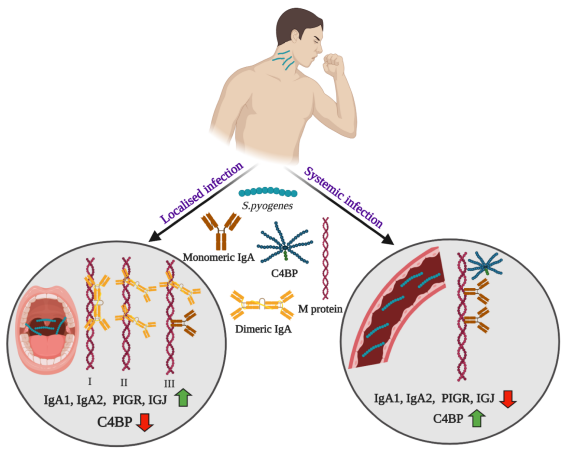


Figure 13: Overview of immune evasion strategies adopted by M28 in different host microenvironments. Reproduced from *Chowdhury et.al., (2021)- "Streptococcus pyogenes Forms Serotype-and Local Environment-Dependent Interspecies Protein Complexes", mSystems, vol 6, issue 5*. During localised infection M28 binds either i) one dimeric IgA, ii) two dimeric IgA or iii) one monomeric and one dimeric IgA. During systemic infection M28 binds IgA and C4BP.

Conclusion

The combination of MS-based methods demonstrated that there is substantial difference in the composition of the HP-PPIs formed between A-C type and E type *S. pyogenes* and that these HPP-PPIs were serotype specific. Furthermore, we confirm that the sequence variability and the domain arrangement of the M proteins play an important role in shaping the serotype-specific HP-PPIs. We also report that M28 either binds sIgA or monomeric IgA and C4BP depending on whether *S. pyogenes* cause localized or systemic infections, thereby proposing possible immune evasion strategies in different host-microenvironments.

Paper III

Background

Over 30 human plasma proteins interacting with the M protein have been identified in **Paper II**. Amongst these human proteins, IgG is one such protein that interacts with all the M proteins. The M proteins are known to interact with both Fab or Fc domain of the IgG and the Fab or Fc mediated IgG interaction is governed by the IgG concentration in the host microenvironment. Even with our knowledge of the role of M protein-IgG binding the molecular details of this interactions remains unclear.

Aim

The aim of this work was:

- To characterize the structure and dynamics of the IgG interaction with M proteins.

Result

The M1 serotype is a prevalent *S. pyogenes* serotype, therefore we used integrative structural biology approaches to determine the interaction of the M1 protein with human IgG's. To mimic a scenario of invasive infection, the intact live bacteria was incubated in pooled human plasma, thus allowing the binding of both Fab and Fc domains of IgG to the M1 protein. The native interactions formed were captured by chemical cross-linking. The majority of the identified cross-links supported Fc mediated binding of IgG's to M1. TX-MS was then used to elucidate the binding interface. Two regions on the M1 protein were found to interact with IgG *i.e.* the A domain and a region between S & C-1 domains. Most of the IgG-Fc interaction was with the region between S and C-1 domain of M1 protein and the A domain had Fab and Fc interactions. To evaluate the affinity and stability of the Fc-binding interfaces, we performed surface plasmon resonance (SPR) and molecular dynamics (MD) stimulations. SPR confirmed the presence of two IgG-Fc binding sites on M1 protein. MD stimulations revealed strong interaction between the A domain of M1

and IgG-Fc compared to the interaction between S and C-1 region of M1 and IgG-Fc.

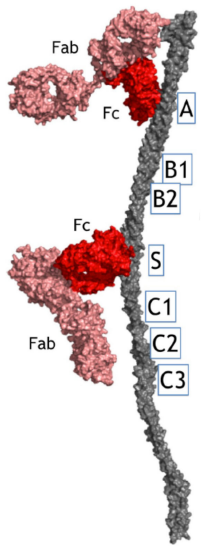


Figure 14: M1 interactions with IgG. Reproduced from *Khakzad et.al., (2021)*- "*Structural determination of Streptococcus pyogenes M1 protein interactions with human immunoglobulin G using integrative structural biology", PLOS Computational Biology, vol 17, issue 1*. The model for M1 interaction with IgG. The different domains of M protein are shown where A domain and the region between S-C1 domain mainly interacts with IgG. M1 is represented in grey colour and IgG in red.

Conclusion

The molecular characterization of the M1-IgG interactions revealed that the interaction of IgG via the Fc-domains with M protein would mask the recognition site of FcγR receptors, hence protecting *S. pyogenes* from phagocytic killing. The M1 binding site on IgG was found to be similar with the streptococcal protein G interaction site on IgG[185,186]. Furthermore, this study identified important M1 peptides that can be further investigated as potential vaccine candidates.

Paper IV

Background

Fibrinogen is a predominant blood protein that reacts to an infection by activating platelets. Platelets are known to recognize and respond to pathogenic bacteria. *S. pyogenes* binding to fibrinogen has been reported to be important for escaping fibrinogen-host antimicrobial functions and facilitate invasion in host. However, not all *S. pyogenes* strains can bind fibrinogen. The highly selective binding pattern of fibrinogen is to a great extent mediated by the different A-C and E type M proteins, as reported in **Paper II**. The role of M1 protein in activation of platelets have been demonstrated to be fibrinogen and anti-M1 specific IgG dependent engaging both Fc and fibrinogen receptors on platelets. However, the roles of the other A-C and E type M proteins are yet to be investigated.

Aim

The aim of this study was:

- To assess the importance and function of fibrinogen acquisition for platelets mediated by M protein from different invasive serotypes of *S. pyogenes*.

Result

The platelet dependent proinflammatory effects of A-C (M1, M3, M5) and E (M28, M49, M89) type M proteins was studied using a combination of multiparameter flow cytometry, enzyme linked immunosorbent assay (ELISA), aggregometry and quantitative mass spectrometry. As expected only M1, M3 and M5 protein bound to fibrinogen in plasma and mediated fibrinogen-IgG dependent platelet activation. However, M28, M49 and M89 did not interact with fibrinogen and therefore failed to activate platelets. M1, M3 and M5 were capable of platelet granule release and platelet aggregation, but not M28, M49 and M89. Furthermore, M1, M3 and M5 could initiate and activate platelet-neutrophil and platelet-monocyte complex, but M28, M49 and M89 failed to mediate initiation and activation of neutrophil and monocytes complex.

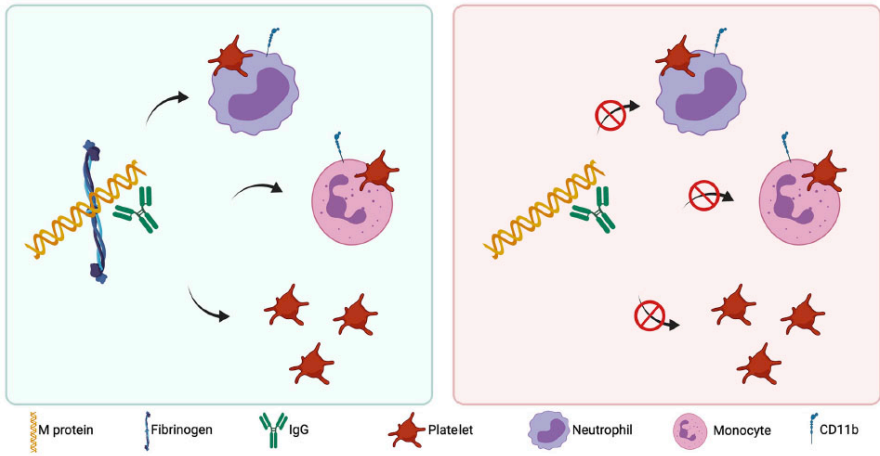


Figure 15: Fibrinogen binding to M protein and its immunomodulatory role. Reproduced from *Palm et.al., (2022)- "Distinct Serotypes of Streptococcal M proteins Mediate Fibrinogen-Dependent Platelet Activation and Proinflammatory Effects", Infection and Immunity, vol 90, issue 2*. Only certain M proteins (M1, M3 and M5) bind fibriongen and IgG-Fabs's mediating activation of platelets and formation and activation of complex with neutrophils and monocytes (Green box). M28, M49 and M89 donot bind fibriongen and therefore fail to activate platelet and its downstream effects (Red box).

Conclusion

The study confirms that M proteins released from only certain *S. pyogenes* serotypes could bind fibrinogen and Fab-IgG thereby causing rapid activation of platelets. The findings show that platelets activated by bacterial proteins release immunomodulatory proteins and engage with neutrophils and monocytes thereby forming platelet-neutrophil and platelet-monocyte complex. We therefore report novel aspects of the immunomodulatory role of fibrinogen acquisition and platelet activation during *S. pyogenes* infections

Discussion and conclusion

Protein-protein interactions are the primary functional process that drives all biological systems. Alterations in the specificity and affinity of these interactions can be the cause of cellular malfunctions and diseases[187]. The common theme of my research is to understand protein-protein interactions in the case of an infection. During an infection, there is a complex interplay between host and pathogenic proteins resulting in the formation of HP-PPIs. The goal of this work was to use a combination of quantitative and structural proteomics to understand the complex networks formed between human and *S. pyogenes* proteins. To adapt itself best in the host settings, *S. pyogenes* forms multimeric protein networks with numerous human proteins. To understand such a wide network of interactions three fundamental aspects of these protein networks were addressed "*What, Why and How*"

During an infection, the humoral immune response is initiated, which is driven by the generation of antibodies against a set of pathogen specific proteins. Even with our better understanding of the adaptive immune response and contribution of functional antibodies, there is a big gap in our understanding of adaptive immune response elicited against the bacterial antigens. One of the major reason for the lack of knowledge is the unavailability of a strategy to systematically profile antigens based on host immune response. This problem has been the foundation of the research question of **Paper I**. To study the link between bacterial antigens and the host antibodies we developed a proteomics workflow, called "systems antigen-omics". The developed systems antigen-omics approach is not only capable of profiling multiple surface attached bacterial antigens but it also offers other added capabilities. Quantifying antibody responses has always been very challenging because of multiple reasons, mainly as antibodies are induced by native proteins and therefore determining antibody specificity on a three-dimensional structure is difficult. Additionally, the affinity issues of antibodies further complicate antibody measurements[188]. This strategy profiles antibody response based on the ability of different antigens to evoke an adaptive response, therefore it opens up an opportunity to quantitatively characterize host antibody responses in an unbiased

fashion. Moreover, as the method reports the antigen intensity measured by MS, therefore systems antigen-omics can be used as an indirect measure of antibody titers. Reverse vaccinology approaches involve the prediction of possible vaccine targets and their localization by large-scale genomic profiling, bioinformatics and in-silico prediction tools followed by *in vivo* immuno assays and *in vitro* animal studies. But, these proteins are not always immunogenic in humans and its ability to induce an antibody response cannot be predicted by screening in animal models. Systems antigen-omics relies on the circulating antibodies in a host to enrich antigens from a pool of secreted, cell wall and surface fractions. This gives not only an idea about the immunogenic antigens in the host but also the sub-cellular localization of the predominant antigens. Moreover, the method has been automated allowing the screening of multiple samples to compare host specific antigenic response. Lastly, this method could be used to profile antigen-ome's in a species independent manner. One application of this method across different species, could be to profile antibodies against host gut microbiota. The gut microbiota plays an important role in maintaining the gut homeostasis and protecting against enteric pathogens. Some symbiotic gut bacteria are capable of entering the systemic circulation and inducing an IgG response, which in turn protects the host from systemic infections by Gram-negative pathogens[189]. Once such microbiome specific proteins capable of triggering an adaptive immune response, are identified they could then be used to boost immunity to combat systemic infections. The method is also being expanded to identify epitopes and quantify antibody binding on the epitopes.

S. pyogenes has been infecting humans in different forms for a very long time but the humoral response of humans to different *S. pyogenes* antigens still remains poorly characterized. Thus, to understand the immune response of different individuals to *S. pyogenes* surface proteins during natural exposure and in infected condition the streptococcal antigen-ome was profiled using the systems antigen-omics approach. An interesting observation of **Paper I** was the host specific response. It was evident that the level of antibodies against an antigen varied to great extent between individuals. Host factors like antigen quantity, B cell precursor frequency, B cell receptor avidity, route of immunization and genetic background play an important role in dictating the generation of antibodies against antigens[188]. Each of these individuals were infected via different routes and the genetic makeup of these individuals were also different, which could well explain the differential level of humoral response. Additionally, the level of antigens expressed are different in different time points of infection[190], which could also account for different

antibody responses. Even though very high levels of antibodies were generated against these antigens, the antibodies were not protective. Antigens are known to possess immuno-dominant and sub-dominant regions. A region in the antigen that elicits the maximum amount of antibody response is said to be immuno-dominant while sub-dominant regions elicit less antibody response[190]. Antibody response against an immuno-dominant region does not always confer protection and weak antibody responses against sub-dominant regions could be possible immune evasion strategies adopted by the pathogens[191-193]. It could thus be that the immuno-dominant regions of these antigens were evoking high antibody titers, therefore even with a high level of circulating antibodies protection against the pathogen was not conferred. Therefore, along with quantifying the antibody response it is equally important to identify immuno-dominant and sub-dominant target sites of the antibody on the antigens. Techniques like TX-MS[180], protein painting[194], epitope excision and epitope extraction[182,195,196] could be used to map the antigen-antibody interface, and HDX-MS could be the used to report the antibody abundance on these sites[183]. Currently, I am working on implementing epitope excision and extraction along with HDX-MS techniques in the systems antigen-omics to identify epitopes for all the *S. pyogenes* antigens and report the antibody occupancy on those epitopes.

Even with the ongoing research for nearly over a century, the development of a safe and effective *S. pyogenes* vaccine still awaits[133]. Amongst the many challenges encountered, one of the major challenges that hinders this development is the lack of knowledge about the correct protein target for a vaccine formulation. For a pathogen like *S. pyogenes* that has evolved to express several virulence factors to overcome the host immune response, a vaccine with one antigen might not be a solution. A vaccine against *S. pyogenes* should rather have a combination of different antigens. Therefore, the compressed panel of 50 *S. pyogenes* antigens generated in **Paper I** could be used to select potential vaccine candidates and test them in *in vivo* and *in vitro* pre-clinical studies. However, this study has also opened up new possibilities of profiling antigens from other *S. pyogenes* infected patient cohorts with different clinical manifestations to compare and expand the *S. pyogenes* antigen-ome. Comparison of antigens eliciting antibody responses in the individuals during different infections could help us determine certain infection specific antigens, which could then be used to design diagnostic kits to identify various *S. pyogenes* infections.

Despite a strong adaptive response against various *S. pyogenes* antigens, the bacterium is capable of causing both localized and systemic infections. It has been observed that certain serotypes are the cause of infections at specific sites revealing

a relationship between *S. pyogenes* serotypes and tissue tropism. The aim of **Paper II** was to identify the protein interactions formed around the different *S. pyogenes* serotypes and to understand whether these protein interactions played an important role in facilitating the bacteria to cause infections at particular sites. We observed that the different serotypes bind specifically to distinct sets of human proteins. As the variability in M protein is the major reason for the rise of different *S. pyogenes* serotypes; hence we investigated the contribution of M proteins in the formation of this serotype specific protein networks in **Paper II**. M proteins were responsible for the serotype-specific protein networks. Based on the mutual exclusive binding pattern of M proteins, it was evident that once an interaction with a human protein is formed, interaction with other proteins are not readily formed. It appears that *S. pyogenes* has evolved to express M proteins with sequence variation and different domain arrangement to form protein interactions with human proteins that can maximize their chance of survival in a particular host microenvironment. One prominent example is the selective binding of IgA and C4BP to M28 protein, an E type M protein. Serotypes expressing E type M protein are generalist strains, capable of causing both localized and systemic infections. *S. pyogenes* bacteria are well known to localize the mucosal surface, and IgA is the predominant antibody in the mucosal environment. The binding of sIgA and IgA to M28 in the mucosal environment could therefore be an immune evasion strategy to overcome the first line of immune defence in the mucosal environment and could also be an adhesion mechanism for the bacteria to adhere to the mucosal surface. Persistent infections of the mucosal membrane by M28 can cause vascular leakage providing access of the bacterium to the systemic circulation, as M28 has been reported to be the leading cause of puerperal sepsis[197,198]. Sepsis however, is a rare condition compared to uncomplicated local conditions, so the evolution of the bacteria to adapt itself in the host takes place in the local host microenvironment rather than the blood[100]. During a mucosal infection, there might be local damage of the mucosal membrane, leading to the leakage of plasma. The bacteria then encounter a drastic change in the protein composition with a certain upsurge of plasma proteins like C4BP. This could then drive the binding of C4BP and IgA to M28, thus preventing the activation of the complement system and clearance by phagocytosis. Characterization of the HP-PPIs for *S. pyogenes* reveals the ability of the pathogen to form specific protein interaction networks with human proteins that facilitates their survival and dissemination in different ecological niches in the host. HP-PPIs is a critical element for all infections, therefore analysis of HP-PPIs for other infections could be detrimental in understanding the host rewiring mechanism adopted by other

pathogens. In this study, a combination of different quantitative and structural proteomics techniques was used to decipher the protein interactions, which could prove to be a valuable tool set for studying HP-PPIs in other infections.

As discussed earlier, antigens have immuno-dominant and sub-dominant sites and it is of utmost importance to map the interaction site of antibodies on the antigen. We then investigated the antigen-antibody interaction in **Paper III**. M1 protein was selected for mapping the IgG-binding site as almost all individuals studied in **Paper I** had circulating IgG's against the M1 protein. TX-MS analysis identified two IgG binding interfaces and MD simulations elucidated the molecular details of the interactions. The M1 protein bound all IgG subclasses at a specific region on the C_H3 domain. This site on the IgG is also the recognition site of FcγR receptors expressed by professional phagocytes. The interaction of M1 with IgG at this site will prevent the bacteria from phagocytic killing, thus confirming the immune evasion mechanism. Repeat regions are proposed to be immuno-dominant regions of the antigen[190], and the IgG's bind to the A and C regions on the M1 protein which are actually repeat regions. Therefore, it could be concluded that antibodies targeting immuno-dominant sites are not always capable of neutralizing the pathogen, which in turn could be an evolutionary mechanism of the pathogen to express such repeat regions to bind host antibodies to trick and circumvent the host immune system for its own benefit. The C_H3 domain on IgG interacting with M1 has also been reported to bind protein G of group C and group G streptococci[185,186]. Thus, it is evident that using immuno-dominant sites to evade immune surveillance is a common phenomenon across other pathogens. TX-MS analysis identified peptide sequences that could capture IgGs via their Fc domains. Blocking these peptides would prevent binding of IgG-Fc thereby facilitating bacterial clearance by the human immune system. These peptides could further be interrogated for their therapeutic capabilities. This study outlines the importance of studying the structural and molecular aspects of an antigen-antibody interactions. Structural and molecular characterization of antigen-antibody complexes for any pathogen to great details have the capability of describing immune evasion mechanisms and identify pathogen-derived peptides that could be potential novel targets for therapeutic strategies to combat infections.

In **Paper II** we report fibrinogen interaction with *S. pyogenes* in a M specific manner, and we thus aimed to understand the role of fibrinogen binding to M protein in **Paper IV**. Fibrinogen binding coupled with Fab-IgG interaction with M1, M3 and M5 mediated rapid platelet activation and complex formation with neutrophils and monocytes while M28, M49 and M89 did not interact with fibrinogen and thus

failed to activate platelets and complex formation. All these *S. pyogenes* strains are responsible for invasive infections, it could thus be inferred that binding of fibrinogen to M protein is not a pre-requisite for invasive infections. During an infection, the blood protein fibrinogen participates in inflammation and tissue repair. At the time of a vascular injury platelets migrate to the site of injury and prevent blood loss by binding to fibrinogen and initiating blood coagulation. Platelets are also involved in immune response [199]. *S. pyogenes* bacteria are known to cause localised infections on the skin and damaged skin can act as a route for the *S. pyogenes* entry into the systemic circulation. As platelets are present at the site of damage therefore, the bacteria have to overcome the immune action of platelets. One way the bacteria does so, is by binding to fibrinogen. In this study, we provide a potential immune manipulation strategy of *S. pyogenes*. We propose that the M proteins are released from the bacterial surface and the interaction with fibrinogen takes place at a distance. As the platelets are activated at a distance, therefore the bacteria could avoid direct contact with antimicrobial substances released by activated platelets or engulfment in platelet aggregates. Thrombocytopenia *i.e.* low platelet count has been reported during infections[200]; it could well be that during an invasive infection *S. pyogenes* depletes platelets by activation and aggregation to facilitate its better survival and dissemination in the host. M proteins not binding to fibrinogen were found to bind vitamin K-dependent protein S (PROS) in **Paper II**. PROS is a plasma protein with anticoagulant activity and it exists both in free form and in complex with C4BP[201-203]. During an invasive infection, the binding of PROS to the M proteins could prevent clot formation and entrapment of the non-fibrinogen binding *S. pyogenes* serotypes in platelet aggregates. Therefore, it seems that various *S. pyogenes* serotypes have developed different mechanisms to overcome the platelet response. This study demonstrates novel serotype-specific consequences of fibrinogen binding and platelet activation during *S. pyogenes* infections.

In summary, the combined work of this thesis reports the diversity of HP-PPIs and the importance of elucidating HP-PPIs to understand the host response to an infection and the pathogens strategies for adaptation in the host. Furthermore, this thesis outlines different proteomics resources which could be integrated to study severe infectious diseases in the future.

Acknowledgements

A lot of people have helped and contributed in this journey of mine either in the form of scientific discussions, administrative help, moral support or just by being by my side. Thank you all!

I would first like to thank my supervisor **Johan Malmström.** Thank you for providing me the opportunity of doing research in your group. I am really grateful for the freedom you gave me during these years to pursue my research ideas. I still remember our first conversation where I said I do not have any experience in proteomics and your answer *"do not worry I am not looking for someone who is an expert in proteomics rather I am looking for someone who has the urge to learn"*. Really, your endless support and guidance throughout these years have not only made me learn a lot about science but has also helped me widen my imagination. I utterly enjoyed the random conversations we had when I knocked at your office door and said, "*Hi, do you have a moment…*". Thank you for inspiring me and making me confident!

To my co-supervisor **Lotta Happonen**. Thank you for everything since day one, from receiving me at the Copenhagen airport to driving me to Lund, helping me get settled. You are the one who introduced me to all the basic laboratory techniques, taught me to prepare mass spectrometer samples, run them on the mass spectrometer and analyze the data, Thank you! I was really lucky to share the office with you in the initial years of my PhD as it was very easy for me to bother you with any sort of questions and you were always there to answer it. I also enjoyed the way we tried to work with our Jupyter notebook scripts.

Anita, for helping in the wide range of problems. Your enthusiasm and organizational skills is highly commendable.

Tirthankar, thank you for sharing the advertisement and also for your support during the initial days regarding my move and stay at Lund. I also enjoyed the random scientific discussions and gym sessions we had.

Tommaso, you make the work environment so lively. Enjoy listening to your jokes across the lab bench while experimenting and of course thank you for the awesome pizzas and tiramisu.

Anahita, I was glad that you were back in the lab and I finally had a co-PhD student. I still remember our first adventure to the autoclave. It is fun working with you and having discussions about food and restaurants and remember the importance of happy tummy!

Alejandro, for all the scientific discussions we have had. I really appreciate how you share interesting research papers with me and you are always ready to talk about science. It was nice collaborating and working with you.

Lars, for all the scientific conversations and your valuable inputs on different bioinformatics ways and tools to analyze data.

Hamed, for a fruitful collaboration and helping out with protein modeling. Even though we haven't meet in person but I enjoyed talking to you on teams.

CK, thank you for all your support and help in the lab and regarding the mass spectrometer.

Carlos, thank you for teaching me spanish and also to quickly identify the number of total proteins in a fasta file.

Aaron, it was great to have you as an office mate and discuss bioinformatics related issues with you.

Di, for being a nice co-PhD friend and for all your help in cleaning the bacteria lab.

Elisabeth, was fun working with you and discussing different ways of analyzing the data.

The IMP group, to all the past and present members for providing a very friendly and supportive work environment in the lab.

Simon, for performing all the HDX-MS experiments and analyzing the data and allowing me to use the pH meter.

Thanks to all the people in **BioMS** and **D13** for such a wonderful work environment and for all the enjoyable talks during lunch.

To all my collaborators **Rolf, Gizem, Oonagh, Frida, Magnus, Anna, Pontus** and **Therese** for sharing precious samples, great scientific inputs, technical assistance and guidance. I really enjoyed working with you all.

To all the people in B14 you were always so welcoming and ready to help me whenever needed.

To my EU PhD program **ViBrANT** supervisors who have guided me in my PhD and have also hosted me in their laboratory. It was a wonderful experience to work with such a diverse group of researchers and PhD students and I learned a lot from the different courses we had. Cheers to all my **ViBrANT PhD friends** for all the fun and nice memories.

All my **teachers and professors** for helping me develop an interest towards research.

To my **Ma and Baba**, please know that what I am today is only because of you both. You made me dream and also taught me to chase my dream. Both of you have taught me the real meaning of life and made me believe in myself. I am really blessed to have such supportive parents who have always instilled faith in me and said that "*I can do*" even in the most difficult times of mine. I am really grateful for such an upbringing and for your unconditional love, care and affection!

To my wife **Chandreyee**, you made this journey much more enjoyable. Believe me you transformed the dullest moment to a bright one and your words "*Don't worry it will be fine soon*" were real energy boasters. For all the fun and adventure, and for always being patient in listening to my research. Kudos to you for understanding me during the final leg of my PhD and being awake with me late in the night which obviously gave you some extra "*me time*".

To my **family back in India**, thank you for being there by my side both physically and emotionally.

To my **friends outside the lab in Sweden**, thank you all for making me feel home away from home and making the life outside research much more joyful.

References

1. Jenner E. An inquiry into the causes and effects of the variolae vaccine. 1798.
2. Koch R. The etiology of tuberculosis (Koch's postulates). The germ theory of disease. 1884:116-118.
3. Neville BA, Forster SC, Lawley TD. Commensal Koch's postulates: establishing causation in human microbiota research. Curr Opin Microbiol. 2018;42:47-52.
4. Molecular Biology of The Cell. 4th ed: Garland Science.
5. Bird BH, Mazet JAK. Detection of Emerging Zoonotic Pathogens: An Integrated One Health Approach. Annu Rev Anim Biosci. 2018;6:121-139.
6. Casadevall A, Pirofski LA. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. Infect Immun. 1999;67(8):3703-3713.
7. Shah PS, Wojcechowskyj JA, Eckhardt M, Krogan NJ. Comparative mapping of host-pathogen protein-protein interactions. Curr Opin Microbiol. 2015;27:62-68.
8. Asensio NC GE, Groot NSD , Burgas MT. Centrality in the host-pathogen interactome is associated with pathogen fitness during infection. Nature Communications. 2017(8):14092.
9. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature. 1999;402(6761 Suppl):C47-52.
10. Nicod C, Banaei-Esfahani A, Collins BC. Elucidation of host-pathogen protein-protein interactions to uncover mechanisms of host cell rewiring. Curr Opin Microbiol. 2017;39:7-15.
11. Yang H, Ke Y, Wang J, et al. Insight into bacterial virulence mechanisms against host immune response via the Yersinia pestis-human protein-protein interaction network. Infect Immun. 2011;79(11):4413-4424.
12. Dyer MD, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog. 2008;4(2):e32.
13. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001;411(6833):41-42.
14. He X, Zhang J. Why do hubs tend to be essential in protein networks? PLoS Genet. 2006;2(6):e88.

15. Durmus Tekir S, Cakir T, Ulgen KO. Infection Strategies of Bacterial and Viral Pathogens through Pathogen-Human Protein-Protein Interactions. Front Microbiol. 2012;3:46.
16. Acharya D, Dutta TK. Elucidating the network features and evolutionary attributes of intra- and interspecific protein-protein interactions between human and pathogenic bacteria. Sci Rep. 2021;11(1):190.
17. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nat Rev Genet. 2015;16(4):224-236.
18. Forst CV. Host-pathogen systems biology. Drug Discov Today. 2006;11(5-6):220-227.
19. Bhavsar AP, Guttman JA, Finlay BB. Manipulation of host-cell pathways by bacterial pathogens. Nature. 2007;449(7164):827-834.
20. Penn BH, Netter Z, Johnson JR, et al. An Mtb-Human Protein-Protein Interaction Map Identifies a Switch between Host Antiviral and Antibacterial Responses. Mol Cell. 2018;71(4):637-648 e635.
21. Mirrashidi KM, Elwell CA, Verschueren E, et al. Global Mapping of the Inc-Human Interactome Reveals that Retromer Restricts Chlamydia Infection. Cell Host Microbe. 2015;18(1):109-121.
22. Chen YF, Xia Y. Structural Profiling of Bacterial Effectors Reveals Enrichment of Host-Interacting Domains and Motifs. Front Mol Biosci. 2021;8:626600.
23. Schweppe DK, Harding C, Chavez JD, et al. Host-Microbe Protein Interactions during Bacterial Infection. Chem Biol. 2015;22(11):1521-1530.
24. Janeways Immunobiology. 9th ed: Garland Science.
25. Nguyen AV, Soulika AM. The Dynamics of the Skin's Immune System. Int J Mol Sci. 2019;20(8).
26. Toke O. Antimicrobial peptides: new candidates in the fight against bacterial infections. Biopolymers. 2005;80(6):717-735.
27. Brass LF, Diamond SL, Stalker TJ. Platelets and hemostasis: a new perspective on an old subject. Blood Adv. 2016;1(1):5-9.
28. Morrell CN, Aggrey AA, Chapman LM, Modjeski KL. Emerging roles for platelets as immune and inflammatory cells. Blood. 2014;123(18):2759-2767.
29. von Hundelshausen P, Weber C. Platelets as immune cells: bridging inflammation and cardiovascular disease. Circ Res. 2007;100(1):27-40.
30. Maouia A, Rebetz J, Kapur R, Semple JW. The Immune Nature of Platelets Revisited. Transfus Med Rev. 2020;34(4):209-220.
31. Speth C, Rambach G, Wurzner R, et al. Complement and platelets: Mutual interference in the immune network. Mol Immunol. 2015;67(1):108-118.

32. Dunkelberger JR, Song WC. Complement and its role in innate and adaptive immune responses. Cell Res. 2010;20(1):34-50.
33. Sarma JV, Ward PA. The complement system. Cell Tissue Res. 2011;343(1):227-235.
34. Beutler B. Innate immunity: an overview. Mol Immunol. 2004;40(12):845-859.
35. Medzhitov R, Janeway C, Jr. Innate immunity. N Engl J Med. 2000;343(5):338-344.
36. Mahanty S, Prigent A, Garraud O. Immunogenicity of infectious pathogens and vaccine antigens. BMC Immunol. 2015;16:31.
37. Bettelli E, Korn T, Oukka M, Kuchroo VK. Induction and effector functions of T(H)17 cells. Nature. 2008;453(7198):1051-1057.
38. Cresswell P, Ackerman AL, Giodini A, Peaper DR, Wearsch PA. Mechanisms of MHC class I-restricted antigen processing and cross-presentation. Immunol Rev. 2005;207:145-157.
39. Bacterial Pathogenesis A Molecular Approach. 2nd ed: ASM Press.
40. Woof JM, Russell MW. Structure and function relationships in IgA. Mucosal Immunol. 2011;4(6):590-597.
41. Wines BD, Sardjono CT, Trist HH, Lay CS, Hogarth PM. The interaction of Fc alpha RI with IgA and its implications for ligand binding by immunoreceptors of the leukocyte receptor cluster. J Immunol. 2001;166(3):1781-1789.
42. Leong KW, Ding JL. The unexplored roles of human serum IgA. DNA Cell Biol. 2014;33(12):823-829.
43. Delacroix DL, Dive C, Rambaud JC, Vaerman JP. IgA subclasses in various secretions and in serum. Immunology. 1982;47(2):383-385.
44. Monto AS, Malosh RE, Petrie JG, Martin ET. The Doctrine of Original Antigenic Sin: Separating Good From Evil. J Infect Dis. 2017;215(12):1782-1788.
45. Tsai CM, Caldera JR, Hajam IA, et al. Non-protective immune imprint underlies failure of Staphylococcus aureus IsdB vaccine. Cell Host Microbe. 2022;30(8):1163-1172 e1166.
46. Mishra RP, Oviedo-Orta E, Prachi P, Rappuoli R, Bagnoli F. Vaccines and antibiotic resistance. Curr Opin Microbiol. 2012;15(5):596-602.
47. van Belkum A, Almeida C, Bardiaux B, et al. Host-Pathogen Adhesion as the Basis of Innovative Diagnostics for Emerging Pathogens. Diagnostics (Basel). 2021;11(7).
48. Rappuoli R, Mandl CW, Black S, De Gregorio E. Vaccines for the twenty-first century society. Nat Rev Immunol. 2011;11(12):865-872.
49. Rodrigues CMC, Pinto MV, Sadarangani M, Plotkin SA. Whither vaccines? J Infect. 2017;74 Suppl 1:S2-S9.

50. Rappuoli R, Santoni A, Mantovani A. Vaccines: An achievement of civilization, a human right, our health insurance for the future. J Exp Med. 2019;216(1):7-9.
51. Rappuoli R. Bridging the knowledge gaps in vaccine design. Nat Biotechnol. 2007;25(12):1361-1366.
52. Plotkin SA, Plotkin SL. The development of vaccines: how the past led to the future. Nat Rev Microbiol. 2011;9(12):889-893.
53. Kennedy DA, Read AF. Why does drug resistance readily evolve but vaccine resistance does not? Proc Biol Sci. 2017;284(1851).
54. Kennedy DA, Read AF. Why the evolution of vaccine resistance is less of a concern than the evolution of drug resistance. Proc Natl Acad Sci U S A. 2018;115(51):12878-12886.
55. Plotkin SA. Correlates of protection induced by vaccination. Clin Vaccine Immunol. 2010;17(7):1055-1065.
56. Micoli F, Bagnoli F, Rappuoli R, Serruto D. The role of vaccines in combatting antimicrobial resistance. Nat Rev Microbiol. 2021;19(5):287-302.
57. Alter G, Dowell KG, Brown EP, et al. High-resolution definition of humoral immune response correlates of effective immunity against HIV. Mol Syst Biol. 2018;14(3):e7881.
58. Antimicrobial Resistance C. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629-655.
59. Serruto D, Bottomley MJ, Ram S, Giuliani MM, Rappuoli R. The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens. Vaccine. 2012;30 Suppl 2:B87-97.
60. Schwartz M. Historical streptococci. Adv Exp Med Biol. 1997;418:1-2.
61. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis. 2005;5(11):685-694.
62. 2005 W. The current evidence for the burden of group A streptococcal diseases. WHO report. WHO, Geneva, Switzerland.
63. Joseph J. Ferretti DLS, Vincent A. FIschetti. Streptococcus pyogenes Basic Biology to Clinical Manifestations.
64. Ramos M, Valle R, Reaves EJ, et al. Outbreak of Group A beta hemolytic Streptococcus pharyngitis in a Peruvian military facility, April 2012. MSMR. 2013;20(6):14-17.
65. Turner CE, Dryden M, Holden MT, et al. Molecular analysis of an outbreak of lethal postpartum sepsis caused by Streptococcus pyogenes. J Clin Microbiol. 2013;51(7):2089-2095.

66. Ertugrul BM, Erol N, Emek M, et al. Food-borne tonsillopharyngitis outbreak in a hospital cafeteria. Infection. 2012;40(1):49-55.

67. Sumby P, Porcella SF, Madrigal AG, et al. Evolutionary origin and emergence of a highly successful clone of serotype M1 group a Streptococcus involved multiple horizontal gene transfer events. J Infect Dis. 2005;192(5):771-782.

68. Aziz RK, Kotb M. Rise and persistence of global M1T1 clone of Streptococcus pyogenes. Emerg Infect Dis. 2008;14(10):1511-1517.

69. Fittipaldi N, Beres SB, Olsen RJ, et al. Full-genome dissection of an epidemic of severe invasive disease caused by a hypervirulent, recently emerged clone of group A Streptococcus. Am J Pathol. 2012;180(4):1522-1534.

70. Beres SB, Sylva GL, Sturdevant DE, et al. Genome-wide molecular dissection of serotype M3 group A Streptococcus strains causing two epidemics of invasive infections. Proc Natl Acad Sci U S A. 2004;101(32):11833-11838.

71. Fittipaldi N, Olsen RJ, Beres SB, Van Beneden C, Musser JM. Genomic analysis of emm59 group A Streptococcus invasive strains, United States. Emerg Infect Dis. 2012;18(4):650-652.

72. Lau EH, Nishiura H, Cowling BJ, Ip DK, Wu JT. Scarlet fever outbreak, Hong Kong, 2011. Emerg Infect Dis. 2012;18(10):1700-1702.

73. Yang P, Peng X, Zhang D, et al. Characteristics of group A Streptococcus strains circulating during scarlet fever epidemic, Beijing, China, 2011. Emerg Infect Dis. 2013;19(6):909-915.

74. Cunningham MW. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev. 2000;13(3):470-511.

75. Henningham A, Barnett TC, Maamary PG, Walker MJ. Pathogenesis of group A streptococcal infections. Discov Med. 2012;13(72):329-342.

76. Mortensen R, Nissen TN, Blauenfeldt T, Christensen JP, Andersen P, Dietrich J. Adaptive Immunity against Streptococcus pyogenes in Adults Involves Increased IFN-gamma and IgG3 Responses Compared with Children. J Immunol. 2015;195(4):1657-1664.

77. O'Connor SP, Darip D, Fraley K, Nelson CM, Kaplan EL, Cleary PP. The human antibody response to streptococcal C5a peptidase. J Infect Dis. 1991;163(1):109-116.

78. Shet A, Kaplan EL, Johnson DR, Cleary PP. Immune response to group A streptococcal C5a peptidase in children: implications for vaccine development. J Infect Dis. 2003;188(6):809-817.

79. Falconer AE, Carson R, Johnstone R, Bird P, Kehoe M, Calvert JE. Distinct IgG1 and IgG3 subclass responses to two streptococcal protein antigens in man: analysis of antibodies to streptolysin O and M protein using standardized subclass-specific enzyme-linked immunosorbent assays. Immunology. 1993;79(1):89-94.

80. Kaplan EL, Markowitz M. The fall and rise of rheumatic fever in the United States: a commentary. Int J Cardiol. 1988;21(1):3-10.

81. Cole JN, Barnett TC, Nizet V, Walker MJ. Molecular insight into invasive group A streptococcal disease. Nat Rev Microbiol. 2011;9(10):724-736.

82. Walker MJ, Barnett TC, McArthur JD, et al. Disease manifestations and pathogenic mechanisms of Group A Streptococcus. Clin Microbiol Rev. 2014;27(2):264-301.

83. Cue D, Dombek PE, Lam H, Cleary PP. Streptococcus pyogenes serotype M1 encodes multiple pathways for entry into human epithelial cells. Infect Immun. 1998;66(10):4593-4601.

84. Moses AE, Wessels MR, Zalcman K, et al. Relative contributions of hyaluronic acid capsule and M protein to virulence in a mucoid strain of the group A Streptococcus. Infect Immun. 1997;65(1):64-71.

85. Hasty DL, Ofek I, Courtney HS, Doyle RJ. Multiple adhesins of streptococci. Infect Immun. 1992;60(6):2147-2152.

86. Fischetti VA. Streptococcal M protein: molecular design and biological behavior. Clin Microbiol Rev. 1989;2(3):285-314.

87. Ghosh P. Variation, Indispensability, and Masking in the M protein. Trends Microbiol. 2018;26(2):132-144.

88. Berge A, Bjorck L. Streptococcal cysteine proteinase releases biologically active fragments of streptococcal surface proteins. J Biol Chem. 1995;270(17):9862-9867.

89. Facklam RF, Martin DR, Lovgren M, et al. Extension of the Lancefield classification for group A streptococci by addition of 22 new M protein gene sequence types from clinical isolates: emm103 to emm124. Clin Infect Dis. 2002;34(1):28-38.

90. Ringdahl U, Svensson HG, Kotarsky H, Gustafsson M, Weineisen M, Sjobring U. A role for the fibrinogen-binding regions of streptococcal M proteins in phagocytosis resistance. Mol Microbiol. 2000;37(6):1318-1326.

91. Smeesters PR, McMillan DJ, Sriprakash KS. The streptococcal M protein: a highly versatile molecule. Trends Microbiol. 2010;18(6):275-282.

92. McMillan DJ, Dreze PA, Vu T, et al. Updated model of group A Streptococcus M proteins based on a comprehensive worldwide study. Clin Microbiol Infect. 2013;19(5):E222-229.

93. Sanderson-Smith M, De Oliveira DM, Guglielmini J, et al. A systematic and functional classification of Streptococcus pyogenes that serves as a new tool for molecular typing and vaccine development. J Infect Dis. 2014;210(8):1325-1338.

94. Happonen L, Hauri S, Svensson Birkedal G, et al. A quantitative Streptococcus pyogenes-human protein-protein interaction map reveals localization of opsonizing antibodies. Nat Commun. 2019;10(1):2727.

95. Macheboeuf P, Buffalo C, Fu CY, et al. Streptococcal M1 protein constructs a pathological host fibrinogen network. Nature. 2011;472(7341):64-68.

96. Whitnack E, Dale JB, Beachey EH. Common protective antigens of group A streptococcal M proteins masked by fibrinogen. J Exp Med. 1984;159(4):1201-1212.

97. Gubbe K, Misselwitz R, Welfle K, Reichardt W, Schmidt KH, Welfle H. C repeats of the streptococcal M1 protein achieve the human serum albumin binding ability by flanking regions which stabilize the coiled-coil conformation. Biochemistry. 1997;36(26):8107-8113.

98. Malmstrom J, Karlsson C, Nordenfelt P, et al. Streptococcus pyogenes in human plasma: adaptive mechanisms analyzed by mass spectrometry-based proteomics. J Biol Chem. 2012;287(2):1415-1425.

99. Pleass RJ, Areschoug T, Lindahl G, Woof JM. Streptococcal IgA-binding proteins bind in the Calpha 2-Calpha 3 interdomain region and inhibit binding of IgA to human CD89. J Biol Chem. 2001;276(11):8197-8204.

100. Nordenfelt P, Waldemarson S, Linder A, et al. Antibody orientation at bacterial surfaces is related to invasive infection. J Exp Med. 2012;209(13):2367-2381.

101. Thern A, Stenberg L, Dahlback B, Lindahl G. Ig-binding surface proteins of Streptococcus pyogenes also bind human C4b-binding protein (C4BP), a regulatory component of the complement system. J Immunol. 1995;154(1):375-386.

102. Morfeldt E, Berggard K, Persson J, et al. Isolated hypervariable regions derived from streptococcal M proteins specifically bind human C4b-binding protein: implications for antigenic variation. J Immunol. 2001;167(7):3870-3877.

103. Horstmann RD, Sievertsen HJ, Knobloch J, Fischetti VA. Antiphagocytic activity of streptococcal M protein: selective binding of complement control protein factor H. Proc Natl Acad Sci U S A. 1988;85(5):1657-1661.

104. Dinkla K, Sastalla I, Godehardt AW, et al. Upregulation of capsule enables Streptococcus pyogenes to evade immune recognition by antigen-specific antibodies directed to the G-related alpha2-macroglobulin-binding protein GRAB located on the bacterial surface. Microbes Infect. 2007;9(8):922-931.

105. Dale JB, Washburn RG, Marques MB, Wessels MR. Hyaluronate capsule and surface M protein in resistance to opsonization of group A streptococci. Infect Immun. 1996;64(5):1495-1501.

106. Foley MJ, Wood WB, Jr. Studies on the pathogenicity of group A streptococci. II. The antiphagocytic effects of the M protein and the capsular gel. J Exp Med. 1959;110:617-628.

107. Cole JN, Pence MA, von Kockritz-Blickwede M, et al. M protein and hyaluronic acid capsule are essential for in vivo selection of covRS mutations characteristic of invasive serotype M1T1 group A Streptococcus. mBio. 2010;1(4).

108. Stollerman GH, Dale JB. The importance of the group a streptococcus capsule in the pathogenesis of human infections: a historical perspective. Clin Infect Dis. 2008;46(7):1038-1045.

109. Akesson P, Sjoholm AG, Bjorck L. Protein SIC, a novel extracellular protein of Streptococcus pyogenes interfering with complement function. J Biol Chem. 1996;271(2):1081-1088.

110. Fernie-King BA, Seilly DJ, Willers C, Wurzner R, Davies A, Lachmann PJ. Streptococcal inhibitor of complement (SIC) inhibits the membrane attack complex by preventing uptake of C567 onto cell membranes. Immunology. 2001;103(3):390-398.

111. Lukomski S, Hoe NP, Abdi I, et al. Nonpolar inactivation of the hypervariable streptococcal inhibitor of complement gene (sic) in serotype M1 Streptococcus pyogenes significantly decreases mouse mucosal colonization. Infect Immun. 2000;68(2):535-542.

112. Fernie-King BA, Seilly DJ, Davies A, Lachmann PJ. Streptococcal inhibitor of complement inhibits two additional components of the mucosal innate immune system: secretory leukocyte proteinase inhibitor and lysozyme. Infect Immun. 2002;70(9):4908-4916.

113. Frick IM, Akesson P, Rasmussen M, Schmidtchen A, Bjorck L. SIC, a secreted protein of Streptococcus pyogenes that inactivates antibacterial peptides. J Biol Chem. 2003;278(19):16561-16566.

114. Bhakdi S, Tranum-Jensen J, Sziegoleit A. Mechanism of membrane damage by streptolysin-O. Infect Immun. 1985;47(1):52-60.

115. Timmer AM, Timmer JC, Pence MA, et al. Streptolysin O promotes group A Streptococcus immune evasion by accelerated macrophage apoptosis. J Biol Chem. 2009;284(2):862-871.

116. Bricker AL, Cywes C, Ashbaugh CD, Wessels MR. NAD+-glycohydrolase acts as an intracellular toxin to enhance the extracellular survival of group A streptococci. Mol Microbiol. 2002;44(1):257-269.

117. Madden JC, Ruiz N, Caparon M. Cytolysin-mediated translocation (CMT): a functional equivalent of type III secretion in gram-positive bacteria. Cell. 2001;104(1):143-152.

118. Akesson P, Moritz L, Truedsson M, Christensson B, von Pawel-Rammingen U. IdeS, a highly specific immunoglobulin G (IgG)-cleaving enzyme from Streptococcus pyogenes, is inhibited by specific IgG antibodies generated during infection. Infect Immun. 2006;74(1):497-503.

119. Collin M, Olsen A. Effect of SpeB and EndoS from Streptococcus pyogenes on human immunoglobulins. Infect Immun. 2001;69(11):7187-7189.

120. Collin M, Svensson MD, Sjoholm AG, Jensenius JC, Sjobring U, Olsen A. EndoS and SpeB from Streptococcus pyogenes inhibit immunoglobulin-mediated opsonophagocytosis. Infect Immun. 2002;70(12):6646-6651.

121. Nyberg P, Rasmussen M, Bjorck L. alpha2-Macroglobulin-proteinase complexes protect Streptococcus pyogenes from killing by the antimicrobial peptide LL-37. J Biol Chem. 2004;279(51):52820-52823.

122. Cleary PP, Prahbu U, Dale JB, Wexler DE, Handley J. Streptococcal C5a peptidase is a highly specific endopeptidase. Infect Immun. 1992;60(12):5219-5223.

123. Wexler DE, Chenoweth DE, Cleary PP. Mechanism of action of the group A streptococcal C5a inactivator. Proc Natl Acad Sci U S A. 1985;82(23):8144-8148.

124. Edwards RJ, Taylor GW, Ferguson M, et al. Specific C-terminal cleavage and inactivation of interleukin-8 by invasive disease isolates of Streptococcus pyogenes. J Infect Dis. 2005;192(5):783-790.

125. Zinkernagel AS, Timmer AM, Pence MA, et al. The IL-8 protease SpyCEP/ScpC of group A Streptococcus promotes resistance to neutrophil killing. Cell Host Microbe. 2008;4(2):170-178.

126. Norrby-Teglund A, Thulin P, Gan BS, et al. Evidence for superantigen involvement in severe group a streptococcal tissue infections. J Infect Dis. 2001;184(7):853-860.

127. Mitchell TJ. The pathogenesis of streptococcal infections: from tooth decay to meningitis. Nat Rev Microbiol. 2003;1(3):219-230.

128. Sjoholm K, Karlsson C, Linder A, Malmstrom J. A comprehensive analysis of the Streptococcus pyogenes and human plasma protein interaction network. Mol Biosyst. 2014;10(7):1698-1708.

129. Sjoholm K, Kilsgard O, Teleman J, Happonen L, Malmstrom L, Malmstrom J. Targeted Proteomics and Absolute Protein Quantification for the Construction of a Stoichiometric Host-Pathogen Surface Density Model. Mol Cell Proteomics. 2017;16(4 suppl 1):S29-S41.

130. Fritzer A, Senn BM, Minh DB, et al. Novel conserved group A streptococcal proteins identified by the antigenome technology as vaccine candidates for a non-M protein-based vaccine. Infect Immun. 2010;78(9):4051-4067.

131. Bensi G, Mora M, Tuscano G, et al. Multi high-throughput approach for highly selective identification of vaccine candidates: the Group A Streptococcus case. Mol Cell Proteomics. 2012;11(6):M111 015693.

132. Reglinski M, Gierula M, Lynskey NN, Edwards RJ, Sriskandan S. Identification of the Streptococcus pyogenes surface antigens recognised by pooled human immunoglobulin. Sci Rep. 2015;5:15825.

133. Davies MR, McIntyre L, Mutreja A, et al. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. Nat Genet. 2019;51(6):1035-1043.
134. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. Lancet Infect Dis. 2009;9(10):611-616.
135. Quinn A, Kosanke S, Fischetti VA, Factor SM, Cunningham MW. Induction of autoimmune valvular heart disease by recombinant streptococcal m protein. Infect Immun. 2001;69(6):4072-4078.
136. Cunningham MW. Molecular Mimicry, Autoimmunity, and Infection: The Cross-Reactive Antigens of Group A Streptococci and their Sequelae. Microbiol Spectr. 2019;7(4).
137. Luo YH, Chuang WJ, Wu JJ, et al. Molecular mimicry between streptococcal pyrogenic exotoxin B and endothelial cells. Lab Invest. 2010;90(10):1492-1506.
138. Dale JB, Beachey EH. Epitopes of streptococcal M proteins shared with cardiac myosin. J Exp Med. 1985;162(2):583-591.
139. Beachey EH, Gras-Masse H, Tarter A, et al. Opsonic antibodies evoked by hybrid peptide copies of types 5 and 24 streptococcal M proteins synthesized in tandem. J Exp Med. 1986;163(6):1451-1458.
140. Dale JB, Chiang EY, Lederer JW. Recombinant tetravalent group A streptococcal M protein vaccine. J Immunol. 1993;151(4):2188-2194.
141. Dale JB. Multivalent group A streptococcal vaccine designed to optimize the immunogenicity of six tandem M protein fragments. Vaccine. 1999;17(2):193-200.
142. Dale JB, Simmons M, Chiang EC, Chiang EY. Recombinant, octavalent group A streptococcal M protein vaccine. Vaccine. 1996;14(10):944-948.
143. Hu MC, Walls MA, Stroop SD, Reddish MA, Beall B, Dale JB. Immunogenicity of a 26-valent group A streptococcal vaccine. Infect Immun. 2002;70(4):2171-2177.
144. Dale JB, Penfound TA, Chiang EY, Walton WJ. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. Vaccine. 2011;29(46):8175-8178.
145. Kotloff KL, Corretti M, Palmer K, et al. Safety and immunogenicity of a recombinant multivalent group a streptococcal vaccine in healthy adults: phase 1 trial. JAMA. 2004;292(6):709-715.
146. McNeil SA, Halperin SA, Langley JM, et al. Safety and immunogenicity of 26-valent group a streptococcus vaccine in healthy adult volunteers. Clin Infect Dis. 2005;41(8):1114-1122.

147. Batzloff MR, Hayman WA, Davies MR, et al. Protection against group A streptococcus by immunization with J8-diphtheria toxoid: contribution of J8- and diphtheria toxoid-specific antibodies to protection. J Infect Dis. 2003;187(10):1598-1608.

148. Brandt ER, Sriprakash KS, Hobb RI, et al. New multi-determinant strategy for a group A streptococcal vaccine designed for the Australian Aboriginal population. Nat Med. 2000;6(4):455-459.

149. Postol E, Alencar R, Higa FT, et al. StreptInCor: a candidate vaccine epitope against S. pyogenes infections induces protection in outbred mice. PLoS One. 2013;8(4):e60969.

150. Sabharwal H, Michon F, Nelson D, et al. Group A streptococcus (GAS) carbohydrate as an immunogen for protection against GAS infection. J Infect Dis. 2006;193(1):129-135.

151. Ji Y, Carlson B, Kondagunta A, Cleary PP. Intranasal immunization with C5a peptidase prevents nasopharyngeal colonization of mice by the group A Streptococcus. Infect Immun. 1997;65(6):2080-2087.

152. Ulrich RG. Vaccine based on a ubiquitous cysteinyl protease and streptococcal pyrogenic exotoxin A protects against Streptococcus pyogenes sepsis and toxic shock. J Immune Based Ther Vaccines. 2008;6:8.

153. Turner CE, Kurupati P, Wiles S, Edwards RJ, Sriskandan S. Impact of immunization against SpyCEP during invasive disease with two streptococcal species: Streptococcus pyogenes and Streptococcus equi. Vaccine. 2009;27(36):4923-4929.

154. Henningham A, Chiarot E, Gillen CM, et al. Conserved anchorless surface proteins as group A streptococcal vaccine candidates. J Mol Med (Berl). 2012;90(10):1197-1207.

155. Lyon WR, Caparon MG. Trigger factor-mediated prolyl isomerization influences maturation of the Streptococcus pyogenes cysteine protease. J Bacteriol. 2003;185(12):3661-3667.

156. Lyon WR, Gibson CM, Caparon MG. A role for trigger factor and an rgg-like regulator in the transcription, secretion and processing of the cysteine proteinase of Streptococcus pyogenes. EMBO J. 1998;17(21):6263-6275.

157. Sharma A, Arya DK, Sagar V, Bergmann R, Chhatwal GS, Johri AK. Identification of potential universal vaccine candidates against group A Streptococcus by using high throughput in silico and proteomics approach. J Proteome Res. 2013;12(1):336-346.

158. Rodriguez-Ortega MJ, Norais N, Bensi G, et al. Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. Nat Biotechnol. 2006;24(2):191-197.

159. James P. Protein identification in the post-genome era: the rapid rise of proteomics. Q Rev Biophys. 1997;30(4):279-331.

160. Milo R. What is the total number of protein molecules per cell volume? A call to rethink some published values. Bioessays. 2013;35(12):1050-1055.

161. Wilkins MR, Sanchez JC, Gooley AA, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev. 1996;13:19-50.

162. Wilhelm M, Schlegl J, Hahne H, et al. Mass-spectrometry-based draft of the human proteome. Nature. 2014;509(7502):582-587.

163. Kim MS, Pinto SM, Getnet D, et al. A draft map of the human proteome. Nature. 2014;509(7502):575-581.

164. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng. 2009;11:49-79.

165. Smith LM, Kelleher NL, Consortium for Top Down P. Proteoform: a single term describing protein complexity. Nat Methods. 2013;10(3):186-187.

166. Vandermarliere E, Mueller M, Martens L. Getting intimate with trypsin, the leading protease in proteomics. Mass Spectrom Rev. 2013;32(6):453-465.

167. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989;246(4926):64-71.

168. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422(6928):198-207.

169. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods. 2007;4(9):709-712.

170. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature. 2016;537(7620):347-355.

171. Old WM, Meyer-Arendt K, Aveline-Wolf L, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics. 2005;4(10):1487-1502.

172. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007;4(3):207-214.

173. Gillet LC, Navarro P, Tate S, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012;11(6):O111 016717.

174. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol. 2018;14(8):e8126.

175. Rost HL, Rosenberger G, Navarro P, et al. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol. 2014;32(3):219-223.

176. Piersimoni L, Kastritis PL, Arlt C, Sinz A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions horizontal line A Method for All Seasons. Chem Rev. 2022;122(8):7500-7531.

177. Chowdhury S, Happonen L, Khakzad H, Malmstrom L, Malmstrom J. Structural proteomics, electron cryo-microscopy and structural modeling approaches in bacteria-human protein interactions. Med Microbiol Immunol. 2020;209(3):265-275.

178. Liu F, Rijkers DT, Post H, Heck AJ. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nat Methods. 2015;12(12):1179-1184.

179. Rappsilber J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. J Struct Biol. 2011;173(3):530-540.

180. Hauri S, Khakzad H, Happonen L, Teleman J, Malmstrom J, Malmstrom L. Rapid determination of quaternary protein structures in complex biological samples. Nat Commun. 2019;10(1):192.

181. Chowdhury S, Khakzad H, Bergdahl GE, et al. Streptococcus pyogenes Forms Serotype- and Local Environment-Dependent Interspecies Protein Complexes. mSystems. 2021;6(5):e0027121.

182. Opuni KFM, Al-Majdoub M, Yefremova Y, El-Kased RF, Koy C, Glocker MO. Mass spectrometric epitope mapping. Mass Spectrom Rev. 2018;37(2):229-241.

183. Stander S, L RG, Scarselli M, Norais N, Rand K. Epitope Mapping of Polyclonal Antibodies by Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS). Anal Chem. 2021;93(34):11669-11678.

184. Engen JR, Botzanowski T, Peterle D, Georgescauld F, Wales TE. Developments in Hydrogen/Deuterium Exchange Mass Spectrometry. Anal Chem. 2021;93(1):567-582.

185. Akerstrom B, Brodin T, Reis K, Bjorck L. Protein G: a powerful tool for binding and detection of monoclonal and polyclonal antibodies. J Immunol. 1985;135(4):2589-2592.

186. Sauer-Eriksson AE, Kleywegt GJ, Uhlen M, Jones TA. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. Structure. 1995;3(3):265-278.

187. Nero TL, Morton CJ, Holien JK, Wielens J, Parker MW. Oncogenic protein interfaces: small molecules, big challenges. Nat Rev Cancer. 2014;14(4):248-262.

188. Angeletti D, Yewdell JW. Understanding and Manipulating Viral Immunity: Antibody Immunodominance Enters Center Stage. Trends Immunol. 2018;39(7):549-561.

189. Zeng MY, Cisalpino D, Varadarajan S, et al. Gut Microbiota-Induced Immunoglobulin G Controls Systemic Infection by Symbiotic Bacteria and Pathogens. Immunity. 2016;44(3):647-658.

190. Lindahl G. Subdominance in Antibody Responses: Implications for Vaccine Development. Microbiol Mol Biol Rev. 2020;85(1).

191. Burton DR, Parren PW. Vaccines and the induction of functional antibodies: time to look beyond the molecules of natural infection? Nat Med. 2000;6(2):123-125.

192. Stalhammar-Carlemalm M, Waldemarsson J, Johnsson E, Areschoug T, Lindahl G. Nonimmunodominant regions are effective as building blocks in a streptococcal fusion protein vaccine. Cell Host Microbe. 2007;2(6):427-434.

193. Lannergard J, Gustafsson MC, Waldemarsson J, Norrby-Teglund A, Stalhammar-Carlemalm M, Lindahl G. The Hypervariable region of Streptococcus pyogenes M protein escapes antibody attack by antigenic variation and weak immunogenicity. Cell Host Microbe. 2011;10(2):147-157.

194. Luchini A, Espina V, Liotta LA. Protein painting reveals solvent-excluded drug targets hidden within native protein-protein interfaces. Nat Commun. 2014;5:4413.

195. Coffey JA, Jennings KR, Dalton H. New antigenic regions of streptokinase are identified by affinity-directed mass spectrometry. Eur J Biochem. 2001;268(19):5215-5221.

196. Roth AJ, Ooi JD, Hess JJ, et al. Epitope specificity determines pathogenicity and detectability in ANCA-associated vasculitis. J Clin Invest. 2013;123(4):1773-1783.

197. Gaworzewska E, Colman G. Changes in the pattern of infection caused by Streptococcus pyogenes. Epidemiol Infect. 1988;100(2):257-269.

198. Eriksson BK, Norgren M, McGregor K, Spratt BG, Normark BH. Group A streptococcal infections in Sweden: a comparative study of invasive and noninvasive infections and analysis of dominant T28 emm28 isolates. Clin Infect Dis. 2003;37(9):1189-1193.

199. Luyendyk JP, Schoenecker JG, Flick MJ. The multifaceted role of fibrinogen in tissue injury and inflammation. Blood. 2019;133(6):511-520.

200. Shannon O. The role of platelets in sepsis. Res Pract Thromb Haemost. 2021;5(1):27-37.

201. Dahlback B, Frohm B, Nelsestuen G. High affinity interaction between C4b-binding protein and vitamin K-dependent protein S in the presence of calcium. Suggestion of a third component in blood regulating the interaction. J Biol Chem. 1990;265(27):16082-16087.

202. Garcia de Frutos P, Alim RI, Hardig Y, Zoller B, Dahlback B. Differential regulation of alpha and beta chains of C4b-binding protein during acute-phase response resulting in stable plasma levels of free anticoagulant protein S. Blood. 1994;84(3):815-822.
203. Rezende SM, Simmonds RE, Lane DA. Coagulation, inflammation, and apoptosis: different roles for protein S and the protein S-C4b binding protein complex. Blood. 2004;103(4):1192-1201.

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

About the author

Sounak Chowdhury was always fascinated by Biology, which resulted him to pursue a Master of Technology in Biotechnology. During his Master's he developed an interest towards infections and the human immune system. His inquisitiveness towards infections further lead him to continue his studies as a Doctoral student at Lund University, Sweden. This thesis is his efforts to untangle and understand protein interactions during *Streptococcus pyogenes* infections.

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences, Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2022:134
ISBN 978-91-8021-296-0
ISSN 1652-8220