Pneumococcal interactions with mucin

by

Vanessa Sofia Agostinho Terra BSc

Thesis Submitted for the Degree of

Doctor of Philosophy

Department of Infection, Immunity and Inflammation,

University of Leicester

January 2011

Title: Pneumococcal interactions with mucin

Author: Vanessa Sofia Agostinho Terra

The nasopharynx is covered by mucin. Mucin is a glycoprotein and the carbohydrate moieties are potential fermentable substrates for *Streptococcus pneumoniae*. Previous *in vitro* studies showed that *S. pneumoniae* can grow on mucin. This study was undertaken to investigate how *S. pneumoniae* degrades the mucin carbohydrates to mono and di-saccharides for subsequent fermentation. *In silico* search of pneumococcal genome identified fourteen putative glycosidases. Pneumococcal mutants of each were made and tested for growth in defined medium with mucin as the only source of carbon. Of these, two genes *SPD0065* and *SPD0247* were chosen for further study. Consequently two novel glycosidases were described, β-galactosidase (BgaC), encoded by gene *SPD0065* and a 6-phospho-β-glucosidase (BglA) encoded by gene *SPD0247*. The knocked-out mutants SPD0065M and SPD0247M could not grow in Sicard's defined medium supplemented with mucin and exhibited decreased enzymatic activity when compared to the wild type D39. Since gene *SPD0562* had been previously identified as encoding a β-galactosidase (BgaA), the relative contribution of BgaA and BgaC to total β-galactosidase activity was investigated by introducing mutations in these genes individually and together. Mutation in the individual genes resulted in significant decrease in the enzymatic activity but the double mutation did not totally abolished activity. BgaC had specificity for galactose (β1,3)-N-acetylgalactosamine. Furthermore, BgaC released galactose from desialylated fetuin. The expression of *SPD0065* and *SPD0247* in *S. pneumoniae* grown in mucin containing medium or harvested from tissues from infected animals was significantly up-regulated compared to growth in glucose containing medium. When the mutants were tested *in vivo*, it was noted that SPD0065M had attenuated growth in the nasopharynx and SPD0247M in the lungs. In this study was demonstrated that BgaC has a role in the sequential deglycosylation of host glycoproteins and that BglA is involved in the further degradation of mucin-derived sugars.

This journey started in the 7th of January 2007 and it seems to be ending right about now. Throughout this journey I met and was in close contact with many people that turned out to be very important to me. So, it is with great pleasure that I am now seated at my desk writing what it seems to be not enough words to thank them.

I would like to start by thanking my supervisor, Professor Peter Andrew. And to him I have to say thank you for a couple of things. Firstly, thank you for accepting me in your laboratory, secondly thank you for being available to listen to my doubts and problems and thirdly thanks a lot for reading this thesis. It was a hard journey but if I had to do it again I would still do it in this lab! Also, thanks for introducing me to Leicester!

I would also like to thank Dr Hasan Yesilkaya. He was a constant presence throughout my PhD. Hasan, was a great co-supervisor and had always a nice word to say as well as many incentive words. I would also like to thank him for telling jokes on the long nights of work. Thanks a lot Hasan, for everything.

I would like to thank Dr. Karen Homer and Ms Susmitha Rao for helping with all the mass spec data. Thanks to both. I would like to thank Dr Marc PrudHomme from the Centre National de la Recherche Scientific (CNRS Toulouse, France) for kindly donate plasmids, pR410, pR412 and pCEP. I would also like to thank Dr Jack Thompson from the National Institute of Health (Bethesda, MD) for providing me with fluorogenic and chromogenic substrates that enable many of the enzymatic reactions. I would also like to thank Dr Brian Spratt for providing us with the *S.pneumoniae* strains from serotypes 6B and 19F. A big thank you goes to the University of Leicester for allowing me to pursue my studies by paying my tuition fees.

I would like to dedicate this thesis to both my mum and dad. They were always this driving force pushing me forward. I would like to thank them for always believing in me and for being there for me every single day of my life. For all that, thank you so very much. Obrigado Mae e Pai. Adoro-vos muito!

I would also like to thank my sister for being this amazing person that took care of me fulfilling her big sister duties. She started when I was born and she never stopped taking care of me. She has a huge role in my life and if today I am here finishing this thesis she is someone I really have to thank to. Also, I have to thank my sister and her husband, Sergio for giving me my amazing nephew. I love you Titi!

I would like to thank you, Yannis, for everything. For being who you are, for being there for me, for never letting me give up and for making me smile every day. I would like you to know that since the day you got into my life I am a much happier person! I love you very much!

I would like to thank Sarah Smeaton (Fred) and Sam for all the support they gave me during this 3 years. They were great friends, to whom I have much to thank to. I would also like to thank Rana, Sarah, Mafalda and Tania for all the kind words, technical insight and friendship. Thanks guys!

I would like to thank one very special person that was in my life when I was only 6 years old and that convinced me that I could do anything I wanted as long as I worked hard and put my mind to it. This thank you is for my primary school teacher Maria Amelia Vasconcelos which I am still in contact with and that has a very special place in my heart.

Finally, there is one final thank you for someone that is no longer with us. Thank you granny for being the best person I have ever met!

Table of contents:

Abbreviations

aa - amino acid

BAB – blood agar base

BHI – Brain heart infusion

BSA – Bovine Serum Albumine

c.f.u. – colony forming units

CSP1 - competence stimulating peptide 1

DNA – deoxyribonucleic acid

dNTP - deoxynucleotide triphosphate

EDTA- Ethylenediaminetetraacetic acid

Fuc – Fucose

Gal- galactose

GalNac- N-acetylgalactosamine

GlcNac- N-acetylglucosamine

Glu- glucose

IPTG- Isopropyl β-D-1-thiogalactopyranoside

Kb - kilo base

kDa – kilo Dalton

LA – Luria-Bertani agar

LB- Luria-Bertani broth

NCM – nitrocellulose membrane

nt- nucleotide

O.D. – optical density

PBS – phosphate buffered saline

PCR – polymerase chain reaction

pNP- 4-nitrophenyl

RNA – ribonucleic acid

SDS – sodium dodecyl sulphate

TE buffer – Tris-EDTA buffer

UV – ultraviolet

A Introduction

I Overview

In this chapter, all the noteworthy information related directly or indirectly with the work described in this thesis is introduced. This introductory chapter consists of four main sections. Section one is a summary of all the general characteristics of the studied bacteria *Streptococcus pneumoniae*, section two states all the main characteristics and functions of the mucin glycoproteins. Section three is an overview of *S. pneumoniae* most prominent virulence factors and section four explains in some detail its metabolism and how it correlates to virulence.

A.1 *Streptococcus pneumoniae* General Characteristics

I.1 General characteristics

Streptococcus pneumoniae is a gram positive bacterium, α-haemolytic when grown in aerobic conditions and β-haemolytic when grown anaerobically due to the activity of the pneumolysin O, only encountered on the *Streptococcus* genus. *S. pneumoniae* cells are usually oval or spherical, generally presenting a size that ranges from 0.5 to 1.25 μm appearing typically in pairs (diplococcus) occasionally singly or in short terms (Buchanan *et al.*, 1975).

This bacterium is heavily encapsulated with polysaccharide. The major polymeric components of the cell wall are peptidoglycan and the choline-ribitol teichoic acid complex (Mosser & Tomasz, 1970) which constitutes the species specific substance.

Streptococcus pneumoniae depends on choline when grown in defined media. It is a fermentative organism, with a growth optimum pH of 7.8, comprised in the range of 6.5-8.3. It ferments glucose, galactose, fructose, lactose, maltose, raffinose, glycogen and inulin. It is a facultative anaerobe with marked tendency to accumulate hydrogen peroxide when cultivated aerobically. The temperature range of growth is 25-42 °C. *Streptococcus pneumoniae* is sensitive to ethylhydrocupreine (optochin), an antibiotic used to differentiate the bacteria from other streptococci.

I.2 Impact of *Streptococcus* on human health

Streptococcus pneumoniae, widely known as pneumococcus, is the leading cause of morbidity and mortality amongst children worldwide, with emphasis in the developing countries (Scott, 2008, Mulholland, 2007).

A total of 10.6 million children under the age of 5 years old are thought to present pneumococcal disease every year (Black *et al.*, 2003). The most common form of disease is seen in very young children (under 2 years old) and in senior individuals (over 65 years old) in the form bacteraemic pneumonia. This is followed by pneumococcal meningitis, sepsis and otitis media in young children and infants. In the developing countries, this bacterium is a leading cause of childhood mortality due to acute respiratory infection (ARI). Additionally, it is a leading cause of non-epidemic childhood meningitis (Mulholland, 1999, Peltola, 2001).

Although the disease can reach all age groups, the more affected are young children and senior individuals, along with people suffering from a range of chronic conditions and immune deficiencies.

Streptococcus pneumoniae natural reservoir is the human nasopharynx; from there it can be transmitted through drops to other individuals. Almost every child is colonised with one or more strains of *S. pneumoniae* and becomes a carrier during the first years of life. In the majority of cases carriage is asymptomatic, the disease occurs only in a small number of people.

In some developing countries, such as Southern India, 50% of the infants are colonised by the age of 2 months with 80% of them being carriers by the age of 6

months (Coles *et al.*, 2001). A study from Africa demonstrated that the prevalence of carriage was 30%, 44%, 51% and 61% in children aged 6 weeks, 10 weeks, 14 weeks and 9 months respectively (Mbelle *et al.*, 1999). In industrialised countries, carriage is observed at an average of 6 months of age. Pneumococcal disease is thought to cause 1.2 million infant deaths per year (Berkley *et al.*, 2005, Denny & Loda, 1986). This indicates a significant public health problem.

I.3 Pneumococcal Serotype Epidemiology

I.3.1 Serotype distribution of carriage isolates

One of the most remarkable characteristics of pneumococcal epidemiology is the consistency observed in the carriage serotypes, in very different environments, at completely different times. For instance, serotypes 6, 14, 19 and 23 are almost invariably present in isolates from healthy young children. This fact has been true for: children in London in the 1950's (MASTERS *et al.*, 1958), at an American day centre in the 1970's (Loda *et al.*, 1975), for aboriginal children living in Australia and for infants from Alabama (Gray *et al.*, 1980, Hansman *et al.*, 1985). Surprisingly, this is still true for children from Papa New Guinea (Hansman, 1972, Montgomery *et al.*, 1990), Gambia (Lloyd-Evans *et al.*, 1996), Malawi (Feikin *et al.*, 2003), Central Asia, Vietnam (Parry *et al.*, 2000), the Netherlands (Bogaert *et al.*, 2001), Finland in the 1990's (Syrjanen *et al.*, 2001) and finally for children from Hertfordshire UK and Mozambique in the last 5 years (Valles *et al.*, 2006, Hussain *et al.*, 2005). These serotypes also account for the majority of all the isolates cultured from samples from the nasopharynx.

Other serotypes also found amongst children are 3, 4, 9, 11, 13, 15, 18 and 33. Since the prevalence of colonisation amongst adults is less than half than the one found in

children, it is harder to precisely determine colonisation patterns (Hansman *et al.*, 1985, Adegbola *et al.*, 2006, Foy *et al.*, 1975, Regev-Yochay *et al.*, 2004, Rountree *et al.*, 1967). Nevertheless, it appears that the serotypes found amongst adults, differ significantly from the ones found in children, while having a broader serotype distribution.

In order to take the carriage epidemiology results into consideration, it is important to contemplate some constraints that hamper the data. First, the sensitivity methods used vary from serotype to serotype. Secondly, it has been estimated that around 30% of all pneumococcal carriers harbour more than one serotype simultaneously (Gratten *et al.*, 1989), making the numerically dominant serotype difficult to identify. Finally, the size and duration of the carriage studies never do justice to the complexity of a system with over 90 serotypes in a highly dynamic environment (Hausdorff *et al.*, 2008). Due to these technical and methodological considerations, it is clear that it is not possible to be quantitatively accurate in epidemiology studies. Despite that, it is possible to infer that some serotypes are very rarely present in carriage studies. For instance, it has been noted as early as the beginning of the 20th century that serotypes 1 and 2 cause invasive pneumococcal disease and are almost never isolated from nasopharynges of healthy individuals (Dochez & Avery, 1915).

These observations suggest that determined serotypes have a high invasive disease attack rate, which in turn poses another very intriguing question: how are these serotypes maintained in the community if they are rarely carried? The answers to this question could be that they would be transmitted directly from one patient with invasive pneumoniae to another (Smillie & Jewett, 1940). The association between highly invasive serotypes and the onset of invasive pneumonia gives more credibility to the

idea of person to person spread. Possibly, a low number of carriers of these highly invasive serotypes, the majority of them amongst ill individuals, would be enough to maintain these serotypes alive in the human community (Hausdorff *et al.*, 2008).

I.3.2 Invasive serotypes distribution

The invasive potential is a measure of the ability of *S. pneumoniae* to progress from nasopharyngeal carriage to invasive disease; this is not the same as virulence since this term is usually applied to the ability of a pathogen to cause disease in laboratory animals.

There are marked differences between serotypes and their ability to cause invasive pneumococcal disease Table 1 presents some serotypes and their relation to carriage and invasive pneumococcal disease (Hausdorff *et al.*, 2008).

Table 1 - Relationship between serotypes and ability to cause pneumococcal invasive disease. The light pink shade indicates serotypes present in carriage; the light grey shade indicates serotypes known for causing invasive pneumococcal disease.

Serotype	Carriage	Invasive disease
1		✓
3	✓	
4		✓
5		✓
6A	✓	
6B	✓	
7		✓
7F		✓
9V		✓
14		✓
15	✓	
18C		✓
19	✓	
19A		✓
19F	✓	
23	✓	

What is important to infer from this data, is that there seems to be a reverse correlation between carriage strains and invasive strains. The most invasive serotypes are less prevalent in the carriage data, and the most carried serotypes are the less prevalent as cause of pneumococcal invasive disease (Brueggemann *et al.*, 2004).

It was suggested in the early 80's, by Gray and co-workers that there was a natural adaptation in children. The common serotypes causing disease in the pre-antibiotic era (serotypes 1 and 2), were highly invasive when compared with the common colonisers

(serotypes 6, 19 and 23). He proposed that the tolerance shown by young children to prolonged carriage would somehow confer them some protection against colonisation of the most invasive serotypes (Gray *et al.*, 1980). According to this study, the cost of the natural adaptation was disease, but at a much lower rate and caused by the colonisers themselves. This idea supports the theory of competition amongst serotypes for nasopharynx colonisation, which has been observed in animal models (Lipsitch *et al.*, 2000). Some serotypes such as 15B, 15C, 22F and 33F (Gonzalez *et al.*, 2006), that seem to have an increased frequency as cause of invasive pneumococcal disease (IPD), following the introduction of Pneumococcal Conjugated Vaccine 7 (PCV-7), have presented lower attack rates than the vaccine serotype (Kaplan *et al.*, 2004). Serotypes that have attack rates that are, the same or very similar to vaccine serotypes (Byington *et al.*, 2006, Sleeman *et al.*, 2006, Steenhoff *et al.*, 2006, Beall *et al.*, 2006) have also been reported. It does seem that the serotypes replacement that occurred after the wide use of PCV-7, should not end in invasive disease replacement in the general population, unless the replacing serotypes are as much or more invasive than the original serotypes. To finalise, it has to be mentioned that even the serotypes commonly known as pacific colonisers, with low level of invasiveness, can and will cause invasive disease if such serotypes become highly prevalent in the carriage population (Hausdorff *et al.*, 2008).

I.4 Therapy and Resistance to Antibiotics

Pneumococcal diseases have been treated with penicillin since its introduction in the 1940s. Although initially efficient, this therapy has been threatened by the development of resistance. Resistance has been observed for all major groups of antibiotics apart from vancomycin. In this antibiotic era, both morbidity and mortality have remained

high (Henriques-Normark, 2007) fact which proves the significant public health threat that this organism represents.

I.4.1 High transformability equals higher resistance capability

A major part of the antibiotic resistance has evolved from horizontal transfer of genes or even gene fragments from commensal pneumococci, which was then followed by a spread within the pneumococcal community and final dissemination of the successful clones into the society. Therefore, variation in ability to colonise, propensity to accept DNA through transformation and mutation frequency may all be factors contributing to the development of resistance in particular serotypes (Henriques-Normark, 2007).

I.4.2 Developments in Antibiotic Resistance

Recent data acquired from Global Surveillance indicate that high level of isolates with resistance to penicillin, presenting a MIC ≥ 2mg/L, are spread all over the world (Henriques-Normark, 2007). Rates exceed 20% in USA, Mexico, Japan, Saudi Arabia, Israel, Spain, France, Greece, Hungary and the Slovak Republic. In countries like Hong Kong, Taiwan and South Korea, the frequencies of resistance exceed 50%. In some of the Asian countries (Felmingham *et al.*, 2004) the incidence of penicillin non-susceptible pneumococcal strains with intermediate susceptibility (MIC 0.12-1 mg/ml) are found in a prevalence of 70-80%.

In the beginning of the 21st century, a European study in pneumococcal resistance was performed in eight countries: Belgium, Austria, France, Germany, Italy, Portugal, Spain and Switzerland (Reinert *et al.*, 2005). In this study, a total of over 2200 isolates were analysed and many resistances to a variety of antibiotics were found. Figure 1 illustrates the resistances found in the above study.

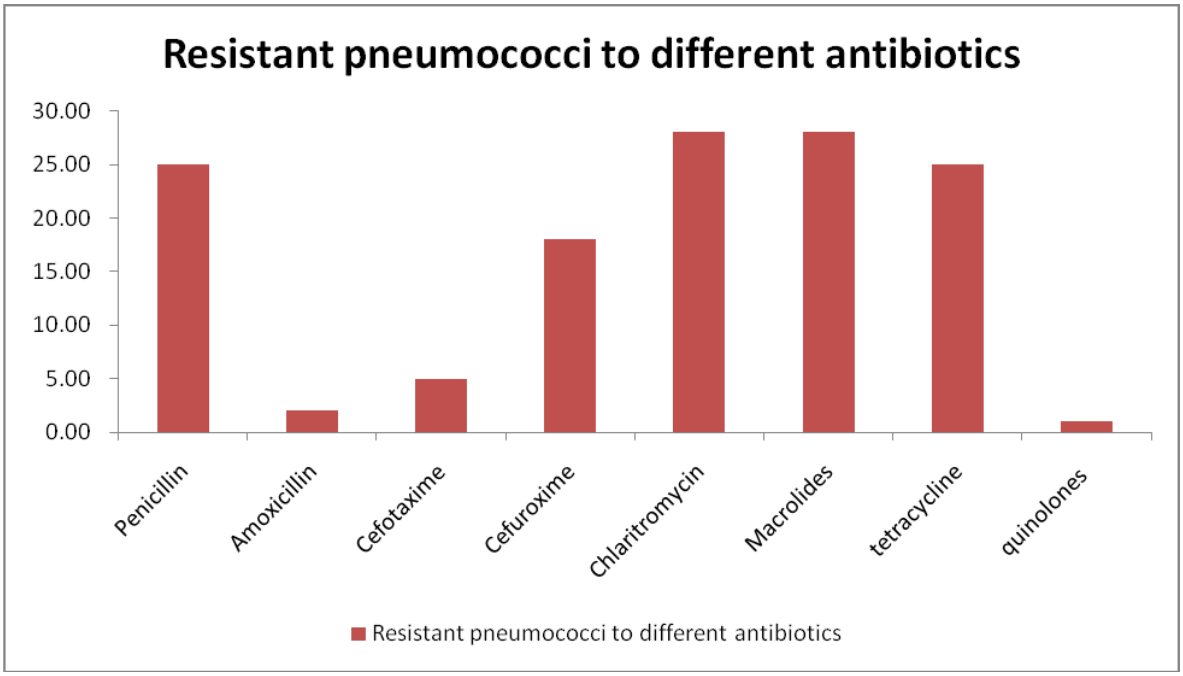


Figure 1- Percentage of pneumococci found resistant to different antibiotics out of the 2200 studies in eight different countries (Reinert *et al.*, 2005).

The resistance rate varied hugely between countries, however the highest rate of penicillin and macrolides resistance was found to be in Spain and France. Serotype 14 was the dominating serotype amongst penicillin G and macrolide resistance in this European study.

It is known that the macrolide resistance has a worldwide spread and it has even surpassed penicillin G resistance in many areas (Felmingham *et al.*, 2002). This coincides with increased use of macrolides, due to an increase of other pathogens causing atypical pneumonia. The highest rates of macrolide resistance are seen in Asian countries such as Korea and Japan, where the resistance exceeds 80%. In the European countries, the resistance to macrolides is higher in countries such as France and Italy (Felmingham *et al.*, 2002; Perichon *et al.*, 1997).

Unfortunately, during the course of the years, other resistances appeared and antibiotics that once were effective have become less effective, with examples like

timethoprim/sulfomeoxazole, which has a resistance rate of about 30% in the world and in countries such as Mexico, South Korea and Spain presents an even higher resistance rate of about 50%. The resistance to tetracycline is also increasing, virtually almost all isolates in Taiwan are resistant to it, and the rates in the rest of the world are also very high (>25%) in Japan, South Africa, Western Russia, Saudi Arabia, Belgium, Spain, France, Italy, Greece, Hungary, the Slovak Republic and finally Poland (Felmingham, 2004). Some American reports presented alarming results as well; resistance to tetracycline is about 20% and to Timethoprim/sulfomeoxazole is 35% (Jacobs, 2004). The reports from US also state that multi-drug resistance is becoming increasingly common. Drugs successfully used in the United States are amoxicillin, amoxicillin-clavulanate, fluoroquinolones and clindamycin. So far, vancomycin is the only agent that has no resistant clones detected.

Another increasing resistance is to fluoroquinolones which at the moment is still considered low (Perichon *et al.* 1997). However, an increase has been noticed in the last few years. Despite that fact, resistance to fluoroquinolones in Hong Kong (Ho *et al.* 1999, Ho *et al.* 2001) Ireland and some areas of Spain has reached levels of 18%, 15% and 5% respectively (Perichon *et al.*, 1997, Ho *et al.*, 1999, Ho *et al.*, 2001, Daly *et al.*, 2004). It has been shown by Pallares and co-workers that resistance to fluoroquinolones can develop during treatment (Pallares *et al.*, 2003). Although antibiotic resistance is a major problem, an even bigger problem is the multidrug-resistance, and it is now observed in more than 50% of pneumococcal isolates from Hong Kong, Taiwan and South Korea (Felmingham, 2004). High frequencies are also present in South Africa (32.5%), Mexico (13.9%), Saudi Arabia (13.3%) as well as United States and Russia (10.6%). The most striking case is observed in Sweden where, despite the fact that it does not present high frequency of resistant pneumococci it does present a high

frequency of multi resistance, presenting 40-50% in PNSP clones, suggesting clonal spread of international clones.

In conclusion, increase incidence of antibiotic resistance has been registered worldwide. The increase resistance to macrolides and the appearance of multi-resistance may hinder the treatment of pneumococcal infections. These results clearly demonstrate how dangerous this pathogen can be, and highlight its ability to evolve and adapt.

I.5 Vaccines

I.5.1 Pneumococcal 7-valent conjugated vaccine

At the present time, there are two available vaccines, the 23-valent polysaccharide vaccine (PS) and a 7-valent conjugate vaccine (PCV-7). Despite the fact that the PS vaccine is useful for immune competent individuals over 5 years of age, it exhibits several disadvantages and has very limited efficacy when administered to young children under the age of 2 and immunodeficient individuals (Bogaert *et al.*, 2007).

It was noted back in 1929 by Avery and co-workers, that the immunogenicity of the polysaccharides was highly increased, by the covalent binding of capsular polysaccharides to proteins (Poland, 1999). This action comprises linkage of capsular polysaccharides to a protein carrier either by covalent binding or through reactive groups (Poland, 1999; Klein, 2000). The difference in using this type of vaccine, when compared to the pure PS vaccination method, is the switch observed in the immune response to a TD response promoted by proteins. This leads to the induction of memory B-cells and a newer and improved B cell response. Additionally, it leads to an improved immune response in early stages of development. This different type of response and its efficacy is directly related to the step by step maturation of the immune system which

will mature a TD response earlier than an anti(poly)saccharide response (TI-2 response) (Bogaert *et al.*, 2007). Unfortunately, there is a limitation; too much carrier protein may impair the antibody response by antigen competition or carrier-mediated epitope suppression (Di John *et al.*, 1989). Therefore, the protection is limited to a determined number of serotypes. The actual vaccine, Prevnar (Wyeth, USA), and Prevenar (Wyeth, Europe) includes polysaccharides types 4, 6B, 9V, 14, 18C, 19F and 23F. This vaccine has a potential coverage of 85% of the isolates for the USA, 60-70% in Europe and 55% in Asia (Pelton *et al.*, 2003). Despite being effective, it is already being threatened by serotype replacement. The effect of replacement in the invasiveness is not yet understood, although some vaccination studies, particularly one in the USA, reported 25% increase of non vaccine serotypes causing pneumococcal invasive disease (Bogaert *et al.*, 2007).

I.5.2 23-valent polysaccharide vaccine (PS)

The majority of the bacteria that cause invasive disease, especially the ones causing bacteraemia, are protected from innate immune host immunity because they have on their cell surfaces polysaccharides (PS). *S. pneumoniae* is the prime example of this fact, with all its strains corresponding to something as much as 90 different serotypes, each with one chemically and antigenically different capsular polysaccharide (PS) (Lees *et al.*, 2008) .

Bacterial polysaccharides are composed of thousands of carbohydrate repeat units resulting in polydisperse polymers that can have molecular masses up to the millions of Daltons. Over 70 years ago, purified PS with high molecular mass were found to induce protective opsonic antibodies when administered to human adults, leading to the introduction of the 14-valent PS vaccine in 1977, which was substituted later on by the 23-valent PS vaccine in 1983 (Lees *et al.*, 2008).

The most significant limitation of this vaccination system is the fact that does not confer protection to children under 2 years old. The currently licensed 23-valent vaccine accounts for serotypes 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F and 33F. The structures of the 23 serotypes of PS used in Pneumovax have been summarised by Jennings, 1990, among others. This vaccine is indicated for use in children above 2 years old and adults.

I.6 Host defence against pneumococcal disease

Capsular polysaccharides are highly immunogenic, and the antibodies against these polysaccharides protect only against the homologous serotype. The capsule antigenicity is type-specific but cross-reaction is possible due to shared polysaccharides (Bruyn *et al.*, 1992). The following layer is the cell wall, composed by polysaccharides, teichoic acid and several cell wall associated proteins. The intense inflammatory response observed is due to the cell wall, since it stimulates the flux of inflammatory cells. Additionally, it also activates the production of cytokines (Bruyn *et al.*, 1992) and activates the alternative complement pathway. In the cell wall, the most immunogenic part is the phosphocholine moiety of the teichoic acid, which plays a very important role in the inflammatory process (Bogaert *et al.*, 2007).

The process of clearing pneumococci is accomplished by phagocytosis and intracellular killing of bacteria by neutrophils and alveolar macrophages. This process occurs only in the presence of type specific antibodies (IgG1, IgG2, IgM and IgA) and active complement. The antibody initiated complement-dependent opsonisation, which activates the classical complement pathway, is believed to be the major process to clear pneumococci from the host. When the correct type of immunoglobulins is not present, pneumococci enter the bloodstream causing bacteraemia (Paton *et al.*, 1993).

The clearance mechanism from blood seems to depend on the interaction of type specific antibodies, complement and phagocytic cells in the liver and spleen. Therefore, absence of the spleen or liver impairment leads to severe predisposition to pneumococcal disease. Different types of congenital deficiencies of immunoglobulins or complement are also associated with predisposition to pneumococcal infection (Bruyn *et al.*, 1992). It is thought that antibodies directed not to the capsule but to the cell wall may play a function in host response against pneumococcal infection. Although the cell wall is completely covered by the capsule, some proteins can penetrate the capsule and may therefore be identified by the immune system (Bogaert *et al.*, 2007). Thus far, proteins such as PspA, PsaA and pneumolysin have been shown to elicit protection in animal models (Paton *et al.*, 1993).

The immune response is T-cell dependent (TD), which is present since birth and stimulates production of immunoglobulins IgG1. Type-specific immunoglobulins will be more effective if belonging to subclass IgG2. The adult response against pneumococci mainly generates this type of antibody, however in children this is not the case, and the antibodies generated are from subclass IgG1 (Freijd *et al.*, 1984). This may result in increase susceptibility to pneumococcal infection (Rijkers *et al.*, 1998). Thus the type of vaccination induces the type of immune response and the corresponding level of protection.

A.2 Mucin

II Overview

The main function of airway mucus is to defend the respiratory tract against pathogenic and environmental challenges; at least that was the mainstream knowledge until now. While this is true, it was also demonstrated by us that the mucin, the major oligomeric component of mucus, and specifically the carbohydrate content of its side chains, provides nutrients for growth and persistence of pathogens (Yesilkaya *et al.*, 2008). In this section, the general characteristics of the mucin structure and its function on mucosal immunity are discussed.

II.1 General characteristics of the mucus

II.1.1 The Mucus barrier

Mucus is a dynamic semi-permeable barrier (Cone, 2009) composed mainly by mucins and inorganic salts suspended in water. The mucus layer coats the non-keratinised epithelial surfaces of humans and other animals. Although, many pathogens enter the body by specific binding to the mucous membranes followed by penetration, some non-specific mechanisms tend to prevent this entry, hence playing a major role in the innate defence against pathogens (Thornton *et al.*, 2008). The composition of this dynamic layer changes depending of the environment where it is being secreted. In the airway, mucus performs three key roles: 1) entraps the foreign particles and pathogens, 2) dissolve harmful gases, 3) expels both, by synchronous movement of cilia (Knowles & Boucher, 2002), promoting the expelling of, for example pathogens entrapped in the mucus.

When a mucus gel suffers a pressure, like cough, the mucin fibres stretch rather than pull apart from each other, this property is called elasticity, and it is vital for the mucociliary transport (Cone, 2009). The mucus discard from the eyes, sinuses, middle ears, nose and lungs are continuously transported by the cilia in the epithelial surfaces into the pharynx. The particles and bacteria entrapped in this mucus are transported to the stomach, where they are immediately inactivated by acid. The mucus present in the tears is cleaned and restored in this way every 10 seconds, the nasal mucus every 10 minutes and the respiratory mucus is transported out of the lungs at a rate of 100 μm/s, managing to clear the lungs within minutes to hours (Schuhl, 1995).

Apart from being a movable physical barrier, this multifaceted gel-like secretion also contains various host defences that exhibit antimicrobial effects, such as: cell wall

degrading lysozyme, the iron-chelating lactoferrin, specific membrane-permeabilising members of the defensin and cathelicidin and pentraxin families. In a healthy and normal lung, these proteins ensure an easy and smooth elimination of possible threats (Rogan *et al.*, 2006). The major component of the mucus is what can be called the giant biopolymers known as mucins (Thornton *et al.*, 2008). Mucin will be described in detail in the following sections.

II.1.2 Mucins

Mucins are defined as high molecular weight glycoproteins (2-20 x 10^5Da) that contain one and sometimes multiple protein domains (TR repeats) extensively glycosylated (O-glycan attachments). Therefore, the main components of these glycoproteins are carbohydrates, and its content can be as much as 80% of the total weight. Currently, two types of mucin are known. One that is believed to be monomeric and mainly but not only, located on the cell surface. And a second type, oligomeric, believed to be secreted and responsible for the gel like properties of the mucus by forming the disulphide bonds critical for the gel-forming properties through the cysteine-rich domains present at both N and C termini (Thornton *et al.*, 2008).

These mucin-like domains are known to have several repeats of serine and threonine, which are the sites for the attachment of the sugar N-acetylgalactosamine. Mucin glycoproteins may also contain apart from N-acetylgalactosamine sugars such as N-acetylglucosamine, galactose, fucose and sialic acids, and the sulphate molecules.

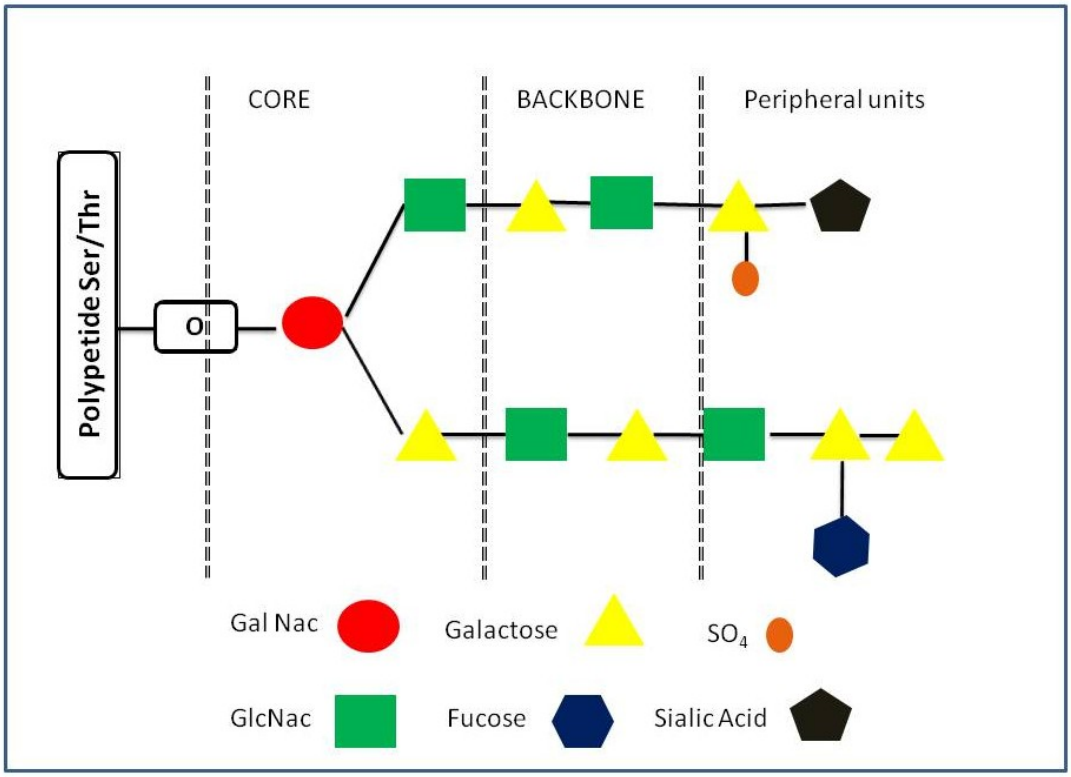


Figure 2- Mucin structure and composition representation; GalNac- N-acetylgalactosamine; GlcNac- N-acetylglucosamine; Figure adapted from Rose & Voynow, 2006.

Mucins have usually low isoelectric point, and this characteristic is attributed to their anionic nature, which is attributed to the sialic acid and sulphate groups present in the side chains of the mucin oligosaccharide (Rose, 1992). When the final addition is fucose, the result is an increase in hydrophobicity (Rose & Voynow, 2006). These two characteristics, charge and hydrophobicity, are thought to influence the physical and biological properties of mucin.

One of the most impressive properties of these oligosaccharides is their structural diversity, which suggests that different structures may be related to different roles. One probable role related to this diversity is, the possibility of enhancement of bacterial binding to the mucus layer facilitating their removal by mucociliary transport (Thornton & Sheehan, 2004). Additionally, mucins also provide competing receptors for cell surface glycoconjugates contributing to trap bacteria and to prevent them from

colonising the epithelium. Therefore, the diversity of mucins being expressed in an individual may play a role in its susceptibility to infection (Thornton & Sheehan, 2004). Accordingly, in a state of disease, alterations to terminal sugars can occur, potentially altering the biochemical and physical properties of the mucin (Rose & Voynow, 2006).

The mucin family is a combination of polymerising and non-polymerising forms of secreted and tethered mucins (Thornton *et al.*, 2008). The secreted mucin in the airways is synthesised by specific cells in the surface of the epithelium, called the goblet cells and in the submucosal glands in cell known as mucous cells (Hovenberg *et al.*, 1996, Wickstrom *et al.*, 1998). The macromolecules formed are kept in a dehydrated state in the secretory granules and are released by regulated secretion both constitutively and as a response to a change in the surrounding conditions (Thornton *et al.*, 2008). Hence, the epithelium is in position to rapidly respond to a variety of environmental signals without the requirement of *de novo* synthesis of the mucin, which usually takes about 1-2 hours (Sheehan *et al.*, 2004).

II.1.3 Mucin glycosylation and cores

One of the major parts of mucin biosynthesis is the O-glycosylation which requires the presence of particular enzymes, such as N-acetylgalactosaminyl peptidyltransferase and several glycosyltransferases depending on the final glycan structure (Thornton *et al.*, 2008, Rose & Voynow, 2006). The process of O-glycosylation is initiated in the Golgi apparatus, when N-acetylgalactosamine (GalNAc) is transferred by an N-acetylgalactosaminyl peptidyltransferase to a serine or a threonine as soon as nascent MUC polypeptide chains transverse the Golgi (Brockhausen, 2000). This is followed by the elongation of the chain by the addition of hexoses such as galactose (Gal), N-

acetylglucosamine (GlcNAc), fucose or sialic acid by the pertinent glycosyl transferase (Rose & Voynow, 2006).

To date, four different major core structures have been identified in O-glycans isolated from mucins (Hounsell, 1994). The transfer of a galactose to the carbon 3 (C3) of N-acetylgalactosamine produces a structure known as core 1 or 2. The following transferase will either elongate by adding a galactose (core 1), or a GlcNAc to form a branch (core 2). Following the same elongation if GlcNAc is added to GalNAc then it will form core 3 and finally, if it is added to form a branch will become core 4 (Figure 2).

The elongation of each individual core is done by the actions of Gal and GlcNAc transferases which will produce the Galβ1-3/4GlcNAc units that can be terminated by the addition of a sialic acid, a sulphate or a fucose.

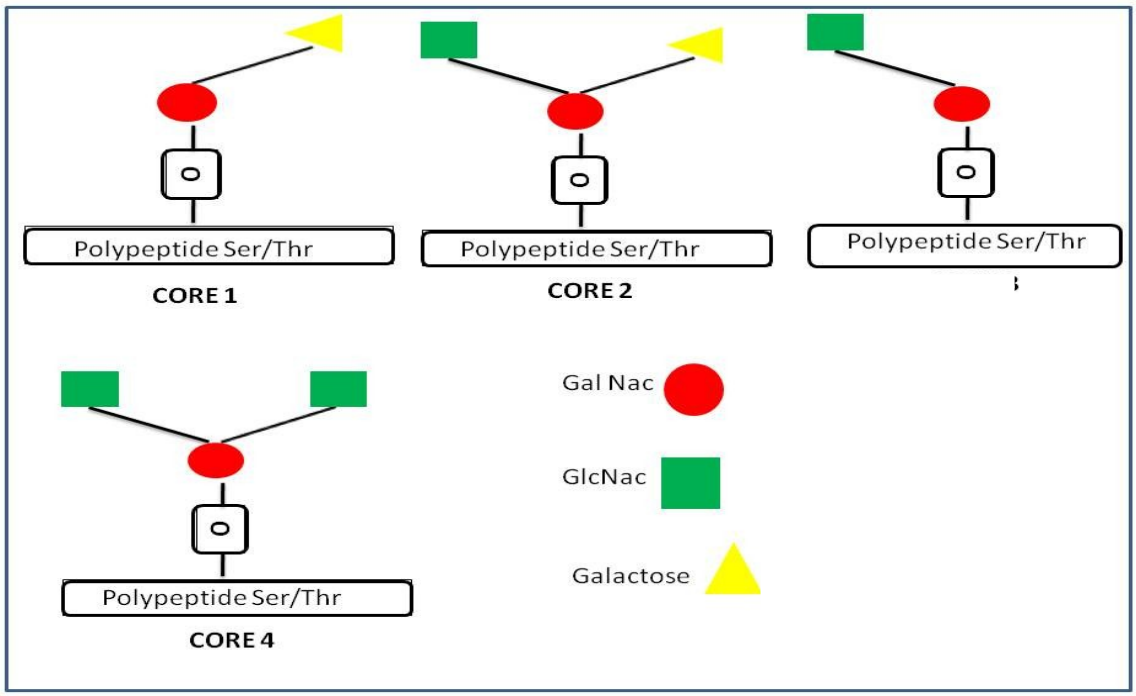


Figure 3- Mucin cores representation. Gal- galactose; GalNac- N-acetylgalactosamine; GlcNac- N-acetylglucosamine; Figure adapted from Rose & Voynow, 2006.

II.1.4 Mucin Composition and Classification

The TR (tandem repeats of serine and threonine) domains are the main feature distinguishing mucins from other glycoproteins (Rose & Voynow, 2006). Mucins are classified by their protein backbone MUC, which is, encoded by what is known as a *muc* gene. The products that result from the translation of a *muc* gene can differ tremendously in size from each other, a protein encoded by a *muc* gene can contain anything from 377 amino acid residues to 11000. At present, there is still some controversy to what are a mucin glycoprotein and a mucin-like glycoprotein. If the definition of the mucin gene includes only protein backbones containing TRs, then the list of *muc* genes consists of 18 entries. MUC1, MUC2, MUC3A, MUC3B, MUC4, MUC5AC, MUC5B, MUC6, MUC7, MUC8, MUC9, MUC11, MUC12, MUC13, MUC16, MUC17, MUC19, and MUC20 and their product will be a mucin glycoprotein. On the other hand, if the mucin gene definition includes proteins that contain numerous serine and threonine but not a single repeat then MUC14 (Kinoshita *et al.*, 2001), MUC15 (Pallesen *et al.*, 2002) and MUC18/CD146 (Lapensee *et al.*, 1997) would also have to be added to the list of *muc* genes.

Another of the mucin characteristics used for classification is the presence of modular motifs at the carboxy termini of the proteins backbone, this characteristic enables classification as membrane tethered, or secretory mucin and within the latter category, cysteine rich or cysteine poor (Rose & Voynow, 2006).

II.1.5 Importance of TRs for MUC protein.

The TRs are normally encoded by a single central exon on a *muc* gene, however, majority of *muc* genes are polymorphic with alleles having variable number of TRs

(Fowler *et al.*, 2001). Moreover, other variations can occur resulting in minimal changes on the length of the repeat unit and in the repeating sequence, hence different individuals may express different forms of the same muc gene product.

TRs have a unique sequence and size, which are the defining characteristics of each MUC protein. Amongst different mucins, TRs can differ in length, presenting anything from 5 to 375 amino acid residues or in the number of repeats which can be from 5 to 395 repeats. The two main airway mucins MUC5AC and MUC5B have TR domains repeated 4 and 5 times respectively and these are separated by cysteine rich domains.

Therefore, the mucin ability to be glycosylated is defined by the TRs. If the TRs are long or many then the more chances there will be for glycosylation. The number of TRs also determines mucin size as it is elongated, therefore increasing its molecular weight (Rose & Voynow, 2006).

II.2 Fetuin –model glycoprotein

Fetuin is a major protein present on the embryonic plasma (Dziegielewska *et al.*, 1987). This glycoprotein has a molecular mass of about 68 kDa and contains a single polypeptide chain with 590 amino acid residues. It belongs to a family designated as albumin like proteins, which contain other proteins such as L-albumin, vitamin D-binding proteins and HSA. These proteins are highly homologous and contain very similar molecular weights. Fetuin is produced by both the yolk sack and the foetal liver. Of this family at least two are heavily glycosylated. Fetuin is a model glycoprotein, frequently used in substitution of mucins since its structure is significantly less complex than the mucin structure and also because the constituent sugars are subject to little or no modification by, for example, sulphation which is usually extensive in mucins

(Green *et al.*, 1988) and known for limiting glycosidic activity. Additionally, fetuin contains galactose linked in both β(1,3) and β(1,4), which makes it the ideal substrate to use for unknown galactosidases specificities.

Fetuin contains six carbohydrate moieties per molecule, three of this present O-linkage to ser/thr, and the remaining three are linked to an asparagine through an N-linkage. The O-linked carbohydrates are thought to account for 20% of the total carbohydrates bound to fetuin. The remaining 80% are N-linked carbohydrates. The N-linked oligosaccharides were shown to be tribranched with the structure shown in Figure 4.

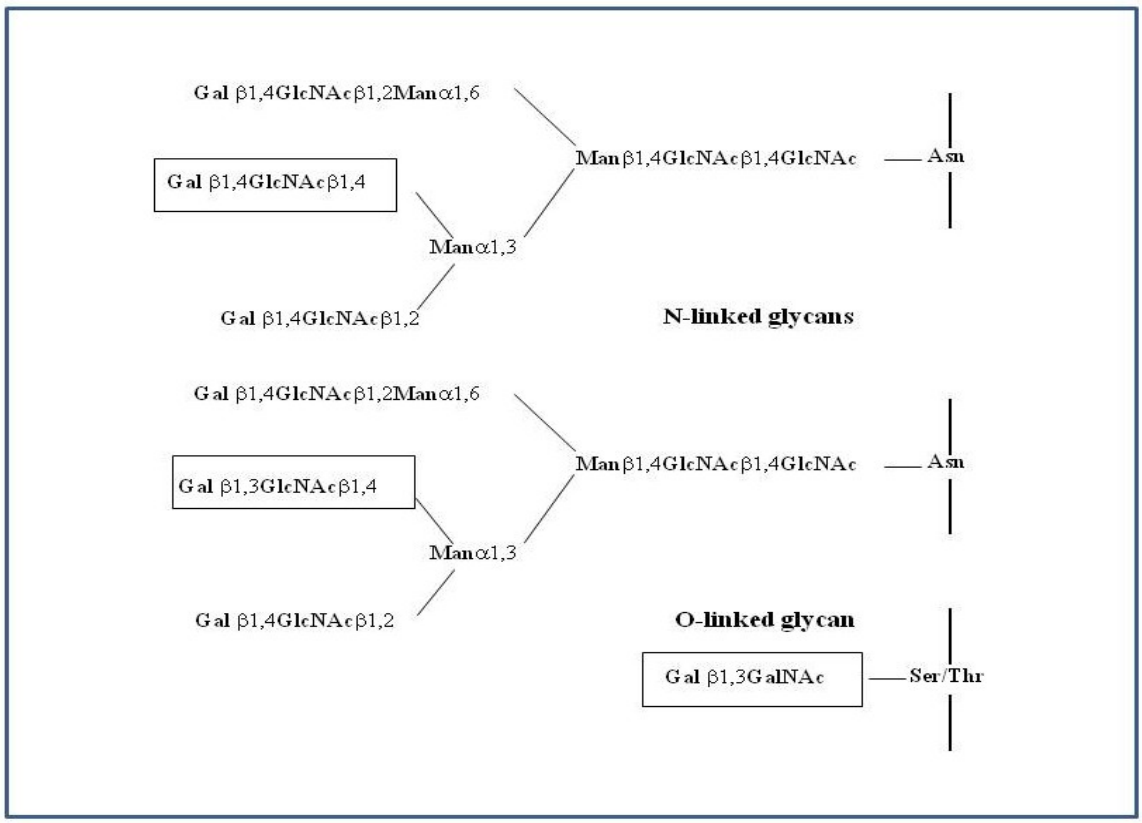


Figure 4- Glycan structures of desialylated fetuin. (Constructed from data presented in (Green *et al.*, 1988b, Takasaki & Kobata, 1986)

There are also reports that show that a minor structure of dibranched fetuin exists. In this case, one of the peripheral branches is not present, more specifically the branch originated from the linkage β(1,4) to mannose. In the tribranched structures, a minority

was found to have a particular peripheral structure containing sialic acid. Due to these characteristics, fetuin was the ideal glycoprotein to substitute mucin in some of the experiments performed throughout the work done for this thesis.

A.3 Virulence factors

III Overview

Many virulence factors have been described for *Streptococcus pneumoniae*, and although some may be more important than others, it is well established that its full virulence is due to a concerted action of many factors. This section of the Introduction chapter is divided into two sub-sections. The first section contains a description of classical virulence factors such as the capsule, pneumolysin, pneumococcal surface protein A (PspA), choline binding protein (CbpA) and pneumococcal surface adhesion A (PsaA).

The second section is describing a newer class of exoglycosidases believed to act and contribute mainly to the colonisation of the nasopharynx such as neuraminidases, with special attention to neuraminidase A (NanA), β-galactosidase A (BgaA), hexosaminidase (StrH) and endo-α-N-acetylgalactosamine (Eng). These exoglycosidases have been shown to be involved in utilisation of host glycoproteins, and as a result these are related directly to the enzymes described in this thesis.

III.1 The capsule

The capsule of *the pneumococcus* is a diverse group of polymers that play a very important role in its virulence. A total of ninety different polysaccharides have been described to date (Henrichsen, 1995), and the chemical structures for about 60% of them have been determined (van Dam *et al.*, 1990). The capsule is the outer layer of *S. pneumoniae* and its thickness has been reported as being in between 200-400 nm (Skov Sorensen *et al.*, 1988). It is covalently attached to the peptidoglycan of the cell wall in most serotypes with exception for the serotype 3 (Sorensen *et al.*, 1990).

The simplest capsular polysaccharides (CPS) types are linear polymers with repeating units composed of two or more polysaccharides whereas the most complicated types are composed of branched polysaccharides with repeat unit backbones composed of one to six monosaccharides with additional side chains (Hakenbeck & Chhatwal, 2007). The polysaccharide capsule is an essential element for virulence, given that all the isolated clinical strains are encapsulated, and spontaneous mutants without it are virtually avirulent (Kadioglu *et al.*, 2008).

The particulars of how the capsule is contributing to virulence are not yet fully understood. However, it is known to have strong antiphagocytic properties in non-immune hosts (Austrian, 1981). Despite some known differences between capsules, they all have the same role of reducing the opsonophagocytosis by limiting access of phagocytic receptors to complement bound to the *S. pneumoniae* cell wall. The thickness of the cell wall is directly related to the ability of causing disease in a particular strain and serotype (MacLEOD & KRAUS, 1950). Nevertheless, pneumococci from the different CPS serotypes differ strikingly in their capability of

causing disease (Austrian, 1981). Capsules are usually highly charged at physiological pH, and this fact may have a direct interference on its interactions with phagocytes (Lee *et al.*, 1991). Cell wall teichoic acid is capable of activating the alternative complement pathway. Additionally, the antibodies created against this and other surface proteins can activate the classical pathway, as does the interaction of the teichoic acid with the C reactive protein.

The capsule also forms an inert shield that seems to prevent both the Fc region of IgG or iC3b binding to deeper cell surface structures and then interacting with receptors on phagocytic cells (Winkelstein, 1981, Musher, 1992). The capsule has also been shown to prevent the mechanical removal of the pneumococci by the mucus action (Nelson *et al.*, 2007), to reduce autolysis and the exposure to antibiotics (van der Poll & Opal, 2009).

III.2 Pneumococcal surface protein A

PspA is located on the cell wall of pneumococci (Lopez *et al.*, 1992), and so far has been found in every single strain of pneumococci discovered (Crain *et al.*, 1990). PspA is located on the surface of the cell and has a molecular mass that can vary from 67 to 99 kDa (Waltman *et al.*, 1990). This protein has four domains, an N-terminal highly charged with a helical region, a proline rich domain, a domain with ten highly conserved 20 amino acid residues repeat and finally a tail of 17 hydrophobic amino acids (Yother & Briles, 1992).

Antibodies raised against PspA are protective against *S. pneumoniae* disease. Its high charge is thought to inhibit the binding of the complement proteins (Jedrzejas *et al.*, 2001) preventing the associated clearing by the complement. PspA has also been shown

to be involved in the reduction of phagocytosis of *S. pneumoniae* (Briles *et al.*, 1997). PspA is a lactoferrin binding protein, and through this activity is thought to protect the bacteria against the bactericidal apolactoferrin (Kadioglu *et al.*, 2008). All protective antibodies reacted with the N-terminal half of the protein showing that this part is most likely exposed on the surface (McDaniel *et al.*, 1992). PspA has been chosen to be included in a future conjugated pneumococcal vaccine because the antibody response against this particular protein seems to be serotype independent (Nabors *et al.*, 2000). This particularity makes this enzyme a prime candidate since vaccination that covers all ninety *S. pneumoniae* serotypes is very unlikely.

III.3 Hyaluronidase

Hyaluronidase belongs to a broad group of enzymes, hyaluronidases, which are responsible for degrading components of the extracellular matrix, hyaluronan, conferring to them a property called primarily as “spreading factor” (Duran-Reynals, 1933).

Hyaluronidase is another main surface protein of *S. pneumoniae*, with variable antigenic properties that may be essential for pneumococcal virulence (Jedrzejas, 2001). Hyaluronidase is produced by most strains of *S. pneumoniae* (Humphrey, 1948, Crain *et al.*, 1990, Hynes & Ferretti, 1989). It has a molecular mass of approximately 107 kDa but when expressed in *E.coli*, 89 KDa (Beveridge & Graham, 1991, Berry *et al.*, 1994, Jedrzejas *et al.*, 1998). The C- terminus of the mature enzyme contains a hydrophobic tail followed by the sequence LPXTG, and a group of charged residues that are responsible for covalently anchoring the enzyme to the peptidoglycan cross-bridges (Jedrzejas, 2004).

The enzyme degrades hyaluronan, which is a main component of the extracellular matrix of tissues. This is believed to aid penetration of the host physical barriers by the pneumococci (Boulnois, 1992, Busse, 1991, Johnston, 1991). Studies *in vitro* showed that this enzyme appears both in cell-associated fractions and in the culture supernatant, suggesting that it is released by the pathogen to the surrounding tissues. This property, most likely, assists the infection and invasion of the host (Beveridge & Graham, 1991). In recent studies, the sequences of these enzymes have been determined for both hyaluronidases from eukaryotes and prokaryotes allowing the determination of functional properties. For example, the hyaluronidase from *S. pneumoniae* cleaves the β (1,4) glycosidic linkage between N-acetyl-β-D-glucosamine and D-glucuronic acid residues in hyaluronan demonstrating its degrading effect on the structure and also in some chondroitins and chondroitin sulphate (Kelly *et al.*, 2001, Nukui *et al.*, 2003). The results obtained so far are consistent with it having a role in infection, especially the crossing of barriers between tissues and spreading of the bacterium throughout the host (Berry & Paton, 2000).

III.4 Pneumolysin

Pneumolysin is a 52 kDa protein member of the cholesterol dependent cytolysin family (Alouf, 2000, Palmer, 2001, Gilbert *et al.*, 2000), a major group of proteins that attacks membranes containing cholesterol, forming pores on the membrane mediating cell death. It is a cytoplasmic enzyme that is released after the action of surface protein autolysin. Therefore, the virulence properties of pneumolysin (ply) are directly dependent of autolysin action.

The pneumolysin is a soluble protein that oligomerises in the membrane of the target cells in order to form a large ring shaped transmembrane pore. This pore has normally a

diameter of 260Å and it is composed for 40 monomer subunits. It is known that during the process of conversion from soluble protein to membrane inserted oligomer pneumolysin undergoes drastic structural changes (Tilley & Saibil, 2006). These oligomers have been connected to the cytolytic activity of the toxin and to the abundance of cell modulatory activities that are present at sublytic concentrations. As part of its activities the following were described: inhibition of ciliary beating on respiratory epithelium and brain ependyma; inhibition of phagocyte respiratory burst and induction of cytokines synthesis and $CD4^+$ T-cell activation and chemotaxis (Hirst *et al.*, 2002, Kadioglu *et al.*, 2004). It was also shown that pneumolysin activates the classical complement pathway independently of its cell-modulatory activity (Mitchell *et al.*, 1991). *S. pneumoniae* that expresses a mutated pneumolysin which lacked haemolytic and complement activating activity was shown to be more virulent than pneumococci in which the pneumolysin gene was deleted (Alexander *et al.*, 1998) indicating that there is an unidentified function of pneumolysin that remains to be discovered. There are some suggestions for this unidentified function. The first possible function is thought to be some kind of interaction with TLR4 since one pneumococcal mutant that produced a defective pneumolysin (without haemolytic activity) could still activate TLR4 dependent response (Malley *et al.*, 2003). The second possible function is linked to reports by Baba and colleagues (Baba *et al.*, 2002) which showed that a non-cytolytic pneumolysin stimulated the production of interferon γ. However the strains that were expressing this deficient pneumolysin were not more virulent than pneumococci which had a knocked out pneumolysin (Bortoni, unpublished).

One interesting fact, regarding pneumolysin and virulence, is that despite the evidence supporting its role in pneumococcal virulence, its importance seems to vary from strain to strain as it was reported by Alexander and co workers (Alexander *et al.*,

1994). The group observed and reported that mice immunised with pneumolysin were conferred protection against nine strains but not against a tenth indicating a strain specific nature.

It has been shown that pneumolysin has an essential role in survival of pneumococci in the upper and lower respiratory tract using an acute pneumonia model (Kadioglu *et al.*, 2002, Orihuela *et al.*, 2004). There is also evidence that pneumolysin is necessary for bacterial spread from lungs into the bloodstream (Kadioglu *et al.*, 2002, Orihuela *et al.*, 2004, Berry *et al.*, 1999, Berry *et al.*, 1989). In bacteraemia, it was reported that when pneumolysin is expressed there are high numbers of pneumococci in the bloodstream and the host succumbs to infection (Orihuela *et al.*, 2004, Berry *et al.*, 1999, Berry *et al.*, 1989). However, when pneumolysin is not expressed there is high tolerance of pneumococci in the blood and there is a chance for chronic bacteraemia to develop (Benton *et al.*, 1995).

III.5 Autolysin (Lyt A)

Autolysins belong to a group of enzymes that degrades the peptidoglycan backbone of bacterial organisms. Their actions lead eventually to the lysis of the cell (Rogers *et al.*, 1980). Autolysins are located on the cell envelope and are thought to play many physiological roles, such as cell wall growth, cell turnover and finally cell separation (Tomasz, 1984). One of the best characterised enzymes belonging to this group is N-acetylmuramoyl-L-alanine amidase, also known as LytA amidase (Lopez *et al.*, 1992, Ronda *et al.*, 1987).

The gene of pneumococcal LytA has been cloned in *E.coli* (Garcia *et al.*, 1990, Garcia *et al.*, 1986). The protein has a molecular mass of around 36 kDa and it is

organised in two domains; a choline binding domain at the C-terminus composed by 20-21 amino acid residues repeats and a second domain at the N-terminus that is directly responsible for the lytic activity against the pneumococcal peptidoglycan (Lopez *et al.*, 1992, Ronda *et al.*, 1987). However, there is a functional connection between these domains because the interactions of the enzyme with the choline in the cell wall are essential for the lytic activity (Tomasz *et al.*, 1971, Briese & Hakenbeck, 1985). It has been suggested that the carboxy terminal attachment module influences the activity of the enzyme by either stabilising or inducing the active conformation of the enzyme (Briese & Hakenbeck, 1985).

One of the implications for virulence of this enzyme is the release of the cell wall components that can be highly inflammatory in animals (Tuomanen, 1999), the indirect implication its related to the fact that brings about the release of intracellular proteins such as pneumolysin (Mitchell *et al.*, 1997). Although there is evidence to support these two properties, its role in virulence it is not yet fully understood (Tomasz *et al.*, 1988). When a loss of function mutant was tested in an animal model it was observed that there was a significant decrease in the virulence in the absence of LytA (Berry & Paton, 2000). Additionally, it has been demonstrated that animals immunised with autolysin had significant longer survival on challenge, demonstrating a protective role (Berry *et al.*, 1994, Berry *et al.*, 1992, Canvin *et al.*, 1995). Therefore, autolysin has been proposed as a candidate for inclusion in a future vaccine (Berry *et al.*, 1989, Lock *et al.*, 1992).

III.6 Pneumococcal surface antigen A (PsaA)

Pneumococcal surface antigen A (PsaA) is part of an ATP Binding Cassette (ABC) type manganese transport system (Dintilhac *et al.*, 1997). PsaA is a protein with a

molecular mass of around 37 kDa and it is composed of 309 amino acid residues. The protein has an N-terminal signal peptide and a lipoprotein recognition sequence LX1X2C at the C-terminus (Sampson *et al.*, 1994). This characteristic signature sequence is responsible for the covalent attachment to a lipid component of the cell membrane via a diacylglyceryl cysteine. The likely function of PsaA is to transport Mn^{2+} and Zn^{2+} to the bacterial cytoplasm (Dintilhac *et al.*, 1997).

It was first thought that PsaA was an adhesion protein (Berry & Paton, 1996). This conclusion was supported by the fact that loss of function mutants were attenuated in virulence in an intraperitoneal and intranasal infection model mice and because cell culture experiments showed a decrease in adherence (Berry & Paton, 1996, Marra *et al.*, 2002). However the PsaA mutants also showed reduced transformability and increased susceptibility to oxygen damage (Tseng *et al.*, 2002). It is known that pneumococcal cultures require zinc and manganese for normal growth and phenotype, and the phenotypic deficiencies observed were then attributed to altered concentration of these metals in the mutant cell. In another words, lack of zinc and manganese have pleiotropic effects on the pneumococcal cell. However, PsaA might still have a role in adherence and studies to investigate this property are underway (Hammerschmidt, 2007). Recent reports have shown that antibodies against PsaA reduce adherence of pneumococci to nasopharyngeal cells (Romero-Steiner *et al.*, 2003). The PsaA protein that was initially identified due to its reaction with monoclonal antibodies also elicits protection against systemic infection and nasopharyngeal carriage (Talkington *et al.*, 1996, Briles *et al.*, 2000). All these data combined suggest that PsaA, most likely, also works as an adhesion protein (Hammerschmidt, 2007).

III.7 Choline binding protein A

Choline binding protein A (CbpA), also termed pneumococcal Surface Protein C (PspC) or more recently pneumococcal surface protein SpsA (Balachandran *et al.*, 2002, Hammerschmidt *et al.*, 1997, Rosenow *et al.*, 1997), is another example of a pneumococcal surface-exposed protein which is attached through a specific choline binding motif. The choline binding motif is composed of ten 20 amino acid residues repeats (Jedrzejas, 2001). This motif has been identified to bind to terminal choline residues present in teichoic or lipoteichoic acid structures (Rosenow *et al.*, 1997). This protein was identified and characterised as a major choline binding protein CBP with properties very similar to PspA (Rosenow *et al.*, 1997). Antibody tests showed that the protein is indeed surface exposed demonstrating a major ability to react both with human convalescent antibody and mouse protective anti-CBP serum.

This protein is composed by 663 amino acid residues and has a molecular mass of around 75 kDa. Sequence analysis determined that the molecule has a functional N-terminal module followed by a proline linker region and by the choline binding repeat region (Rosenow *et al.*, 1997). Although, the similarities between PspC and PspA are striking, there are also notable differences such as the primary sequence of the N-terminal module. Its function is directly related to adherence, making it the first known pneumococcal adhesion molecule (Cundell *et al.*, 1995). A loss of function PspC mutant was shown to bind less well to epithelial cells and sialic acid *in vitro*, and showed reduced nasopharyngeal colonisation when compared to the wild type (Rosenow *et al.*, 1997). PspC also binds to the polymeric immunoglobulin receptor that normally transports IgA. Hence, one of the other names SpsA (secretory pneumococcal surface protein A) (Rosenow *et al.*, 1997, Zhang *et al.*, 2000). It is thought that this

activity could be the first stage of translocation across the respiratory epithelium, which is in accordance with the results obtained by Jounblat and colleagues (Jounblat *et al.*, 2003) who demonstrated that a loss of function PspC mutant was less virulent in a mouse model of pneumonia. In addition, serotype 2 and 3 PspC loss of function mutants are less virulent in a sepsis model (Iannelli *et al.*, 2004). This conclusion is supported by other studies (Quin *et al.*, 2005).

One additional feature of this multifunctional protein is the binding to factor H (a component of the alternative complement pathway that is involved in the regulation of the complement activation) (Janulczyk *et al.*, 2000, Dave *et al.*, 2004). This is believed to prevent the formation of C3b on pneumococci surface, thus preventing opsonisation (Quin *et al.*, 2005). PspC also binds the C3 complement factor (Cheng *et al.*, 2000). PspC proteins are highly polymorphic (Orihuela *et al.*, 2004) and can be divided into two main structural groups, in addition to PspC that binds to phosphorylcholine, there is also the variant that was first identified as Hic (Janulczyk *et al.*, 2000) which is anchored to the bacterial cell through the motif LPXTG. As with PspC Hic can interact with the H factor, providing resistance to complement.

IV Exoglycosidases

The enzymes described in this section are, apart from neuraminidase, recently described as being involved in pneumococcal virulence. It is consensus that all of them have a very important role in the deglycosylation of the glycoproteins covering the upper respiratory tract. Although the mechanism of action in cleaving the glycoproteins is understood, the intention remains yet unclear. However, it has been shown by several groups including our group in Leicester that the action of these enzymes in the

glycoprotein structure has a direct impact in pathogenesis (Marion *et al.*, 2009, King *et al.*, 2006, King *et al.*, 2005, Terra *et al.*, 2009). The modification of the glycoproteins contributes to the provision of nutrients (Yesilkaya *et al.*, 2008). This ability of cleaving the glycoproteins may also contribute to unmasking of receptors that will elicit the adherence to the epithelial cells. Using model glycoproteins (e.g. mucin and fetuin), cleavage sites to release monosaccharides from these structures have been attributed to each of the enzymes described in the following subsections (Figure 5).

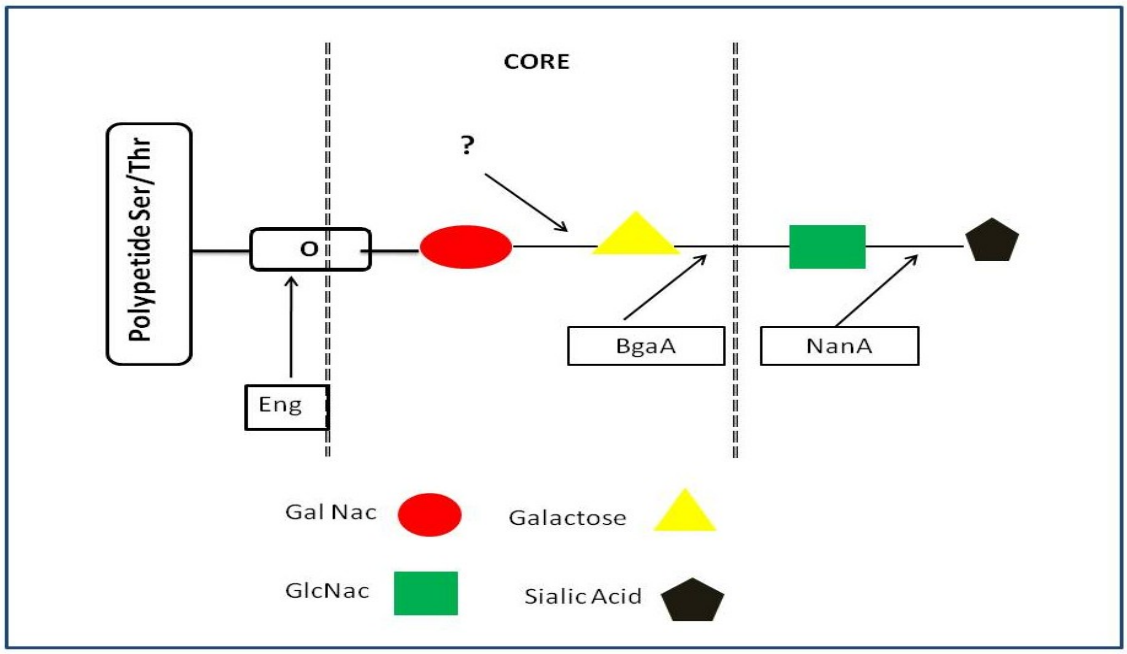


Figure 5-Schematic representation of core 1 O-linked glycan structures commonly present in human glycoconjugates. The arrows represent possible cleavage sites of exoglycosidases. The question mark is labeling the possible cleavage site for the glycosidase described in this thesis. This figure scheme was based on the scheme presented by Marion and co-workers (Marion *et al.*, 2009).

IV.1 Neuraminidases

Streptococcus pneumoniae encodes three neuraminidases, a NanA, NanB and the recently discovered NanC. While all strains encode for NanA, and virtually all strains encode for NanB, NanC only has been found in approximately 50% of strains (Pettigrew *et al.*, 2006).

Neuraminidase is an enzyme, that is present in all strains of pneumococcus studied (Berry *et al.*, 1996). Neuraminidases are also known as sialidases, because they cleave terminal sialic acid residues from cell surface glycans such as mucins, glycoproteins and glycolipids. These actions will change the glycosylation patterns of the host cell surface, revealing receptors for pneumococcal adhesion and other processes (Shakhnovich EA *et al.*, 2002; Krivan *et al.*, 1988). Structurally, NanA has a molecular mass of 108 kDa (Camara *et al.*, 1994) and NanB has 75 kDa (Berry *et al.*, 1996). NanA, but not NanB, contains a C-terminal LPXTGX motif, for the covalent binding of NanA to the peptidoglycan structures (Camara *et al.*, 1994, Schneewind *et al.*, 1995). It is important to note that these enzymes also have different pH optima. NanA has an activity pH of 7 and NanB has an activity pH of 5 (Berry *et al.*, 1994). The fact that these two enzymes have different sizes and different activity pHs and also different locations in the pneumococcal cell suggests that they are specialised for activity in different environments, perhaps one being more important to invasion while the other is more important for colonisation. Experiments performed with loss of function mutants in a mouse model of acute pneumonia showed that both NanA and NanB are important for survival in the upper and lower respiratory tract as well as survival in blood (Manco *et al.*, 2006). At the moment there is no experimental evidence for a biological role for NanC, however, this enzyme is more commonly found in isolates from the cerebrospinal fluid than in carriage isolates (Pettigrew *et al.*, 2006). Neuraminidases are thought to be involved in the degradation of host glycoproteins by being responsible for the cleavage of the terminal sialic acid, allowing for the sequential degradation of the glycoproteins by the remaining glycosidases.

IV.2 β-galactosidase A, (BgaA)

BgaA was identified by Zahner and co-workers (Zahner & Hakenbeck, 2000). This enzyme was found by performing a homology search on the *S. pneumoniae* capsular type four strain genome and it was discovered to have 26% amino acid identity compared to the β-galactosidase of *S. termophilus*. Within the region of similarity, two motifs characteristic of glycosyl hydrolase family 2 are present (Henrissat, 1991), however this protein was shown to have some differences when compared to β-galactosidases described previously, for example, all other β-galactosidases described were smaller being composed of 1000 amino acids. Additionally, all the conserved motifs between this and the previously described β-galactosidases were located in the first half of the protein. A putative signal peptide was also found at the N-terminus of BgaA. The second half of the protein did not show similarity with any other protein, except for the extreme C-terminus where a LPXTG motif was found, preceding a hydrophobic domain. This is a frequent feature of *S. pneumoniae* surface proteins (Navarre & Schneewind, 1999). This enzyme was found to be a surface protein (Zahner & Hakenbeck, 2000). It was also demonstrated by Zeleny and co-workers (Zeleny *et al.*, 1997) that this enzyme has specificity for Galβ1-4GlcNAc, a linkage existent in abundance in the mucin glycoprotein. This enzyme has been identified as being able to degrade host glycoproteins and being responsible for the cleavage of the Galβ1-4GlcNAc in glycoproteins (King *et al.*, 2006).

IV.3 β -N-Acetylglucosaminidase, StrH

The β-N-Acetylglucosaminidase (StrH) was first cloned and characterised by Clarke *et al*., (1995) and co-workers. It is encoded by a gene of 3939 bp which translates into a

protein with 1312 amino acid residues. Within this amino acid sequence a putative signal peptide was found at the N-terminus while at the C-terminus a sortase/ anchorage motif (Clarke *et al.*, 1995).

A remarkable and uncommon feature was found in this protein, a tandem repeat, within which there are 30 amino acids that remained conserved in other six hexosaminidases, giving the impression that this region might be important for substrate orientation (Clarke *et al.*, 1995). This enzyme was demonstrated to have broad substrate specificity, using N-acetylglucosamine linked to galactose (GlcNAc-Gal) by both β–1,3 and β–1,6 linkages. GlcNAc-Gal β-1,3 type of linkage is commonly found in mucins, pinpointing a possible role for this enzyme in glycoprotein degradation. In order to reinforce that this enzyme has a role in the degradation of the host glycoproteins, it was later demonstrated by King and co-workers that it has a role in the deglycosylation of host glycoproteins (King *et al.*, 2006).

IV.4 Endo-α-N-acetylgalactosaminidase, Eng

This enzyme was first described in pneumococcus by Bhavanandan and co-workers. It was found to be 160 kDa and had an optimum pH of 7.6 (Bhavanandan *et al.*, 1976). More recently this enzyme was studied by Marion and co-workers (Marion *et al.*, 2009), and the gene encoding for this enzyme as well as some of its more distinct characteristics were described. The group demonstrated that this enzyme is necessary to cleave sialylated core-1-O linked glycans.

Eng is an O-glycosidase encoded by gene *SP0368* (from *S. pneumoniae* D39) and this enzyme seems to belong to a newly identified glycosyl hydrolase family 101 (Fujita *et al.*, 2005). The biological relevance of this protein is again directly related to the

deglycosylation of glycoproteins and it was shown by Marion and co-workers that this enzyme is involved in colonisation of the upper respiratory tract and adherence to the epithelial cells. Their studies demonstrated that Eng is a surface associated protein in a sortase-dependent manner. They also discovered that the gene encoding for this enzyme was found in all available genomes and the activity of this enzyme was detectable in all strains tested, suggesting that this enzyme activity is common to all strains.

The discovery of this enzyme expanded the previous knowledge that *S. pneumoniae* can sequentially deglycosylate N-linked glycans (King *et al.*, 2006) and that can also deglycosylate O-linked glycans appearing in mucin and mucin like proteins. Finally, a role in pathogenesis of *S. pneumoniae* was indicated by Marion and co-workers, since a reduced ability for colonisation of the upper respiratory tract was shown by loss of function mutants (Marion *et al.*, 2009).

A.4 Metabolism

V Overview

The main focus of this thesis is carbohydrate metabolism, by *Streptococcus pneumoniae,* specifically the identification of pneumococcal enzymes able to degrade host glycoproteins in order to provide carbohydrates essential for growth and persistence. In order to understand better how these enzymes act, this section contains a summary of the most important concepts to keep in mind, such as glycolysis, mixed acid fermentation and the transportation method used to transport the sugar molecules inside the bacterial cell.

V.1 Carbohydrate metabolism

As all lactic acid bacteria, *S. pneumoniae* is a nutritionally fastidious facultative anaerobe (Poolman, 1993) requiring a complex medium to grow. The pneumococci are known by maintaining a strictly fermentative metabolism even in the presence of oxygen. This is due to the lack of genes required for respiration in the presence of oxygen (Poolman, 1993). *S. pneumoniae* can obtain energy only from the fermentation of carbohydrates. Mainly, these carbohydrates are oxidised to pyruvate via glycolysis (Hoskins *et al.*, 2001).

It is known that, for example, strain R6 can grow in the presence of monosaccharides such as glucose, mannose, fructose and galactose, the disaccharides sucrose, lactose trehalose, maltose and cellobiose and the trisaccharides raffinose and inulin (Tettelin *et al.*, 2001). For example, glucose, mannose and galactose are widely available to pneumococci in the host since they are main constituents of the mucin-carbohydrate side chains.

V.2 Transport of sugars by *Streptococcus pneumoniae*

There are three main sugar uptake systems in bacteria (including *S. pneumoniae*): first, the phosphoenolpyruvate (carbohydrate) phosphotransferase system (PTS) which is involved in both transport and phosphorylation of the substrate at the expense of a molecule of phosphoenolpyruvate (PEP), with the outcome of an accumulation of the resulting carbohydrate–phosphate (Postma *et al.*, 1993). Second, the ion-linked sugar transporter, which is a secondary sugar uptake system driven by an ion gradient (Poolman, 1993) and third the carbohydrate transport ATPases, primary transport systems that combine hydrolysis of ATP with translocation of the sugar molecule (ABC transporters) (Fath & Kolter, 1993).

The PTS system it is probably, the most efficient transport system energetically, since the sugar is transported and phosphorylated in one step with consumption of one molecule of PEP. This is equivalent to one molecule of ATP since during the glycolytic pathway one molecule of ATP is derived from one molecule of PEP. When sugars are transported by non-PTS systems, more than one ATP equivalent has to be spent for transport and phosphorylation (Poolman, 1993).

In the case of *Streptococcus pneumoniae,* sugar transportation is performed primarily through phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS) and ATP binding cassette (ABC) transport system (Tettelin *et al.*, 2001). There are 21 known genes encoding for PTS systems in pneumococci, once again emphasizing the role of carbohydrates for the fitness of the bacterium (Tettelin *et al.*, 2001). There are specific PTS systems for fructose, glucose, lactose, mannose mannitol, trehalose, N-acetylglucosamine and finally sucrose, while additionally there are many more PTS systems for which specificity remains yet undetermined.

V.2.1 Phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS)

The PTS transporter is a complex enzyme system that is responsible for the detection, transmembrane transport and phosphorylation of numerous sugars in both Gram negative and Gram positive bacteria (Postma & Lengeler, 1985). Generally, regardless of the organism or the carbohydrate, all PTS systems catalyse the following overall reaction:

$$\text{P-enolpyruvate}_{(in)} + \text{carbohydrate}_{(out)} \xrightarrow{PTS} \text{Pyruvate}_{(in)} + \text{carbohydrate-P}_{(in)}$$

This reaction occurs through a multitude of processes performed by a variety of enzymes (Postma *et al.*, 1993). A much more realistic depiction is the following:

$$P-enolpyruvate + enzymeI\,(EI) \Leftrightarrow P-EI + pyruvate \quad (1)$$

$$P-EI + H\,\text{Pr} \Leftrightarrow P-H\,\text{Pr} + EI \quad (2)$$

$$P-H\,\text{Pr} + EIIA \Leftrightarrow P-EIIA + H\,\text{Pr} \quad (3)$$

$$P-EIIA + EIIB \Leftrightarrow P-EIIB + EIIA \quad (4)$$

$$P-EIIB + carbohydrate \xrightarrow{EIIC} EIIB + carbohydrate-P \quad (5)$$

In the majority of cases, the enzymes EI and HPr are soluble and cytoplasmic. These proteins have an active participation in the phosphorylation of all PTS transported carbohydrates; therefore they are named the general PTS proteins. On the other hand, the enzymes EII are carbohydrate specific. Hence, bacteria usually contain many different EII (Deutscher *et al.*, 2006). Each EII complex consists of one or two hydrophobic integral membrane protein (domains C and D) and two hydrophilic

domains (domain A and B), which together are responsible for the transport of the carbohydrate across the membrane and its phosphorylation.

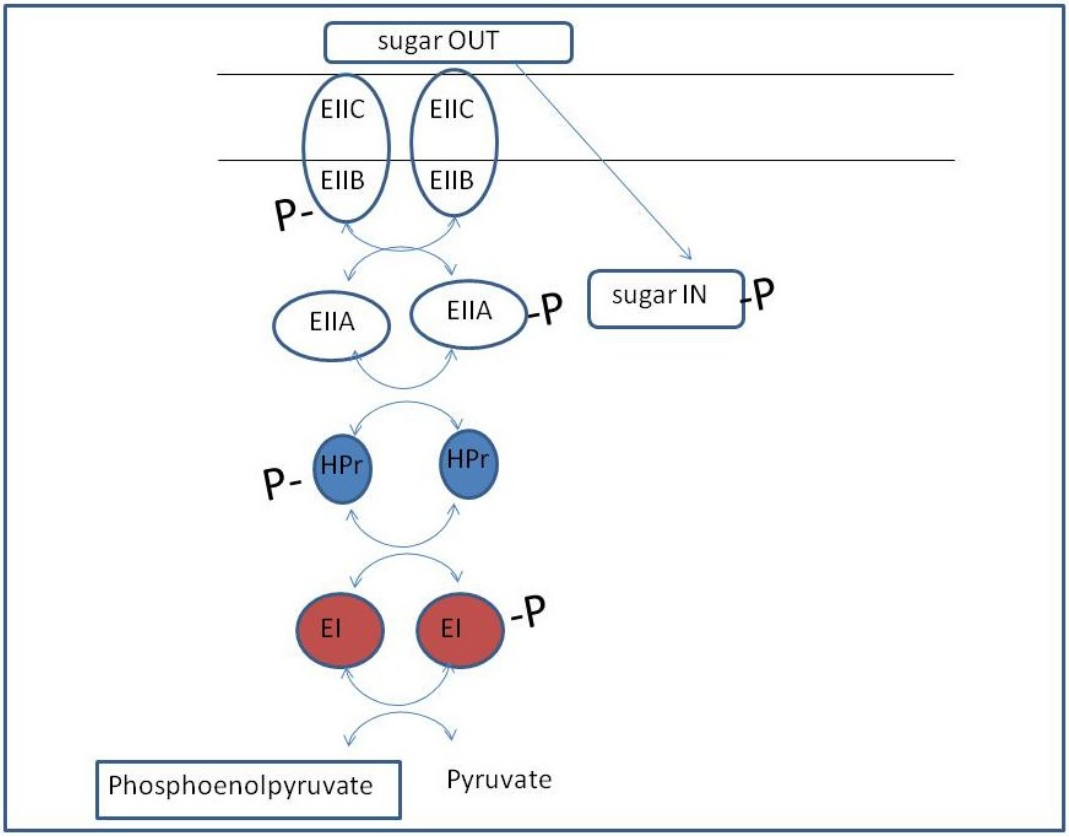


Figure 6-Carbohydrate transport and phosphorylation by the PTS. Carbohydrates are transported and simultaneously phosphorylated by the PTS.

Depending on the sugars entering the cell through PTS system, they either enter the glycolysis directly, or go into Leloir or Tagatose pathway and then into the glycolytic pathway (Cocaign-Bousquet *et al.*, 1996).

V.2.2 ATP Binding Cassette (ABC) transporters

ABC transporters are found in all prokaryotic and eukaryotic cells. They use ATP energy to translocate molecules through the membrane. The translocator component is

composed of two multi-transmembrane and two intracellular ATP binding subunits (Higgins, 1992).

ABC transport is mediated by specific binding proteins which in Gram negative bacteria are free in the periplasm. However, in Gram positive bacteria, these specific binding proteins are anchored via an N-terminal lipid motif to the outer surface of the cell membrane due to the lack of periplasm (Gilson *et al.*, 1988). Nevertheless, once the sugar molecule binds to the specific binding proteins, they still interact with the membrane spanning protein exactly in the same way as Gram negative bacteria. This kind of transport system includes three main proteins: the specific binding proteins, the membrane spanning transporter and finally the ATP hydrolysing protein which is the one providing the energy by hydrolysing ATP (Madigan *et al.*, 2003). ABC transporters are energetically less efficient, since one molecule of ATP is consumed in the process of transport through the membrane (Hoskins *et al.*, 2001). Despite this diminished efficiency, in *S. pneumoniae* there are a variety of sugars that enter the bacterial cell using this process of transport. Figure 7 is a schematic representation of the ABC transporter.

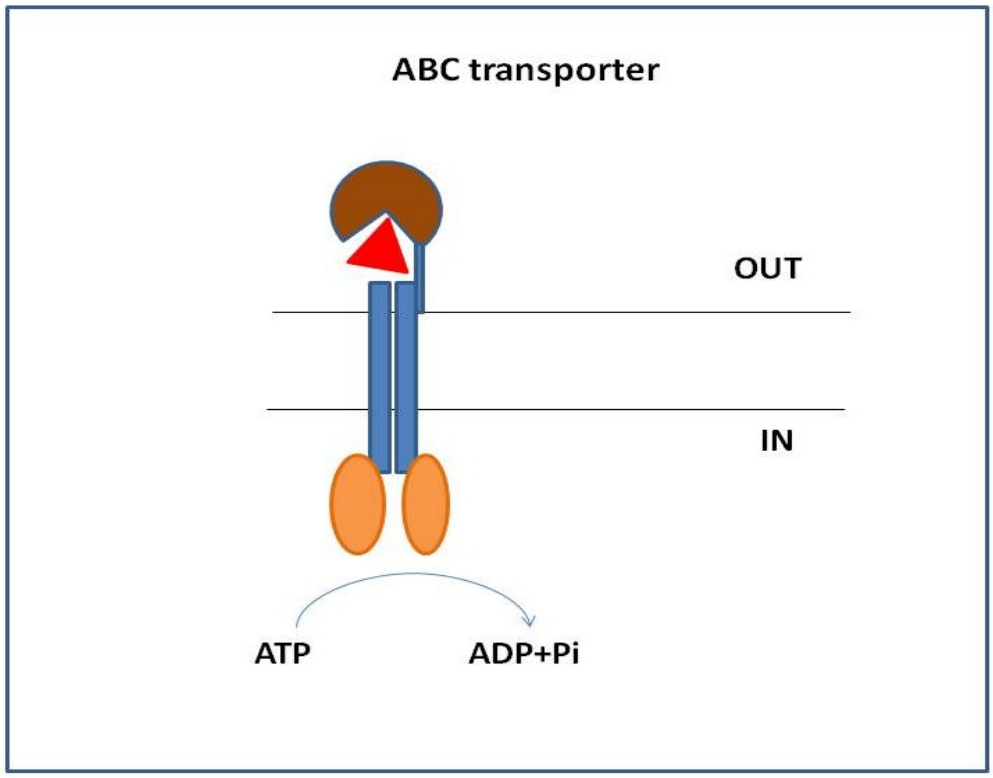


Figure 7- Carbohydrate transport by the ABC transporter. Carbohydrates are transported at the expense of ATP. Red represents the sugar molecule, brown represents the specific binding proteins, blue represents the membrane spanning transporter and yellow represents the ATP hydrolyzing protein.

V.3 Glycolysis

The glycolytic pathway is common to virtually all eukaryotic and prokaryotic cells (Berg *et al.*, 2001). This pathway can be divided into three stages. Stage one where fructose-1,6-bisphosphate is produced. Stage two where the cleavage of the fructose-1,6-bisphosphate in two three interconvertible carbon fragments occurs and stage three where ATP is used in the oxidation of the three carbon fragments into pyruvate, NADH and there is a net gain of two ATP per molecule of substrate (Neves *et al.*, 2002, Berg *et al.*, 2001).

Glycolysis is the most important metabolic process inside the cell and holds a central role. Its main product, pyruvate, can quickly be directed into other metabolic processes. Figure 8 is a scheme of the glycolysis pathway.

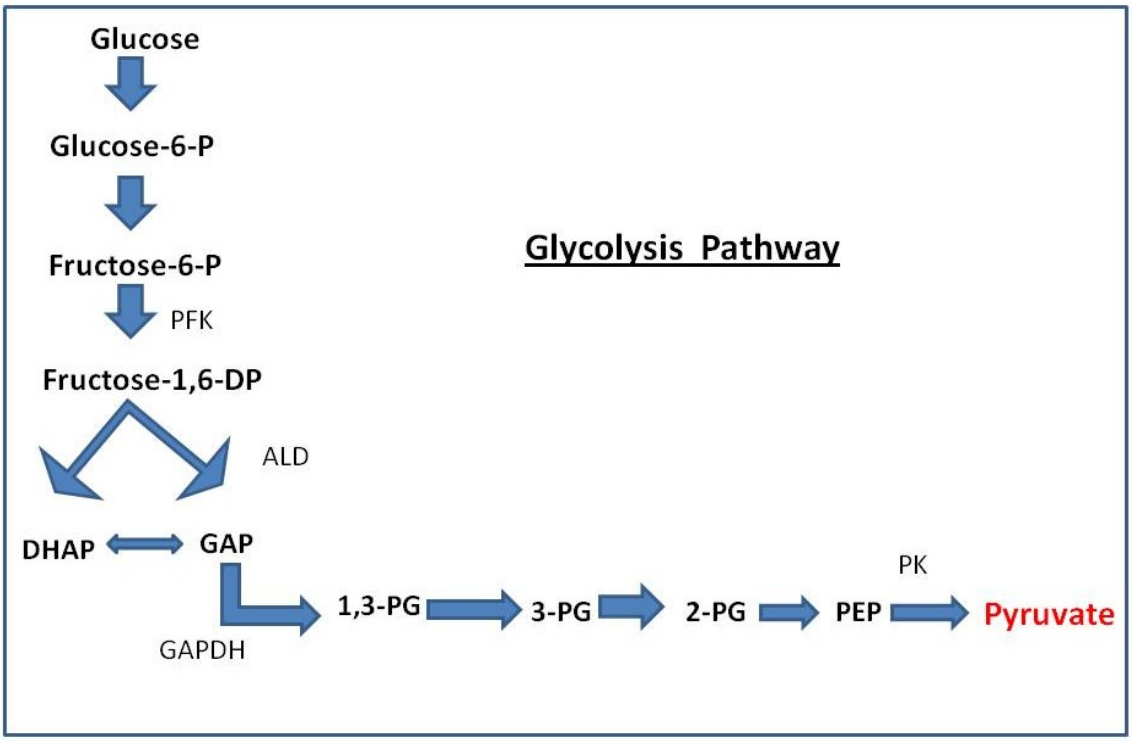


Figure 8-Schematic representation of the glycolytic pathway. PFK- Phosphofructokinase; ALD –Aldolase; DHAP-dihydroxiacetone phosphate; GAP-glyceraldehyde-3-phosphate; GAPDH –Glyceraldehyde-3-phosphate dehydrogenase; 1,3 –PG–1,3- bisphosphoglycerate; 3-PG – 3- phosphoglycerate; 2-PG- 2-phosphoglycerate; PEP – phosphoenolpyruvate; PK- Pyruvate Kinase.

V.3.1 Glycolysis – Growth in Glucose

When glucose is the main carbon source, it is translocated inside the cell through the PTS system EII $^{(Gluc)}$ entering the classical pathway of glycolysis (Figure 8) as glucose-6-phosphate (G6P) (Cocaign-Bousquet *et al.*, 1996). Glucose may also enter the cell through an ABC transporter; in any case the entry point in the glycolytic pathway will be the same except it must be phosphorylated before. The pyruvate formed from the subsequent glycolytic activities is converted into lactate by lactate dehydrogenase with

NADH being oxidized to NAD^+. This process is known as homolactic fermentation (Figure 9).

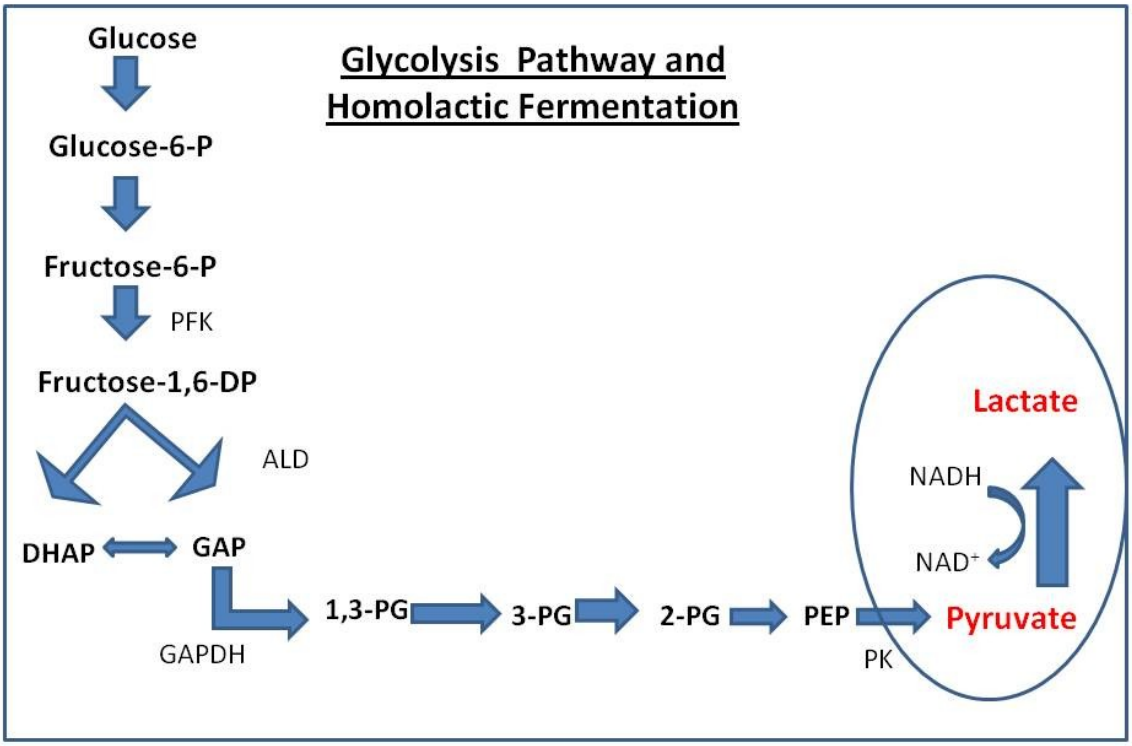


Figure 9- Glycolytic pathway followed by homolactic fermentation. The final main product in homolactic fermentation is lactate. PFK- Phosphofructokinase; ALD – Aldolase; DHAP-dihydroxiacetone phosphate; GAP-glyceraldehyde-3-phosphate; GAPDH –Glyceraldehyde-3-phosphate dehydrogenase; 1,3 –PG–1,3-bisphosphoglycerate; 3-PG – 3- phosphoglycerate; 2-PG- 2-phosphoglycerate; PEP – phosphoenolpyruvate; PK- Pyruvate Kinase.

V.3.2 Glycolysis –Growth in Galactose

The metabolism of galactose is different from that of glucose in many aspects. Firstly, galactose does not enter glycolysis directly, and secondly galactose is not a preferred sugar. Therefore, rather than homolactic fermentation, mixed acid fermentation following glycolysis is very likely to occur (Neves *et al.*, 2002).

Depending if the sugar is translocated through the PTS system or ABC transporter it will enter the Tagatose (Bissett & Anderson, 1973, Bissett & Anderson, 1974a, Bissett & Anderson, 1974b) or Leloir (Frey, 1996) pathways respectively.

The Tagatose pathway occurs in the presence of either lactose or galactose translocated into the cell through the PTS system, before entering glycolysis. This pathway converts Gal-6-P into Tagatose-6-P (Tag-6-P) through the action of galactose-phosphate-isomerase. Tag-6-P is then converted into Tagatose-1, 6-diphosphate which will finally be converted into the DHAP and GAP by the action of the enzyme tagatose-bisphosphate-aldolase. DHAP and GAP are the intermediaries that enter glycolysis.

The Leloir pathway occurs when galactose enters the cell through a permease. In this system, the sugar needs to be phosphorylated first, being converted into galactose-1-phosphate (Gal-1P) followed by conversion into glucose-1-phosphate (Glu-1P), by the enzyme galactose uridylyl transferase. Subsequently Glu-1-P is converted by phosphoglycerate mutase into glucose-6-phosphate (Glu-6-P) (Cocaign-Bousquet *et al.*, 2002).

If lactose is the fermentation sugar it is cleaved by β-galactosidase to release the galactose and glucose moiety. The glucose will enter glycolysis and the galactose enters

Leloir pathway as shown on Figure 10. This cleavage of lactose is necessary for both pathways (Figure 10).

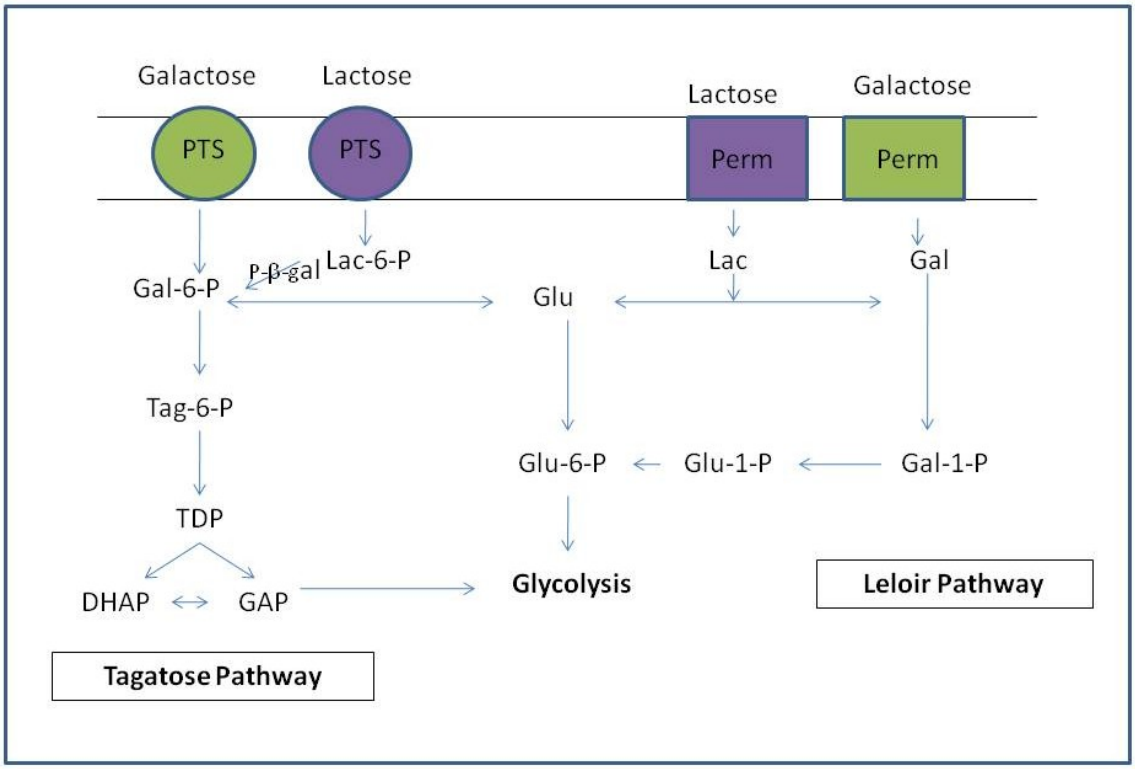


Figure 10- Schematic representation of the interaction between Glycolysis, the Tagatose and Leloir Pathways This figure was based on (Garrigues *et al.*, 1997).

V.4 Mixed acid fermentation

As previously stated above, pneumococci convert pyruvate formed during glycolysis into lactic acid in homolactic fermentation. However, when deprived from the preferred sugar glucose, in anaerobiosis or in sugar limitation, a shift to mixed acid fermentation occurs with products other than lactate, such as ethanol acetate and formate being formed as shown in Figure 11 (Neves *et al.*, 2002, Melchiorsen *et al.*, 2000, Neijssel *et al.*, 1997).

In bacteria, mixed acid fermentation results from two different processes, one involves PFL (pyruvate formate lyase) and the other one requires PDHC (pyruvate

dehydrogenase complex) (Neves *et al.*, 2002, Melchiorsen *et al.*, 2000, Neijssel *et al.*, 1997).

In aerobiosis, PDHC catalyses the oxidative descarboxylation of pyruvate to acetyl-CoA and CO_2. In anaerobiosis or microaerobic environments, on the other hand the majority of the pyruvate is converted to acetyl-CoA and formate by the action of oxygen-sensitive pyruvate formate lyase (PFL) (Neves *et al.*, 2002, Zhu & Shimizu, 2004). In *S. pneumoniae* no genes encoding for PDHC have been identified or any pyruvate dehydrogenase activity is detected in cell free extracts (Hoskins *et al.*, 2001, Tettelin *et al.*, 2001, Smith *et al.*, 2002). Therefore, mixed acid fermentation is expected to occur through PFL (Yesilkaya *et al.*, 2009).

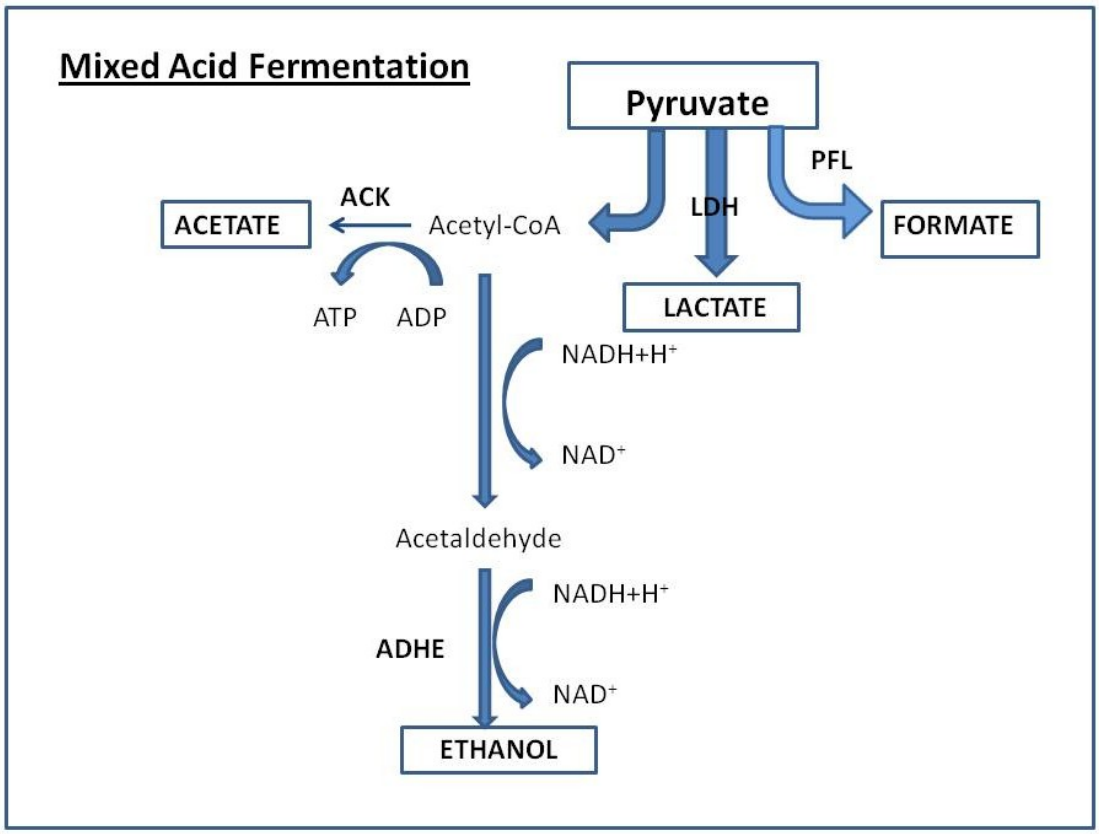


Figure 11- Schematic representation of Mixed acid fermentation. This figure was based on (Cocaign-Bousquet *et al.*, 2002). LDH - lactate dehydrogenase; PFL- pyruvate formate lyase; ACK- acetate kinase; ADHE- alcohol dehydrogenase.

But the question is what does regulate this metabolic shift? Initially, this shift was explained by allosteric modulation of two enzymes competing for pyruvate, pyruvate lactate dehydrogenase (LDH) and pyruvate formate lyase (PFL) (Cocaign-Bousquet *et al.*, 1996). It was thought then that during homolactic fermentation, the pool of FDP could activate pyruvate kinase (PK) and LDH, directing the flux to lactate production (Neves *et al.*, 2005). On the other hand, the formation of end products different from lactate was thought to be due to the reduction of LDH activity caused by the low levels of the effector FDP, and to the relief of pyruvate formate lyase inhibition by the concomitant DHAP and glyceraldehyde-3-phosphate.

Recently, the observation that intracellular FDP is, under all conditions tested, sufficiently high to ensure full activation of the LDH, makes the role of FDP in the regulation of pyruvate distribution between LDH and PFL unclear (Garrigues *et al.*, 1997).

It was then proposed by Garrides and co-workers that there was a role for the NADH/NAD^+ ratio in the modulation of both GAPDH and LDH suggesting that homolactic fermentation was a result of a high NADH/NAD^+ ratio (Garrigues *et al.*, 1997). This high ratio would restrain GAPDH but not LDH and would cause an accumulation of metabolites upstream of the former enzyme, such as FDP or the sugar trioses, which will inhibit PFL. Therefore, the high NADH/NAD^+ ratio had as a consequence homolactic fermentation.

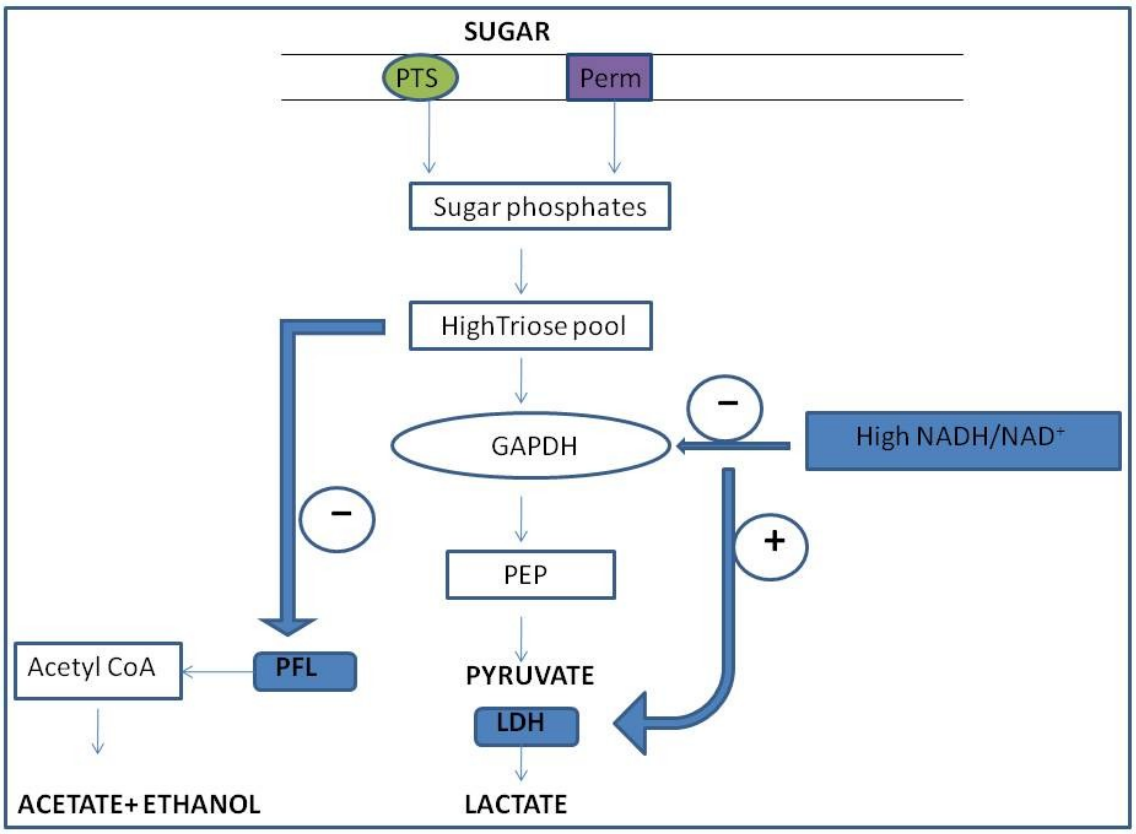


Figure 12-Model of glycolysis regulation. Based on (Garrigues *et al.*, 2001). Perm for permease, PTS - Phosphotransferase system. GAPDH - glyceraldehydes-3-phosphate-dehydrogenase; PEP – phosphenolpyruvate; PFL - pyruvate formate lyase; LDH-Lactate dehydrogenase.

The mixed acid fermentation, observed for galactose and lactose, was a consequence of a lower NADH/NAD^+ ratio liberating GAPDH of the inhibition and limiting LDH directing the fermentation to mixed acid fermentation (Garrigues *et al.*, 1997).

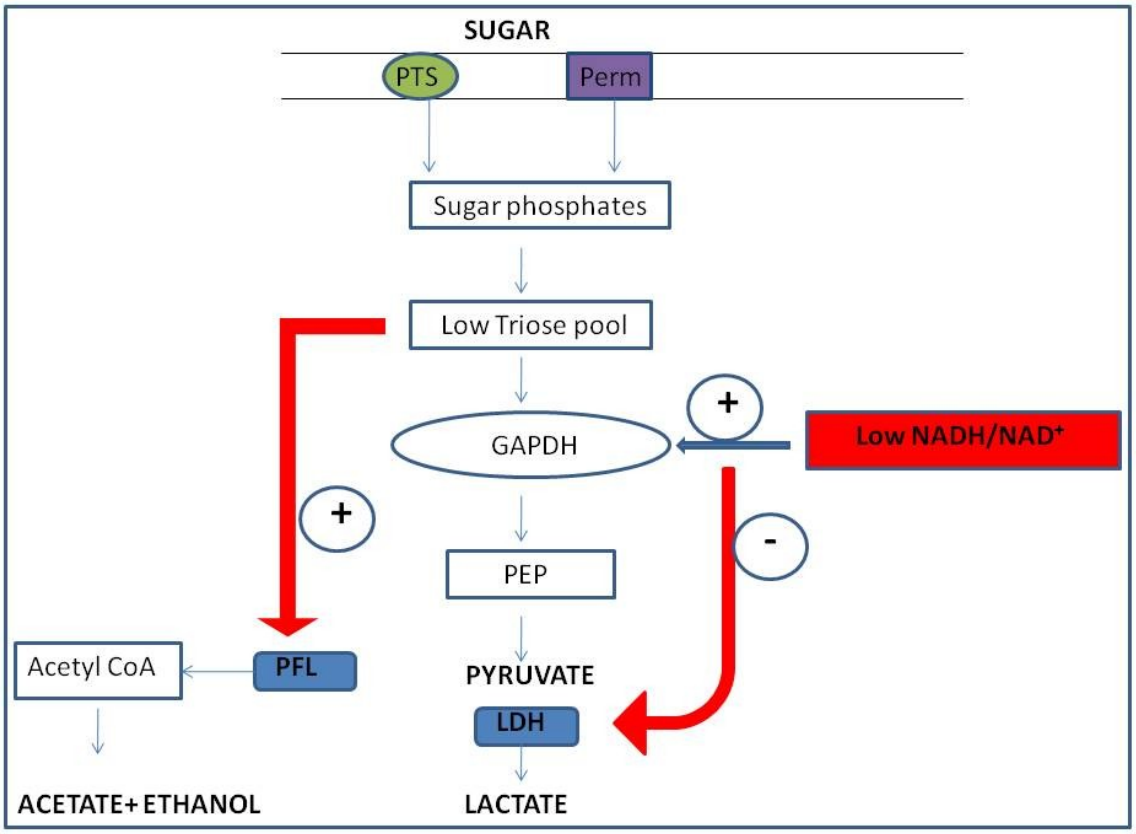


Figure 13- Model of glycolysis regulation when the NADH/NAD^+ ratio is low, directing the fermentation to mixed acid fermentation. Based on (Garrigues *et al.*, 2001). Perm – permease; PTS - Phosphotransferase system; GAPDH - glyceraldehydes-3-phosphate-dehydrogenase; PEP – phosphenolpyruvate; PFL- pyruvate formate lyase; LDH - Lactate dehydrogenase.

Although all these possible regulation mechanisms are established (allosteric modulation of enzymes, metabolite levels and more recently transcript and protein levels), the reasons behind the metabolic switch to mixed acid fermentation remain unknown. However, allosteric mechanisms involving the phosphorylated pools and the redox balance are very likely to be involved (Garrigues *et al.*, 1997).

VI Carbon Catabolite Repression

It is known that a bacterium expresses its genes differentially based on the extracellular environment, expressing only the ones it needs to propagate. For instance, if a particular substrate is absent, the genes necessary for its uptake and metabolism are often repressed. The subsequent availability of substrate then leads to the relief of the subset of enzymes (Deutscher *et al.*, 2006).

Over 100 years ago, it was observed that growth in glucose can lower the activity of other enzymes; this phenomenon was known as the glucose effect. This meant that some organisms, in this particular case *Bacillus subtilis* would use glucose first as a preferential sugar and use the other available sugars only when all the glucose had been metabolised. It appears that for each organism a particular hierarchy of utilisation of carbohydrates is observed, with glucose appearing as the one having the primary role.

In the following years, it was found that as long as glucose, fructose or sucrose are present in the media, there is a repression of the enzymes encoding for the metabolism and uptake of the less favourable carbon sources. This phenomenon is known as carbon catabolite repression (CCR) (Contesse *et al.*, 1970).

When the primary sugar is exhausted, bacteria first need to synthesise all the enzymes necessary for the transport and metabolism of the less preferential carbon sources and while this is happening, a lag phase is observed on the growth. The glucose effect observed in the beginning of the last century can therefore be considered CCR. CCR has been, ever since, extensively studied and it was found that the preferential use of one particular carbon source over the others involves either activation or repression

of transcription (Deutscher *et al.*, 2006) of the genes encoding for the utilisation of other carbon sources

VI.1 PTS system and its role in Carbon Catabolite Repression

It is now known that PTS-PEP transport system plays a very important role in bacterial CCR (Deutscher *et al.*, 2006). Despite the fact that this system is involved both in Gram positive and Gram negative bacteria, the mechanisms implicated are different. In both gram positive and gram negative bacteria there are different proteins from the PTS system involved (Saier *et al.*, 1995).

VI.1.1 PTS and Gram positive (low GC content)

The most PTS mediated CCR mechanisms respond to the phosphorylation level of a PTS protein, which is controlled by the metabolic state of the cell (Deutscher *et al.*, 2006). In Firmicutes, however, the protein HPr is the one exerting this role contrarily to the phylum Enterobacteriaceae where the main focus is on $EIIA^{Glu}$. For this particular phylum, HPr is phosphorylated not only in His15 in the PEP dependent reaction, but also on the Ser46 in an ATP-requiring reaction. This protein therefore exists in four different forms each one of them with a particular regulatory role in a particular metabolic state (Deutscher *et al.*, 2006).

VI.1.2 Carbon Catabolite Repression, the mechanism

In Firmicutes, CCR is directly related to the phosphorylation level of the protein HPr. As it was mentioned before, this particular protein can either be phosphorylated on the His15 in a PEP-dependent reaction or on Ser-46 in an ATP-requiring reaction. In this particular subsection, the later is the main focus, since it has been shown to be the one relevant for the regulation of CCR although, HPr(His-P) also has a relevant role in the presence of glucose (Deutscher, 2008).

The phosphorylation of Ser46 occurs when the intracellular concentrations of FDP and ATP are high, thus enabling the cell to recognise the presence of preferential carbon source in the surrounding environment (Gorke & Stulke, 2008). HPr protein phosphorylated on Ser46 (HPr(ser-P)) binds to Ccpa protein acting as an effector (Jones *et al.*, 1997). This interaction is also enhanced by the glycolytic pathway intermediates FDP and Glucose-6-phosphate (Glu-6-P) (Monedero *et al.*, 2001). This complex, once formed, binds to the catabolite responsive elements (*cre*) sites and can either repress or activate the transcription of genes (Deutscher *et al.*, 1995) encoding for enzymes involved in the catabolism. The activation or repression of the transcription of the genes depends of the location of the *cre*. If *cre* is located upstream -35 region of the promoter it will lead to CCA, carbon catabolite activation. If however the cre region is located within or downstream of the promoter, it will lead to CCR, carbon catabolite repression (Fujita, 2009).

On the other hand, in case of poor nutritional supply, high inorganic phosphate, low ATP and low FDP will accumulate in the cell, HPr Kinase, which is also responsible for the dephosphorylation of the HPr protein on Ser 46 will act. As a result of the phosphate group being cleaved, the enzymes necessary for the utilisation of not preferential carbon sources will be synthesised (Gorke & Stulke, 2008), relieving the CCR.

In addition to this, HPr (His-P) also contributes to CCR. In the absence of glucose, HPr(His-P) will phosphorylate Glycerol kinase (GlpK) and transcriptional regulators that contain phosphoenolpyruvate-carbohydrate phosphotransferase system regulatory domains (PRD's), essential for their activity. In the presence of glucose, the level of

phosphorylation of GlpK is not sufficient leading to the inactivation of GlpK therefore leading to inducer exclusion (Deutscher *et al.*, 2006).

VI.2 CCR in *Streptococcus pneumoniae*

Due to the fact that *Streptococcus pneumoniae* colonises the nasopharynx, an environment known by being rich in sugars other than glucose, it is important to try to establish a relation between the availability of substrate and propagation of the bacteria and in a final note establish a relationship between metabolism and virulence.

It is believed that CCR occurs in *Streptococcus pneumoniae* in a similar manner to what happens in *Bacillus subtilis*. CCR in *Bacillus subtilis* is described in section VI.1.2 of this chapter. However similar the mechanisms, there are important differences that during the course of years have been described. In this section the characterised differences of CCR in *Streptococcus pneumoniae* are described.

A decade ago Rosenow and co-workers, described the raffinose utilisation system and following the finding of a *cre* consensus region upstream of the translation initiation codon for their α-galactosidase (*aga*) hypothesised how carbon catabolite repression would be, since in the vast majority of gram positive bacteria it is mediated through Ccpa protein. Obviously, they started by identifying and inactivating a homolog of Ccpa from *Bacillus subtilis* in *S. pneumoniae*. However, the inactivation of this gene had no effect on the regulation of α-galactosidase (*aga)*. Because inactivation of Ccpa has the same phenotype in *Streptococcus mutans,* it was concluded by them that catabolite repression in *Streptococcus pneumoniae* is either mediated by PTS directly, or by some unidentified protein (Rosenow *et al.*, 1999).

Just a few years later, Giammarinaro & Paton, showed that CcpA homologue RegM is involved in *S. pneumoniae* virulence, which will be discussed in the next section in detail and that there is a regulation role for this homologue of CcpA in the regulation of carbon metabolism. Their studies also showed that the disruption of the gene *regM* led to alteration of the growth rate in semi-synthetic media supplemented with different carbon sources. Additionally, induction of sugar metabolism enzymes and loss of virulence was observed. The authors speculated that, the apparent involvement in the control of endogenous β-galactosidase and α-glucosidase and the change in growth rates as a function of the carbon source added would indicate a role for RegM in regulation of sugar metabolism. However, and contrarily to what has been described for other bacteria, CcpA homologue does not appear to be involved in the glucose mediated repression of β-galactosidase expression. Finally, and regarding the role of RegM in carbon catabolite repression, Giammarinaro and Paton located *cre* like sequences in the promoters and coding regions of potential β-galactosidases and α-glucosidases genes (Giammarinaro & Paton, 2002).

In 2005, Iyer and co-workers, produced remarkable observations, when they proved that CcpA is not a universal regulator of CCR. They have demonstrated that there is a second regulatory system independent of CcpA by proving that even in the absence of CcpA, α-galactosidase and β-glucosidase activities were still subject to carbon catabolite repression. Alternatively, the loss of CcpA in the presence of sucrose led to a very pronounced CCR effect in α-glucosidase was also seen, once again showing an independent regulatory system. Iyer and co-workers demonstrated that, in *S. pneumoniae,* CcpA is not a universal regulator of CCR (Iyer *et al.*, 2005).

Finally, in 2007, Kaufman and Yother, showed that a point mutation located in the promoter of the *pts* gene upstream of the BgaA gene altered the ability of CcpA to bind the *cre*. In fact this point mutation was within the *cre* sequence and consisted in a transversion of a G to a C. This point mutation had a major impact, since it was on a universal conserved residue in *cre* sequences. Additionally, it had as a consequence decreased affinity of CcpA for the *cre* sequence, resulting in a reduced ability to inhibit the binding of RNA polymerase leading to constitutive expression of β-galactosidase. It is known that CcpA represses β-gal activity by binding to the promoter region of the *pts* operon, therefore a second repressor was revealed when CcpA was deleted in D39 parental strain and in the BgaA constitutive mutant.

The weak binding of CcpA in this BgaA constitutive mutant seemed to inhibit somehow the binding of the second repressor, therefore leading to high levels of *pts-bgaA* expression and β-galactosidase activity. When CcpA was deleted both from the mutant and parental strain D39, a reduction on the *pts-bga* expression was observed for the mutant, showing that the second repressor is able to bind in the absence of CcpA. In the case of D39 even though some repression was observed, it was not of the same level as when CcpA was present. This experiment also determined that the second repressor may bind to a different site than CcpA, since it binding was observed in both the mutated site in the constitutive mutant and D39 parental strain. The main conclusion of these studies was that glucose mediated repression occurs both through CcpA and the second repressor in D39.

Even though these results were objective, they do contradict results obtained previously by Giammarinaro & Paton, (2002), where it was seen that CcpA-glucose mediated repression of β−galactosidase activity did not occur in D39. This discrepancy

is most likely related to the effect of the second repressor which may have obscured the effect of glucose in their studies. Even when CcpA is deleted the binding of the second repressor allows some glucose mediated repression (Kaufman & Yother, 2007). In conclusion, the regulation of β-galactosidase expression seems to involve multiple levels of repression, including the presence of glucose and the activities of CcpA and second repressor.

Despite differences to previous studies, glucose mediated repression follows a logic pathway, in the sense that in niches where glucose levels are low, such as the nasopharynx, repression of *pts-bgaA* genes is relieved, resulting in higher level of expression of β-galactosidase and pts. However, if glucose concentration rises in sites of infection that would lead to a repression of *pts-bgaA* and the bacterial cell would use glucose as carbon source. Thus, it seems likely that CcpA and the second repressor may therefore play a central role in regulating the transition from colonisation to systemic infection (Kaufman & Yother, 2007).

VI.3 CcpA relation with virulence in *S. pneumoniae*

The same research groups that were involved in unravelling the process of CcpA regulation in *S. pneumoniae*, did important work in relating CcpA and virulence. Giammarinaro and Paton, found that in the absence of RegM gene, the homologous of CcpA, pneumococcal cells struggle to grow in a semi-synthetic medium supplemented with glucose. Since glucose is the main sugar present in the blood, this would mean an inability to grow in the presence of glucose. However, fresh heparinised mouse blood was used to grow both the control D39 and D39 carrying the mutation in RegM no differences were observed. Thus, some other factors were responsible for the attenuation in virulence (Giammarinaro & Paton, 2002).

One possible explanation for this observation was that RegM is required for the optimal expression of essential virulence factors. However, after using Western blot to detect some of the best characterised virulence factors they failed to detect any differences of expression between wild type and mutant. Nonetheless, the group demonstrated that deletions of RegM down-regulates the transcription of the capsule locus. Thus, it was proposed that the avirulent phenotype observed for the RegM mutant was largely due to the inability to produce sufficient capsule to survive in the host. However, the possibility of the involvement of unknown factors regulated by RegM was not excluded.

Giammarinaro and Paton also demonstrated that capsule gene expression is dependent upon sugar source, since, when using the same semi-synthetic media supplemented with lactose, the capsule genes were also down-regulated. RegM may be a regulatory link between these sugar pathways and capsule production.

It was concluded that RegM directly or indirectly regulates the expression of endogenous proteins such as β-gal and α-glu, and it is the first product of a gene outside the *cps* locus (capsule locus) to be directly linked to regulation in capsule biosynthesis (Giammarinaro & Paton, 2002).

Iyer and co-workers also studied the relationship between CcpA and virulence, since *S. pneumoniae* is an obligate commensal and clinically important pathogen. They observed that, in contrast to the *in vitro* observations where CcpA could be expendable, the CcpA strain was severely attenuated in infection of the lungs and colonisation of the nasopharynx. Upon complementation of the mutant in trans, the virulence in the lungs was restored and the virulence in the nasopharynx was partially restored, even though was still significant. The group attributed this observation to aberrant expression of

CcpA from the complementation of the construct. It is known that CcpA autoregulates itself in other bacteria (Mahr *et al.*, 2000, Egeter & Bruckner, 1996), containing a *cre* consensus sequence upstream of the gene. Thus, the gene was inspected for the presence of a *cre* sequence and it was found that CcpA gene in *S. pneumoniae* does contain a candidate for a *cre* sequence 140bp upstream of the start codon. Therefore, it was suggested that CcpA autoregulation might be required for the optimal regulation of target genes *in vivo* and that CcpA is directly or indirectly regulating genes that are required for *in vivo* fitness.

When the list of up and down regulated genes was analysed, the proteins associated with functions in central and intermediary metabolism were in prominent place, thus the changes in basic metabolism could be responsible for the defects observed *in vivo*. Definitely, the loss of CcpA caused down-regulation of glycolytic enzymes like enolase known for contributing for pneumococcal virulence. It is therefore imaginable that misregulation of these factors can harmfully affect colonisation of the nasopharynx and survival in the lungs during infection leading to complete and rapid clearance (Iyer *et al.*, 2005).

Further studies are required in the elucidating the metabolic pathways of carbohydrate metabolism of *S. pneumoniae*, and in particular in investigating the strict relation that seems to exist between CcpA and virulence, however it seems clear that besides the regulatory function in Carbon Catabolite repression, CcpA also plays a central role in virulence.

A.5 Aims

The aims of this project were to identify novel sugar hydrolases involved in mucin degradation, in an attempt to better understand how pneumococci persist in the host where the levels of glucose are low. In parallel, we aimed to fully characterise the targeted enzymes. This process entailed confirmation of annotation, identification of specific substrates both using a synthetic substrate and a glycoprotein and sub cellular localisation. The usefulness of the chosen enzymes as protective agents against pneumococcal disease was also investigated. We also aimed to establish a link between the chosen enzymes and virulence of pneumococci.

In this manner this thesis is organised in seven chapters. Chapter one where there is a broad introduction focusing in four main aspects, the bacteria, the mucin, virulence factors in pneumococci and a last section where sugar metabolism is reviewed. Chapters three, four, five and six are results chapters where a description of the path followed to fully characterise the two enzymes chosen was done. Finally Chapter seven is a discussion of all data and observations reported in this thesis.

B. Material and Methods

B.1 Chemicals and Biologic Material

Unless otherwise stated all chemical used in this study were obtained from Sigma, Fisher or Oxoid. The bacterial strains and plasmids used in this study are listed in Table 1 and Table 2 respectively; the mutants are listed in Table 3.

Table 1- List of the bacteria used throughout this project.

Strains	**Genotype**	**Source**
E.coli BL21 DE3	*F-ompT hsdsB (rb- mb-) gal dcm (DE3)*	Novagen
E. coli Fusion-blue competent cells	*endA1, hsdR17 (rK12-, mK12+), supE44, thi-1, recA1, gyrA96, relA1, lac'[proA+B+,acIqZΔM15::Tn10(tetR)]*	Clontech
Streptococcus pneumoniae (D39)	Virulent type 2 strain	The National Collection of Type Cultures, London, UK Strain NTCC 7466
S. pneumoniae	Serotype 6B strain IO11966	(Enright & Spratt, 1998)
S. pneumoniae	Serotype 19F, ST type 177, strain OXC1261	(Enright & Spratt, 1998)

Table 2- List of plasmids used throughout this study.

Plasmids	Study	Source
Donor plasmid pR412 (spectinomycin resistance)	Mariner mutagenesis	Dr Marc PrudHomme CNRS-Universite Paul Sabatier Toulouse France
Donor plasmid pR410 (kanamycin resistance)	Mariner mutagenesis	Dr Marc PrudHomme CNRS-Universite Paul Sabatier Toulouse France
p*Ecoli* - Nterm 6X HN	Cloning	Clontech
pCep	Complementation	Dr Marc PrudHomme CNRS-Universite Paul Sabatier Toulouse France (Guiral *et al*., 2006)

Table 3- List of the mutations performed on putative glycosidases in D39 background.

The mutants presented in Table 3 were constructed by Dr Hasan Yesilkaya prior the commencement of the work for this thesis.

Mutations	Putative glycosidase	Origin
SPD0063	β-N-acetylhexosaminidase	This study
SPD0065	β-galactosidase 3	This study
SPD0247	6-phospho-β-glucosidase	This study
SPD0277	cell division protein	This study
SPD0427	6-phospho-β-galactosidase	This study
SPD0444	endo-β-Nacetylglucosaminidase	This study
SPD0503	6-phospho-β-glucosidase	This study
SPD0562	β-galactosidase	This study
SPD0853	endo-β-Nacetylglucosaminidase	This study
SPD1046	6-phospho-β-galactosidase	This study
SPD1678	α- galactosidase	This study
SPD1898	hypothetical protein	This study
SPD1969	glycosyl hydrolase related protein	This study

B.2 Growth conditions and media

All media were prepared using distilled water and were sterilised by autoclaving at 121°C at 15 psi (pounds per square inch) for 15 min.

Escherichia coli strains were grown in Luria Broth (LB: 10 g/l NaCl, 5 g/l yeast extract and 10 g/l trypticase peptone) with shacking at 37 °C or on Luria Agar (LA) plates (LB with added 1.5 % w/v bacteriological agar).

S. pneumoniae strains were grown in Brain Heart Infusion broth (BHI, Oxoid) or in Blood Agar base plates (BA, Oxoid) statically at 37 °C in a gas jar containing a candle to remove the oxygen and promote the anaerobic growth. When appropriate Sicard's defined medium (SICARD, 1964) was used. The composition of the defined medium is described in Table 4-7.

Table 4- Composition of salt solution used for Sicard's defined medium

A. Minerals and Buffer	**Quantities (g)**
NaCl	10
NH_4Cl	4
KCl	0.8
Na_2HPO_4	0.24
$MgSO_4$	0.048
$CaCl_2$	0.002
$FeSO_4.7H_2O$	0.00110
Tris	9.68
H_2O	1L

Table 5- Amino acids necessary to add to salt solution A for Sicard's defined medium

B. Amino Acids	**Quantities (mg)**
L-Arginine	400mg
L-Asparagine	20mg
Glycine	240mg
L-Histidine	300mg
L-Isoleucine	13.1mg
L-Leucine	13.1mg
L-Lysine	840mg
L-Methionine	360mg
L-Threonine	350mg
L-Valine	11.7mg
L-cysteine- HCl	200 mg
L-glutamine	400 mg

Table 6-Composition of vitamins solution to add to Sicard's defined medium

C. Vitamins	**Quantities (mg)**
Biotin	0.030mg
Choline	10mg
Nicotinamide	1.2mg
Pantothenate	4.8mg
Pyridoxal HCL	1.2mg
Riboflavine	0.6mg
Thiamine	1.2mg

Table 7- Other components necessary to prepare Sicard's defined medium.

D. Other	Quantities (mg)
Pyruvate, sodium	1.6g
Uracil	2mg

The salt solution (A) was prepared first followed by the adjustment of the pH to 7.55. Then, all amino acids (except glutamine and cysteine) and uracil were dissolved in this balanced salt solution (A). The salt solution was then autoclaved. The vitamin solution (C) was prepared in a final volume of 10 ml. To the vitamin solution (C) glutamine, cysteine and sodium pyruvate were added. The vitamin solution was then filter sterilised. When required, 1% PGM (porcine gastric mucin) was added to the medium.

Blood agar base was made according the manufacturer's instructions, sterilised and allowed to cool until 45 °C, before addition of 5% v/v horse blood (Oxoid). Brain Heart Infusion was also prepared according to the manufacturer's instructions. Where necessary media were supplemented with appropriate antibiotics at the following concentrations: 100 μg/ml ampicillin (Sigma, UK), 100 μg/ml spectinomycin (Sigma, UK) and 500 μg/ml of kanamycin (Sigma, UK). 1 mM IPTG was also added when required.

B.3 Methods

I. *In vitro* mariner mutagenesis **(Akerley *et al.*, 1998)**

All knock out mutants studied in this project were generated by *in vitro* mariner mutagenesis and were prepared prior to the commencement of the work for this thesis by Dr Hasan Yesilkaya.

In order to produce the knock out mutants the target genes were amplified by PCR. The products were then incubated with transposase *Himar I* (Lampe *et al.*, 1996) and with plasmid pR412 which contains the mariner minitransposon conferring spectinomycin resistance. The other components of the reaction are described in Table 8. The reaction was incubated for 6 h at 30°C in the heating block. After the incubation, the DNA was recovered using a QIAquick spin column (Qiagen, UK). The DNA was finally eluted in 30 µl of TE buffer.

Table 8- Components necessary for the transposition reaction.

Components	**Volume (µl)**
Transposase *Himar 1*	0.5
10 X transposition buffer	2.0
glycerol	2.0
Acetylated BSA 10 mg/ml	0.5
PCR Product	5.0
Donor plasmid pR412 (spectinomycin resistant)	5.0
Nanopure H_2O	6.0
Total Volume	20.0

I.1 Gaps Repairing (Akerley *et al.*, 1998)

The gaps repair in the reaction mixture was setup as shown in Table 9. The reaction mixture was incubated for 30 min at 16 °C and then stopped by heating at 75 °C for 10 min.

Table 9- Components of the repair gaps reaction

Components	Volume (µl)
NEB buffer 2	3.5
BSA (NEB)	0.15
dNTP's (2mM)	1.5
T4 DNA polymerase (3U/µl)	1.5
DNA eluate (from Section I)	30
Total	36.15

I.2 Ligation

The ligation reaction was done by adding the items described in Table 10 to the above reaction (Table 9). The reaction was incubated overnight at 16 °C. The ligation was terminated by heating for 10 min at 75 °C. The product of the transposition reaction was purified to remove salts using a QIAquick spin column as described by the manufacturer (Qiagen, UK).

Table 10 - Components of the ligation reaction added to the gaps repair reaction.

Components	Volume (µl)
10X *E .coli* DNA ligase buffer	4
E. coli DNA ligase	2

I.3 Transformation of *Streptococcus pneumoniae* (Bricker & Camilli, 1999)

S. pneumoniae D39 was grown overnight in BHI at 37 °C. In the morning the cultures were diluted 1:50 and 1:100 in 10 ml of fresh BHI broth. The bacterial suspension was incubated at 37 °C until the OD_{595nm} was 0.03-0.05. At this stage 860 µl were taken and put into a clear microcentrifuge tube and 100 µl 100 mM NaOH, 10 µl 20% (w/v) BSA, 10 µl 100mM $CaCl_2$ and 2 µl 50 ng/ml CSP1 (Alloing *et al.*, 1996) were added. The reaction was then incubated at 37°C, statically for 3 hours. 330 µl were plated onto BA plates supplemented with100 µg/ml spectinomycin, every hour over the 3 incubation hours. The plates were incubated overnight in a CO_2 jar at 37 °C.

II. Growth curves

100 µl glycerol stock of *S. pneumoniae* D39 were inoculated into 10 ml of the Sicard's (SICARD, 1964) defined medium with glucose, or with mucin replacing glucose, and 1% vitamins. The cultures were grown over 14 hours, taking 20 µl every 2 hours to assess the growth by means of c.f.u counts (Miles & Misra, 1938). All the procedure was done in triplicates starting from three independent stocks.

II.1 Miles and Misra colony forming unit counts (Miles & Misra, 1938)

The Miles and Misra method is a technique to count the number of colony forming units in a bacterial suspension.

20 µl of *S. pneumoniae* bacterial suspension were inoculated in 180 µl of PBS in a 96 well microtitre plate, followed by serial dilution. These dilutions were then platted onto BA plates supplemented with antibiotic when necessary. The BA plates were previously divided into 6 sections. Three drops of 10 µl each were then plated from

each serial dilution in each one of these sections. The plates were allowed to dry and incubated overnight in a CO_2 jar. In the following day the plates were counted. The counted sectors are the ones that have between 30-300 colonies. The number of colony forming units per ml is calculated using the following formula:

$$\frac{n.cA + n.c.B + n.cC}{30} \times section\ dilution \times 1000$$

Where n.cA, B and C are the number of colonies counted from each drop, 30 is the volume plated.

III. Cell free extract preparation in BHI and mucin

100 µl of frozen glycerol stock of bacteria were used to inoculate 10 ml of BHI broth (Oxoid, UK). The cultures were grown until late exponential phase (OD600nm ≈ 0.8). Then, the cultures were centrifuged at 2360 g, for 10 min at 4 °C. The supernatant was removed and the pellet was resuspended in 1 ml ice cold 50 mM Tris-HCl buffer, pH 7.5. Following this the pellets were freeze thawed, and then sonicated (Yesilkaya *et al.*, 2000a) with five pulses of 15 seconds each with 30 seconds rest between each burst. Finally the cell lysates were aliquoted and frozen at –80 °C to be used when necessary.

For the preparation of mucin-derived cell free extracts, the same protocol was followed, but when the BHI was discarded the cultures were exposed to mucin for 6 hours. After these 6 hours the mucin was discarded and the pellets were resuspended in 1ml ice cold 50 mM Tris-HCl buffer, pH 7.5. The protocol continued as described above.

IV. Enzyme assays

IV.1 pH optima determination

The pH optimum was determined by assaying the cell free extracts in different buffers with different pH. 0.2 M Sodium citrate buffer ranging from pH 3 to 6 with increments of 0.5, sodium phosphate buffer ranging from pH 6 to 8 with 0.5 increments and Tris-HCl for pH 7.5 to 8.9. The reaction mixture to determine presence or absence of activity at a specific pH was prepared according the composition presented on Table 11.

Table 11- Composition of the reaction mixture of the enzymatic assay performed to determine pH optima.

	Volume (µl)	
Components	Reaction	Negative Control
Buffer 0.2 M (pH range)	35	45
pNP-linked substrate	10	10
Distilled water	45	45
Cell lysate	10	-

The reaction mixtures were incubated at 37 °C for 1 to 5 hours depending on the enzyme being assayed. After the incubation period, the reaction was stopped by addition of 100 µl of ice cold 1M sodium bicarbonate buffer pH 10.2. The samples were then centrifuged for 1 min 13000 rpm, to remove the cell debris. 180 µl of the reaction was then transferred to a 96 well microtitre plate and the absorbance was read at 415 nm in Bio-Rad plate reader model 680. The enzymatic activity was calculated by using the Lambert-Beer Law ($A= \varepsilon * b*c$, where $\varepsilon = 18300\ M^{-1}cm^{-1}$, b = 0.5 cm, A=

Absorbance reading and c is the unknown concentration). 1U is defined as 1 mM of product formed per minute.

IV.2 Activity assay using the synthetic substrate pNP-β-galactopyranoside and pNP-N-acetyl-β-D-glucosaminide.

The assay was performed following Miller's method (Miller, 1972). The reaction mixture was setup according to what is described in Table 12. The enzymatic reaction was incubated for 5 hours at 37 ºC and it was stopped by the addition of 100 µl of ice cold 1M sodium bicarbonate buffer pH 10.2. The samples were then centrifuged down in a Micro centaur bench top microcentrifuge for 1 min at 13000 rpm, to remove the cell debris. 180 µl were transferred to a 96 well microtitre plate and the absorbance was read at 415 nm in Bio-Rad plate reader model 680.

Table 12- Composition of the enzymatic reaction for determination of absence or presence of activity against pNP-linked substrates.

	Volume (µl)	
Components	Reaction	Negative Control
0.2 M Phosphate buffer pH 6.5	35	45
10mM pNP-β-galactopyranoside	10	10
Distilled water	45	45
Cell lysate	10	-

The reaction composition when using substrate pNP-N-acetyl-β-D-glucosaminide was exactly the same with the exception for the buffer used which was 0.2 M citrate buffer pH 5.0.

V. RNA extraction (Stewart *et al.*, 2002)

An overnight culture of *S. pneumoniae* was prepared as described in section II The following morning, fresh BHI was inoculated with enough amount of the overnight culture to give an $OD_{500\ nm}$ of 0.1. The culture was monitored until the $OD_{500\ nm}$ reached 0.6 (late exponential phase) at this stage. 40 ml of GTC solution (5M guanidine isothiocyanate, 0.5% sodium-N-lauryl sarcosine, 25 mM tri-sodium citrate pH 7.0, 0.1M 2-mercaptoethanol or DTT, 0.5 % (v/v) tween 80) were added to 10 ml of culture and mixed rapidly by swirling. The solution was then centrifuged at 5000 g for 20 min. The supernatant was discarded and the pellet was resuspended in 1 ml of GTC solution. Next, the suspension was transferred to a microcentrifuge tube and centrifuged at 12000 g for 30 sec. The supernatant was discarded and the pellet resuspended in 1.2 ml of RNazol (Sigma, UK). The suspension was immediately transferred to a ribolyser blue matrix tube and processed in the Ribolyser Hybaid apparatus at 6.5 power for 45 sec. The tube was then left at room temperature for 5 min. Then, 200 µl of chloroform was added and the mixture was vortexed for 15 sec. The tube was then left at room temperature for 3 min, followed by centrifugation at 12000 g for 15 min at 4°C. The upper phase was transferred to a clear tube. An equal volume of chloroform was added to the sample and mixed by vortex for 15 sec. The sample was again centrifuged at 12000 g for 15 min, 4 °C and the upper phase was once more transferred to a clear tube. 500 µl of isopropanol was added to the sample for DNA precipitation. The sample was left at room temperature for 15 min, followed by centrifugation at 12000 g for 10min at 4 °C. Finally all the supernatant was removed and the pellet was resuspended in 87.5 µl of nanopure water RNase free.

The procedure was then continued by the RNA clean up, consisting in eliminating the DNA from the sample. 2.5 µl of DNAse A and 10 µl of DNAse A (Qiagen, UK) buffer were added and the sample was left at room temperature for 15 min being subsequently purified with an RNeasy Mini Kit (Qiagen UK). The RNA concentration was measured by absorbance reading at 260 nm knowing that an OD_{260nm} of 1 corresponds to 40 µg/ml of RNA.

V.1 Extraction of pneumococcal RNA from infected tissues.

Outbred 9-week-old female MF1 mice (Harlan Olac, Bicester, UK) were infected intranasally with 50 µl PBS containing 1 x 10^6 passaged D39 pneumococci (Yesilkaya *et al.*, 2000a, Yesilkaya *et al.*, 2006). When the mice became severely lethargic they were anesthetised and the blood was collected by cardiac puncture. After sacrificing the animals by cervical dislocation, the lungs and nasopharynx were removed and homogenized on ice in 10 ml of sterile PBS using a tissue homogenizer. To separate pneumococci from host cells, lung homogenates and blood samples were centrifuged at 900 *g* for 6 min at 4 ºC. Afterwards the supernatants were subsequently centrifuged at 15,500 *g* for 2 min at 4 ºC, and the bacterial pellet was stored at -80ºC until further processing. Prior to pelleting, 20 µl of homogenate was removed, serially diluted in PBS and plated onto blood agar base (BA) agar in order to enumerate pneumococci and to exclude the presence of contaminating microflora. RNA extraction and purification were done as described in the previous section V.

V.2 Synthesis of complementary DNA

The first strand cDNA synthesis was performed on approximately 1 µg of DNAse-treated total RNA, immediately after isolation, using 200 U of SuperScript II reverse

transcriptase (Invitrogen, Paisley, UK) at 42 ^{o}C for 55 min, and random hexamers (Yesilkaya *et al.*, 2008) cDNA (2 µl) was amplified in 20 µl reaction volume that contained 1 x SYBR Green PCR master mix (Applied Biosystems, Foster City, USA) and 3 pmol of each primer (indicated with 'RTF' or 'RTR' tags). The transcription level of specific genes was normalised to *gyrB* transcription, amplified in parallel with SPD0709F (5' TCGTGTGGCTGCCAAGCGTG 3') and SP0709R (5' GGCTGATCCACCAGCTGAGTC 3') primers. The PCR conditions consisted of: 10 min at 95 °C, 30 seconds at 95 °C, 1 min at 55 °C and 30 sec at 72 °C, for 40 cycles. The results were analysed by the comparative C_T method (Livak & Schmittgen, 2001).

VI. Pneumococcal chromosomal DNA preparation (SAITO & MIURA, 1963)

Pneumococcal chromosomal DNA extraction was performed according to the method of Saito and Maiura (1963). 200 µl of overnight culture of *S. pneumoniae* strain D39 were used to inoculate 10 ml of BHI and this was grown at 37 °C for 6- 8 hours statically and in a CO_2 jar. The cells were pelleted at 2000 g for 10 min at 4°C. The pellet was resuspended in 200 µl TE buffer pH 8.0 containing 25% w/v sucrose. 30 µl of 500 mM EDTA, 20 µl of SDS and 1µl of proteinase K (12.5 mg/ml) were also added to the bacterial cell suspension. Cells were incubated at 37 °C overnight to obtain a clear cell lysate. The cell debris was removed by centrifugation in a Micro centaur bench top microcentrifuge at 6000 rpm for 5 min. The supernatant was transferred to a clear tube and an equal volume of liquefied phenol washed in Tris buffer pH 7.6 (Fisher Scientific) was added to the mixture and gently mixed until an emulsion formed, this was followed by centrifugation in a Micro centaur bench top micro centrifuge at 6000 rpm for 10 min. The upper aqueous phase was transferred to a fresh tube without

disturbing the white protein layer formed in between the two phases. An equal volume of chlorophorm:iso-amylalchool (24:1) was added and centrifuged as before. This procedure was repeated several times until no protein residue was seen at the interface. The upper phase was retained and mixed with 2.5 volumes of ethanol 100% and 0.1 volume 3M sodium acetate pH 5.2. The sample was then centrifuged in a Micro centaur bench top microcentrifuge at 13000 rpm for 15 min. The supernatant was discarded and the pellet washed with 500 µl of ethanol 70 % (v/v). The pellet was then resuspended in 50 µl miliQ water. The DNA concentration was measured by reading the absorbance at 260 nm knowing that an OD_{260nm} of 1 corresponds to 50 µg/ml of DNA.

VII. PCR amplification (Mullis *et al.*, 1986)

The PCR reaction began with a 95° C heating, to separate the double strand DNA in two individual single strands, then the annealing allowed the hybridization of the specific primers with their complementary sequences in the target DNA single strand, according to the base - pairing rules. The incorporation of the dNTP's to the 3' end of each primer attached to the DNA strand by action of the *Taq polymerase* occured at 72°C (activation temperature of the enzyme). This was followed by the extension, the last step, where in each cycle two new strands complementary to the initial single DNA strand are synthesized. Therefore, in the end of the polymerase chain reaction, the amount of copies of DNA amplified is exponentially duplicated. Table 13 shows the composition of the PCR reaction performed in a DNA Engine DYAD TM, Peltier Thermal Cycler.

Table 13- Composition of the PCR reaction.

PCR mixture		
Components	Sample	Negative control
HotStar Taq ®Plus Master Mix 2x (QIAGEN)	10 µl	10 µl
Primer mix	2 µl	2 µl x 3 (each primer)
Pneumococcal DNA sample (12.5 ng/µl)	2 µl	----------
Nanopure H_2O	6 µl	4 µl
Total volume	20 µl	20 µl

In this study the PCR conditions were set, like as follows, denaturing: 45 min at 95º C; annealing: 1 min at 55ºC; extension: 2 min at 72º C;

At the end of the 35 cycles of denaturing, annealing and extension, the samples remained for an additional 3 min at 72 ºC. Finally the PCR products were maintained at 4 ºC. This technique was used to amplify genes *SPD0247*, *SPD0065* and *SPD0853* from *S. pneumoniae*, to subsequently clone the three amplified gene products. The primers used are described in Table 14.

Table 14- In fusion primer sequences used for PCR amplification of the genes of interest. The area highlighted in bold indicates the 15 base pairs of homology with the vector.

Primers	Sequence 5' to 3'	Origin
SPD0065IFF	**TAAGGCCTCTGTCGAC**ACACGATTTGAGATACGAGA	This study
SPD0065IFR	**CAGAATTCGCAAGCTT**TTAAAATCCATTATTATCGC	This study
SPD0247IFF	**TAAGGCCTCTGTCGAC**CTAAGATTTCCAAAGGATT	This study
SPD0247IFR	**CAGAATTCGCAAGCTT**TTAAAATCCATTATTATCGC	This study
SPD0853IFF	**TAAGGCCTCTGTCGAC**AATTTAGGAGAATTTTGGTAC	This Study
SPD853IFR	**CAGAATTCGCAAGCTT**CTAATCTTTGCCACCTAGCT	This study

After PCR cycling was complete, the products were analysed by agarose electrophoresis gel to confirm the presence of a single band of DNA product, and to estimate the concentration of the PCR product. The amount of DNA was quantified by comparing band intensity against a molecular weight marker ladder that was run in the same gel.

VII.1 Purification and extraction of the PCR products

All the buffers used in the PCR products purification and extraction were provided by the QIAquick Gel Extraction Kit. The PCR products were extracted and purified using the QIAquick Gel Extraction Kit (QIAGEN, UK). The protocol is described below.

The DNA fragments corresponding to the genes of interest were excised from the agarose gel with a clean, sharp scalpel. The slices were weighed in microcentrifuge tubes, and 3 volumes of Buffer QG were added, considering 1 volume of gel 100 mg to be 100 µl. To dissolve the gel slices, the tubes were incubated in a 50º C water bath for 10 min, and to help dissolving every 2 min were mixed by vortexing. After well dissolved, the colour was checked to ensure the right pH for the DNA adsorption by the QIAquick membrane. The colour yellow showed that the pH was 7.5, the desired value to ensure an efficient DNA adsorption by the membrane. One volume of isopropanol was added to the sample mix, to increase the yield of DNA fragments.

The sample preparations were applied to the QIAquick column and centrifuged for 1 min at 13000 rpm. After discarding the flow-through, 0.5 ml of Buffer QG was added and the tubes centrifuged at 13000 rpm for 1 min, to remove all traces of agarose gel. To wash the membranes, 0.75 ml of Buffer PE was added to the QIAquick column which was left for 2 min and then centrifuged for 1 min at 13000 rpm. After discarding the flow-through the column was centrifuged for one additional minute to remove completely the ethanol 70%. Finally the DNA was eluted with 50 µl of Buffer EB, letting stand in the column 1 min before centrifugation to increase the efficient elution of bounded DNA. Eluted PCR products were kept at 4 ºC in the freezer, for further use.

The confirmation and quantification of the purification and extraction of the PCR products was done by size comparison with the molecular weight ladders run in the same electrophoresis agarose gel.

VIII. Cloning reaction

The technique used in this particular part of the work consisted in using the In-Fusion Dry Down PCR Cloning Kit (Clontech, Japan). This is designed to join various pieces of DNA that have 15 base pair homology at their linear ends. The advantage of this technique is the unnecessary cut of the vector or insert with restriction digestion, nevertheless, in this case both vector and PCR product were cut. The first step in this was the design of the primers (Table 14) which have to contain 15 base pair homology with the sequences flanking the insertion area in the cloning vector, this is followed by PCR amplification of the gene in study with those same primers. The PCR product is then combined with the linearised vector in the microcentrifuge tube for the cloning reaction.

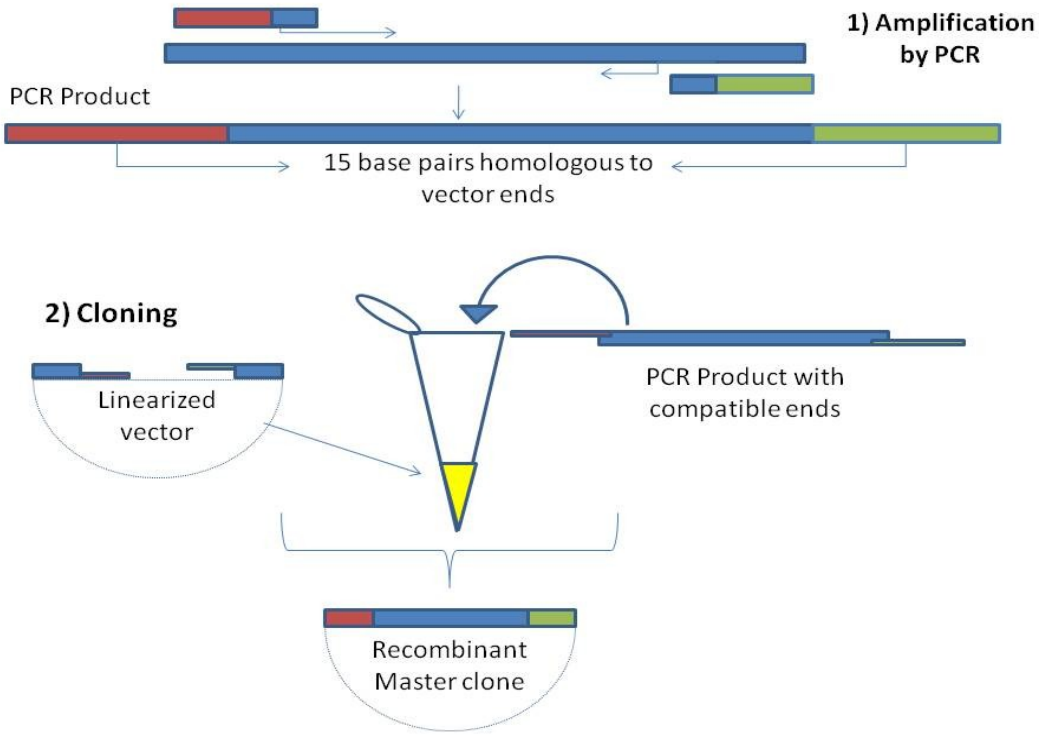


Figure 1- Cloning procedure when using the In-Fusion™ Dry-Down PCR Cloning Kit.

The circular plasmid p*Ecoli* - Nterm 6X HN was cut by two restriction enzymes, *Sal I* and *Hind III*. The same enzymes also cut the PCR fragments for *SPD0065* and *SPD0247* and *SPD0853*, since the restriction sites were introduced when designing the primers.

To set up the cloning procedure, the previously linearised vector plus insert were added to the In-fusion dry down mixture which contains ligase and ligation buffer, described in Table 15. After pipetting up and down to mix, the ligation reaction was incubated at 42º C for 30 min. The reaction was then kept for 5 minutes on ice, followed by dilution with 40 µl of TE buffer (10 mM Tris-HCl + 1 mM EDTA, pH 8.0).

Table 15 – Ligation reaction composition.

Reaction No.	**1**	**2**	**3**
Purified PCR insert (DNA target)	SPD0247 2 µl	SPD0065 1.5 µl	SPD0853 2 µl
p*Ecoli* - Nterm 6X HN vector	1.5 µl	1.5 µl	1.5 µl
nH_2O	6.5 µl	7 µl	6.5 µl
Add to In-Fusion Dry-Down pellet	yes	yes	yes

TE buffer pH 8.0 composition was performed as described in the In-Fusion™ Dry-Down PCR Cloning Kit User Manual.

VIII.1 Transformation of Fusion –Blue cells

The transformation procedure was done following the protocol described on the Fusion-Blue Competent Cells Manual (Clontech).

One vial of frozen Fusion-Blue™ cells was thawed on ice. Then, from the diluted ligation reaction (from section VIII.) 3 µl were pipetted into one vial tube. After mixing gently to ensure even distribution of DNA, the preparation was incubated on ice for 30 min. The transformation tube was then immersed in a water bath at 42º C for 45 sec, and immediately transferred into ice for 2 min. Then, 450 µl of SOC (2% (w/v) Bacto Tryptone, 0.5% (w/v) Bacto Yeast Extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 20mM glucose) medium was added to the cells, and the tube placed in the incubator at 37° C for 60 min with 250 rpm shaking to promoter recovery. After this step, dilutions of 1:4 and 1:2 were prepared by addition of SOC medium and spread in LA plates supplemented with ampicillin (100µg/ml).

The remaining cells, from the same transformation reaction from where the dilutions were made, were centrifuged during 1 min at 13000 rpm in a Micro centaur bench top centrifuge, and the pellet resuspended in 100 µl of medium supernatant, this was the neat culture which has subsequently been spread on LA plates supplemented with ampicillin (100µl/ml).

All plates were incubated at 37 ºC overnight. In the morning, twenty large and defined colonies were streaked into new LA plates supplemented with ampicillin (100 µg/ml), for plasmid preparation.

VIII.2 Extraction of plasmid DNA (Mini Prep)

All the buffers used in this protocol were provided by the QIAprep® Miniprep Kit from QIAGEN. The plasmid DNA extraction was made following the QIAquick mini - prep protocol from QIAGEN, as described below.

From the freshly streaked selective plates, one single colony was picked, and used to inoculate a 10 ml LB supplemented with ampicillin (100µg /ml). The culture was incubated at 37° C with vigorous shaking overnight. In the morning, 5ml of culture was transferred into a new universal tube, and the bacterial cells were harvested by centrifugation at 3082 g during 10 min at room temperature. The pellet was then resuspended in 250 µl Buffer P1*, and transferred to a micro-centrifuge tube (1.5 ml). Then 250 µl of Buffer P2* was added and the preparation mixed by inverting the tube gently, 4 to 6 times (without vortexing because this could shear the genomic DNA), to give a homogeneously coloured suspension.

To precipitate DNA, 350 µl of Buffer N3* was added, followed by immediately mixing the tube by inversion 4 to 6 times. The preparation became cloudy. The suspension was then centrifuged for 10 min at 13000 rpm in a Micro centaur bench top centrifuge, and the supernatant transferred to the QIAprep column, by decanting and pipetting the traces. The tube was once again centrifuged for 1 min at 13000 rpm in a Micro centaur bench top centrifuge and the flow-through discarded. To remove any trace of nucleases, 0.5 ml of Buffer PB* was added to the column, followed by centrifugation at 13000 rpm in a Micro centaur bench top centrifuge for 1 min, and discarding the flow-through after. Then the column was washed by adding 0.75 ml of Buffer PE* and centrifuging twice, discarding the flow-through between and after each centrifugation. These steps eliminated residual buffers and ethanol that could interfere with the subsequent enzymatic reactions. Finally the column was placed into a new microcentrifuge tube and to elute DNA 50 µl of Buffer EB* was added and left to stand for 1 min at room temperature. This was followed by centrifugation in a Micro centaur bench top for 1 min at 13000 rpm, and stored in the freezer (-20°C) for further use.

VIII.3 Restriction Enzymes Digestion

Restriction digestion is a method of cutting plasmid DNA by the use of enzymes (endonucleases) that recognise specific sequences in the DNA, usually palindromic sequences. In a general manner a restriction digestion is set up as described on Table 16, followed by incubation for 3 hours in a 37º C water bath, with the tubes sealed with nescofilm. Once the colonies were confirmed as being positive the genes were sent for sequencing.

Table 16-Restriction digestion reaction composition and quantities added in the microcentrifuge tube.

Restriction reaction		
Components	**Sample**	**Control**
Miniprep plasmid DNA	6 µl (Plasmid + insert)	6 µl (Empty plasmid)
Enzyme	1 µl	----------
Enzyme buffer NEB	2 µl	2µl
Nanopure water	10 µl	12 µl
Total reaction volume:	20 µl	20 µl

The restriction enzymes used were *Hind III* and *SalI* (New England Biolabs, UK) and the enzyme buffer was number 2, which is compatible for both enzymes. Table 17 shows the recognition sequences of both enzymes.

Table 17-*HindIII* and *SalI* characteristics: recognition sequences, cutting site, New England Biolabs correspondent buffer and working temperature.

Enzyme	Sequence	Buffer	Temperature
Hind III	AAGCTT	2	37° C
Sal I	GTCGAC	3	37 °C

VIII.4 Transformation of *E. coli* BL21 DE3

Transformation of expression bacterial cells *E.coli* BL21 DE3 was carried out as instructed by the manufacturer.

Two aliquots of competent cells BL21 DE3 were thawed on ice and mixed gently in order to ensure that the cells were evenly suspended. 1 µl of plasmid DNA was added directly onto the cells and gently stirred. The tubes were then put on ice for 5 min, followed by a heat shock for 30 sec at 42 °C. Following this procedure the tubes were put on ice for 2 min.

200 µl of SOC medium was added to each vial of bacteria and incubated at 37 °C for 1 hour with 250 rpm shacking to promote the recovery of bacteria. Finally the bacteria were plated into LA supplemented with 100 µg/ml of ampicillin and incubated at 37 °C overnight.

VIII.5 Expression of recombinant proteins

One colony carrying the desired construct was inoculated into 10 ml of LB supplemented with 100 µg/ml ampicillin. The culture was grown overnight at 37 ºC with 250 rpm shaking. In the next morning, dilutions of the overnight culture were

made into fresh media supplemented with 100 μg/ml of ampicillin on a factor of 1/20and 1/50 in a final volume of 50ml. The cultures were incubated at 25 ºC with shaking at 250 rpm till the $OD_{500\ nm}$ was approximately 0.6- 0.8.Two sets of cultures were incubated (set A and set B), one set that would be induced and a second set that would not be induced, functioning as a negative control for expression When the OD was reached, 1 ml of the cultures was taken and it was labelled "before induction". The remaining cultures from set A were induced by addition of 1ml IPTG for a final concentration of 1mM and the growth continued to be monitored for 5 hours in both set A and set B, collecting 1ml of each set of cultures every hour over 5 hours. After five hours all the samples were processed by centrifuging at 3000 g for 15 min at 4 ºC. The supernatant was discarded and the pellets frozen at –80 ºC for further analysis.

VIII.6 Analysis of total expression

The frozen pellets obtained from the 1 ml collections (section VIII.5) were defrosted on ice and resuspended in $1/10^{th}$ of the original culture volume using PBS. 5 μl of these pellets were collected and an equal amount of SDS Loading buffer was added and mixed well. Since the samples were very viscous they were centrifuged for 30 sec, followed by addition of 1 Kunitz of DNAse. The samples were then heated at 95 ºC for 5 minutes, centrifuged briefly at 5000 rpm in a Micro centaur bench top centrifuge before analysis by SDS PAGE.

VIII.7 Preparation of cell pellets for determination of solubility

The cell pellets obtained from section VIII.5 were resuspended in 1X equilibration/wash buffer (50 mM Sodium phosphate, 300 mM NaCl, pH 7.0) to give $1/10^{th}$ of the original volume of the culture. 0.75 mg/ml of lysozyme was added in order

facilitate the lysis process and the samples were left on ice for 30 min in an attempt to minimise proteolytic activity from other sources The samples were then sonicated on three pulses of 10 seconds with 30 seconds pause in between each pulse. The cell extract was centrifuged 10 –12.000 g for 20 min at 4°C to pellet the insoluble material. The supernatant was transferred to a clear microcentrifuge tube (clarified sample). The remaining pellet containing the insoluble material was resuspended in the original volume. 5 µl were collected and an equal amount of SDS sample buffer was added. The resuspended pellet was sonicated for 30 s, heated at 95 °C for 3 min and centrifuged briefly at 5000 rpm in a Micro centaur bench top before analysis by SDS PAGE.

IX. Purification of recombinant proteins

IX.1 Sample preparation to isolate native proteins

The cell pellets obtained from the expression procedure (section VIII.5) were harvested by centrifugation at 3000 g for 15 min at 4 ºC and the supernatant was removed. The cell pellet was then resuspended in 2 ml of chilled 1X equilibration /wash buffer (50 mM Sodium phosphate, 300 mM NaCl, pH 7.0) containing 1mM EDTA. Lysozyme at a final concentration of 0.75 mg/ml was added. The suspension was left at 4 ºC for 45 min. After, the sample was sonicated with three pulses of 10 seconds each with 30 seconds pause between. The cell extract was centrifuged at 10 - 12000 g for 20 min at 4 ºC to pellet any insoluble material. The supernatant was carefully transferred to a clean tube (clarified sample).

IX.2 Purification Protocol

The purification procedure started with the thorough resuspension of the TALON Metal Affinity Resin (Clontech, catalogue no. 635501) by vortexing. Immediately after, 8 ml of resin were transferred to a tube that was able to accommodate 10-20 times of the resin bed volume (2 ml of resin originate 1 ml of bed volume). The tube was then centrifuged at 700 g for 2 min to pellet the resin. The supernatant was discarded. Afterwards, 10 bed volumes of 1X Equilibration/wash buffer (50 mM Sodium phosphate, 300 mM NaCl, pH 7.0) were added and the suspension was mixed briefly by vortexing to pre-equilibrate the resin. The resin was then centrifuge at 700 g for 2 min. The supernatant was discarded. The two last steps were repeated twice. Following, the clarified protein obtained from section IX.1, was added to the resin. This mixture was agitated at room temperature for 20 min in order to allow the poly-histidine tagged protein to bind the resin. After, the mixture was centrifuged again at 700 g for 5 min and the supernatant was discarded. The resin was then washed by adding 20 bed volumes of 1X Equilibration Wash buffer. The suspension was gently agitated at room temperature for 10 min on a platform shaker. Following centrifugation at 700 g for 5 min was done and the supernatant was discarded. The last three steps were repeated once. 1 bed volume of 1X Equilibration/wash buffer (50 mM Sodium phosphate, 300 mM NaCl, pH 7.0) was added and the resin was resuspended. The mixture was transferred to the column with the end cap on and allowed to settle. The end cap was removed and the buffer drained from the column. The column was washed with 5 bed volumes of 1X Equilibration wash buffer (50 mM Sodium phosphate, 300 mM NaCl, 20 mM imidazole pH 7.0). Finally the sample was eluted by adding 5 bed volumes of

1X elution buffer (50mM Sodium phosphate, 300mM NaCl,150mM imidazole pH 7.0) to the column. The fractions were collected in 500 µl volumes.

IX.3 Quantification of protein using Bradford Assay (Bradford, 1976)

The Bradford assay was performed following the protocol provided with the Bradford reagent from Bio-Rad, UK. It was assayed as micro procedure.

The procedure started with the preparation of the BSA (Bovine Serum Albumin) standards. Standards ranging between 0 mg/ml to 0.8mg/ml were prepared and assayed. 10 µl of each standard was pipetted into a 96 well microtitre plate, also 10 µl of each of the unknown proteins were pipetted into the same microtitre plate followed by addition of 200 µl of Bradford reagent in working concentration (diluted 1:5) The reaction incubated 5 min at room temperature. Finally the OD was read at 595 nm in a Bio-Rad microplate reader Model 680.

The standards were plotted and the concentration of protein determined by using the equation from the trendline adjusted to the points obtained by plotting the standards.

IX.4 SDS PAGE

A Sodium dodecyl sulphate gel (SDS PAGE) was necessary both for the analysis of expression of recombinant proteins and also in order to perform the western blot. The gel was made according to what is described in Table 18 and 19.

Table 18 - Composition of the 10% SDS PAGE gel.

SDS gel 10 %	
Components	**Quantities**
Distilled water	4.0 ml
30% acrilamide bisacrilamide	3.3 ml
1.5 M pH 8.8 Tris HCl	2.5 ml
10 % SDS	0.1 ml
10% APS	0.1 ml
Temed	0.004 ml

The 5 % gel was made following what is described in Table 19.

Table 19- Composition of the 5% SDS-PAGE GEL.

SDS gel 5 %	
Components	**Quantities**
Distilled water	3.4 ml
30% acrilamide bisacrilamide	0.83 ml
1.5 M pH 8.8 Tris HCl	0.63 ml
10 % SDS	0.05 ml
10% APS	0.05 ml
Temed	0.005 ml

X. Substrate specificity- Endpoint Experiments

In order to determine the substrate specificity of each one of the recombinant proteins, endpoint experiments were done. The enzymes were tested against 8 different synthetic substrates which were:

- PNP-β-galactopyranoside
- PNP-α-galactopyranoside
- N-acetyl-β-D-glucosaminide
- 4-Methylumbeliferyl-α-glucoside
- PNP-6-phospho-β -glucoside (Thompson *et al.*, 1997)
- PNP-α-mannoside (Thompson *et al.*, 1997)
- PNP-6-phospho- β-galactopyranoside (Thompson *et al.*, 1997)
- PNP-α-6-phospho-glucoside (Thompson *et al.*, 1997)

The endpoint experiments were performed with purified enzymes and their activities against the synthetic substrates were determined. The reaction composition is shown on Table 20.

Table 20- Description of the components of the assay. The samples were incubated 37ºC.

Substrate (μl) 10mM	**Buffer (μl) 10 mM, pH 6.5**	**Distilled water (μl)**	**Enzyme (μl)**
10	35	45	10

XI. Kinetic characterisation of the recombinant protein encoded by gene *SPD0065*

The gene *SPD0065* encodes for a β-galactosidase, therefore the kinetic characterisation of the recombinant protein was done. The synthetic substrate ONP-β-

galactopyranoside was used and two different conditions were tested; variation of enzyme concentration and variation of concentration of substrate.

For the enzyme variation it was assayed 5, 10 and 15 µl of enzyme were assayed from a fraction that contained 345 µg/ml of protein. So 1.7, 3.4 and 5.2 µg of protein was used in each assay. Each experiment was made in triplicates.

For the substrate variation, a substrate concentration range from 0 mM to 4.0 mM was used. This assay was performed after analysis of the previous results where it was concluded that 5.2 µg was the optimum amount of recombinant protein to use.

The reaction mixture was prepared as described in Table 21. The final volume of the reactions was 100 µl. The reactions were followed for 1 hour at 37 °C in a Bio-Rad plate reader model 680. The absorbance was read at 415nm every minute.

Table 21- Composition of the enzymatic reaction.

Substrate concentration (mM)	**Substrate (µl)**	**Buffer (µl)**	**Distilled water (µl)**	**Enzyme (µl)**
0	0	35	50	15
1.0	10	35	40	15
1.5	15	35	35	15
2.0	20	35	30	15
2.5	25	35	25	15
3.0	30	35	20	15
3.5	35	35	15	15
4.0	40	35	10	15

XII. Kinetic characterisation of the recombinant protein encoded by gene *SPD0247*

For the second recombinant enzyme a similar study was done. The gene *SPD0247* encodes for a hypothetical 6-phospho-β-glucosidase, therefore the kinetic characterisation of the recombinant protein was performed. The synthetic substrate PNP-6-phospho-β-glucoside was used and two conditions were tested; variation of enzyme concentration and variation of concentration of substrate

For the enzyme variation it was assayed 5, 10 and 15 μl of enzyme was assayed from a fraction that contained 242 μg/ml of protein. This translates into 1.2, 2.4 and 3.6 μg of protein in each assay. Each experiment was performed in triplicates. For the substrate variation, a range from 0 mM to 2.0 mM was used. The reaction mixture was prepared as described on Table 22; the final volume of the reactions was 100 μl. The reactions were followed for 1 hour at 37 °C on the Bio-Rad plate reader model 680, monitoring the assay every minute and registering the absorbance readings at 415nm. The reaction was stopped when the saturation phase was achieved.

Table 22- Composition of the enzymatic reaction

Substrate concentration (mM)	Substrate (μl)	50mM Tris-HCl pH 7.5	Enzyme (μl)
0	0	35	10
0.5	5	30	10
1.0	10	25	10
1.5	25	20	10
2.0	20	15	10

XII.1 Identification of linkage specificity

In order to identify the specificity of enzyme encoded by *SPD0065* in a natural substrate a digestion of desialylated fetuin was set up as specified in Table 23. The digestion was incubated overnight at 37°C. In the following day, the samples were ran in an SDS gel at 0.03 mA for 1 h 30 min and blotted onto the nitrocellulose membrane (Millipore, UK). Next, the membranes were incubated at 4°C overnight in blocking reagent diluted 1:10 in TBS (0.05 M Tris–HCl, 0.15 M NaCl, pH 7.5) (DIG Glycan Differentiation Kit, Roche, UK).

Table 23- Composition of the fetuin digestion reaction.

Concentration of fetuin mg/ml	**Volume of fetuin (µl)**	**Phosphate buffer pH 6.5**	**Enzyme (µg)**
0.01	1	98	1
0.02	2	97	1
0.04	4	95	1
0.08	8	91	1
0.12	12	87	1
0.15	15	84	1
Control 0.15	15	85	No enzyme

The membrane was then washed twice, for 10 min duration with 50 ml TBS (0.05 M Tris –HCl, 0.15 M NaCl, pH 7.5) and once with buffer 1 (0.05 M Tris –HCl, 0.15 M NaCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM $CaCl_2$, pH 7.5). The following step was the lectin incubation, where 10 µl DSA (datura stramonium agglutinin) which recognises Gal (1,4) GluNAc, and 100 µl of PNA (peanut agglutinin) which recognises core

disaccharide Gal(1,3) GalNAc were added to 10 ml of buffer 1 and the membrane was incubated with this solution for 1 h. The blot was then washed 3 times for 10 min each with 50 ml TBS buffer. Following, the membrane was incubated for 1 h with the anti-Digoxigenin-AP (10 µl of conjugate were added to 10 ml of TBS). The blot was then washed 3 times for 10 min each in 50 ml of TBS. Finally, the staining reaction was performed by adding 10 ml buffer 2 (0.1 M Tris-HCl, 0.05 M $MgCl_2$) containing 200 µl of NBT/BCIP. The membrane was immersed in this solution until the grey bands started to develop. To stop the reaction the membrane was washed abundantly with nanopure water.

XIII. Analysis of glycoprotein degradation

The following procedure was done by Dr Karen Homer King's College London, Dental Institute in collaboration with our group in the University of Leicester (Terra *et al.*, 2009).

The pneumococcal strains were grown in modified Sicard's medium supplemented with mucin for 48 h. After the incubation period, cells were pelleted by centrifugation at 1,700 *g* for 10 min and the supernatants were retained and stored at -80ºC until required. Controls for the analysis of monosaccharide removal from mucin due to the action of pneumococcal glycosidases comprised uninoculated, mucin-supplemented medium. Low-molecular weight medium components were removed from mucin-containing supernates by exhaustive dialysis against distilled water. Residual sialic acids (N-acetylneuraminic and N-acetylglycolylneuraminic acids) and neutral/amino sugars (galactose, fucose, N-acetylglucosamine and N-acetylgalactosamine) were released by treatment with 40 mM and 4 M trifluoroacetic acid (Pierce, UK),

respectively, essentially as previously described (Byers *et al.*, 1999). Released sialic acids and neutral/amino sugars were analysed in separate runs by high-pH anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using a DX500 HPLC system (Dionex, Camberley, Surrey, UK), as previously described (Byers *et al.*, 1999). Monosaccharide concentrations were calculated by comparison of peak areas with those of the authentic compounds (Sigma, UK) as standards, over the concentration range 0-50 μM. Monosaccharide concentrations of mucin after growth of pneumococcal strains were related to the protein concentrations of individual samples, to take into account sample losses or volume changes occurring during dialysis. Significant removal of mucin monosaccharides by *S. pneumoniae* strains was determined relative to uninoculated medium controls using Student's *t*-test ($p<0.05$).

The release of monosaccharides from PGM, fetuin and their desialylated forms by recombinant SPD0065 was also monitored by HPAEC-PAD, as described above.

XIV. Protection studies using recombinant proteins.

Three groups of MF1 mice (Harlan Olac, Bicester, UK) were given 25 μg doses of recombinant protein with 67 μl PBS and 33 μl Imject Alum adjuvant (Pierce, UK). The recombinant proteins and adjuvant were mixed at room temperature for 1 hour prior injection. Mice were immunised by intraperitoneal injection with 100 μl of the mixture.

The mice were re-injected every two weeks and blood was collected after a period of 8 weeks. The serum was prepared by leaving the blood at room temperature for one hour and was recovered by centrifugation at 5000 g for 10 min. The serum was subsequently stored at –80 °C until needed.

XV. Localisation assay - western Blot

In order to assess the cellular localisation of the proteins in study, a localisation assay was performed. The protocol was started with the cell lysates preparation. From the original cell lysate, cell fractions containing the membrane proteins, the cell wall proteins and the cytoplasmic proteins were prepared.

Six hours cultures of *S. pneumoniae* D39 were harvested by centrifugation, 2366 g for 15 min. Subsequently the pellet was resuspended in 1 ml of the supernatant and a centrifugation at maximum speed (13000 rpm in a Micro centaur bench top centrifuge, 10 min) was performed. The supernatant was removed and the pellet was resuspended in TE buffer +25% (w/v) sucrose + 12.5 mg/ml lysozyme. The suspensions were left to incubate overnight at 37 °C until a clear lysate was observed. Afterwards, the cell lysate was centrifuged at 2366 g for 5 min. The supernatant was transferred to a new tube; those were the cell wall proteins. To the remaining pellet 500 µl of PBS was added and the mixture cooled on ice for 10 min, followed by a sonication period of 3 pulses of 15 seconds with 30 second intervals in between each burst. The cell lysates were centrifuged for 10 min 3500 rpm in a Micro centaur bench top centrifuge. The supernatant containing the cytoplasmic proteins was transferred to a new tube. The remaining pellet contains the membrane proteins.

This procedure was followed by analysis by SDSPAGE electrophoreses gel (section IX.4) and subsequent western blot (section XV.1).

XV.1 Western blot

The western blot was done using an Electroblot (Bio-Rad). The filter papers and the sponges were equilibrated in transfer buffer (48 mM Tris Base, 39 mM Glycine, 20% (v/v) methanol, 0.037% (w/v) SDS) for 30 min before the starting of the transfer. Then the reaction was assembled according to the following scheme (Figure 2):

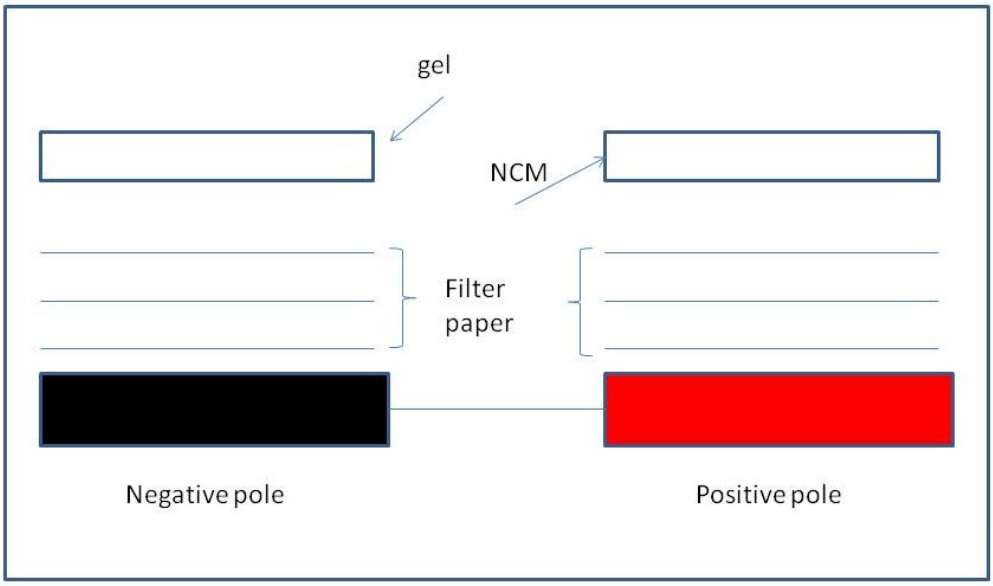


Figure 2- Schematic explanation of western blot assembly. The blue arrows are pointing towards the gel and the nitrocellulose membrane (NCM), the green olive arrow is pointing towards the filter paper and the red and black squares represent the cassette used to assemble the western blot. The black square is the negative pole and the red square is the positive pole.

The electroblot was performed for 1 h at 0.25 mA. The membrane was then blocked overnight in TBS 5 % (w/v) milk solution, on the next day the NCM membrane was washed thoroughly with TBST (100 mM Tris HCl, 0.15 M NaCl, 0.05 % Tween 20, pH 8.0) solution, followed by 2 hours incubation with the serum collected from the immunised animals (section XIV), which contained the primary antibody, in a proportion of 1:1000. The membrane was then, once again, thoroughly washed with TBST solution. The next step was the incubation with anti mouse IgG Fc specific-

alkaline phosphatase (Sigma, UK) 1:2000 for a period of 2 h. Once the incubation was finished the NCM membrane was washed three times with TBST and developed by the addition of 5 ml of BCIP/NBT to the membrane. When the bands were clear and visible the membrane was washed with distilled water to stop the reaction. The membrane was dried and scanned.

XVI. *In vivo* virulence studies

XVI.1 Stocks of pneumococci

New plates of blood agar base supplemented with 5% of horse blood were prepared. Bacteria were streaked for single colonies and optochin sensitivity was confirmed. The plates were incubated overnight in a CO_2 jar at 37°C. In the next day, a sweep of colonies was inoculated in 10 ml of fresh BHI. The culture was incubated statically overnight at 37 °C to $OD_{500\ nm}$ had reach 1.4-1.6. The cultures were then centrifuged at 3000 rpm for 15 min the supernatant was discarded and the pellet was resuspended in 1 ml of 80% BHI and 20% fetal calf serum filter sterilised. A part of this (700 µl) were added to fresh BHI serum broth and the OD_{500} was checked to confirm that it was 0.70. The cultures were incubated at 37 °C for 5 hours. Subsequently the OD was checked to see if it had reached 1.6, and 500 µl aliquots were prepared and frozen at -70 C. After 24 hours at -70 C the viability, the sensitivity and any contaminations were tested.

XVI.2 Animal passage of pneumococci

100 µl of bacterial suspension was injected into mice intraperitoneally using a 0.5 ml fine insulin syringe. After 22-28 hrs the animals were 2+ starry. These animals were anaesthetised with 5% (v/v) inhaled anaesthetic combined with 1.5 L O_2/min in an

anaesthetic box. The animals were then exsanguinated by cardio puncture with a 23G needle, followed by sacrifice of the animals by dislocation of the neck.

XVI.3 Survival experiments

Female MF1 outbred mice (Harlan Olac), were used for virulence testing. Standardised inoculum was prepared as described previously (Yesilkaya *et al.*, 2006, Yesilkaya *et al.*, 2000b). In order to determine virulence of the pneumococcal strains, mice were lightly anesthetized with 2.5% (v/v) fluothane (Zeneca, Macclesfiled, United Kingdom) over oxygen (1.5 to 2 litres/min). A 50 μl sample of PBS containing approximately 5 x 10^5 *S. pneumoniae* colony forming units was given via intranasal. The inoculum dose was confirmed by viable counting after platting on blood agar plates. Mice were monitored for disease signs of progressively starry coat, hunched and lethargic for 7 days, and those reaching the point of being severely lethargic were humanely sacrificed. The time point to this was considered survival time. To express the signs of disease numerically a score of 1 or 2, 3 or 4 and 5 or 6 were given got hunched, starry coat, or lethargic respectively.

In order to determine the bacteraemia in each mouse, approximately 20 μl of venous blood was taken by tail bleeding from the intranasal infected mice at predetermined time points after infection. Viable counts were determined by serial dilutions in sterile PBS and plating onto blood agar plates, supplemented with 5% (v/v) defibrinated horse blood with appropriate antibiotic. Survival times were analysed by the Man-Whitney U test.

XVI.4 Kinetic experiments

In vivo growth of bacterial strains was also determined in the nasopharynx, lungs and blood. To accomplish this goal, at predetermined interval times following intranasal infection, groups of mice were severely anesthetised as blood was collected by cardiac puncture, being subsequently sacrificed by cervical dislocation. The lungs and nasopharynx were transferred separately into 10 ml of sterile PBS, weighed and homogenised in a Stomacher-Lab blender (Seward Medical, London, United Kingdom) (Yesilkaya *et al.*, 2000b).Viable counts in homogenates and blood were performed as above. The data were analysed by analysis of variance followed by Benferroni post-test. Statistical significance was considered as p values of < 0.05.

XVII. Complementation of the mutants

In order to understand the potential effect of the knock out mutations, genetic complementation of the mutants was attempted using the technique described by Guiral *et al.*, (2006). This method uses a plasmid that is capable of replication in *E. coli*, but not in *S. pneumoniae,* pCEP has an expression/selection cassette that contains a maltose-inducible promoter, separated from a kanamycin-resistance gene by *Nco*I and *Bam*HI cloning sites. To facilitate homologous recombination, expression/selection the cassette is flanked by more than 2 kb of pneumococcal DNA. Therefore, pCEP allows insertion of the gene of interest under the control of a maltosaccharide-inducible promoter, in a chromosomally silent site, downstream of *amiA* operon (Alloing *et al.*, 1990).

XVII.1 Isolation of plasmid pCEP (miniprep)

The miniprep of the plasmid pCEP was performed following the protocol described in section VIII.2.

XVII.2Digestion of plasmid pCEP with *Bam*HI

The plasmid was digested according to the procedure described on section VIII.3, the particular digestion was set up as described on Table 24.

Table 24- Restriction digestion components

Reaction components	Reaction	Controls
pCEP	6 µl	6 µl
BamHI	1.5 µl	—
Buffer	3 µl	3µl
Distilled water	9.5 µl	11 µl

XVII.3Amplification of interest genes

The genes of interest were amplified using the following primers, designed to add *Bgl*II restriction sites to the PCR products to later digest with this enzyme. The plasmid was digested with *Bam*HI, but the inserts were digested with BglII which presents compatible ends with *Bam*HI. *Bgl*II digestion site was introduced in the amplicons because the genes contained *Bam*HI sites within their sequences. The primers used are presented on Table 25.

Table 25- Primers used on the amplification of the interest genes creating compatible ends with the plasmid pCEP.

Genes	Primers	Source/origin
SPD0065	GCGAGATCTTGATGACACGATTTGAGATACGAGAT (Forward Primer SPD0065CF)	Present study
	GCGAGATCTTGTCATAAGTTTTCCCCCTTTATATGTT (Reverse primer SPD0065CR)	Present study
SPD0247	GCGAGATCTTGATGCTAAGATTTCCAAAGGATTTTG (forward primer SPD0247CF)	Present study
	GCGAGATCTTGTTAAAATCCATTATTATCGCTTAATTC (reverse primer SPD0247CR)	Present study

XVII.4Digestion of the interest genes with *Bgl*II

The inserts were digested with the enzyme *Bgl*II which recognises de following sequence **AGATCT**. Using this approach the inserts have compatible ends with the plasmid. The digestion was set as follows on Table 26.

Table 26- Components of the digestion reaction performed for further ligation**.**

Components	Volumes (µl)
PCR product	15
*Bgl*II (NEB)	1.5
Buffer	3.0
Nanopure water	10.5

The ligation reaction was performed as described on Table 27.

Table 27- Components of the ligation reaction performed in the complementation attempts.

Components	Volumes (µl)
pCEP	5
Gene of interest	15
Buffer	4
Ligase	2
Nanopure water	14

XVII.5Transformation

The transformation of the mutants SPD0065M and SPD0247M in order to achieve complementation was done according to the method described in this Chapter section I.

XVII.6Confirmation of the transformation

In order to confirm the successful transformation the DNA of one single colony was extracted following the next protocol.

One colony was inoculated in BHI broth (OXOID) and let to grow overnight at 37 °C. In the morning 400 µl of the culture were transferred to a microcentrifuge tube. The culture was centrifuged at 13000 rpm for 1 min followed by discarding of the supernatant. The pellet was then resuspended in 250 µl of nanopure water. The suspension was boiled for 4 min and put on ice for 2 additional minutes. The final step was to spin down at 13000 rpm for 1 min. The DNA was then used in the PCR.

The successful introduction of the intact copy of the gene was confirmed by PCR using *malF* and pCEPR primers (Table 28) whose recognition sites are localized in immediately up and downstream of the cloning site, respectively. The complemented strain was designated SPD0065REV.

Table 28- Primers used for confirmation of complementation

Primers	Sequence
malF	GCTTGAAAAGGAGTATACTT
pCEPR	AGGAGACATTCCTTCCGTATC

XVIII. Construction of *SPD0065* knock-out mutation in different backgrounds

The SPD0065 homolog in strains strain IO11966 from serotype 6B and ST type 177, strain OXC1261 from serotype 19F was amplified by PCR using the primers SPD0065F and SPD0065R described in Table 29. The amplified region was transformed in strains IO11966 and OXC1261 following the protocol described in this Chapter section I.3. If an homologue of gene *SPD0065* was present the PCR product would recombine with the genome, being consequently knocked out by insertion of the spectinomycin cassette. The successful insertion of the mutation was confirmed by PCR.

Table 29-Primers used for amplification of gene *SPD0065*.

Primers	Sequence
SPD0065F	CTGCTAAACTGCCCAGAGATG
SPD0065R	GGATACGCAACGTAAGGGAA

XIX. Construction of double mutant SPD0562K /0065S

To construct the double galactosidase mutant in strain D39 a mutant carrying an insertion mutation in *SPD00562* was made following the method described in this Chapter section I, and I.2 with the exception that plasmid pR410 conferring kanamycin resistance was used. *SPD0065* gene already mutated by the insertion of a spectinomycin cassette was used as template for PCR amplification with primers SPD0065F and SPD0065R (Table 28). The amplified region was transformed following the procedure described in I.3. Transformants were selected for the presence of spectinomycin and kanamycin resistance.

XX. Biofilm studies

The ability of *S. pneumoniae* D39 and SPD00247M to adhere to plastic was investigated. Glycerol stocks of *S. pneumoniae* were inoculated (5 µl) in 195 µl of BHI into a micro titration plate with flat bottoms. The bacteria were incubated overnight at 37 °C in a CO_2 jar. On the following day, the cells in suspension that did not adhere were collected, serial diluted and plated for further enumeration. The wells were then washed with 200 µl BHI three times and one final volume of 100 µl of fresh BHI was added to the wells. The micro titration plate was then sonicated for 3 seconds. This allowed bacteria do be free in suspension. This suspension was then serial diluted and plated in BAB containing spectinomycin for SPD0247M and BAB only for the wild type D39. The cells that adhered were counted and normalised against the ones that did not. The data were analysed using Mann Whitney test.

For the assay where recombinant protein was added the procedure was exactly the same except that 2 µg of protein were added to the wells previous to 16 h incubation.

C RESULTS

I. Overview

Several putative glycosyl hydrolase knock-out mutants were generated, using the *in vitro* mariner mutagenesis technique. It was hypothesised that these genes would be involved in pneumococcal nutrition and in order to clarify this point pneumococci were grown in defined medium supplemented with mucin as the only source of carbon. In an attempt to better understand the role of each of these enzymes, cell lysates of the mutants were prepared and assayed for glycosyl activity using β and α-linked synthetic substrates. Following the results of growth in Sicard's defined medium supplemented with mucin, the expression of these genes *in vitro* and *in vivo* was also measured.

I.1 Selection of genes

The mutations were done based on the *in silico* analysis of the *S. pneumoniae* D39 genome. Putative sugar hydrolase genes were identified and it was aimed to produce knock-out mutants in these genes in order to assess their importance for the growth and virulence of the bacterium. The mutants had been produced by Dr Hasan Yesilkaya before commencement of work in this thesis (see table 3 of the Material and Methods Chapter).

I.2 Growth in Sicard's defined medium supplemented with mucin.

The first analysis after the construction of the glycosyl hydrolases knock-out mutants was the growth in defined medium with mucin as the only source of carbon.

The growth curves were performed in Sicard's defined medium supplemented with 1% (w/v) mucin (as described in the Material and Methods Chapter section II). For these assays the mucin suspension was left to settle and only the upper part of the mucin suspension, containing the mucin soluble part was used, this avoided the problems observed in prior experiments in which mucin aggregated with the bacteria making it impossible to quantify colony forming units.

Thirteen mutants and wild type D39 were tested. Knock-out mutant SPD0247M did not grow in Sicard's defined medium during a 14 hour experiment. The results obtained are shown in Figure **2**1, Figure 2 and Figure 3.

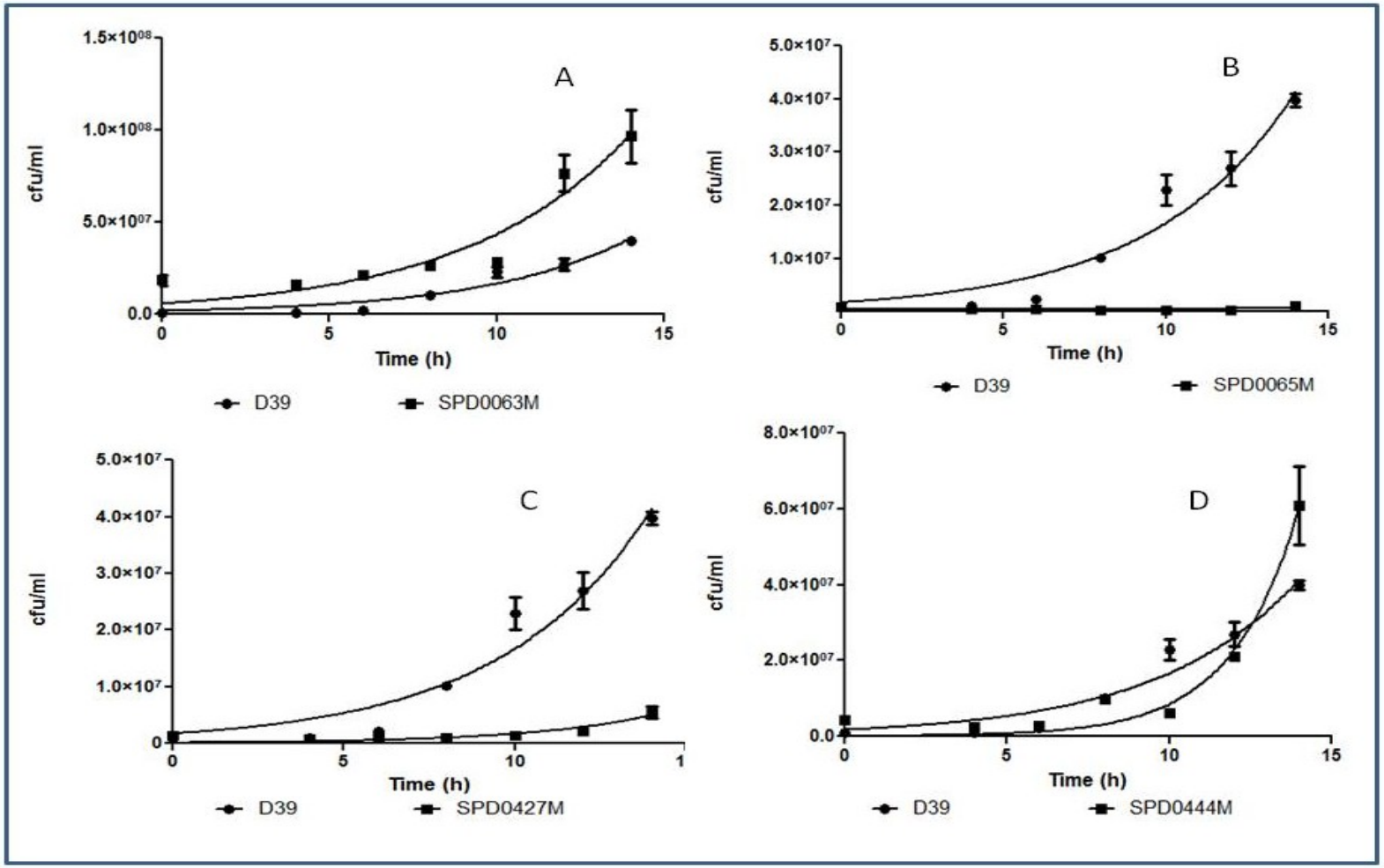


Figure 1- Growth curves performed in Sicard's defined medium supplemented with 1% (v/v) mucin as carbon source. The lines are a best fit adjusted to the experimental points. A- Mutant SPD0063M and wild type D39.B- Mutant SPD0065M and D39. C- Mutant SPD0427M and D39. D- Mutant SPD0444M and D39. Each point is the mean of three replicates and the vertical bars represent the standard error of mean.

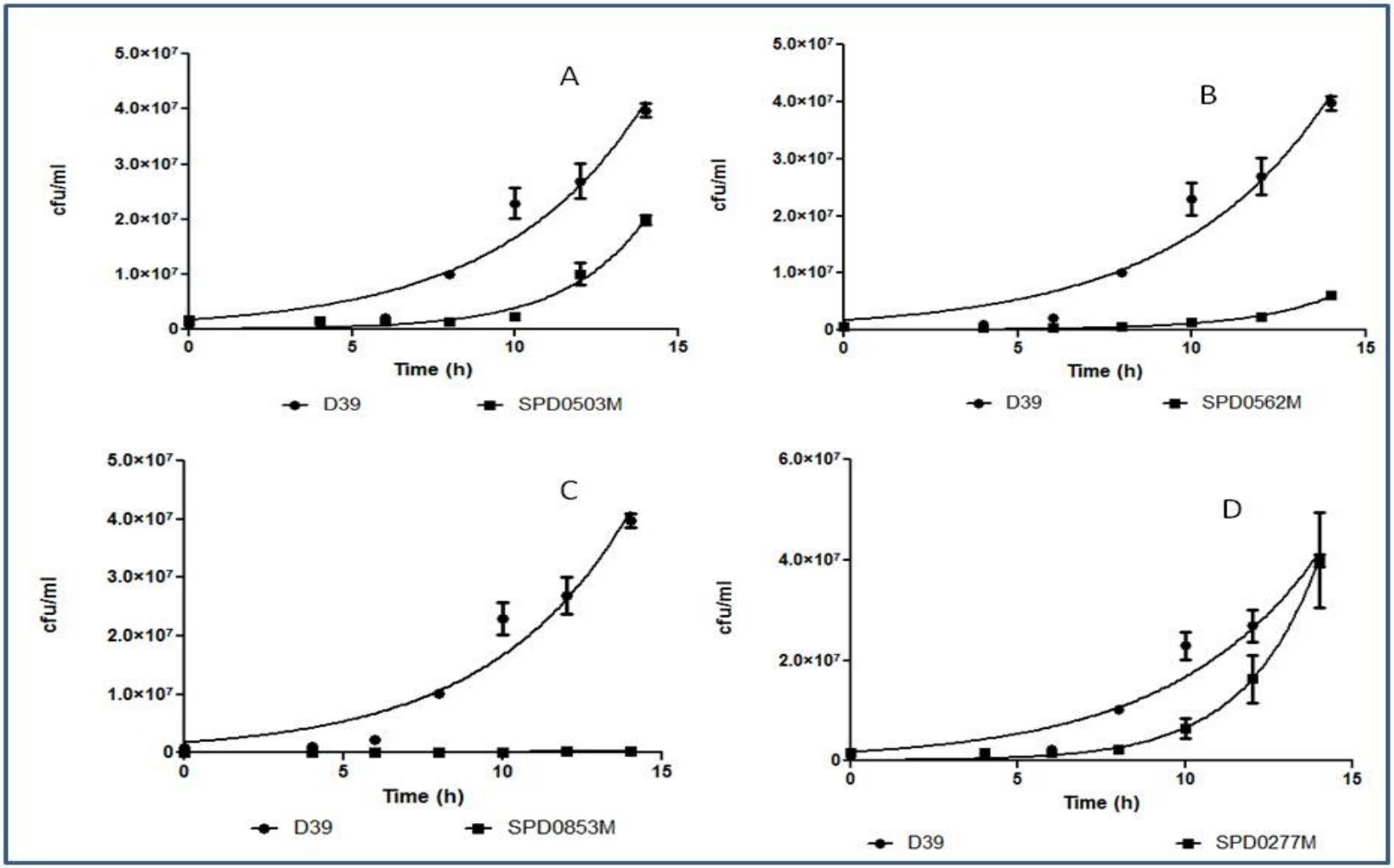


Figure 2 - Growth curves performed in Sicard's defined medium supplemented with 1% (v/v) mucin as carbon source. The lines are a best fit adjusted to the experimental points. A- Mutant SPD0503M and wild type D39.B- Mutant SPD0562M and D39. C- Mutant SPD0853M and D39. D- Mutant SPD0277M and D39. Each point is the mean of three replicates and the vertical bars represent the standard error of mean.

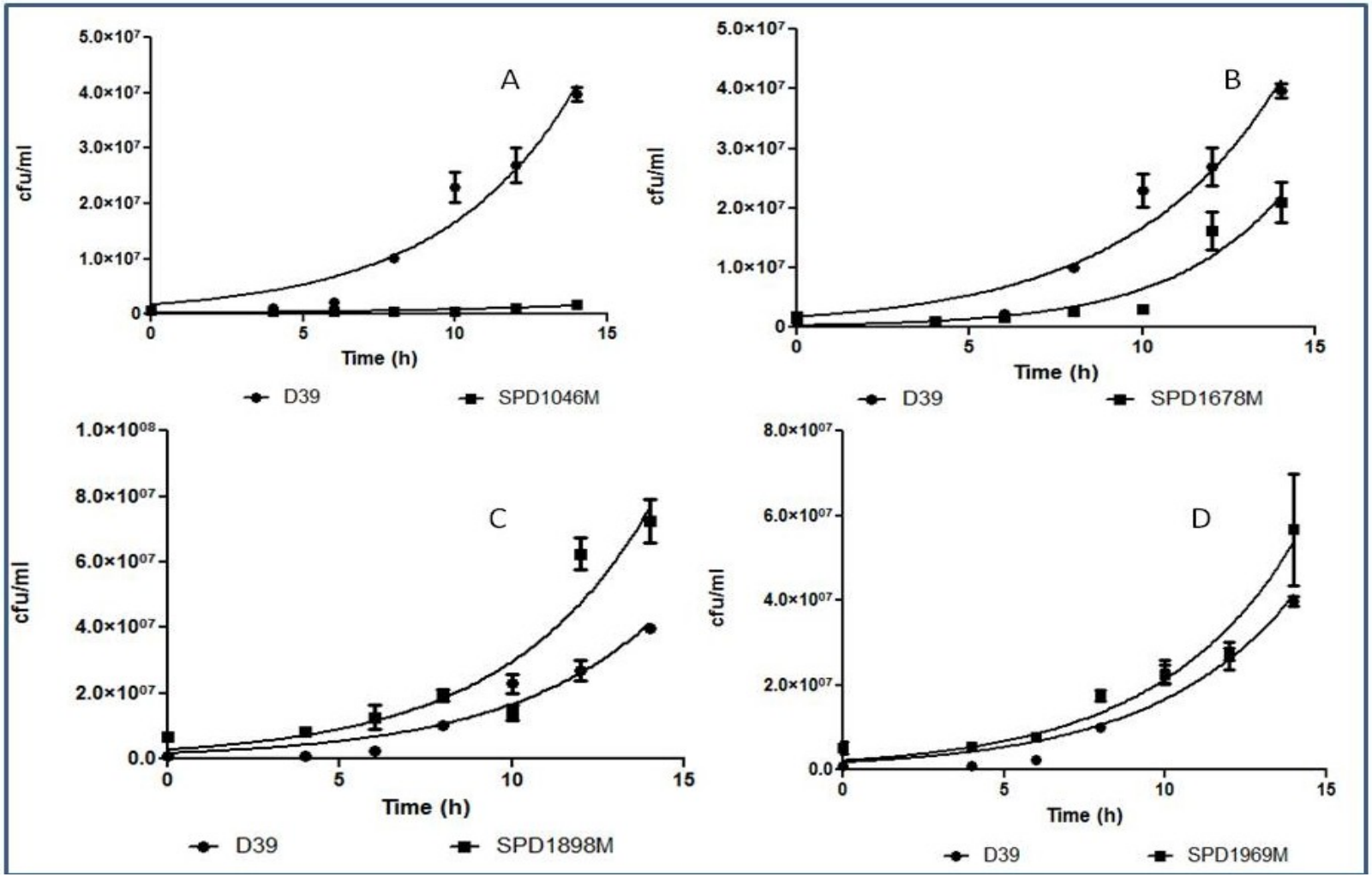


Figure 3- Growth curves performed in Sicard's defined medium supplemented with 1% (v/v) mucin as carbon source. The lines are a best fit adjusted to the experimental points. A- Mutant SPD1046M and wild type D39.B- Mutant SPD1678M and D39. C- Mutant SPD1898M and D39. D- Mutant SPD1969M and D39. Each point is the mean of three replicates and the vertical bars represent the standard error of mean.

The growth rates of the fourteen mutants are presented in Table 1.

Table 1- Growth rates obtained for the fourteen mutants and the wild type D39 in Sicard's defined medium supplemented with mucin. Values marked with (*) are significantly different from the wild type ($p < 0.05$). The data are the mean of three independent experiments.

Mutants	Growth rate h^{-1}	Standard deviation
D39	0.227	0.011
SPD0063M	0.203	0.033
SPD0065M	0.035*	0.027
SPD0247M	No growth	No growth
SPD0277M	0.384	0.105
SPD0427M	0.272	0.066
SPD0444M	0.487*	0.086
SPD0503M	0.413*	0.045
SPD0562M	0.399	0.088
SPD0853M	0.047*	0.001
SPD1046M	0.148	0.038
SPD1678M	0.312	0.048
SPD1898M	0.236	0.024
SPD1969M	0.235	0.059

After a lag phase of 6 hours the wild type D39 started to grow exponentially at a rate of 0.227 ± 0.011/h. In contrast with the parental strain the mutants SPD0065M and SPD0247M showed significantly ($p<0.05$) decreased growth rate in Sicard's defined medium supplemented with mucin as sole carbon source. SPD0853M also exhibit a decreased growth rate (0.047 ± 0.001) when grown in the presence of mucin. A 1.5 log cfu/ml increase in bacterial yield was observed for the wild type D39. On the other hand the yield for both mutants decreases over 12h. Mutants SPD0444M and SPD0503M showed an increase in growth rate when compared to the wild type (0.487± 0.086, 0.413±

0.045, 0.227/h ± 0.011 respectively, $p<0.05$). The other mutants SPD0063M, SPD0277M , SPD0427M, SPD0562M, SPD1046M, SPD1678M, SPD1898M and SPD1969M presented no significant difference in growth rate when compared to the wild type D39 ($p>0.05$).

The analysis of these results suggested that some glycosyl hydrolases seem to have more impact in mucin utilisation and /or degradation than others. These observations led to the investigation of the action of these enzymes and the role of these sugar hydrolases in carbohydrate utilisation. It was decided to assess the action against synthetic substrates of lysates of cells grown in BHI and in the presence of mucin.

I.3 Enzyme assays

Enzyme assays were carried out for better understanding the role of the sugar hydrolases in the degradation of polysaccharides. For this synthetic pNP-β-linked substrates were utilised. The optimum pH was first determined as described in the Material and Methods Chapter section IV.1.

I.4 pH Optima

The pH optimum was assayed using the lysate of D39 grown in BHI broth (see Material and Methods Chapter section III). This assay was designed in an attempt of determining the pH optimum for β-galactosidases. Figure 4 shows the activity against the synthetic substrate ONP-β-galactopyranoside between pH 3.0-8.9. As it can be seen in Figure 4 the highest activity was achieved at the pH range of 6.5-7.0 using phosphate buffer.

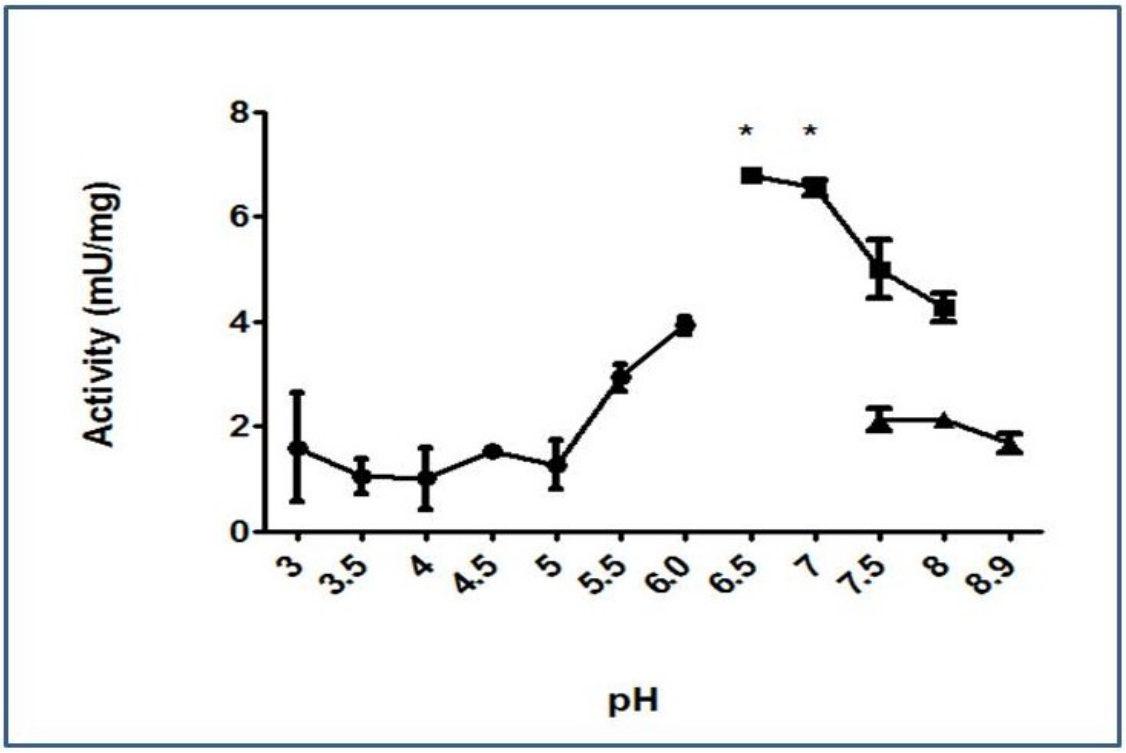


Figure 4– Activity of cell lysate of D39 using the synthetic substrate ONP-β-galactopyranoside, over a range of pH. Buffers used were: (●) Citrate buffer (pH 3-6), (■) Phosphate buffer (pH 6.5-8.0), and (▲) Tris-HCl Buffer (pH 7.5-8.9). For each pH value three independent measures were made. A significantly different activity value from all other values of pH was obtained for 6.5 and 7.0 ($p<0.05$).

In a similar way, the pH optimum for the enzyme N-acetyl-β-D-glucosaminidase was also assayed following the procedure described in the Material and Methods Chapter section IV.1. The pH optimum for the activity of the lysates against synthetic substrate pNP-N-acetyl-β-D-glucosaminide was in the pH range of 5- 5.5 (Figure 5).

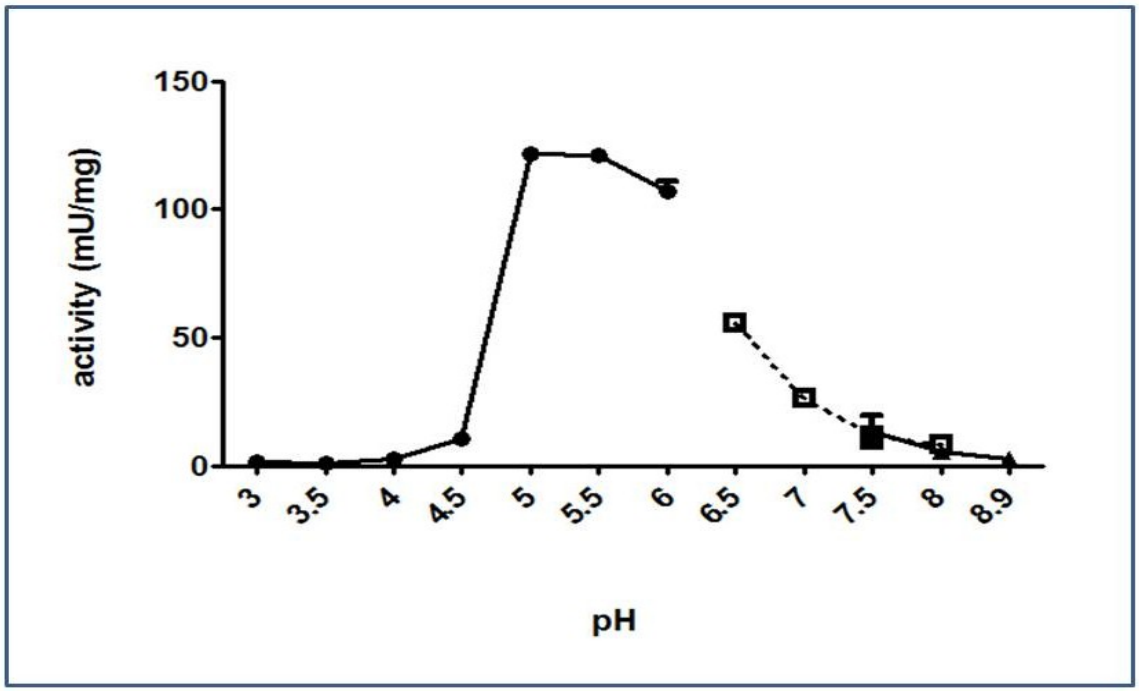


Figure 5 -Activity with cell free extract of D39 using the synthetic substrate pNP-N-acetyl-β-D-glucosaminide, over a range of pH. (●) Citrate buffer (pH 3-6); (■) Phosphate buffer (pH 6.5-8.0); and (▲) Tris-HCl Buffer pH (7.5A-8.9A). For each pH value three independent measures were made. 1 way ANOVA was performed, the activity values for all pH other than 5 and 5.5 are significantly different from the activity values obtained for 5 and 5.5 ($p<0.05$).

The pH optima assays determined the optimum pH of D39 cell free extracts against synthetic substrates ONP-β-galactopyranoside, and pNP-N-acetyl-β-D-glucosaminide. These results indicate an optimum pH for the enzymes β-galactosidase and N-acetyl-β-hexosaminidase. It is noteworthy that they have very different optimum values.

I.5 Determination of activity of the cell free extracts of knock-out mutants using synthetic substrates ONP- β-galactopyranoside and pNP-N-acetyl-β-D-glucosaminide

Once the optimum pH for the different synthetic substrates was determined, cell free extracts of pneumococcal mutants were assayed using ONP-β-galactopyranoside and pNP-N-acetyl-D-glucosaminide. This experiment was performed to understand the impact of the absence of each particular enzyme on the activity against the substrates. The cell extracts were prepared in BHI and in mucin, as described in the Material and Methods Chapter

section III, to determine if there was an effect of mucin on the total glycosyl hydrolases activity.

Figure 6 shows the activity obtained when using cell free extracts prepared from cells grown in BHI against the substrate ONP-β-galactopyranoside. Extracts of SPD0562M, SPD1678M, SPD1898M and SPD1969M were significantly different from D39 wild type cell free extract. However, as it can be seen in Figure 6 whereas mutant SPD0562M and SPD1969M showed a decrease in activity ($p<0.05$), mutants SPD1678M, SPD1898M showed a significant increase in activity ($p<0.05$) (Figure 6). For the remaining of the mutants there were no significant differences when compared to the wild type ($p>0.05$).

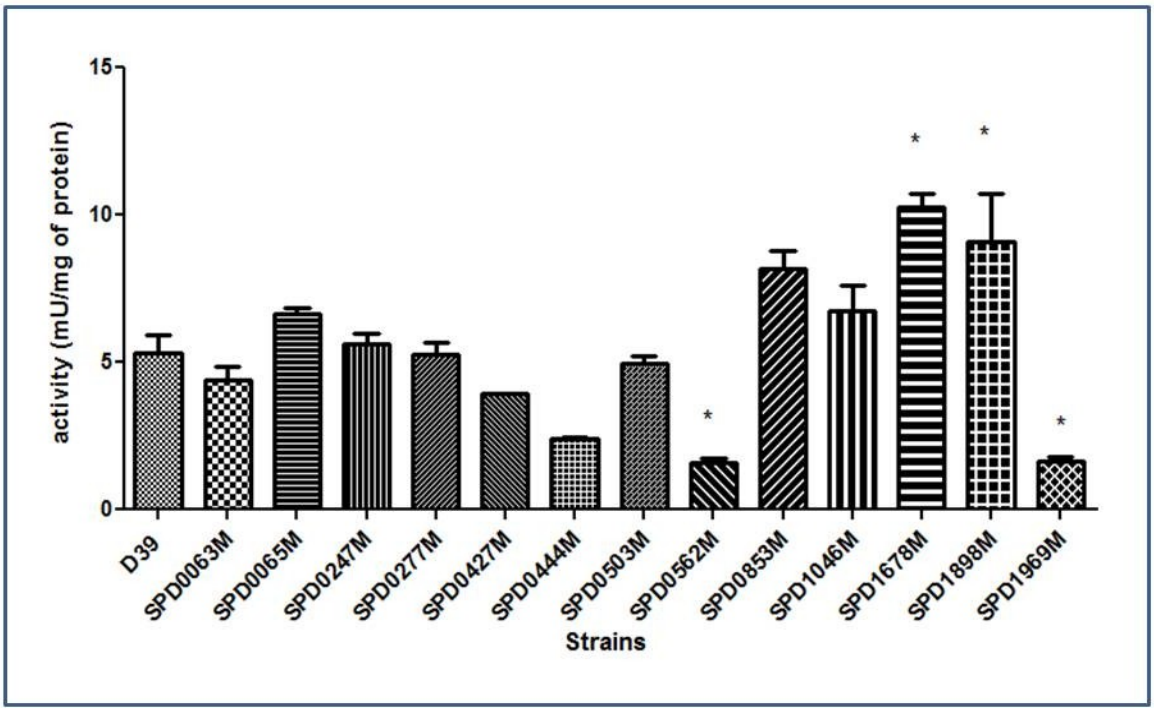


Figure 6- Activity of cell free extracts prepared from the knock-out mutants grown in BHI against the synthetic substrate ONP-β–galactopyranoside * represents $p< 0.05$ calculated using one way Anova and with a Bonferroni post test. The values are the mean of three independent replicates; the vertical bars correspond to the standard error of mean.

Figure 7 shows the activity obtained when using cell free extracts prepared from bacteria grown in BHI and exposed to mucin against the substrate ONP-β-galactopyranoside. SPD1969M was the only mutant to register an activity significantly

different from D39 wild type cell free extract ($p<0.05$). D39 had an activity of 17.7 ± 2.1 mU/mg of protein, SPD1969M showed an increase in activity to 42.2 ± 11.4 mU/mg of protein ($p<0.05$), but the remainder of the mutants were not significantly different from the wild type ($p>0.05$). The reduction observed for the knock-out mutants SPD0065M (10.3 ± 0.7 mU/mg of protein), and SPD0562M (1.6 ± 0.4 mU/mg of protein) in mucin was expected since the annotation for both of these genes was of a β-galactosidase.

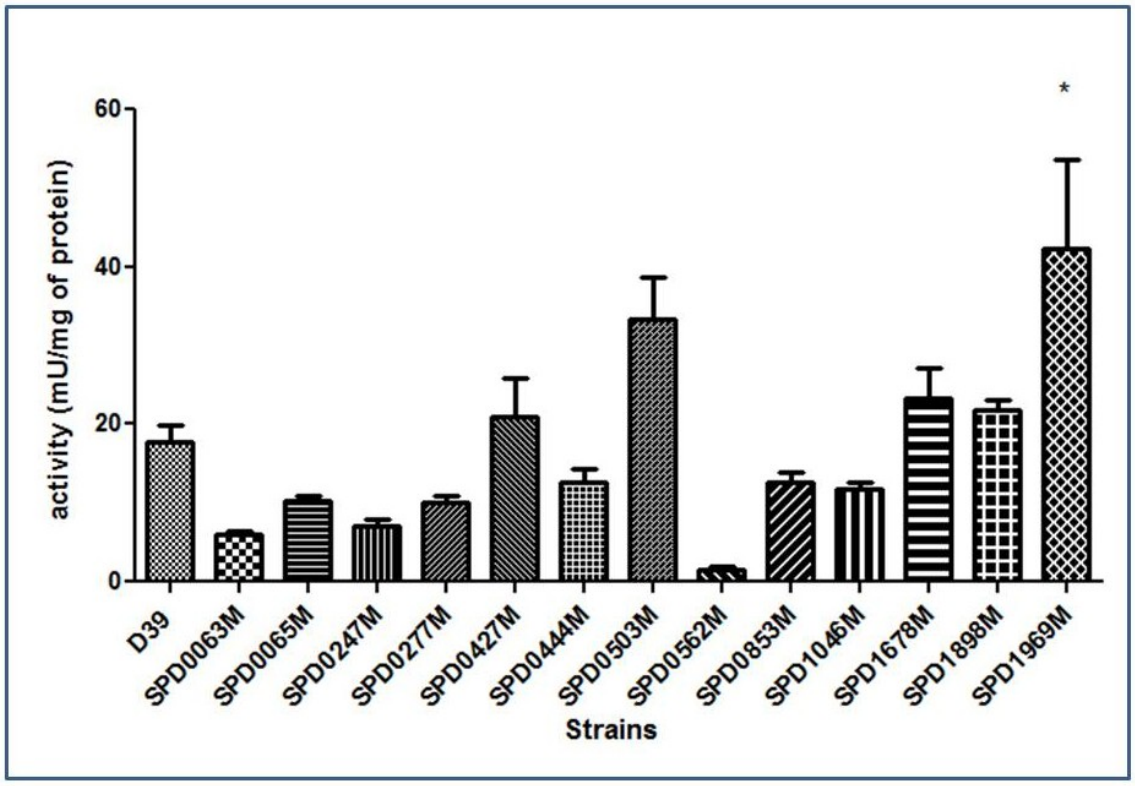


Figure 7- Activity of cell free extracts prepared from the knock-out mutants grown in BHI and exposed for 6 hours to mucin against the synthetic substrate ONP-β−galactopyranoside * represents $p< 0.05$ calculated using one way Anova and with a Bonferroni post test. The values are the mean of three independent replicates; the vertical bars correspond to the standard error of mean.

Figure 8 shows the comparison of the activities against ONP-β-galactopyranoside obtained with cell extracts prepared from pneumococci grown in BHI and mucin. The asterisk marks ONP-β-galactopyranoside activity that is significantly different in BHI-grown cell free extract compared to mucin-derived extract. For example, cell free extract of D39 grown in BHI had an activity of 5.3 ± 0.6 mU/mg of protein whereas it had 17.7±

2.1 mU/mg in mucin derived extract. Mutants SPD0427M, SPD0503M, SPD1678M, SPD1898M, SPD1969M showed significantly more activity in mucin derived extracts than in extracts prepared in BHI ($p<0.05$).

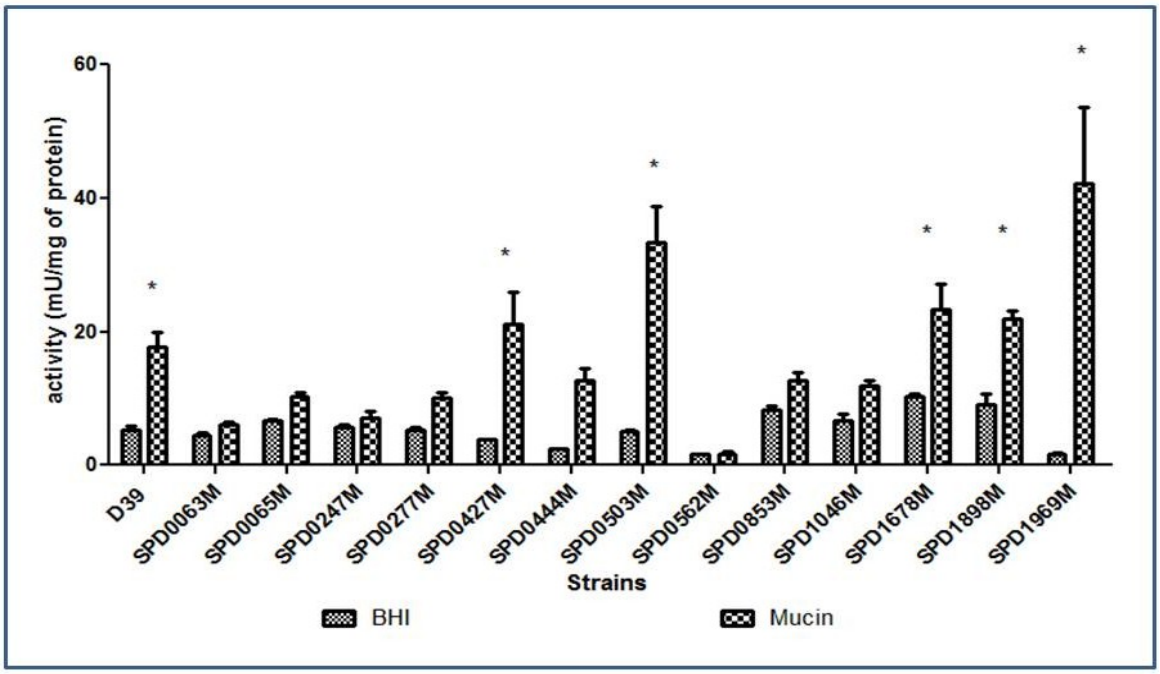


Figure 8 – Activity of cell free extracts prepared from the knock-out mutants grown in BHI and grown in BHI and exposed for 6 hours to mucin against the synthetic substrate ONP-β–galactopyranoside * represents $p< 0.05$ calculated using Two way Anova and with a Bonferroni post test. The values are the mean of three independent replicates; the vertical bars correspond to the standard error of mean.

Although, some cell free extracts prepared from mutant pneumococci, grown and exposed to mucin presented a lower activity when compared to the wild type grown in the same medium, when the cell free extract from the same mutants grown in different media were compared in relation to its activities an increase in activity was observed for the mucin derived extracts. For example, SPD0065M extract from cells grown in BHI had an activity of 6.6 ±0.2 mU/mg of protein whereas the SPD0065M extract from cells grown and exposed to mucin had an activity of 10.3 ± 0.4 mU/mg of protein ($p>0.05$). The same trend was observed for mutants SPD0427M ($p<0.05$), SPD0503M ($p<0.05$), SPD1678M ($p<0.05$), SPD1898M ($p<0.05$), SPD1969M ($p<0.05$), as well as for D39, for extracts from cells grown in BHI the activity was of 5.3 ± 0.6 mU/mg of protein and for the extract

prepared from cells grown and exposed to mucin the activity was of 17.7± 2.1mU/mg of protein, $p<0.05$ (Figure 8). These results strongly suggest that mucin is capable of inducing glycosidases.

Figure 9 shows the activity obtained when using cell free extracts prepared from pneumococcal cells grown in BHI against the substrate PNP-N-acetyl-β-D-glucosaminide to determine activity of possible glucosaminidases within the glycosidases mutants. Cell extracts of SPD0063M, SPD0444M and SPD1968M showed significantly reduced activity when compared to the activity obtained for D39.cell free extract. D39 had an activity of 122.4 ± 2.5 mU/mg of protein whereas mutant SPD0063M had an activity of 6.7 ± 0.9 mU/mg ($p < 0.05$). The reduction observed in the activity of this mutant's cell free extract was expected since this enzyme had been identified as a β-N-acetylhexosaminidase (King *et al.* 2006). Cell extracts of mutants SPD0065M ($p<0.05$), SPD0277M ($p<0.05$), SPD0427M ($p<0.05$) and SPD1898M ($p<0.05$) showed an increase in activity when compared to the activity obtained for D39 cell extract.

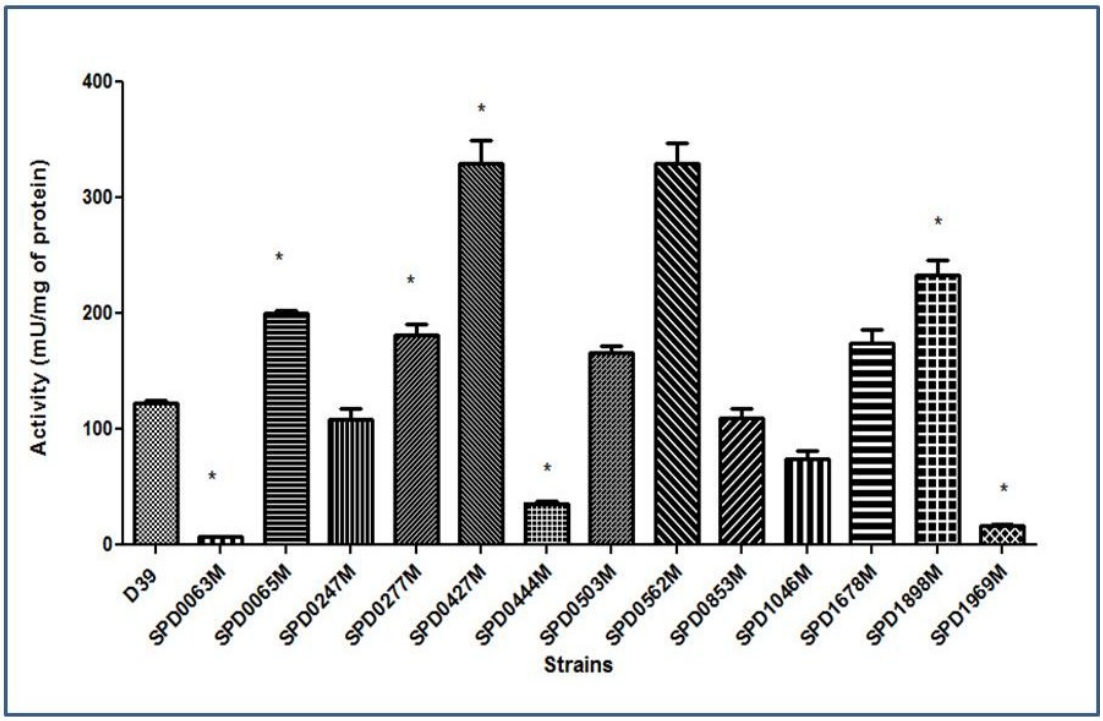


Figure 9- Activity for cell free extracts prepared from the knock-out mutants grown in BHI against the synthetic substrate PNP-N-acetyl-β-D-glucosaminide * represents a p value < 0.05 calculated using one way anova and with a Bonferroni post test. The values plotted are the mean of three independent replicates. The vertical bars indicate the standard error of mean.

Figure 10 shows the activity obtained when using cell free extracts prepared from pneumococcal cells grown in BHI and subsequently exposed to mucin against the synthetic substrate PNP-N-acetyl-β-D-glucosaminide SPD0503M and SPD1969M showed a significant increase in activity ($p<0.05$) when compared to the value of activity obtained for the mucin derived cell free extract of D39. As expected, SPD0063M had a lower activity when compared to the wild type, however, this was not significantly different ($p>0.05$). The remainder of the mutants did not register any significant differences in activity values when compared to the activity obtained for the mucin derived wild type cell extract ($p>0.05$).

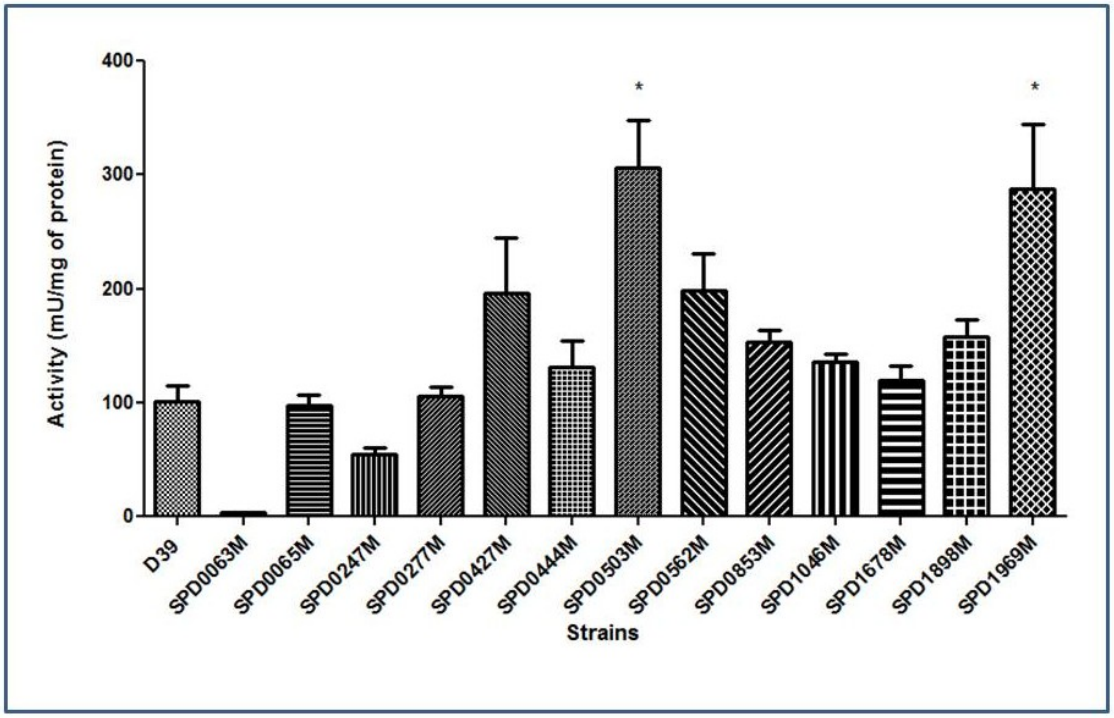


Figure 10- Activity for cell free extracts prepared from the knock-out mutants exposed to mucin when using the synthetic substrate PNP-N-acetyl-β-D-glucosaminide * represents a p value < 0.05 calculated using one way anova and with a Bonferroni post-test. The values plotted are the mean of three independent replicates.

Figure 11 shows the activity obtained when using cell free extracts prepared from *S.pneumoniae* D39 grown in both in BHI and in BHI with subsequent exposure to mucin against the synthetic substrate PNP-N-acetyl-β-D-glucosaminide. These results did not demonstrate a consistent pattern of induction of activity by mucin. Indeed higher activity was seen in BHI derived extracts for mutants SPD0065M, SPD0427M and SPD0562M ($p<0.05$). Whereas, mutants SPD0444M, SPD0503M and SPD1969M showed an increase of activity for the cell free extract prepared from cells grown in BHI with subsequent exposure to mucin ($p<0.05$).

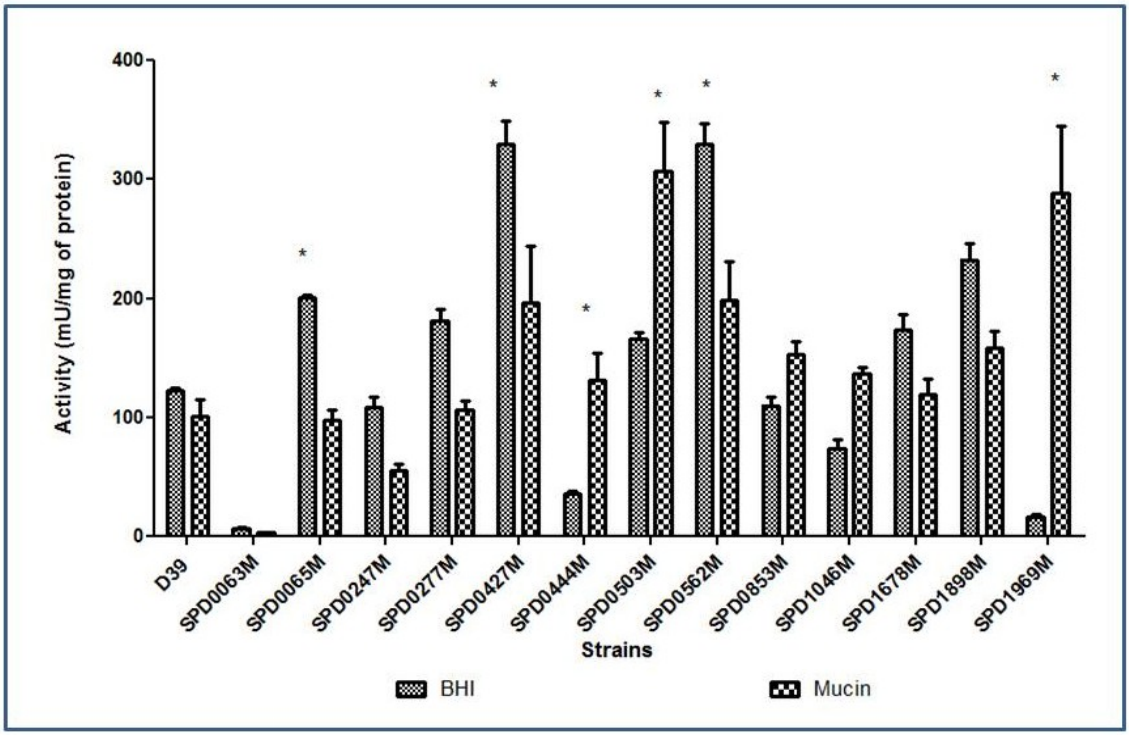


Figure 11- Activity for all cell free extracts prepared from the knock-out mutants when using the synthetic substrate PNP-N-acetyl-β-D-glucosaminide * represents a p value < 0.05, calculated using 2 way Anova and with a Bonferroni post test. The values plotted are the mean of three independent replicates. The vertical bars indicate the standard error of mean.

I.6 Assay of gene expression by quantitative Real Time PCR

Due to the previously observed phenotype (defective growth in Sicard's defined medium and reduced enzymatic activity), genes *SPD0065* and *SPD0247* were investigated for influence of mucin in their expression. For this, the expression of genes *SPD0065* and *SPD0247* was measured in D39. When compared to growth in glucose, in Sicard's defined medium supplemented with mucin the mRNA level for gene *SPD0065* increased 189 ± 11 fold, (n=3) whereas *SPD0247* increased 5 ± 0.2 fold, (n=3). In addition, the level of mRNA was also measured in tissues recovered from infected animals compared to bacteria grown *in vitro* in glucose-containing Sicard's medium. In bacteria recovered from the nasopharynx, *SPD0065* mRNA increased 2.2 ± 0.1 fold, (n=3) and *SPD0247* mRNA showed an increase of 8.3 ±.0.4 fold, (n=3). In bacteria recovered from the lungs, the levels of mRNA for *SPD0065* and *SPD0247* were up 35 ± 0.4 fold, (n=3) and 17 ± 0.7 fold, (n=3) respectively. The mRNA for these targets was also measured in the blood from infected animals. It was observed that expression went up 6.5 ± 1.1 fold, (n=3) for *SPD0065*, however there were no differences from the values obtained for gene *SPD0247* when compared to cells grown *in vitro* in Sicard's containing glucose. The increase in expression noted for these targets both in mucin and *in vivo*, suggests a role for these enzymes in glycoprotein degradation and/or utilisation.

II. Summary

Genes *SPD0247*, *SPD0065* and *SPD0853* were chosen for further analysis. Knock-out mutants SPD0065M, SPD0853M and SPD0244M presented defective or no growth in Sicard's defined medium supplemented with mucin as the only source of carbon. The cell free extracts of knock-out mutants SPD0065M and SPD0247M showed decreased enzymatic activity when tested against ONP-β-galactopyranoside. Furthermore expression of these genes was influenced by mucin.

Knock-out mutant SPD0853M was also carried on for the following stage of the work. This gene was picked for two reasons: firstly its poor growth in Sicard's defined medium supplemented with mucin and secondly, due to its annotation which was of an endo-β-N-acetylglucosaminidase, suggesting an involvement in glycoprotein degradation. The remainder of targets that presented different phenotypes from the one seen for D39 will be investigated in the future.

D RESULTS

I. Overview

After analysis of the results generated from the experiments performed with the mutant collection, three hydrolases were picked for further characterisation. In this chapter the work done in order to clone the three genes of interest is described. Additionally, the attempts at complementation of knock out mutants SPD0065M and SPD0247M are also described. Mutational work for target gene *SPD0065* was also attempted in different serotype backgrounds, namely serotype 6B and serotype 19.

I.1 Pneumococcal DNA extraction and purification

I.1.1 Agarose gel electrophoresis

The pneumococcal DNA was extracted following the protocol described on the Material and Methods Chapter, section VI. In order to analyse if the pneumococcal chromosomal DNA preparation was successful an electrophoresis agarose gel was ran. In Figure 1 a smear of DNA can be seen in lane 1 showing that pneumococcal chromosomal DNA preparation was successful. From the absorbance at 260nm the concentration was calculated as 587ng/µl.

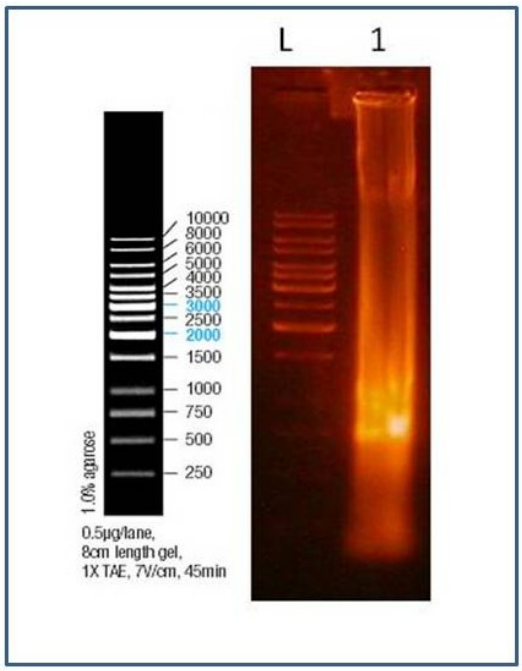


Figure 1- Agarose gel electrophoresis showing the isolated pneumococcal chromosomal DNA: L- GeneRuler™ 1 Kb DNA Ladder; Lane 1- D39 DNA sample from pneumococcal preparation.

I.2 Amplification of the genes *SPD0247*, *SPD0853* and *SPD0065* by PCR

With the objective of producing the recombinant proteins, the work was started with the amplification of the genes of interest by PCR. The results are shown in Figure 2. The amplified DNA was confirmed to have the expected size of 1380 (lane 1, *SPD0247*), 2166

(lane 2 *SPD0853*) and 1788 bp (lane 3 *SPD0065*) respectively. The PCR conditions are described in the Material and Methods Chapter, section VII.

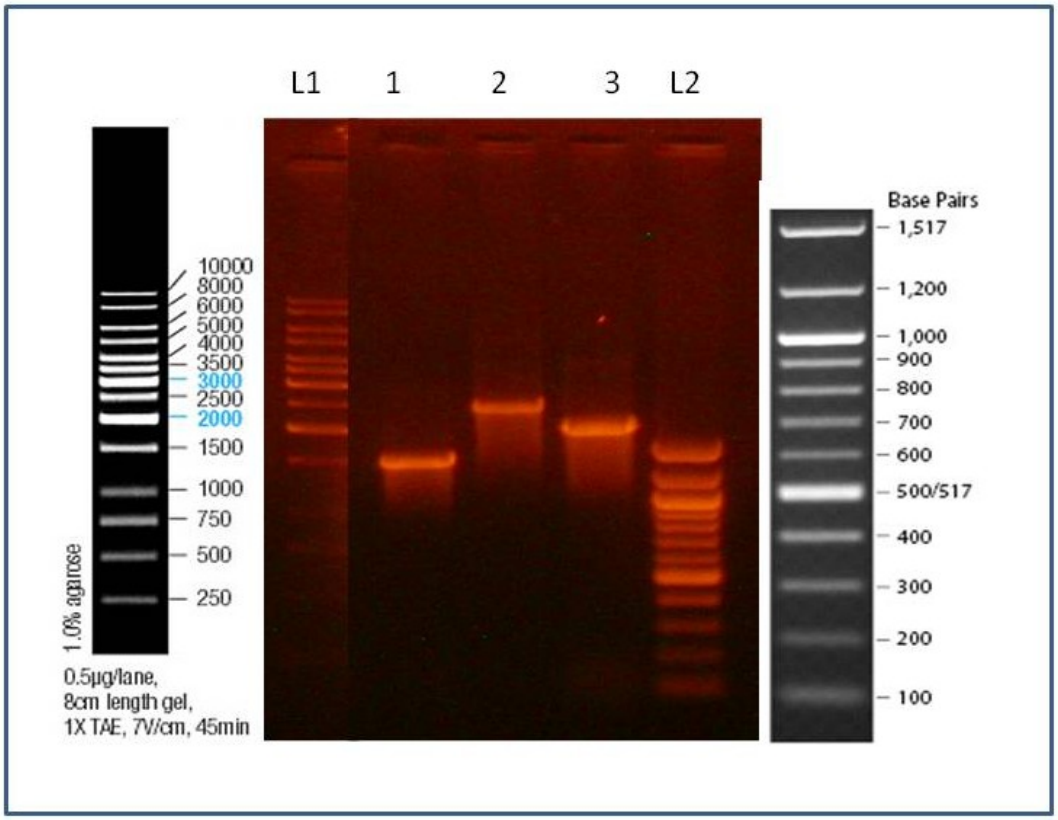


Figure 2 - Agarose gel electrophoresis confirming PCR amplification:: L_1-1 Kb DNA marker; Lane 1 - Amplified *SPD0247* gene; Lane 2 - Amplified *SPD0853* gene; Lane 3- Amplified *SPD0065* gene; L_2-100 bp DNA marker.

I.3 Purification and extraction of the PCR products

The products of PCR reaction were subjected to an agarose gel electrophoresis and the bands excised for purification as described on the Material and Methods chapter, section VII.1. From the purified products, 5 µl were analysed to confirm that the products were still present (Figure 3). As it can be seen in lanes 1, 2 and 3 there is a band corresponding to the size of amplified genes, namely 1380 (lane 1, *SPD0247*), 1788 bp (lane 2 *SPD0065*) and 2166 (lane 3 *SPD0853*). Thus was confirmed that the purification had been successful.

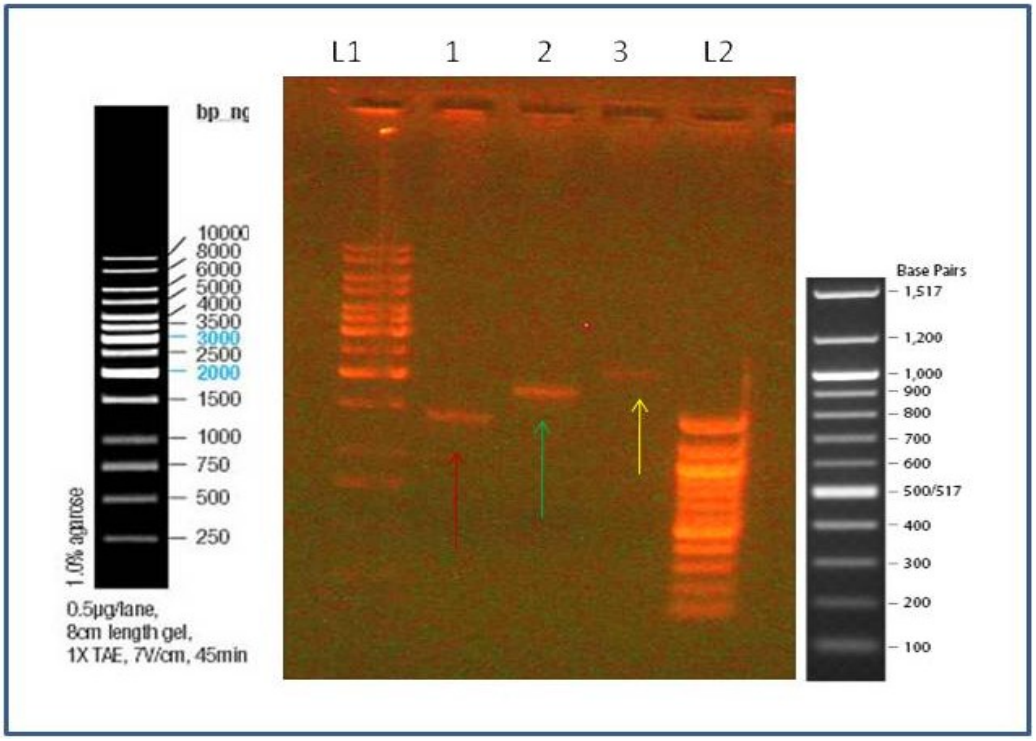


Figure 3- Agarose gel electrophoresis of purified PCR products: L_1-1Kb DNA marker,; Lane 1 - *SPD0247* purified (red arrow) ; Lane 2 - *SPD0065* purified (green arrow); Lane 3 - *SPD0853* purified (yellow arrow); L_2-100bp DNA marker,. The sizes of the ladders are also shown.

I.4 Cloning and analysis by restriction digestion of clones containing *SPD0247*

Gene SPD0247 was cloned into p*Ecoli* - Nterm 6X HN for expression of the recombinant protein. The cloning procedure was done following the protocol described in the Material and Methods chapter section VIII. The confirmation of the successful cloning was done by restriction digestion using restriction enzymes *Hind*III and *Sal*I since these sites would be restored if the cloning process was successful. From the transformants two possibly containing the plasmid p*Ecoli* - Nterm 6X HN and the insert were picked for restriction digestion analysis. The results obtained are shown in Figure 4. As it can be seen in lanes 1 and 6,transformant 1 was cut with *Hind*III and *Sal*I respectively. From the obtained size (5758 bp from the empty plasmid + 1380bp for the gene), it was concluded that this plasmid contained the insert. On the other hand, the second transformant to be

analysed (lanes 3 and 7) , did not contain the insert as it can be seen by its size (5758bp) when comparing to the 1Kb DNA marker. The plasmid in transformant 1 was named pVCT1.

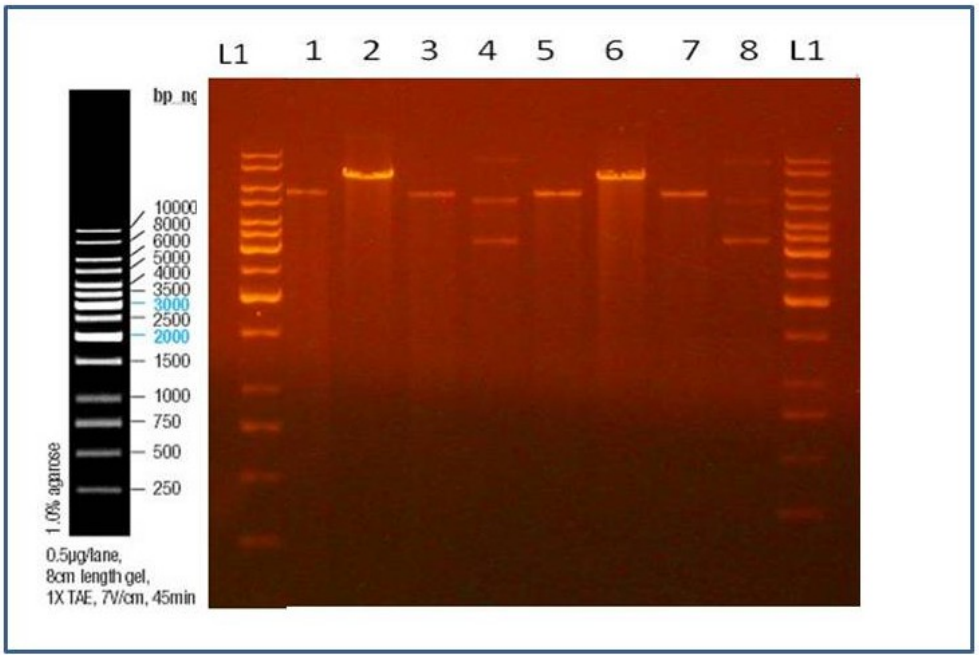


Figure 4 - Restriction digestion with *Hind*III and *Sal*I of the plasmid with *SPD0247* in p*Ecoli* - Nterm 6X HN (pVCT1): L_1-1Kb DNA marker; Lane 1 – Empty plasmid (p*Ecoli* - Nterm 6X HN) PVT3 cut with *Hind*III; Lane 2 – plasmid DNA obtained from transformant1 cut with *Hind*III; Lane 3 – plasmid DNA obtained from transformant 2 cut with *Hind*III; Lane 4 – control no enzyme; Lane 5 – PVT3 cut with *Sal*I; Lane 6 – plasmid DNA obtained from transformant1 cut with *Sal*I; Lane 7 – plasmid DNA obtained from transformant 2 cut with *Sal*I; Lane 8- control no enzyme; L_1 – 1Kb DNA marker. The size of the ladder is also shown.

Despite the fact that the pVCT1 had the correct size, and that the restriction sites for *Sal*I and *Hind*III were present, when sequenced it was noticed that there was a point mutation in the sequence (see Appendix 1, Figure 1) . At position 454 the base A was replaced by G, forming a new codon GGC instead of GAC. GGC translates as glycine instead of the original aspartic acid.

I.5 Analysis of transformants by restriction digestion

Due to the results obtained upon sequencing of pVCT1, two new transformants were tested by restriction digestion for the presence of the gene *SPD0247*. The restriction digestions of these two new transformants were analysed by an agarose gel electrophoresis . The results obtained are shown in Figure 5.

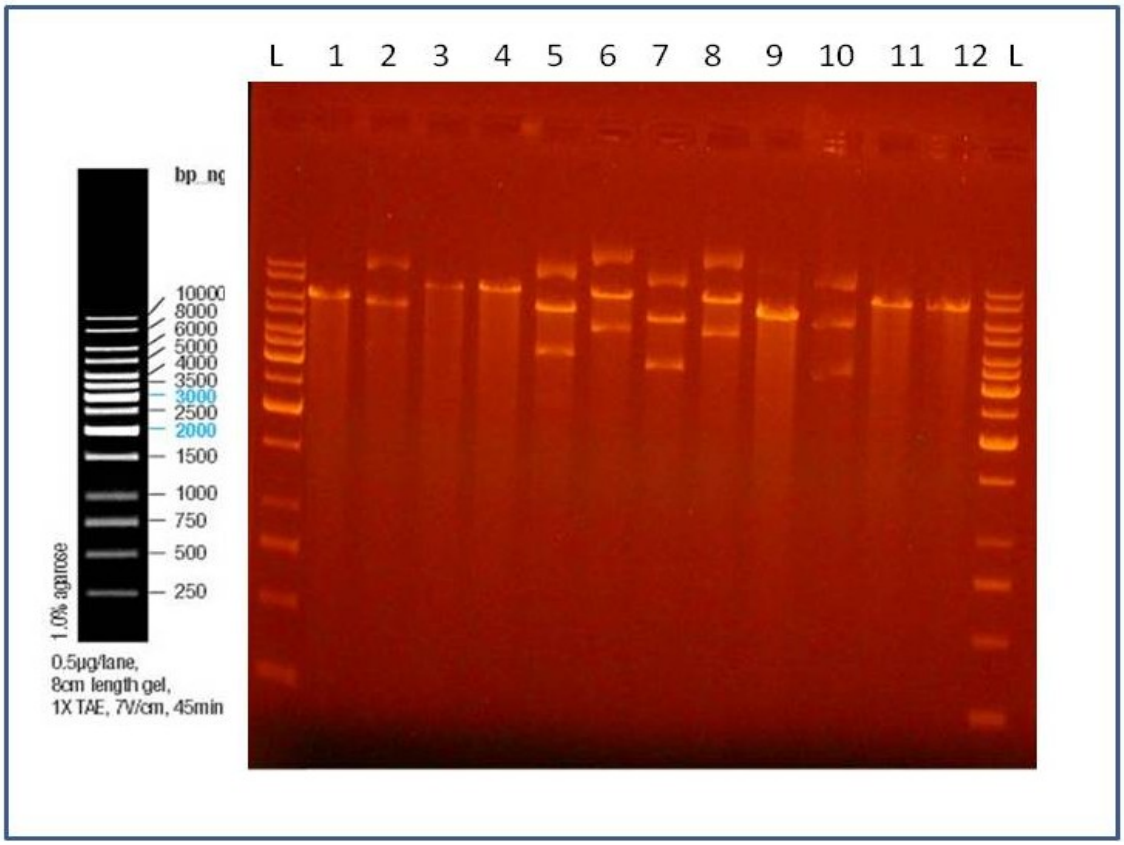


Figure 5- Restriction digestion with *Hind*III and *Sal*I of gene *SPD0247* gene into p*Ecoli* - Nterm 6X HN (pVCT2): L- 1Kb DNA marker; Lane 1 – Empty plasmid (PVT3) cut with *Hind*III; Lane 2-control no enzyme; Lane 3 – plasmid DNA obtained from transformant 1 cut with *Hind*III; Lane 4 – plasmid DNA obtained from transformant 2 cut with *Hind*III; Lane 5 and 6- plasmid DNA obtained from transformant 1 no enzyme added; Lanes 7 and 8– plasmid DNA obtained from transformant 2 no enzymes added; Lane 9–PVT3 cut with SalI; Lane 10 – control no enzyme; Lane 11 – plasmid DNA obtained from transformant 1 cut with *Sal*I; Lane 12- plasmid DNA obtained from transformant 2 cut with *Sal*I; L– 1Kb DNA marker. The size of the ladder is also shown.

Both transformant 1 (lane 3 and 11, cut with *Hind*III and *Sal*I, respectively) and transformant 2 (lane 4 and 12, cut with *Hind*III and *Sal*I, respectively) proved to be positive for the insertion of gene *SPD0247*, since they showed the expected size. The

construct was around 7000 bp (5758 bp from the empty plasmid + 1380 bp for the gene). The DNA from tranformant 2 pVCT2 was sequenced (see Appendix 1, Figure2) and results showed an exact match with the gene of interest. The work was therefore continued with plasmid pVCT2.

I.6 Cloning, and analysis by restriction digestion, of clones containing gene *SPD0065*

The cloning procedure was performed following the protocol described on the Material and Methods Chapter, section VIII.3. In order to determine if the cloning had been successful the recombinant plasmids were analysed by restriction digestion using enzymes *Sal*I and *Hind*III. As it can be seen in Figure 6, all transformants tested positive for the insertion of gene *SPD0065* in plasmid p*Ecoli* - Nterm 6X HN. pVCT3. Transformant 1 (lane 4 and 5, cut with *Hind*III and *Sal*I, respectively), transformant 2 (lane 6 and 7, cut with *Hind*III and *Sal*I, respectively) and transformant 3 (lane 8 and 9, cut with *Hind*III and *Sal*I, respectively) proved to be positive, since they showed the expected size when compared to the 1Kb DNA marker. The construct was around 7600 bp (5758 bp from the empty plasmid + 1788 bp for the gene). The DNA from one of this transformants pVCT3 was sent for sequencing (see result in Appendix 1 Figure3).

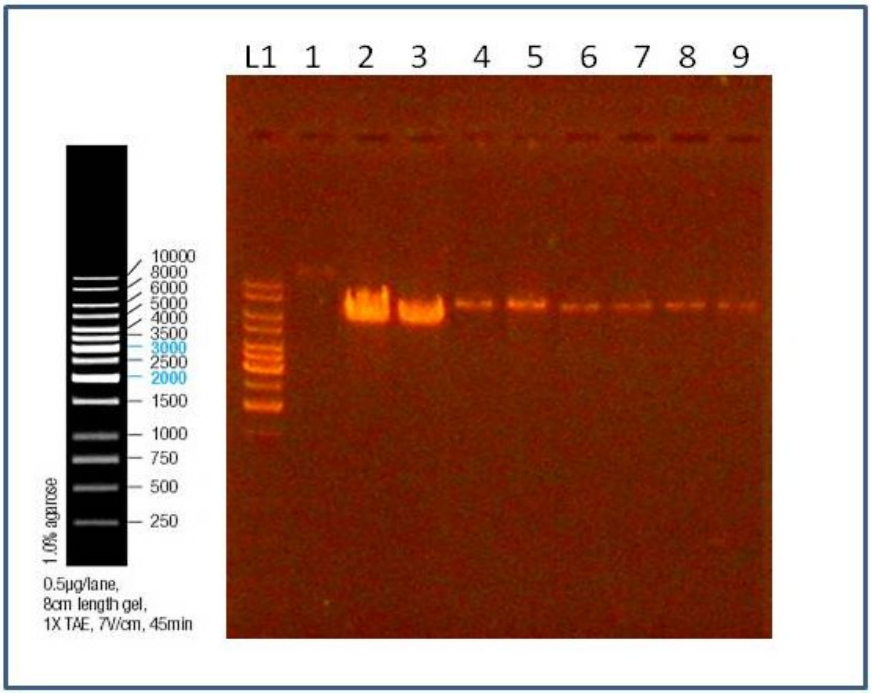


Figure 6 - Restriction digestion with *Hin*dIII and *Sal*I of the cloned *SPD0065* gene in p*Ecoli* - Nterm 6X HN(pVCT3); L_1-1Kb DNA marker; Lane 1– control no enzyme; Lane 2 – Empty plasmid cut with *Hind*III (PVT3); Lane 3 – PVT3 cut with *Sal*I; Lane 4 – plasmid DNA obtained from transformant 1 cut with *Hind*III cut; Lane 5– plasmid DNA obtained from transformant 1 cut with *Sal* I; Lane 6 – plasmid DNA obtained from transformant 2 cut with *Hind*III; Lane 7- plasmid DNA obtained from transformant 2 cut with *Sal*I; Lane 8 – plasmid DNA obtained from transformant 3 cut with *Hind*III; Lane 9 – plasmid DNA obtained from transformant3 cut with *Sal*I. The size of the ladder is also shown.

I.7 Sequencing of clone containing gene *SPD0065*

The DNA from plasmid pVCT3 was sequenced (see Appendix 1, Figure 3) and results showed an exact match with the gene of interest. The work was therefore continued with plasmid pVCT3.

I.8 Cloning and analysis by restriction digestion of clones containing gene *SPD0853*

The results of electrophoresis after restriction digestion of the transformants containing gene *SPD0853* are shown in Figure 7 and Figure 8. The plasmid in transformant 1 did not contain *SPD0853* gene since the DNA band (lanes 5 and 6) had the same size of the empty plasmid, 5758 bp. The cloning, mini prep and restriction digestion procedures were

attempted several times (see Figure 8). However despite these efforts the cloning of gene *SPD0853* was unsuccessful.

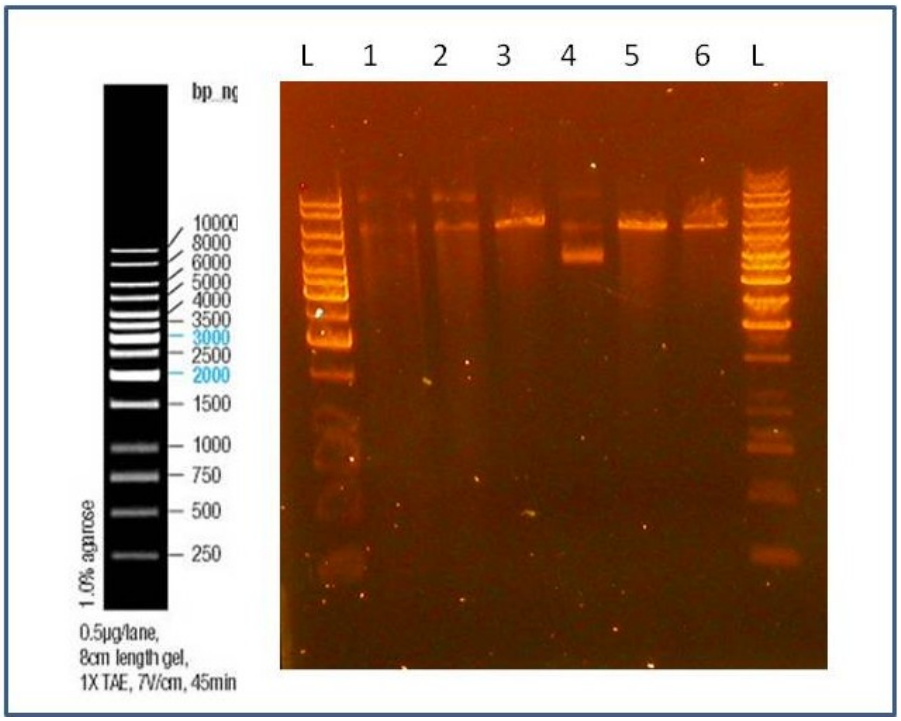


Figure 7- Restriction digestion with HindIII and SalI of the cloned *SPD0853* gene into p*Ecoli* - Nterm 6X HN: L_1- 1Kb DNA marker; Lane 1- no enzyme control; Lane 2- empty vector (PVT3) cut with *Sal*I; Lane 3 – PVT3 cut with *Hind*III; Lane 4– plasmid DNA obtained from transformant 1 no enzyme added; Lane 5 – plasmid DNA obtained from transformant 1 cut with *Sal*I; Lane 6 – plasmid DNA obtained from transformant 1 cut with *Hind*III. The size of the ladder is also shown.

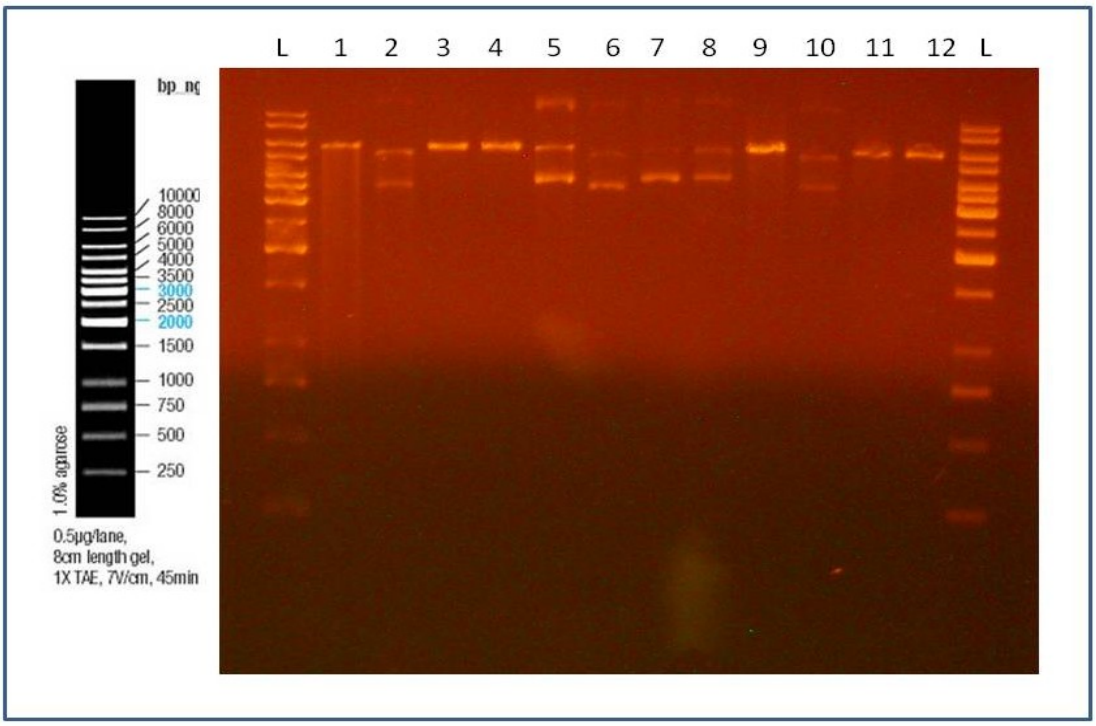


Figure 8- Restriction digestion with *Hind*III and *Sal*I of the cloned *SPD0853* gene into p*Ecoli* - Nterm 6X HN L_1-1Kb DNA marker; Lanes 1 to 4– plasmid DNA obtained from transformants 1, 2, 3 and 4 cut with *Sal*I, Lanes 5 to 8 - plasmid DNA obtained from transformants 1, 2, 3 and 4 no enzymes added; Lanes 9 to 12 – plasmid DNA obtained from transformants 1, 2, 3 and 4 cut with *Hind*III.

I.9 Complementation of mutants

I.9.1 Complementation of knock out mutant SPD0065M

In order to confirm that the mutation of *SPD0065* did not introduce polar effects, SPD0065M was complemented with an intact copy of the gene using pCEP, which is a non-replicative plasmid that allows controlled gene expression following ectopic integration into the chromosome (Guiral *et al.* 2006).

I.9.2 Amplification of interest gene *SPD0065*

SPD0065 was amplified with the primers described in Table 25 in Material and Methods Chapter, which incorporated *Bgl*II sites into the 5'- and 3'- ends of the gene. Figure 9 shows the amplified SPD0065 gene.

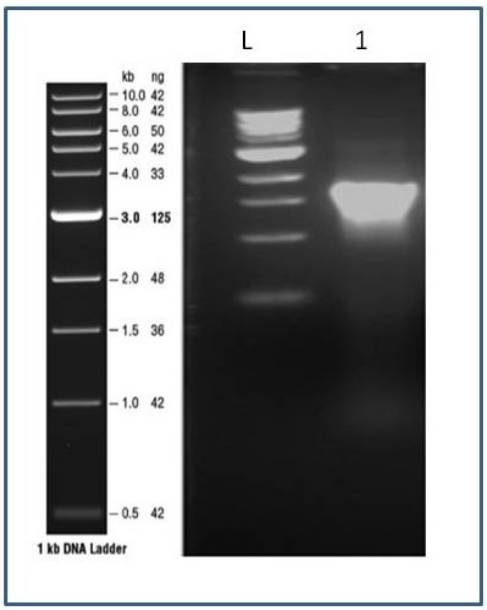


Figure 9- Amplification of gene *SPD0065* with primers SPD0065CF and SPD0065CR. L- 1Kb DNA marker New England Biolabs, Lane 1- Amplified *SPD0065*

I.10 Mini prep and digestion of plasmid pCEP with BamHI

The mini prep of the plasmid pCEP was performed according the protocol described in the Material and Methods chapter, section XVII.1. pCEP was digested with *BamH*I following the protocol described in the Material and Methods chapter, section XVII.2. The restriciton digestion results are shown in Figure 10.

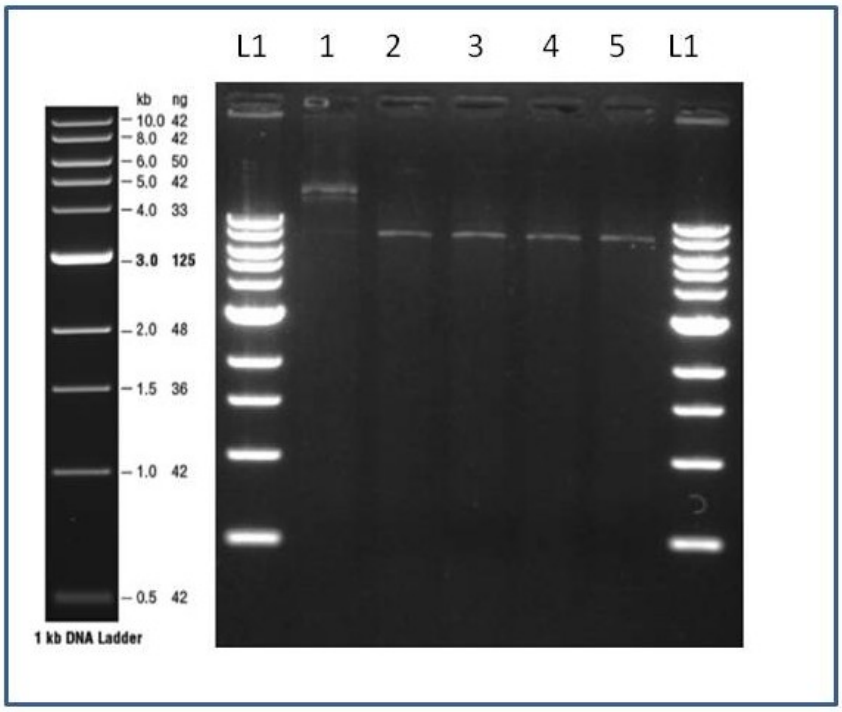


Figure 10- pCEP digested with *BamH*I. L- 1Kb DNA marker New England Biolabs; Lane 1- pCEP not digested; Lanes 2-5 pCEP digested with *BamH*I; L1- Ladder 1 Kb New England Biolabs

I.11 Transformation of isogenic mutant SPD0065M

Following the amplification of the *SPD0065* gene with *Bgl*II compatible ends and the digestion of plasmid pCEP with *BamH*I, the fragments were ligated as described in Table 27 of the Material and Methods chapter, the ligation was possible since the overhangs of both *Bgl*II and *BamH*I are compatible. The resulting plasmid was termed pVCT4 In an attempt to complement knock out mutant SPD0065M the plasmid pVCT4 was used to transform the isogenic mutant using the procedure of Bricker & Camilli (1999) described in the Material and Methods Chapter, section I. The transformants were selected in the presence of spectinomycin (100 μg/ml) and kanamycin (500 μg/ml).

I.12 Confirmation of the transformation SPD0065M

The successful introduction of the intact copy of the gene *SPD0065* into SPD0065M was confirmed by PCR using primers *malF* and pCEPR (Table 28, Material and Methods

chapter, section XVII.6) whose recognition sites are localized immediately up and downstream of the cloning site, respectively. The results obtained are shown in Figure 11. As it can be confirmed by looking at Figure 11, transformants 4 to 7 (lanes 4 to 7) were positive for introduction of a copy of the gene *SPD0065* into the knock-out mutant SPD0065M (1788 bp + 263bp of kanamycin cassette) . These colonies were used for further work. The complemented strain was designated SPD0065REV

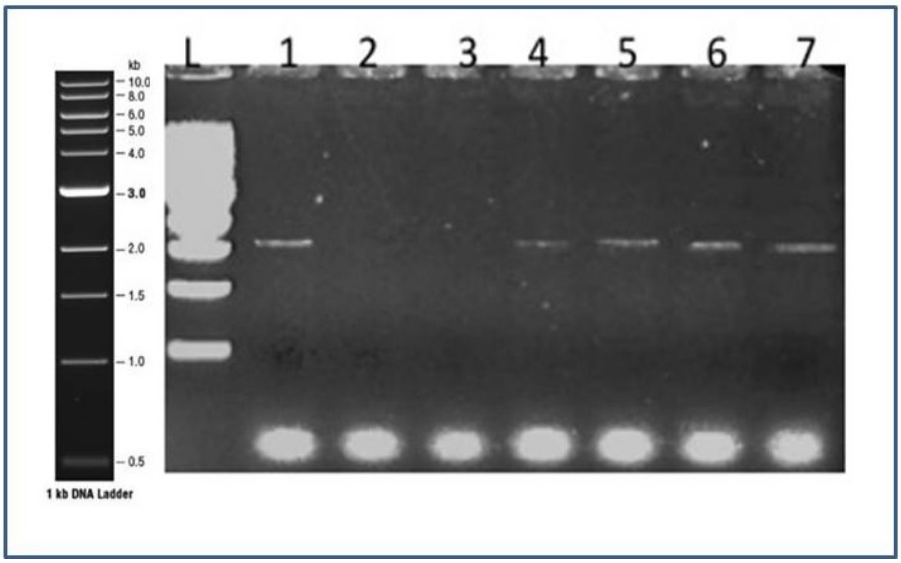


Figure 11–Confirmation by PCR of the introduction of a copy of the gene *SPD0065* into knock out mutant SPD0065M. Primers used were *malF* and pCEPR. L-1Kb DNA marker; Lane 1- positive control; Lane 2-7–Transformants with possible introduction of the intact copy of the gene *SPD0065*.

I.13 Complementation of knock out mutant SPD0247M

SPD0247 was amplified with the primers described in Table 25 in the Material and Methods chapter. These primers incorporate, *Bgl*II sites into the 5'- and 3'- ends of the gene. Figure 12 shows the amplified gene *SPD0247*.

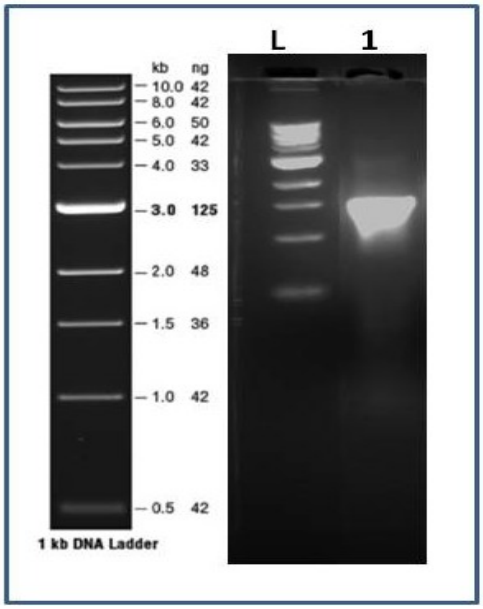


Figure 12- Amplification of gene *SPD0247* with primers SPD0247CF and SPD0247CR; L-1Kb DNA marker, Lane 1- Amplified *SPD0247*.

I.14 Transformation of isogenic mutant SPD0247M

Following the amplification of the *SPD0247* gene with *Bgl*II compatible ends and the digestion of plasmid pCEP with *BamH*I (Figure 10) the fragments were ligated following what is described in Table 27 from the Material and Methods chapter. In an attempt to complement the knock-out mutant SPD0247M, the plasmid named pVCT5 (containing pCEP plasmid + *SPD0247* gene) was used to transform the isogenic mutant SPD0247M using the procedure described by Bricker & Camilli (1999) (Material and Methods chapter, section I). The transformants were selected in the presence of spectinomycin and kanamycin (100 μg/ml, 500 μg/ml respectively).

I.15 Confirmation of the transformation of SPD0247M

The introduction of the intact copy of the gene was unsuccessful as confirmed by PCR using *malF* and pCEPR primers (Table 28 from the Material and Methods section XVII.6) whose recognition sites are localised immediately up and downstream of the cloning site. The results obtained are shown in Figure 13 As it can be confirmed from Figure 13 all transformants are negative for the presence of the gene *SPD0247,* which would present a

band of around 1643bp (1380 bp insert + 263 bp of the kanamycin cassette). The complementation was attempted several times but all the attempts proved unsuccessful.

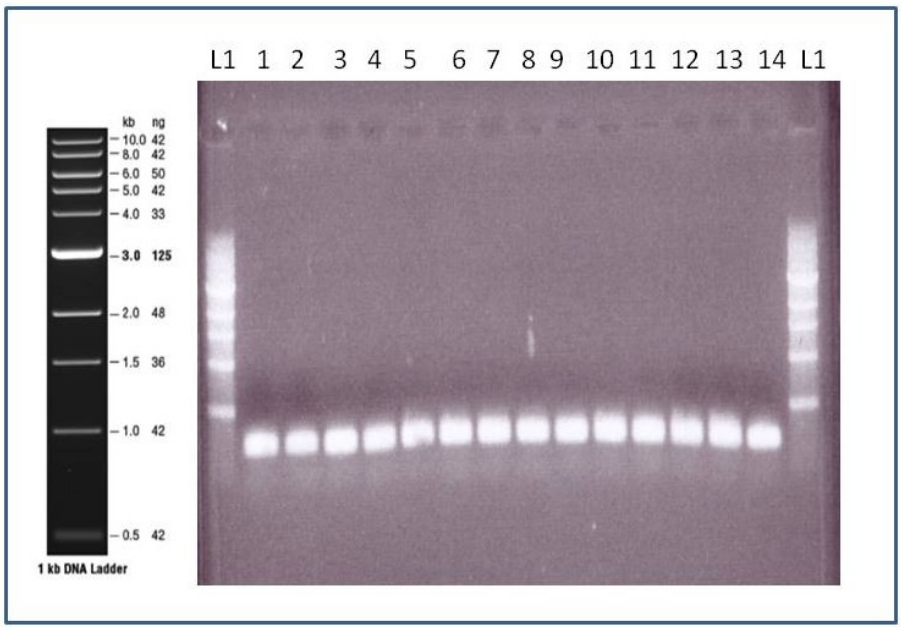


Figure 13 – Confirmation by PCR of the introduction of an intact copy of the gene *SPD0247* into SPD0247M knock-out mutant using primers *malF* and pCEPR. L- 1Kb DNA marker New England Biolabs, Lane 1-14 - transformants with possible introduction of the intact copy of gene *SPD0247*.

I.16 Knock out mutation in different backgrounds

I.16.1 Knock out mutations of gene *SPD0065* in serotypes 6B and 19

In order to understand if the effect of the mutation in *SPD0065* gene was extended to other serotypes, knock out mutants in two different backgrounds were constructed using the *in vitro* mariner mutagenesis method (Material and Methods chapter section I).

To achieve this goal the mutated region was amplified from SPD0065M with primers SPD0065F and SPD0065R (Table 29) and transformed into serotype 6B and 19F pneumococci, using the method described in the Material and Methods Chapter, section I. The presence of the insertion was confirmed by PCR. The results obtained are shown in Figure 14. A band of approximately 3Kb was expected (1146bp-minitranposon carrying the spectinomycin resistance and the gene 1788bp). Figure 14 shows that transformants 3,

4 and 5 are positive for the knock-out mutation.Transformant 3 was named pVS1 and was used in further studies.

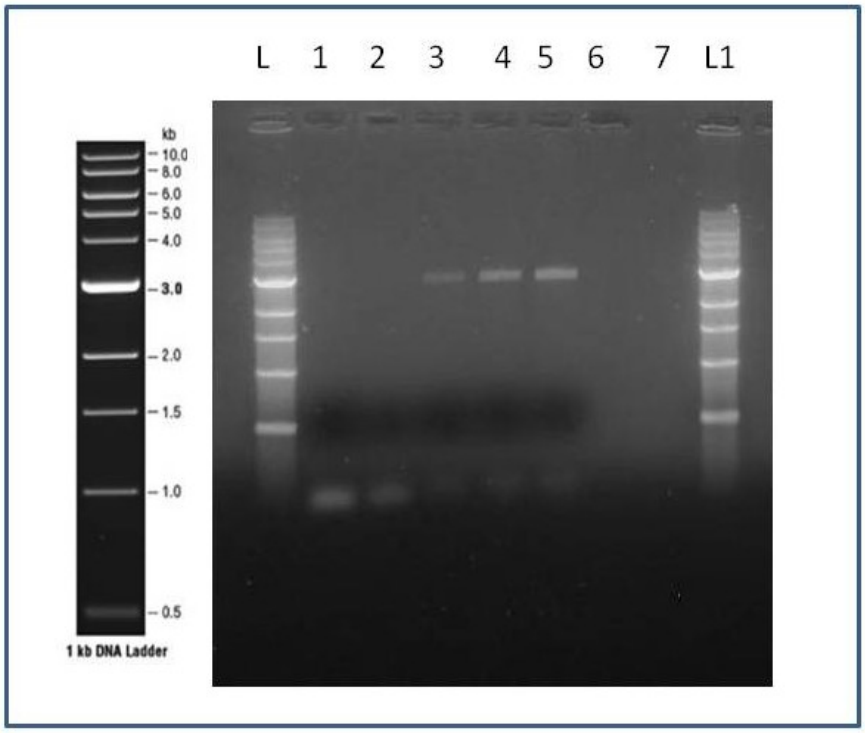


Figure 14 – Evidence of knock out mutation in gene *SPD0065* in a serotype 19 of pneumococcus. L- 1Kb DNA marker; Lane 1-Negative control (using the primers that amplified SPD0065 mutated region with DNA of serotype 19); Lanes 2 to 6- possible transformants; L1- 1Kb DNA marker.

Following the same procedure, the mutation was also performed in a serotype 6B pneumococcus. The results obtained are shown in Figure 15. A band of approximately 3Kb was expected (1146bp-minitranposon carrying the spectinomycin resistance and the gene 1788bp). Figure 15 shows that one colony (lane 3) was a positive transformant. This transformant was named pVS2 and was used in further studies.

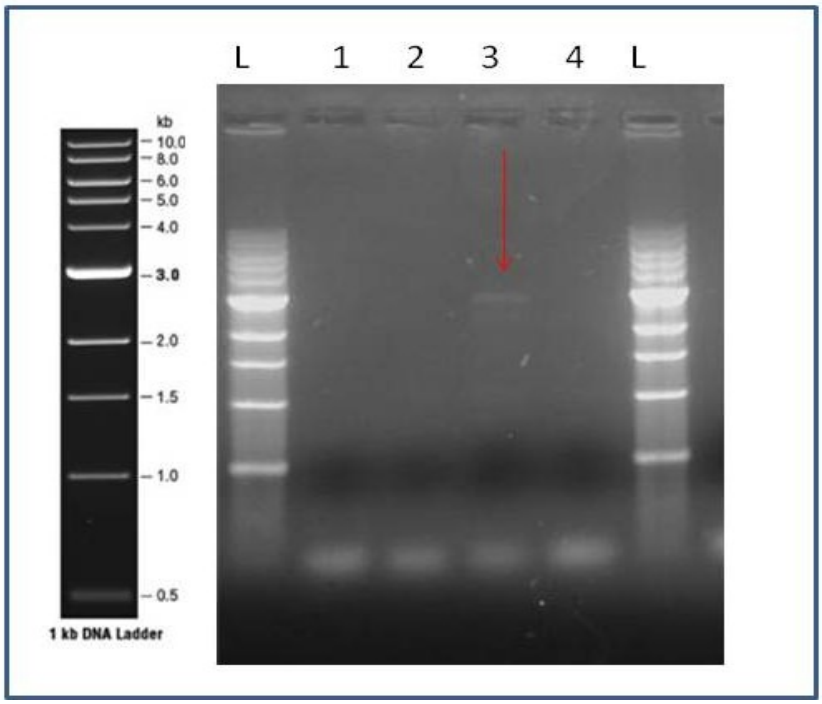


Figure 15- Evidence of knock-out mutation in gene *SPD0065* in a serotype 6B pneumococcus. L- 1Kb DNA marker; Lanes 1- negative control (using the primers that amplified SPD0065 mutated region with DNA of serotype 6B), Lanes 2 to 4- possible transformants; L1- 1Kb DNA marker. The arrow represents the positive transformant.

E RESULTS

I. Overview

This chapter starts with the description of the expression and purification of the enzyme encoded by gene *SPD0065*. Once purified this enzyme was characterised in terms of subcellular localisation, linkage and substrate specificity, and kinetic parameters determination. Through this process the enzyme encoded by *SPD0065* was confirmed as a β-galactosidase (BgaC).

Subsequent to the *in vitro* characterisation of the enzyme, virulence, immunogenicity and enzyme assays using cell lysates were performed in an attempt to understand the function of this enzyme *in vivo*.

I.1 Synthesis of protein encoded by gene *SPD0065*

The expression procedure had been optimised for the protein encoded by the gene *SPD0247* and this was tested and adopted for *SPD0065* because the same expression system was used. Expression of the gene *SPD0065* originated a protein with a polyhistidine tag and this was done in *E.coli* BL21 DE3 following the protocol described in Section VIII.5 of the Material and Methods Chapter. The analysis of the resulting protein by SDS-PAGE showed that it had a molecular weight of approximately 69 kDa, which was predicted from the deduced amino acid sequence.

The samples were prepared according to the protocol described on the Material and Methods Chapter section VIII.6. Following the expression, the pellets containing the mixture of proteins were frozen at -80°C. The resultant protein mixture was analysed by SDS-PAGE and the results are shown in Figure 1. As it can be seen, in Figure 1A, in lanes 6 and 9 there is a distinct band of about 69 kDa corresponding to the recombinant protein expressed 1 and 2 hours after induction with IPTG, whereas in lanes 5 and 8, for example, no bands at that molecular weight range were observed. These lanes contain the proteins from empty vectors that have also been induced for the purpose of serving as a negative control. In Figure 1B, lanes 3, 6 and 9 a band of about 69 kDa was observed, this band corresponds to the expressed recombinant protein 3, 4 and 5 hours after induction with IPTG. In a similar manner, lanes 2, 5 and 8 contain the expression of empty vectors that were also induced with IPTG, however in these lanes no band in the 69kDa molecular weight range is observed. The expression of the recombinant protein was considered successful. It was concluded that 25°C was the temperature to use for this expression system and it was adopted until the end of the project.

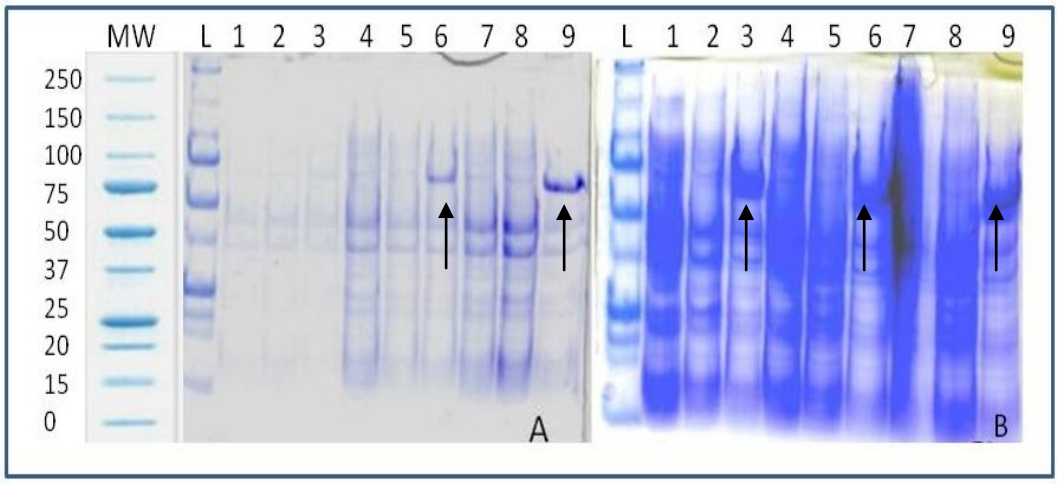


Figure 1A- Expression of recombinant protein encoded by gene *SPD0065*. MW- Prestained all blues protein marker. L- protein molecular weight marker; Lanes 1, 4 and 7– Empty vector (pVT3) non-induced before induction and 1 and 2 h after induction respectively; Lanes 2,5 and 8- pVT3 before induction and 1 and 2 hours after induction; Lane 3, 6 and 9- *E.coli* containing *SPD0065* construct (pVCT2) before induction and 1 and 2 hours after induction respectively; **B**- Lanes 1,4 and 7 pVT3 non induced 3, 4 and 5 hours after induction respectively; Lanes 2, 5 and 8- pVT3 3, 4 and 5 hours after induction Lanes 3, 6 and 9- pVCT2 3, 4 and 5 hours after induction respectively; The black arrow indicates the target protein.

I.2 Determination of protein solubility

Since proteins can accumulate in insoluble inclusion bodies it is important to determine the distribution of the protein in soluble and insoluble forms. In order to understand if the protein under study was in the soluble or insoluble fractions in *E.coli,* cell extracts were prepared following the protocols described in the Material and Methods Chapter, section VIII.7. The samples were then analysed by SDS-PAGE gels and the results are shown in Figure 2.

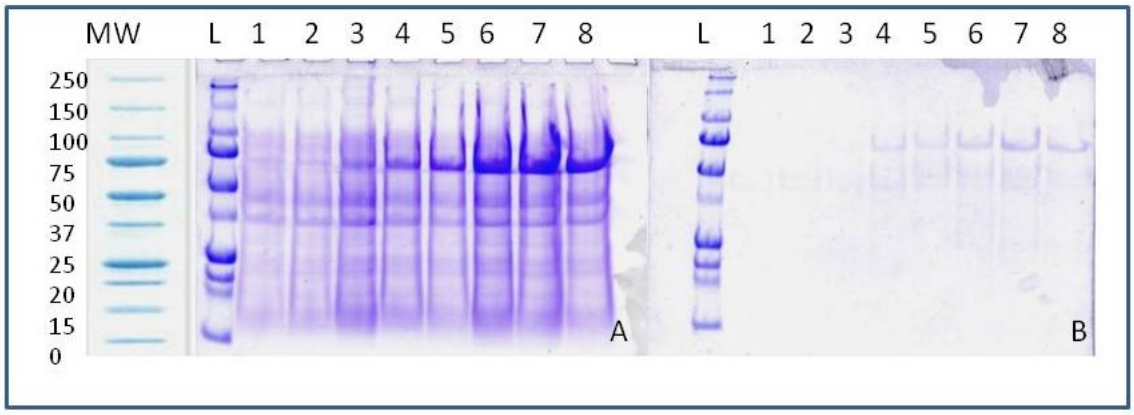


Figure 2- Determination of protein solubility. **A** insoluble proteins; MW- Prestained all blues protein marker L- protein molecular weight marker; Lane1-Empty vector (pVT3) before induction; Lane 2- PVT3 non induced; Lane 3- construct containing *SPD0065* (pVCT2) before induction; Lanes 4-8- pVCT2 after induction 1 to 5 hours respectively **B**- soluble proteins L- protein molecular weight marker; Lane 1-pVT3 before induction; Lane 2- pVT3 non induced; Lane 3- construct containing *SPD0065* (pVCT2) before induction; Lanes 4-8- pVT2 1 to 5h after induction;

In Figure 2A, (lanes 4 to 8), a very intense band is visible around 69 kDa suggesting that the recombinant protein is in the insoluble fraction. In Figure 2 B, bands in the 69 kDa region (lane 4 to 8) are also visible, however, these bands are much fainter than the ones seen in Figure2 A indicating that the majority of the recombinant protein accumulates in the insoluble fraction.

Although the SDS-PAGE gels indicate that the majority of the recombinant enzyme was in the insoluble fraction, the purification was still attempted on the soluble fraction where the protein was also detectable. From this point onwards all the expression pellets were prepared assuming that sufficient recombinant protein was in the soluble fraction.

I.3 Analysis of the purification

Before starting the purification protocol, the pellets from *E.coli* expressing plasmid pVCT2 were prepared and treated under mild conditions, 50mM sodium phosphate, 300mM NaCl, pH 7.0, in order to isolate the native recombinant protein (see section IX.1 from the Material and Methods Chapter). Attempting the purification in more

stringent conditions, e.g. by adding 6M guanidinium, would have increased the recombinant protein solubility; however it would have involved the denaturing of the enzyme.

The purification protocol was done following the procedure described in the Material and Methods Chapter, section IX.2. A column containing immobilised cobalt was used to enable the binding of the polyhistidine-tagged recombinant protein. This principle is based on the reversible interactions between the amino acid side chains and the immobilised cobalt. Elution occurs when the imidazole nitrogen is protonated, generating a positively charged ammonium ion, which is repelled by the positively charged metal. This resin offered two practical advantages. First there is practically no binding of background proteins to the resin, and secondly, the use of such a resin allows the elution of the tagged protein under non-stringent conditions, for example with the use of a slightly higher pH.

The eluate was collected in 500 μl fractions, from which 10 μl from the first 9 fractions were run on an SDS PAGE gel Figure 3. The purified protein can be seen in lanes 6 to 9 (Figure 3) as expected, since it was indicated by the manufacturers (Clontech, Japan) that the majority of the polyhistidine-tagged protein should be recovered in the first 4 to 5 ml of eluate. Fractions 6 to 9 containing the recombinant protein were quantified using a Bradford assay (see Material and Methods Chapter section IX.3). The protein concentration of each fraction was determined as: fraction 6, 4.49 mg/ml, fraction 7, 4.19 mg/ml, fraction 8, 3.88 mg/ml and fraction 9, 2.24 mg/ml.

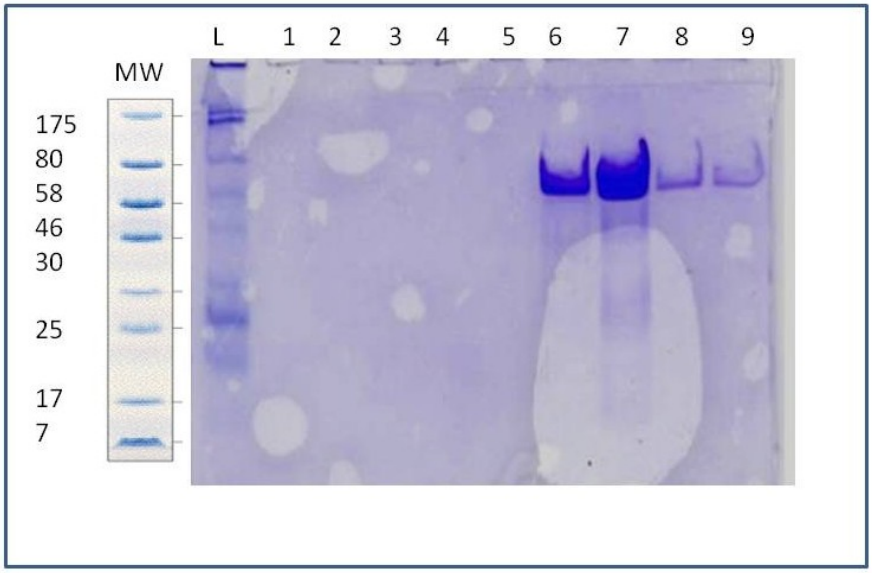


Figure 3 – Determination of the purity of the recombinant protein SPD0065. MW- Prestained protein marker P7708S New England Biolabs; L- protein molecular weight marker; Lanes 1–9 sequential fractions eluted from a sepharose-Co column. Fractions 6 to 9 contain the purified recombinant protein.

I.4 Mass Spectrometry for the protein encoded by *SPD0065*

In order to identify this protein as being a putative β-galactosidase the purified recombinant protein was analysed by matrix assisted laser desorption ionization–time-of-flight (MALDI-TOF) mass spectrometry of the tryptic digests of the products by PNACL at the University of Leicester. The enzyme was identified as being a β-galactosidase (see Figure 5, Appendix 1).

I.5 Substrate specificity

The substrate specificity was determined next, by performing endpoint experiments following the protocol described in Material and Methods section X. The enzyme was assayed with eight synthetic substrates and the results are shown in Table 1.

Table 1- Activity of β-galactosidase against 4-methylumbelliferyl- and pNP- linked substrates. Minus sign indicates the absence and plus sign indicate the presence of activity against each substrate.

Substrates	Activity
ONP-β-D-Galactopyranoside	+
4-Methylumbelliferyl-N-acetyl-β-D-glucosaminide	-
4-Methylumbelliferyl-β-D-glucopyranoside	-
O-Nitrophenyl-α-D-galactopyranoside	-
pNP-β-D-Glucoside-6-phosphate	-
pNP-α-D-Mannose-6-phosphate	-
pNP-α-D-Galactose-6-phosphate	-
pNP-β-D-Glucose-6-phosphate	-

The β-galactosidase recombinant protein only showed activity against the synthetic substrate ONP-β-galactopyranoside. It is also noteworthy that even in the presence the substrate ONP-α-galactopyranoside there was no detectable activity. This result suggests a high specificity for the β-linkage in enzyme-substrate recognition. In addition, no detectable activity was seen in the presence of the substrate pNP-α-D-galactose-6-phosphate. In a similar manner, no enzyme activity was seen when the substrate 4-methylumbelliferyl-β-D-glucopyranoside was used. This result suggests a high specificity for the sugar galactose. Hence, these data agree with the previous identification of the enzyme by mass spectrometry as being a β- galactosidase.

I.6 Kinetic characterisation of enzymes SPD0065

Once the most appropriate synthetic substrate for the recombinant enzyme was determined as being ONP-β-galactopyranoside, the following step was the determination of its kinetic parameters. Kinetic experiments were designed and performed following the protocol described on Material and Methods Chapter section XI.

K_M and V_{max} were calculated in order to understand the affinity level for the synthetic substrate used. The reason why these studies were performed using a synthetic substrate is mainly due to the fact that the potential biological substrate (mucin) does not have a reporting system either a cromophore or a fluorophore. K_{cat} was not calculated because there is no structural information for the enzyme and therefore the number of catalytic site would have to be assumed.

The first step to determine the biochemical properties of this protein was to understand if its kinetics obeyed Michaelis Menten kinetics. When the substrate concentration is low, the velocity (v) increases linearly with the concentration of substrate, but eventually as the concentration is increased, this linearity is lost. This point which is considered the saturating concentration of substrate is where velocity tends to a limiting value named V_{max} (maximum velocity) (Fersht, 1999).

Once the saturation condition was achieved the points were plotted and a non-linear regression was adjusted to these experimental points. The results obtained determined that the recombinant enzyme BgaC followed saturation kinetics therefore allowing parameters K_M and V_{max} to be determined using Michaelis Menten kinetics (Figure 4).

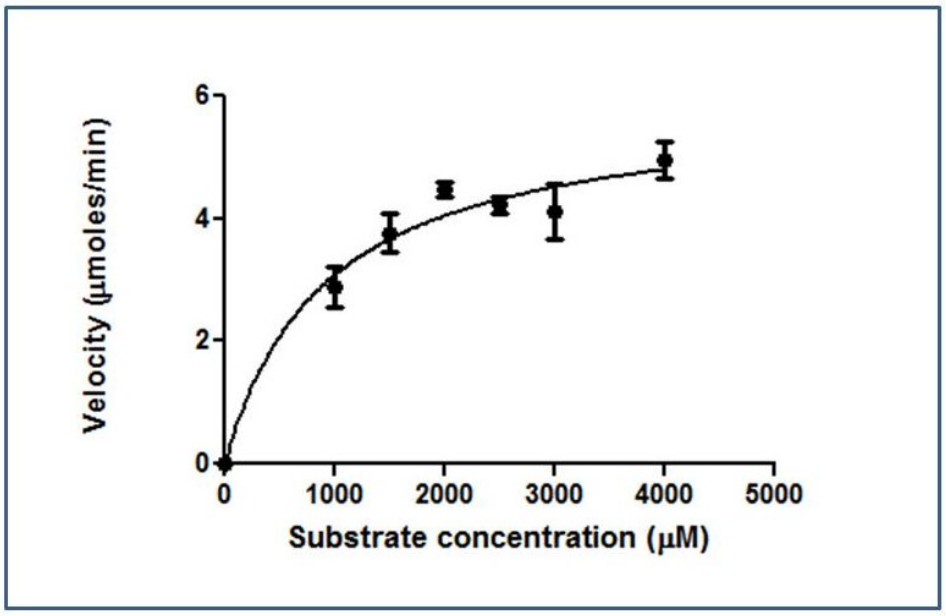


Figure 4- Determination of Michaelis Menten kinetic parameters against substrate ONP-β-D-galactopyranoside. Each datum point represents the mean of three independent experiments. Vertical bars indicate standard deviation. K_M and Vmax were calculated using non linear regression using the program GraphPad Prism version 5.00 for Windows, (GraphPad Software, San Diego California USA, www.graphpad.com).

In order to perform a complete analysis, the results were also plotted using both Lineweaver Burk (Figure 5) and Eadie Hofstee (Figure 6) approaches, and K_M and V_{max} were calculated. The results obtained with the adjustment of a non linear approach were confirmed by the use of Lineweaver Burk and Eadie Hofstee approaches. The results obtained from all three methods are summarised in Table 2.

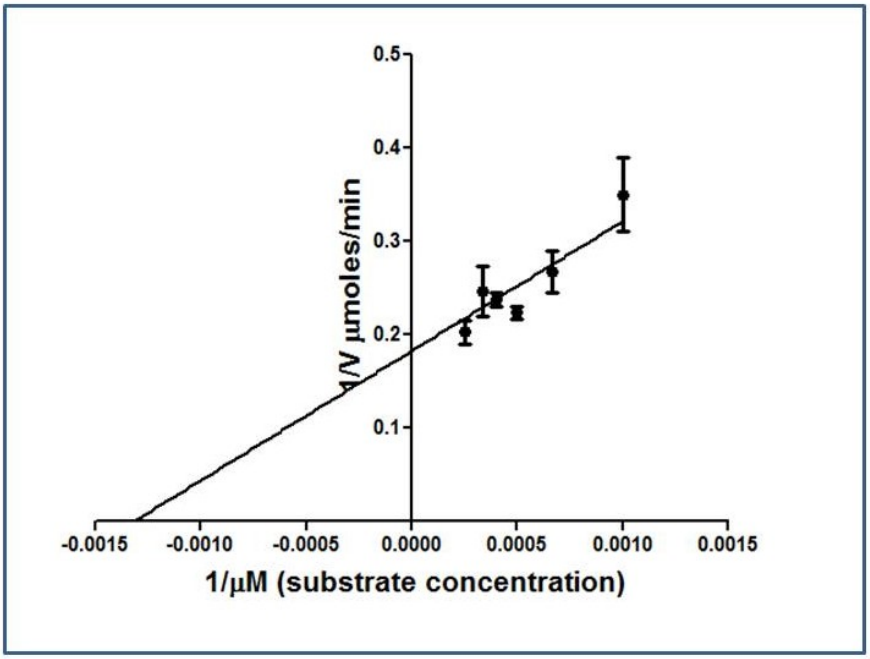


Figure 5-Graphic determination of K_M and V_{max} parameters using a Lineweaver Burk plot. Each datum point corresponds to the mean of three independent experiments. The vertical bars represent the standard deviation.

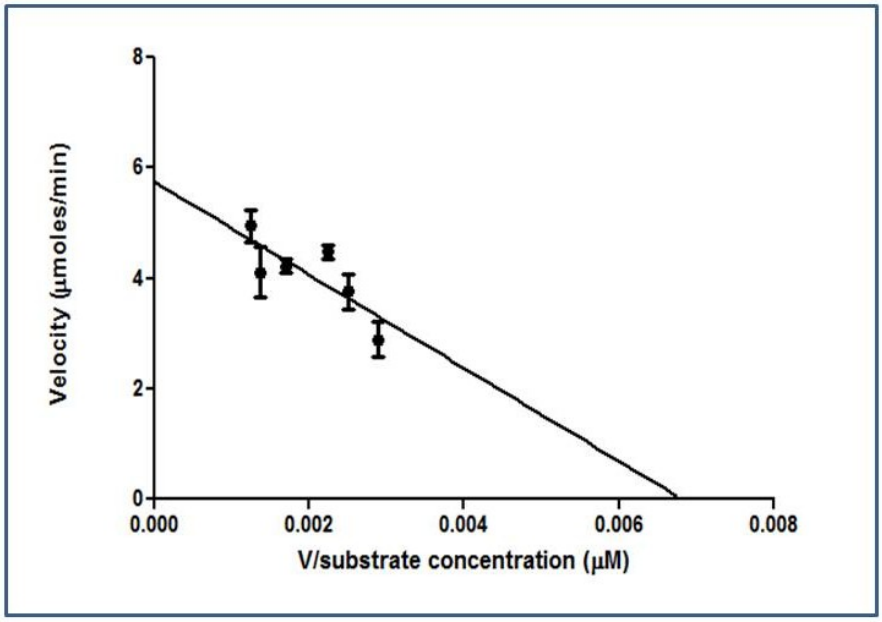


Figure 6- Graphic determination of K_M and Vmax parameters using an Eadie Hofstee plot. Each datum point corresponds to the mean of three independent experiments. The vertical bars represent the standard deviation.

Table 2- Kinetic parameters determination using the Michaelis Menten approach and two linearisation methods (Lineweaver Burk and Eadie-Hofstee).

	Interpretation of kinetic parameters based on graphical representation			**Experimental values**	
Approaches	**Intercept the x axis**	**Intercept the y axis**	**Slope**	**K_M (μM)**	**Vmax (μmol.min^{-1}.mg^{-1})**
Michaelis Menten (Figure 4)	_	_	_	907 ± 294	0.11 ± 0.01
Lineweaver Burke (Figure 5)	-1/Km	1/Vmax	Km/Vmax	758	0.10
Eadie Hofstee (Figure 6)	_	Vmax	-Km	843	0.11

I.7 Analysis of glycoprotein degradation by the wild type and knock-out mutant SPD0065M

The following results were acquired by Dr Karen Homer King's College Dental Institute, London, using the mutants prepared in this study.

Pig gastric mucin (PGM) is a heavily glycosylated protein. The supplier's (Sigma) data indicate that, as supplied, it contains 10% (w/v) fucose, 37% (w/v) hexosamines, 27% (w/v) galactose and 6% (w/v) neuraminic acids, with the protein backbone making up the remaining 20% (w/v). The analysis of the composition of PGM following dialysis and centrifugation was broadly in agreement with these figures: 7% (w/v)

fucose; 44% (w/v) hexosamines; 14% (w/v) galactose and 3% (w/v) neuraminic acids. We were able to define the composition of this glycoprotein further and showed that the hexosamines comprised N-acetylglucosamine and N-acetylgalactosamine in a molar ratio of approximately 2:1. In addition to sialic acid, PGM also contained relatively small amounts of N-glycolylneuraminic acid (a molar ratio of approximately 6:1). During growth of the wild-type D39 strain in PGM-containing medium, significant amounts of both N-acetyl- and N-glycolylneuraminic acids were removed in a time-dependent manner. After 48 h of growth, approximately 40-50% of both sialic acids had been cleaved from the glycoprotein backbone. Galactose and N-acetylgalactosamine concentrations in PGM were reduced by 30% and 35%, respectively. There was limited cleavage of fucose and N-acetylglucosamine following 48 h growth of bacteria and the composition of PGM in the supernates, with respect to these two sugars, did not differ significantly from non-inoculated medium controls.

The isogenic knock out mutant, SPD0065M, which lacked the putative β-galactosidase, released the same amount of sialic acids from PGM as the wild-type strain after 48 h growth however, galactose cleavage was reduced from 30% by strain D39 to only 13% of the total by this mutant and there was no detectable removal of N-acetylgalactosamine, fucose or N-acetylglucosamine from PGM by this mutant.

I.8 Analysis of glycoprotein degradation by the recombinant protein BgaC

The following results were acquired in by Dr Karen Homer King's College Dental Institute London, using the recombinant enzyme produced in this study.

The ability of recombinant β-galactosidase to release sugars from glycoproteins was studied using PGM and fetuin as biological substrates, both with or without prior

treatment with immobilised sialidase. Fetuin was chosen as an alternative substrate to mucin because the structural complexity of its oligosaccharides is substantially less than in mucin (Green *et al.*, 1988). Also its sugars are subject to little modification, for example it does not suffer sulphation that is extensive in many mucins and which is known to limit glycosidic activity. Because the removal of sialic acids has been shown to be essential for further removal of subsequent sugars (King *et al.*, 2006), suggesting a sequential release of sugars, sialic acid was removed from PGM using immobilised sialidase in the assays.

Treatment of PGM with sialidase resulted in the removal of approximately 10% of sialic acid, reflecting perhaps O-acetylation of the majority of these residues decreasing therefore their susceptibility to cleavage. On the other hand, with fetuin there was removal of greater than 95% of sialic acid. This treatment resulted in a glycoprotein containing N- and O- linked glycans (Figure 7) the majority having reducing terminal galactose residues.

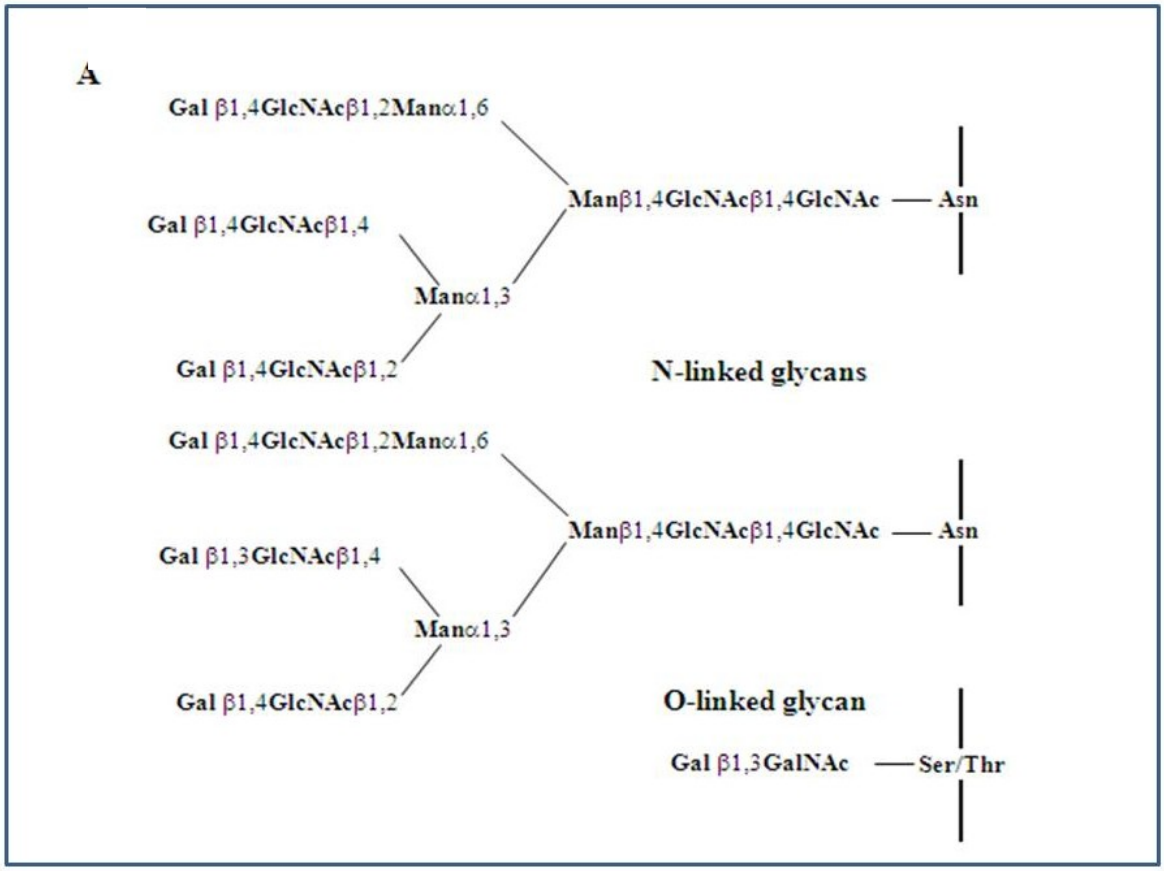


Figure 7- Glycan structures of desialylated fetuin. Adapted from references ((Green *et al.*, 1988, Takasaki & Kobata, 1986).

Treatment of PGM with intact or partially desialylated with recombinant protein encoded by *SPD0065* did not result in any detectable release of any of the neutral or amino sugars which are found in the glycoproteins, probably because galactose residues are blocked by the presence of other monosaccharides in this complex glycoprotein.

We demonstrated, however, that recombinant SPD0065 acted on desialylated fetuin, releasing galactose from the glycans (Figure 8). Activity of the enzyme against fully sialylated fetuin was minimal, with the release of only a trace amount of galactose, indicating that sialic acid removal from oligosaccharides to expose galactose was essential for the activity of the enzyme.

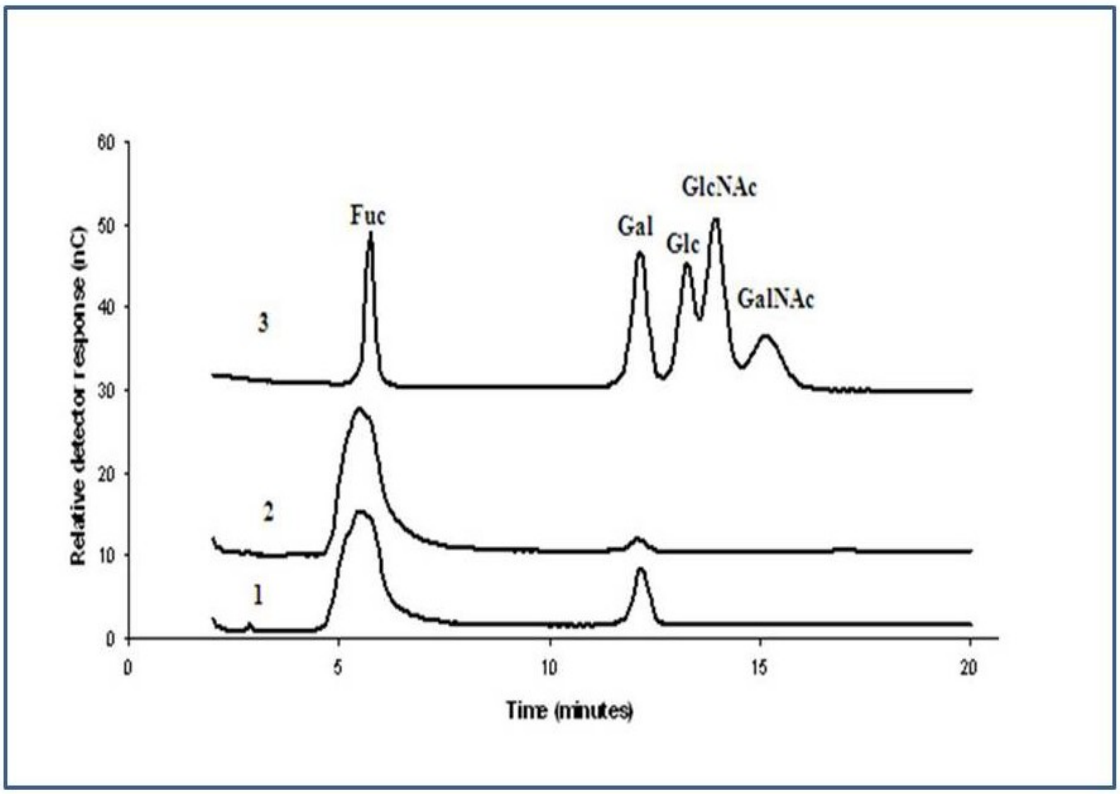


Figure 8- HPAEC-PAD separation and detection of monosaccharides. Traces 1 and 2 show monosaccharide release from desialylated and intact fetuin, respectively, following treatment with recombinant protein SPD0065. Trace 3 shows the separation of monosaccharide standards (Fuc, fucose; Gal, galactose, Glc, glucose; GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine). Glucose, while not a component of PGM or fetuin oligosaccharides, was included in the standard mixture as it can be a contaminant introduced during processing of glycoproteins.

I.9 Identification of linkage specificity of enzyme SPD0065

The method used for the linkage specificity is based on the principal of lectins that bind specific carbohydrates structures because they contain two or more binding sites for carbohydrates units. Lectins and carbohydrates are linked by weak interactions that ensure specificity however unlinkage is also possible when necessary (Berg *et al.*, 2001).

For these particular assay lectins PNA (peanut agglutinin) and DSA (datura stramonium agglutinin) were used. The reason for these choices was that PNA recognises core disaccharide galactose β(1-3) N-acetylgalactosamine and thus is suitable for identifying O-glycosidically linked carbohydrate chains whereas DSA

recognises galactose β(1-4) N-acetylglucosamine in N-glycans and O-glycans. Also, these lectins are supplied with a reporting system consisting of a conjugated digoxigenin which enables immunological detection of the bound lectins (DIG Glycan differentiation Kit, Roche).

As it was mentioned before, the enzyme encoded by gene *SPD0065* was identified as β-galactosidase. Knowing that the previously identified β-galactosidase encoded by gene *SPD0562* cleaves the linkage galactose β (1-4) N-acetylglucosamine, it was necessary to identify if the function of these two enzymes was different. In order to answer that question, linkage specificity was investigated. The protocol followed is described in the Material and Methods Chapter section XII.1. The results obtained are shown in Figure 9.

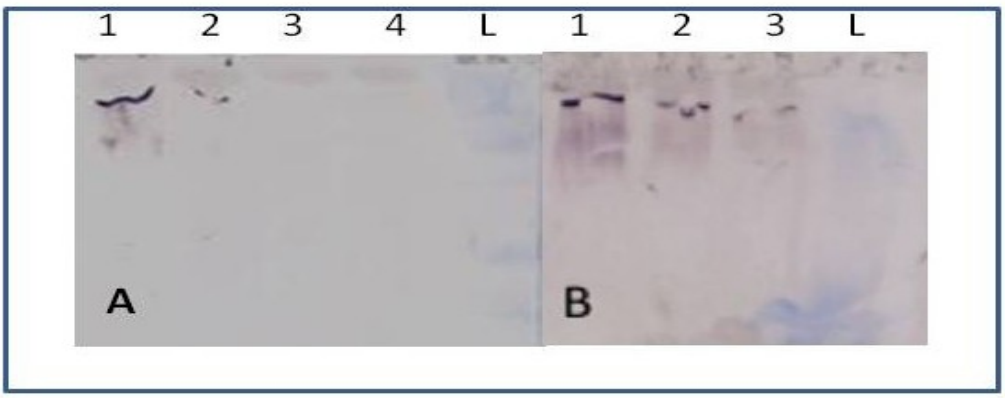


Figure 9- Determination of linkage specificity of BgaC. A Lane 1-Positive control (non digested fetuin); Lane 2- 0.04mg/ml of digested fetuin; Lane 3- 0.02mg/ml digested fetuin; Lane 4- 0.01mg/ml digested fetuin; L- Ladder, A was stained with Peanut agglutinin (PNA) which recognises the core disaccharide galactose β(1-3) N-acetylgalactosamine.; B- Lane 1-0.04mg/ml of digested fetuin; Lane 2- 0.02mg/ml digested fetuin; Lane 3- 0.01mg/ml digested fetuin; B was stained with Datura stamonium agglutinin (DSA) which recognises the core disaccharide galactose β(1-4) N-acetylglucosamine.

The analysis of the westerns blots shown in Figure 9, suggests that the enzyme encoded by *SPD0065* has linkage specificity for galactose β(1-3) N-acetylgalactosamine. In this assay, absence of bands means that the lectin binding site has been digested by the added enzyme but on the other hand if such linkage is not digested a band will appear. In Figure 9A, lanes 3 and 4, no band is visible. This is due to the action of the enzyme BgaC suggesting that the linkage galactose β(1-3) N-acetylgalactosamine is recognised and digested by BgaC. Lane 1, Figure 9A is the positive control and in this lane no BgaC was added and therefore it shows that if the linkage is not digested a band is visible. The same exact digestion is shown on the western blot of Figure 9 B however, in this blot, the digestion was stained with DSA which recognises galactose β(1-4) N-acetylglucosamine, which was previously identified to be substrate for BgaA. The appearance of bands in all 3 lanes demonstrates that this linkage is not a substrate for BgaC. This result suggested that BgaC and BgaA recognise and digest different linkages.

I.10 Localisation of β-galactosidase

In order to determine the subcellular localisation of β- galactosidase encoded by gene *SPD0065*, western blotting was used to analyse D39 cellular fractions. The SDS-PAGE gel was blotted and incubated with antibodies previously raised against the recombinant protein (the antibodies were raised as described in the Material and Methods Chapter, section XIV). The protocol followed is described in Chapter 2 section XV. The results are described in Figure 10.

The serum raised against BgaC reacted both with the cell wall and cell membrane fractions, but there is no evidence for cytoplasmic localisation, therefore the results point to a cell surface localisation.

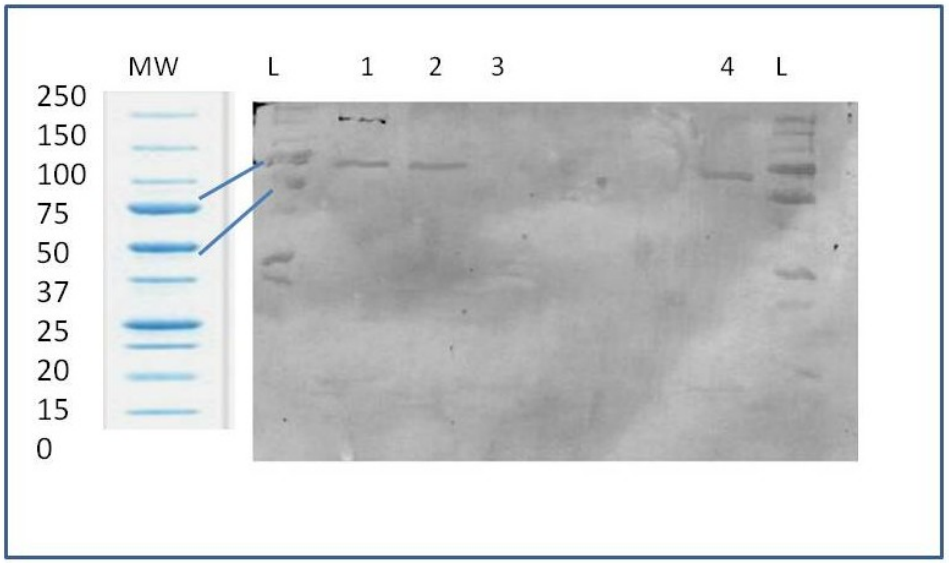


Figure 10-Determination of β-galactosidase subcellular localisation in D39 by immunoblotting. MW- Prestained Bio-Rad all blues protein markers; L-Protein marker; Lane 1- cell wall extract; Lane 2 -membrane extract; Lane 3- cytoplasm extract; Lane 4- whole cell extract; L-Protein marker. The blue lines indicate the visible bands of the protein marker.

I.11 Contribution of β-galactosidase C to the total β-galactosidase activity

In order to understand how this newly characterised β-galactosidase contributes to the total β-galactosidase activity, cell lysates of *S. pneumoniae* D39 strain mutated in both *SPD0065* and *SPD0562* genes was prepared following the protocol described in the Material and Methods Chapter section XIX. Subsequently, an enzyme assay was performed according to the method described in the Material and Methods Chapter section IV.2 using mutated strains, SPD0065M, SPD0562M, SPD0562K/65S, the parental strain and control D39 and a complemented strain SPD0065REV. The results (Figure 11) showed that there was detectable activity for both knock out mutants (SPD0065M, 10.43±0.57, n=6, SPD0562M, 9.15 ± 0.39, n=6, $p<0.05$) although, it was significantly lower when compared to the wild type D39 (19.03 ± 0.53, n=6, $p<0.05$). This indicates that in the absence of one of the β-galactosidases the other can metabolise the available substrate. However, an unexpected result was also observed. In the absence of both β-galactosidases, when using the cell lysate prepared from the double

knock-out mutant (SPD0562K/65S) there was still detectable activity ($p<0.05$) compared to the wild type. The complementation of mutant SPD0065M (SPD0065REV), showed a complete restoration of the activity against substrate ONP-β-galactopyranoside. The results obtained are shown in Figure 11.

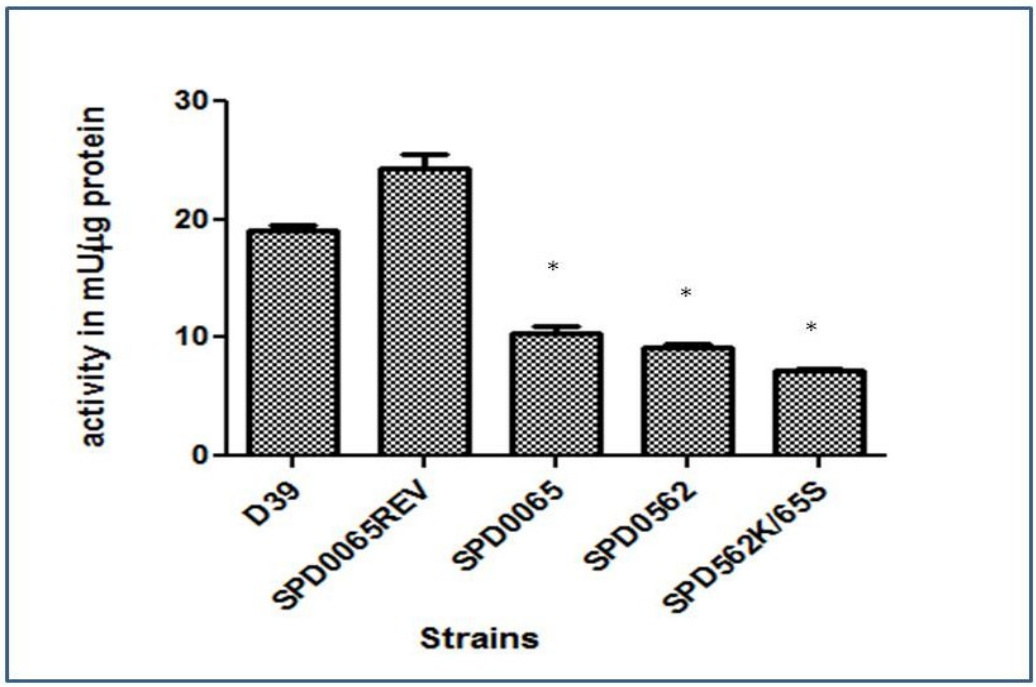


Figure 11- Contribution of β-galactosidase C (SPD0065) to total β-galactosidase activity. Each bar represents the mean of 6 datum points. The vertical bar represents the standard error of mean.* represents $p<0.05$ compared to the wild type D39.

I.12 Virulence of mutant SPD0065M

Isogenic pneumococcal knock out mutant SPD0065M was made in the University of Leicester by Dr Hasan Yesilkaya, prior the commencement of the work for this thesis.

The virulence of mutant SPD0065M was determined by intranasal infection. The protocol followed is described in the Material and Methods chapter, section XVI.3 and the collection of tissues and blood was done following the protocol described in the Material and Methods Chapter section XVI.4. In order to determine the importance of the studied protein for the virulence and infection process, two experiments were performed. In the first experiment disease signs were monitored after intranasal

infection, once the animals reached 2+ lethargic they were humanely sacrificed and that time recorded as survival time. The results obtained are shown in Figure 12. The median survival time of mice infected intranasally with SPD0065M (52 h ±18, n=10) was similar to the wild type infected cohort (48 h ±15, n=18) ($p>0.05$).

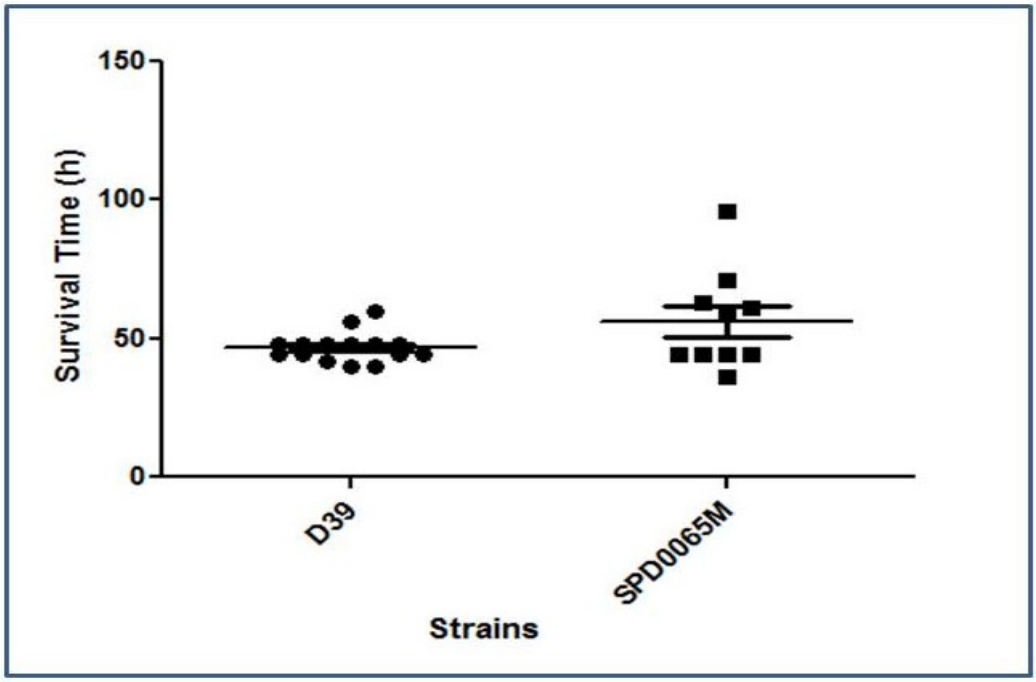


Figure 12- Survival of mice infected intranasally with (●) *S. pneumoniae* D39 (parental strain) and (■) isogenic mutant strain SPD0065M. For each group, 10 mice were used. Vertical bars indicate the standard error of mean. Horizontal bar indicates the median survival time. Each point corresponds to an individual animal.

In the second experiment lungs, nasopharynx and blood were collected from groups of mice every 12 hours after intranasal infection (with D39 or SPD0065M) and cfus in the homogenates were determined (see section II.1 from the Material and Methods Chapter). The number of pneumococci retrieved from blood, lungs and nasopharynx from mice infected are shown in Figure 13, Figure 14 and Figure 15 respectively.

While bacteraemia occurred sometime between 4 and 8 h after intranasal infection with the wild type, in the SPD0065M-infected the numbers at this earliest point were lower. Once in blood, SPD0065M grew with a similar rate as the wild type (Figure 13).

In addition, the progression of infection also was monitored in the lungs and nasopharynx Figure 14 and Figure15 respectively.

In the lungs (Figure 14), SPD0065M grew as well as the wild type at all time points ($p>0.05$). In the nasopharynx (Figure 15), SPD0065M colony counts (log_{10} 1.75 ±0.21, log_{10} 1.99 ±0.25, log_{10} 2.32 ±0.47; n=10, for 12, 24 and 48 h post-infection, respectively) were less than D39 at each time point (log_{10} 2.57 ±0.09, log_{10} 2.84 ±0.11, log_{10} 3.43 ±0.16; n=20, at 12, 24, and 48 h post-infection, respectively) ($p<0.01$ for 12 and 24 h, and $p<0.001$ for 48 h).These results suggested a link between β-galactosidase and the colonization of the nasopharynx.

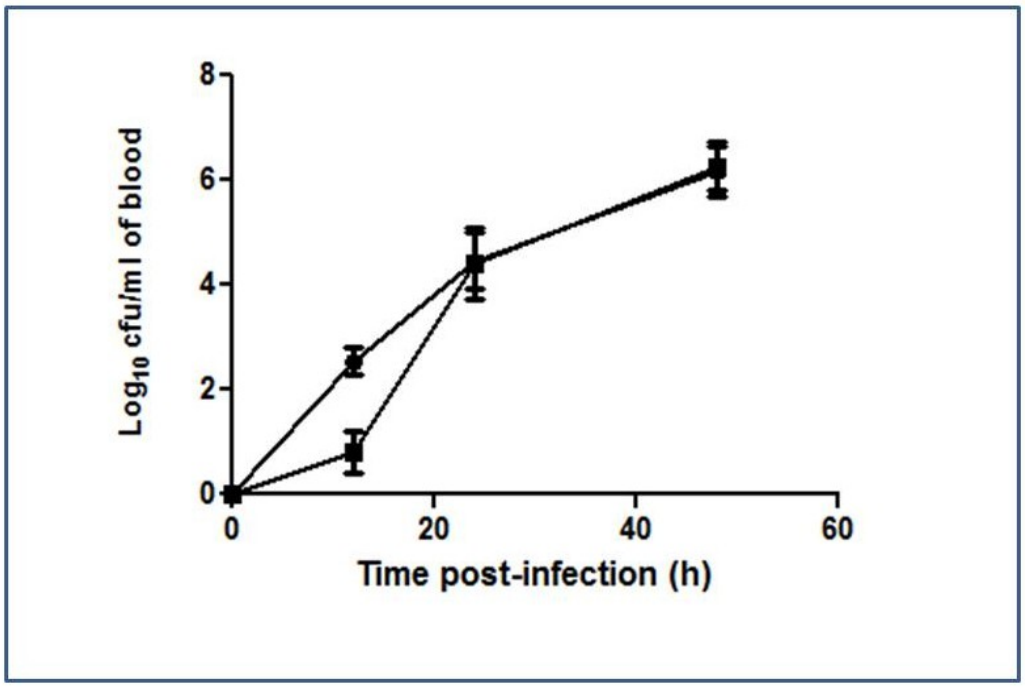


Figure 13- Numbers of pneumococci retrieved from blood of animals infected intranasally with (●) *S.pneumoniae* D39 (parental strain), (■) isogenic mutant strain SPD0065M. For each time point, between 6 to 10 mice were used for SPD0065M (■) and 15 to 20 for D39 (●). Vertical bars indicate the standard error of the mean.

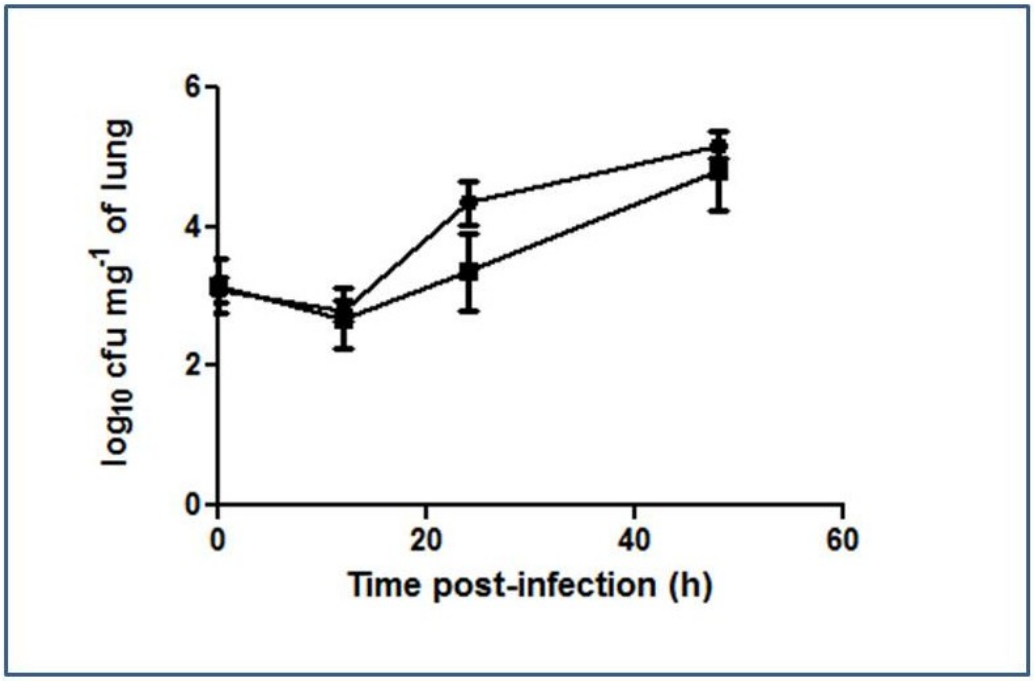


Figure 14- Numbers of pneumococci retrieved from lung homogenate of animals infected intranasally with (●) *S.pneumoniae* D39 (parental strain), (■) isogenic mutant strain SPD0065M. For each time point, between 6 to 10 mice were used for SPD0065M (■) and 15 to 20 for D39 (●).Vertical bars indicate the standard error of the mean.

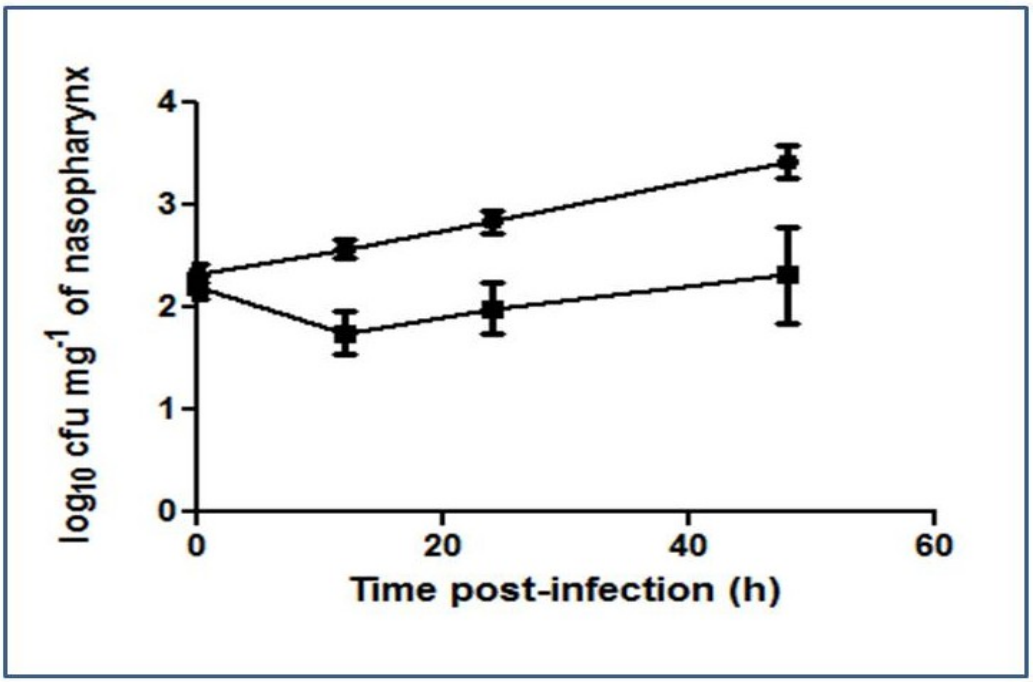


Figure 15- Numbers of pneumococci retrieved from nasopharynx homogenate of animals infected intranasally with (●) *S. pneumoniae* D39 (parental strain), (■) isogenic mutant strain SPD0065M. For each time point, between 6 to 10 mice were used for SPD0065M (■) and 15 to 20 for D39 (●). Vertical bars indicate the standard error of the mean.

I.13 Virulence of Double knock out mutant SPD562K/65S

Because the level of bacteraemia observed with SPD0065M was lower than the one observed for wild type in the early hours after infection it was important to determine the virulence of the double knock-out mutant SPD0562K/65S. In this case, a slight increase on the survival time was observed (Figure 16) with the median survival time of mice infected intranasally with SPD0562K/65S being 59 ± 4 h, n=7 and the wild type infected cohort being 45 ± 2 h, n=10 ($p<0.01$).

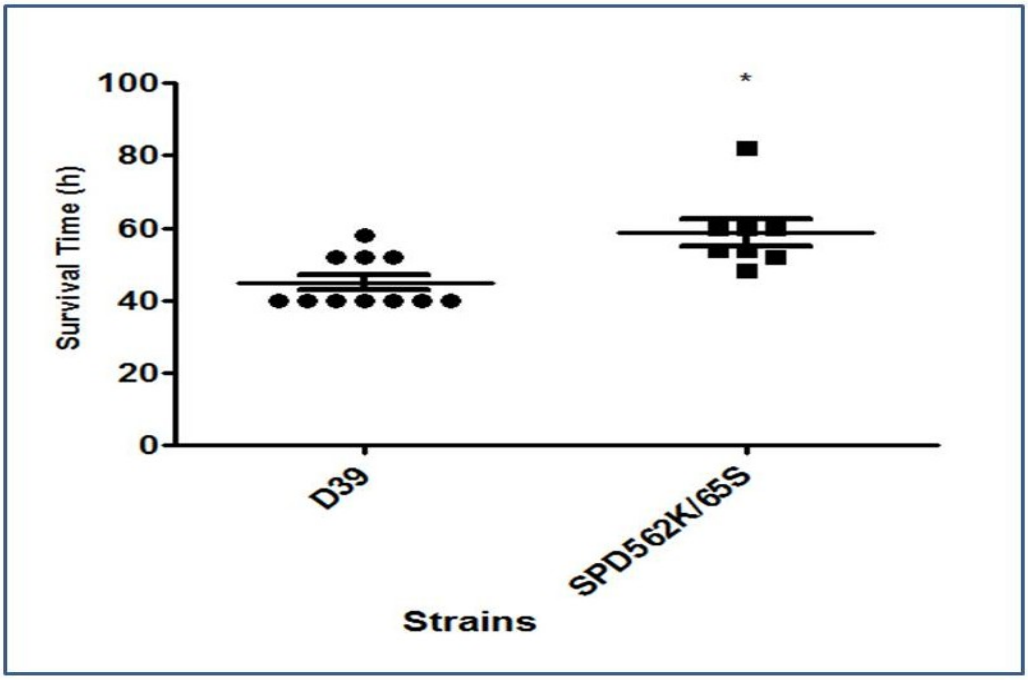


Figure 16- Survival of animals of animals infected intranasally with. (●) *S. pneumoniae* D39 (parental strain) or (■) isogenic mutant strain SPD562K/65S. For each group 7 mice were used for SPD0562K/65S (■) and 10 mice for D39 (●). Horizontal bars indicate the median survival time.* indicates $p<0.01$.

Bacteraemia was also followed (Figure 17). For SPD562K/65S the colony counts ($\log_{10}$ 0.23 ±0.21, $\log_{10}$ 2.59 ±0.30, $\log_{10}$ 5.20 ± 0.90; n=10, for 12, 24 and 36 h post-infection, respectively) were lower than D39 ($\log_{10}$ 0.92 ± 0.27, $\log_{10}$ 4.93 ± 0.40, $\log_{10}$ 6.70 ± 0.60; n=10, for 12, 24 and 36 h post-infection, respectively) ($p>0.05$ for 12 and 36h, and $p<0.01$ for 24h).

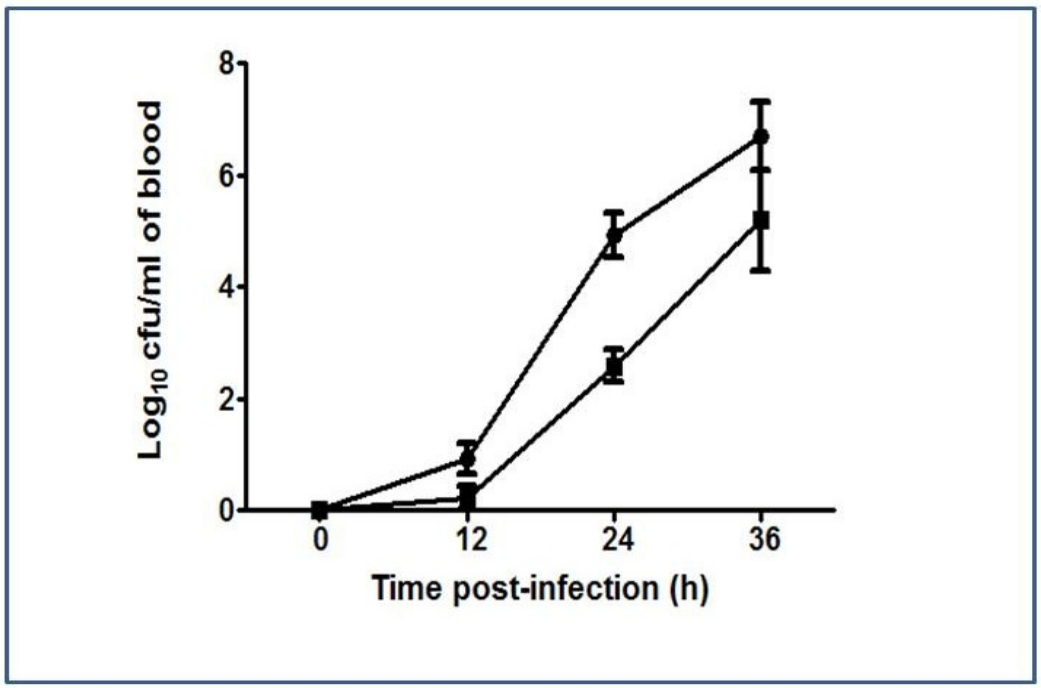


Figure 17- Numbers of pneumococci retrieved from blood of animals infected intranasally with (●) *S. pneumoniae* D39 (parental strain) or (■) isogenic mutant strain SPD565K/65S M.For each time point 10 mice were used for SPD0562K/65S M (■) and D39 (●). Vertical bars indicate the standard error of mean.

I.14 Immunisation studies

In order to determine the protective potential of this new β−galactosidase, an immunisation assay was performed as described in the Material and Methods Chapter section XIV. 10 mice were immunised every 2 weeks and their blood collected just before each immunisation. After 8 weeks the mice were challenged with *S. pneumoniae* D39 and the disease signs were observed. The results obtained are shown in Figure 18.

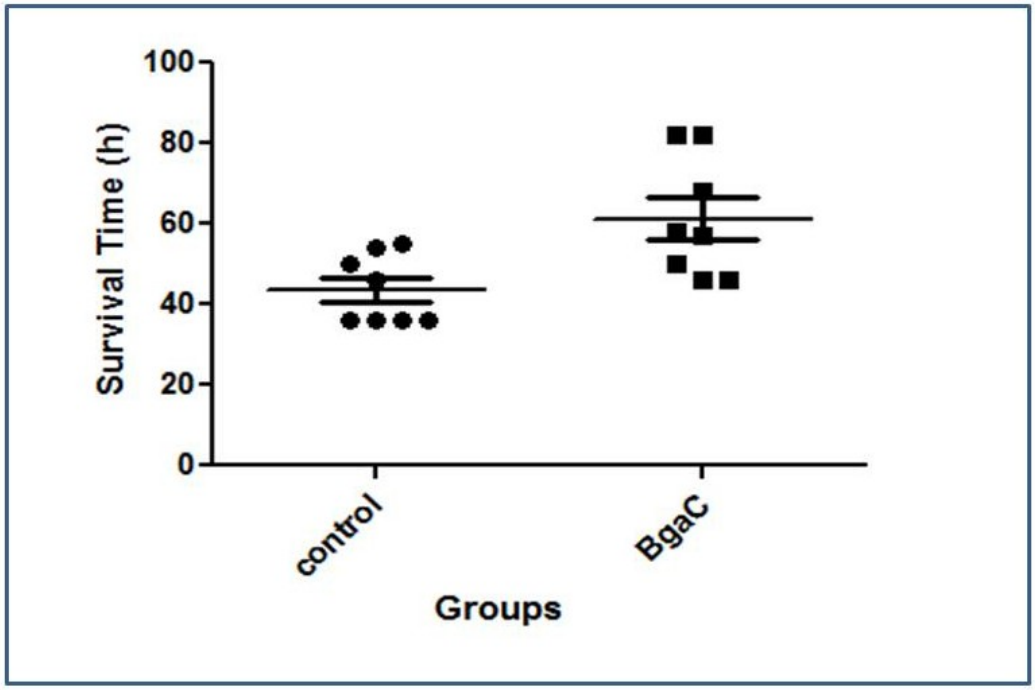


Figure 18- Survival of mice challenged intranasally with *S.pneumoniae* D39 after previously immunised with (■) recombinant protein β-galactosidase in alum. (●) Control- mice were given alum in PBS. 8 mice were used, for the immunised and the control groups. Horizontal bars indicate the median survival time. Vertical bars indicate the standard error of the mean. Each point corresponds to one animal.

The analysis of the results using the Mann Whitney test showed a significant difference between the groups. The control group had a median survival time of 43.6 ± 3.0h whereas the immunised group has a median survival of 61.3 ± 5.2 h ($p<0.05$).

At the same time, the colony forming units in blood were measured (Figure 19). The values observed for this experiment are very similar to the values obtained for the experiments performed with SPD0065M mutant (Figure 15) with the numbers of pneumococci in blood significantly lower ($p<0.01$) than in the controls at 12h post-infection (two way ANOVA, Bonferroni post-test) however, there was no significant difference for times 0, 24 and 48 hours ($p>0.05$). These results seem to indicate that immunisation with recombinant protein β-galactosidase delays the appearance of pneumococci in blood.

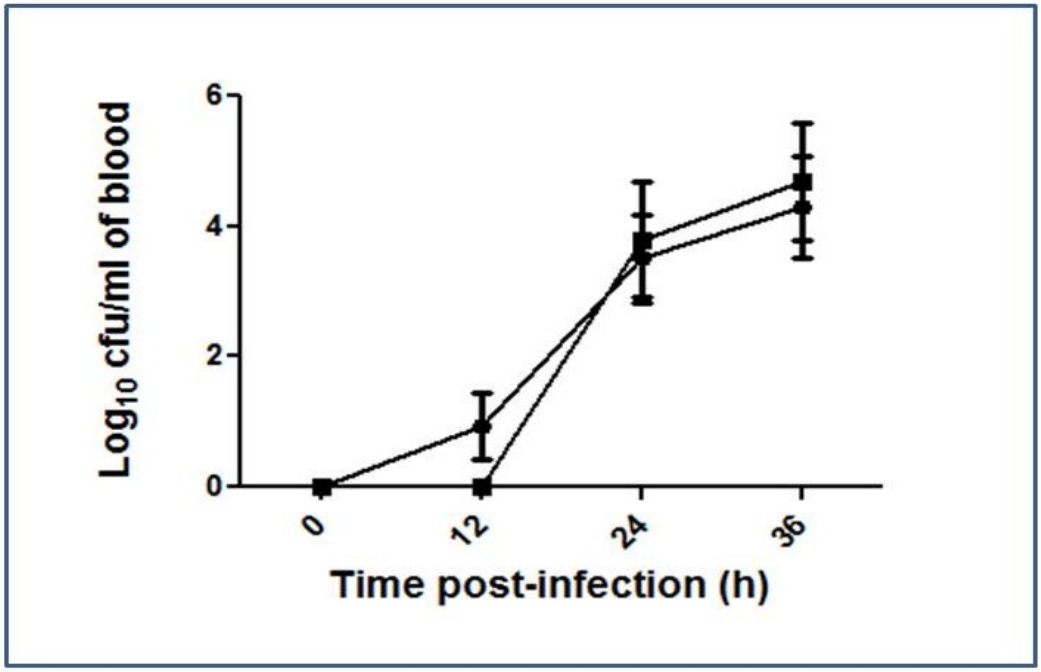


Figure 19- Pneumococci numbers retrieved from blood of infected animals. (●) represents control group infected with *S. pneumoniae* D39, (■) represents immunised group infected with *S. pneumoniae* D39. For each time point, 10 mice were used for both groups. Vertical bars indicate the standard error of the mean.

I.15 Activity against ONP-β-galactopyranoside in pneumococci of different serotypes

The knock out mutants were also made in different serotypes, as described in the Material and Methods Chapter, section XVIII to determine if the β-galactosidase function established for serotype 2 D39 could be extended to different serotypes. For this, knock-out mutants and the wild types of a serotype 6B strain IO11966 and a serotype 19F ST type 177, strain OXC1261 were tested for their activity against the synthetic substrate ONP-β-galactopyranoside.

The activities for the cell lysates was calculated using the Lambert Beer law as before. The results obtained are shown in Figure 20. The results suggested that the strain IO11966 serotype 6B exhibits a similar pattern of activity to D39 since there was a significant decrease in activity in the knock-out mutant made in serotype 6B background when compared to the activity obtained with its wild type ($p<0.005$).

However, the same pattern was not observed for the serotype 19F. There was no difference in the activity of the knock-out mutant when compared to its wild type ($p>0.05$). The results showed that the homologous enzyme of SPD0065 is most likely synthesised by serotype 6B, however it is likely that serotype 19F does not synthesise such enzyme. Also it is noteworthy that the activity for the wild type of serotype 19F was lower than the one of wt 6B. It is also possible that another β-galactosidase compensates for the lack of this enzyme.

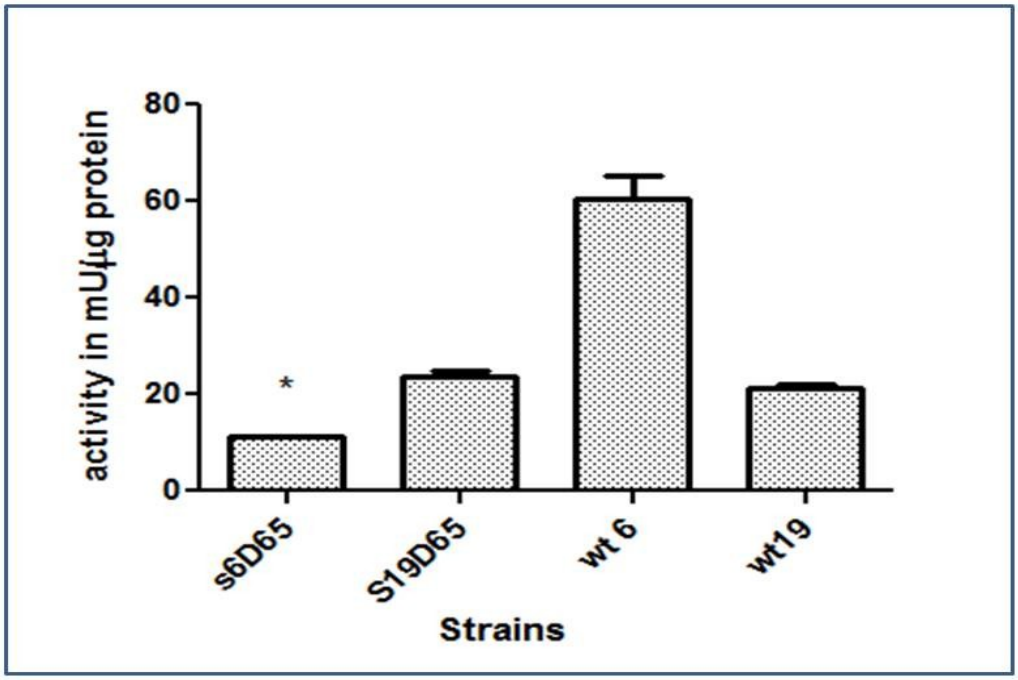


Figure 20 – Activity against substrate ONP-β-galactopyranoside of cell free extracts of knock-out mutants and serotypes 6B and 19F wild type. S6D65- knock-out mutant in serotype 6B background; S19D65- knock-out mutant in serotype 19F background; wt 6B- wild type serotype 6B; wt19- wild type serotype 19. Each bar is the mean of three independent experiments. The vertical bars indicate the standard error of mean.*- represents $p < 0.05$.

F RESULTS

I. Overview

This chapter starts with the description of the expression and purification of the enzyme encoded by gene *SPD00247*. Once the purified recombinant protein was obtained its subcellular localisation and substrate specificity was determined, followed by kinetic characterisation. The enzyme encoded by *SPD00247* is a 6-phospho-β-glucosidase. In addition to the characterisation of this enzyme, virulence studies using the knock out mutant (SPD0247M) and immunogenic properties were investigated, as were its possible contribution to biofilm formation.

I.1 Synthesis of protein encoded by *SPD0247*

Expression of the recombinant protein with a polyhistidine tag product of construct containing the gene *SPD0247* in *E.coli* BL21 DE3 was performed following the protocol described in Section VIII.5 of the Material and Methods Chapter. The analysis of the resulting protein by SDS-PAGE showed that it had a molecular weight of approximately 50 kDa, which is in line with the predicted molecular weight from the deduced amino acid sequence (Figure 1).

The samples were prepared according to the protocol described on the Material and Methods Chapter, section VIII.6. The expression of the recombinant protein was first optimised by executing the expression protocol at three different temperatures, 25°C, 30°C and 37°C. Following the expression, the pellets were frozen at -80°C.The resultant protein mixture was analysed by SDS-PAGE and the results are shown in Figure 1. In Figure 1A the expression of recombinant protein encoded by gene *SPD0247* can be seen by more intense staining in lanes 3, 5, 7 and 9. Additionally it was also observed that the expression of the recombinant protein improved over time. The biggest yield is observed at 4h after induction of the expression (lane 9 Figure 1A). Lanes 2, 4, 6 and 8 show the proteins from *E.coli* containing the empty vector and as it can be seen there is no band of the same molecular weight as the recombinant protein. In Figure 1B, in lanes 7 and 9 a 50 kDa band is visible, corresponding to the recombinant protein at 30°C. These lanes have samples from 1 and 2 hours after induction of expression. In Figure 1C in lanes 2, 4 and 6 the recombinant protein expression after 3, 4 and 5 hours of the induction is visible. However, a large quantity of contaminant proteins also is visible. Once again, lanes 1, 3 and 5 show the proteins from *E.coli* containing the empty vector; no band is visible on the 50 kDa range. As it can be seen by the more intense staining in lanes 2, 5, 7 and 9 in Figure 1D, the biggest yield of protein was observed at a temperature of 25°C. This is the temperature

that showed a constant increase in expression through time. Followed the analysis of these results from this point forward all expression procedures were carried out at the temperature of 25ºC.

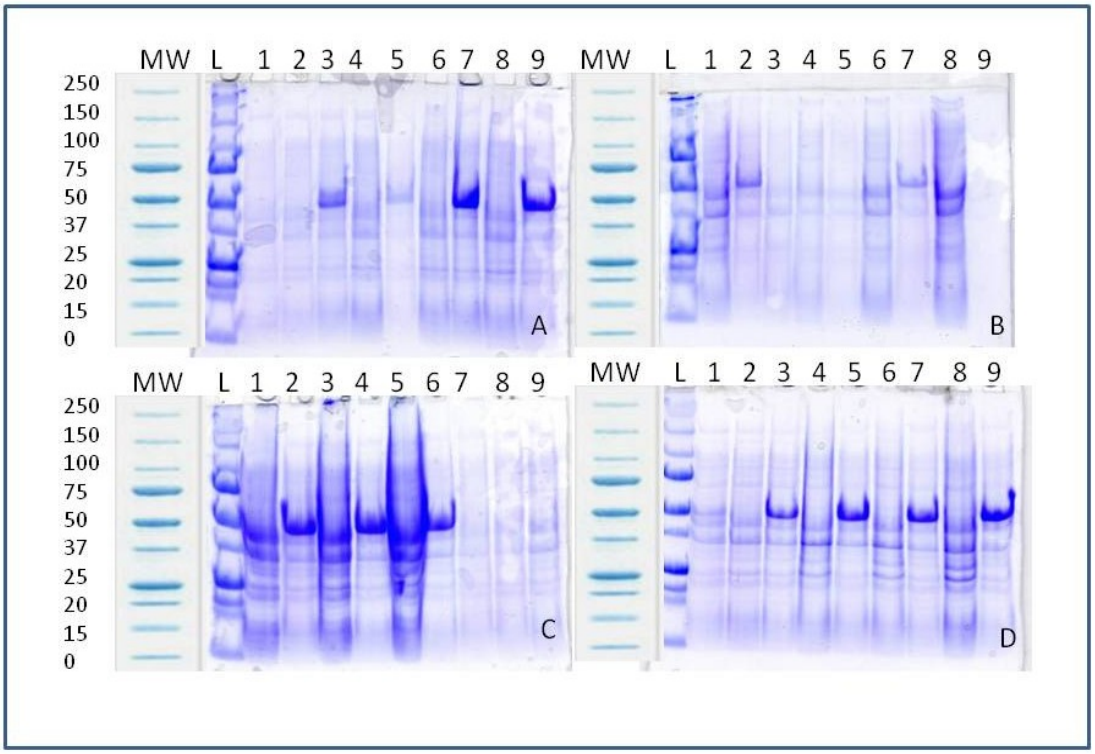


Figure 1-Optimisation steps done to determine the optimum temperature of expression of recombinant protein encoded by gene *SPD0247*. **A**-MW- Prestained Bio-Rad all blues protein marker; L- protein molecular weight marker; Lane1-construct containing SPD0247 before induction pVCT1; Lane 2-Empty vector (pVT3) before induction; Lanes 3, 5, 7, 9–pVCT1 1,2,3 and 4 hours after induction respectively; Lanes 4, 6 and 8 –pVT3 1,2 and 3 hours after induction respectively; The cell extract was prepared after growth of bacteria *E.coli* BL21 at 37ºC **B**–MW- Prestained Bio-Rad all blues L- protein molecular weight marker; Lanes 1and 3- PVT3 4 and 5 hours after induction respectively at 37°C; Lane 2-pVCT1 5 hours after induction; Lane 4-pVCT1 before induction at 30 ºC; Lane 5-pVT3 before induction at 30 ºC; Lanes 6 and 8-pVCT1 1 and 2 hours after induction respectively at 30 ºC; Lanes 7 and 9-pVT3 1 and 2 h after induction respectively at 30 ºC **C**- MW-Prestained Bio-Rad all blues; L- protein molecular weight marker; Lanes 1,3 and 5–pVT3 after induction 3,4 and 5 hours respectively at 30 ºC; Lanes 2,4 and 6–pVCT1 3, 4 and 5 hours after induction at 30 ºC; Lane 7-pVT3 before induction at 25 ºC; Lane 8–pVCT1 before induction at 25 ºC; Lane 9–pVT3 1h after induction at 25 ºC; **D**– MW- Prestained Bio-Rad all blues L– protein molecular weight marker; Lanes 1, 3, 5, 7 and 9–pVCT1 after induction 1,2,3,4 and 5 hours respectively at 25ºC; Lanes 2,4,6 and 8–pVT3 2,3,4 and 5 hours after induction at 25ºC.

I.2 Determination of protein solubility

In order to understand if the protein under study was in the soluble or insoluble fractions in *E.coli* cell extracts were prepared following the protocols described on Material and methods chapter, section VIII.7. The samples were then analysed by SDS-PAGE gels and the results are shown in Figure 2.

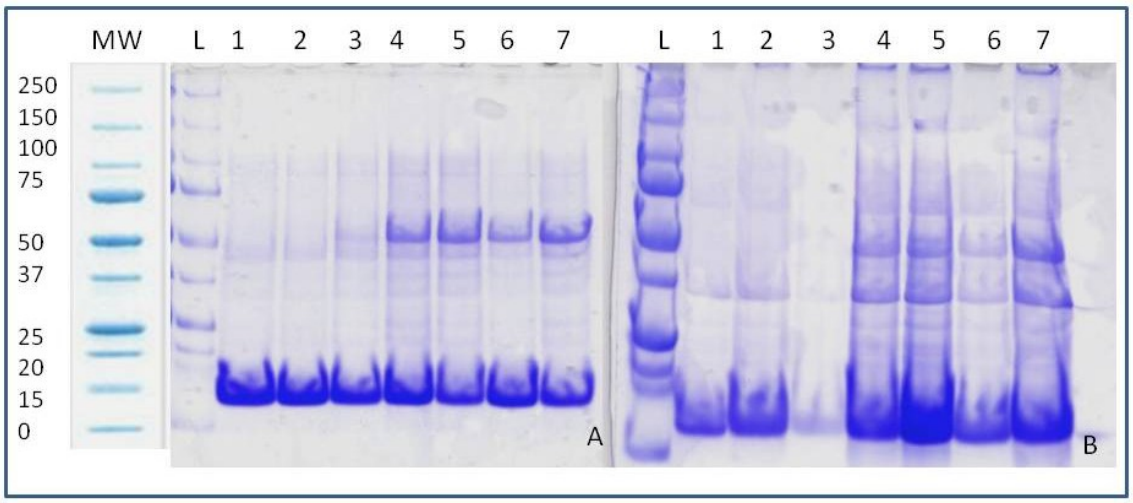


Figure 2- Determination of protein solubility **-A**-Soluble expression - MW- Prestained Bio-Rad all blues protein marker; L-protein molecular weight marker; 1-empty vector before induction (PVT3) at 25ºC; 2- construct containing SPD0247 gene before induction (PVCT1) at 25º C; 3- PVCT1 1h after induction at 25ºC; 4- PVCT1 2h after induction at 25º C; 5 – PVCT1 3h after induction at 25 ºC; 6- PVCT1 4h after induction at 25 ºC; 7- PVCT1 5h after induction at 25ºC **B** –Insoluble expression; L- protein molecular weight marker; 1- PVT3 before induction at 25ºC; 2- PVCT1 before induction at 25º C; 3- PVCT1 1h after induction at 25ºC; 4- PVCT1 2h after induction at 25ºC; 5 – PVCT1 3h after induction at 25ºC; 6- PVCT1 4h after induction at 25ºC ; 7- PVCT1 5h after induction at 25ºC

In Figure 2A, (lanes 4 to 7), a very intense band is visible around 50 kDa range suggesting that the recombinant protein is in the soluble fraction. In Figure 2 B, a band at the 50 kDa region is also visible (lanes 4 to 7), however, it is a much fainter band than in Figure 2A , indicating most likely that the band is a result of cross-contamination from the recombinant protein in soluble fraction. From this point onwards all the pellets resulting from expression of pVCT1 in *E.coli* were prepared assuming that the recombinant protein

was in the soluble fraction. A 15 kDa band is also visible in both Figure 2A and B, however the fact that this protein is present in lanes 1 and 2 which contains the products of expression of an empty construct and non induced sample suggests that this is some other over-expressed protein but not the product of pVCT1.

I.3 Purification

Before starting the purification protocol, pellets from *E.coli* expressing plasmid pVCT1 were prepared and treated under mild conditions, 50mM Na_2PO_4, 300mM NaCl, pH 7.0, in order to isolate the native recombinant protein (see section IX.1 from the Material and Methods Chapter). The use of more stringent conditions, i.e. the addition of 6M guanidinium, would enhance the solubility of the recombinant protein, however at the same time it would denature the protein. The purification protocol was done as described on the Material and Methods Chapter, section IX.2 A column containing immobilised cobalt was used to enable binding of the polyhistidine tag from the recombinant protein. The eluate was collected in 500 µl fractions, from which 10 µl from the first nine fractions were analysed by SDS PAGE. A gel in which the fractions were analysed by SDS-PAGE can be seen on Figure 3.The purified protein can be seen in lanes 3 and 4 (Figure 3). This was expected, since it was indicated by the manufacturers that the majority of the polyhistidine-tagged protein should be recovered in the first 4 ml of eluate. Fractions 3 and 4 were quantified using a Bradford assay (see Material and Methods Chapter, section IX.3). Using the standard curve equation, the protein concentration of each fraction was determined, Fraction 3 had a concentration of 2.83 mg/ml and Fraction 4 had a concentration of 3.75 mg/ml.

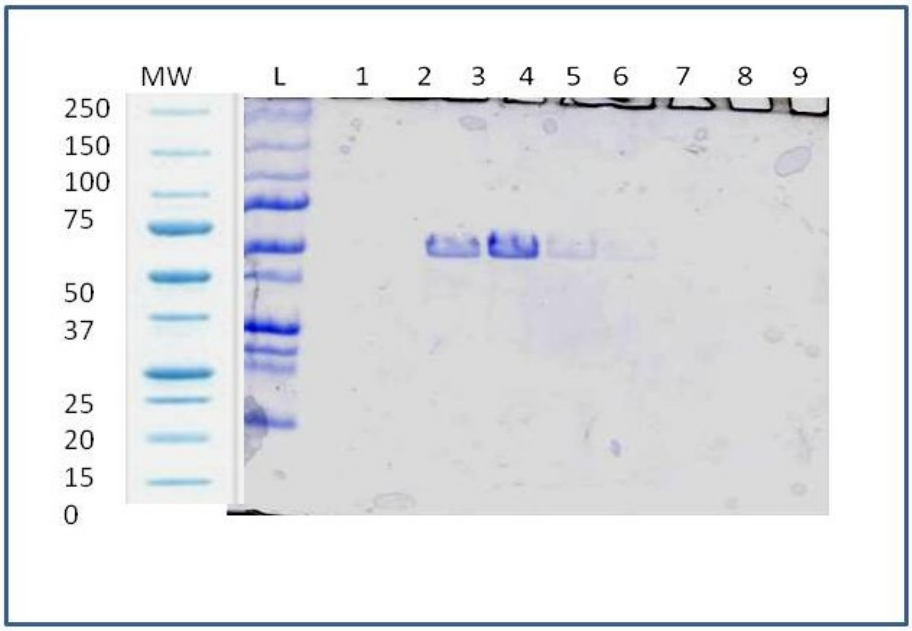


Figure 3-Analysis of purified protein by SDS-PAGE. MW- Prestained Bio-Rad all blues protein marker; L-protein molecular weight marker; Lanes 1 to 9 – sequential fractions eluted from a sepharose-Co column. Fractions 3 and 4 contain the purified recombinant protein.

I.4 MALDI TOF analysis for protein identification

In order to identify the purified protein as being a putative 6-phospho-β-glucosidase, it was analysed by matrix assisted laser desorption ionization–time-of-flight (MALDI-TOF) mass spectrometry by PNACL at the University of Leicester. The enzyme was identified as being a 6-phospho-β-glucosidase (see Figure 4, Appendix 1).

I.5 Substrate specificity

The substrate specificity was determined following the protocol described in the Material and Methods Chapter section X. The enzyme was assayed with eight synthetic substrates and the results are shown in Table 1 .

Table 1- Activity of 6-Phospho-β-glucosidase against 4-methylumbelliferyl- and pNP-linked substrates. Minus sign indicates the absence and plus sign indicate the presence of activity against each substrate.

Substrate	Activity
ONP-β-D-galactopyranoside	-
4-Methylumbelliferyl-N-acetyl-α-D-glucosaminide	-
4-Methylumbelliferyl-β-D-glucopyranoside	-
4-Methylumbelliferyl-α-D-glucopyranoside	-
O-Nitrophenyl-α-D-galactopyranoside	-
pNP-α-D-glucoside-6-phosphate	-
pNP-α-D-mannose-6-phosphate	-
pNP-α-D-galactose-6-phosphate	-
pNP-β-D-glucose-6-phosphate	+

The purified protein only showed activity against the synthetic substrate pNP-β-D-glucose-6-phosphate. This enzyme has an activity of 3.37μU/min/μg of protein against this substrate. Also it is noteworthy that even in the presence other α-linked glucoside substrates there was no detectable activity. There was also no detectable activity in the presence of 4-methylumbelliferyl-β-D-glucopyranoside and 4-methylumbelliferyl-α-D-glucopyranoside, indicating that the presence of a phosphate group was also a requirement for enzyme-substrate recognition. The data suggest that this enzyme is a 6-phospho-β-glucosidase.

I.6 Kinetic characterisation of enzyme 6-P-β-glucosidase

After pNP-β-D-glucose-6-phosphate was determined to be a suitable synthetic substrate for the recombinant enzyme, the next step was the determination of its kinetic parameters. Kinetic experiments were designed and performed following the protocol described on Material and Methods, section XII. Kcat was not calculated because there is no structural information for the enzyme; therefore assumptions of number of catalytic sites would have to be made. Following the confirmation that this enzyme obeys Michaelis Menten kinetics, the experimental values were analysed. The graphs and calculations were performed in GraphPad Prism version 5.00 for Windows.

The first step of the analysis was the adjustment of a non linear regression that is a best fit to the experimental values. From there is possible to obtain Km and V_{max} values. The results obtained for 6-P-β-glucosidase when tested against the synthetic substrate 6-phospho-β-glucoside are plotted in the Figure 4. The results were also linearised using Lineweaver Burk (Figure 5) and Eadie Hofstee approaches (Figure 6).

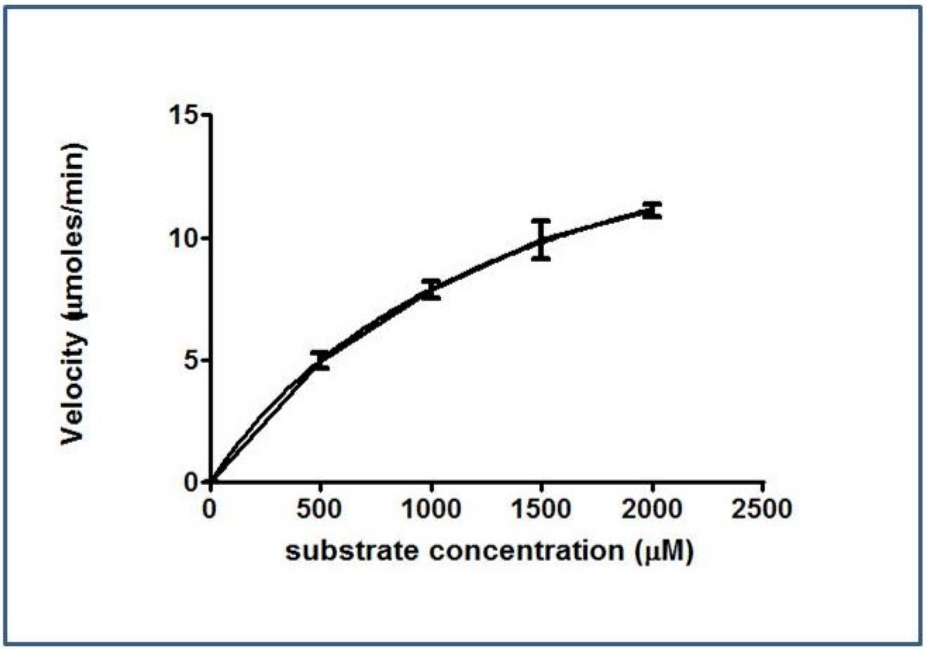


Figure 4- Determination of Michaelis Menten Kinetic parameters against substrate 6-phospho-β-D-glucoside. Each datum point represents the mean of three independent experiments. Vertical bars indicate standard deviation. K_M and V_{max} were calculated using non linear regression using the program GraphPad Prism version 5.00 for Windows, (GraphPad Software, San Diego California USA, www.graphpad.com).

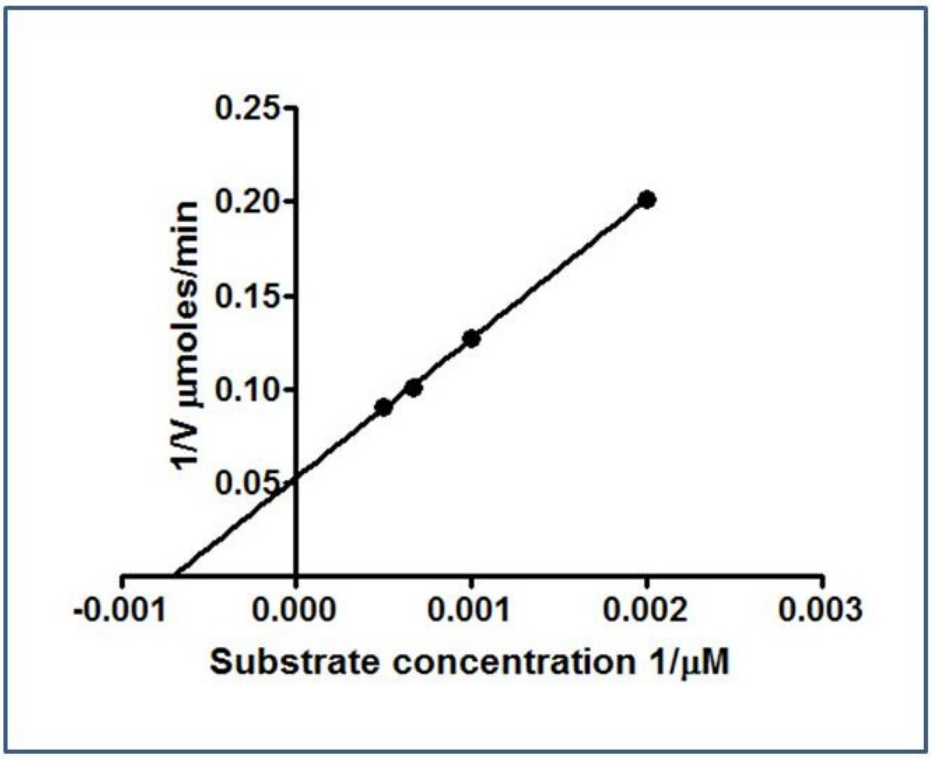


Figure 5- Graphic determination of K_M and V_{max} parameters using a Lineweaver Burk plot. Each datum point corresponds to the mean of three independent experiments.

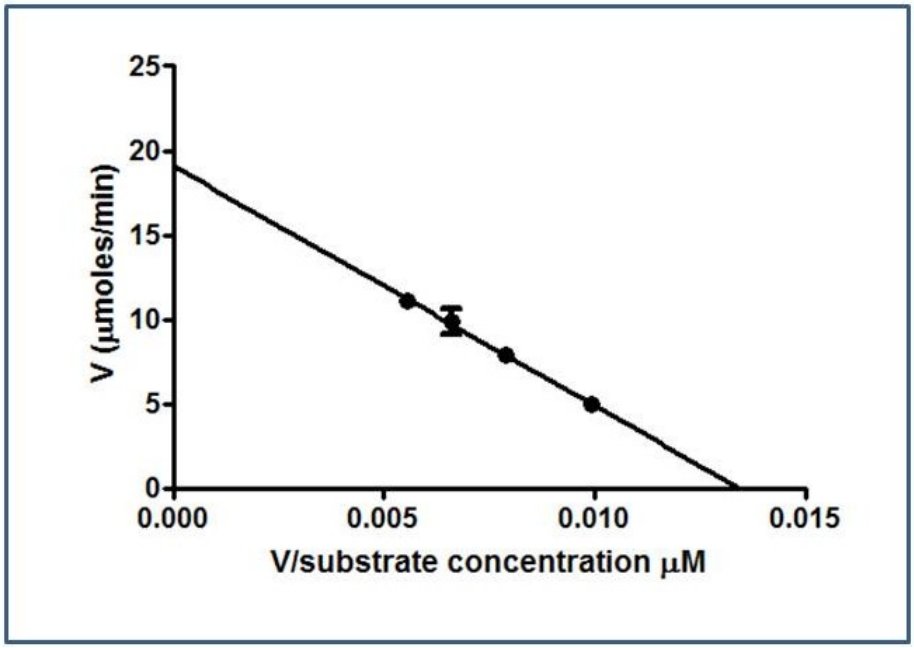


Figure 6-Graphic determination of K_M and Vmax parameters using Eadie Hofstee plot. Each datum point corresponds to the mean of three independent experiments. The vertical bars represent the standard deviation.

The results obtained from all three methods are summarised in Table 2.

Table 2- Kinetic parameters determination using Michaelis Menten approach and two linearisation methods (Lineweaver Burk and Eadie-Hofstee).

	Interpretation of Kinetic parameters based in graphic representation			**Experimental Values**	
Approaches	**Intercepts the x axis**	**Intercepts the y axis**	**Slope**	**Km (μM)**	**Vmax (μmol.min^{-1}.mg^{-1})**
Michaelis Menten (Figure 5)	–	–	–	1416±319	0.795±0.09
Lineweaver Burke (Figure 6)	-1/Km	1/Vmax	Km/Vmax	1418	0.787
Eadie Hofstee (Figure 7)	–	Vmax	-Km	1419	0.788

I.7 Subcellular localisation of protein SPD0247

Localisation of recombinant protein *SPD0247* was done by means of immunoblotting using polyclonal serum raised against the enzyme. The cell extracts were prepared according the protocol described on Material and Methods Chapter, section XV, and the Western blot was performed following the protocol described in Material and Methods Chapter, section XV.1.

It was established that the protein is membrane associated, since the antibodies raised against this protein only reacted with the extract containing the membrane proteins (lane 2 Figure 7). The band appearing on lane 4 corresponds to the positive control D39 whole cell

extract. The fractionation quality was assessed by determining activities for lactate dehydrogenase, known for having intracellular activity, and neuraminidase which has a cell wall localisation. The results showed that the fractionation was of good quality with activity for lactate dehydrogenase observed in the cytoplasmic fraction, but not in the other two fractions. Neuraminidase activity was seen in the cell wall fraction but not on the other two fractions, and finally for the membrane fraction no activity was seen for any of the above enzymes, denoting a free of contaminations fraction.

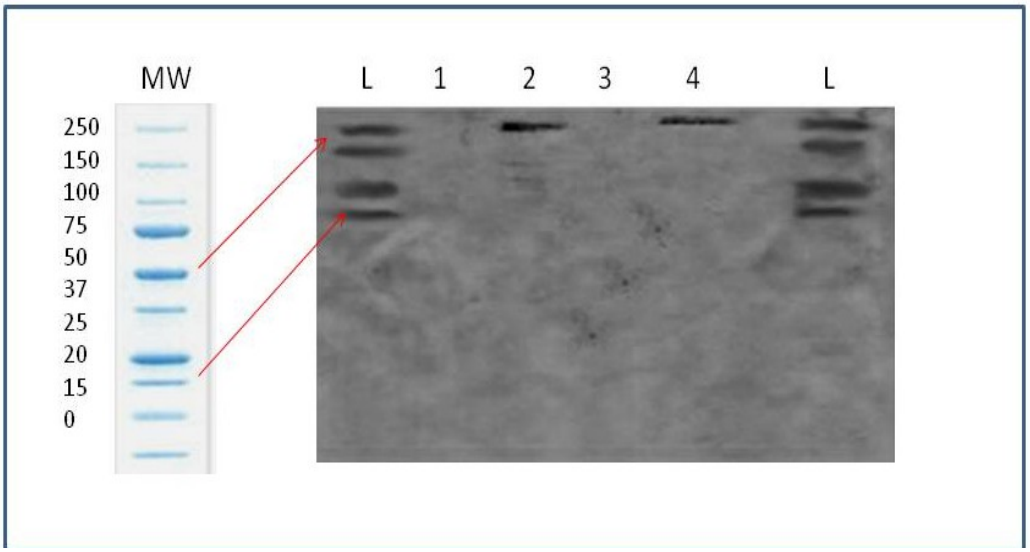


Figure 7- Determination of 6-phospo-β-glucosidase subcellular localisation by immunoblot. Cellular fractions of pneumococcal D39 MW- Prestained Bio-Rad all blues protein marker; L-Protein Marker; Lane 1- cell wall extract; Lane 2 -membrane extract; Lane 3- cytoplasm extract; Lane 4- Whole cell extract from D39; L-Protein Marker. The red arrows indicate the visible bands of the protein marker.

I.8 Determination of virulence of pneumococcal mutant SPD0247M

Isogenic pneumococcal knock-out mutant was made in the University of Leicester by Dr Hasan Yesilkaya, prior the commencement of the work for this thesis.

The virulence of pneumococcal mutant SPD0247M was determined by intranasal infection. The followed protocol is described in the Material and Methods Chapter section XVI.3 and the collection of tissues and blood was done following the protocol described in

the Material and Methods Chapter XVI.4 In order to determine the importance of the studied protein for the virulence and infection process, two experiments were performed. In the first experiment, disease signs were monitored after intranasal infection. When the animals reached 2+ lethargic they were humanely sacrificed and the time recorded as survival time. The results obtained are shown on Figure 8. SPD0247M-infected mice survived significantly longer (72h ±12, n=6) than the D39-infected cohort (48 h ±15 n=18) ($p<0.01$).

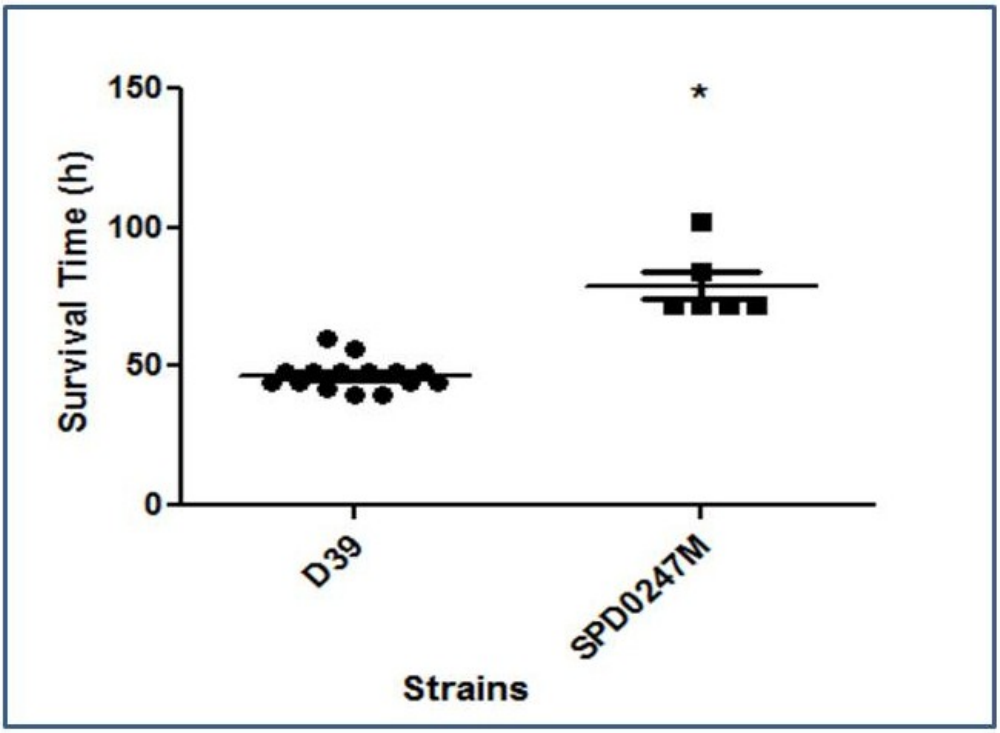


Figure 8-Survival of mice infected intranasally with (●) *S.pneumoniae D39* (parental strain) and (■) isogenic mutant strain SPD0247M. For each group, 6 mice for SPD0247M (■) and 15 for D39 (●) were used. Vertical bars indicate the standard error of the mean. The horizontal bar indicates the median survival time. Each point corresponds to an individual animal.

In the second experiment lungs, nasopharynx and blood were collected from groups of mice every 12 hours after intranasal infection. The number of pneumococci retrieved from blood, lungs and nasopharynx from mice infected with D39 and SPD0247M are shown on Figure 9, Figure 10 and Figure 11 respectively.

In blood there was no significant difference in the pneumococci numbers retrieved when compared at times 0 and 12h ($p>0.05$), however there was significant difference at times 24h and 48 hours ($p<0.05$). In the lungs at 24 h (log_{10} 2.18 ±0.19, n=10) and 48 h (log_{10} 3.38 ±0.37, n=10) post-infection the colony counts were significantly lower than for the wild type (log_{10} 4.34 ±0.3 and 5.17 ±0.19, n=20, for 24 h and 48 h, respectively) ($p<0.001$), however there was no significant difference at times 0 and 12 hours ($p>0.05$, Figure 10). SPD0247M colonised the nasopharynx as well as the wild type with no significant differences ($p>0.05$) throughout the experiment (Figure 11). These results suggested that there may be a tissue specific action for the enzyme being investigated.

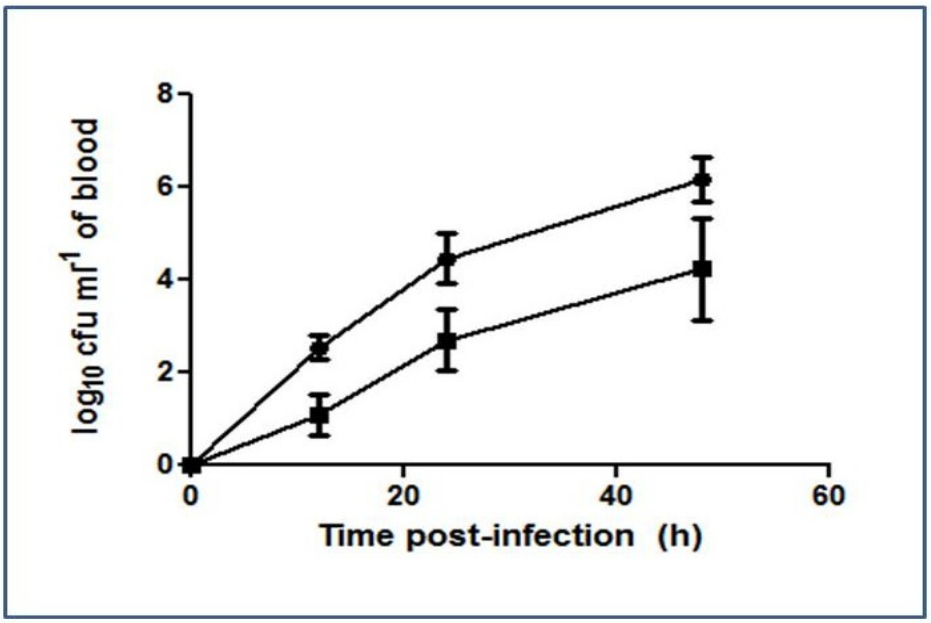


Figure 9- Number of pneumococci per ml retrieved from blood of intranasally infected mice. (●) *S.pneumoniae* D39 (parental strain); (■) Isogenic mutant strain SPD0247M. For each time point, between 6 to 10 mice for SPD0247M (■) and 15 to 20 for D39 (●) were used. Vertical bars indicate the standard error of the mean.

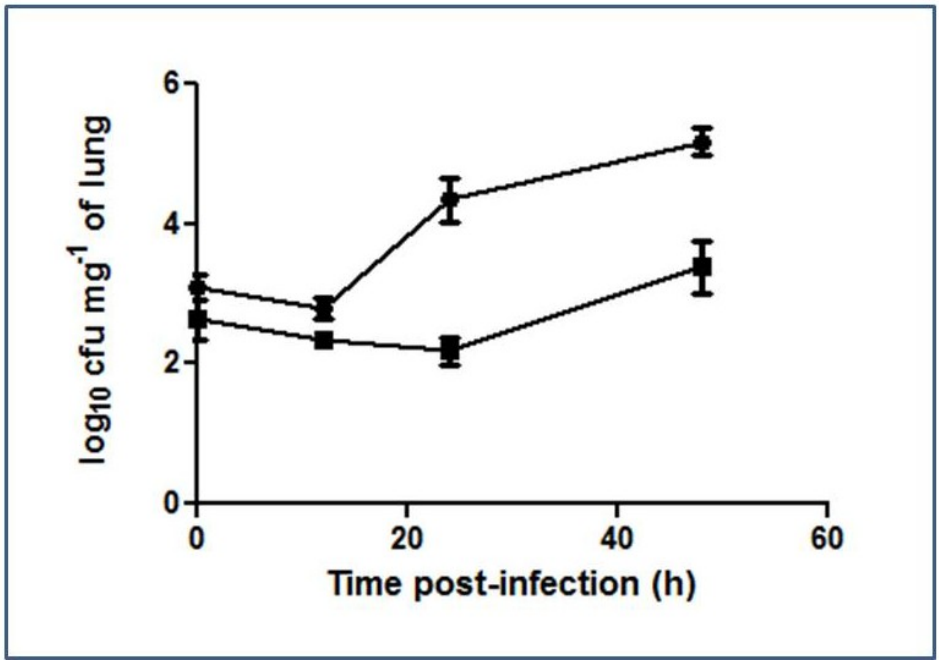


Figure 10- Number of pneumococci per mg retrieved from lungs of intranasally infected mice. (●) *S. pneumoniae* D39 (parental strain); (■) Isogenic mutant strain SPD0247M. For each time point, between 6 to 10 mice for SPD0247M (■) and 15 to 20 for D39 (●) were used. Vertical bars indicate the standard error of the mean.

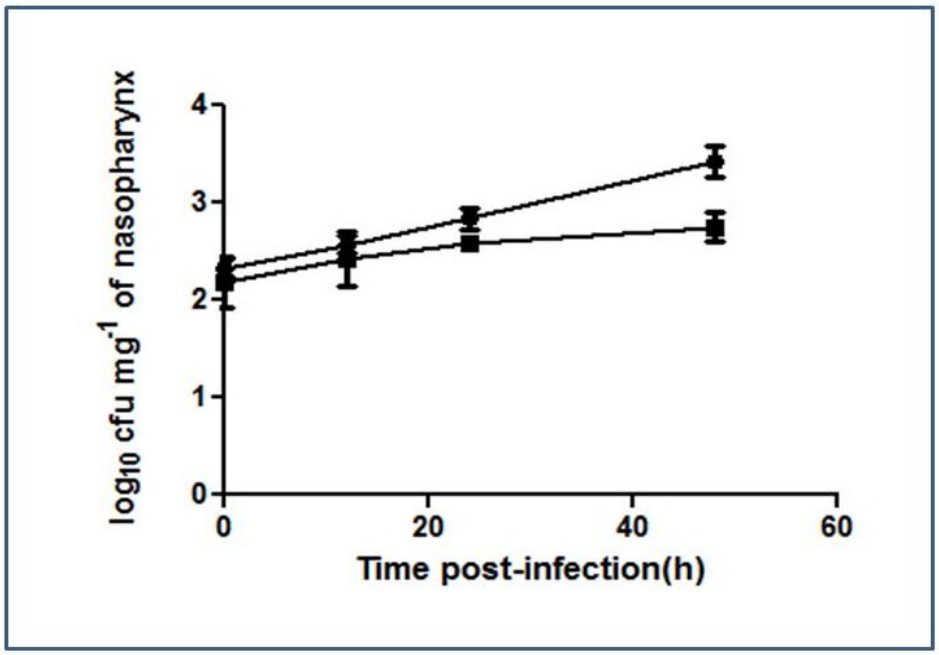


Figure 11- Number of pneumococci per mg retrieved from nasopharynx of intranasally infected mice (●) *S.pneumoniae* D39 (parental strain); (■) Isogenic mutant strain SPD0247M. For each time point, between 6 to 10 mice for SPD0247M (■) and 15 to 20 for D39 (●) were used. Vertical bars indicate the standard error of the mean.

I.9 Determination of use of 6-phospho-β-glucosidase as protective immunogen

In order to determine what was the protective potential of this newly identified 6-P-β–glucosidase an immunisation assay was performed as described in the Material and Methods Chapter, section XIV. 10 mice were immunised every 2 weeks for a period of 8 weeks and their blood collected just before each immunisation to isolate serum containing antibodies. After this period the mice were challenged with *S.pneumoniae* D39 and the disease signs were observed. The results obtained are described in Figure 12.

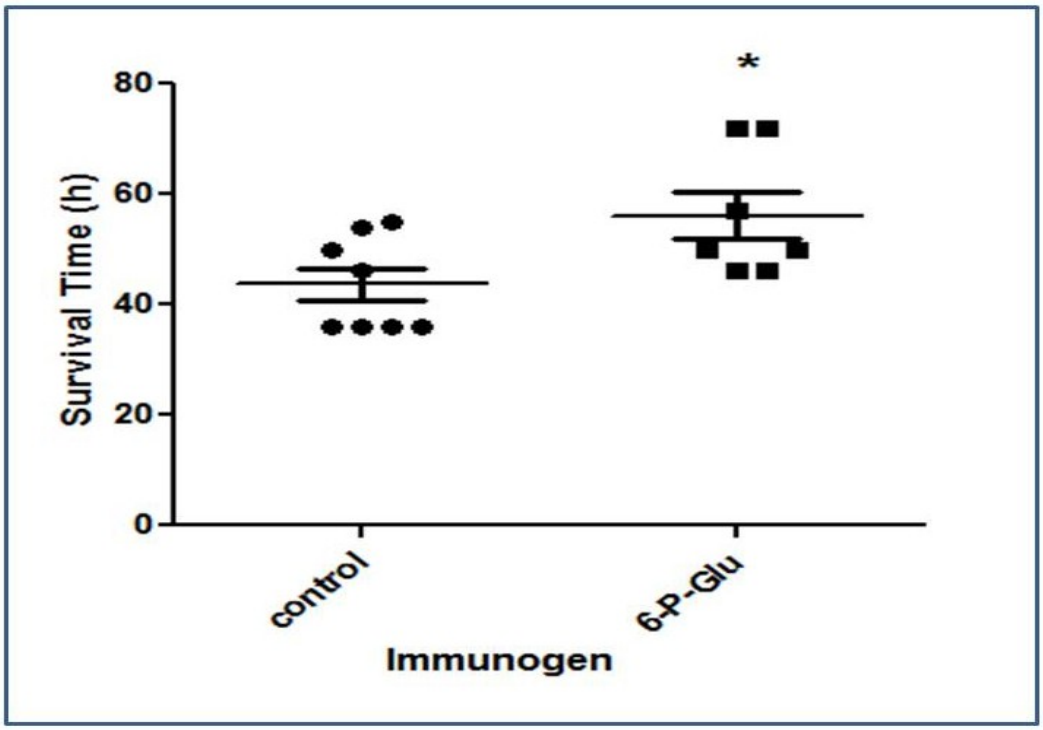


Figure 12- Survival of mice challenged with *S.pneumoniae* D39 previously immunised with (■) recombinant protein 6-P-β-glucosidase and Alum. (●) Control-animals injected with Alum and PBS only. For the control group 8 mice were used, for the immunised group 7 mice were used. Vertical bars indicate the standard error of the mean. Each point corresponds to one animal.

The analysis of the results showed differences between the groups. The control group has a median survival of 43.6 ± 3.0 hours, whereas the immunised group has a median survival of 56.1 ± 4.3 hours, ($p<0.05$). Even though there was a difference in survival ($p<0.05$) when the pneumococci numbers were measured in blood in both control group

and immunised group there was no significant difference ($p>0.05$). The bacteria seeded into blood at the same time, and they grew with a similar rate (Figure 13). The survival results suggest that the administration of this enzyme as a protective agent increases survival slightly, however it does not fully protect against pneumonia and bacteraemia caused by *S. pneumoniae*. The results obtained for the colony forming units numbers retrieved from blood are shown on Figure 13.

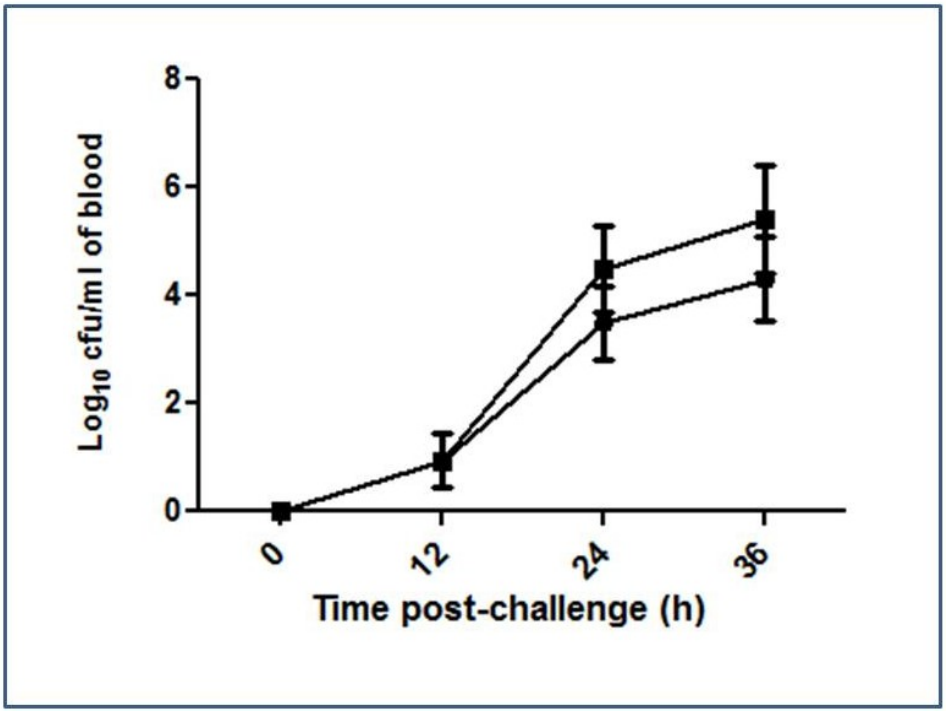


Figure 13- Number of pneumococci per ml retrieved from blood of intranasally challenged mice. (●) Control group challenged with *S.pneumoniae* D39; (■) Immunised group challenged with *S. pneumoniae* D39. Vertical bars indicate the standard error of the mean.

I.10 Adherence studies

Adherence assessment was done because of the observations published by Kilic *et al.* 2004, who demonstrated that genes associated with beta-glucoside metabolism may regulate adhesion and biofilm formation in *Streptococcus gordonii*. Although the gene under study does not encode a β-glucosidase but a 6-phospho-β-glucosidase it was hypothesised that if this enzyme plays a role in β-glucoside metabolism it may also play a

role in adherence. In order to test this hypothesis an adherence assay was performed following the protocol described in the Material and Methods Chapter section XX. The results obtained are shown in Figure 14.

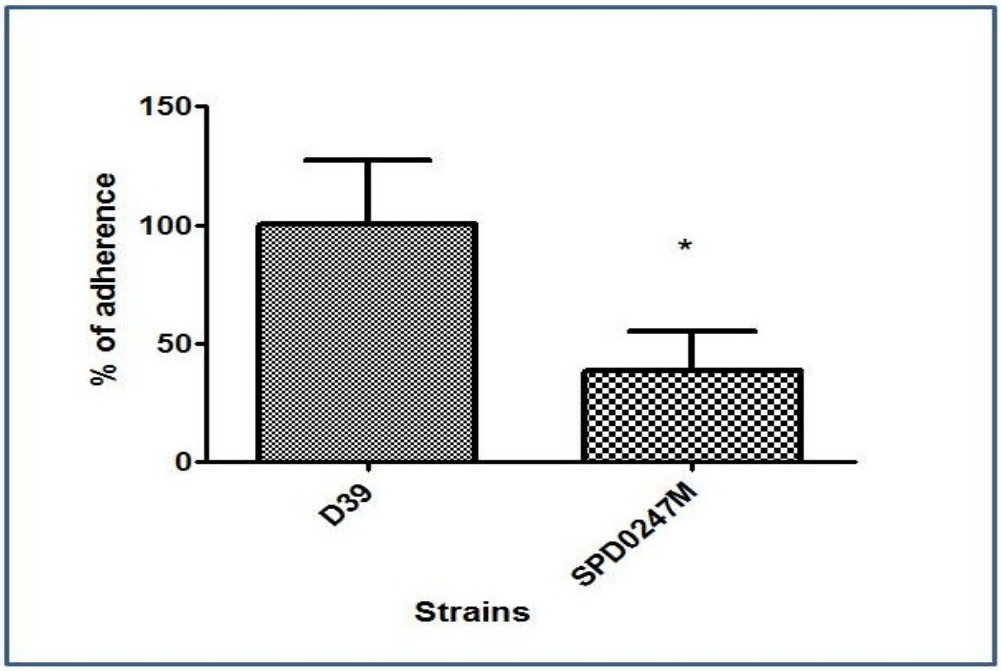


Figure 14- % of adherence of mutant SPD0247M when compared to D39. The vertical bars indicate the standard error of mean. * indicates $p<0.05$.

There was a clear reduction on the ability to form biofilm on plastic by the isogenic mutant SPD0247M ($p<0.05$). The results suggest a role for this enzyme in ability to adhere to this surface. Providing the recombinant protein externally was also attempted, however there was no improvement on the biofilms formation ability for mutant SPD0247M.

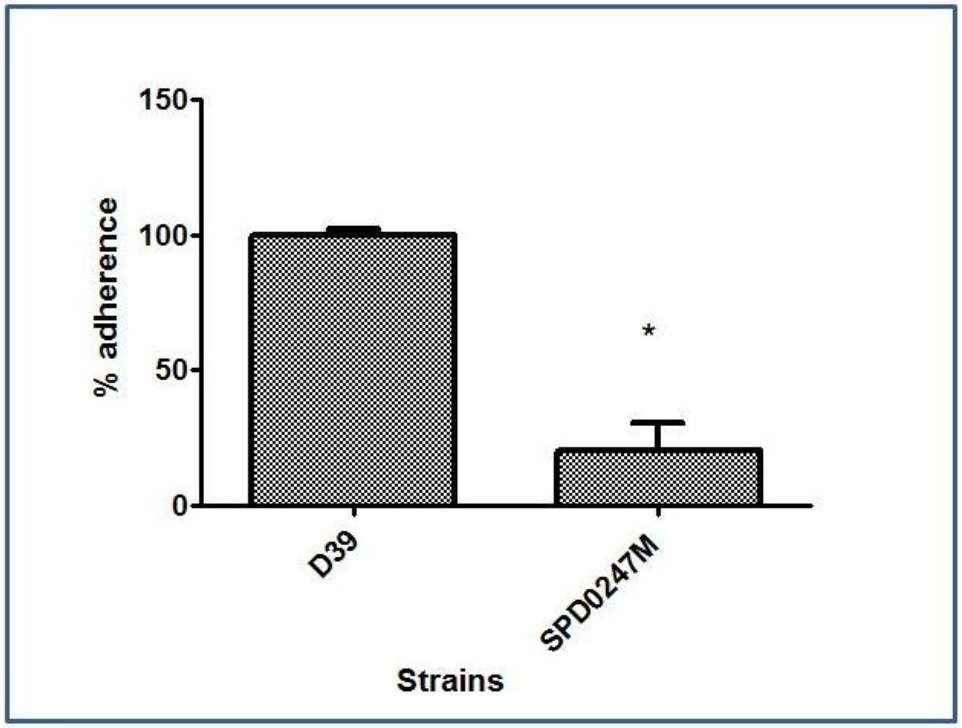


Figure 15-% of adherence for mutant SPD0247M when compared to D39. The vertical bars indicate the standard error of mean. * indicates $p<0.05$.

G Discussion

I. Overview

Streptococcus pneumoniae is a Gram positive bacterium responsible for 1.6 million deaths each year, worldwide (World Health Organisation, 2008). This bacterium is known as the causative agent of pneumonia, otitis media, meningitis and septicaemia (Mitchell *et al.*, 1997). This bacterium can colonise the upper respiratory tract asymptomatically. It is thought that colonisation of the human airway by *S. pneumoniae* is an essential step in the development of disease (Bogaert *et al.*, 2004). At the present time, the mechanisms used by the bacterium to establish itself, to propagate and colonise are not yet fully understood.

Streptococcus pneumoniae uses carbohydrates to fulfil nutritional requirements, however little is known about their metabolism. Other groups showed that pneumococcal modification of host glycoproteins can contribute to revelation of adherence receptors, modification of host defense molecule function (Tong *et al.*, 1999, Tong *et al.*, 2000), progression of the bacteria through the mucin layer (King *et al.*, 2006), biofilm formation and interspecies competition (Shakhnovich EA *et al.*, 2002). As concentrations of readily available carbohydrates, such as glucose, are low *in vivo* (Homer *et al.*, 1994), it is likely that the pneumococci rely on other carbon sources to multiply and colonise the host environment. The need to understand how pneumococci utilise carbohydrates for nutritional requirements led to studies of the utilisation of host glycoproteins, such as mucins, within the nasopharynx. The nasopharynx is covered by a layer of mucin which contains N- and O- linked carbohydrates (Rose & Voynow, 2006). *Streptococcus pneumoniae* expresses glycosidases, which have the ability to degrade N- and O-linked glycoproteins and glycosaminoglycans abundant in the human

host (Burnaugh *et al.*, 2008). Despite the fact that four glycosidases involved in mucin degradation, NanA, StrH, BgaA and EngO, are already known and characterised, analysis of *S. pneumoniae* strain TIGR4 genome revealed a bacterium which has 33% of its genome devoted to sugar metabolism, encodes for at least 10 exoglycosidases and has 30% of its transporters responsible for sugar transportation to the inside of the cell (Tettelin *et al.*, 2001). Further analysis of *S. pneumoniae* sequenced strains TIGR4 and R6, uncovered more ORFs whose expression products can potentially be involved in carbohydrate metabolism (Hoskins *et al.*, 2001), for example, β-D-galactosidase and N-acetyl-β-D-galactosaminidase. These facts, combined with the desirability of understanding *S. pneumoniae* carbohydrate metabolism, led to the investigation of the role of genes annotated as glycosidases in the growth of *S. pneumoniae* in the presence of glycoproteins.

Previously the ability of pneumococci to utilise mucin as a carbon source was demonstrated by Yesilkaya *et al.*, (2008). A recent publication suggested that the host glycoproteins are deglycosylated in a sequential manner (Figure 1) (King *et al.*, 2006). It is thought that the first stage of mucin degradation is the cleavage of sialic acid by the action of a neuraminidase A (NanA) (Corfield *et al.*, 1992). However, if sialic acid is O-acetylated on its hydroxyl groups, the rate of cleavage by NanA is considerably slower than if sialic acid is not acetylated. The O-acetylation problem may be overcome by removal of the acetyl moiety by a group of enzymes known as sialate O-acetylesterases (Corfield *et al.*, 1992), although this has not been proven. Fourteen pneumococcal clinical isolates and D39 were investigated for the presence of sialate O-acetylesterase activity. The authors observed that all strains studied exhibited sialate O-acetylesterase activity (Yesilkaya *et al.*, 2008) which means that *S. pneumoniae* is likely to be able to overcome the acetylation of sialic acid when in the host. Once this moiety

is removed, the assumption is that glycoprotein deglycosylation continues with the action of NanA, BgaA and StrH, (Figure 1) working in a sequential fashion resulting in the release of fermentable carbon sources (Burnaugh *et al.*, 2008).

Following what was known already about sugar hydrolases, the aim of this project was to identify genes essential for growth in mucin and characterisation of the enzymes encoded by these genes in host glycoproteins' degradation. The studies performed in this thesis led to the identification of two novel glycosidases that appear to be necessary for pneumococci growth in mucin.

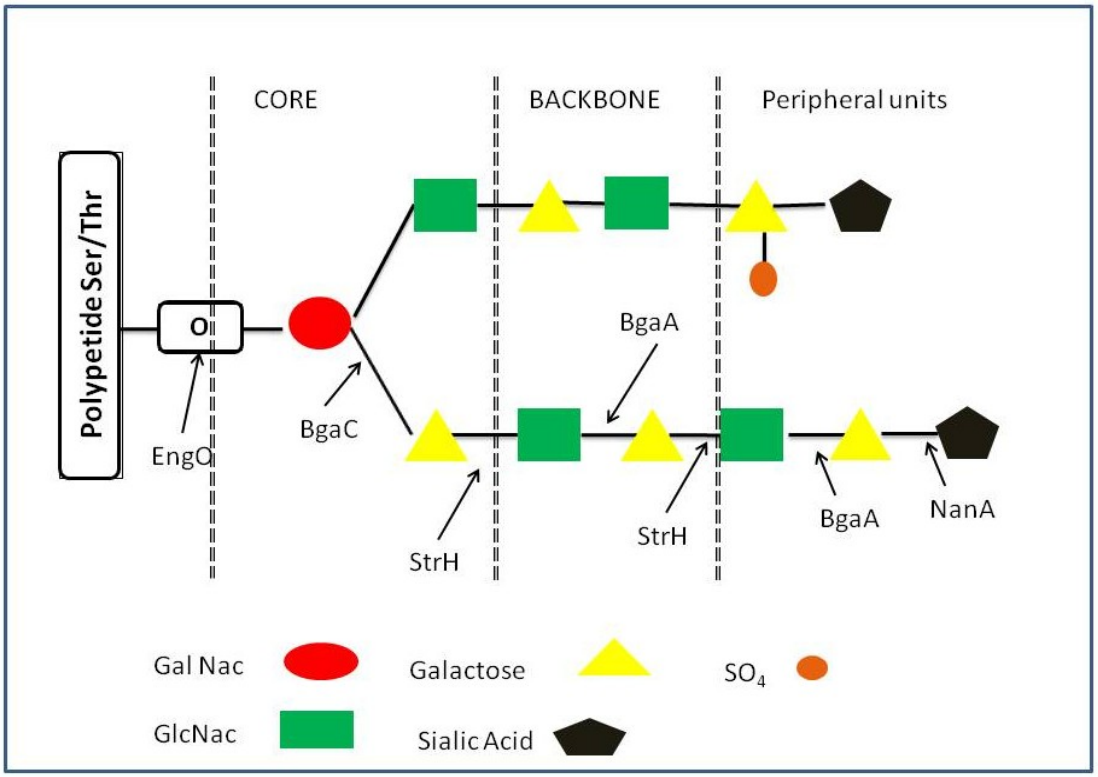


Figure 1- Schematic representation of mucin and possible cleavage site of the described exoglycosidases. GlcNac- N-acetylglucosamine; GalNac- N-acetylgalactosamine; SO_4-sulphate group; BgaA- β-galactosidase A, BgaC -β-galactosidase C, StrH- N-acetylglucosaminidase, NanA- neuraminidase A, EngO- O-glycosidase.

I.1 Identification of genes necessary for growth in mucin

In silico analysis of the *S. pneumoniae* TIGR4 genome revealed other genes that were predicted to be involved in carbohydrate metabolism. A collection of fourteen

isogenic knock-out mutants was generated in these different putative glycosyl hydrolases, by Dr Hasan Yesilkaya, prior the commencement of this thesis. The objective was to identify which genes would encode for enzymes directly involved in mucin-derived carbohydrate metabolism. The hypothesis was that if the knocked out genes encoded for an enzyme directly related with utilisation or degradation of the carbohydrates of mucin, a growth deficiency would be observed. Identification of involvement of these genes in mucin-derived carbohydrate utilisation was assessed by growth of the loss of function mutants in Sicard's defined medium supplemented with mucin as the source of carbon. Previously, it had been shown by our group and others that both *S. pneumoniae* and *S. oralis* can efficiently grow in the presence of glycoconjugates such as porcine gastric mucin and alpha-1-acid glycoprotein (Yesilkaya *et al.*, 2008a, Byers *et al.*, 1999), validating the approach used for selection of genes involved in glycoproteins degradation. The ability of *S. pneumoniae* to utilise glycoproteins as a nutrient source involves three stages: degradation of the glycoproteins carbohydrate content, transport, and utilisation inside the cell. A growth assay does not allow identification of a definite role for the enzymes; however it allowed for discrimination between involvement and no involvement in mucin degradation and or utilisation.

In this study it was demonstrated that mutant SPD0247M, encoding for a putative glycosyl hydrolase, exhibited a complete inability to grow on mucin and mutants SPD0065M, SPD0057M and SPD0853M (see Figure 1 and 2 of Chapter 3) exhibited a significant decreased growth rate when compared to the wild type. Genes *SPD0065*, *SPD0057* and *SPD0853* had been annotated as β-galactosidase, β-N-acetylhexosaminidase and endo-β-N-acetylglucosaminidase respectively. The growth deficiency observed for the loss of function mutants suggested a role for these enzymes

(β-galactosidase, β-N-acetylhexosaminidase and endo-β-N-acetylglucosaminidase) in the ability to degrade and utilise mucin-derived carbohydrates, which given the annotation of this enzymes was plausible, because the sugars present in mucin could be substrate for these enzymes.

Isogenic mutants SPD0427M (6-phospho-β-galactosidase), SPD0562M (β-galactosidase A), SPD1046M (6-phospho-β-galactosidase), SPD1678M (α-galactosidase), SPD1898M (hypothetical protein) and SPD1969M (putative glycosyl hydrolase) presented no significant difference in growth compared to the wild type. These growth results suggested that these enzymes had no obvious role in the mucin degradation and or further utilisation. These results were plausible because assuming the mucin structure described in Introduction chapter section II.1, these enzymes were not thought to have any obvious role in the mucin degradation because the sugars present were not substrate for these enzymes.

Mutants SPD0444M (endo-N-acetylglucosaminidase) and SPD0503M (6-phospho-β-glucosidase) exhibited significantly better growth when compared to the wild type. These results were unexpected. However, at the moment it is not possible to conclusively offer an explanation for the phenotype observed but there are two explanations that can be put forward. Either there was some mechanism of compensation or the introduction of cassette in the gene created a polar effect. The mechanism of compensation could, for example, be an over expression of other genes that encode for enzymes involved in mucin degradation creating effectively a mutant that had an increased ability to degrade or utilise mucin. As for the possible polar effect it could be that the insertion of the antibiotic cassette interfered with the expression of the genes downstream of the interest gene with possible pleiotropic effects leading to an

abnormal growth in the presence of mucin. Both of these explanations are speculation, however they could be tested. Genetic complementation of these mutants could be done using pCEP plasmid previously described by Guiral *et al.*, (2006) to eliminate the possibility of polar effect. Possible over expression of other genes could be assayed using microarray technique where all the over-expressed genes would be measured and it would be possible to understand if there are genes over-expressed in these two mutants possibly involved in mucin degradation.

Despite the lack of explanation for the phenotype observed, it is possible to conclude that the enzymes encoded by genes *SPD0444* and *SPD503* do not appear to have direct involvement in mucin-derived carbohydrates metabolism which according the hypothesis adopted for this work would result in growth deficiency. From this first screen three mutants were chosen for further analysis, SPD0065M, SPD0244M and SPD0853M.

I.2 Mutant SPD0065M ability to grow in Sicards defined medium supplemented with mucin

Gene *SPD0065* was thought to encode for a β-galactosidase. When this gene was mutated by insertion of an antibiotic cassette the mutant could not grow in Sicard's defined medium supplemented with mucin, however as a control the mutant was grown in Sicard's defined medium supplemented with glucose and no differences were observed in comparison with the wild type. This observation suggested that β-galactosidase was essential for the ability of pneumococci to grow in the presence of mucin as carbon source. The carbohydrate content of the mucin side chains is composed of galactose, N-acetylgalactosamine, N-acetylglucosamine and sialic acid (Rose & Voynow, 2006). The data collected for mutant SPD0065M suggested that the SPD0065

enzyme was necessary to cleave one of these sugars from the mucin structure; therefore in its absence the mutant would have a diminished ability to cleave the carbohydrates and subsequently difficulties in growth. An example of this type of analysis is the work of Yesilkaya *et al.*, (2008) which showed that pneumococci are able to grow in the presence of mucin, but when a loss of function mutant in NanA was tested for the ability to grow in Sicard's defined medium supplemented with mucin it was not able to grow. Their observations support the data acquired for the mutant SPD0065M by demonstrating that when glycosidases involved in the cleavage of mucin are not present, pneumococci has decreased ability to grow or no ability at all. The β-galactosidase C (BgaC) as NanA. BgaA and StrH (Burnaugh *et al.*, 2008) appears to have a role in the efficient growth of pneumococci in the presence of glycoproteins.

In order to confirm that the growth pattern in mucin was not caused by a polar effect created by the insertion of the spectinomycin cassette in gene *SPD0065* a complementation experiment was done. Insertion mutation can lead to problems in genes downstream of the interest gene and a polar effect occurs when the mutation leads to the termination of transcription. Consequently any gene located downstream of the mutation without its own promoter, in an operon organisation, will not be expressed. If the gene of interest is the first gene of a predicted operon then the probability of a polar effect caused by an insertion mutation is higher. *SPD0065* (*bgaC*) is predicted to be such a gene.

The method adopted for complementation uses a plasmid (pCEP) that is capable of replication in *E. coli*, but not in *S. pneumoniae.* pCEP has an expression/selection cassette that contains a maltose-inducible promoter, separated from a kanamycin-resistance gene by *Nco*I and *Bam*HI cloning sites (Figure 2A). To facilitate homologous

recombination, the expression/selection cassette is flanked by more than 2 kb of pneumococcal DNA allowing the entrance of the engineered segment at a much higher frequency than the complete pCEP. Therefore, pCEP allows insertion of the gene of interest under the control of a maltosaccharide-inducible promoter, in a chromosomally silent site, downstream of *amiA* operon (Alloing *et al.*, 1990). This promoter also functions constitutively without induction by maltose. While the 2 kb of pneumococcal DNA flanking the cassette allows its integration into a region of the pneumococcal chromosome which originally does not have any expressed sequences (Figure 2B, green arrows) (Guiral *et al.*, 2006), it can also give rise to high numbers of false positive clones harbouring only the resistance cassette.

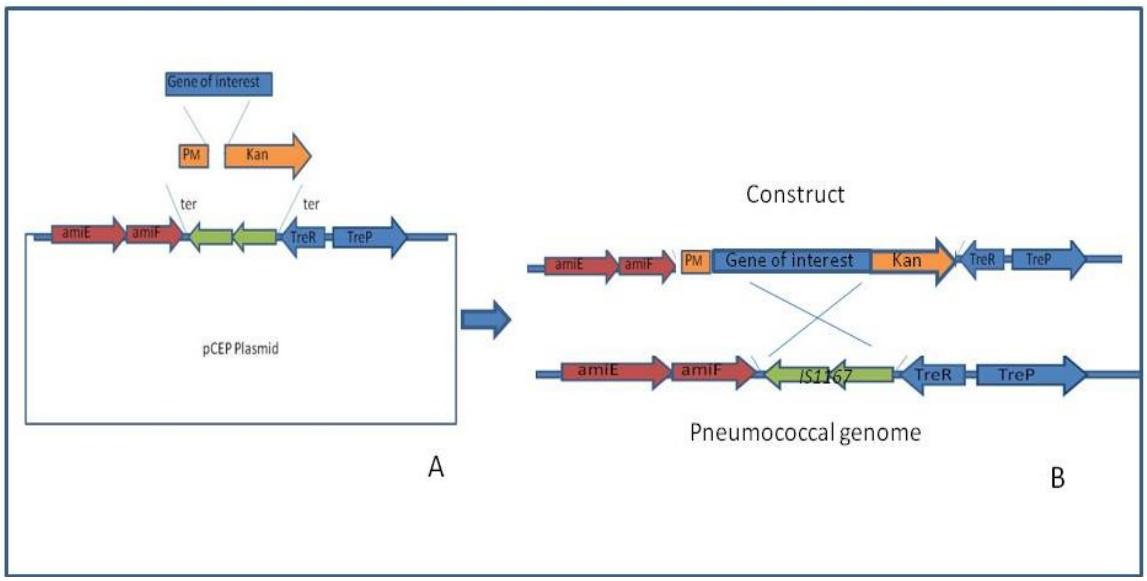


Figure 2- A- pCEP plasmid. Chromosomal expression organisation (CEP); P_M-maltose inducible promoter, Kan-*aphIII* gene which confers kanamycin resistance for selection of single copy mutants. The green arrows represent the IS*1167* truncated region that was substituted by P_M-Kan (the orange square and arrow). The gene of interest is to be cloned between P_M and Kan. The blue arrows (TreR and TreP) are part of the predicted trehalose utilisation system. Ter -transcription terminator. B- Strategy

of pCEP recombination with pneumococcal genome. Figure based on Guiral *et al.*, (2006).

The complementation of mutant SPD0065M was successful, as confirmed by PCR using primers pCEPR and MalF. The positive colonies were tested for both growth in Sicard's defined medium supplemented with mucin as the sole source of carbon and enzymatic activity with substrate ONP-β-galactopyranoside. The complemented mutant SPD0065REV showed growth properties similar to those obtained for the wild type D39. A complete restoration of the wild type phenotype was also observed when the cell free extracts were tested for β-galactosidase activity with ONP-β-galactopyranoside. These results confirmed that the insertion mutation of the transposon previously done to knock out gene *SPD0065* did not create a polar effect. A restoration of the phenotype was also observed by Guiral *et al.*, (2006), who after analysis of the wild type strain R800 transformed with pCEP in *S. pneumoniae,* reported similar growth properties between the complemented mutant and wild type (Guiral *et al.*, 2006).

I.3 Mutant SPD0247M ability to grow in Sicard's defined medium supplemented with mucin

Another mutant that showed a complete inability to grow in Sicard's defined medium supplemented with mucin was mutant SPD0247M, however in Sicard's defined medium supplemented with glucose it grew as well as the wild type. At the time of this experiment this mutant's annotation was still of a putative glycosyl hydrolase, however throughout the course of the work the annotation for gene *SPD0247* changed

to 6-phospho-β-glucosidase. Considering this new annotation and the carbohydrate content of mucin, which does not include a phosphorylated hexose, it appeared unlikely that that this enzyme would be involved in mucin degradation.

The fact that, despite the lack of substrate for 6-phospho-β-glucosidase in the mucin structure the mutant SPD0247M was still unable to grow, led to speculation of whether this enzyme would be involved in the further utilisation of the carbohydrates released from the mucin side chains. The data acquired in the growth experiment combined with the annotation, suggested that the role of this enzyme could be an involvement in the degradation of phosphorylated disaccharides intracellularly. This results in monosacharides and disaccharides being produced. The disaccharides would then be transported through a PTS system, being phosphorylated simultaneously and thus becoming a substrate for 6-phospho-β-glucosidase.

As with mutant SPD0065M, complementation was also attempted with mutant SPD0247M. This was done to eliminate the possibility of a polar effect created by the insertion of the antibiotic cassette in the interest gene. However despite all efforts the complementation was always unsuccessful.

In the original description of pCEP, Guiral *et al*., (2006) described efficient integration of a *luc* gene into the pneumococcal genome, in which 10 out of 10 analysed transformants carried CEP^{luc} (**C**hromosomal **E**xpression **P**latform composed of maltose promoter, *luc* and *aph*III, which confers kanamycin resistance) downstream of *amiA*. Guiral *et al*. (2006) used *BamH*I/*Nco*I digested pCEP and insert for ligation. However, in this study pCEP was digested with *BamH*I, and the *SPD0247* gene with *Bgl*II, because *SPD0247* contains a *BamH*I site in its sequence. Once pCEP was digested with *BamH*I it would produce GATC cohesive ends. Incorporation of *Bgl*II site in the PCR

product for further digestion would have added an additional T to the new construct. In order to prevent formation of a stop codon, a TTG codon was fashioned by addition of TG nucleotides to the *SPD0247* primers **GCGAGATC<u>TTG</u>**. (The highlighted section indicates *Bgl*II site, and the newly created codon is underlined). The initial three nucleotides were added because certain restriction digestion enzymes cannot work efficiently if there are no additional nucleotides before the recognition site). Even though the ligation strategy adopted for this work differed from the one suggested by Guiral *et al.* (2006) study, it seems unlikely that it is the cause of the failure to complement SPD0247M because the same strategy was adopted for the complementation of SPD0065M, and at that instance it was successful.

In this study, the ligation products were directly transformed into *S. pneumoniae* as recommended by Guiral *et al*., (2006). As an alternative, an intermediate cloning step could have been performed in *E. coli*. However, this was not done because, firstly transformation efficiency of the pneumococcus is higher than that of *E. coli* and secondly cloning *S. pneumoniae* products in *E. coli* is known to be problematic because of toxicity of pneumococcal fragments to *E. coli* (Martin *et al*., 1989). In conclusion, the reason why the complementation of SPD0247M was unsuccessful is yet unknown however in the future, this work can be repeated using different compatible cloning sites. As an alternative to genetic complementation, phenotypic complementation can also be attempted. Previously, King and co-workers successfully complemented pneumococcal mutants with addition of recombinant β-galactosidase and neuraminidase, however both of this enzymes are localised at the cell surface which makes the phenotypic complementation possible (King *et al*, 2006). Although, phenotypic complementation may be successful for extracellular enzymes, it is doubtful that it would work for intracellular enzymes such as 6-phospho-β-glucosidase.

I.4 Mutant SPD0853M ability to grow in Sicard's defined medium supplemented with mucin

The mutant SPD0853M had reduced ability to grow in Sicard's defined medium supplemented with mucin. This was an expected result since the enzyme encoded by gene *SPD0853* was annotated as an endo-β-N-acetylglucosaminidase. As mentioned previously, mucin contains N-acetylglucosamine in its structure linked to other sugars via a β-linkage, which would make that sugar a substrate for the enzyme encoded by *SPD0853*. Therefore, when a reduction in the ability to grow in the presence of mucin as carbon source was observed, the initial thought was that this gene would most likely encode for an enzyme involved in the mucin degradation. However, because this enzyme also had proposed to be involved in the separation of cells by cleaving the bond between the N-acetylglucosamine and the N-acetylmuramic acid (De Las Rivas *et al.*, 2002) other questions were raised. Was this endo-β-N-acetylglucosaminidase only releasing N-acetylglucosamine from the bond between N-acetylglucosamine and N-acetylmuramic acid or would it cleave N-acetylglucosamine from any other glycoprotein regardless of the neighbouring moiety as long as the two were joined through a β-linkage?

To answer these questions an assay that identified the cleavage specificity could be done through the use of specific lectins. For example, for the determination of the cleavage on N-acetylglucosamine in the peptidoglycan structure, a mannose binding lectin could be used. This lectin is known for recognising the peptidoglycan by binding to the N-acetylglucosamine moiety. On the other hand, if this enzyme would cleave an N-acetylglucosamine, a lectin such as DSA could be used as long as this moiety was linked through a β-linkage to the following moiety. If the enzyme were found to be able

to cleave N-acetylglucosamine β-linked to any other moiety, then the decrease in the ability to grow would be an expected result because that would mean that the mucin contained a substrate for it. However, if this enzyme would only be able to cleave the specific bond present in the peptidoglycan, then the reason behind the decreased growth would not be clear. Nevertheless, it would be possible that, in minimal medium, the lack of the endo-β-N-acetylglucosaminidase would have a more prominent effect in cell division leading to a growth deficiency.

As with mutants SPD0065M, SPD0247M, cloning of gene *SPD0853* was attempted for further expression and characterization. However, the cloning failed despite numerous attempts. Several reasons may explain this: low DNA concentration in the cloning reaction resulting in low ligation efficiency, low DNA retrieval after purification of PCR product and possible incorrect primers sequence. These possibilities were investigated and eliminated. Despite further several attempts, the cloning process was unsuccessful. Another explanation for the difficulties encountered came from the work of De Las Rivas *et al.*, (2002) which identified *SPR0867* (a homolog of *SPD0853,* from strain R6, which has 100% identity in the overlapping sequence) as the gene encoding for a LytB, an enzyme responsible for cleavage of peptidoglycan. The authors observed that SPR0867contained a signal peptide. When they attempted to clone this gene in this unprocessed form, they were always unsuccessful. However, when the DNA for the signal peptide was cleaved the cloning was possible. The authors hypothesised that when the complete gene was present the signal peptide was processed by *E. coli* and the protein exported to the outside of the cell where it would degrade the peptidoglycan, causing cell lysis (De Las Rivas *et al.*, 2002). Gene *SPD0853* was found to encode a LytB (De Las Rivas *et al.*, 2002). Bioinformatics analysis demonstrated that *SPD0853* also contains DNA for a signal

peptide. The failure to clone *SPD0853* when the signal peptide sequence was present, and the fact that the enzyme was already characterised as being a murein hydrolase (LytB) with the implication that then this enzyme would not be involved in the degradation of glycoproteins, led to the abandonment of this target. This enzyme's homolog from strain R6, was shown to be involved in the degradation of the peptidoglycan, cleaving the N-acetylglucosamine being identified as the first endo-N-acetylglucosaminidase in *S. pneumoniae* (De Las Rivas *et al.*, 2002). The research efforts were focused only on genes *SPD0247* and *SPD0065*.

I.5 pH optimum determination

Cell free lysates from D39 were used to determine the pH optimum when using synthetic substrates ONP-β-galactopyranoside and with pNP-N-acetyl-β-D-glucosaminide. The pH optimum was determined to allow enzymatic assays to be done with cell free extracts prepared from the fourteen isogenic mutants. Such an assay would help understanding and establishing relationships between annotations and possible enzyme functions. The pH optimum determination was indispensible since if not known the activity results would be misleading. However, this type of determination has its own drawbacks; for example, if there was an unidentified β-galactosidase with a pH optimum of pH 5.0, performing the assay at 6.5 may ignore this enzyme's contribution to the total β-galactosidase activity.

The synthetic substrates used were based on the sugars present in mucin, namely galactose and N-acetylglucosamine. The pH optimum obtained for cell free extracts when using the substrate ONP-β-galactopyranoside was in the range of 6.5-7.0, while with pNP-N-acetyl-β-D-glucosaminide the range was 5.0-5.5. The pH obtained when substrate pNP-N-acetyl-β-D-glucosaminide was used, was in agreement with data from Homer *et al.*, (1994) where the authors demonstrated that the optimum pH for enzyme N-acetyl-β-D-glucosaminidase from *Streptococcus intermedius* was in the range of 5.0-5.5. However, the pH optimum for cleavage of substrate ONP-β-galactopyranoside by β-galactosidase from *S. intermedius* was in the range of 5.5-6.0 (Homer *et al.*, 1994), slightly lower than that found for β-galactosidase from *S. pneumoniae.*

I.6 Determination of loss of function mutants' activity with substrates ONP-β-galactopyranoside and -N-acetyl-β-D-glucosaminide

The collection of isogenic, loss of function mutants was tested against the synthetic substrates ONP-β-galactopyranoside and PNP-N-acetyl-β-D-glucosaminide at the pH determined as described in the Material and Methods Chapter section IV.1 SPD0065M showed reduced enzymatic activity when the substrate ONP-β-galactopyranoside was used. This was an expected result for this mutant since the annotation for gene *SPD0065* was for β-galactosidase. SPD0247M was a different case. At the time of these experiments the annotation for this enzyme was for a glycosyl hydrolase, therefore the identity of the substrate was unclear. The fact that there was a reduction in enzymatic activity when using ONP-β-galactopyranoside as substrate, suggested that this enzyme was a β-galactosidase but during the course of this work the annotation for gene *SPD0247* changed to a 6-phospho-β-glucosidase. Considering this new annotation the reduction in activity when using substrate ONP-β-galactopyranoside was an unexpected result. However, maybe the result observed is due to some indirect role of this enzyme in the expression of surface proteins, namely the β-galactosidases. If this is the case, than a reduction in the activity of the cell free lysate of mutant SPD0247M with substrate ONP-β-galactopyranoside would be an expected result.

SPD0063M had reduced activity against the substrate pNP-N-acetyl-β-D-glucosaminide, an expected result since Clarke and co-workers (Clarke *et al.*, 1995) demonstrated that this enzyme is a β-acetylhexosamindase and therefore likely to use the tested substrate.

SPD0853 had been annotated as an endo-β-N-acetylglucosaminidase (Hoskins *et al.*, 2001). SPD0853M was tested against the synthetic substrate pNP-N-acetyl-β-D-glucosaminide but the cell free lysate did not show any significant difference in activity when compared to wild type. This result was consistent with the fact that this enzyme is an endoglycosidase (Hoskins *et al.*, 2001) and therefore the enzyme would only cleave internal sugars of a polysaccharide. In order to check that the annotation was not incorrect (and that one glycosidase would not be missed) the assay was done with mutant SPD0853M. Also in support of the annotation for gene *SPR0867*, a homolog of gene *SPD0853* as an endo-N-acetylglucosaminidase, is the work of De Las Rivas *et al.*, (2002). They reported that this enzyme is also named as murein hydrolase and it was shown to have a role in the cleavage of N-acetylglucosamine from the peptidoglycan (De Las Rivas *et al.*, 2002).

I.7 *SPD0065* and *SPD0247* gene expression

Expression of gene *SPD0065* was measured both *in vitro*, in the presence of PGM, and *in vivo*, in the lungs, nasopharynx and blood of infected animals. An increase in gene expression was seen in the tissues tested and in the presence of mucin when pneumococci were grown in Sicard's defined medium when compared to expression in Sicard's defined medium supplemented with glucose. The increase in expression suggests that this enzyme has an important role in mucin degradation. The conclusion that the in the presence of glycoproteins pneumococci over-express or de-repress genes that encode for proteins related to glycoprotein degradation is supported by the work of Yesilkaya *et al.* (2008). They found that expression of *nanA* was increased in the presence of mucin (Yesilkaya *et al.*, 2008). Because the concentration of glucose *in vivo* is low (Homer *et al.*, 1994) and gene expression analysis in the tissues showed

expression of gene *SPD0065* is increased when compared to the gene expression in the presence of glucose it seemed reasonable to hypothesise that this gene may be regulated by catabolite control protein A. In Gram positive bacteria, the catabolite control protein A (CcpA) is implicated in carbon catabolite repression (CCR). Kaufman & Yother, (2007) demonstrated that CcpA plays a role in the regulation of pneumococcal BgaA (*SPD0562*) (see Introduction chapter sectionVI.2). The genes or operons believed to be regulated by CcpA, contain an operator sequence known as the catabolite responsible element (*cre*), where CcpA binds. This particular sequence can be found both in the promoter region or in the coding sequence of the target gene, and it is known to have a consensus sequence WWTGNAARCGNWWWCAWW (R for G or A; W for A or T; N for any base) (Miwa *et al.*, 2000). An exact match of the *cre* consensus sequence was identified 37 bases upstream of the start codon of the gene *SPD0065* (ATGAAAGCGCAAACT). It is therefore possible, that this gene is regulated by CcpA. This would mean that enzymes involved in the metabolism of secondary carbon sources i.e. galactose would be at baseline expression when glucose is available. On the other hand, when the glucose concentration is low and the levels of secondary sugars are high the repression would be relieved allowing these enzymes to be expressed and metabolise the available sugars.

In order to test such a hypothesis, a mutation in the *cre* consensus sequence could be made. This mutation would disrupt the ability of CcpA to bind to the *cre* sequence not allowing repression to occur, and consequently there would be a constitutive expression of β-galactosidase demonstrating that repression of this gene was under CcpA control.

The expression of gene *SPD0247* was measured in Sicard's defined medium supplemented with mucin as well as in lungs, nasopharynx and blood of infected

animals. A significant increase in the *in vivo* expression of *SPD0247* was observed in the lungs and nasopharynx. *SPD0247* mRNA level was also increased *in vitro* in Sicard's defined medium supplemented with mucin. This suggested that this gene is inducible by glycoproteins present in the lungs and nasopharynx, and *in vitro* by mucin. These data support the growth data which showed that lack of the enzyme encoded by gene *SPD0247* completely abolished the ability of *S. pneumoniae* to degrade or utilise mucin as nutrient. If gene *SPD0247* has a relevant role in the degradation or further utilisation of glycoproteins it seems plausible that its expression would be increased when in the presence of the host glycoproteins.

This ability of host glycoproteins to induce the expression of streptococcal genes that encode for enzymes that may be responsible for cleaving or utilising the mucin derived carbohydrates, have been reported previously by Homer *et al.,*(1994). Homer and co-workers observed that when *S. intermedius* was grown in minimal medium supplemented with mucin as a source of fermentable carbon, high levels of β-D-galactosidase, N-acetyl-β-D-glucosaminidase, N-acetyl-β-D-galactosaminidase and neuraminidase were seen. In *S. pneumoniae* a recent report demonstrated that *nanA* is up regulated in the presence of PGM (Yesilkaya *et al.*, 2008). This is significant because NanA is responsible for the cleavage of terminal sialic acid from mucin (Corfield *et al.*, 1992). Contrary to gene *SPD0065*, gene *SPD0247* does not contain a *cre* consensus sequence, and therefore it is not likely that this gene is regulated by CcpA.

In conclusion, in this study two more enzymes are being reported as over-expressed in the presence of host glycoproteins establishing a link between over-expression and ability to use glycoproteins.

I.8 SPD0065M and SPD0247M abilities to degrade glycoproteins

The work was performed in collaboration with Dr Karen Homer, King's College Dental Institute London.

The ability of *S. pneumoniae* to degrade glycoproteins has been previously demonstrated by King *et al*.,(2006) The group reported that cell cultures of D39 were sufficient to expose mannose in human secretory component (hSC), however when mutants in BgaA (β-galactosidase A), StrH (β-N-acetylglucosaminidase) and NanA (neuraminidase A) were tested, mannose was not exposed. Furthermore, mannose exposure was achieved when recombinant BgaA, NanA and StrH were added to the mutants demonstrating that these three enzymes are surface exposed. King and co-workers concluded that enzymes such as BgaA, StrH and NanA were essential for the exposure of mannose in human secretory component (hSC) (King *et al*., 2006).

In an attempt to understand if BgaC has a similar role in the mucin degradation, as that reported by King and co-workers for the β-galactosidase A, a 48h culture of loss of function mutant SPD0065M in Sicard's defined medium supplemented with mucin was analysed by high-pH anion exchange chromatography using a HPLC system (as described in the Material and Methods Chapter, section XVII). The ability of SPD0065M to release galactose and N-acetylgalactosamine from the mucin glycoprotein was decreased when compared to the wild type however the mutant and the wild type released the same amount of sialic acid. The reduction in the ability to cleave galactose is good evidence that this enzyme is responsible for the release of galactose from the mucin side chains.

In the mucin structure, galactose is linked to N-acetylgalactosamine (β1,3) in the core and more peripherally to N-acetylglucosamine (β1,4). Assuming the proposal that deglycosylation occurs sequentially is correct, a reduction in release of both galactose and N-acetylgalactosamine suggested that BgaC preferentially targets the sugar residues present in the core-1 disaccharide rather than the more peripheral galactose linked to N-acetylglucosamine (β1,4). Since the non cleavage of galactose which is linked to N-acetylgalactosamine would render N-acetylgalactosamine inaccessible. It is interesting that there was a partial reduction in galactose release rather than a complete abolition or no impact. The partial reduction implies that BgaC works with BgaA in mucin degradation.

Recently Marion and co-workers (Marion *et al.*, 2009) showed that another glycosidase is involved in glycoprotein degradation by pneumococci. The authors reported that a loss of function mutant in gene *SP0368,* encoding for an O-glycosidase, was able to release sialic acid from core-1 O-linked glycans but was unable to release galactose (β1,3) N-acetylgalactosamine. They also found that a double loss of function mutant in O-glycosidase and NanA was unable to modify O-linked glycans (Marion *et al.*, 2009), demonstrating that release of sialic acid is paramount for the sequential deglycosylation to occur and that pneumococci is able to release disaccharides such as galactose (β1,3) N-acetylgalactosamine by the action of this O-glycosidase, producing an available substrate for BgaC.

BgaC was found to cleave only ONP-β-galactopyranoside. Interestingly, the enzyme did not cleave the substrate ONP-α-galactopyranoside thereby showing strict substrate specificity for β-linkage. The recombinant enzyme BgaC was tested for release of galactose using porcine gastric mucin as a test glycoprotein, however

although there is β-linked galactose on the side chains of this mucin, there was no detectable release of galactose. This was a possible outcome of the experiment because PGM may have its side chains protected by terminal sialic acid residues. When the sialic acid is present, and assuming the sequential deglycosylation of glycoproteins suggested by King *et al.*, (2006), the absence of previous treatment with NanA would mean that the galactose present was inaccessible.

As an alternative complex substrate, desialylated fetuin was used (Figure 3). Unlike other model glycoproteins that have been used in the study of pneumococcal galactosidases e.g. human α-1 acid glycoprotein, which contain galactose residues predominantly in β-1,4-linkage to N-acetylglucosamine (Burnaugh *et al*., 2008), fetuin oligosaccharides contain this sugar in both β-1,4- and β-1,3 linkage to N-acetylglucosamine (N-linked) (Figure 3, red rectangles and blue oval) and in β-1,3 -linkage to N-acetylgalactosamine (O-linked) (Figure 3, green oval).

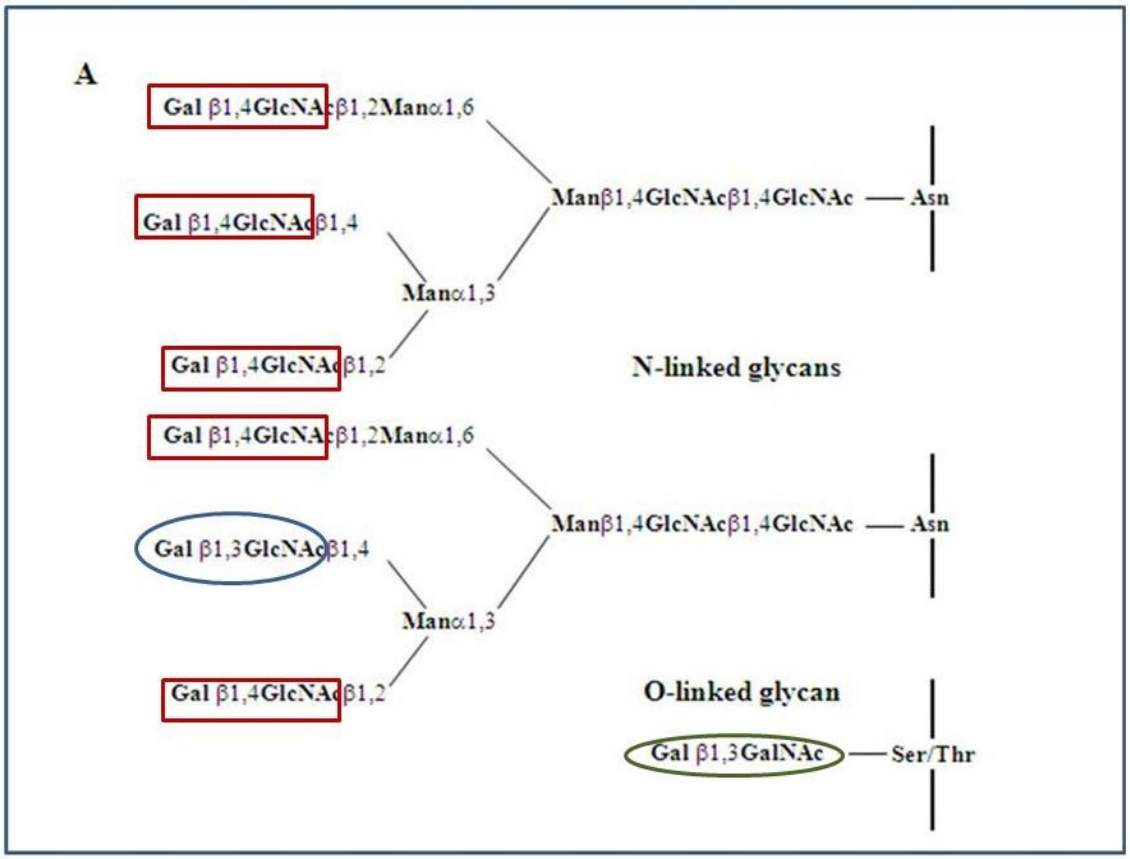


Figure 3- Glycan structures of desialylated fetuin. Adapted from references (Green *et al.*, 1988, Takasaki & Kobata, 1986)

Galactose β-1,3 N-acetylglucosamine (N-linked) is a peripheral sugar accessible once sialic acid has been removed. Galactose β-1, 3 N-acetylgalactosamine is core-1 O-linked structure present in fetuin but also present in the human airway, including on IgA making it a relevant substrate. In core-1, once the sialic acid is removed, galactose is also accessible. Using desialylated fetuin, it was possible to demonstrate the release of galactose from the glycoprotein. However, in spite of being a method used for determination of the role of glycosidases in the degradation of glycoproteins, this approach to measure the sugars release does not allow determination of the linkage from where the galactose is being released (β-1,4- or β-1,3-linkage).

In an attempt to determine if the galactose released was β-1,4 or β-1,3 linked, an assay using lectins that bind specific bonds between sugars was performed. By using

PNA (peanut agglutinin) and DSA (*Datura stramonium* agglutinin), it was possible to identify that BgaC cleaves galactose linked β1, 3 to N-acetylgalactosamine. However, due to the lack of a lectin that would recognise this disaccharide it was not possible to determine if the galactose was also being released from galactose linked β-1, 3 to N-acetylglucosamine. Previously BgaA was identified as being responsible for the cleavage of galactose β-1, 4 N-acetylglucosamine (King *et al.*, 2006). Using the lectin binding assay it was possible to rule out galactose β-1, 4 N-acetylglucosamine as a substrate for BgaC. Similar observations were made by Jeong *et al.*, (2009). The authors observed that BgaC could release galactose linked β1, 3 but it could not release galactose linked β1, 4. The authors also reported that BgaA could only release galactose linked β1, 4. Their observations for BgaA, combined with the strict specificity of BgaC for galactose linked β(1,3), eliminated the possibility of redundant functions for these enzymes. The observations made in this thesis, and supported by the study of Jeong *et al*., (2009) regarding the ability of BgaC to remove galactose from previously desialylated glycoproteins, show for the first time that there is more than one β-galactosidase in *S. pneumoniae* strain D39 responsible for the cleavage of galactose from the mucin side chains.

Knock-out mutant SPD0247M showed reduced, but significant, release of sialic acids, with approximately 20% of it removed after 48h of incubation with mucin. At this point in the project, gene *SPD0247* was annotated as a glycosidase and therefore release of other sugars was also investigated. However, no significant release of galactose, N-acetylgalactosamine, fucose or N-acetylglucosamine from PGM was observed. This result was expected since it was observed in the 14h growth experiments performed in Sicard's defined medium supplemented with mucin that mutant SPD0247M was unable to grow therefore implying that this mutant was either unable to

cleave the sugars present in the mucin structure or it was unable to utilise them. Thus the observed release of sialic acid may have been achieved when pneumococci were first exposed to mucin. However, because the mutant SPD0247M was unable to grow in Sicard's defined medium supplemented with mucin, it was not possible to assign a role for this enzyme based in this experiment.

In an attempt to overcome the impediment of the inability of mutant SPD0247M to grow on mucin a similar experiment was performed but using the recombinant protein.

The *SPD0247* gene was shown to encode a 6-phospho-β-glucosidase using the synthetic substrate pNP-6-phospho-β-glucoside. It was also reported in this thesis that there was no detectable activity against non-phosphorylated hexoses. In an attempt to determine its linkage specificity, the enzyme was tested using pNP-6-phospho-α-glucoside and pNP-6-phospho-β-glucoside substrates. The results determined that this enzyme is specific for phosphorylated glucose with a β linkage. It seems that an *in vivo* substrate for this enzyme must contain a phosphorylated hexose but thus far, there are no reports showing that there are phosphorylated sugars in the mucin side chains and therefore based in the evidence this would make mucin an improbable substrate for BglA.

Despite the apparent lack of possible substrates for BglA in the mucin structure when this gene was knocked out the bacteria showed an inability to grow in Sicard's defined medium supplemented with mucin as the source of nutrient. Could this enzyme be involved in degradation of mono- or di-saccharides resultant from the degradation of the glycoprotein's polysaccharide?

The observation that this enzyme acted only on phosphorylated hexoses suggests that this enzyme cleaves disaccharides that have entered the cell through a PTS system.

When a disaccharide enters the cell through a PTS system it is transported and phosphorylated simultaneously. *S. pneumoniae* has many PTS transporters including cellobiose PTS transporters (Hoskins *et al.*, 2001) despite the fact that cellobiose is not found in mammalian tissues or fluids. However, mammalian tissues are rich in glycosaminoglycans (i.e. hyaluronic acid) which contain β-linked disaccharide repeating units. Structural analogues of cellobiose are released when glycosaminoglycans are degraded (Masuda, 1984). Degradation of hyaluronic acid by hyaluronate lyase also produces structural analogues of cellobiose (Pritchard *et al.*, 1994). These analogues may then be transported into the cell through the cellobiose PTS systems (McKessar & Hakenbeck, 2007). It seems reasonable to assume that the cellobiose PTS systems are used to transport structural analogues of cellobiose since these are the ones present in the host. In light of these studies, a possible function for the 6-P-β-glucosidase, BglA, could be cleavage of disaccharides resulting from glycoprotein degradation by the action of other enzymes. These disaccharides would then enter the cell through the PTS systems, which would phosphorylate the disaccharides as they are transported allowing the substrate recognition by BglA. The released monosaccharides could then be directed into the glycolytic pathway.

Despite the lack of evidence that BglA is involved in the initial stages of glycoprotein degradation, a final experiment was performed. Recombinant BglA was assessed for cleavage of PGM. Release of monosaccharides was not detected when desialylated PGM was treated with recombinant BglA. Although it is believed that this was due to the requirement of phosphorylated substrates, this could also have been due to complex structure of PGM where sulphation and acetylation are common, contributing to masking access to the substrate (Sheehan *et al.*, 1991, Corfield *et al.*, 1992). In order to exclude this possibility, the model glycoprotein desialylated fetuin

was used. The sugars present in fetuin are subject to little modification therefore sulphation and acetylation is not extensive as it is in mucin. As with mucin, release of sugars from desialylated fetuin by BglA was not detected. Based in the substrate specificity determined for this enzyme, and in the data acquired with both mucin and fetuin it appears that host glycoproteins do not contain a substrate for BglA and therefore the enzymes' involvement in the utilisation of glycoproteins does not appear to be related to degradation. It is proposed that its role is intracellular breakdown of products resulting from the action of other glycosidases able to degrade glycoproteins.

I.9 Contribution of β-galactosidase C to total pneumococcal β-galactosidase activity

In the light of the previous findings and in combination with the results obtained in the enzyme activity experiments, two other isogenic mutants were constructed: a double knock-out mutant SPD0065/0562M (where *SPD0065* encodes for β-galactosidase C and *SPD0562* encodes for β-galactosidase A) and a complemented mutant which was called SPD0065REV. The double knock-out mutant and the single mutants were used to determine the contribution of each β-galactosidase to the total β-galactosidase activity of pneumococci. Using this approach, a significant contribution from both β-galactosidases BgaA and BgaC to the total β-galactosidase activity was identified. Thus both BgaA and BgaC contribute, to the cleavage of β-galactose from the substrate. However, an unexpected result was that in the absence of both β-galactosidases (BgaA and BgaC), detectable residual activity remained, presumably due to the presence of other, currently unknown, genes coding for β-galactosidases or due to other enzymes with more relaxed substrate specificity that are able to metabolise the substrate ONP-β-galactopyranoside. In strain D39, BgaA activity makes up 46% and BgaC contributes 39% for the total β-galactosidase activity. Therefore, in this strain 15% of β-

galactosidase activity remains unaccounted. These results are not in agreement with those reported by Zahner and co-workers, where in strain TIGR4, 85% of the total activity of β-galactosidase was attributed to the homolog of *SPD0562* (BgaA) (Zahner & Hakenbeck, 2000). The authors reported that with a BgaA mutant, it was shown that BgaA-independent β-galactosidase activity was approximately seven-fold lower than that in the wild type, in the presence of lactose, which induces the activity of the enzyme BgaA (Zahner & Hakenbeck, 2000). The clear difference between the results of these studies indicates that different strains TIGR4 and D39 may use different combinations of enzymes to metabolise the same substrate or may use the same enzymes but in different proportions.

These observations led to the investigation of BgaC in *S. pneumoniae* serotypes 6B IO11966 strain and 19F 177 ST type OXC1261 strain. Since the two strains used for this study had not been sequenced, the approach used for the gene disruptions was different from the one used to construct the mutants in the D39 strain. The disrupted β-galactosidase gene was amplified from SPD0065M (knock out mutant constructed by insertion of a spectinomycin cassette) and the transformation was done as described in the Material and Methods chapter section III.3. The recombination into the genome was expected to disrupt any *SPD0065* homolog. Following this procedure, the mutants were tested for β-galactosidase activity using the synthetic substrate ONP-β-galactopyranoside. The results demonstrated a clear reduction on the activity of β-galactosidase for the serotype 6B IO11966 strain, demonstrating that the synthesis of this enzyme is not restricted only to the strain D39, and that this enzyme (BgaC homolog) contributes to the total β-galactosidase activity in the serotype 6B strain IO11966 as in the D39 strain.

As for the mutant constructed in the serotype 19F strain OXC1261, there was no significant difference between the wild type (OXC1261) and the loss of function mutant. This activity result suggested that the homolog of *SPD0065* in strain OXC1261 is not involved in the cleavage of ONP-β-galactopyranoside. In serotype 2 strain D39, for example, there is another β-galactosidase BgaA encoded by gene *SPD0562* also contributing to the total β-galactosidase activity. This may well be the case for serotype 19F, where the S*PD0065* homolog does not seem to contribute to the total β-galactosidase activity. The strain OXC1261 used for this study was not a sequenced strain, however it is possible that if this strain contained a homolog of *SPD0562* this would be the main contributor for the β-galactosidase activity similarly to what has been reported in serotype 4 strain TIGR4 where the homolog of *SPD0562* is responsible for 85% of the β-galactosidase activity.

Analysis of the pneumococcal genome sequences available at the Sybil website (http://strepneumo-sybil.igs.umaryland.edu/cgi-bin) showed that a β-galactosidase C appears in thirty four of the sequenced thirty five pneumococcal strains with amino acid sequence identity of 99.1% and that the β-galactosidase A is present in thirty three of these pneumococcal strains, with an amino acid sequence of 93.2%. These data suggest that both BgaA and BgaC are present and conserved across different strains of *S. pneumoniae* implying that these enzymes are important for pneumococcal maintenance. The differences in activity observed between strains may be due to proportional contribution of each β-galactosidase to the combined total β-galactosidase activity It would be interesting to study the evolutionary relevance of this proportional utilisation of different β-galactosidases.

II. Characterisation of recombinant proteins encoded by genes *SPD0065* and *SPD0247*

II.1 Kinetic characterisation for BgaC and BglA

Kinetic parameters, *Km* and Vmax, were determined for BgaC (β-galactosidase) and BglA (6-P-β-glucosidase) using synthetic substrates. It was determined that the *Km* for BgaC when using ONP-β-galactopyranoside substrate was 0.91 mM and the Vmax was 0.11 μmol.min.mg^{-1}. These values are similar to those reported for the *Streptococcus cremoris* (*Km*, 0.38 mM), *Streptococcus thermophilus* (*Km*, 0.25 mM) and *Streptococcus intermedius* (*Km*, 1.56mM) β-galactosidases (Homer *et al.*, 1994, Jagota *et al.*, 1981, Rao, 2006). BglA had a *Km* of 1.42 mM and Vmax of 0.80 μmol.min.mg^{-1}, which is comparable to a 6-P-β-glucosidase from *E.coli* (*Km* 0.44mM) (Thompson *et al.*, 1999) and to a 6-P-β-glucosidase from *Pectobacterium carotovorum* (*Km* 0.21 mM) using the same substrate used to determine *Km* for the work in this thesis with BglA (An *et al.*, 2005).

So far there are no kinetic parameters available for the enzymes involved in the degradation of the host glycoproteins, however it would be useful to do a complete kinetic parameters determination in order to better understand how these processes are happening in the host. For example, it would be extremely useful in order to be able to draw a comparison between BgaA and BgaC to know the *Km* for BgaA. This would help answer which of the enzymes had a higher affinity for the substrate aiding in the understanding of the importance of each one of them for the total β-galactosidase activity. A more profound comprehension of pneumococcal metabolism is necessary and for that one step forward would be to determine kinetic parameters to all enzymes

involved in host glycoproteins degradation giving a realistic view of how efficient these processes are and which enzymes are more active and where.

Using a computational approach it is now possible to adjust a non linear trendline to the experimental points of a kinetic assay and determine the kinetic parameters *Km* and Vmax directly from the Michaelis Menten plot. However, these parameters were also calculated using conventional approaches, such as Lineweaver Burk and Eadie Hofstee, which allow the linearisation of the Michaelis Menten plot. Lineweaver Burk is a double reciprocal plot and is the most widely used approach. The disadvantage of this method is that it places undue weight on the experimental values obtained at lower concentrations of substrate but at these points the calculations of the rates are less precise because there are fewer products produced. The Eadie Hofstee approach was also used because it overcomes the problem of undue weight seen in the Lineweaver Burk approach; however, this method has its own disadvantage. In this method neither of the coordinates is independent, since it is a plot of reaction velocity vs. velocity/ substrate concentration, so any experimental error associated with rate determination will be present in both axes. Notwithstanding their disadvantages both approaches confirmed the results obtained when using the computational method described above.

II.2 Subcellular localization of BgaC and BglA

Pneumococcal glycosidases known to be involved on the sequential degradation of glycoproteins, such as NanA, BgaA and StrH, have been shown to have a surface localisation being anchored to the cell wall through a LPXTG motif (Bergmann & Hammerschmidt, 2006).The results demonstrated a cell surface localisation for β-galactosidase (BgaC) and a membrane association for 6-P-β-glucosidase (BglA); however the anchoring system requires further investigation. The findings presented in

this thesis for BgaC localisation are in accordance with those described by Jeong and co-workers, who reported the same subcellular localisation (Jeong *et al.*, 2009). Both this study and that of Jeong and co-workers observed that despite the apparent lack of signal peptide or typical membrane anchorage these proteins occur on the cell surface.

Two common systems for anchoring surface proteins are through a LPXTG motif and via choline binding domains (Fischetti *et al*., 2000, Yother & White, 1994). Anchoring through LPXTG implies the presence of an amino-terminal signal peptide, a carboxy-terminal hydrophobic region containing a charged tail where the LPXTG motif is found. This LPXTG motif is then recognised by a transpeptidase called sortase which anchors these proteins covalently to the peptidoglycan of the cell wall (Navarre & Schneewind, 1999). The use of choline binding repeats relies on the interaction of the choline binding repeats with the phosphorylcholine of the lipoteichoic acid of the membrane (Yother & White, 1994). Bioinformatic analysis showed that BgaC and BglA do not possess a signal peptide, a LPXTG motif or choline binding domains. Recently, it was reported that there is a new class of surface proteins in Gram positive bacteria that do not have the typical anchorage systems (Chhatwal, 2002). An example is the pneumococcal surface protein PavA which despite lacking obvious motifs for anchorage and also lacking a signal peptide, is displayed on the surface (Bergmann & Hammerschmidt, 2006). Another example is α-enolase, which despite the lack of signal peptide, LPXTG motif or choline binding repeats is displayed on the surface of pneumococci (Bergmann *et al.*, 2001). The authors hypothesised that the protein is secreted despite the lack of signal peptide and is then re-associated by interacting with receptors in the surface of both encapsulated and unencapsulated pneumococcal strains. The unknown mechanism used by pneumococcal cells for secretion and re-association of α-enolase and PavA with the pneumococcal surface could be one of the mechanisms

used by the bacterial cell to transport and make β-galactosidase C a surface protein. According to these findings it seems possible that pneumococci may utilise some novel and yet uncharacterised mechanism of export and surface retention of its proteins to anchor β-galactosidase C.

III. Pneumococcal Virulence

III.1 BgaC and BglA contribution to pneumococcal virulence

The influence of BgaC on pneumococcal virulence was investigated. The loss of function mutant was tested for the ability to cause disease when compared to the wild type pneumococcal strain D39. It was observed that there were no significant differences in terms of survival of intranasally infected mice. However, there was a delay in the appearance of bacteraemia in the group infected with mutant SPD065M. The differences were significant when the numbers of pneumococci in the nasopharynx were measured. For the group infected with SPD0065M, there were significantly less numbers in the nasopharynx at 12, 24 and 48 hours post-infection, suggesting a site specific role for BgaC. However, in the other analysed tissues (lungs and blood) there was no significant difference in the levels of pneumococci measured. Previously it was reported by Manco *et al.*, (2006) and Orihuela *et al.*, (2004) that depending on the route of infection an enzyme may or not play a role in the development of disease. In order to clarify if the role of BgaC is restricted to the intranasal route of infection and to the nasopharynx, the virulence of mutant SPD0065M could be investigated by using three other different routes of infection: intravenously, intraperitoneally and intratracheally.

The virulence-related findings in this thesis disagree with those of Jeong and co-workers (Jeong *et al.*, 2009). In their case the loss of function mutant of BgaC increases

nasopharyngeal colonisation, and the onset of bacteraemia occurred earlier in the mice infected intranasally with the mutant compared to the wild type (Jeong *et al.*, 2009). Jeong and co-workers used CD1 mice and 2 x 10^7 cfu/mouse in 10 µl and here MF1 mice and 5 x10^5 cfu/mouse in 50 µl were used. The differences observed between the data presented in this thesis and the data presented by Jeong *et al.*, (2009) could be attributed to the mice strain the cfus and the volume of the dose. Differences in mouse strain were shown to have a severe impact in how the pneumococcal disease progresses. Gingles *et al.*, (2001) investigated the progress of pneumococcal disease in inbred strains and concluded that the mouse strains used could be grouped into resistant, intermediate and susceptible based in the survival times and load of pneumococci in blood. This establishes that different mice strains exposed to the same load of pneumococci can develop disease in a different way. Both MF1 and CD1 mice are outbred mice. MF1 mice are widely used in pneumococcal infections because they are well known for their susceptibility to both intranasal (Gingles *et al.*, 2001, Canvin *et al.*, 1995, Chiavolini *et al.*, 2003) and intravenous infections (Iannelli *et al.*, 2004). CD1 mice are also susceptible to invasive pneumococcal disease via the intranasal route (Morona *et al.*, 2006); however this is a less well-known model. Differences in dose and dose volume are also a possible reason for the differences observed. A higher dose such as the one used by Jeong and co-workers has the ability to mask any virulence differences. Moreover, there is the remarkable difference in the inoculums volume. A 10 µl volume confines the infection to the nasopharynx whereas in this study a 50 µl dose was used which allows fulminant infection with concomitant sepsis (Canvin *et al.*, 1995). Differences in obtained phenotypes have been widely reported when using same loss of function mutants but different doses and different strains of mice. For example, Manco *et al.*, (2006) demonstrated a role for NanA and NanB in the development of

both upper and lower respiratory tract infection however, these findings were in sharp contrast with those of King and co-workers (King *et al.*, 2004) where the authors reported that there was no differences in the levels or persistence in the nasopharynx between the wild type and the loss of function mutant NanA. In their study, the route of infection was the same as the one used by Manco *et al.,* (2006) but there was clear discrepancy in the animals' age. There was also a dose difference, King and co-workers (King *et al.*, 2004) used a high dose ($1.8x10^7$CFU) which could have masked the virulence differences. The same was observed in our study and in contrast with that of Jeong *et al.*, (2009). In their study a dose of 2 x 10^7 CFU/mouse was used whereas in the data presented in this thesis a lower dose was used 5 x 10^5 CFU/mouse. This difference could be enough to enhance nuances in virulence; a higher dose may mask small differences observed throughout the onset of disease.

Following the observations reported in this thesis regarding the roles of BgaC and BgaA in glycoprotein degradation, a double knock-out mutant was constructed and its virulence was assessed. A 14h increase of survival ($p<0.01$) was observed in mice infected with the double knock-out mutant SPD0562K/65SM when compared to the wild type. For mice infected with the single knock-out mutant SPD0065M the increase in survival was of 4h. Bacteraemia was also followed in mice infected with the double knock-out mutant. The results showed lower numbers of pneumococci in blood throughout the experiment: this numbers never reached the levels obtained for the wild type. One interesting fact is that for the single mutant SPD0065M there was a delay in pneumococci seeding into blood, however once in blood they grew as well as D39 contrary to what was observed for the double knock-out mutant. Nevertheless the numbers of double knock-out mutant observed in the blood, although low, were sufficient to cause disease and subsequently death of the animals.

These observations seem to show that the combination of knock-out mutations in pneumococci increased the survival of mice when infected intranasally, however it did not prevent the ultimate death of the animals. The observation of lower levels of the double knock out mutant in blood could have two explanations. It could point to a role for BgaA in the blood since lower numbers of pneumococci in blood were not observed when mice were infected with the single knock-out mutant SPD0065M. Alternatively it may reflect a role for BgaA in the capacity of pneumococcal transfer from the lungs to the blood.

Even if BgaC and BgaA are involved in the maintenance of pneumococci in the nasopharynx through their role of degrading host glycoproteins to provide nutrient, a complete loss of the ability to persist and cause disease was not the expected result, because the presence of other enzymes was predicted to enable some deglycosylation to continue. For example, pneumococci would still have functional NanA, StrH and O-glycosidase, enzymes previously shown to have a role in glycoprotein degradation. In support of this contention, Manco *et al.*, (2006) demonstrated that mice infected intranasally with a loss of function mutant of NanA did not succumb to infection. Since NanA is responsible for the release of sialic acid typically the first sugar of host glycoproteins than loss of this enzyme would be expected to have a more profound impact than the loss of any enzyme involved in subsequent steps in the degradation of glycoprotein.

Involvement of BglA in pneumococcal virulence was also investigated. A significant increase in survival of mice ($p<0.01$) was observed when infected with a BglA mutant. Infection with the mutant strain SPD0247M resulted in a survival time of 72 ± 12h whereas after infection with the wild type the survival time was 48 ± 15h. The

differences of numbers of SPD0247M pneumococci compared to the wild type in the blood and lungs were also significant. The colony forming units from the lungs were different at time 24 and 48h ($p<0.001$) of the experiment but not at times 0 and 12h ($p>0.05$). The numbers of SPD0247M at the earlier time points of the experiment were not different from those of the wild type in the lungs suggesting a difficulty to grow once in the lungs rather than a difficulty in invading the lungs. Similar observations were made for the numbers of SPD0247M in blood, where they were significantly lower at times 24 and 48 h post-infection ($p<0.05$). These data suggest a role for BglA at both sites. If the proposed role for the BglA, namely cleavage of disaccharides which are transported through the PTS systems is correct, then due to the presence of abundant β-glucosides in the lungs and blood it is possible that the lack of this enzyme in the single knock-out mutant will interfere with the cleavage of the disaccharides diminishing the flux of nutrients and therefore impairing the establishment of pneumococci in the lungs consequently having an effect in the overall virulence of the mutant.

IV. Immunisation with BgaC and BglA

Immunisations were carried out after observing delay in the onset of disease upon infection with single pneumococcal knock-out mutants SPD0247M and SPD0065M. Immunisation with BgaC and BglA delayed the onset of disease after infection with the wild type D39 ($p<0.05$). Thus the immunisation experiments supported the conclusions of the virulence experiments performed with the mutants that these enzymes are involved in the pathogenesis of disease. Either these immunisations result in antibodies that neutralised the enzyme, thereby creating a functional mutant, or they are opsonic (these are not mutually exclusive possibilities).Whatever is the mechanism of action of the antibodies one would expect them to be active against secreted or surface exposed

proteins. Therefore, the results obtained in the immunisation experiments performed with the BgaC were expected. On the other hand, since BglA is thought to be an intracellular enzyme associated with the membrane, responsible for the cleavage of disaccharides after transport through the PTS systems, a delay in the onset of disease was an unexpected result. The observed phenotype led to speculation regarding the localisation of this enzyme. Is this enzyme placed in various different locations in the pneumococcal cell? Alternatively, could this be a case of antibody cross reactivity? The cross reactivity hypothesis seems less likely because the localisation of the enzyme in the pneumococcal cell was performed by using antibodies raised during these immunisation/infection experiments and only a single band was seen in Western blots. On the other hand, it may be possible that, when in the host, BglA is found at more than one site in the bacterium. So far and with the data obtained in this thesis, it is not possible to understand if the multi-placement is the reason for the phenotype observed without further experiments.

Presently, the best option for serotype independent immunity against pneumococcal disease is the protective potential of individual proteins. Recently, many efforts have been made to assess the protective potential of pneumococcal proteins such pneumolysin, CbpA and PspA, either singly or in combinations (Ogunniyi *et al.*, 2000). Immunisation with BgaC and BglA was not enough to prevent the death of the animals. In order to increase protection conferred by BgaC and BglA a combined immunisation would be a possibility. Previously Ogunniyi *et al.*, (2000) observed that mice immunised with the combination of a pneumolysin toxoid and PsaA survived significantly longer than animals immunised with the antigens alone, demonstrating that there was an additive protection. More recently Wu *et al.,*(2010) showed that antigens that seem to provide none or little protection against pneumococcal disease can, when

given in combination with others known to be highly immunogenic, protect (Wu *et al.*, 2010). For example, Wu and co-workers used combinations of ATP-dependent caseinolytic (ClpP) pneumolysin mutant (Al46Ply) and putative lipoate-preotein ligase (Lpl) to immunise mice, and showed that singly these enzymes could prolong the life of infected mice for half a day, however once combined the survival time of the infected mice increased to 10 days. Therefore, based in the evidence provided by Wu and co-workers it would be advantageous to combine two or more surface proteins known to be involved in glycoprotein degradation with a known immunogenic protein, such as pneumolysin or PsaA, and assay for protection. Because the studied enzymes are present in various pneumococcal strains from different serotypes these could, in addition to other pneumococcal proteins, be included in vaccination studies for serotype-independent protection against pneumococcal disease.

V. Mutant SPD0247M ability to form biofilm

Biofilm formation at the mucosal surface is important for pneumococcal colonisation (Moscoso *et al.*, 2009). It has been shown that loss of function mutants in enzymes involved in glycoprotein degradation, such as NanA, BgaA and BgaC, are reduced in their ability to form *in vitro* biofilms (King, 2010). Recently Trappetti *et al.*, (2009) reported that an increase in free sialic acid contributes to efficient formation of biofilm and that the ability to form biofilm could be diminished when NanA inhibitors were added, directly implicating the glycosidase NanA in biofilm formation. In the work performed in this thesis, a reduced ability to adhere to plastic *in vitro* was seen for the mutant SPD0247M which does not express BglA. These observations were in line with the data reported by Kilic *et al*., (2004). This group showed that the *gom* locus is involved in *S. gordonii* adherence to plastic (Kilic *et al.*, 2004). The *gom* locus contains 15 ORFs, including a *fucA* (fucosidase), a *manA* (mannosidase), a *gomA* (hypothetical protein), *manB* (second mannosidase), a *gomR* (regulator protein), a *gomF* and *gomG* which are binding proteins dependent sugar transport membrane components, a *gomE* (sugar binding protein), a *gomB* which is an hypothetical protein a *gomH*, a two component sensor histidine kinase a *gomD* which is a two component response regulator, a *Bglj* which is a β-glucosidase, a *gomL* (VirJ-like sensor histidine kinase) and a *glmS* (glucosamine-6-P synthase). The *gom* regulon has been described as responsible for the metabolism of glycoprotein-derived oligosaccharides (Kilic *et al.*, 2004). These data support the growth data collected for the SPD0247M mutant in Sicard's defined medium supplemented with mucin, in which a mutant in BglA could not grow. The *gom* locus also codes for BglJ a homolog, of BglA with 88% identity at the amino acid level (Figure 4). When Bglj was mutated, Kilic and co-workers saw a

decrease of the ability of the mutant to adhere to plastic. Similar observations were made in this thesis, when a loss of function mutant in BglA was tested for the ability to adhere to plastic, a reduction was seen. Phylogenetic analysis by Kilic *et al*., (2004) showed that Bglj is more similar to BglA from *S. pneumoniae* than to the other three β-glucosidases present in *S. gordonnii* suggesting that BglA could have a similar role in adherence as the one observed for Bglj.

Reduction in adherence by SPD0247M mutant was not an expected result. BglA is thought to be an intracellular enzyme and therefore less likely to have a role in biofilms formation, furthermore recombinant enzyme provided externally did not restore the ability of adherence of mutant SPD0247M supporting the proposed intracellular role for BglA. One question that can be raised is how could an intracellular protein have such an impact in the ability of *S. pneumoniae* to form biofilm? Would the lack of this enzyme shape the pneumococcal surface in a way that other adhesion proteins necessary for biofilms would not be present? Previously, it was seen that pyruvate oxidase has the ability to enhance nasopharyngeal colonisation in the transparent variant of pneumococci (Overweg *et al.,* 2000). Pyruvate oxidase is responsible for the descarboxylation of pyruvate to acetyl, H_2O_2 and CO_2. When pyruvate oxidase is not present, a reduction in the ability of pneumococci to colonise the nasopharynx was seen suggesting an indirect role for pyruvate oxidase in virulence. The same could be happening with BglA, perhaps the lack of this enzyme or of a product resulting from BglA enzymatic activity would be responsible for shaping the pneumococcal surface having an indirect role in the ability to form biofilm. This could be tested by comparing the proteome of the surface proteins in SPD0247M and to the proteome of the same surface proteins in the wild type. This comparison would give an answer as to what

proteins are present in the mutant and the wild type and of the ones that are not what their role could be in biofilm formation.

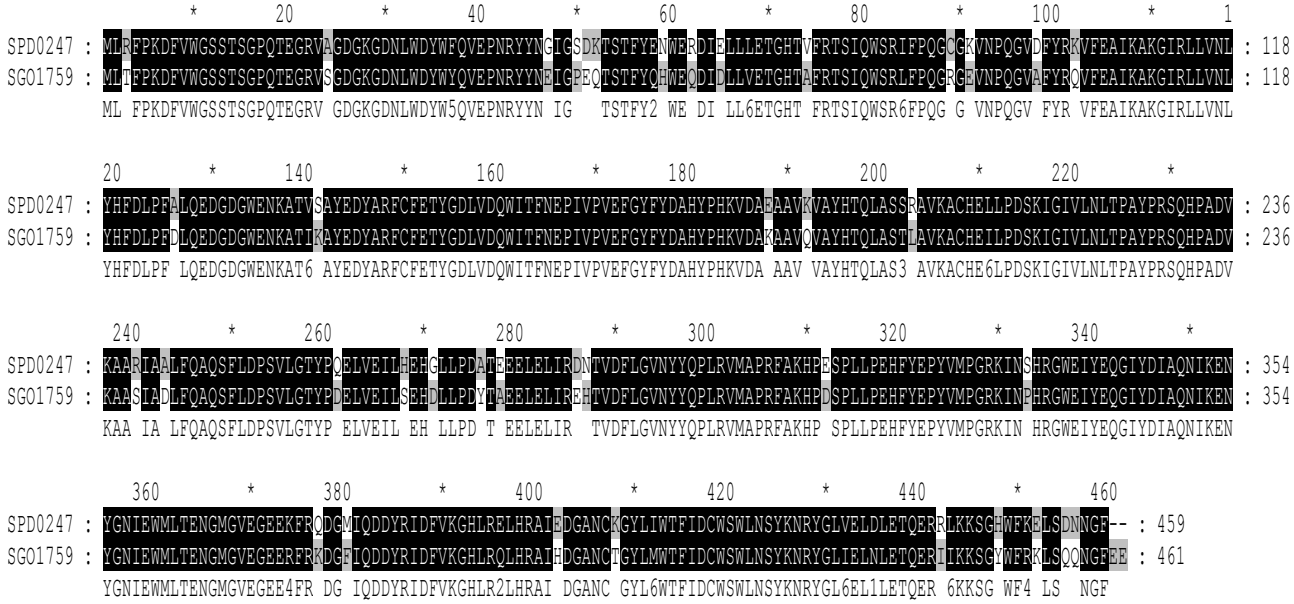


Figure 4 – Alignment of 6-P- β-glucosidase (BglA) encoded by gene *SPD0247* from *S. pneumoniae* and 6-P- β-glucosidase from *S. gordonii* (Bglj) encoded by gene *SGO1759* Alignment of amino acid sequences was done using GeneDoc: Analysis and Visualization of Genetic Variation

H Summary

In the work done for this thesis it was demonstrated that gene *SPD0065* encodes for a β-galactosidase C (BgaC) with a *Km* 0.91mM and a Vmax of 0.11 μmol.min.mg^{-1} and with an optimum pH in the range 6.5-7.0. BgaC was found to have a surface localisation despite the lack of signal peptide, LPXTG motif and choline binding domains. BgaC contributes 39% of the total β-galactosidase activity. BgaC was found to be responsible for the release of galactose linked β(1,3) to O- and N-linked glycoproteins. A BgaC homolog was found to contribute to total β-galactosidase activity in a serotype 6B strain but not in a serotype 19F strain type. Gene *SPD0065* was found to be over-expressed *in vitro* in the presence of mucin and *in vivo* in the lungs, blood and nasopharynx.

SPD0065M, a loss of function mutant in the gene *SPD0065*, had a decreased ability to grow in Sicard's defined medium supplemented with mucin as sole carbon source and it was shown to release significantly less galactose than the wild type from the model glycoprotein, fetuin. Furthermore, the survival of mice infected with SPD0065M was similar to the ones infected with the wild type strain but when the numbers of mutant pneumococci were measured in the infected tissues, they were significantly lower in the nasopharynx ($p<0.05$) than the pneumococcal numbers when infected with the wild type. Mice immunised with BgaC survived significantly longer than those immunised with alum adjuvant alone. The numbers of pneumococci were significantly lower in blood at 12h post challenge ($p<0.05$), however there was no significant difference at 0, 24 and 48h ($p>0.05$). Mice infected with the double mutant SPD0562K/65S also survived significantly longer than those infected with the wild type

with significantly lower numbers of pneumococci at 24h post infection. Finally, SPD0065M was complemented and a complete restoration of the wild type phenotype was observed.

The enzyme encoded by gene *SPD0247* was also characterised in this thesis. SPD0247 was found to encode for a 6-phospho-β-glucosidase (BglA), with a *Km* of 1.42mM and Vmax of 0.80 μmol.min.mg^{-1}. BglA is an enzyme associated with the membrane despite the lack of the conventional motifs. Gene *SPD0247* was over-expressed in the lungs and nasopharynx of infected animals but not in the blood. This gene was also over-expressed in the presence of mucin. The loss of function mutant SPD0247M was unable to grow in Sicard's defined medium supplemented with mucin as the source of carbon. SPD0247M was also unable to release carbohydrates from mucin and fetuin. It was then hypothesised in this thesis, that BglA is involved in intracellular degradation of disaccharides that enter the bacterial cell through a PTS system being phosphorylated as they are transported. Mice infected with loss of function mutant SPD0247M survived significantly more than those infected with the wild type. When the disease kinetics was followed it was observed that the numbers of pneumococci were significantly lower at 0 and 12h post infection in the lungs ($p<0.001$) and blood ($p<0.05$) but there was no significant difference in the nasopharynx. The protective potential of BglA was also investigated. Mice immunised with BglA survived significantly longer ($p<0.05$) than the animals immunised with alum adjuvant in PBS. The involvement of BglA in adherence was also considered. SPD0247M had a decrease ability to adhere to plastic suggesting a role for BglA in the ability to adhere.

I Future Work

The understanding of pneumococcal sugar metabolism when in the host requires further studies. It is now clear that *S. pneumoniae* needs glycosidases to degrade the host glycoproteins in order to provide nutrients. In this context, further work with the collection of mutants described in Chapter three of this thesis could be done. It would be interesting to pursue the reasons of why isogenic knocked-out mutants SPD044M and SPD0503M grew significantly better than the wild type in the presence of mucin.

Another interesting path to follow would be to study the regulation of the metabolic enzymes in the presence of glucose and of secondary sugars to clarify if the enzymes are regulated by CcpA.

In a more immediate future, it would be interesting to ascertain a definite function for BglA. In order to achieve that aim growth in Sicard's defined medium in the presence of disaccharides such as cellobiose or lactose and in the presence of hyaluronic acid could be done. If the enzyme is responsible for the intracellular breakdown of disaccharides phosphorylated as they are transported through the PTS systems a decrease in growth for mutant SPD0247M should be observed in the presence of cellobiose and hyaluronic acid.

Another interesting follow-up study would be to do investigate the degree of immunity conferred by combinations of enzymes such as BgaC or BglA and pneumolysin, since it was demonstrated in this study that immunisation with BgaC and BglA increased survival of MF1 mice.

Genes sent for sequencing. Figure 1 and Figure 2 are relative to gene *SPD0247*.

Figure 3 is relative to the other gene of interest gene *SPD0065*.

As mentioned in Chapter 4 the first colony presented a mutation. The mutation is marked in red.

Alignment with forward read sequence

```
Vanessa         CTAAGATTTCCAAAGGATTTTGTCTGGGGATCCTCTACTTCTGGACCGCAGACAGAAGGA 60
SPD0247         CTAAGATTTCCAAAGGATTTTGTCTGGGGATCCTCTACTTCTGGACCGCAGACAGAAGGA 60
                ************************************************************

Vanessa         CGTGTAGCTGGTGACGGTAAGGGAGACAATCTCTGGGATTACTGGTTCCAAGTGGAGCCA 120
SPD0247         CGTGTAGCTGGTGACGGTAAGGGAGACAATCTCTGGGATTACTGGTTCCAAGTGGAGCCA 120
                ************************************************************

Vanessa         AATCGTTACTATAATGGGATTGGTTCAGATAAGACATCGACTTTTTATGAAAATTGGGAG 180
SPD0247         AATCGTTACTATAATGGGATTGGTTCAGATAAGACATCGACTTTTTATGAAAATTGGGAG 180
                ************************************************************

Vanessa         CGGGATATTGAGCTTTTGTTAGAGACTGGTCACACAGTCTTTCGGACTTCTATTCAGTGG 240
SPD0247         CGGGATATTGAGCTTTTGTTAGAGACTGGTCACACAGTCTTTCGGACTTCTATTCAGTGG 240
                ************************************************************

Vanessa         TCACGGATTTTTCCACAAGGCTGTGGAAAAGTCAACCCTCAAGGTGTGGATTTTTATCGT 300
SPD0247         TCACGGATTTTTCCACAAGGCTGTGGAAAAGTCAACCCTCAAGGTGTGGATTTTTATCGT 300
                ************************************************************

Vanessa         AAGGTCTTTGAGGCTATTAAGGCTAAAGGAATTCGTCTGTTAGTCAATCTCTATCATTTT 360
SPD0247         AAGGTCTTTGAGGCTATTAAGGCTAAAGGAATTCGTCTGTTAGTCAATCTCTATCATTTT 360
                ************************************************************

Vanessa         GATTTACCTTTTGCCCTTCAAGAGGATGGTGATGGTTGGGAAAATAAGGCGACAGTCTCA 420
SPD0247         GATTTACCTTTTGCCCTTCAAGAGGATGGTGATGGTTGGGAAAATAAGGCGACAGTCTCA 420
                ************************************************************

Vanessa         GCCTATGAAGACTATGCTCGTTTTTGTTTTGAGGCTTATGGAGATTTAGTGGATCAATGG 480
SPD0247         GCCTATGAAGACTATGCTCGTTTTTGTTTTGAGACTTATGGAGATTTAGTGGATCAATGG 480
                ********************************* ************************

Vanessa         ATTACCTTTAACGAGCCCATCGTTCCTGTAGAATTTGGTTATTTTTATGATGCCCATTAT 540
SPD0247         ATTACCTTTAACGAGCCCATCGTTCCTGTAGAATTTGGTTATTTTTATGATGCCCATTAT 540
                ************************************************************

Vanessa         CCACATAAGGTGGATGCAGAGGCAGCTGTTAAGGTAGCCTATCATACACAATTGGCCAGC 600
SPD0247         CCACATAAGGTGGATGCAGAGGCAGCTGTTAAGGTAGCCTATCATACACAATTGGCCAGC 600
                ************************************************************

Vanessa         AGCCGGGCTGTTAAGGCTTGCCATGAACTTTTGCCTGATTCCAAGATTGGGATTGTCCTC 660
SPD0247         AGCCGGGCTGTTAAGGCTTGCCATGAACTTTTGCCTGATTCCAAGATTGGGATTGTCCTC 660
                ************************************************************

Vanessa         AACTTGACACCGGCTTATCCACGTAGCCAGCATCCTGCTGATGTCAAGGCAGCTCGTATT 720
SPD0247         AACTTGACACCGGCTTATCCACGTAGCCAGCATCCTGCTGATGTCAAGGCAGCTCGTATT 720
                ************************************************************

Vanessa         GCGGCCCTTTTTCAGGCCCAATCTTTCTTANATCCATCTGTCTTGGGGACTTATCCACAG 780
SPD0247         GCGGCCCTTTTTCAGGCCCAATCTTTCTTAGATCCATCTGTCTTGGGGACTTATCCACAG 780
                ***************************** ****************************

Vanessa         GAGTTGGTAGAAATCTTGCATGAACACGGTCTTTTACCTGATGCTACAGA---------- 830
SPD0247         GAGTTGGTAGAAATCTTGCATGAACACGGTCTTTTACCTGATGCTACAGAGGAAGAGTTG 840
                ************************************************
```

Alignment with T7 terminator primer

```
Vanessa         ------------------------------------------------------------
SPD0247         CTAAGATTTCCAAAGGATTTTGTCTGGGGATCCTCTACTTCTGGACCGCAGACAGAAGGA 60

Vanessa         ------------------------------------------------------------
SPD0247         CGTGTAGCTGGTGACGGTAAGGGAGACAATCTCTGGGATTACTGGTTCCAAGTGGAGCCA 120

Vanessa         ------------------------------------------------------------
SPD0247         AATCGTTACTATAATGGGATTGGTTCAGATAAGACATCGACTTTTTATGAAAATTGGGAG 180

Vanessa         ------------------------------------------------------------
SPD0247         CGGGATATTGAGCTTTTGTTAGAGACTGGTCACACAGTCTTTCGGACTTCTATTCAGTGG 240

Vanessa         ------------------------------------------------------------
SPD0247         TCACGGATTTTTCCACAAGGCTGTGGAAAAGTCAACCCTCAAGGTGTGGATTTTTATCGT 300

Vanessa         ------------------------------------------------------------
SPD0247         AAGGTCTTTGAGGCTATTAAGGCTAAAGGAATTCGTCTGTTAGTCAATCTCTATCATTTT 360

Vanessa         ------------------------------------------------------------
SPD0247         GATTTACCTTTTGCCCTTCAAGAGGATGGTGATGGTTGGGAAAATAAGGCGACAGTCTCA 420

Vanessa         --------------------GTTTTGTTTTGANGCTTATGNAGATTTAGTGGATCAATGG 40
SPD0247         GCCTATGAAGACTATGCTCGTTTTTGTTTTGAGACTTATGGAGATTTAGTGGATCAATGG 480
                           ***********  ****** ******************

Vanessa         ATTACCTTTAACGAGCCCNTCGTTCCTGTAGAATTTGGTTATTTTTATGATGCCCATTAT 100
SPD0247         ATTACCTTTAACGAGCCCATCGTTCCTGTAGAATTTGGTTATTTTTATGATGCCCATTAT 540
                ****************** ****************************************

Vanessa         CCACATAAGGTGGATGCAGAGGCAGCTGTTAAGGTAGCCTATCATACACAATTGGCCAGC 160
SPD0247         CCACATAAGGTGGATGCAGAGGCAGCTGTTAAGGTAGCCTATCATACACAATTGGCCAGC 600
                ************************************************************

Vanessa         AGCCGGGCTGTTAAGGCTTGCCATGAACTTTTGCCTGATTCCAAGATTGGGATTGTCCTC 220
SPD0247         AGCCGGGCTGTTAAGGCTTGCCATGAACTTTTGCCTGATTCCAAGATTGGGATTGTCCTC 660
                ************************************************************

Vanessa         AACTTGACACCGGCTTATCCACGTAGCCAGCATCCTGCTGATGTCAAGGCAGCTCGTATT 280
SPD0247         AACTTGACACCGGCTTATCCACGTAGCCAGCATCCTGCTGATGTCAAGGCAGCTCGTATT 720
                ************************************************************

Vanessa         GCGGCCCTTTTTCAGGCCCAATCTTTCTTAGATCCATCTGTCTTGGGGACTTATCCACAG 340
SPD0247         GCGGCCCTTTTTCAGGCCCAATCTTTCTTAGATCCATCTGTCTTGGGGACTTATCCACAG 780
                ************************************************************

Vanessa         GAGTTGGTAGAAATCTTGCATGAACACGGTCTTTTACCTGATGCTACAGAGGAAGAGTTG 400
SPD0247         GAGTTGGTAGAAATCTTGCATGAACACGGTCTTTTACCTGATGCTACAGAGGAAGAGTTG 840
                ************************************************************

Vanessa         GAACTCATTCGTGATAATACGGTGGACTTCCTTGGTGTGAACTACTATCAACCTTTGCGT 460
SPD0247         GAACTCATTCGTGATAATACGGTGGACTTCCTTGGTGTGAACTACTATCAACCTTTGCGT 900
                ************************************************************

Vanessa         GTTATGGCTCCTCGATTTGCTAAGCATCCAGAGAGTCCACTCTTACCAGAACATTTTTAC 520
SPD0247         GTTATGGCTCCTCGATTTGCTAAGCATCCAGAGAGTCCACTCTTACCAGAACATTTTTAC 960
                ************************************************************

Vanessa         GAGCCTTATGTGATGCCTGGACGTAAAATCAATTCTCATCGTGGTTGGGAGATTTATGAG 580
SPD0247         GAGCCTTATGTGATGCCTGGACGTAAAATCAATTCTCATCGTGGTTGGGAGATTTATGAG 1020
                ************************************************************

Vanessa         CAAGGGATTTATGACATCGCCCAAAATATCAAGGAAAATTATGGCAATATTGAGTGGATG 640
SPD0247         CAAGGGATTTATGACATCGCCCAAAATATCAAGGAAAATTATGGCAATATTGAGTGGATG 1080
                ************************************************************
```

```
Vanessa         TTGACTGAGAATGGTATGGGTGTTGAAGGGGAAGAAAAATTCCGTCAAGATGGAATGATT 700
SPD0247         TTGACTGAGAATGGTATGGGTGTTGAAGGGGAAGAAAAATTCCGTCAAGATGGAATGATT 1140
                ************************************************************

Vanessa         CAAGATGATTACCGTATTGACTTTGTAAAAGGTCATCTTCGTGAACTTCACCGTGCCATT 760
SPD0247         CAAGATGATTACCGTATTGACTTTGTAAAAGGTCATCTTCGTGAACTTCACCGTGCCATT 1200
                ************************************************************

Vanessa         GAAGATGGTGCCAACTGTAAGGGCTACTTGATTTGGACCTTTATTGACTGCTGGTCATGG 820
SPD0247         GAAGATGGTGCCAACTGTAAGGGCTACTTGATTTGGACCTTTATTGACTGCTGGTCATGG 1260
                ************************************************************

Vanessa         CTCAACAGCTATAAAAATCGCTATGGTTTGGTCGAATTAGACTTGGAAACGCAAGAACGT 880
SPD0247         CTCAACAGCTATAAAAATCGCTATGGTTTGGTCGAATTAGACTTGGAAACGCAAGAACGT 1320
                ************************************************************

Vanessa         CGTCTGAAGAAATCAGGCCACTGGTTCAAAGAATTAAGCGATAATAATGGATTTTAAAAG 940
SPD0247         CGTCTGAAGAAATCAGGCCACTGGTTCAAAGAATTAAGCGATAATAATGGATTTTAA--- 1377
                ************************************************************

Vanessa         CTTGCGAATTCTGGCGGCCGCTTAATTAATTA 972
SPD0247         --------------------------------
```

Figure 1- Sequence of gene *SPD0247* in the clone presenting a mutation.

Alignment with forward read sequence

```
Vanessa          CTAAGATTTCCAAAGGATTTTGTCTGGGGATCCTCTACTTCTGGACCGCAGACAGAAGGA 60
SPD0247r         CTAAGATTTCCAAAGGATTTTGTCTGGGGATCCTCTACTTCTGGACCGCAGACAGAAGGA 60
                 ************************************************************

Vanessa          CGTGTAGCTGGTGACGGTAAGGGAGACAATCTCTGGGATTACTGGTTCCAAGTGGAGCCA 120
SPD0247r         CGTGTAGCTGGTGACGGTAAGGGAGACAATCTCTGGGATTACTGGTTCCAAGTGGAGCCA 120
                 ************************************************************

Vanessa          AATCGTTACTATAATGGGATTGGTTCAGATAAGACATCGACTTTTTATGAAAATTGGGAG 180
SPD0247r         AATCGTTACTATAATGGGATTGGTTCAGATAAGACATCGACTTTTTATGAAAATTGGGAG 180
                 ************************************************************

Vanessa          CGGGATATTGAGCTTTTGTTAGAGACTGGTCACACAGTCTTTCGGACTTCTATTCAGTGG 240
SPD0247r         CGGGATATTGAGCTTTTGTTAGAGACTGGTCACACAGTCTTTCGGACTTCTATTCAGTGG 240
                 ************************************************************

Vanessa          TCACGGATTTTTCCACAAGGCTGTGGAAAAGTCAACCCTCAAGGTGTGGATTTTTATCGT 300
SPD0247r         TCACGGATTTTTCCACAAGGCTGTGGAAAAGTCAACCCTCAAGGTGTGGATTTTTATCGT 300
                 ************************************************************

Vanessa          AAGGTCTTTGAGGCTATTAAGGCTAAAGGAATTCGTCTGTTAGTCAATCTCTATCATTTT 360
SPD0247r         AAGGTCTTTGAGGCTATTAAGGCTAAAGGAATTCGTCTGTTAGTCAATCTCTATCATTTT 360
                 ************************************************************

Vanessa          GATTTACCTTTTGCCCTTCAAGAGGATGGTGATGGTTGGGAAAATAAGGCGACAGTCTCA 420
SPD0247r         GATTTACCTTTTGCCCTTCAAGAGGATGGTGATGGTTGGGAAAATAAGGCGACAGTCTCA 420
                 ************************************************************

Vanessa          GCCTATGAAGACTATGCTCGTTTTTGTTTTGAGGCTTATGGAGATTTAGTGGATCAATGG 480
SPD0247r         GCCTATGAAGACTATGCTCGTTTTTGTTTTGAGGCTTATGGAGATTTAGTGGATCAATGG 480
                 ************************************************************

Vanessa          ATTACCTTTAACGAGCCCATCGTTCCTGTAGAATTTGGTTATTTTTATGATGCCCATTAT 540
SPD0247r         ATTACCTTTAACGAGCCCATCGTTCCTGTAGAATTTGGTTATTTTTATGATGCCCATTAT 540
                 ************************************************************

Vanessa          CCACATAAGGTGGATGCAGAGGCAGCTGTTAAGGTAGCCTATCATACACAATTGGCCAGC 600
SPD0247r         CCACATAAGGTGGATGCAGAGGCAGCTGTTAAGGTAGCCTATCATACACAATTGGCCAGC 600
                 ************************************************************

Vanessa          AGCCGGGCTGTTAAGGCTTGCCATGAACTTTTGCCTGATTCCAAGATTGGGATTGTCCTC 660
SPD0247r         AGCCGGGCTGTTAAGGCTTGCCATGAACTTTTGCCTGATTCCAAGATTGGGATTGTCCTC 660
                 ************************************************************
```

```
Vanessa          AACTTGACACCGGCTTATCCACGTAGCCAGCATCCTGCTGATGTCAAGGCAGCTCGTATT 720
SPD0247r         AACTTGACACCGGCTTATCCACGTAGCCAGCATCCTGCTGATGTCAAGGCAGCTCGTATT 720
                 ************************************************************

Vanessa          GCGGCCCTTTTTCAGGCCCAATCTTTCTTANATCCATCTGTCTTGGGGACTTATCCACAG 780
SPD0247r         GCGGCCCTTTTTCAGGCCCAATCTTTCTTAGATCCATCTGTCTTGGGGACTTATCCACAG 780
                 ***************************** ****************************

Vanessa          GAGTTGGTAGAAATCTTGCATGAACACGGTCTTTTACCTGATGCTACAGA---------- 830
SPD0247r         GAGTTGGTAGAAATCTTGCATGAACACGGTCTTTTACCTGATGCTACAGAGGAAGAGTTG 840
                 **********************************************
```

Figure 2- Sequence of the gene *SPD0247* on the second clone chosen for sequencing.

SPD0065 alignment with T7 promoter primer

```
Vanessa          ACACGATTTGAGATACGAGATGATTTCTATCTCGATGGAAAATCATTTAAGATTTTATCT 60
SPD0065          ACACGATTTGAGATACGAGATGATTTCTATCTCGATGGAAAATCATTTAAGATTTTATCT 60
                 ************************************************************

Vanessa          GGTGCCATTCATTATTTTAGGATTCCTCCAGAGGATTGGTATCATTCGCTCTATAACTTG 120
SPD0065          GGTGCCATTCATTATTTTAGGATTCCTCCAGAGGATTGGTATCATTCGCTCTATAACTTG 120
                 ************************************************************

Vanessa          AAGGCTCTTGGTTTTAATACGGTAGAGACTTATGTTGCTTGGAATTTACACGAGCCTCGT 180
SPD0065          AAGGCTCTTGGTTTTAATACGGTAGAGACTTATGTTGCTTGGAATTTACACGAGCCTCGT 180
                 ************************************************************

Vanessa          GAAGGTGAGTTTCATTTTGAAGGTGATCTGGATTTAGAGAAATTTCTCCAAATAGCGCAG 240
SPD0065          GAAGGTGAGTTTCATTTTGAAGGTGATCTGGATTTAGAGAAATTTCTCCAAATAGCGCAG 240
                 ************************************************************

Vanessa          GATTTGGGTCTCTACGCAATTGTGCGTCCGTCTCCATTTATCTGTGCGGAATGGGAATTC 300
SPD0065          GATTTGGGTCTCTACGCAATTGTGCGTCCGTCTCCATTTATCTGTGCGGAATGGGAATTC 300
                 ************************************************************

Vanessa          GGTGGCTTACCAGCTTGGCTCTTGACCAAGAACATGCGAATTCGCTCATCCGACCCAGCA 360
SPD0065          GGTGGCTTACCAGCTTGGCTCTTGACCAAGAACATGCGAATTCGCTCATCCGACCCAGCA 360
                 ************************************************************

Vanessa          TATATCGAGGCAGTTGGTCGCTACTATGATCAGTTATTGCCAAGACTGGTGCCTCGTTTG 420
SPD0065          TATATCGAGGCAGTTGGTCGCTACTATGATCAGTTATTGCCAAGACTGGTGCCTCGTTTG 420
                 ************************************************************

Vanessa          TTGAACAATGGTGGCAATATTCTCATGATGCAGGTTGAAAATGAGTATGGTTCTTACGGA 480
SPD0065          TTGAACAATGGTGGCAATATTCTCATGATGCAGGTTGAAAATGAGTATGGTTCTTACGGA 480
                 ************************************************************

Vanessa          GAAGATAAGGCTTACCTGAGAGCGATTCGACAGCTAATGGAAGAGTGTGGCGTAACCTGT 540
SPD0065          GAAGATAAGGCTTACCTGAGAGCGATTCGACAGCTAATGGAAGAGTGTGGCGTAACCTGT 540
                 ************************************************************

Vanessa          CCCCTCTTTACATCAGATGGTCCATGGCGAGCTACTCTGAAAGCTGGAACCTTAATTGAA 600
SPD0065          CCCCTCTTTACATCAGATGGTCCATGGCGAGCTACTCTGAAAGCTGGAACCTTAATTGAA 600
                 ************************************************************

Vanessa          GAGGACCTCTTTGTAACAGGAAACTTTGGNTCTAAGGCACCTTACAACTTTTCGCANATG 660
SPD0065          GAGGACCTCTTTGTAACAGGAAACTTTGGTTCTAAGGCACCTTACAACTTTTCGCAGATG 660
                 ***************************** ************************ ***

Vanessa          CAGGAATTCTTTGATGA-CATGGTAAGAAATGGCCACTCATGTGTATGGAGTTCTGGGAT 719
SPD0065          CAGGAATTCTTTGATGAACATGGTAAGAAATGGCCACTCATGTGTATGGAGTTCTGGGAT 720
                 ***************** ****************************************

Vanessa          GGTTGGTTCAATCGCTGGAAAGAACCGATTATCACACGGGATCCTAAGGAATTGG----- 774
SPD0065          GGTTGGTTCAATCGCTGGAAAGAACCGATTATCACACGGGATCCTAAGGAATTGGCAGAT 780
                 ************************************************
```

```
Vanessa         ------------------------------------------------------------
SPD0065         GCAGTTCGAGAGGTTTTGGAACAAGGCTCTATCAATCTTTACATGTTCCACGGTGGTGCA 840

Vanessa         ------------------------------------------------------------
SPD0065         AACTTTGGTTTCATGAATGGTTGCTCAGCTCGAGGAACTTTGGACCTGCCACAAGTTACA 900

Vanessa         ------------------------------------------------------------
SPD0065         TCTTATGATTACGATGCCCTTCTGGATGAAGAAGGAAATCCAACTGCTAAATATCTTGCA 960

Vanessa         ------------------------------------------------------------
SPD0065         GTCAAGAAGATGATGGCAACACATTTTTCAGAGTATCCGCAGTTGGAACCACTCTACAAA 1020

Vanessa         ------------------------------------------------------------
SPD0065         GAGAGTATGGAGTTGGATGCTATTCCACTAGTTGAAAAAGTTTCTTTGTTTGAAACCTTA 1080

Vanessa         ------------------------------------------------------------
SPD0065         GATAGCTTGTCAAGTCCTGTAGAAAGTCTCTATCCTCAAAAGATGGAGGAGCTGGGACAA 1140

Vanessa         ------------------------------------------------------------
SPD0065         AGTTATGGCTACCTACTTTATCGAACAGAAACAAACTGGGATGCAGAAGAAGAAAGACTT 1200

Vanessa         ------------------------------------------------------------
SPD0065         CGTATCATTGATGGTCGAGATAGGGCCCAGCTGTATGTCGATGGTCAGTGGGTTAAAACT 1260

Vanessa         ------------------------------------------------------------
SPD0065         CAATATCAGACAGAGATTGGGGAAGATATTTTTTATCAAGGTAAAAAGAAAGGGCTATCT 1320

Vanessa         ------------------------------------------------------------
SPD0065         AGGTTAGATATCTTGATAGAAAATATGGGGCGTGTCAACTATGGGCATAAGTTCTTAGCG 1380

Vanessa         ------------------------------------------------------------
SPD0065         GATACGCAACGTAAGGGAATTCGGACAGGGGTCTGTAAGGATCTGCATTTCTTACTAAAC 1440

Vanessa         ------------------------------------------------------------
SPD0065         TGGAAACACTATCCACTCCCACTAGACAATCCTGAGAAAATTGATTTTTCAAAAGGATGG 1500

Vanessa         ------------------------------------------------------------
SPD0065         ACTCAAGGACAACCAGCCTTTTACGCTTATGACTTTACAGTCGAAGAGCCAAAAGATACT 1560

Vanessa         ------------------------------------------------------------
SPD0065         TACCTAGACTTGTCTGAGTTTGGTAAGGGAGTTGCCTTTGTCAATGGGCAGAATCTAGGA 1620

Vanessa         ------------------------------------------------------------
SPD0065         CGTTTTTGGAACGTTGGCCCAACTCTCTCACTTTATATCCCTCATAGCTATCTCAAGGAA 1680

Vanessa         ------------------------------------------------------------
SPD0065         GGTGCCAACCGTATCATTATCTTTGAAACAGAAGGTCAATATAAAGAAGAGATTCATTTA 1740

Vanessa         ---------------------------------------------
SPD0065         ACTCGTAAACCTACACTAAAACATATAAAGGGGGAAAACTTATGA 1785
```

Figure 3- Sequence of gene SPD0065 from colony 1.

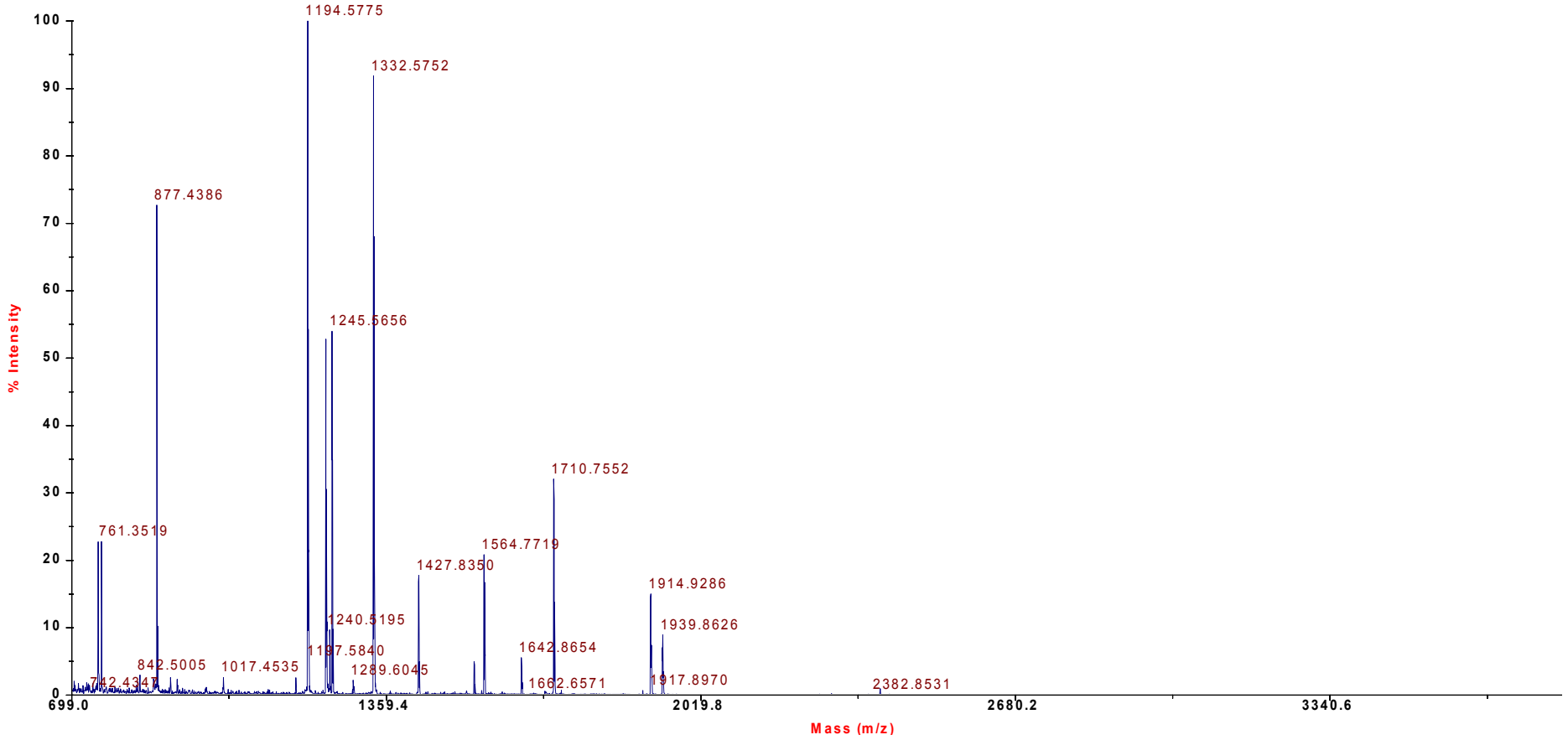


Figure 4- Spectrum obtained from the mass spectrometry for the purified BglA.

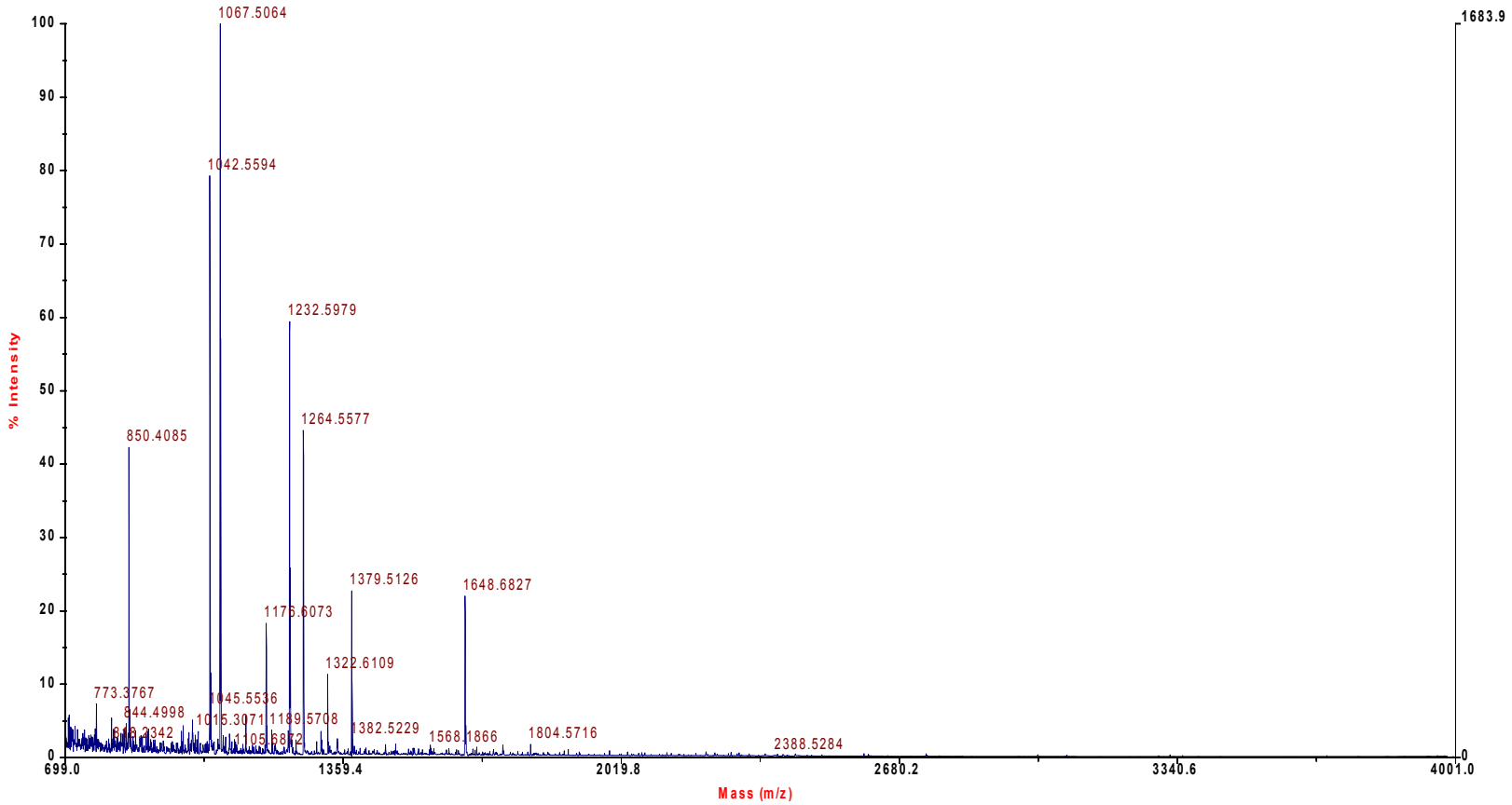
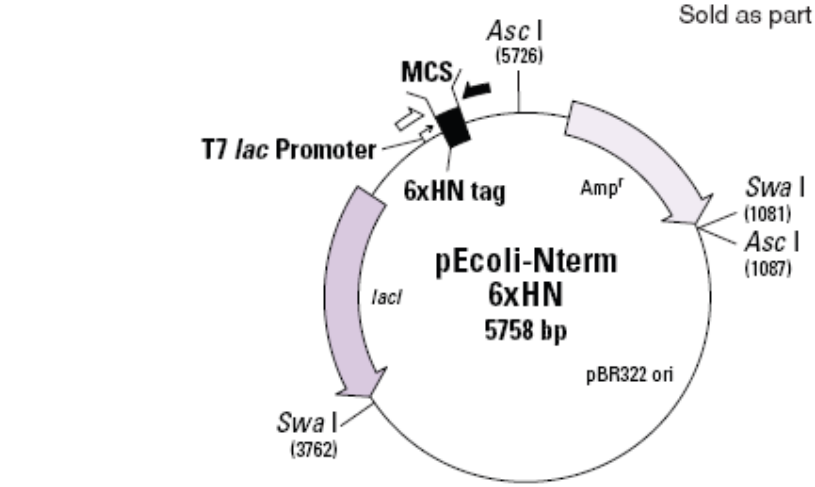


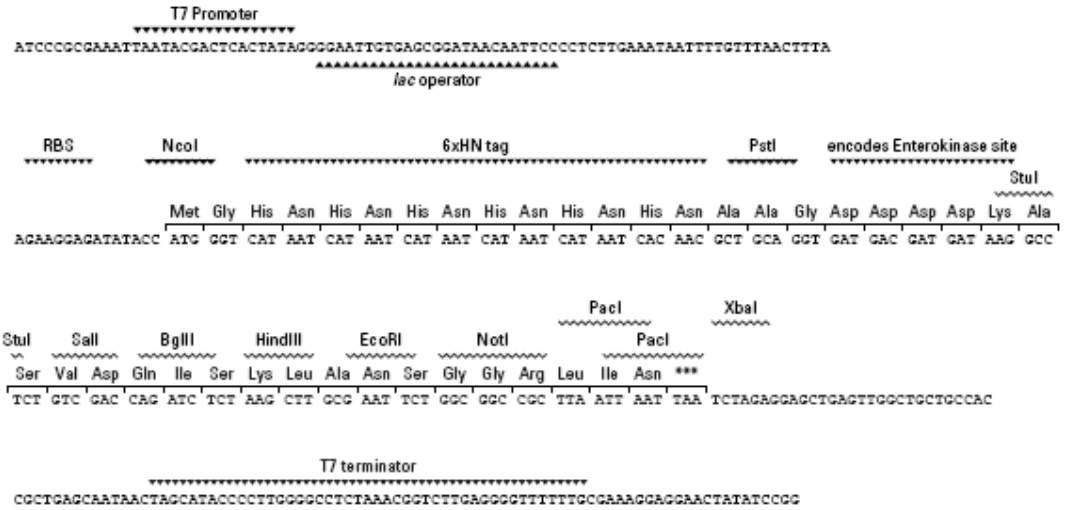
Figure 5- Spectrum obtained from the mass spectrometry for the purified BgaC.

- p*Ecoli* - Nterm 6X HN vector with ampicillin resistance marker

pEcoli-Nterm 6xHN Vector Information PT3868-5

Sold as part of Catalog No. 635639





Restriction Map and Multiple Cloning Site (MCS) of pEcoli-Nterm 6xHN Vector. All sites shown are unique.

Gene *SPD0065* nucleotide sequence (1788 nt)

atgacacgatttgagatacgagatgatttctatctcgatggaaaatcatttaagattttatctggtgccattcattattttaggatt
cctccagaggattggtatcattcgctctataacttgaaggctcttggttttaatacggtagagacttatgttgcttggaatttaca
cgagcctcgtgaaggtgagtttcattttgaaggtgatctggatttagagaaatttctccaaatagcgcaggatttgggtctcta
cgcaattgtgcgtccgtctccatttatctgtgcggaatgggaattcggtggcttaccagcttggctcttgaccaagaacatgc
gaattcgctcatccgacccagcatatatcgaggcagttggtcgctactatgatcagttattgccaagactggtgcctcgtttgt
tgaacaatggtggcaatattctcatgatgcaggttgaaaatgagtatggttcttacggagagataaggcttacctgagagc
gattcgacagctaatggaagagtgtggcgtaacctgtcccctctttacatcagatggtccatggcgagctactctgaaagct
ggaaccttaattgaagaggacctctttgtaacaggaaactttggttctaaggcaccttacaacttttcgcagatgcaggaatt
ctttgatgaacatggtaagaaatggccactcatgtgtatggagttctgggatggttggttcaatcgctggaaagaaccgatt
atcacacgggatcctaaggaattggcagatgcagttcgagaggttttggaacaaggctctatcaatctttacatgttccacg
gtggtacaaactttggtttcatgaatggttgctcagctcgaggaactttggacctgccacaagttacatcttatgattacgatg
cccttctggatgaagaaggaaatccaactgctaaatatcttgcagtcaagaagatgatggcaacacatttttcagagtatc
cgcagttggaaccactctacaaagagagtatggagttggatgctattccactagttgaaaaagtttctttgtttgaaacctta
gatagcttgtcaagtcctgtagaaagtctctatcctcaaaagatggaggagctgggacaaagttatggctacctactttatc
gaacagaaacaaactgggatgcagaagaagaaagacttcgtatcattgatggtcgagatagggcccagctgtatgtcg
atggtcagtgggttaaaactcaatatcagacagagattggggaagatatttttttatcaaggtaaaaagaaagggctatcta
ggttagatatcttgatagaaaatatggggcgtgtcaactatgggcataagttcttagcggatacgcaacgtaagggaattc
ggacaggggtctgtaaggatctgcatttcttactaaactggaaacactatccactcccactagacaatcctgagaaaattg
atttttcaaaaggatggactcaaggacaaccagccttttacgcttatgactttacagtcgaagagccaaaagatacttacct
agacttgtctgagtttggtaagggagttgcctttgtcaatgggcagaatctaggacgtttttggaacgttggcccaactctctc
actttatatccctcatagctatctcaaggaaggtgccaaccgtatcattatctttgaaacagaaggtcaatataaagaagag
attcatttaactcgtaaacctacactaaaacatataaaggggggaaaacttatga

BgaC amino acid sequence (595 aa):

MTRFEIRDDFYLDGKSFKILSGAIHYFRIPPEDWYHSLYNLKALGFNTVETYVAWNLHEPREGEFHFEGDLDL
EKFLQIAQDLGLYAIVRPSPFICAEWEFGGLPAWLLTKNMRIRSSDPAYIEAVGRYYDQLLPRLVPRLLNNGG
NILMMQVENEYGSYGEDKAYLRAIRQLMEECGVTCPLFTSDGPWRATLKAGTLIEEDLFVTGNFGSKAPYN
FSQMQEFFDEHGKKWPLMCMEFWDGWFNRWKEPIITRDPKELADAVREVLEQGSINLYMFHGGTNFGF
MNGCSARGTLDLPQVTSYDYDALLDEEGNPTAKYLAVKKMMATHFSEYPQLEPLYKESMELDAIPLVEKVSL
FETLDSLSSPVESLYPQKMEELGQSYGYLLYRTETNWDAEEERLRIIDGRDRAQLYVDGQWVKTQYQTEIGE
DIFYQGKKKGLSRLDILIENMGRVNYGHKFLADTQRKGIRTGVCKDLHFLLNWKHYPLPLDNPEKIDFSKGW

TQGQPAFYAYDFTVEEPKDTYLDLSEFGKGVAFVNGQNLGRFWNVGPTLSLYIPHSYLKEGANRIIIFETEGQ
YKEEIHLTRKPTLKHIKGENL

Gene *SPD0247* nucleotide sequence (1380 nt):

atgctaagatttccaaaggattttgtctggggatcctctacttctggaccgcagacagaaggacgtgtagctggtgacggta
agggagacaatctctgggattactggttccaagtggagccaaatcgttactataatgggattggttcagataagacatcga
ctttttatgaaaattgggagcgggatattgagcttttgttagagactggtcacacagtctttcggacttctattcagtggtcacg
gatttttccacaaggctgtggaaaagtcaaccctcaaggtgtggatttttatcgtaaggtctttgaggctattaaggctaaag
gaattcgtctgttagtcaatctctatcattttgatttacctttgcccttcaagaatggtgatggttgggaaaataaggcgacagt
ctcagcctatgaagactatgctcgtttttgttttgagacttatggagatttagtggatcaatggattcctttaacgagcccatcgtt
cctgtagaatttggttatttttatgatgcccattatccacataaggtggatgcagaggcagctgttaaggtagcctacatacac
aattggccagcagccgggctgttaaggcttgccatgaacttttgcctgattccaagattgggattgtcctcaacttgacaccg
gctatccacgtagccagcatcctgctgatgtcaaggcagctcgtattgcggccctttttcaggcccaatctttcttagatccat
ctgtcttggggacttatcacaggagttggtagaaatcttgcatgaacacggtcttttacctgatgctacagaggaagagttg
gaactcattcgtgataatacggtggacttccttggtgtgaactactatcaacctttgcgtgttatggctcctcgatttgctaagc
atccagagagtccactcttaccagaacatttttacgagccttatgtgatgcctggacgtaaaatcaattctcatcgtggttgg
gagatttatgagcaagggatttatgacatcgcccaaaatatcaaggaaaattatggcaatattgagtggatgttgactgag
aatggtatgggtgttgaaggggaagaaaaattccgtcaagatggaatgattcaagatgattaccgtattgactttgtaaaa
ggtcatcttcgtgaacttcaccgtgccattgaagatggtgccaactgtaagggctacttgatttggacctttattgactgctggt
catggctcaacagctataaaaatcgctatggtttggtcgaattagacttggaaacgcaagaacgtcgtctgaagaaatca
ggccactggttcaaagaattaagcgataataatggattttaa

BglA amino acid sequence (459 aa):

MLRFPKDFVWGSSTSGPQTEGRVAGDGKGDNLWDYWFQVEPNRYYNGIGSDKTSTFYENWERDIELLLE
TGHTVFRTSIQWSRIFPQGCGKVNPQGVDFYRKVFEAIKAKGIRLLVNLYHFDLPFALQEDGDGWENKATV
SAYEDYARFCFETYGDLVDQWITFNEPIVPVEFGYFYDAHYPHKVDAEAAVKVAYHTQLASSRAVKACHELL
PDSKIGIVLNLTPAYPRSQHPADVKAARIAALFQAQSFLDPSVLGTYPQELVEILHEHGLLPDATEEELELIRDN
TVDFLGVNYYQPLRVMAPRFAKHPESPLLPEHFYEPYVMPGRKINSHRGWEIYEQGIYDIAQNIKENYGNIE
WMLTENGMGVEGEEKFRQDGMIQDDYRIDFVKGHLRELHRAIEDGANCKGYLIWTFIDCWSWLNSYKNR
YGLVELDLETQERRLKKSGHWFKELSDNNGF

Gene *SPD0562* nucleotide sequence (6687 nt):

atggggaaaggccattggaatcggaaaagagtttatagcattcgtaagtttgctgtgggagcttgctcagtaatgattggg
acttgtgcagttctattaggaggaaatatagctggagaatctgtagtttatgcggatgaaacacttattactcatactgctgag
aaacctaaagaggaaaaaatgatagtagaagaaaaggctgataaagctttggaaactaaaaatgtagttgaaaggac
agaacaaagtgaacctagttcaactgaggctattgcatctagaagaaagaagatgaagccgtaactccaaaagagga
aaaagtgtctgctaaaccggaagaaaaagctccaaggatagaatcacaagcttcagtcaagaaaaaccgctcaagg
aagatgctaaagctgtaacaaatgaagaagtgaatcaaatgattgaaaacaggaaagtggattttaatcaaaattggta
ctttaaactcaatgcaaattctaaggaagccattaaacctgatgcagacgtatctacgtggaaaaaattagatttaccgtat
gactggagtatctttaacgatttcgatcatgaatctcctgcacaaaatgaaggtggacagctcaacggtggggaagcttgg
tatcgcaagactttcaaactagatgaaaaagacctcaagaaaaatgttcgccttacttttgatggcgtctacatggattctca
agtttatgtcaatggtcagttagtggggcattatccaaatggttataaccagttctcatacgatatcaccaaataccttacaa
agatggtcgtgagaatgtgattgctgtccatgcagtcaacaaacagccaagtagccgttggtattcaggaagtggtatcta
tcgtgatgtgactttacaagtgacagataaggtgcatgttgagaaaaatgggacaactattttaacaccaaaacttgaaga
acaacaacatggcaaggttgaaactcatgtgaccagcaaaatcgtcaatacggacgacaaagaccatgaacttgtag
ccgaatatcaaatcgttgaacgaggtggtcatgctgtaacaggcttagttcgtacagcgagtcgtaccttaaaagcacatg
aatcaacaagcctagatgcgattttagaagttgaaagaccaaaactctggaccgttttaaatgacaaacctgccttgtacg
aattgattacgcgtgtttaccgtgacggtcaattggttgatgctaagaaggatttgtttggttaccgttactatcactggactcca
aatgaaggtttctctttgaatggtgaacgtattaaattccatggagtatccttgcaccacgaccatggggcgcttggagcag
aagaaaactataaagcagaatatcgccgtctcaaacaaatgaaggagatgggagttaactccatccgtacaacccac
aaccctgctagtgagcaaaccttgcaaatcgcagcagaactaggtttactcgttcaggaagggcctttgatacttggtatg
gtggcaagaaaccttatgactatggacgtttctttgaaaaagatgccactcacccagaagctcgaaaaggtgaaaaatg
gtctgattttgacctacgtaccatggtcgaaagaggcaaaaacaaccctgctatcttcatgtggtcaattggtaatgaaata
ggtgaagctatggtgatgcccactctttagcaactgttaaacgtttggtcaaggttatcaaggatgttgataagactcgctat
gttaccatgggagcagataaattcgtttcggtaatggtagcggagggcatgagaaaattgctgatgaactcgatgctgttg
gatttaactattctgaagataattacaaagcccttagagctaagcatccaaaatggttgatttacggttcagaaacatcatca
gcaacccgtacacgaggaagttactatcgccctgaacgtgaattgaaacatagcaatggacctgagcgtaattatgaac
agtcagattatggaaatgatcgtgtgggttgggggaaaacagcaaccgcttcatggacttttgaccgtgacaacgctggct
atgctggacagtttatctggacaggtacggactatattggtgaacctacaccatggcacaaccaaaatcaaactcccgtta
agagctcttactttggtatcgtagatacagccggcattccaaaacatgacttctatctctaccaaagccaatgggtttctgtta
agaagaaaccgatggtacaccttcttcctcactggaactgggaaaacaaagaattagcatccaaagtagctgactcag
aaggtaagattccagttcgtgcttattcgaatgcttctagtgtagaattgttcttgaatggaaaatctcttggtcttaagactttca
ataaaaaacaaaccagcgatgggcggacttaccaagaaggtgcaaatgctaatgaactttatcttgaatggaaagttgc
ctatcaaccaggtaccttggaagcaattgctcgtgatgaatctggcaaggaaattgctcgagataagattacgactgctgg
taagccagcggcagttcgtcttattaaggaagaccatgcgattgcagcagatggaaaagacttgacttacatctactatga
aattgttgacagccaggggaatgtggttccaactgctaataatctggttcgcttccaattgcatggccaaggtcaactggtc

ggtgtagataacggagaacaagccagccgtgaacgctataaggcgcaagcagatggttcttggattcgtaaagcattta
atggtaaaggtgttgccattgtcaaatcaactgaacaagcagggaaattcacccttactgcccactctgatctcttgaaatc
gaaccaagtcactgtctttactggtaagaaagaaggacaagaaaagactgttttggggacagaggtgccaaaagtaca
gaccattattggagaggcacctgaaatgcctaccactgttcctttgtatacagtgatggtagtcgtgcagaacgtcctgtaa
cctggtctttagtagatgtgagcaagcctggtattgtaacggtgaaaggtatggctgacggacgagaagtagaagctcgt
gtagaagtgattgctcttaaatcagagctaccagttgtgaaacgtattgctccaaatactaacttgaattctgtagacaaatct
gtttcctatgttttgactgatggaagtgtacaagagtatgaagtggacaagtgggagattgccgaagaagataaagctaa
gttagcaattccaggttctcgtattcaagcgaccggttatttagaaggtcaaccaattcatgcaacccttgtggtagaagaa
ggcaatcctgcagcacctgtagtgccaactgttactgttggaggtgaagcagtaacaggtcttactagtcgacaaccaat
gcaatatcgtactctatcttatggtgcccaatgccagaagtcacagcaagtgctgaaaatgctgatgtgacagttcttcaag
caagcgcagcaaacggcatgcgtgcgagcatctttattcagccagatggtggccctcttcaaacctatgcaattcaattcc
ttgaagaagcgccaaaaattgctcacttgagcttgcaagtggaaaaagctgacagtctcaaagaagaccaaactgtca
aattgtcggttcgagctcactatcaagatggaacgcaagctgtattaccagctgataaagtaaccttctctacaagtggtga
aggggaagtcgcaattcgtaaaggaatgcttgagttgcataagccaggagcagtcactctgaacgctgaatatgaggg
agctaaaggccaagttgaactcactatccaagccaatactgagaagaagattgcgcaatctatccgtcctgtaaatgtag
tcacagatttgcatcaggaaccaagtcttccagcaacagtaacagttgagtatgacaaaggtttccctaaaactcataaa
gtcacttggcaagctattccgaaagaaaaactagactcctatcaaatatttgaagtactaggtaaagttgaaggaattgac
cttgaagcgcgtgcaaaagtctctgtagaaggtatcgtttcagttgaagaagtcagtgtgacaactccaatcgcagaagc
accacaattaccagaaagcgttcggacatatgattcaaatggtcacgtttcatcagctaaggttgcatgggatgcgattcgt
ccagagcaatacgctaaggaaggtgtctttacagttaatggtcgcttagaaggtactcaattaacaactaaacttcatgttc
gcgtatctgctcaaactgagcaaggtgcaaacatttctgaccaatggaccggttcagaattgccacttgcctttgcttcagat
tcaaatccaagcgacccagtttcaaatgttaatgacaagctcatttcctacaataaccaaccagccaatcgttggacaaa
ctggaatcgtagtaatccagaagcttcagtcggtgttctgtttggagattcaggtatcttgagcaaacgctccgttgataatct
aagtgtcggattccacgaagaccatggagttggtgcaccgaagtcttatgtgattgagtattatgttggtaagactgtccca
acagctcctaaaaaccctagttttgttggtaatgaggaccatgtctttaatgatctgccaactggaaaccagttactaatcta
aaagcccctgctcaactcaaggctggagaaatgaaccactttagctttgataaagttgaaacctatgctattcgtattcgca
tggttaaagcagataacaagcgtggaacgtctatcacagaggtacaaatctttgcgaaacaagttgcggcagccaaac
aaggacaaacaagaatccaagttgacggcaaagacttagcaaacttcaaccctgatttgacagactactaccttgagtc
tgtagatggaaaagtccggcagtcacagcaaatgttagcaacaatggtctcgctaccgtcgttccaagcgttcgtgaagg
tgagccagttcgtgtcatcgcgaaagctgaaaatggcgacatcttaggagaataccgtctgcacttcactaaggataaga
acttactttctcataaaccagttgctgcggttaaacaagctcgcttgctacaagtaggtcaagcacttgaattgccgactaag
gttccagtttacttcacaggtaaagacggctacgaaacaaaagacctgacagttgaatgggaagaagttccagcggaa
aatctgacaaaagcaggtcaatttactgttcgaggccgtgtccttggtagtaaccttgttgctgaggtcactgtacgagtgac
agacaaacttggtgagactctttcagataaccctaactatgatgaaaacagtaaccaggcctttgcttcagcaaccaatg
atattgacaaaaactctcatgaccgcgttgactatctcaatgacggagatcattcagaaaatcgtcgttggacaaactggt
caccaacaccatcttctaatccagaagtatcagcgggtgtgatcttccgtgaaaatggtaagattgtagaacggactgttg

ctcaagccaaacttcacttctttgcagatagtggtacggatgcaccatctaaactcgttttagaacgctatgtcggcccagg
ctttgaagtacctacctactattcaaactaccaagcctacgaatctggacatccatttaacaatccagaaaattgggaagct
gtgccttatcgtgcggataaagacatcgcagctggtgatgaaatcaacgtaacatttaaagctgtcaaagccaaagtcat
gagatggcgtatggagcgtaaagctgacaagagcggtgttgcgatgattgagatgaccttccttgcaccaagtgaattgc
ctcaagaaagcactcaatcaaagattcttgtagatggaaaagaacttgctgatttcgctgaaaatcgtcaagactatcaaa
ttacctataaggtcaacggccaaaagtctcagttgaagaaaacaatcaagtagcttcaactgtggtagatagtggagaa
gatagccttccagtacttgttcgcctgtttcagaaagtggaaaacaagtcaaggaataccgtatccagttgactaaggaaa
aaccagtttctgctgtacaagaagatcttccaaaactcgaatttgttgaaaaagatttggcctacaagacagttgagaaaa
aagattcaacactgtatctaggtgaaactcgtgtagaacaagaaggaaaagtggaaaagaacgtatctttacagtgatta
atcctgatggaagtaaggaagaaaaactccgtgaagtggtagaagttccgacagaccgcatcgtcttggttggaaccaa
accagtagctcaagaagctaaaaaaccacaagtgtcagaaaaagcagatacaaaaccaattgattcaagtgaagctg
atcaaactaataaagcccagttaccaaatacaggtagtgcggcaagccaagcagcagtagcagcaggtttagctcttct
aggtttgagtgcaggattagtagttactaaaggtaaaaaagaagactag

BgaA amino acid sequence (2228 aa):

MGKGHWNRKRVYSIRKFAVGACSVMIGTCAVLLGGNIAGESVVYADETLITHTAEKPKEEKMIVEEKADKA
LETKNVVERTEQSEPSSTEAIASEKKEDEAVTPKEEKVSAKPEEKAPRIESQASSQEKPLKEDAKAVTNEEVNQ
MIENRKVDFNQNWYFKLNANSKEAIKPDADVSTWKKLDLPYDWSIFNDFDHESPAQNEGGQLNGGEAW
YRKTFKLDEKDLKKNVRLTFDGVYMDSQVYVNGQLVGHYPNGYNQFSYDITKYLYKDGRENVIAVHAVNK
QPSSRWYSGSGIYRDVTLQVTDKVHVEKNGTTILTPKLEEQQHGKVETHVTSKIVNTDDKDHELVAEYQIVE
RGGHAVTGLVRTASRTLKAHESTSLDAILEVERPKLWTVLNDKPALYELITRVYRDGQLVDAKKDLFGYRYYH
WTPNEGFSLNGERIKFHGVSLHHDHGALGAEENYKAEYRRLKQMKEMGVNSIRTTHNPASEQTLQIAAEL
GLLVQEEAFDTWYGGKKPYDYGRFFEKDATHPEARKGEKWSDFDLRTMVERGKNNPAIFMWSIGNEIGEA
NGDAHSLATVKRLVKVIKDVDKTRYVTMGADKFRGNGSGGHEKIADELDAVGFNYSEDNYKALRAKHPKW
LIYGSETSSATRTRGSYYRPERELKHSNGPERNYEQSDYGNDRVGWGKTATASWTFDRDNAGYAGQFIWT
GTDYIGEPTPWHNQNQTPVKSSYFGIVDTAGIPKHDFYLYQSQVSVKKKPMVHLLPHWNWENKELASKVA
DSEGKIPVRAYSNASSVELFLNGKSLGLKTFNKKQTSDGRTYQEGANANELYLEWKVAYQPGTLEAIARDES
GKEIARDKITTAGKPAAVRLIKEDHAIAADGKDLTYIYYEIVDSQGNVVPTANNLVRFQLHGQGQLVGVDNG
EQASRERYKAQADGSWIRKAFNGKGVAIVKSTEQAGKFTLTAHSDLLKSNQVTVFTGKKEGQEKTVLGTEV
PKVQTIIGEAPEMPTTVPFVYSDGSRAERPVTWSLVDVSKPGIVTVKGMADGREVEARVEVALKSELPVVKR
IAPNTNLNSVDKSVSYVLTDGSVQEYEVDKWEIAEEDKAKLAIPGSRIQATGYLEGQPIHATLVVEEGNPAAP
VVPTVTVGGEAVTGLTSRQPMQYRTLSYGAQLPEVTASAENADVTVLQASAANGMRASIFIQPKDGGPLQ
TYAIQFLEEAPKIAHLSLQVEKADSLKEDQTVKLSVRAHYQDGTQAVLPADKVTFSTSGEGEVAIRKGMLELH
KPGAVTLNAEYEGAKGQVELTIQANTEKKIAQSIRPVNVVTDLHQEPSLPATVTVEYDKGFPKTHKVTWQAI
PKEKLDSYQIFEVLGKVEGIDLEARAKVSVEGIVSVEEVSVTTPIAEAPQLPESVRTYDSNGHVSSAKVAWDAI

RPEQYAKEGVFTVNGRLEGTQLTTKLHVRVSAQTEQGANISDQWTGSELPLAFASDSNPSDPVSNVNDKLI
SYNNQPANRWTNWNRSNPEASVGVLFGDSGILSKRSVDNLSVGFHEDHGVGAPKSYVIEYYVGKTVPTAP
KNPSFVGNEDHVFNDSANWKPVTNLKAPAQLKAGEMNHFSFDKVETYAIRIRMVKADNKRGTSITEVQIF
AKQVAAAKQGQTRIQVDGKDLANFNPDLTDYYLESVDGKVPAVTANVSNNGLATVVPSVREGEPVRVIAK
AENGDILGEYRLHFTKDKNLLSHKPVAAVKQARLLQVGQALELPTKVPVYFTGKDGYETKDLTVEWEEVPAE
NLTKAGQFTVRGRVLGSNLVAEVTVRVTDKLGETLSDNPNYDENSNQAFASATNDIDKNSHDRVDYLNDG
DHSENRRWTNWSPTPSSNPEVSAGVIFRENGKIVERTVAQAKLHFFADSGTDAPSKLVLERYVGPGFEVPT
YYSNYQAYESGHPFNNPENWEAVPYRADKDIAAGDEINVTFKAVKAKVMRWRMERKADKSGVAMIEMT
FLAPSELPQESTQSKILVDGKELADFAENRQDYQITYKGQRPKVSVEENNQVASTVVDSGEDSLPVLVRLVSE
SGKQVKEYRIQLTKEKPVSAVQEDLPKLEFVEKDLAYKTVEKKDSTLYLGETRVEQEGKVGKERIFTVINPDGS
KEEKLREVVEVPTDRIVLVGTKPVAQEAKKPQVSEKADTKPIDSSEADQTNKAQLPNTGSAASQAAVAAGLA
LLGLSAGLVVTKGKKED

Adegbola, R.A., Hill, P.C., Secka, O., Ikumapayi, U.N., Lahai, G., Greenwood, B.M., Corrah, T., 2006. Serotype and antimicrobial susceptibility patterns of isolates of *Streptococcus pneumoniae* causing invasive disease in The Gambia 1996-2003. *Tropical Medicine & International Health.* **11,** 1128-1135.

Alexander, J.E., Berry, A.M., Paton, J.C., Rubins, J.B., Andrew, P.W., Mitchell, T.J., 1998. Amino acid changes affecting the activity of pneumolysin alter the behaviour of pneumococci in pneumonia. *Microbial Pathogenesis.* **24,** 167-174.

Alexander, J.E., Lock, R.A., Peeters, C.C., Poolman, J.T., Andrew, P.W., Mitchell, T.J., Hansman, D., Paton, J.C., 1994. Immunization of mice with pneumolysin toxoid confers a significant degree of protection against at least nine serotypes of *Streptococcus pneumoniae*. *Infection and Immunity.* **62,** 5683-5688.

Alloing, G., Granadel, C., Morrison, D.A., Claverys, J.P., 1996. Competence pheromone, oligopeptide permease, and induction of competence in *Streptococcus pneumoniae*. *Molecular Microbiology.* **21,** 471-478.

Alloing, G., Trombe, M.C., Claverys, J.P., 1990. The ami locus of the gram-positive bacterium *Streptococcus pneumoniae* is similar to binding protein-dependent transport operons of gram-negative bacteria. *Molecular Microbiology.* **4,** 633-644.

Akerley, B.J., Rubin, E.J., Camilli, A., Lampe, D.J., Robertson, H.M., Mekalanos, J.J., 1998. Systematic identification of essential genes by in vitro mariner mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America.* **95,** 8927-8932.

Alouf, J.E., 2000. Cholesterol-binding cytolytic protein toxins. *International Journal of Medical Microbiology.* **290,** 351-356.

Austrian, R., 1981. Some observations on the pneumococcus and on the current status of pneumococcal disease and its prevention. *Reviews of Infectious Diseases.* **3 Suppl,** S1-17.

Baba, H., Kawamura, I., Kohda, C., Nomura, T., Ito, Y., Kimoto, T., Watanabe, I., Ichiyama, S., Mitsuyama, M., 2002. Induction of gamma interferon and nitric oxide by truncated pneumolysin that lacks pore-forming activity. *Infection and Immunity*. **70,** 107-113.

Balachandran, P., Brooks-Walter, A., Virolainen-Julkunen, A., Hollingshead, S.K., Briles, D.E., 2002. Role of pneumococcal surface protein C in nasopharyngeal carriage and pneumonia and its ability to elicit protection against carriage of *Streptococcus pneumoniae*. *Infection and Immunity*. **70,** 2526-2534.

Beall, B., McEllistrem, M.C., Gertz, R.E.,Jr, Wedel, S., Boxrud, D.J., Gonzalez, A.L., Medina, M.J., Pai, R., Thompson, T.A., Harrison, L.H., McGee, L., Whitney, C.G., Active Bacterial Core Surveillance Team, 2006. Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. *Journal of Clinical Microbiology*. **44,** 999-1017.

Benton, K.A., Everson, M.P., Briles, D.E., 1995. A pneumolysin-negative mutant of *Streptococcus pneumoniae* causes chronic bacteremia rather than acute sepsis in mice. *Infection and Immunity*. **63,** 448-455.

Berg, J.M., Tymoczko, J.L. and Stryer, L., 2001. *Biochemistry*. 5th Edition ed. New York: W.H. Freeman and Company.

Bergmann, S. & Hammerschmidt, S., 2006. Versatility of pneumococcal surface proteins. *Microbiology*. **152,** 295-303.

Bergmann, S., Rohde, M., Chhatwal, G.S., Hammerschmidt, S., 2001. alpha-Enolase of *Streptococcus pneumoniae* is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Molecular Microbiology*. **40,** 1273-1287.

Berkley, J.A., Lowe, B.S., Mwangi, I., Williams, T., Bauni, E., Mwarumba, S., Ngetsa, C., Slack, M.P., Njenga, S., Hart, C.A., Maitland, K., English, M., Marsh, K., Scott, J.A., 2005. Bacteremia among children admitted to a rural hospital in Kenya. *The New England Journal of Medicine*. **352,** 39-47.

Berry, A.M., Lock, R.A., Hansman, D., Paton, J.C., 1989. Contribution of autolysin to virulence of *Streptococcus pneumoniae*. *Infection and Immunity*. **57,** 2324-2330.

Berry, A.M., Lock, R.A., Paton, J.C., 1996. Cloning and characterization of nanB, a second *Streptococcus pneumoniae* neuraminidase gene, and purification of the NanB enzyme from recombinant Escherichia coli. *Journal of Bacteriology*. **178,** 4854-4860.

Berry, A.M., Lock, R.A., Thomas, S.M., Rajan, D.P., Hansman, D., Paton, J.C., 1994. Cloning and nucleotide sequence of the *Streptococcus pneumoniae* hyaluronidase gene and purification of the enzyme from recombinant Escherichia coli. *Infection and Immunity*. **62,** 1101-1108.

Berry, A.M., Ogunniyi, A.D., Miller, D.C., Paton, J.C., 1999. Comparative virulence of *Streptococcus pneumoniae* strains with insertion-duplication, point, and deletion mutations in the pneumolysin gene. *Infection and Immunity*. **67,** 981-985.

Berry, A.M. & Paton, J.C., 2000. Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infection and Immunity*. **68,** 133-140.

Berry, A.M. & Paton, J.C., 1996. Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. *Infection and Immunity*. **64,** 5255-5262.

Berry, A.M., Paton, J.C., Hansman, D., 1992. Effect of insertional inactivation of the genes encoding pneumolysin and autolysin on the virulence of *Streptococcus pneumoniae* type 3. *Microbial Pathogenesis*. **12,** 87-93.

Berry, A.M., Yother, J., Briles, D.E., Hansman, D., Paton, J.C., 1989. Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. *Infection and Immunity*. **57,** 2037-2042.

Beveridge, T.J. & Graham, L.L., 1991. Surface layers of bacteria. *Microbiological Reviews*. **55,** 684-705.

Bhavanandan, V.P., Umemoto, J., Davidson, E.A., 1976. Characterization of an endo-alpha-N-acetyl galactosaminidase from *Diplococcus pneumoniae. Biochemical and Biophysical Research Communications.* **70,** 738-745.

Bissett, D.L. & Anderson, R.L., 1974a. Genetic evidence for the physiological significance of the D-tagatose 6-phosphate pathway of lactose and D-galactose degradation in S*taphylococcus aureus. Journal of Bacteriology.* **119,** 698-704.

Bissett, D.L. & Anderson, R.L., 1974b. Lactose and D-galactose metabolism in group N streptococci: presence of enzymes for both the D-galactose 1-phosphate and D-tagatose 6-phosphate pathways. *Journal of Bacteriology.* **117,** 318-320.

Bissett, D.L. & Anderson, R.L., 1973. Lactose and D-galactose metabolism in *Staphylococcus aureus*: pathway of D-galactose 6-phosphate degradation. *Biochemical and Biophysical Research Communications.* **52,** 641-647.

Black, R.E., Morris, S.S., Bryce, J., 2003. Where and why are 10 million children dying every year? *Lancet.* **361,** 2226-2234.

Bogaert, D., de Groot, R. and Hermans, P.W.M., 2007. *Streptococcus pneumoniae:* From colonization and Infection towards prevention strategies. In Hakenbeck, R. & Chhatwal, S., eds, *Molecular Biology of Streptococci.* First Edition ed. Great Britain: Horizon Bioscience. 252-260.

Bogaert, D., De Groot, R., Hermans, P.W., 2004. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *The Lancet Infectious Diseases.* **4,** 144-154.

Bogaert, D., Engelen, M.N., Timmers-Reker, A.J., Elzenaar, K.P., Peerbooms, P.G., Coutinho, R.A., de Groot, R., Hermans, P.W., 2001. Pneumococcal carriage in children in The Netherlands: a molecular epidemiological study. *Journal of Clinical Microbiology.* **39,** 3316-3320.

Boulnois, G.J., 1992. Pneumococcal proteins and the pathogenesis of disease caused by *Streptococcus pneumoniae*. *Journal of General Microbiology.* **138,** 249-259.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry.* **72,** 248-254.

Bricker, A.L. & Camilli, A., 1999. Transformation of a type 4 encapsulated strain of *Streptococcus pneumoniae*. *FEMS Microbiology Letters.* **172,** 131-135

Briese, T. & Hakenbeck, R., 1985. Interaction of the pneumococcal amidase with lipoteichoic acid and choline. *European Journal of Biochemistry.* **146,** 417-427.

Briles, D.E., Ades, E., Paton, J.C., Sampson, J.S., Carlone, G.M., Huebner, R.C., Virolainen, A., Swiatlo, E., Hollingshead, S.K., 2000. Intranasal immunization of mice with a mixture of the pneumococcal proteins PsaA and PspA is highly protective against nasopharyngeal carriage of *Streptococcus pneumoniae*. *Infection and Immunity.* **68,** 796-800.

Briles, D.E., Hollingshead, S.K., Swiatlo, E., Brooks-Walter, A., Szalai, A., Virolainen, A., McDaniel, L.S., Benton, K.A., White, P., Prellner, K., Hermansson, A., Aerts, P.C., Van Dijk, H., Crain, M.J., 1997. PspA and PspC: their potential for use as pneumococcal vaccines. *Microbial Drug Resistance.* **3,** 401-408.

Brockhausen, I., 2000. O-linked chain glycosyltransferases. *Methods in Molecular Biology.* **125,** 273-293.

Brueggemann, A.B., Peto, T.E., Crook, D.W., Butler, J.C., Kristinsson, K.G., Spratt, B.G., 2004. Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. *The Journal of Infectious Diseases.* **190,** 1203-1211.

Bruyn, G.A., Zegers, B.J., van Furth, R., 1992. Mechanisms of host defense against infection with *Streptococcus pneumoniae*. *Clinical Infectious Diseases.* **14,** 251-262.

Buchanan, R.E., Gibbons, N.E., Cowan, S.T., Holt, J.G., Liston, J., Murray, R.G.E., Niven, C.F., Ravin, A.W. and Stainer, R.Y., eds, 1975. *Bergey's Manual of Determinative Bacteriology.* 8th edition ed. Baltimore: The Williams & Wilkins Company.

Burnaugh, A.M., Frantz, L.J., King, S.J., 2008. Growth of *Streptococcus pneumoniae* on human glycoconjugates is dependent upon the sequential activity of bacterial exoglycosidases. *Journal of Bacteriology.* **190,** 221-230.

Busse, W.W., 1991. Pathogenesis and sequelae of respiratory infections. *Reviews of Infectious Diseases.* **13 Suppl 6,** S477-85.

Byers, H.L., Tarelli, E., Homer, K.A., Beighton, D., 1999. Sequential deglycosylation and utilization of the N-linked, complex-type glycans of human alpha1-acid glycoprotein mediates growth of *Streptococcus oralis*. *Glycobiology.* **9,** 469-479.

Byington, C.L., Korgenski, K., Daly, J., Ampofo, K., Pavia, A., Mason, E.O., 2006. Impact of the pneumococcal conjugate vaccine on pneumococcal parapneumonic empyema. *The Pediatric Infectious Disease Journal.* **25,** 250-254.

Camara, M., Boulnois, G.J., Andrew, P.W., Mitchell, T.J., 1994. A neuraminidase from *Streptococcus pneumoniae* has the features of a surface protein. *Infection and Immunity.* **62,** 3688-3695.

Canvin, J.R., Marvin, A.P., Sivakumaran, M., Paton, J.C., Boulnois, G.J., Andrew, P.W., Mitchell, T.J., 1995. The role of pneumolysin and autolysin in the pathology of pneumonia and septicemia in mice infected with a type 2 pneumococcus. *The Journal of Infectious Diseases.* **172,** 119-123.

Cheng, Q., Finkel, D., Hostetter, M.K., 2000. Novel purification scheme and functions for a C3-binding protein from *Streptococcus pneumoniae*. *Biochemistry.* **39,** 5450-5457.

Chiavolini, D., Memmi, G., Maggi, T., Iannelli, F., Pozzi, G., Oggioni, M.R., 2003. The three extra-cellular zinc metalloproteinases of *Streptococcus pneumoniae* have a different impact on virulence in mice. *BMC Microbiology.* **3,** 14.

Chhatwal, G.S., 2002. Anchorless adhesins and invasins of Gram-positive bacteria: a new class of virulence factors. *Trends in Microbiology*. **10,** 205-208.

Clarke, V.A., Platt, N., Butters, T.D., 1995. Cloning and expression of the beta-N-acetylglucosaminidase gene from *Streptococcus pneumoniae*. Generation of truncated enzymes with modified aglycon specificity. *The Journal of Biological Chemistry*. **270,** 8805-8814.

Cocaign-Bousquet, M., Even, S., Lindley, N.D., Loubiere, P., 2002. Anaerobic sugar catabolism in *Lactococcus lactis*: genetic regulation and enzyme control over pathway flux. *Applied Microbiology and Biotechnology*. **60,** 24-32.

Cocaign-Bousquet, M., Garrigues, C., Loubiere, P., Lindley, N.D., 1996. Physiology of pyruvate metabolism in *Lactococcus lactis*. *Antonie van Leeuwenhoek.* **70,** 253-267.

Coles, C.L., Kanungo, R., Rahmathullah, L., Thulasiraj, R.D., Katz, J., Santosham, M., Tielsch, J.M., 2001. Pneumococcal nasopharyngeal colonization in young South Indian infants. *The Pediatric Infectious Disease Journal*. **20,** 289-295.

Cone, R.A., 2009. Barrier properties of mucus. *Advanced Drug Delivery Reviews.* **61,** 75-85.

Contesse, G., Crepin, M., Gros, F., Ullmann, A. and Monod, J., 1970. On the Mechanism of Catabolite Repression. In Beckwith, J. R. & Zipser, D., eds, *The Lactose Operon.* first ed. USA: Cold Spring Harbor Laboratory. 401-414.

Corfield, A.P., Wagner, S.A., Clamp, J.R., Kriaris, M.S., Hoskins, L.C., 1992. Mucin degradation in the human colon: production of sialidase, sialate O-acetylesterase, N-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. *Infection and Immunity*. **60,** 3971-3978.

Crain, M.J., Waltman, W.D.,2nd, Turner, J.S., Yother, J., Talkington, D.F., McDaniel, L.S., Gray, B.M., Briles, D.E., 1990. Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*. *Infection and Immunity.* **58,** 3293-3299.

Cundell, D.R., Gerard, N.P., Gerard, C., Idanpaan-Heikkila, I., Tuomanen, E.I., 1995. *Streptococcus pneumoniae* anchor to activated human cells by the receptor for platelet-activating factor. *Nature.* **377,** 435-438.

Daly, M.M., Doktor, S., Flamm, R., Shortridge, D., 2004. Characterization and prevalence of MefA, MefE, and the associated msr(D) gene in *Streptococcus pneumoniae* clinical isolates. *Journal of Clinical Microbiology.* **42,** 3570-3574.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M.K., McDaniel, L.S., 2004. Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H. *Journal of Immunology.* **173,** 471-477.

Denny, F.W. & Loda, F.A., 1986. Acute respiratory infections are the leading cause of death in children in developing countries. *The American Journal of Tropical Medicine and Hygiene.* **35,** 1-2.

Deutscher, J., 2008. The mechanisms of carbon catabolite repression in bacteria. *Current Opinion in Microbiology.* **11,** 87-93.

Deutscher, J., Francke, C., Postma, P.W., 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiology and Molecular Biology Reviews.* **70,** 939-1031.

Deutscher, J., Kuster, E., Bergstedt, U., Charrier, V., Hillen, W., 1995. Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in gram-positive bacteria. *Molecular Microbiology.* **15,** 1049-1053.

Di John, D., Wasserman, S.S., Torres, J.R., Cortesia, M.J., Murillo, J., Losonsky, G.A., Herrington, D.A., Sturcher, D., Levine, M.M., 1989. Effect of priming with carrier on response to conjugate vaccine. *Lancet.* **2,** 1415-1418.

Dintilhac, A., Alloing, G., Granadel, C., Claverys, J.P., 1997. Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Molecular Microbiology.* **25,** 727-739.

Dochez, A.R. & Avery, O.T., 1915. The occurrence of carriers of disease-producing types of pneumococcus. *The Journal of ExperimentalMedicine.* **22,** 105-113.

Duran-Reynals, F., 1933. Studies on a certain spreading factor existing in bacteria and its significance for bacterial invasivness. *Journal of Experimental Medicine.* **58,** 161-181.

Dziegielewska, K.M., Mollgard, K., Reynolds, M.L., Saunders, N.R., 1987. A fetuin-related glycoprotein (alpha 2HS) in human embryonic and fetal development. *Cell and Ttissue Research.* **248,** 33-41.

Egeter, O. & Bruckner, R., 1996. Catabolite repression mediated by the catabolite control protein CcpA in *Staphylococcus xylosus*. *Molecular Microbiology.* **21,** 739-749.

Enright, M.C. & Spratt, B.G., 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology.* **144,** 3049-3060.

Fath, M.J. & Kolter, R., 1993. ABC transporters: bacterial exporters. *Microbiological Reviews.* **57,** 995-1017.

Feikin, D.R., Davis, M., Nwanyanwu, O.C., Kazembe, P.N., Barat, L.M., Wasas, A., Bloland, P.B., Ziba, C., Capper, T., Huebner, R.E., Schwartz, B., Klugman, K.P., Dowell, S.F., 2003. Antibiotic resistance and serotype distribution of *Streptococcus pneumoniae* colonizing rural Malawian children. *The Pediatric Infectious Disease Journal.* **22,** 564-567.

Felmingham, D., 2004. Comparative antimicrobial susceptibility of respiratory tract pathogens. *Chemotherapy.* **50 Suppl 1,** 3-10.

Felmingham, D., Farrell, D.J., Reinert, R.R., Morrissey, I., 2004. Antibacterial resistance among children with community-acquired respiratory tract infections (PROTEKT 1999-2000). *The Journal of Infection.* **48,** 39-55.

Felmingham, D., Reinert, R.R., Hirakata, Y., Rodloff, A., 2002. Increasing prevalence of antimicrobial resistance among isolates of *Streptococcus pneumoniae* from the PROTEKT surveillance study, and compatative in vitro activity of the ketolide, telithromycin. *The Journal of Antimicrobial Chemotherapy.* **50 Suppl S1,** 25-37.

Fersht, A., 1999. *Structure and Mechanism in Protein Science.* 4th Edition ed. W.H. Freeman and Company.

Fowler, J., Vinall, L., Swallow, D., 2001. Polymorphism of the human muc genes. *Frontiers in Bioscience : a Journal and virtual library.* **6,** D1207-15.

Foy, H.M., Wentworth, B., Kenny, G.E., Kloeck, J.M., Grayston, J.T., 1975. Pneumococcal isolations from patients with pneumonia and control subjects in a prepaid medical care group. *The American Review of Respiratory Disease.* **111,** 595-603.

Freijd, A., Hammarstrom, L., Persson, M.A., Smith, C.I., 1984. Plasma anti-pneumococcal antibody activity of the IgG class and subclasses in otitis prone children. *Clinical and Experimental Immunology.* **56,** 233-238.

Frey, P.A., 1996. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *The FASEB Journal: official publication of the Federation of American Societies for Experimental Biology.* **10,** 461-470.

Fujita, K., Oura, F., Nagamine, N., Katayama, T., Hiratake, J., Sakata, K., Kumagai, H., Yamamoto, K., 2005. Identification and molecular cloning of a novel glycoside hydrolase

family of core 1 type O-glycan-specific endo-alpha-N-acetylgalactosaminidase from *Bifidobacterium longum*. *The Journal of Biological Chemistry.* **280,** 37415-37422.

Fujita, Y., 2009. Carbon catabolite control of the metabolic network in *Bacillus subtilis*. *Bioscience, Biotechnology, and Biochemistry.* **73,** 245-259.

Garcia, P., Garcia, J.L., Garcia, E., Lopez, R., 1986. Nucleotide sequence and expression of the pneumococcal autolysin gene from its own promoter in *Escherichia coli*. *Gene.* **43,** 265-272.

Garcia, P., Garcia, J.L., Garcia, E., Sanchez-Puelles, J.M., Lopez, R., 1990. Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene.* **86,** 81-88.

Garrigues, C., Loubiere, P., Lindley, N.D., Cocaign-Bousquet, M., 1997. Control of the shift from homolactic acid to mixed-acid fermentation in *Lactococcus lactis*: predominant role of the NADH/NAD+ ratio. *Journal of Bacteriology.* **179,** 5282-5287.

Garrigues, C., Mercade, M., Cocaign-Bousquet, M., Lindley, N.D., Loubiere, P., 2001. Regulation of pyruvate metabolism in *Lactococcus lactis* depends on the imbalance between catabolism and anabolism. *Biotechnology and Bioengineering.* **74,** 108-115.

Giammarinaro, P. & Paton, J.C., 2002. Role of RegM, a homologue of the catabolite repressor protein CcpA, in the virulence of *Streptococcus pneumoniae*. *Infection and Immunity.* **70,** 5454-5461.

Gilbert, R.J., Jimenez, J.L., Chen, S., Andrew, P.W., Saibil, H.R., 2000. Structural basis of pore formation by cholesterol-binding toxins. *International Journal of Medical Microbiolog.* **290,** 389-394.

Gilson, E., Alloing, G., Schmidt, T., Claverys, J.P., Dudler, R., Hofnung, M., 1988. Evidence for high affinity binding-protein dependent transport systems in gram-positive bacteria and in *Mycoplasma*. *The EMBO Journal.* **7,** 3971-3974.

Gingles, N.A., Alexander, J.E., Kadioglu, A., Andrew, P.W., Kerr, A., Mitchell, T.J., Hopes, E., Denny, P., Brown, S., Jones, H.B., Little, S., Booth, G.C., McPheat, W.L., 2001. Role of genetic resistance in invasive pneumococcal infection: identification and study of susceptibility and resistance in inbred mouse strains. *Infection and Immunity.* **69,** 426-434.

Gonzalez, B.E., Hulten, K.G., Lamberth, L., Kaplan, S.L., Mason, E.O.,Jr, U.S. Pediatric Multicenter Pneumococcal Surveillance Group, 2006. *Streptococcus pneumoniae* serogroups 15 and 33: an increasing cause of pneumococcal infections in children in the United States after the introduction of the pneumococcal 7-valent conjugate vaccine. *The Pediatric Infectious Disease Journal.* **25,** 301-305.

Gorke, B. & Stulke, J., 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews .Microbiology.* **6,** 613-624.

Gratten, M., Montgomery, J., Gerega, G., Gratten, H., Siwi, H., Poli, A., Koki, G., 1989. Multiple colonization of the upper respiratory tract of Papua New Guinea children with *Haemophilus influenzae* and *Streptococcus pneumoniae*. *The Southeast Asian Journal of Tropical Medicine and Public Health.* **20,** 501-509.

Gray, B.M., Converse, G.M.,3rd, Dillon, H.C.,Jr, 1980. Epidemiologic studies of *Streptococcus pneumoniae* in infants: acquisition, carriage, and infection during the first 24 months of life. *The Journal of Infectious Diseases.* **142,** 923-933.

Green, E.D., Adelt, G., Baenziger, J.U., Wilson, S., Van Halbeek, H., 1988a. The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy. *The Journal of Biological Chemistry.* **263,** 18253-18268.

Guiral, S., Henard, V., Laaberki, M.H., Granadel, C., Prudhomme, M., Martin, B., Claverys, J.P., 2006. Construction and evaluation of a chromosomal expression platform (CEP) for ectopic, maltose-driven gene expression in *Streptococcus pneumoniae*. *Microbiology.* **152,** 343-349.

Hakenbeck, R. and Chhatwal, S., eds, 2007. *Molecular Biology of Streptococci.* 1st Edition ed. Great Britain: Horizon Bioscience.

Hammerschmidt, S., 2007. Pneumococcal Virulence Factors. In Hakenbeck, R. & Chhatwal, S., eds, *Molecular Biology of Streptococci.* First Edition ed. Great Britain: Horizon Bioscience. 178-179.

Hammerschmidt, S., Talay, S.R., Brandtzaeg, P., Chhatwal, G.S., 1997. SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Molecular Microbiology.* **25,** 1113-1124.

Hansman, D., 1972. Type distribution and antibiotic sensitivity of pneumococci from carriers in Kiriwina, Trobriand islands (New Guinea). *The Medical Journal of Australia.* **2,** 771-773.

Hansman, D., Morris, S., Gregory, M., McDonald, B., 1985. Pneumococcal carriage amongst Australian aborigines in Alice Springs, Northern Territory. *The Journal of Hygiene.* **95,** 677-684.

Hausdorff, W.P., Brueggemann, A.B., Hackell, J.G. and Scott, A.G., 2008. Pneumococcal Serotype Epidemiology. In Siber, G. R., Klugman, K. P. and Makela, P. H., eds, *Pneumococcal Vaccines.* First Edition ed. United States of America: ASM Press. 140-150.

Henrichsen, J., 1995. Six newly recognized types of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology.* **33,** 2759-2762.

Henriques-Normark, B., 2007. Molecular EPidemiology and Mechanisms for Antibiotic Resistance in Streptococcus pneumoniae. In Hakenbeck, R. & Chhatwal, S., eds, *Molecular Biology of Streptococci.* First Edition ed. Great Britain: Horizon Bioscience. 269-284.

Henrissat, B., 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical Journal.* **280,** 309-316.

Higgins, C.F., 1992. ABC transporters: from microorganisms to man. *Annual Review of Cell Biology.* **8,** 67-113.

Hirst, R.A., Yesilkaya, H., Clitheroe, E., Rutman, A., Dufty, N., Mitchell, T.J., O'Callaghan, C., Andrew, P.W., 2002. Sensitivities of human monocytes and epithelial cells to pneumolysin are different. *Infection and Immunity.* **70,** 1017-1022.

Ho, P.L., Que, T.L., Tsang, D.N., Ng, T.K., Chow, K.H., Seto, W.H., 1999. Emergence of fluoroquinolone resistance among multiply resistant strains of *Streptococcus pneumoniae* in Hong Kong. *Antimicrobial Agents and Chemotherapy.* **43,** 1310-1313.

Ho, P.L., Yung, R.W., Tsang, D.N., Que, T.L., Ho, M., Seto, W.H., Ng, T.K., Yam, W.C., Ng, W.W., 2001. Increasing resistance of *Streptococcus pneumoniae* to fluoroquinolones: results of a Hong Kong multicentre study in 2000. *The Journal of Antimicrobial Chemotherapy.* **48,** 659-665.

Homer, K.A., Whiley, R.A., Beighton, D., 1994. Production of specific glycosidase activities by *Streptococcus intermedius* strain UNS35 grown in the presence of mucin. *Journal of Medical Microbiology.* **41,** 184-190.

Hoskins, J., Alborn, W.E.,Jr, Arnold, J., Blaszczak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R.,Jr, Skatrud, P.L., Glass, J.I., 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. *Journal of Bacteriology.* **183,** 5709-5717.

Hounsell, E.F., 1994. Physicochemical analyses of oligosaccharide determinants of glycoproteins. *Advances in Carbohydrate Chemistry and Biochemistry.* **50,** 311-350.

Hovenberg, H.W., Davies, J.R., Carlstedt, I., 1996. Different mucins are produced by the surface epithelium and the submucosa in human trachea: identification of MUC5AC as a major mucin from the goblet cells. *The Biochemical Journal.* **318,** 319-324.

Humphrey, J.H., 1948. Hyaluronidase production by pneumococci. **55,** 273-275.

Hussain, M., Melegaro, A., Pebody, R.G., George, R., Edmunds, W.J., Talukdar, R., Martin, S.A., Efstratiou, A., Miller, E., 2005. A longitudinal household study of *Streptococcus pneumoniae* nasopharyngeal carriage in a UK setting. *Epidemiology and Infection.* **133,** 891-898.

Hynes, W.L. & Ferretti, J.J., 1989. Sequence analysis and expression in *Escherichia coli* of the hyaluronidase gene of *Streptococcus pyogenes* bacteriophage H4489A. *Infection and Immunity*. **57,** 533-539.

Iannelli, F., Chiavolini, D., Ricci, S., Oggioni, M.R., Pozzi, G., 2004. Pneumococcal surface protein C contributes to sepsis caused by *Streptococcus pneumoniae* in mice. *Infection and Immunity*. **72,** 3077-3080.

Iyer, R., Baliga, N.S., Camilli, A., 2005. Catabolite control protein A (CcpA) contributes to virulence and regulation of sugar metabolism in *Streptococcus pneumoniae*. *Journal of Bacteriology*. **187,** 8340-8349.

Jacobs, M.R., 2004. *Streptococcus pneumoniae*: epidemiology and patterns of resistance. *The American Journal of Medicine.* **117 Suppl 3A,** 3S-15S.

Jagota S., Ramana M. V., Dutta S. M., 1981. Beta-Galactosidase of *Streptococcus cremoris*. *Journal of Food Science.* **46,** 161-168.

Janulczyk, R., Iannelli, F., Sjoholm, A.G., Pozzi, G., Bjorck, L., 2000. Hic, a novel surface protein of *Streptococcus pneumoniae* that interferes with complement function. *The Journal of Biological Chemistry*. **275,** 37257-37263.

Jedrzejas, M.J., 2004. Extracellular virulence factors of *Streptococcus pneumoniae*. *Frontiers in Bioscience.* **9,** 891-914.

Jedrzejas, M.J., 2001. Pneumococcal virulence factors: structure and function. *Microbiology and Molecular Biology Reviews.* **65,** 187-207.

Jedrzejas, M.J., Lamani, E., Becker, R.S., 2001. Characterization of selected strains of pneumococcal surface protein A. *The Journal of Biological Chemistry.* **276,** 33121-33128.

Jedrzejas, M.J., Mewbourne, R.B., Chantalat, L., McPherson, D.T., 1998. Expression and purification of *Streptococcus pneumoniae* hyaluronate lyase from *Escherichia coli*. *Protein Expression and Purification.* **13,** 83-89.

Jennings, H.J., 1990. Capsular polysaccharides as vaccine candidates. *Current Topics in Microbiology and Immunology.* **150,** 97-127.

Jeong, J.K., Kwon, O., Lee, Y.M., Oh, D.B., Lee, J.M., Kim, S., Kim, E.H., Le, T.N., Rhee, D.K., Kang, H.A., 2009. Characterization of the *Streptococcus pneumoniae* BgaC protein as a novel surface beta-galactosidase with specific hydrolysis activity for the Galbeta1-3GlcNAc moiety of oligosaccharides. *Journal of Bacteriology.* **191,** 3011-3023.

Johnston, R.B.,Jr, 1991. Pathogenesis of pneumococcal pneumonia. *Reviews of Infectious Diseases.* **13 Suppl 6,** S509-17.

Jones, B.E., Dossonnet, V., Kuster, E., Hillen, W., Deutscher, J., Klevit, R.E., 1997. Binding of the catabolite repressor protein CcpA to its DNA target is regulated by phosphorylation of its corepressor HPr. *The Journal of Biological Chemistry.* **272,** 26530-26535.

Jounblat, R., Kadioglu, A., Mitchell, T.J., Andrew, P.W., 2003. Pneumococcal behavior and host responses during bronchopneumonia are affected differently by the cytolytic and complement-activating activities of pneumolysin. *Infection and Immunity.* **71,** 1813-1819.

Kadioglu, A., Coward, W., Colston, M.J., Hewitt, C.R., Andrew, P.W., 2004. CD4-T-lymphocyte interactions with pneumolysin and pneumococci suggest a crucial protective role in the host response to pneumococcal infection. *Infection and Immunity.* **72,** 2689-2697.

Kadioglu, A., Taylor, S., Iannelli, F., Pozzi, G., Mitchell, T.J., Andrew, P.W., 2002. Upper and lower respiratory tract infection by *Streptococcus pneumoniae* is affected by pneumolysin deficiency and differences in capsule type. *Infection and Immunity.* **70,** 2886-2890.

Kadioglu, A., Weiser, J.N., Paton, J.C., Andrew, P.W., 2008. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nature Reviews. Microbiology.* **6,** 288-301.

Kaplan, S.L., Mason, E.O.,Jr, Wald, E.R., Schutze, G.E., Bradley, J.S., Tan, T.Q., Hoffman, J.A., Givner, L.B., Yogev, R., Barson, W.J., 2004. Decrease of invasive pneumococcal infections in children among 8 children's hospitals in the United States after the introduction of the 7-valent pneumococcal conjugate vaccine. *Pediatrics.* **113,** 443-449.

Kaufman, G.E. & Yother, J., 2007. CcpA-dependent and -independent control of beta-galactosidase expression in *Streptococcus pneumoniae* occurs via regulation of an upstream phosphotransferase system-encoding operon. *Journal of Bacteriology.* **189,** 5183-5192.

Kelly, S.J., Taylor, K.B., Li, S., Jedrzejas, M.J., 2001. Kinetic properties of *Streptococcus pneumoniae* hyaluronate lyase. *Glycobiology.* **11,** 297-304.

Kilic, A.O., Tao, L., Zhang, Y., Lei, Y., Khammanivong, A. and Herzberg, M.C., 2004. Involvement of Streptococcus gordonii beta-glucoside metabolism systems in adhesion, biofilm formation, and in vivo gene expression. *Journal of Bacteriology.,* **186,** 4246-4253.

King, S.J., 2010. Pneumococcal modification of host sugars: a major contributor to colonization of the human airway? *Molecular Oral Microbiology.* **25,** 15-24.

King, S.J., Hippe, K.R., Weiser, J.N., 2006. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by *Streptococcus pneumoniae*. *Molecular Microbiology.* **59,** 961-974.

King, S.J., Whatmore, A.M., Dowson, C.G., 2005. NanA, a neuraminidase from *Streptococcus pneumoniae*, shows high levels of sequence diversity, at least in part through recombination with Streptococcus oralis. *Journal of Bacteriology.* **187,** 5376-5386.

King, S.J., Hippe, K.R., Gould, J.M., Bae, D., Peterson, S., Cline, R.T., Fasching, C., Janoff, E.N., Weiser, J.N., 2004. Phase variable desialylation of host proteins that bind to *Streptococcus pneumoniae* in vivo and protect the airway. *Molecular Microbiology.* **54,** 159-171.

Kinoshita, M., Nakamura, T., Ihara, M., Haraguchi, T., Hiraoka, Y., Tashiro, K., Noda, M., 2001. Identification of human endomucin-1 and -2 as membrane-bound O-sialoglycoproteins with anti-adhesive activity. *FEBS Letters.* **499,** 121-126.

Klein, J.O., 2000. The pneumococcal conjugate vaccine arrives: a big win for kids. *The Pediatric Infectious Disease Journal.* **19,** 181-182.

Knowles, M.R. & Boucher, R.C., 2002. Mucus clearance as a primary innate defense mechanism for mammalian airways. *The Journal of Clinical Investigation.* **109,** 571-577.

Krivan, H.C., Roberts, D.D., Ginsburg, V., 1988. Many pulmonary pathogenic bacteria bind specifically to the carbohydrate sequence GalNAc beta 1-4Gal found in some glycolipids. *Proceedings of the National Academy of Sciences of the United States of America.* **85,** 6157-6161.

Lapensee, L., Paquette, Y., Bleau, G., 1997. Allelic polymorphism and chromosomal localization of the human oviductin gene (MUC9). *Fertility and Sterility.* **68,** 702-708.

Lampe, D.J., Churchill, M.E., Robertson, H.M., 1996. A purified mariner transposase is sufficient to mediate transposition in vitro. *The EMBO Journal.* **15,** 5470-5479.

Lee, C.J., Banks, S.D., Li, J.P., 1991. Virulence, immunity, and vaccine related to *Streptococcus pneumoniae*. *Critical Reviews in Microbiology.* **18,** 89-114.

Lees, A., Puvanesarajah, V. and Frasch, C.E., 2008. Conjugation Chemistry. In Siber, G., Klugman, K. and Makela, P. H., eds, *Pneumococcal Vaccines.* first ed. USA: ASM Press. 163-172.

Lipsitch, M., Dykes, J.K., Johnson, S.E., Ades, E.W., King, J., Briles, D.E., Carlone, G.M., 2000. Competition among *Streptococcus pneumoniae* for intranasal colonization in a mouse model. *Vaccine.* **18,** 2895-2901.

Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* **25,** 402-408.

Lloyd-Evans, N., O'Dempsey, T.J., Baldeh, I., Secka, O., Demba, E., Todd, J.E., Mcardle, T.F., Banya, W.S., Greenwood, B.M., 1996. Nasopharyngeal carriage of pneumococci in Gambian children and in their families. *The Pediatric Infectious Disease Journal.* **15,** 866-871.

Lock, R.A., Hansman, D., Paton, J.C., 1992. Comparative efficacy of autolysin and pneumolysin as immunogens protecting mice against infection by *Streptococcus pneumoniae*. *Microbial Pathogenesis.* **12,** 137-143.

Loda, F.A., Collier, A.M., Glezen, W.P., Strangert, K., Clyde, W.A.,Jr, Denny, F.W., 1975. Occurrence of *Diplococcus pneumoniae* in the upper respiratory tract of children. *The Journal of Pediatrics.* **87,** 1087-1093.

Lopez, R., Garcia, J.L., Garcia, E., Ronda, C., Garcia, P., 1992. Structural analysis and biological significance of the cell wall lytic enzymes of *Streptococcus pneumoniae* and its bacteriophage. *FEMS Microbiology Letters.* **79,** 439-447.

MacLEOD, C.M. & KRAUS, M.R., 1950. Relation of virulence of pneumococcal strains for mice to the quantity of capsular polysaccharide formed in vitro. *The Journal of Experimental Medicine.* **92,** 1-9.

Madigan, M.T., Martinko, J.M. and Parker, J., 2003. *Brock Biology of Microorganisms.* 10th Edition ed. USA: Prentice Hall.

Mahr, K., Hillen, W., Titgemeyer, F., 2000. Carbon catabolite repression in *Lactobacillus pentosus*: analysis of the ccpA region. *Applied and Environmental Microbiology.* **66,** 277-283.

Malley, R., Henneke, P., Morse, S.C., Cieslewicz, M.J., Lipsitch, M., Thompson, C.M., Kurt-Jones, E., Paton, J.C., Wessels, M.R., Golenbock, D.T., 2003. Recognition of pneumolysin by Toll-like receptor 4 confers resistance to pneumococcal infection. *Proceedings of the National Academy of Sciences of the United States of America.* **100,** 1966-1971.

Manco, S., Hernon, F., Yesilkaya, H., Paton, J.C., Andrew, P.W., Kadioglu, A., 2006. Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infection and Immunity.* **74,** 4014-4020.

Marion, C., Limoli, D.H., Bobulsky, G.S., Abraham, J.L., Burnaugh, A.M., King, S.J., 2009. Identification of a pneumococcal glycosidase that modifies O-linked glycans. *Infection and Immunity.* **77,** 1389-1396.

Marra, A., Lawson, S., Asundi, J.S., Brigham, D., Hromockyj, A.E., 2002. In vivo characterization of the psa genes from *Streptococcus pneumoniae* in multiple models of infection. *Microbiology.* **148,** 1483-1491.

Masters, P.L., Brumfitt, W., Mendez, R.L., Likar, M., 1958. Bacterial flora of the upper respiratory tract in Paddington families. 1952-4. *British Medical Journal.* **1,** 1200-1205.

Masuda, H., 1984. The preparation of heparan sulfate from the mitral valve of the human heart. *The International Journal of Biochemistry.* **16,** 99-103.

Mbelle, N., Huebner, R.E., Wasas, A.D., Kimura, A., Chang, I., Klugman, K.P., 1999. Immunogenicity and impact on nasopharyngeal carriage of a nonavalent pneumococcal conjugate vaccine. *The Journal of Infectious Diseases.* **180,** 1171-1176.

McDaniel, L.S., Sheffield, J.S., Swiatlo, E., Yother, J., Crain, M.J., Briles, D.E., 1992. Molecular localization of variable and conserved regions of pspA and identification of additional pspA homologous sequences in *Streptococcus pneumoniae*. *Microbial Pathogenesis.* **13,** 261-269.

McKessar, S.J. & Hakenbeck, R., 2007. The two-component regulatory system TCS08 is involved in cellobiose metabolism of *Streptococcus pneumoniae* R6. *Journal of Bacteriology.* **189,** 1342-1350.

Melchiorsen, C.R., Jokumsen, K.V., Villadsen, J., Johnsen, M.G., Israelsen, H., Arnau, J., 2000. Synthesis and posttranslational regulation of pyruvate formate-lyase in *Lactococcus lactis*. *Journal of Bacteriology.* **182,** 4783-4788.

Miles, A.A. & Misra, S.S., 1938. The estimation of the bactericidal power of the blood. *Journal of Hygiene*. **38,** 732-749.

Miller, J.H., 1972. *Experiments in Molecular Genetics.* 1st ed. New York USA: Cold Spring Harbor Laboratory.

Mitchell, T.J., Alexander, J.E., Morgan, P.J., Andrew, P.W., 1997. Molecular analysis of virulence factors of *Streptococcus pneumoniae*. *Society for Applied Bacteriology symposium series.* **26,** 62S-71S.

Mitchell, T.J., Andrew, P.W., Saunders, F.K., Smith, A.N., Boulnois, G.J., 1991. Complement activation and antibody binding by pneumolysin via a region of the toxin homologous to a human acute-phase protein. *Molecular Microbiology.* **5,** 1883-1888.

Monedero, V., Kuipers, O.P., Jamet, E., Deutscher, J., 2001. Regulatory functions of serine-46-phosphorylated HPr in *Lactococcus lactis*. *Journal of Bacteriology.* **183,** 3391-3398.

Montgomery, J.M., Lehmann, D., Smith, T., Michael, A., Joseph, B., Lupiwa, T., Coakley, C., Spooner, V., Best, B., Riley, I.D., 1990. Bacterial colonization of the upper respiratory

tract and its association with acute lower respiratory tract infections in Highland children of Papua New Guinea. *Reviews of Infectious Diseases.* **12 Suppl 8,** S1006-16.

Morona, J.K., Morona, R., Paton, J.C., 2006. Attachment of capsular polysaccharide to the cell wall of *Streptococcus pneumoniae* type 2 is required for invasive disease. *Proceedings of the National Academy of Sciences of the United States of America.* **103,** 8505-8510.

Moscoso, M., Garcia, E., Lopez, R., 2009. Pneumococcal biofilms. *International Microbiology.* **12,** 77-85.

Mosser, J.L. & Tomasz, A., 1970. Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. *The Journal of Biological Chemistry.* **245,** 287-298.

Mulholland, K., 2007. Childhood pneumonia mortality - a permanent global emergency. *Lancet.* **370,** 285-289.

Mulholland, K., 1999. Magnitude of the problem of childhood pneumonia. *Lancet.* **354,** 590-592.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology.* **51 Pt 1,** 263-273.

Musher, D.M., 1992. Infections caused by *Streptococcus pneumoniae*: clinical spectrum, pathogenesis, immunity, and treatment. *Clinical Infectious Diseases.* **14,** 801-807.

Nabors, G.S., Braun, P.A., Herrmann, D.J., Heise, M.L., Pyle, D.J., Gravenstein, S., Schilling, M., Ferguson, L.M., Hollingshead, S.K., Briles, D.E., Becker, R.S., 2000. Immunization of healthy adults with a single recombinant pneumococcal surface protein A (PspA) variant stimulates broadly cross-reactive antibodies to heterologous PspA molecules. *Vaccine.* **18,** 1743-1754.

Navarre, W.W. & Schneewind, O., 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Review.* **63,** 174-229.

Neijssel, O.M., Snoep, J.L., Teixeira de Mattos, M.J., 1997. Regulation of energy source metabolism in streptococci. *Society for Applied Bacteriology symposium series.* **26,** 12S-19S.

Nelson, A.L., Roche, A.M., Gould, J.M., Chim, K., Ratner, A.J., Weiser, J.N., 2007. Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infection and Immunity.* **75,** 83-90.

Neves, A.R., Pool, W.A., Kok, J., Kuipers, O.P., Santos, H., 2005. Overview on sugar metabolism and its control in *Lactococcus lactis* - the input from in vivo NMR. *FEMS Microbiology Reviews.* **29,** 531-554.

Neves, A.R., Ventura, R., Mansour, N., Shearman, C., Gasson, M.J., Maycock, C., Ramos, A., Santos, H., 2002. Is the glycolytic flux in *Lactococcus lactis* primarily controlled by the redox charge? Kinetics of NAD(+) and NADH pools determined in vivo by 13C NMR. *The Journal of Biological Chemistry.* **277,** 28088-28098.

Nukui, M., Taylor, K.B., McPherson, D.T., Shigenaga, M.K., Jedrzejas, M.J., 2003. The function of hydrophobic residues in the catalytic cleft of *Streptococcus pneumoniae* hyaluronate lyase. Kinetic characterization of mutant enzyme forms. *The Journal of Biological Chemistry.* **278,** 3079-3088.

Orihuela, C.J., Gao, G., Francis, K.P., Yu, J., Tuomanen, E.I., 2004. Tissue-specific contributions of pneumococcal virulence factors to pathogenesis. *The Journal of Infectious Diseases.* **190,** 1661-1669.

Pallares, R., Fenoll, A., Linares, J., Spanish Pneumococcal Infection Study Network, 2003. The epidemiology of antibiotic resistance in *Streptococcus pneumoniae* and the clinical relevance of resistance to cephalosporins, macrolides and quinolones. *International Journal of Antimicrobial Agents.* **22 Suppl 1,** S15-24.

Pallesen, L.T., Berglund, L., Rasmussen, L.K., Petersen, T.E., Rasmussen, J.T., 2002. Isolation and characterization of MUC15, a novel cell membrane-associated mucin. *European Journal of Biochemistry.* **269,** 2755-2763.

Palmer, M., 2001. The family of thiol-activated, cholesterol-binding cytolysins. *Toxicon.* **39,** 1681-1689.

Parry, C.M., Diep, T.S., Wain, J., Hoa, N.T., Gainsborough, M., Nga, D., Davies, C., Phu, N.H., Hien, T.T., White, N.J., Farrar, J.J., 2000. Nasal carriage in Vietnamese children of *Streptococcus pneumoniae* resistant to multiple antimicrobial agents. *Antimicrobial Agents and Chemotherapy.* **44,** 484-488.

Paton, J.C., Andrew, P.W., Boulnois, G.J., Mitchell, T.J., 1993. Molecular analysis of the pathogenicity of *Streptococcus pneumoniae*: the role of pneumococcal proteins. *Annual Review of Microbiology.* **47,** 89-115.

Peltola, H., 2001. Burden of meningitis and other severe bacterial infections of children in africa: implications for prevention. *Clinical Infectious Diseases.* **32,** 64-75.

Pelton, S.I., Dagan, R., Gaines, B.M., Klugman, K.P., Laufer, D., O'Brien, K., Schmitt, H.J., 2003. Pneumococcal conjugate vaccines: proceedings from an interactive symposium at the 41st Interscience Conference on Antimicrobial Agents and Chemotherapy. *Vaccine.* **21,** 1562-1571.

Perichon, B., Tankovic, J., Courvalin, P., 1997. Characterization of a mutation in the parE gene that confers fluoroquinolone resistance in *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy.* **41,** 1166-1167.

Pettigrew, M.M., Fennie, K.P., York, M.P., Daniels, J., Ghaffar, F., 2006. Variation in the presence of neuraminidase genes among *Streptococcus pneumoniae* isolates with identical sequence types. *Infection and Immunity.* **74,** 3360-3365.

Poland, G.A., 1999. The burden of pneumococcal disease: the role of conjugate vaccines. *Vaccine.* **17,** 1674-1679.

Poolman, B., 1993. Energy transduction in lactic acid bacteria. *FEMS Microbiology Reviews.* **12,** 125-147.

Postma, P.W. & Lengeler, J.W., 1985. Phosphoenolpyruvate:carbohydrate phosphotransferase system of bacteria. *Microbiological Reviews.* **49,** 232-269.

Postma, P.W., Lengeler, J.W., Jacobson, G.R., 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiological Reviews.* **57,** 543-594.

Pritchard, D.G., Lin, B., Willingham, T.R., Baker, J.R., 1994. Characterization of the group B streptococcal hyaluronate lyase. *Archives of Biochemistry and Biophysics.* **315,** 431-437.

Quin, L.R., Carmicle, S., Dave, S., Pangburn, M.K., Evenhuis, J.P., McDaniel, L.S., 2005. In vivo binding of complement regulator factor H by *Streptococcus pneumoniae*. *The Journal of Infectious Diseases.* **192,** 1996-2003.

Rao M.V.R., D., S.M., 2006. Purification and properties of beta-galactosidase from *Streptococcus thermophilus*. *Journal of Food Science.* **46,** 1419-1423.

Regev-Yochay, G., Raz, M., Dagan, R., Porat, N., Shainberg, B., Pinco, E., Keller, N., Rubinstein, E., 2004. Nasopharyngeal carriage of *Streptococcus pneumoniae* by adults and children in community and family settings. *Clinical Infectious Diseases.* **38,** 632-639.

Reinert, R.R., Reinert, S., van der Linden, M., Cil, M.Y., Al-Lahham, A., Appelbaum, P., 2005. Antimicrobial susceptibility of *Streptococcus pneumoniae* in eight European countries from 2001 to 2003. *Antimicrobial Agents and Chemotherapy.* **49,** 2903-2913.

Rijkers, G.T., Sanders, E.A., Breukels, M.A., Zegers, B.J., 1998. Infant B cell responses to polysaccharide determinants. *Vaccine.* **16,** 1396-1400.

Rogan, M.P., Geraghty, P., Greene, C.M., O'Neill, S.J., Taggart, C.C., McElvaney, N.G., 2006. Antimicrobial proteins and polypeptides in pulmonary innate defence. *Respiratory Research.* **7,** 29.

Rogers, H., Perkins, H.R. and Ward, J.B., 1980. Formation of cell wall polymers. *Microbial Cell Wall and Membranes.* In C.Nombela ed. London: Chapman and Hall. pp.: 437-460.

Romero-Steiner, S., Pilishvili, T., Sampson, J.S., Johnson, S.E., Stinson, A., Carlone, G.M., Ades, E.W., 2003. Inhibition of pneumococcal adherence to human nasopharyngeal epithelial cells by anti-PsaA antibodies. *Clinical and Diagnostic Laboratory Immunology.* **10,** 246-251.

Ronda, C., Garcia, J.L., Garcia, E., Sanchez-Puelles, J.M., Lopez, R., 1987. Biological role of the pneumococcal amidase. Cloning of the lytA gene in *Streptococcus pneumoniae*. *European Journal of Biochemistry.* **164,** 621-624.

Rose, M.C., 1992. Mucins: structure, function, and role in pulmonary diseases. *The American Journal of Physiology.* **263,** L413-29.

Rose, M.C. & Voynow, J.A., 2006. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiological Reviews.* **86,** 245-278.

Rosenow, C., Maniar, M., Trias, J., 1999. Regulation of the alpha-galactosidase activity in *Streptococcus pneumoniae*: characterization of the raffinose utilization system. *Genome Research.* **9,** 1189-1197.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., Masure, H.R., 1997. Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Molecular Microbiology.* **25,** 819-829.

Rountree, P.M., Beard, M.A., Arter, W., Woolcock, A.J., 1967. Further studies on the nasal flora of people of Papua-New Guinea. *The Medical Journal of Australia.* **1,** 967-969.

Saier, M.H.,Jr, Chauvaux, S., Deutscher, J., Reizer, J., Ye, J.J., 1995. Protein phosphorylation and regulation of carbon metabolism in gram-negative versus gram-positive bacteria. *Trends in Biochemical Sciences.* **20,** 267-271.

Saito, H. & Miura, K.I., 1963. Preparation of transforming deoxyribonucleic acid by phenol treatment. *Biochimica et Biophysica Acta.* **72,** 619-629.

Sampson, J.S., O'Connor, S.P., Stinson, A.R., Tharpe, J.A., Russell, H., 1994. Cloning and nucleotide sequence analysis of psaA, the *Streptococcus pneumoniae* gene encoding a 37-kilodalton protein homologous to previously reported *Streptococcus sp*. adhesins. *Infection and Immunity*. **62,** 319-324.

Schneewind, O., Fowler, A., Faull, K.F., 1995. Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science.* **268,** 103-106.

Schuhl, J.F., 1995. Nasal mucociliary clearance in perennial rhinitis. *Journal of investigational Allergology & Clinical Immunology.* **5,** 333-336.

Scott, J.A., 2008. The global epidemiology of childhood pneumonia 20 years on. *Bulletin of the World Health Organization.* **86,** 494-496.

Shakhnovich EA, FAU - King, S.J., King SJ, FAU - Weiser, J.N., Weiser JN, 2002. Neuraminidase expressed by *Streptococcus pneumoniae* desialylates the lipopolysaccharide of *Neisseria meningitidis* and *Haemophilus influenzae:* a paradigm for interbacterial competition among pathogens of the human respiratory tract. - *Infection and Immunity. **70**: 7161-4.*

Sicard, A.M., 1964. A New Synthetic Medium for *Diplococcus Pneumoniae,* and its use for the study of reciprocal transformations at the Ami A Locus. *Genetics.* **50,** 31-44.

Sheehan, J.K., Kirkham, S., Howard, M., Woodman, P., Kutay, S., Brazeau, C., Buckley, J., Thornton, D.J., 2004. Identification of molecular intermediates in the assembly pathway of the MUC5AC mucin. *The Journal of Biological Chemistry.* **279,** 15698-15705.

Sheehan, J.K., Thornton, D.J., Somerville, M., Carlstedt, I., 1991. Mucin structure. The structure and heterogeneity of respiratory mucus glycoproteins. *The American Review of Respiratory Disease.* **144,** S4-9.

Skov Sorensen, U.B., Blom, J., Birch-Andersen, A., Henrichsen, J., 1988. Ultrastructural localization of capsules, cell wall polysaccharide, cell wall proteins, and F antigen in pneumococci. *Infection and Immunity*. **56,** 1890-1896.

Sleeman, K.L., Griffiths, D., Shackley, F., Diggle, L., Gupta, S., Maiden, M.C., Moxon, E.R., Crook, D.W., Peto, T.E., 2006. Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. *The Journal of Infectious Diseases.* **194,** 682-688.

Smillie, W.G. & Jewett, O.F., 1940. The relationship of immediate family contact to the transmission of type-specific pneumococci. *American Journal of Epidemiology.* **32,** 79-88.

Smith, A.W., Roche, H., Trombe, M.C., Briles, D.E., Hakansson, A., 2002. Characterization of the dihydrolipoamide dehydrogenase from *Streptococcus pneumoniae* and its role in pneumococcal infection. *Molecular Microbiology.* **44,** 431-448.

Sorensen, U.B., Henrichsen, J., Chen, H.C., Szu, S.C., 1990. Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of *Streptococcus pneumoniae* revealed by immunochemical methods. *Microbial Pathogenesis.* **8,** 325-334.

Steenhoff, A.P., Shah, S.S., Ratner, A.J., Patil, S.M., McGowan, K.L., 2006. Emergence of vaccine-related pneumococcal serotypes as a cause of bacteremia. *Clinical Infectious Diseases.* **42,** 907-914.

Stewart, G.R., Wernisch, L., Stabler, R., Mangan, J.A., Hinds, J., Laing, K.G., Young, D.B., Butcher, P.D., 2002. Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology.* **148,** 3129-3138.

Syrjanen, R.K., Kilpi, T.M., Kaijalainen, T.H., Herva, E.E., Takala, A.K., 2001. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Finnish children younger than 2 years old. *The Journal of Infectious Diseases.* **184,** 451-459.

Takasaki, S. & Kobata, A., 1986. Asparagine-linked sugar chains of fetuin: occurrence of tetrasialyl triantennary sugar chains containing the Gal beta 1-3GlcNAc sequence. *Biochemistry.* **25,** 5709-5715.

Talkington, D.F., Brown, B.G., Tharpe, J.A., Koenig, A., Russell, H., 1996. Protection of mice against fatal pneumococcal challenge by immunization with pneumococcal surface adhesin A (PsaA). *Microbial Pathogenesis.* **21,** 17-22.

Terra, V.S., Homer, K.A., Rao, S.G., Andrew, P.W., Yesilkaya, H., 2009. Characterization of a novel {beta}-galactosidase activity contributing to glycoprotein degradation and virulence in *Streptococcus pneumoniae*. *Infection and Immunity.***78**, 348-57.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K., Fraser, C.M., 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science.* **293,** 498-506.

Thompson, J., Robrish, S.A., Bouma, C.L., Freedberg, D.I., Folk, J.E., 1997. Phospho-beta-glucosidase from *Fusobacterium mortiferum:* purification, cloning, and inactivation by 6-phosphoglucono-delta-lactone. *Journal of Bacteriology.* **179,** 1636-1645.

Thompson, J., Ruvinov, S.B., Freedberg, D.I., Hall, B.G., 1999. Cellobiose-6-phosphate hydrolase (CelF) of *Escherichia coli*: characterization and assignment to the unusual family 4 of glycosylhydrolases. *Journal of Bacteriology.* **181,** 7339-7345.

Thornton, D.J., Rousseau, K., McGuckin, M.A., 2008. Structure and function of the polymeric mucins in airways mucus. *Annual Review of Physiology.* **70,** 459-486.

Thornton, D.J. & Sheehan, J.K., 2004. From mucins to mucus: toward a more coherent understanding of this essential barrier. *Proceedings of the American Thoracic Society.* **1,** 54-61.

Tilley, S.J. & Saibil, H.R., 2006. The mechanism of pore formation by bacterial toxins. *Current Opinion in Structural Biology.* **16,** 230-236.

Tomasz, A., 1984. Buiding and Breaking of bonds in the cell wall of bacteria- the role for autolysin. *Microbial Cell Wall and Autolysins.* In C. Nombela ed. The Netherlands: Elsevier Science Publishers, Amsterdam. 3-12.

Tomasz, A., Moreillon, P., Pozzi, G., 1988. Insertional inactivation of the major autolysin gene of *Streptococcus pneumoniae*. *Journal of Bacteriology.* **170,** 5931-5934.

Tomasz, A., Zanati, E., Ziegler, R., 1971. DNA uptake during genetic transformation and the growing zone of the cell envelope. *Proceedings of the National Academy of Sciences of the United States of America.* **68,** 1848-1852.

Tong, H.H., Li, D., Chen, S., Long, J.P., DeMaria, T.F., 2005. Immunization with recombinant *Streptococcus pneumoniae* neuraminidase NanA protects chinchillas against nasopharyngeal colonization. *Infection and Immunity*. **73,** 7775-7778.

Tseng, H.J., McEwan, A.G., Paton, J.C., Jennings, M.P., 2002. Virulence of *Streptococcus pneumoniae*: PsaA mutants are hypersensitive to oxidative stress. *Infection and Immunity*. **70,** 1635-1639.

Tuomanen, E., 1999. Molecular and cellular biology of pneumococcal infection. *Current Opinion in Microbiology.* **2,** 35-39.

Valles, X., Flannery, B., Roca, A., Mandomando, I., Sigauque, B., Sanz, S., Schuchat, A., Levine, M., Soriano-Gabarro, M., Alonso, P., 2006. Serotype distribution and antibiotic susceptibility of invasive and nasopharyngeal isolates of *Streptococcus pneumoniae* among children in rural Mozambique. *Tropical Medicine & International Health.* **11,** 358-366.

van Dam, J.E., Fleer, A., Snippe, H., 1990. Immunogenicity and immunochemistry of *Streptococcus pneumoniae* capsular polysaccharides. *Antonie van Leeuwenhoek.* **58,** 1-47.

van der Poll, T. & Opal, S.M., 2009. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet.* **374,** 1543-1556.

Vilei, E.M., Correia, I., Ferronha, M.H., Bischof, D.F., Frey, J., 2007. Beta-D-glucoside utilization by *Mycoplasma mycoides* subsp. *mycoides* SC: possible involvement in the control of cytotoxicity towards bovine lung cells. *BMC Microbiology.* **7,** 31.

Waltman, W.D., McDaniel, L.S., Gray, B.M., Briles, D.E., 1990. Variation in the molecular weight of PspA (pneumococcal surface protein A) among *Streptococcus pneumoniae*. *Microbial Pathogenesis.* **8,** 61-69.

Wickstrom, C., Davies, J.R., Eriksen, G.V., Veerman, E.C., Carlstedt, I., 1998. MUC5B is a major gel-forming, oligomeric mucin from human salivary gland, respiratory tract and endocervix: identification of glycoforms and C-terminal cleavage. *The Biochemical Journal.* **334,** 685-693.

Winkelstein, J.A., 1981. The role of complement in the host's defense against *Streptococcus pneumoniae*. *Reviews of Infectious Diseases.* **3,** 289-298.

Wu, K., Zhang, X., Shi, J., Li, N., Li, D., Luo, M., Cao, J., Yin, N., Wang, H., Xu, W., He, Y., Yin, Y., 2010. Immunization with a combination of three pneumococcal proteins confers additive and broad protection against *Streptococcus pneumoniae* Infections in Mice. *Infection and Immunity.* **78,** 1276-1283.

Yesilkaya, H., Kadioglu, A., Gingles, N., Alexander, J.E., Mitchell, T.J., Andrew, P.W., 2000a. Role of manganese-containing superoxide dismutase in oxidative stress and virulence of *Streptococcus pneumoniae*. *Infection and Immunity*. **68,** 2819-2826.

Yesilkaya, H., Manco, S., Kadioglu, A., Terra, V.S., Andrew, P.W., 2008. The ability to utilize mucin affects the regulation of virulence gene expression in *Streptococcus pneumoniae*. *FEMS Microbiology Letters.* **278,** 231-235.

Yesilkaya, H., Soma-Haddrick, S., Crennell, S.J., Andrew, P.W., 2006. Identification of amino acids essential for catalytic activity of pneumococcal neuraminidase A. *Research in Microbiology*. **157,** 569-574.

Yesilkaya, H., Spissu, F., Carvalho, S.M., Terra, V.S., Homer, K.A., Benisty, R., Porat, N., Neves, A.R., Andrew, P.W., 2009. Pyruvate formate lyase is required for pneumococcal fermentative metabolism and virulence. *Infection and Immunity*. **77**, 5418-2.

Yother, J. & Briles, D.E., 1992. Structural properties and evolutionary relationships of PspA, a surface protein of *Streptococcus pneumoniae*, as revealed by sequence analysis. *Journal of Bacteriology*. **174,** 601-609.

Zahner, D. & Hakenbeck, R., 2000. The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. *Journal of Bacteriology*. **182,** 5919-5921.

Zeleny, R., Altmann, F., Praznik, W., 1997. A capillary electrophoretic study on the specificity of beta-galactosidases from Aspergillus oryzae, *Escherichia coli*, *Streptococcus pneumoniae*, and *Canavalia ensiformis* (jack bean). *Analytical Biochemistry*. **246,** 96-101.

Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S., Ohwaki, M., Tuomanen, E., 2000. The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell*. **102,** 827-837.

Zhu, J. & Shimizu, K., 2004. The effect of pfl gene knockout on the metabolism for optically pure D-lactate production by *Escherichia coli*. *Applied Microbiology and Biotechnology*. **64,** 367-375.