1 Introduction

1.1 The pneumococcus

Streptococcus pneumoniae, or simply “the pneumococcus”, is a Gram-positive diplococci that colonizes the upper respiratory tract (nasopharynx) of many healthy individuals (up to 80% in some settings) without causing disease [1]. However, it is also a bacterial pathogen able to breach the host defences thus causing disease. The mechanisms by which *S. pneumoniae* causes disease are not fully understood but often occur secondary to another respiratory tract infection making it an opportunistic pathogen [2, 3]. The pneumococcus causes a wide range of diseases in its host including less serious but more frequent diseases such as otitis media and sinusitis to life-threatening diseases such as meningitis, bacteraemia and sepsis [4, 5]. It is therefore, a very important cause of mortality and morbidity globally especially in children under the age of 5, patients with cardiopulmonary disease, immunocompromised patients as well as elderly people [6, 7].

S. pneumoniae has almost 100 known serotypes based on the antisera binding pattern [8]. The capsule is one of the most important virulence determinants of the pneumococcus and some capsular types are known to be more important than others in causing invasive disease [6]. Much of the diversity of these immunogenic capsules is believed to be caused by the selective pressure exerted by the host immune system [9]. It is on the basis of this knowledge that the currently licenced vaccines have been developed. These vaccines currently target only a subset of the most virulent capsules. This has now led to a reduction in carriage and disease of serotypes included in the vaccines (vaccine type (VT) serotypes) and an increase in carriage and disease of non-vaccine type (NVT) serotypes [10-12]. This phenomenon known as serotype replacement is apparent in many vaccinated populations including The Gambia [9, 12-15]. Further, serotype switching (strains acquiring a different set of capsule synthesis genes), has also been observed in vaccinated populations and vaccine pressure is thought to play some role in this [16, 17], since the currently licensed vaccines all target the capsule. Consistently, Croucher *et al*. [18] have shown that recombination hotspots seems to be concentrated around antibiotic resistant

genes (*tetM*) and surface exposed proteins, which are potential vaccine candidates such as *pspA, psrP* and *pspC* as well as the capsule locus.
However, serotype switching is a natural part of pneumococcal evolution and has been known to have occurred decades before the introduction of vaccines [19]. Indeed, a study has revealed that the serotype switching event to serotype 19A, by a 23F lineage not previously found to have serotype 19A capsule, is thought to have occurred more than 10 years prior to the introduction of the PCV7 vaccine which targets 23F [20]. This shows that the 19A variants existed prior to PCV7 and have expanded to detectable levels after the selective reduction of VT serotypes by the vaccine.

The pneumococcus has a single circular chromosome that is approximately 2.1Mb in size with a G + C content of about 40% [21, 22]. Genome annotations have identified over 2000 protein coding genes in both TIGR4 and the un-encapsulated R6 strain but only over 60% of these have been assigned a biological function, leaving a great number of genes that could have vital roles in both disease and colonisation yet to be discovered [21]. Non-coding RNAs are also present with both TIGR4 and R6 having 4 rRNA operons and several other tRNAs [21, 22]. The pneumococcus has many insertion sequences (IS), which make up approximately 5% of the TIGR4 genome. Additionally, they possess a wide array of ATP-dependent transporters including the iron transporters, zinc transporters and manganese transporters. However, the most abundant transporters are the sugar transporters, which make up about 30% of all the transporters [21]. Some of these proteins are essential for full virulence of the pneumococcus therefore, they are being investigated as potential vaccine candidates [23, 24].
Also, the capsule synthesis genes, which determine the serotype are flanked by two conserved genes, *aliA* and *dexB* [25].

1.2 Pneumococcal colonisation

Colonisation is a prerequisite for infection and is more prevalent in younger children (<5 years of age) and can be as high as 80% in some countries but reduces to less than 10% as they reach adulthood [1, 26]. Most children in developing countries, would have been colonised with the pneumococcus at some stage of their childhood [27].

Carriage strains can be horizontally transferred from one individual to another and some of the risk factors for this include crowded areas such as day-care centres, hospitals, and prisons. Most of this horizontal transfer is believed to occur within children who are the main reservoirs of carriage [4]. Serotypes can be carried singly or simultaneously with other serotypes in the nasopharynx and this may last from a few weeks to months before being cleared and replaced by another type [1]. Although the length of carriage of different serotypes varies with some serotypes found much less often in carriage than others [28], this trait however does not affect a serotype's invasiveness, with some serotypes rarely found in colonisation studies shown to be amongst the most invasive [28, 29]. Furthermore, strains do compete against one another for colonisation. Some pneumococci produce strain-specific pneumocins which are inhibitory to other pneumococcal strains thus out competing them during colonisation [30]. This is why, when these more prevalent strains are cleared out due to vaccination, they are replaced by the previously suppressed strains [31].

Even though the capsular polysaccharide is the main determinant of immunogenicity, it is less important during carriage, thus the observed prevalence of transparent strains (strains with less capsule expression) in carriage [1, 4]. Consequently, the pneumococcus is known to express an array of proteins beneath the capsule that are essential for adhesion and thus colonisation [4]. These proteins interact with the host epithelial cells and ensure that the bacteria are anchored sufficiently to prevent innate immune clearance by ciliary movement [32].
Conversely, strains with increased expression of capsular polysaccharides are more often isolated in invasive disease because the capsule helps protect the pneumococcus against phagocytosis [1, 4]. The importance of capsule in invasive disease is further supported by the fact that un-encapsulated pneumococci are rarely if ever seen in invasive disease [1, 4].

It is worth mentioning that even though colonisation precedes disease, most colonised individuals do not go on to develop disease. The reason for this is not fully understood but transition to disease often requires the generation of local inflammation factors including tissue necrotic factor F and interleukin 1 [33]. Subsequently, this increases the number of receptors on their target cell (host cells) including the platelet-activation factor (PAF) receptor. The pneumococcus takes advantage of this scenario to bind to

the PAF receptor which facilitates internalisation *en route* to causing invasive disease as depicted in Fig. 1.1 [1, 4, 33]. The host immune system may also play a role as disease burden is greatest among young children whose immune system hasn't fully developed, the elderly and immunocompromised individuals [34].

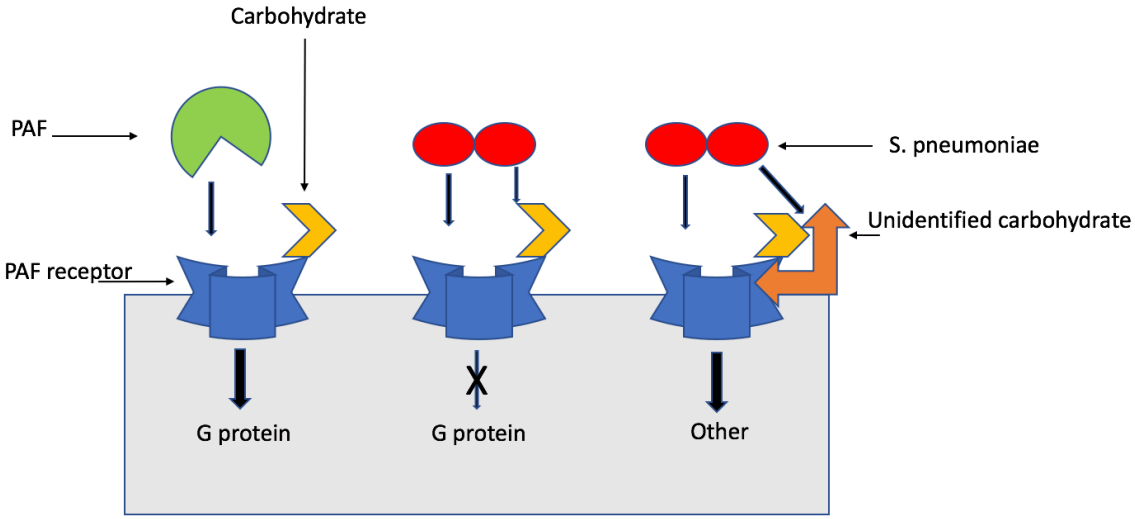

Figure 1.1 Interaction of the pneumococcus with PAF receptor.
The left illustrates PAF binding to the PAF receptor in a choline-depending fashion, thereby eliciting a G protein signal. The middle and the right diagram illustrate two proposed mechanisms of pneumococcal binding to the PAF receptor. The middle proposes that it engages both the PAF receptor and a carbohydrate on the PAF receptor or as depicted on the diagram on the right, it binds to the PAF receptor and another carbohydrate from an unidentified receptor that co-caps with the PAF receptor.
Picture adapted from [33].

1.3 Natural immune response to pneumococcal colonisation

There is evidence that the cytokine IL17A, which activates neutrophils, plays a significant role as an effector of rapid pneumococcal clearance in mice when challenged with non-encapsulated pneumococcal cells [35]. This clearance is antibody independent and it has also been demonstrated in *in vitro* studies that human IL17A cytokines were independently sufficient to induce pneumococcal killing by neutrophils [36, 37]. Further, there is proof of the role of CD4+ T cells in providing protection against colonisation and invasive pneumococcal disease in mouse models [38, 39]. Malley *et al*. argued that CD4+ T cells are the main source of protection against recolonization rather than antibodies and that this protection is serotype independent [39]. They further disputed the absolute necessity of antibodies in protecting against colonisation due to the fact that children often build resistance against invasive pneumococcal disease (IPD) from all serotypes before the appearance of measurable levels of anti-capsular antibodies [39].

1.4 Epidemiology and Burden

Pneumococcal diseases are a major problem worldwide but more so in resource limited countries. In 2000, the burden of serious pneumococcal diseases was estimated at 14.5 million cases worldwide which resulted in about 826,000 deaths in children between the ages of 1 and 59 months [40]. Furthermore, mortality due to pneumococcal diseases is estimated to account for about 11% of all deaths in HIV-negative children under the age of 5 globally and unsurprisingly, more than 60% of deaths occurred in sub-Saharan Africa and south Asia [40]. Nasopharyngeal colonization precedes invasive disease and pneumococcal carriage in healthy Gambian children under the age of 5 was shown to be 80% [26]. However, other studies have shown that the prevalence of carriage can vary between countries or even between cities of the same country [41, 42].

The most prevalent IPD is pneumonia, accounting for over 95% of pneumococcal disease/cases with pneumococcal meningitis reported to account for only 0.7% of all IPD worldwide [40]. IPD is responsible for most of the mortality caused by the pneumococcus however, some serotypes have a higher propensity to cause invasive disease than others [43]. Before the introduction of conjugate vaccines, PCV-7

serotypes accounted for about 90% of IPD in American and Canadian young children and at least 60% in all other regions except Asia, where they account for only 45% of IPD [6]. Nevertheless, the contribution of these serotypes to IPD was significantly lower in Europe than in USA and Canada or in Oceania, perhaps due to the significantly higher prevalence of serotype 1 and 5 in Europe than in these other regions [6].

1.5 Clinical Disease

As mentioned above, the pneumococcus causes a wide range of diseases, also depicted in Fig. 1.2.

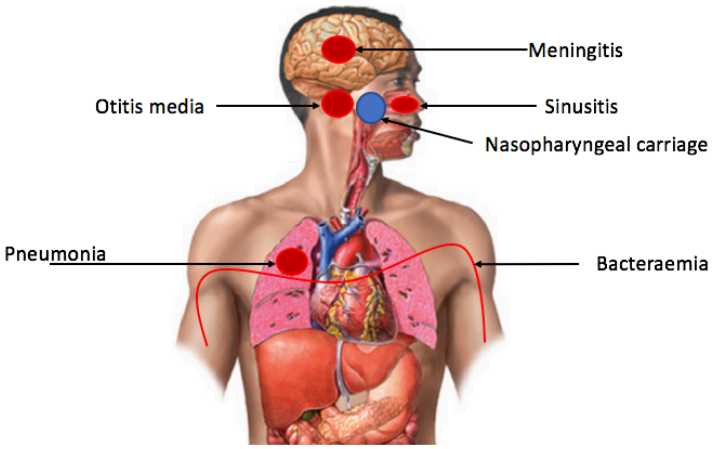

Figure 1.2 Pneumococcal diseases and their anatomic sites.
Schematic diagram showing the diseases caused by the pneumococcus and their site of infection.
Adapted from (https://www.slideshare.net/meningitis/human-28372932)

1.5.1 Pneumonia

Pneumonia is an infection of the lower respiratory tract often caused by bacteria but also caused by viruses [44]. *S. pneumoniae* has been implicated as the leading bacterial cause of pneumonia and this may be accompanied with bacteraemia in some instances [45]. Clinically, the signs and symptoms of pneumonia include fever and chills, rapid or difficult breathing, coughs and chest pain [46].

Pneumococcal pneumonia is by far the most common form of IPD globally [34]. Before the introduction of the PCVs, pneumococcal pneumonia was responsible for about 53% of all IPD in the USA [7]. *S. pneumoniae* is also the leading bacterial cause of community acquired pneumonia in European adults [47]. Furthermore, etiological studies of lower respiratory tract infections done in Zimbabwe and The Gambia both showed *S. pneumoniae* to be the commonest bacterial cause of pneumonia with incidence rates of 46% and 61% of patients respectively [44, 48]

1.5.2 Meningitis

Meningitis, as the name implies is the inflammation of the meninges and *S. pneumoniae* was the second most prevalent bacterial cause of meningitis before the introduction of conjugate vaccines against *Haemophilus influenzae* type b. However, it is now the number 1 cause in many countries [49]. Pneumococcal meningitis is the most devastating IPD that leads to death in up to 50% of patients without treatment. Also, long term consequences such as hearing loss, learning difficulties, seizures as well as brain damage may occur [7, 50]. In a 2010 meningitis outbreak in a Ghanaian district, *S. pneumoniae* was the leading cause of meningitis accounting for almost half (49%) of all bacteria isolated in the study [50]. Indeed, similar observations have been made previously in Ghana and neighbouring Burkina Faso between 2000- 2003 and 2002-2005 respectively, as well as in Malawi [51-53]. In all these studies, the case-fatality rate was approximately 40% with serotype 1 being the most prevalent serotype [51, 52]. In a more recent meningitis outbreak in Ghana (2015-2016), *S. pneumoniae* was again the leading cause accounting for 77% of all bacterial isolates of which 80% were serotype 1 [51, 54]. This recent outbreak is cause for concern because it occurred three years post-PCV13 introduction in Ghana [54]. Nonetheless, less than

5% of cases occurred in children under 5 years, which is the vaccinated group [54]. It is interesting to note however that these outbreaks resemble meningococcal meningitis serogroup A outbreaks, showing high levels of seasonality with peak incidences occurring between March and April. The fact that these countries fall within the meningitis belt, which runs from Senegal in the West to part of Ethiopia in the East further suggests that climate may be a factor [51, 52, 55]. Unfortunately, in Burkina Faso, the current licensed conjugate vaccines are not adequate for most of the serotypes implicated in pneumococcal meningitis [55]. PCV7 would cover only 33% of paediatric meningitis serotypes and a meagre 10% of adult meningitis serotypes [55]. Together, these results further support the need for vaccines with broader serotype coverage because the serotypes implicated in meningitis might be different from those mostly found in pneumonia *per se*.

1.5.3 Bacteraemia

Bacteraemia in its simplest term means the isolation of bacteria from blood and it is often secondary bacteraemia as a consequence of severe pneumonia or meningitis as they progress to causing death [45, 56]. When bacteraemia occurs without a known anatomic source of the infecting bacteria, it is called primary bacteraemia [57]. Bacteraemia is a prominent cause of death especially in young children and has been implicated as the cause of death in one-third of infants (<60 days) and a quarter of children older than 1 year in sub-Saharan Africa [58]. Interestingly, in that study and another from the same region, *S. pneumoniae* was the most commonly isolated bacteria from the blood cultures [56, 58]. Additionally, *S. pneumoniae* was amongst the most important pathogens in bacteraemia in studies outside Africa including Australia, England, USA and Denmark, which consistently showed it to be amongst the top 5 most commonly isolated pathogens in each study and overall the third most common pathogen [57].

Host genetics have been recently implicated to play a role in susceptibility to pneumococcal bacteraemia. Through genome-wide association studies (GWAS), an

association with polymorphisms in two neutrophil expressed long intergenic non-coding RNA genes (lincRNA) was found in Kenyan children [56].

1.5.4 Otitis Media

Furthermore, the pneumococcus also causes non-invasive diseases, including otitis media (OT), which is a middle ear infection. Although this infection is less severe than IPDs, it is much more prevalent causing high morbidity in children. Prolonged untreated OT may result in ear and development problems [59, 60]. The burden of OT can be as high as 1.5 million annual cases in some regions of the world [61] and studies have shown *S. pneumoniae* to be the most important pathogen in otitis media [61, 62].

1.6 Antibiotic resistant pneumococcus

Antibiotic resistance is a global problem with many pathogens acquiring resistance to various classes of antibiotics and threatening to lead us to a post-antibiotic era. The data on resistant pneumococcal strains varies between continents and indeed between countries on the same continent [27, 63, 64]. In the Pre-conjugate vaccine era, some countries observed an increase in antibiotic non-susceptible *S. pneumoniae*. In a study in South Africa, as much as 95% of their hospital-acquired strains were resistant to penicillin [64]. Also, this study reported that as much as 9%, 12% and 4% of all their isolates were resistant to chloramphenicol, tetracycline, and erythromycin respectively [64]. An increase in cefotaxime resistance was also reported [64]. Of further concern is the observed increase in resistance to other antibiotics unrelated to penicillin such as vancomycin, which is the last line drug used in pneumococcal diseases [65]. In The Gambia, there was a slight but insignificant increase in penicillin, chloramphenicol or trimethoprim-sulfamethoxazole resistance amongst invasive pneumococcal isolates between 1996-2003 [66]. However, the increasing trend will most likely continue because of antibiotic abuse in many countries including The Gambia, where antibiotics can be obtained without prescription.

1.7 Licenced vaccines

Because of the large disease burden caused by the pneumococcus coupled with its rapidly decreasing susceptibility to the available antibiotics, it is necessary to explore the benefits of vaccination. The current WHO-recommended vaccine schedule of the conjugate vaccines is either three primary doses (the 3+ 0 schedule) or the 2 + 1 schedule, which includes two primary doses and a single booster dose [67]. The choice of schedule is completely dependent on the pneumococcal disease epidemiology of the population, the coverage as well as the timeliness of the vaccine doses [67].

Initially, a 14-valent polysaccharide vaccine was used until it was replaced in 1983 by the still in use 23-valent polysaccharide vaccine. This vaccine contained 23 of the most common serotypes implicated in invasive disease, accounting for about 88% of invasive disease [68]. However, this vaccine is only about 60% efficacious and it is not immunogenic in children younger than 2 years and the elderly (>65 years) who comprise the main risk groups [69]. Consequently, in 2000, the first conjugate vaccine was introduced in the USA, which included 7 serotypes conjugated to a non-toxin form of the diphtheria toxin protein [69].

Although this is more immunogenic in younger children and was licensed to be used in that age group, the conjugation to the protein meant that less serotypes could be incorporated in the vaccine. Nonetheless, this vaccine can induce a T-cell dependent immune response and studies including PCV-7, -9 and -11 have shown a disease risk reduction range between 62 and 89% in children less than 24 months old [68, 70]. Also, PCV13 has been shown to reach approximately 93% efficacy against VT invasive disease in children (<15 years) in Madrid in 2014-2015 [71].

To further improve vaccine coverage in all regions especially Asia, where PCV7 was less efficacious, another vaccine was formulated to include all the serotypes in PCV-7 plus serotype 1 and 5 to make the 9-valent vaccine. Together these serotypes accounted for approximately 66% of all IPD in Asia and >82% of IPD in all the other regions [6]. Furthermore, another two serotypes (3 and 7F) were added to the 9-valent formulation to make the 11-valent vaccine and this further improved the vaccine

coverage of IPD in most regions by 2.6-6.5% and a substantial increase of 9% in Asia [6]. The 10-valent vaccine, which has eight serotypes conjugated to non- *Haemophilus influenzae* protein D and 2 other serotypes conjugated to either tetanus or diphtheria toxoids is also in use [72]. This vaccine has all the PCV9 serotypes plus serotype 7F [72].

Most countries, including The Gambia, have now included the PCV13 to their national immunisation programmes because it improves serotype coverage even further with two additional serotypes (6A and 19A) added to the 11-valent formulation [73-75].
Studies carried in vaccinated populations have seen significant reductions in the rates of IPD incidence as well as a reduction in VT serotype carriage in both vaccinated and unvaccinated individuals [76, 77]. The reduction of VT-carriage in non-vaccinated individuals is due to herd immunity, however, this is masked by the increase in carriage of NVT serotypes thus keeping carriage rates fairly constant [76, 78, 79].

1.8 Limitations of the current licenced vaccines

As indicated above, the currently licensed vaccines are limited in several ways including, low serotype coverage, serotype replacement and an increase in NVT IPD. All these limitations have prompted renewed efforts to identify conserved surface exposed proteins across all serotypes that are capable of inducing sufficient immune responses. Several identified proteins are in advanced stages of vaccine development [65, 68]
Further, the expensive cost of the polysaccharide conjugate vaccines is also a hindrance especially for low-income countries who are most affected by pneumococcal diseases. This makes the development of cheaper vaccines even more important. In the conjugate vaccines, every serotype is conjugated separately to the protein and this is the reason these vaccines are costly and can contain only a limited number of serotypes in one formulation [68]. To circumvent this problem, it is essential to manufacture protein-based vaccines which are relatively cheaper to produce.

1.9 Surface Proteins

Streptococcus pneumoniae like many other bacteria possess several surface exposed proteins that interact with host tissues and are believed to be essential for pathogenicity and survival *in vivo* by helping to conceal the bacterial surface from host defence mechanisms [69]. These proteins can be differentiated and grouped together based on their mechanism of attachment on to the cell surface. These groups include Choline Binding Proteins (CBP), proteins covalently bound to the peptidoglycan, histidine triad family macromolecules and those directly attached to lipids of the cytoplasmic membrane [80]. After sequencing the genome of TIGR4 by Tettelin *et al*., [21], they predicted a total of 36 lipoproteins in that genome. Generally, the pneumococcus possesses many lipoproteins performing many different roles with some expressed on the outer surface of membranes and others remain within the inner membrane [81]. Common to all lipoproteins are their identifiable signal peptides linked to their amino termini, which is essential for their transport through the cell membrane to the outer membrane but it does not exclusively determine which lipoproteins get to the outer membrane [69, 81]. Subsequent to its transport to the outer membrane, it undergoes further modification, which generally occurs in three steps. First, diacylglyceride is transferred to the cysteine sulphydryl group of the unmodified prolipoprotein. Second, signal peptidase II, which is specific for lipoproteins cleaves the signal peptide thus forming an apolipoprotein and finally, there is acetylation of the α-amino group of the conserved N-terminal cysteine residue (Fig.1.3) [81, 82]. This leads to a highly hydrophobic amino terminus, believed to be firmly associated with the hydrophobic membrane [81].

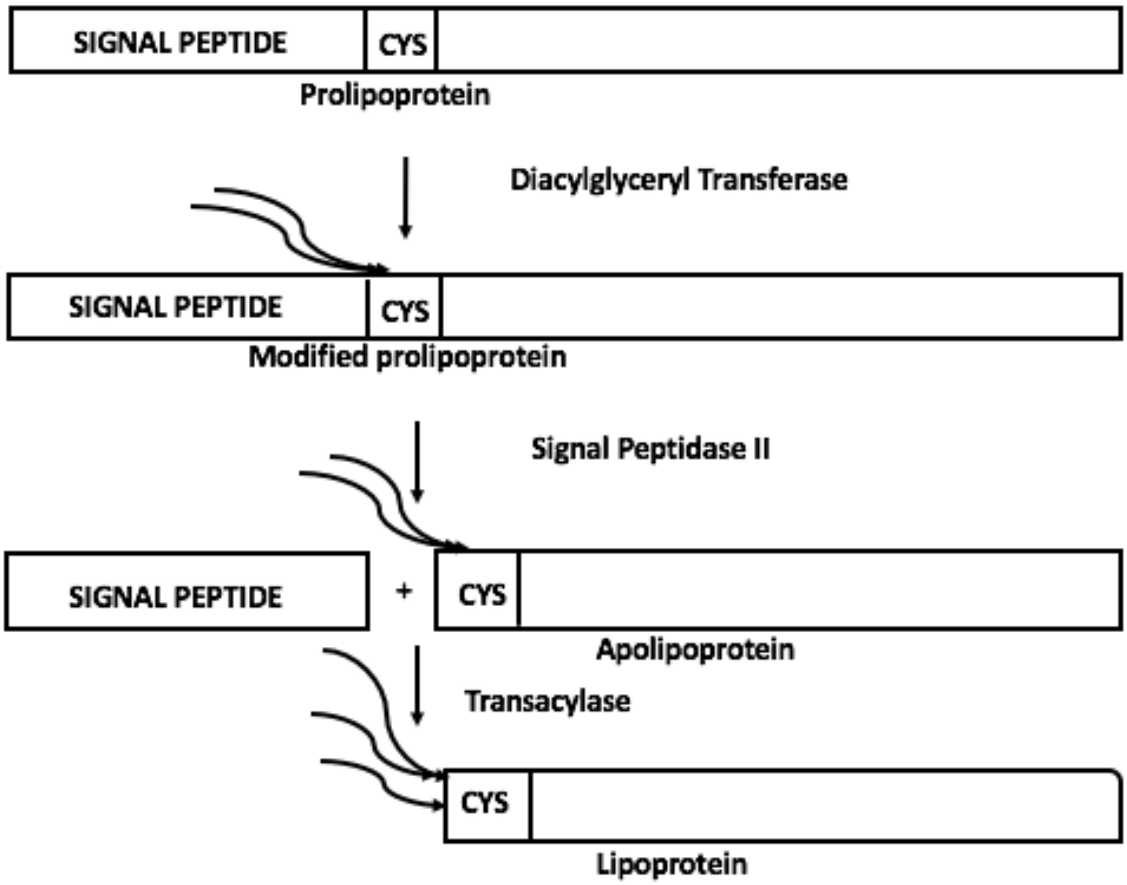

Figure 1.3 Post-translation modification of lipoproteins.
Modification process of the lipoprotein post export to the outer membrane. The signal peptide is to the left of the cysteine and the catalytic enzymes are written to the right of the reaction arrows. Adapted from Juncker *et al*. [82]

1.10 Some Protein Vaccines in the pipeline

As mentioned above, the limitations of the currently licenced vaccines have prompted research into finding alternative vaccine candidates. The most interesting candidates with such potential seem to be the pneumococcal proteins themselves. A number of these proteins have been promising as they have been shown to be essential for full virulence of the pneumococcus and vaccination with these proteins have been protective in mouse models of infection [23, 24, 83]. Two of these proteins were recently examined in a phase-II clinical trial in The Gambia to determine their efficacy against pneumococcal carriage but have been shown to be ineffective [84]. These

proteins, pneumococcal enzyme pneumolysin (Ply) and pneumococcal histidine triad protein D (PhtD) were given in two different doses (10μg or 30μg) alongside pneumococcal 10-valent vaccine conjugated to non-typable *Haemophilus influenzae* protein D (PHiD-CV). The efficacy was measured by the prevalence of non-PHiD-CV serotypes in the nasopharynx but rather than seeing a reduction in these serotypes, there was an increase due to the void left by the reduction of the 10-valent vaccine serotypes [84]. Contrariwise, prior studies involving these proteins in mouse models were very promising with a study showing that passively vaccinated mice with human antibodies raised against these two proteins conferred protection against nasopharyngeal colonisation [85] and vaccinating with recombinant proteins protected against pneumonia [86].

Another pneumococcal surface protein that has been studied extensively with great promise is the pneumococcal surface protein A (PspA) [87, 88]. Recently, mice vaccinated with recombinant PspA (rPspA) and genetically modified pneumolysin (PdT) with recombinant BCG as an adjuvant and subsequently giving a booster of rPspA-PdT has been shown to protect these mice against lethal challenges [87]. The antibodies raised against both proteins were sufficiently high with a favourable shift in antibody class from IgG1 to IgG2. Vaccination also enhanced complement deposition and nullified the cytotoxic effect of Ply [87]. However, it has to be mentioned that even though there have been no experimental data to support this, there is fear that anti-PspA may react with myosin thus leading to an autoimmune disease [89].

Furthermore, pneumococcal choline-binding protein A (PcpA) has been studied for its immunogenicity and protection in animal models. Briefly, this candidate was shown to be protective against invasive pneumococcal diseases and has been shown to be both immunogenic and safe in a Phase I clinical trial when conjugated to PhTD or given alone (monovalent) [90, 91].

Together these are all encouraging steps towards finding an effective vaccine that is cost-effective, serotype independent and immunogenic in the main risk groups. Some candidates are shown to be more protective against colonisation [92], some more effective against systemic challenges [93] and some effective in both models [89]. Some are protective when used alone while some have shown synergistic effects when used in combination formulas [23, 89]. However, there remains a need to explore

other potential proteins that may offer better coverage and might be more immunogenic to use as vaccines.

1.11 Application of Whole Genome Sequencing

The decreasing cost of next generation sequencing has led to more whole genome sequencing (WGS) projects. WGS of pathogens has enabled us to do several simultaneous studies on the same pathogen thus increasing our capacity to answer many biological questions. Since the sequencing of the pneumococcal genome by Tettelin *et al*. [21] in 2001, many studies utilising WGS have been done to improve our understanding of this important pathogen. Studies focused on the evolution of this pathogen have enhanced our understanding of its interaction with its host and other bacteria in the same niche. Further, we have better understanding about the mechanisms that contribute to pneumococcal evolution. Random mutations through point mutations and recombination are the major contributors to evolution, which are then subject to mechanisms such as natural selection and genetic drift [18, 94]. The larger size of DNA involved in recombination means that recombination introduces more diversity than point mutations [95]. A classic example of recombination due to selective pressure is in the event of resistance to antibiotics as these loci have been identified as recombination hotspots [17, 96].

Furthermore, WGS has been utilised in comparative genomic studies to identify virulence determinants that contribute to the different invasive propensities observed between similar serotypes found in different geographical locations [29]. WGS has also been used extensively to study the mechanisms of antibiotic resistance through horizontal gene transfer from other successful bacteria in the same niche [97].

The use of WGS in reverse vaccinology has also been explored. Wizemann *et al*. [80] used surface exposed protein motifs to screen for pneumococcal surface proteins. The identified proteins were cloned and recombinant protein used to immunise mice before being challenged with lethal doses of *S. pneumoniae* to look for protective candidates. The limitation of this study however was the low number of isolates with every serotype represented by only one isolate. This makes it impossible to evaluate the level of conservation of these proteins or even the diversity within the same serotype or between serotypes.

Furthermore, another study used a similar approach to scan for proteins present in only *S. pneumoniae* and absent in other Streptococcus species [98]. Although they chose 13 proteins that were present in all their isolates (51) covering 29 serotypes, these were also found in other Streptococcus species. Similar to the earlier study, their sample size of 51 was too small for conservation studies and they may have missed a lot of potential candidates in their study [98].

1.12 Thesis Aims and Objectives

With the above in mind, I have designed my MPhil to make a holistic evaluation of a particular family of pneumococcal surface proteins, 'the lipoproteins', as potential vaccine candidates.

The interest in lipoproteins is due to the many important roles they play such as in cell signalling, substrate binding and transport, antibiotic resistance as well as protein export and folding [99]. Some lipoproteins have been shown to play a role in both bacterial colonisation and pathogenesis. Pneumococcal lipoprotein PsaA has been shown to have an indirect role in colonisation and it is also essential for full virulence [100]. Furthermore, several other pneumococcal lipoproteins especially metal transporters have also been implicated in the virulence of this pathogen. These include the zinc transporters AdcAI and AdcAII and the ABC iron transporters PiuA and PiaA [23, 101]. Lipoproteins' role in virulence have also been demonstrated in other Gram-positive bacteria such as the LpK of *Mycobacterium leprae,* which induces human interleukin 12 during infection [102]. Also, their role in conjugation, protein secretion and localisation, sensing of their environment as well as their involvement in the spore cycle of *Bacillus subtilis* have also been speculated [103].

In this MPhil thesis, I will utilise the whole genome sequences of *S. pneumoniae* isolated from the Gambia through the Bill and Melinda Gates Foundation-funded Global Pneumococcal Sequencing (GPS) Project available to me at the Sanger Institute to evaluate lipoprotein genes. These samples include both invasive and

carriage samples therefore, this will give me a broader perspective of the protein distribution in both sets of data. The aims are to:

1. Identify all the surface exposed lipoprotein genes in these genomes.
2. Determine their level of conservation (prevalence) and diversity
3. Determine if there is association between alleles of these genes and pathogenic potential (i.e. ratio of prevalence in disease vs. carriage)
4. Use bioinformatics techniques to predict surface exposure and antigenicity of these proteins, which should aid in identifying those that have a greater potential to be successful vaccine candidates.