GENETIC AND BIOCHEMICAL ANALYSIS OF HUMAN PEPTIDOGLYCAN RECOGNITION PROTEIN-S

SHÜNÉ OLIVER

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements of the degree of Master of Science

Johannesburg, 2009

DECLARATION

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

_____

(Signature of candidate)

_____ day of _____ 2009

ABSTRACT

Peptidoglycan Recognition Proteins (PGRPs) are a family of conserved proteins with antimicrobial activity in humans. PGRP-S is a 20kDA monomer with a poorly elucidated bactericidal mechanism. The aim of this project was to examine the genetic variation of the PGRP-S gene in the Sub-Saharan African immunocompromised population. DNA was extracted from the blood of HIV positive patients with and without active Tuberculosis as well as healthy controls. The coding regions of PGRP-S were amplified and subjected to direct automated sequencing. A wild type PGRP-S GST fusion protein was cloned, overexpressed, purified and used to determine protein-protein interactions as well as used in antimicrobial assays. Three polymorphic sites were noted in the population, as well as a stable variation in the genomic DNA from the mRNA sequence. One HIV positive Venda patient had a potentially novel non-synonymous mutation which would lead to a variant of protein with a reduced capacity to bind peptidoglycan. The variation of the genomic DNA to its corresponding nucleotide in the mRNA at position 80 may represent an example of RNA editing. The protein cloned had no discernable activity. Further studies with a larger population group and further analysis of the non-synonymous mutant protein may yield insight on the genetics and biochemistry of PGRP-S.

ACKNOWLEDGEMENTS

My supervisor, Dr Monde Ntwasa

The staff of the HIV clinic at Johannesburg General Hospital for their gracious help without which this project would not have been possible

Dr Lizette Koekemoer and Prof Tracy McLellan for their valuable insight, critique and resources that allowed this project to happen

Arshad Mather and Christopher Arnot for advice and technical assistance

Mpho Rakgotho for his invaluable assistance with automated sequencing and sequencing analysis

Rodney Hull for his continual scientific and intellectual input

The National Research Foundation for student funding

LIST OF ABBREVIATIONS

AIDS	Acquired Immune Deficiency Syndrome
AMPs	Antimicrobial peptides
D element	Diversity element
DAP	Diaminopimelic
DC	Dendritic cell
E. coli	*Escherichia coli*
E. faecalis	*Enterococcus faecalis*
hBD	Human β defensin
HIV	Human Immunodeficiency Virus
HNP	Human Neutrophil Protein
J element	Joining element
L. acidophilus	*Lactobacillus acidophilus*
L. monocytogenes	*Listeria monocytogenes*
LPS	Lipopolysaccharide
LRR	Leucine rich repeat
Lys	Lysine
M. luteus	*Micrococcus luteus*
MHC	Major Histocompatability complex
NAG	N-acetylglucosamine
NAM	N-acetyl muramic acid
Nec	Necrotic
NFκB	Nuclear Factor κ B
PAMPs	Pathogen associated molecular patterns
PGRP	Peptidoglycan recognition protein
PRR	Pathogen recognition receptor
Psh	Persephone
RAG	Recombinase activating gene
S. agalactiae	*Streptococcus agalactiae*
S. epidermis	*Staphylococcus epidermis*
semml	Semmelweis
TH	T-Helper
TLR	Toll like receptor
V element	Variable element
βGRP	β glycan recognition protein

List of figures

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1. Innate and Adaptive immunity

Defence mechanisms against parasites and pathogens are of paramount importance to the survival of an organism. From the simplest to the most complex organisms have evolved some form of protection against unwanted organisms. This defence system is known as the immune system. Two systems of immunity exist: the ancient innate or natural immune system and the younger adaptive or acquired immune system. As the adaptive immune system occurs only in chordates (Pancer and Cooper, 2006), more interest has been placed on this system, assuming that the younger system developed to compensate for the perceived inadequacy of the more ancient innate immune system (Medzhitov and Janeway, 1997a). However, interest in the innate immune system was reawakened by the fact that the innate immune system directs the adaptive immune system (Fearon and Locksley, 1996) and arguments have been given about the importance of the innate immune system (Medzhitov and Janeway, 1997a).

There are several basic differences in the innate and adaptive immune system, as summarised in Table 1. The innate immune system is the first line of defence encountered by foreign bodies. It mounts an immediate maximal and non-specific response. The innate immune system is also capable of perfect differentiation of self from non-self due to the recognition system employed (Janeway, 1992). It is involved in most inflammatory responses as macrophages, mast cells, neutrophils and eosinophils all bear receptors of the innate immune system. These cells become activated during the inflammatory response (Janeway and Medzhitov, 2002).

The evolutionarily younger adaptive immune system is slower than the more generalised innate immune system. This lag in maximal response time is due to

immunoglobulin receptor rearrangement required by the adaptive immune system. This rearrangement leads to specificity towards a unique antigen and generates immunological memory, a hallmark of adaptive immunity. Adaptive immunity is therefore the system required for vaccine induced immunity. However, adaptive immunity is involved in allergy, autoimmunity and graft recognition. All this is due to the recognition system employed by the adaptive immune system (Janeway and Medzhitov, 2002).

Table 1: Basic differences between the innate and adaptive immune systems, based on differences sited in Fearon and Locksley (1996) Medzhitov and Janeway (2001).

Factor	**Innate Immunity**	**Adaptive immunity**
Occurrence	All eukaryotic organisms	Chordates, *Agnatha* onward
Specificity	Non-specific, general response	Specific, pathogen targeted response
Action time	Immediate, maximal response	Delayed activation of effector molecules
Memory	No immunological memory generated	Generation of immunological memory
Receptors	Germline encoded, no rearrangement needed	Encoded in gene segments, rearrangement required
Distribution	Non-clonal	Clonal
Self discrimination	Perfect discrimination of self from non-self	Imperfect, can lead to auto-immunity
Recognition	Pathogen associated molecular patterns (PAMPS) recognised by Pathogen recognition receptors (PRRs)	Details of molecular structure recognised by human leukocyte antigens, lock and key mechanism

2. RECOGNITION BY THE IMMUNE SYSTEM

Adaptive immunity is an almost infinitely flexible system mediated by B and T-lymphocytes. The Variable, Domain and Joining elements of their immunoglobulin and T-cell receptor genes are rearranged by the products of the Recombinase Activing (RAG-1 and RAG-2) genes. This rearrangement can result in the creation of as many as 10^{11} lymphocyte clones which express distinct antigen receptors (Thompson, 1995)

B-lymphocyte receptors recognise protein, carbohydrate and simple chemical group conformations of native antigens, while most T-lymphocytes recognise protein antigens derived from peptides bound to cell surface proteins known as the Major Histocompatibility complex (MHC). Adaptive immune recognition is peptide based allows discrimination of a broader range of molecular structures and T-cells can recognise cytosol-derived non-glycosylated viral encoded proteins. This recognition strategy has been adopted to cope with the genetic variability of possible pathogens. However, this very strategy has resulted in the sacrifice of the ability to differentiate between potential pathogens and innocuous substances (Fearon and Locksley, 1996).

The innate immune system recognises pathogens by conserved products of microbial pathogens that are not present in the host. Foreign objects, particularly pathogens, have molecular patterns that are strongly conserved in evolution and are unlikely to be replaced easily or rapidly. These Pathogen Associated Molecular Patterns (PAMPs) are distinctive features that will not be found in the host, allowing for perfect discrimination of self from non-self. Receptors of the innate immune system recognises signatures such as the prokaryotic cell wall components peptidoglycan and lipoteichoic acid, mannans from yeast and double stranded RNA found in viruses (Medzhitov and Janeway, 1997a).

PAMPs are recognised by a limited number of germline-encoded receptors called Pathogen Recognition Receptors (PRRs). There are three types of PRRs: Humoral proteins circulating in the plasma, transmembrane signalling or endocytic receptors or intracellular receptors. The recognition can either be direct, as with collectins which directly recognise and bind to pathogens with certain carbohydrate signatures. The receptor may also be triggered by the products of PAMP recognition, such as with classical complement pathway which is triggered by either the activation of antigen-antibody-complement complex known as the C1 complex (Medzhitov and Janeway, 1997b).

Invertebrates function extremely well with only an innate immune system, as can be seen by the success of insects inhabiting non-sterile habitats. It has been suggested that the adaptive immune system developed, among many reasons, due to the longevity of chordates and because the adaptive immune system is less metabolically costly (Rolff, 2007). This highlights the importance of the non-clonal system of recognition, as a common hypothesis is that the adaptive immune system developed to overcome the inflexibility of non-clonal recognition. This hypothesis does not take into account that host evasion mechanisms employed by pathogens overcome recognition by specific clonal antigen receptors. Furthermore, it is likely that if a pathogen can avoid innate recognition, it can avoid adaptive recognition (Medzhitov and Janeway, 1997a).

3. THE INNATE IMMUNE SYSTEM OF INSECTS

Functionally, insects possess only an innate immune system. It has proved successfully to be sufficient for these organisms survive and proliferate. In long-lived insects such as cockroaches there are indications of specificity and immunological memory (Gillespie et al, 1997). Archetypal insect defence reactions involve physical barriers (the gut and integument), co-ordinated responses of haemocyte subpopulations and the induction of the synthesis of antimicrobial peptides and

proteins by the fat body, the organ analogous to the mammalian liver (Gillespie et al, 1997).

3.1 The immune response of *Drosophila*

In *Drosophila*, as with most insects, the first line of defence is structural. The cuticle and integument act as physical barriers for pathogens. If an injury does occur the processes of coagulation and melanisation are induced. Melanisation occurs within minutes of injury and is mediated by crystal cells. Pathogens which do manage to escape these early defence mechanisms are phagocytosed by plasmatocytes and encapsulated. Encapsulation is a lamellocyte-mediated process whereby an overlapping sheath of haemocytes is formed around the target. (Lavine and Strand, 2002). Antimicrobial peptides (AMPs) can be considered the last component of the insect's innate response as it takes the longest to respond. They are released in two phases. The peptides are first expressed locally in barrier tissues, and later are distributed systemically (Khush and Lemaitre, 2000).

Gram positive bacteria and fungi activate primarily the Toll pathway while anti Gram negative bacterial responses are mediated by the Imd pathway. This has the implication that *Drosophila* exhibits some specificity in its immune responses (Khush and Lemaitre, 2000).

One of the critical questions of innate immunity was how an extracellular event such as an infection results in the intracellular production of antimicrobial peptides The missing key was the Toll receptor, a transmembrane protein which is linked to the extracellular compartment where signalling takes place.

4. THE TOLL RECEPTOR AND INNATE IMMUNITY IN INSECTS

4.1 Characteristics of the Toll Receptor

The Toll receptor was first described as a critical mediator in fly embryology as the determinant of dorsoventral polarity formation (Hashimoto et al, 1988). In 1995, the Toll receptor was shown to induce an immune response in cell culture (Rosetto et al, 1995). Shortly afterwards, in 1996, the Toll pathway was demonstrated to be the cascade initiating antifungal defence in *Drosophila* (Lemaitre et al, 1996). The control mechanisms leading to the synthesis of antimicrobial peptides were largely unknown until this discovery. In 1995 the same group had described the immune deficient (*imd*) mutation that impaired the inducibility of all antibacterial peptides in *Drosophila*. The mutation, however, did not affect the expression of the antifungal peptide drosomycin. This indicated that different pathways existed to induce antifungal and antibacterial peptide genes. The dorsoventral regulatory pathway was examined as the pathway had parallels to the cytokine-induced NFκB activation cascade found in mammals which is needed for the induction of a number of pro-inflammatory mediators. Several of the proteins involved in the pathway leading to dorsoventral polarity and the activation of NFκB are homologous. The parallels in the signal transduction components were examined when kappa-B binding sites were observed in antibacterial peptide genes. It was demonstrated that any loss of function in the dorsoventral pathway impaired the induction of the antifungal peptide *drosomycin*. Mutations in the *snake* and *easter* genes, which function upstream of *spätzle*, the ligand for the *Drosophila* Toll receptor, did not affect drosomycin inducibility, highlighting that in immunity a different proteolytic cascade was involved in activating *spätzle* and triggering Toll signalling. (Lemaitre et al, 1996).

Demonstrating the speed of progress in the field of Toll research, a human homologue of *Drosophila Toll* was described in 1997. The same study demonstrated that the *Toll* homologue was also involved in innate immunity (Medzhitov et al, 1997). At the same time that Medzhitov's human homologue was described another group had discovered a family of human receptors structurally related to *Drosophila*

Toll (Rock et al, 1998). Both groups began their search for homologues by using expressed-sequence tag databases. However, the human Toll (hToll) described by Medzhitov was identical to the receptor termed Toll-like receptor 4 by Rock and co-workers (Rock et al, 1998). The nomenclature still stands and Toll-like receptor is often thought of as the "original Toll-like receptor" (Vasselon and Detmers, 2002).

Toll is a type I transmembrane receptor containing a leucine rich ectodomain, which when deleted results in a constitutively active protein. The ectodomain also contains one or two cysteine rich regions. The intracellular domain of the receptor is homologous to a similar region in the intracellular region of the Interleuken-1 (IL-1) receptor. The IL-1 receptor leads to activation of the transcription factor NFκ-B and subsequently immune activation. Due to this homology, that particular region of the intracellular domain is known as the Toll/IL-1 or TIR domain (Akira and Takeda, 2004).

TIR domains are not only found in Toll and Toll Like Receptors (TLR); rather an entire TIR superfamily exists. The members of the superfamily are extremely diverse, but are all involved in host defence or inflammation. Members of the TIR superfamily can broadly be divided into three subgroups. The first subgroup contains extracellular immunoglobulin domains. The second subgroup consists of proteins containing leucine rich repeats. The third subgroup consists of three members: MyD88, A46R and A52. These proteins are not receptors, but cytosolic adapter molecules (O'Neill, 2000).

In *Drosophila,* several Toll-related proteins exist. These are Toll3 to 9, 18 wheeler (18W), Mst and STSDm2245. When comparing the conserved TIR domain Toll3 and 4 share 79% identity, Toll and Toll 5 have 60% identity, as does 18w and Toll7. Toll6 is less related to the other *Drosophila* Tolls (Tauszig et al, 2000). The Toll-related receptors are expressed in different tissues during development. Toll5, also known as Tehao, is structurally similar to Toll, and has the most similar expression

pattern. Toll8, also known as Tollo, has an expression pattern similar to 18W (Kambris et al, 2002).

Toll6 is expressed in the nervous system. Toll4 and 9 have a restricted expression pattern. After germ band extension Toll9 is no longer detectable in the embryo. The receptors are also differentially expressed in immune responsive larval tissues. Most Tolls are found in the fat body. In circulating blood cells Toll, 18W, and Toll8 are present, while 18W, Toll4, 8, and 9 are found in the lymph glands. Of the Toll related receptors it is known that 18W, Toll3, 4, 6, 7 and 8 do not activate NFκB related molecules or interact with the critical adapter molecule DmMyD88. (Kambris et al, 2002). However, it has been demonstrated that 18W, although dispensable for an immune response in adult *Drosophila*, is required for the induction of antimicrobial peptides in larvae (Ligoxygakis et al, 2002). Furthermore, Toll5 can activate the drosomycin promoter, although this has only been demonstrated in cell culture (Tauszig et al, 2000). The exception to the theory that *Drosophila* Toll-related receptors are not immune signalling molecules is Toll9. Toll9 has an ectodomain and TIR domain similar to mammalian TLRs (Bilak et al, 2003) and may use the same signalling pathway to activate immune genes (Bettencourt et al, 2004).

5. TOLL LIKE RECEPTORS IN MAMMALS

Toll-Like Receptors (TLRs) are a class of pattern recognition receptors found in mammals and are structurally related to the *Drosophila* Toll receptor. TLRs are defined as paralogues of *Drosophila* Toll. They share a tertiary fold called the Toll-homology domain (TH) domain. The fold consists of a minimum of 128 amino acids. This domain is not only found in Toll and TLRs, but in the IL-1R family, plant disease resistance proteins and MyD88 factors. Toll and TLRs also share a characteristic Leucine Rich Repeat (LRR) ectodomain (Rock et al, 1998).

Rock and co-workers predicted that the existence of LRR indicated an affinity for Spätzle or another cystine-knot related factor (Rock et al, 1998). Indeed, there was some question as to whether Toll-like receptors were true PRRs. *Drosophila* Toll is not a true PRR as pattern recognition takes place upstream of Spätzle, which results in the cleavage and activation of the Toll ligand. No Spätzle-like protein has been described in humans. Although no direct crystallographic evidence exists of TLRs binding to a PAMP, the co-precipitation of TLRs and their PAMP ligands is the evidence currently used to define TLRs as true PRRs (Andersen-Nissen et al, 2007)

There are currently eleven members of the TLR family known in humans, with the latest, TLR-11 being described in 2004. Human TLRs can be divided into 5 subfamilies; the TLR-2, TLR-3, TLR-4, TLR-5 and TLR-9 subfamilies. TLR-1, 2, 6 and 10 constitute the TLR-2 subfamily, TLR-7, 8 and 9 make up the TLR-9 subfamily. TLR-3, 4 and 5 are the only members of their subfamily. The families are based on evolutionary relationships but different subfamilies appear to have evolved to recognise specific ligands. Members of the TLR-2 subfamily are specific for lipopeptide PAMPs including zymosan and Peptidoglycan. Members of TLR-3 family recognise double stranded RNA. LPS is recognised by receptors in the TLR-4 family and members of the TLR-7 family nucleic acid and haem motifs (Roach et al, 2005). There are a limited number of TLR receptors, yet they recognise a wide variety of PAMPs. The recognition repertoires of these receptors are increased by co-operation between TLRs by dimerisation. The cytoplasmic domains of TLR 2 and 1 or 6 associate to form dimers (Ozinsky et al, 2001). TLR dimerisation occurs within members of the same subfamilies only. One of the classifications for subfamilies is that they recognize a similar kind of PAMP, hinting at divergent evolution (Roach et al, 2005). In 2001 Muzio and Mantovani proposed that TLRs could be classed in functional groups based according to expression patterns on a variety of blood cells. TLR1 is classed as ubiquitous, TLRs 2, 4 and 5 as restricted and TLR3 as specific (Muzio and Mantovani, 2001).

The wide distribution as well as the homo and heterodimerization of the receptors allows recognition of a diversity of ligands. TLRs initiate rapid and complex signalling pathways (Ozinsky et al, 2000). These pathways include the production of chemokines, pro-inflammatory cytokines and interferons and lead to an upregulation of co-stimulatory molecules. Most crucially, TLRs are an important bridge between the innate and adaptive immune systems (Clark and Kupper, 2005).

6. THE INNATE IMMUNE SYSTEM DIRECTS ADAPTIVE IMMUNITY

As mentioned previously, Toll is critical receptor in the innate immune response of *Drosophila*, as are the TLRs for mammals. However, TLRs are not merely critical innate receptors, but represent an interface between innate and adaptive immunity. As early as 1992 Charles Janeway jr. had argued that the adaptive immune system did not function in isolation because the delayed activity of this system due to receptor rearrangement was a timely process and if the adaptive immune system relied on specific antigen ligation only, the host would die before the adaptive immune system was mobilised. The immune system therefore did not function on a basis of either innate or adaptive mechanisms. Janeway argued, in addition to his PAMP theory, that a second signal provided by the innate immune system primed the adaptive response (Janeway, 1992). He was later proved correct as by 1996, it had been agreed that the innate immune system directs the adaptive (Fearon and Locksley, 1996).

The theory proposed by Janeway (Janeway, 1992), in its logical simplicity, demonstrated that the idea that the innate and adaptive immune system functioned in isolation is a seriously flawed model. Not only are the receptors of the innate immune system more widespread and distributed on cells that would encounter an infection first, they are also represented by multiple protein families as opposed to receptors that are members of the immunoglobulin superfamily as the adaptive immune system is. The rearrangements of the receptors of the adaptive immune system are directed by somatic mechanisms, and as such, the T- and B-lymphocytes bearing the receptors

have an undefined and unpredicted specificity. The naïve lymphocytes cannot be driven to a specific function by a signal from its receptor alone. They require co-stimulatory signals produced as a product of pattern recognition by the innate immune system. PAMPs induce endogenous signals that function as co-stimulators that activate T-cells. Innate immune signals also mediate the inflammatory response (Medzhitov and Janeway, 1997b).

Innate immunity can direct the adaptive immune system by *trans* or *cis* actions. In a *cis* reaction, a covalent tag is added to an antigen so that it can be recognised by the adaptive immune system. An example of this is complement. The complement system is part of the innate immune system, which has a direct antimicrobial function in itself. The complement cascade is activated by microbial components and when activated C3b binds to nearby proteins or carbohydrates. C3b is then proteolytically processed to the C3dg fragment which can be recognised by CD21 (Cluster of Differentiation 21), a receptor expressed on B-lymphocytes and follicular dendritic cells (DCs) (Fearon, 1999).

DCs are antigen presenting cells which develop in the bone marrow and migrate to tissues in an immature form. In the immature form they cannot serve as antigen presenting cells. DCs are induced to mature by inflammatory cytokines, LPS and Tumour Necrosis Factor. In maturing DCs CD80 and 87 are upregulated and Major Histocompability Complex (MHC) Class II complexes can be trafficked to appropriate intracellular complexes. Mature DCs then turn off their endocytotic functions and the expressed MHCII complexes represent the environment in which they encountered the antigen. The DCs then alter their adhesion proteins and chemokine receptors and migrate to secondary lymphoid organs where they can encounter antigen specific T-cells (Banchereau and Steinman, 1998). Distinct sets of TLRs are expressed on the immature DCs (Kadowaki et al, 2001), and as the recognition molecules for the maturation signals induce DCs maturation and

migration and consequently primes the adaptive immune reaction by stimulating T-lymphocytes. This represents a *trans* reaction (Fearon, 1999).

The nature of the acquired immune response is also directed by the innate immune system. The differentiation of precursor helper T-Cells (T_H) into T_H1 or T_H2 cells is governed by the cytokine milieu. These instructive cytokines are produced early in infection and are produced by signalling cascade induced by PAMP recognition. T_H1 type immunity is essential for the elimination of intracellular pathogens and the development of T_H1 cells are driven by Interleukin 12 and Interferon γ. T_H2 type immunity is required for mucosal parasite infections and IL 13 drives the differentiation of T_H0 cells to T_H2 (Banchereau and Steinman, 1998). Another important mechanism whereby the innate immune system directs the adaptive immune system is by antimicrobial peptides and proteins.

7. ANTIMICROBIAL PEPTIDES AND PROTEINS

The first invertebrate peptide antibiotics were characterised by Boman and colleagues in 1981. As the peptides were obtained from the silkworm *Hyalophora cecropia* the peptide was termed cecropin (Steiner et al, 1981). Since then numerous small, mainly cationic antibiotic peptides from a variety of structurally different families have been identified (Brown and Hancock, 2006).

In mammals, many antimicrobial peptides (AMPs) are constitutively present in cells and stored in secretory granules, while others are induced by pro-inflammatory stimuli. AMPs tend to have large cationic patches on their surface which enable them to depolarize or pierce bacterial membranes which tend to be largely negatively charged (Oppenheim et al, 2003).

AMPs can be broadly classified according to their charge; anionic or cationic. Anionic AMPs are structurally similar to the charge-neutralizing pro-peptides of

larger zymogens, which when synthesized are also antimicrobial. Cationic peptides occur in three principle categories: linear helical, linear proline rich or cysteine stabilized peptides with a β-sheet structure (Brogden et al, 2003). Mammalian AMPs belong to four structural classes. These are the β-sheet peptides stabilised by two to three disulphide bridges (protegrins, defensins), amphipathic α-helical peptides (the cathlecidin LL-37), loop peptides (Bovine neutrophil peptide) and peptides generated by proteolysis of larger proteins in the host (Yan and Hancock, 2001).

7.1 Cathelicidins

Cathelin is an inhibitor of Cathepsin L and a protein which lends its name to a family of antimicrobial proteins. These AMPs have a pro-region highly homologous to the cathepsin L inhibitor cathelin and are subsequently referred to as cathlecidins (Oppenheim et al 2003). Only one cathlecidin, LL-37, has been fully characterized. LL-37 is derived by the proteolysis of human CAP18 protein, and is only active after proteolysis (Golec, 2007). It is a 37 amino acid peptide whose lack of cysteine residues results in a linear structure. The structure then adapts according to its environment, adopting a α helical structure in a hydrophobic environment and a largely random coil conformation in a hydrophilic environment. The peptide is induced by inflammatory or infectious stimuli and has activity against both Gram positive and Gram negativebacteria. It also neutralises LPS, is chemotactic for neutrophils, monocytes and mast cells (in which it can also induce degranulation), and alters macrophage transcriptional regulation. It is also involved in re-epithelisation of skin healing, wound vascularisation and also has anti tumour effects (De Smet and Contreras, 2005).

7.2 Defensins

Typical examples of cysteine stabilised peptides include the protegrins and the defensins. Defensins were initially isolated from rabbit and human neutrophils in an

attempt to characterise phagocyte-derived oxygen independent antimicrobial activities. About 50 defensins have been discovered. Vertebrate defensins are divided into Triple stranded antiparallel β sheets and cyclic peptides containing 6 cysteine residues. The two major classes, α and β defensins, are antiparallel β-sheets (Yang et al, 2001).

Mammalian defensins are cationic, non-glycosylated peptides with arginine as the primary catalytic residue. Their average molecular mass ranges from 3.5-6kDa and usually have three disulphide bridges due to the six cysteine residues characteristic of its class (De Smet and Contreras, 2005). The α or "classic" defensins contain 29-35 residue amino acids and the disulphide bridges result in a triple stranded β-sheet structure with a β-hairpin loop containing the cationic charged molecules (Ganz, 2003). α-defensins are found in the azurophilic granules of neutrophils (Yang et al, 2001) and in humans are consequently referred to as Human Neutrophil Proteins (HNPs). HNP 1-3 are found not only in neutrophils, but B-and Natural Killer Cells as well. HNP 5 and 6 are referred to as Enteric defensins as they are found in granules of Paneth cells of the intestine. α-defensins are expressed as inactive pre-propeptides with the C-terminal part of the peptide accounting for the antimicrobial activity (De Smet and Contreras, 2005) They are active against Gram negativebacteria, mycobacteria, fungi and enveloped viruses (Brogden et al, 2003).

β-defensins are slightly larger than the classic defensins. They also have three disulphide bridges resulting in a β-sheet structure and a β-hairpin loop containing cationic molecules. The first β defensin isolated, human beta defensin-1 (hBD-1) is expressed in epithelial surfaces that interacts directly with the environment or microbial floras. hBD-2 is widely expressed and can be found in epithelial tissue, bone marrow and leukocytes. hBD-3 has been isolated from lesion psoriatic scales. hBD-3 is a broad spectrum antibiotic capable of killing a range of pathogenic bacteria and yeast, while hBD-1 and 3 are predominantly active against Gram

negativebacteria (Ganz, 2003). Besides antimicrobial activity, defensins have a variety of other biological functions. These include anti tumour activity, interference with cell signalling pathways, stimulation of cytokines, stimulation of cell proliferation, chemotaxis of immune cells and adhesion molecule expression (Garcia et al, 2001).

7.3 Lysozyme

Some antibacterial peptides are constitutively expressed in insects. One such protein is lysozyme, which hydrolyses β-1, 4 linkages between N-acetylglucosamine and N-acetylmuramic acid in the peptidoglycan of bacterial cell walls. Insect lysozymes are 14 kDa proteins with sequence similarity to chicken white lysozyme. In lepidopteran species lysozyme is expressed primarily in the fat body, while in the dipteran *Drosophila melanogaster* it is expressed in the midgut where it serves as a digestive enzyme rather than an antibacterial (Gillespie et al, 1997).

Lysozymes are widely occurring enzymes found in many tissues and secretions, as well as in many organisms. They are found in bacteriophages, bacteria, plants, and higher eukaryotes. Several classes of lysozyme exist; with the best known examples being the chicken type (C-type), Goose (G-type) and Viral type (V-type) (Bachali et al, 2002).

Lysozyme ("lytic enzyme") was discovered as a bacteriolytic factor by Alexander Fleming in 1922 (Gordon et al, 1979). The enzyme is believed to function by disrupting the mechanical integrity of the cell wall by degrading peptidoglycan. However, it has been proposed that T4 lysozyme may be more similar to other AMPs in that it causes membrane disruption as non-enzymatically active versions of the protein still maintain antimicrobial activity (Düring et al, 1999).

Lysozyme is believed to be a classical example of divergent evolution as they appear to have evolved from a remote common ancestor (Evrard et al, 1998). A family of proteins that are part of lysozymes' evolutionary divergence are crucial in the recognition and defence against peptidoglycan. These are the peptidoglycan recognition proteins.

8. THE IMMUNE RESPONSE TO PEPTIDOGLYCAN: ANOTHER VITAL PATHOGEN RECOGNITION RECEPTOR ELUCIDATED

8.1 The peptidoglycan recognition system

8.1.1 Peptidoglycan, an archetypal PAMP

Peptidoglycan, also known as murein, is a highly conserved bacterial cell wall component, found in all prokaryotes except the archaebacteria. In Gram positive bacteria, it typically constitutes a large portion of the cell wall, usually a 20 to 80 nm layer outside the plasma membrane. Gram negativebacteria usually have a 2-7 nm layer, making peptidoglycan 5-10 % of their wall weight. The polymer consists of two sugar derivatives, N-acetyl glucosamine and its lactyl ester N-acetyl muramic acid as well as several amino acids not found in proteins (Scheffers and Pinho, 1995). The basic structure of peptidoglycan and the point of variation that separates diaminopimelic type peptidoglycan (found in Gram negatives) from Lysine-type peptidoglycan (found in Gram positives) is illustrated in Figure 1.

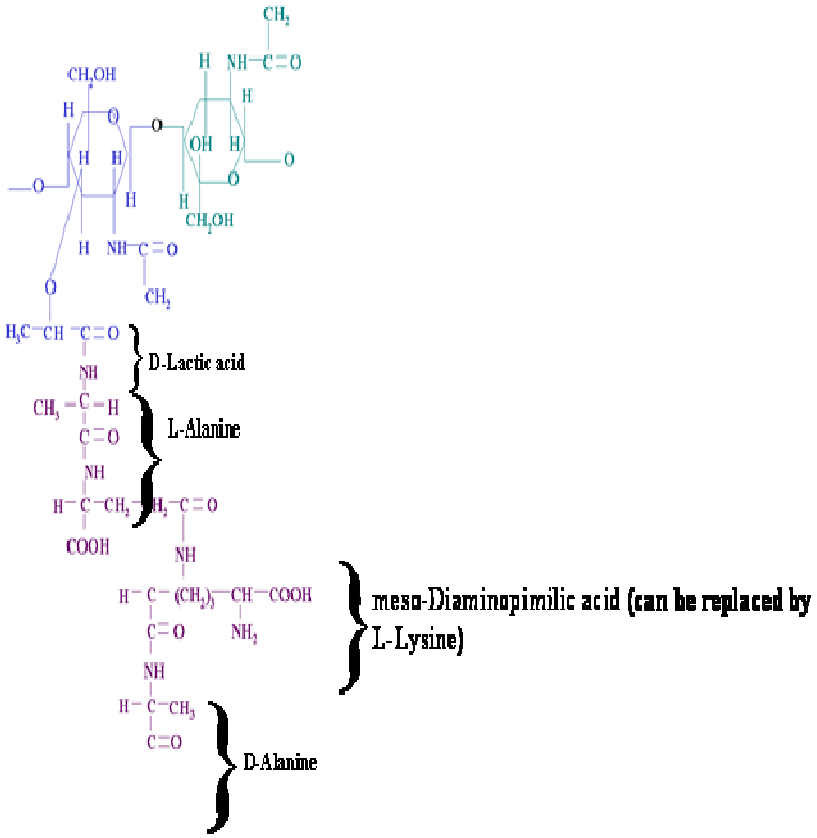

Figure 1: The structure of peptidoglycan. Green represents the N-acetylglucosamine (NAG) residue. Blue represents the lactyl ester of NAG, N-acetyl muramic acid (NAM). Purple represents the tetrapeptide chain (Based on information in Scheffers and Pinho, 1995).

Due to its highly conserved nature and simplicity of its structure, peptidoglycan is an ideal motif for recognition by the innate immune system. It can be considered an archetypal PAMP, and is recognised by a diverse array of PRRs. As such, several innate mechanisms are in place to respond to peptidoglycan detection. It was studying one of these systems, the prophenoloxidase system that lead to the discovery of family of lysozyme-related proteins required for the recognition of peptidoglycan.

8.1.2 The Prophenoloxidase cascade

The Prophenoloxidase (ProPo) cascade is an efficient mechanism of recognising non-self in invertebrates. This system can recognise and respond to concentrations of lipopolysaccharide, β-1, 3 glucans and peptidoglycan as low as picograms per litre. The ProPo cascade causes the parasite to blacken in the haemolymph. This is due to the process of melanisation, the deposition of melanin around an encapsulated object at the site of infection (Söderhall and Cerenius, 1998).

Phenoloxidase (monophenyl L-dopa: oxygen oxidoreductase) is found in the haemolymph or coelom of many invertebrate groups, found in both protostomes and deuterostomes. The enzyme catalyses the oxidation of both mono and diphenols to quinones. The quinones then non-enzymatically polymerise to form melanin (Söderhall et al, 1998). Melanisation plays a role in defence through its action in wound healing. Phenoloxidase also causes the cuticle to harden by sclerotinisation (Gillespie et al, 1997). The intermediates formed during the melanisation process, in addition to the melanin are toxic to micro organisms. In its native form Prophenoloxidase (ProP) can exist as an oligomer. The purified monomer of ProP of several invertebrates has a mass of 70-80 kDa. After cleavage the monomer has a mass of 60-70 kDa. The cleavage of ProP results in the activation of the enzyme. This activation is mediated by the action of several serine proteases. The activating serine proteases isolated from crayfish, silkworm and fruit fly are proteins of about 30 kDa. The enzyme requires cleavage, i.e. it remains inactive, partly as a regulatory measure as the intermediates of the cascade are toxic. As such the ProP zymogens are located in blood cell vesicles (Cerenius and Söderhäll, 2004). As is typical of the innate immune system, the cascade is initiated by pattern recognition by a PRR, which occurs upstream of the activating serine protease cascade. The recognition process and the subsequent production of melanin by the Prophenoloxidase cascade is schematically represented in Figure 2.

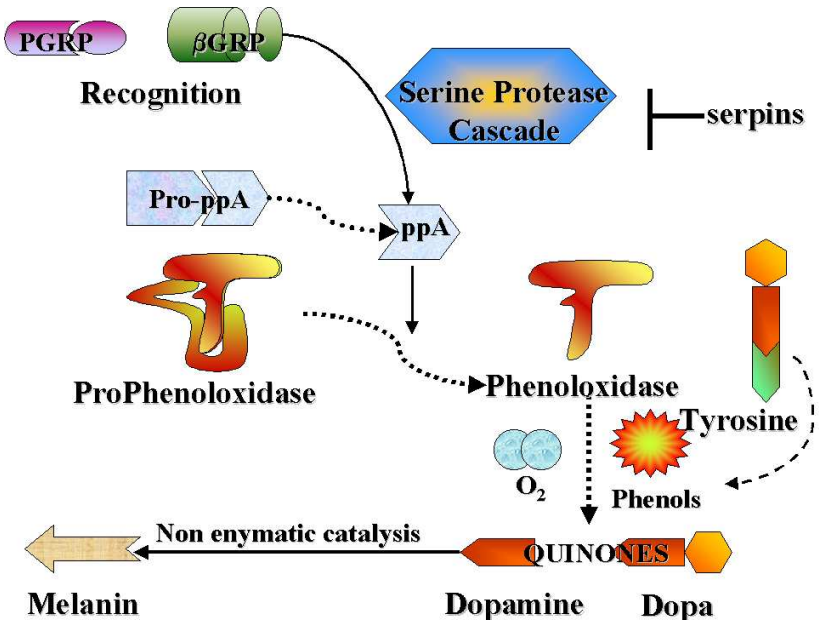

Figure 2: Schematic description of the Prophenoloxidase cascade which leads to melanisation. PGRP, Peptidoglycan recognition protein; βGRP, β-Glucan recognition protein; Pro-ppA, Pro-Prophenoloxidase activating enzyme; ppA, Prophenoloxidase activating enzyme. This diagram is based on descriptions in Söderhäll and Cerenius (1998), Cerenius and Söderhäll, (2004) and Lavine and Strand, (2002).

8.1.3 Purification of a PGRP from *Bombyx mori*

Studies of the ProP cascade led to the discovery of the original recognition receptor for peptidoglycan in the silkworm *Bombyx mori*. The fungal cell wall component β-1,3 glucan and the bacterial cell wall component peptidoglycan were known to trigger the ProP cascade. As the innate immune system works on pattern recognition, it was assumed by Yoshida and co-workers (1996), that pattern recognition receptors existed for these components. They proposed that these receptors, which they termed β-1,3 glucan recognition protein (βGRP) and peptidoglycan recognition protein (PGRP), exist to recognise β-1,3 glucan and peptidoglycan respectively. Their proposed βGRP was isolated in the silkworm, cockroach and crayfish. The group then

proceeded to isolate PGRP from *Bombyx mori.* They demonstrated that Silkworm PGRP is specific, binding only to peptidoglycan and not β-1,3 glucan or chitin. The confirmation of the existence of PGRP showed that the ProP cascade had two points of initiation, β-1,3 glucan and peptidoglycan recognition (Yoshida et al, 1996).

8.2 A family of *Drosophila* and human PGRPs

The discovery of PGRP in *Bombyx mori* began a search for homologous proteins in other organisms. *Drosophila* PGRPs, in contrast to previously discovered PGRPs, had two different classes of the protein, as opposed to the single class in the silkworm The first class, consisting of seven small, exported proteins, are similar to all PGRPs found in *Bombyx mori* and *Trichoplusia ni*. The second class of the proteins are long transcripts, and appeared to be membrane proteins. The PGRPs are widespread across the genome; over eight chromosomal loci on the three major chromosomes of the fly (Werner et al, 2000).

The short PGRPs consist of a PGRP domain preceded by a signal peptide, which was predicted to be processed so that the mature protein consists only of the PGRP domain. This particular class of proteins contains the members PGRP-SA, PGRP-SB1, PGRP-SB2, PGRP-SC1B, PGRP-SC2 and PGRP-SD. The long PGRP genes give rise to longer transcripts which often have several splice forms. This class does not encode a typical signal peptide. Three of the long transcripts, PGRP-LA, PGRP-LC and PGRP-LD have a predicted transmembrane region. PGRP-LE has a highly charged N-terminal domain connected to the PGRP domain, but has no export signal or transmembrane region. PGRP-LB encodes only a PGRP domain attached to a relatively long 5' untranslated region (Werner et al, 2000).

Drosophila PGRPs have differing expression during the lifetime of the fly. During development most PGRPs are detectable in all post embryonic stages. PGRP LD and LE are highly expressed in 0-5 hour embryos. PGRP-SB2 is expressed exclusively in the pre-pupae. After immune challenging with both Gram positive and Gram

Negative bacteria, PGRP LB, SA, SC2 and SD were strongly expressed. PGRP-LB is the only inducible long transcript. The inducible genes were found to be expressed in the fat body. PGRP-SC and SC2 showed high constitutive expression. Of the short PGRP genes, only PGRP SC was constitutively expressed at high levels in the gut (Werner et al, 2000).

In 2001 Liu et al. described the family of four human PGRPs. The 576 amino acid PGRP-L is found on chromosome 19, as is the 196 amino acid PGRP S. PGRP Iα and Iβ, 341 and 343 amino acids respectively, are found on chromosome 1.). Western Blot analysis showed PGRP L, Iα, Iβ and S to be 65, 38, 40 and 24 kDa peptides respectively. . All PGRPs bind peptidoglycan directly, as demonstrated by their ability to bind peptidoglycan bound to agarose. However, the strength of binding is not equal. While PGRP S, L and Iα bind peptidoglycan with high affinity, PGRP Iβ binds the molecule with a low affinity. This was speculated indicate that PGRP Iβ may bind another unknown ligand, or need other ligands for high affinity binding (Liu et al, 2001).

The C-terminal regions of all four PGRPs are highly conserved, containing three PGRP domains. PGRP Iα and Iβ have an additional PGRP domain, designated domain IV, in the N-terminal half. This additional domain is 89% similar to PGRP domain II (Liu et al, 2001). The PGRP domain is also known as the T-phage lysozyme homology domain (Kibardin et al, 2003). At present there are 328 PGRP domains in 284 proteins in the Simple Modular Architecture Research Tool (SMART) database, including *Drosophila* and mammalian PGRP as well as T7 lysozyme (Schultz et al, 1998; Letunic et al, 2004). A landmark study by Lazzaro and co-workers demonstrated that immune competence in wild flies is driven by polymorphisms in 16 innate immunity genes. PGRPs were of these genes, with polymorphisms in the PGRP domain leading to immunocompetence (Lazzaro et al,

2004). This demonstrates the critical role of the PGRP domain in the successful functioning if the protein.
The C-terminal region of all insect and mammalian PGRPs contains several highly conserved residues. Between the second and third domains two cysteines, an arginine, a glutamine, and a histidine residue are found. Conserved asparagine, glycine, isoleucine, phenylalanine and proline residues are found between domain I and II (Liu et al, 2001). This later proved to be important for protein structure and function, especially when the crystal structure for PGRP-S was elucidated (Guan et al, 2005).

8.3 PGRP-S/PGLYRP-1

8.3.1 PGRP-S in *Drosophila*

Short PGRPs in *Drosophila* are usually assigned three major roles; the activation of the Toll pathway (PGRP-SA) (Michel et al, 2001), activation of the Pro-Phenol oxidase cascade (Yoshida et al, 1996) and direct enzymatic activity (PGRP-SC1 and PGRP-SA), with PGRP-SC1 being an N-acetylmuramoyl-L-ala amidase and PGRP-SA being a carboxypeptidase (Chang et al, 2004). Recently, PGRP-SB1 was described as an N-acetylmuramoyl-L-ala amidase as well, and this activity gives it direct antibacterial activity, particularly against *Bacillus megaterium.* Its mode of action is therefore different to vertebrate PGRPs. (Mellroth and Steiner, 2006). Furthermore, short PGRPs were also involved in the regulation of the IMD pathway. Bischoff and co-workers described that PGRP-SC1a and b deficient flies had an over activation of the IMD pathway presumably by degrading peptidoglycan before it reaches PGRP-LC and PGRP-LE (Bischoff et al, 2006).

PGRPs have a significant homology to bacteriophage T7 lysozyme, making it part of the N-acetyl muramoyl-L-alanine amidase superfamily. This enzyme hydrolyses the bond between the N-acetylmuramoyl group in the glycan strand and the L-alanine stem peptide in peptidoglycan. It was reported by Mellroth and colleagues that a

Drosophila PGRP, PGRP-SC1B can degrade peptidoglycan to less immunostimulatory products. However, the degradation takes place to different degrees. PGRP-SC1B degrades the direct cross link between stem peptides, and peptidoglycan containing meso-diaminopimelic acid is the most resistant degradation (Mellroth et al, 2003).

Five conserved amino acids are required for the enzymatic activity of T7 lysozyme. PGRP-SC1B has four of those five residues. The fifth, a threonine 128 substitution for lysine, is consistent with a reduced lysozyme activity. Based on this requirement, there are five potentially enzymatic PGRPs; PGRP-SB1, SB2, SC1A, SC2 and LB. Indeed, PGRP-SC1B has enzymatic activity (Mellroth et al, 2003) as does PGRP-SB1 (Mellroth and Steiner, 2006). Enzymatic PGRPs differ from PGRPs with a purely receptor function by a serine for cysteine 130 substitution, thus removing one of three zinc ligands making them inactive enzymes. The cysteine residue can serve as a marker for PGRP function; the presence of a cysteine being a marker for enzymatic PGRP and the absence of a cysteine residue indicative of a receptor type PGRP. This protein is the first example of a eukaryotic N-acetyl muramyl transamidase (Mellroth et al, 2003).

PGRP-SA adopts a fold unique to the PGRP-S structure. This includes a mixed strand beta sheet and three alpha helices forming the protein core. Two additional β-strands are found at the N-terminus. PGRP-SA as well as PGRP-LB has a conserved L-shaped groove also found in T7 lysozyme. The groove is thought to represent the peptidoglycan binding site. Three highly conserved residues; Cys48, Tyr76, and Thr156 are essential for peptidoglycan recognition in *Drosophila* PGRP. It is known that a mutation of Cys48 abolishes the ability of PGRP-SA to activate the Toll receptor (Michel et al, 2001). This cysteine is involved in a disulphide bond with Cys54 at the edge of the L-shaped groove. This disulphide bond may be needed for the structural integrity of the groove (Reiser et al, 2004).

The discovery of the link between PGRP and Toll activation has been an interesting journey. In a screening of genes involved in the *Drosophila* immune pathway, Reichhart and co-workers isolated a fly line with a reduced ability to induce drosomycin, i.e. a loss of function Toll mutant, in response to Gram positive bacteria only. The mutation was called *semmelweis*, (*seml*) after the Hungarian physician Ignaz Semmelweis. *Seml* was the first mutation discovered which abolished Toll activation for Gram positive bacteria only. *Seml* was shown to act upstream of Toll as *drosomycin* transcription in a gain of function *seml* double mutant could induce drosomycin peptide expression. *Seml* does not play a role in dorso-ventral axis formation, in contrast to Toll pathway mutants, homozygous *seml* females are always fertile (Michel et al, 2001).

The s*eml* mutation was mapped to a gene coding for PGRP-SA, specifically to a single guanine to adenine mutation at base pair 74 resulting in a cysteine 80 to tyrosine mutation. The mutation lies in the PGRP domain. The *seml* phenotype was rescued by the injection of wild type haemolymph into recipient flies. This marked PGRP-SA as a circulating haemolymph protein (Michel et al, 2001).

A complementary mutant affecting only the fungal branch of Toll activation was found as follows. The *necrotic (nec*) gene encodes three serine proteases inhibitors of the serpin family. Mutations in this gene results in necrotic areas on the epidermis, manifesting as brown spots along the body and leg joints. One of these genes, Spn43Ac, can rescue this mutant phenotype. *Nec* mutants have constitutive expression of the antifungal peptide drosomycin, but not the antibacterial peptide cecropin A1. This implicates Spn43Ac in the Toll, but not imd pathway, as drosomycin is an end product of the Toll pathway and cecropin A1 of the imd pathway. Introducing the serpin into nec mutants prevented the constitutive activation of the Toll pathway. The serpin is therefore a negative regulator of the Toll pathway. The 60 kDa protein is found in the haemolymph of the fly. Spn43Ac regulates Toll as an inhibitor of one of the serine proteases required for the cleavage and activation of

the Toll ligand Spätzle. *Nec* flies do not have the cleaved form of Spätzle after immune challenge, yet have no developmental defects (Levashina et al, 1999).

If a serpin deficiency results in the constitutive expression of the Toll pathway, a serine protease is necessary for the regulation of Toll activation. What is the target protease for Spn43Ac? The same group who described the nec mutation discovered a suppressor mutation in the *nec* flies. This mutation was named Persephone (*psh*). *Nec*:*psh* double mutants had a lifespan comparable to wild type flies, as well as normal Toll and Imd functioning. The *psh* mutation was mapped to a region encoding two serine proteases. The serine protease Persephone was the first identified component of the serine protease cascade responsible for the activation of Toll during an immune challenge. Furthermore, the *psh* mutation was the first mutation described which impairs fungal induced drosomycin induction (Ligoxygakis et al, 2002).

8.3.2 Mammalian PGRP-S: the discovery of PGRP function

Although *Drosophila* is a useful model organism for many systems, caution must be exercised when drawing parallels to function in humans. This is particularly evident in the immune system. The Toll and TLR pathways are highly conserved, but TLRs serve no apparent developmental role in mammals. However, one of the first functions postulated for the TLR family was that it may be involved in the development of vertebrate embryos (Rock et al, 1998). It is now generally accepted that TLRs have no developmental functions in vertebrates (Beutler, 2004).

When PGRPs were first discovered it was confidently declared that only PGRP-S was a signalling molecule and other human PGRPs were cytoplasmic. The bold statement was made on a structural modelling prediction (Liu et al, 2001). This was later disproved by a potentially damaging article by Xu et al (2004), which claimed that PGRP-L may be dispensable for innate immunity. This was largely due to the

examination of the behaviour of macrophages in PGRP-L null mice, even though the gene is not expressed in macrophages. The experiment appeared to rely on the assumption that PGRP in mammals also behaved as a signalling molecule. However, the same study found PGRP-L as well as PGRP-Iα and β to be soluble, disproving the well postulated structural model (Xu et al, 2004). Furthermore, with PGRPs proving to be antimicrobial proteins and not signalling molecules, in contrast to the multifunctional *Drosophila* PGRPs, it would be wise to be conservative when drawing comparisons between the immune systems of *Drosophila* and humans. It is important to bear in mind the multifunctional nature of *Drosophila* proteins (e.g. PGRP and Toll) before assuming that the mammalian homologues will behave in the same manner.

Short PGRP has always had a special position in PGRP research. The PGRP discovered in *Bombyx mori* (Yoshida et al, 1996) was a short PGRP, and the first mammalian PGRP discovered and cloned was PGRP-S (Kang et al, 1998). PGRP-S is described as being originally identified and characterised by Kiselev et al in 1998 under the name of Tag7(Cho et al, 2005), at around the same period that the Steiner group discovered PGRP-S by initial differential display studies on *Trichoplusia ni*. This may give the false impression that PGRP-S and Tag 7 is the same molecule, but this will be examined more closely later in a discussion of the Tag 7 molecule. PGRP-S is the protein that leads to the discovery of the antimicrobial nature of the human PGRP family.

The path to the elucidation of PGRP function began when Roman Dziarski, Håkan Steiner and co-workers described that PGRP-S deficient mice had an increased susceptibility to infection with non-pathogenic bacteria, but mice with wild-type PGRP-S were still equally susceptible to infection by pathogenic bacteria (Dziarski et al, 2003). This was the basis of the work leading to the definition of PGRP-S as an antimicrobial protein. Based on the aforementioned study, Dziarski proposed that certain bacteria are non-pathogenic due to the action of PGRP-S (Dziarski et al,

2003). The work demonstrating the bacteriostatic and bactericidal activities of Bovine Oligosaccharide-binding protein towards both Gram positive and Gram negativebacteria (Tydell et al, 2002) as well as studies by Cho et al on the immune function of PGRP-S demonstrated a few critical basic facts of PGRP-S function. Firstly, they demonstrated that PGRP-S was capable of binding both Lys and DAP-type peptidoglycan and thus had broader ligand specificity than other mammalian PGRP-S molecules. They also demonstrated the dose dependent inhibition of *S. aureus* and *E. coli* by radial diffusion assays, and that the inhibition mechanism required peptidoglycan -dependent binding. The synergistic effect of PGRP-S with lysozyme was also demonstrated. The group also highlighted the fact that PGRP-S is microbicidal to organisms containing no peptidoglycan (Cho et al, 2005). The antimicrobial nature of PGRP-S had been determined and from the onset, it proved to be an intricate and complex system.

Shortly after the definitive description of PGRP-S' antimicrobial function, a landmark study described that all human PGRPs are actually antimicrobial and constituted a novel class of human bactericidal proteins. Lu et al (2006) demonstrated several critical facts about human PGRPs. Firstly, all human PGRPs had bactericidal activity. Both pathogenic and non-pathogenic bacteria were killed by PGRPs. Secondly, PGRPs functioned as disulphide linked dimers with different specificities. PGRP Iα and β as well as PGRP Iαβ heterodimers were highly bactericidal towards *Bacillus subtilis*, but not *Lactobacillus acidophilus*. *Listeria monocytogenes* was sensitive to PGRP I α and β homodimers but not heterodimers (Lu et al, 2006)

PGRP-S is highly bactericidal to *L. acidophilus*, but not to *L. monocytogenes*. Normal flora such as *Staphylococcus epidermis, Micrococcus luteus, Enterococcus faecalis,* and *Streptococcus agalactiae* were not sensitive to PGRP killing. The spectrum of PGRP activity corresponded to the localisation of PGRP and thus indicated that PGRPs may play a role in the existence of normal flora. The bactericidal activity of

PGRPs is dependent on interaction with peptidoglycan rather than membrane permeabilization (Lu et al, 2006).

All mammalian PGRPs bind peptidoglycan and have a direct antibacterial function. In 2006 human PGRPs were declared a new class of bactericidal proteins (Lu et al, 2006). They are significantly different to other known AMPs, differing in structure, expression and known mechanism of action. They are of the largest known antimicrobial peptides, indeed, with the smallest (PGRP-S) being a 24kDa monomer, they are proteins rather than peptides. Vertebrate AMPs usually kill by traversing the bacterial cell wall and membrane permeabilization. peptidoglycan-lytic enzymes destroy the integrity of peptidoglycan and thereby induce osmotic lysis. Cell wall targeting antibiotics inhibit peptidoglycan synthesis. The kinetics of PGRP bacterial killing resembles that of the cell wall targeting antibiotic Penicillin, and unlike AMPs, do not target the membrane. There are two proposed mechanisms for the antimicrobial activity of PGRPs. As they strongly bind peptidoglycan, they could prevent cell wall synthesis. The second mechanism is killing by enzymatic digestion of peptidoglycan. This, however, is unlikely as only PGRP-L has amidase activity. (Lu et al, 2006). The exact mechanism of antimicrobial activity of these proteins is thus still largely unknown.

Lu et al also demonstrated the importance of glycosylation and divalent cations for PGRP function. N-glycosylase treatment completely negated PGRP function. They also proposed that PGRPs previously described as only bacteriostatic were only behaved as such because they were purified without calcium ions (Lu et al, 2006). All four PGRPs required zinc ions for bactericidal activity. Calcium can partially replace zinc, but only for the killing of Gram positive bacteria (Wang et al, 2007).

8.4 Tag 7

In 1998 a novel murine cytokine, Tag7, was cloned. The soluble form of the protein triggered apoptosis in murine L929 lines. The protein was assigned an immune function as it was constitutively expressed in lymphoid and haemopoietic tissues, but the authors could not find any significant homologous proteins in the databases of the time (Kiselev et al, 1998). A murine Tag7-like protein (GenBank accession no: AF193843) was fond to have 99.6% identity to murine PGRP (GB: AF0768421) and 99.3% identity to murine Tag7 (GB: X86374), suggesting that Tag7 may be a form of PGRP (Rehman et al, 2001).

It was reported in 1998 that shifts in the reading frame of murine Tag7 made it identical to murine PGRP. Tag7 is therefore a short PGRP (Kang et al, 1998). Tag7 has since demonstrated interesting properties not seen before in short PGRPs. Murine PGRP is intriguingly different to that of humans. As reported by Rehman the molecule they termed Tag7/PGRPS, as it coded identical transcripts, was involved in sleep regulation in rats. Sleep deprived rats had significantly increased mRNA levels of Tag7/PGRPS. (Rehman et al, 2001). In 2008 it was demonstrated that murine PGRP-S was upregulated by NF-κB in a relA dependent manner (Lang et al, 2008). The original purpose of the study was to determine the role of PGRP-S in cerebral ischemia, but the most finding of the study was the first report of a regulation mechanism of a mammalian PGRP. These observations have only been made in murine models, and again it would be wise to exercise caution before extrapolating these findings to humans.

Tag7 also forms a stable cytotoxic complex with Heat Shock Protein 70. It was suggested that PGRPs may play a role in anticancer defence (Saschenko et al, 2004). This was confirmed by a study demonstrating that transfection with *Tag7* cDNA decreased oncogenicity of mammary gland adenocarcinoma in immunodeficient animals (Larin et al, 2004). The last reported results on Tag7 trials stated that phase

I/II trials of gene therapy on human patients with disseminated melanoma and metastatic renal cell carcinoma was underway (Moiseyenko et al, 2005).

8.5 Potential clinical implications for PGRP-S

Thus far, the study of PGRP-S, indeed of the entire family, has been limited to examining its' biological role and functioning. Very little has been published about human PGRPs since the two landmark discoveries marking them as antimicrobial proteins (Lu et al, 2006) with a requirement for cationic molecules to function (Wang et al, 2007). PGRP-Iα and β have been implicated in psoriasis biology (Sun et al, 2006) but to date clinical data about PGRPs has been sparse. The first official clinical study involving PGRP-S was published this year. Rohatgi et al reported that an increased serum level of PGRP-S was associated with older age, and diabetic, hypertensive, actively smoking, and hypercholestrolemic caucasians. These patients also had a history of myocardial infarctions and generally had a higher body mass index (Rohatgi et al, 2008). These are all intriguing factors, but the authors appear to ignore the fact that PGRP-S is an antimicrobial protein. They do not mention anything about the health of their patients with reference to potential infections that may account for increased serum levels of the protein at the time of sampling. They also describe PGRP-S as a PGN scavenger and receptor, according to the work of Cho et al, 2005 and Liu et al in 2000. The landmark work of Liu et al of 2006 describing the molecule as an antimicrobial protein is not cited at all. The results of their findings are noteworthy nonetheless and determining the nature of the relationship of PGRP-S to arthrosclerosis is a possible area of study, particularly as initial northern blots determining the tissue expression of PGRP mRNA did not show PGRP-S to be expressed in the heart (Liu et al, 2001).

9. Aims and Objectives

There are clearly many unanswered questions about PGRPs. As antimicrobial proteins they have potential for future therapeutics. However, their mechanisms of action are poorly understood. Have all their ligands been elucidated? Do they interact with other proteins? How are PGRPs distributed in the population? The aim of this study is to characterise the genetics and biochemistry of PGRP-S by asking basic questions. What genetic variation exists in the PGRP-S gene of the South African population? Does genetic variation lead to different forms of the protein? The biochemical analysis aims to determine the nature of the interaction of PGRP-S with other proteins and to characterise its' antimicrobial effect.

CHAPTER TWO

MATERIALS AND METHODS

2.1 INTRODUCTION

This project involved the analysis of materials obtained from human subjects. This material was gathered according the principles set up by the Ethics committee of the University of the Witwatersrand, ethics clearance number R34/39. All patients were fully informed of what they would be subjected to, both by a formal letter and by trained staff of the Johannesburg General hospital. An example of the ethical consent form can be found in Appendix A. Three groups of patients were analysed; HIV positive (defined as patients at on antiretroviral therapy), HIV and TB positive (patients on antiretroviral therapy with an active tuberculosis infection) and healthy patients not on antiretroviral therapy. The aim of the groupings was to determine whether the immunocompromised population had an increased amount of variation in the PGRP-S gene. The subjects constituted a sample of the appropriate group at the hospital. The patients' family history as well as a brief medical history was collected where possible.

2.2 METHODS

2.2.1 NUCLEIC ACID ANALYSIS

2.2.1.1 RNA Extraction

Fresh blood was collected from patients of the Anti-Retroviral clinic at the Johannesburg General Hospital. Control samples were obtained from healthy volunteers from the University of the Witwatersrand. Once ethical consent was

obtained from the individual, blood was drawn into 5ml EDTA tubes. The blood was processed within 8 hours of collection to ensure the integrity of the RNA.

Total RNA was extracted from whole blood using the QIAamp RNA Blood Mini kit according to manufacturer's specifications (Qiagen) as detailed in the appendix table RNA was quantified by using a Nanodrop® spectrometer. The integrity of RNA to be used in subsequent experiments was determined by formaldehyde agarose gel electrophoresis as described by Sambrook et al (1989).

2.2.1.2 DNA Extraction

Genomic DNA was extracted from whole blood using QIAamp DNA Blood Midi kit according to manufacturer's protocol (Qiagen, South Africa), as detailed in the appendix.

2.2.1.3 Amplification of coding regions of PGRP-S

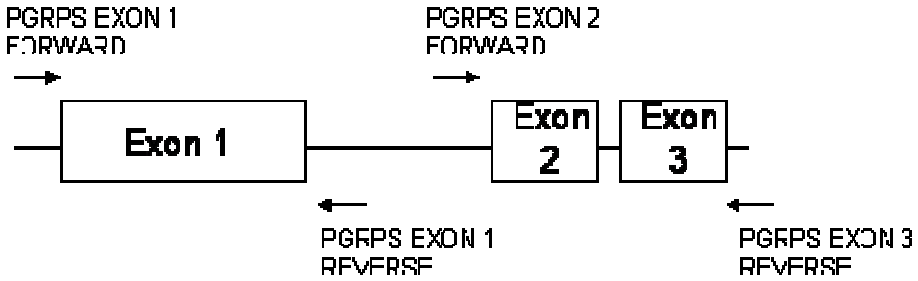

Figure 3: Schematic diagram of the amplification of PGRP-S.

The primers for the amplification of Exon 1 were designed manually to bind 50 base pairs outside the coding region to yield a 400bp product. Due to the 60bp intron separating Exon 2 and 3, the two regions were amplified as a single product with primers binding outside the coding regions for Exon 2 and 3. The coding regions of PGRP-S were amplified using the polymerase chain reaction using the Perkin Elmer GeneAmp PCR system 2400.

Exon 1 was amplified using the following primers:
PGRP-S Exon1 Forward: 5′-GGCGCTCCTAGCGGTCTC-3′
PGRP-S Exon 1 Reverse: 5′- ACTTGCACACCTTTGCCTCT-3′.
The primers were synthesised by Inqaba Biotec. The reaction mixture for Exon 1 contained 0.5U of Taq polymerase (New England Biolabs), 10mM dNTP (Inqaba Biotech), 1.2mM $MgCl_2$, 1µM of forward and reverse primer, 50-500ng of template in a final volume of 50µl.

The amplification consisted of a 5 minutes pre-denaturation at 94^{o}C, followed by thirty five cycles of 94^{o}C for 30 seconds, 60^{o}C for 30 seconds and 72^{o}C for thirty seconds. This was followed by a final extension at 72^{o}C for 7 minutes.

Exon 2 was amplified using the following primers:
PGRP-S Exon 2 Forward: 5'- TGGGAAATTAACGGGTGAGA -3'
PGRP-S Exon 3 Reverse: 5'- TGAGGAACACATTTTGAGCTACAT -3'.
The reaction mixture for Exon 2 contained 0.5U of Taq polymerase (New England Biolabs), 10mM dNTP (Inqaba Biotech), 1.0mM $MgCl_2$, 1µM of forward and reverse primer, 50-500ng of template in a final volume of 50µl.
The amplification consisted of a 5 minutes pre-denaturation at 94^{o}C, followed by thirty five cycles of 94^{o}C for 30 seconds, 54^{o}C for 30 seconds and 72^{o}C for thirty seconds. This was followed by a final extension at 72^{o}C for 7 minutes.

2.2.1.4 Cleanup of PCR amplicons

A 1% agarose gel was prepared in TAE buffer containing ethidium bromide (0.05mg/ml). DNA electrophoresis loading buffer was added to an aliquot of PCR amplicons and electrophoresed for 60 minutes at 100V. The amplicons were sized using the GeneRuler™ 100bp plus ladder as a molecular weight marker. Amplicons that could be clearly visualised when 5µl was electrophoresed were used for subsequent sequencing reactions. Successfully amplified products were

electrophoresed on a 1% agarose gel without ethidium bromide. The electrophoretogram was washed in a 1/10 000X solution of SyBr Gold (Invitrogen). The gel was viewed on Dark viewer (Clare Chemical Research) and the bands excised. This prevented any potential muatation inductions by the intercalation of ethidium bromide and any ultra violet induced damage. The fragment was cleaned up using the GenElute™ Gel extraction kit (Sigma).

2.2.1.5 Sequencing of PCR amplicons

PCR amplicons were prepared for sequencing using the BigDye® Primer Cycle sequencing kit, and sequenced on an ABI-Prism® 3100 (Applied Biosystems) at the HIV Unit of the National Institute for Communicable Diseases (NICD) using the standard cycle sequencing programme recommended for the BigDye® system. The primers used for the amplification of the product were used in the sequencing reaction. Each amplicon underwent two sequencing reactions, one with the forward primer to produce the forward sequence and one with the reverse primer to produce the reverse sequence.

2.2.1.6 Bioinformatics

For all samples, a forward and reverse sequencing reaction was performed and as such, each sample had two sequencing reactions per exon. At the terminal ends of each sequencing reactions, the data is often not useful as the signal is too weak and noisy to obtain the strong, clear peaks required for sequence determination, especially when the sequences are required for the determination of polymorphisms. Assembling the forward and reverse sequences allows the creation of a consensus sequence for the area of interest. Creating a consensus is important as it allows for data to be obtained for the entire area of interest. By aligning the forward and reverse sequences, the terminal ends of the reaction, which often have unreadable data can be supplemented by the complementary sequence. Sequences were assembled using

Sequencher Version 4.5 for Windows. The Basic Local Alignment search tool BLAST (www.ncbi.nlm.nih.gov/BLAST) was used to determine whether the sequences aligned significantly with any other sequences besides the target sequences. DNA alignments and translations, as well as Protein alignments were performed using DNAman version 4.03 and ClustalW (Thompson et al, 1994). SNPs were verified by comparison to SNPs recorded at dbSNP (Sherry et al, 2001), Ensembl (Hubbard et al, 2007), the UCSC genome browser (Kent et al, 2002), SNPer (Riva and Kohane, 2004) and the Japanese SNP database jSNP (Hirakawa et al, 2002) Secondary structure predictions were done using the PsiPred secondary structure prediction server (Bryson et al, 2005).All URLs are provided in the appendix

2.2.2 HETEROLOGOUS EXPRESSION AND INTERACTION STUDIES

2.2.2.1 cDNA synthesis

Reverse Transcriptase PCR was performed using the Access RT-PCR system (Promega) with the reaction mix and conditions described in Table 2

Table 2: Reverse Transcription reaction mix

Reagent	**Final concentration**
AMV/*Tfl* 5X reaction buffer	1X
dNTP mix (10mM each dNTP)	0.2mM
Upstream Primer (PGRP-S forward)	1µM
Downstream Primer (PGRP-S reverse)	1µM
25mM $MgSO_4$	1mM
AMV reverse transcriptase (5u/µl)	0.1u/µl
Tfl DNA polymerase (5u/µl)	0.1u/µl
RNA sample	5µg
Nuclease free water	To a final volume of 50µl

The cycling conditions were as follows: 1 cycle of 48^{o}C for 45 minutes, 1 cycle of 94^{o}C for 2 minutes, and thirty cycles of 94^{o}C for 30 seconds, 57^{o}C for sixty seconds and 57^{o}C for 68^{o}C for 2 minutes. A final extension cycle at 68^{o}C was performed for 2 minutes.

2.2.2.2 HETEROLOGOUS EXPRESSION OF PGRP-S

2.2.2.2.1 Preparation of Competent cells

An overnight *E. coli* BL21 (DE3) or *E. coli* XL1-Blue culture was diluted (1:20) into LB broth containing no antibiotics. Diluted cultures were then grown at 37 ˚C until the absorbance at 600 nm reached 0.3-0.6 units (early log phase). The cells were harvested by centrifugation at 4 500 x g for 15 minutes at 4ºC in sterile centrifuge tubes. The cells were kept on ice at all times for the duration of the preparation. The cells were then resuspended in ice-cold 0.1 M $MgCl_2$ and left on ice for a further 20 minutes. The cells were then collected as above by centrifugation for 15 minutes and resuspended in ice-cold 0.1M $CaCl_2$ and left for 1 hour. Cells were collected and gently resuspended in ice-cold 0.1M $CaCl_2$. The cells were stored in 1 volume of ice-cold sterile glycerol (30 %) at –70 ˚C.

2.2.2.2.2 Preparation of pGEM-T/PGRPS vector

Ligation into pGEM-T easy vector (Promega) was according to the manufacturer's instructions. The ligation cocktail consisted of 2 μg of the PGRP-S cDNA fragment, 50 ng plasmid and 3.5 units of T4 DNA ligase. The T4 DNA ligase was at ratio of 1unit/μg of DNA. The reactions were incubated overnight at 4^{o}C. pGEM-T easy plasmid without an insert was used as a ligation control.

2.2.2.2.3 Preparation of pET41/PGRPS vector

pET41 (a) was linearized by restriction digestion with *EcoR1* and *XhoI*. Calf intestinal phosphatase (CIP) was used to treat linear plasmid molecules at a ratio of 0.5 units per μg DNA in a manufacturer recommended buffer. The reaction was incubated at 37^{o}C for 30 minutes. The tube was heated at 65^{o}C for 15 minutes and was followed by phenol chloroform isoamyl (25:24:1) extraction to stop the reaction as outlined in section 2.2.2.5. The linearization of the plasmid was confirmed by agarose gel electrophoresis

2.2.2.2.4 Transformation of competent cells

An aliquot of competent cells was transferred into a sterile microcentrifuge tube containing 100 ng of plasmid on ice. The tube was flicked gently to mix the contents and placed on ice for 20 minutes. The cells were heat shocked at 42˚C for 45 seconds and immediately returned on ice for a further 2 minutes. Luria Bertani broth was added to the cells and incubated for 2 hours at 37 ˚C with shaking. Cells were plated onto LB agar plate containing appropriate antibiotics; ampicillin (100μg/ml) for XL1-Blue cells and kanamycin for BL21 (DE3)(pLysS) cells. pGEM-T easy colonies were selected on the basis of blue/white screening, which was confirmed by restriction digestion of plasmids from white colonies to confirm the presence of the insert. Similarly, plasmids from colonies of pET-41 transformed *E. coli* BL21 cells were tested for the presence of insert with restriction digestion.

2.2.2.2.5 Small-scale plasmid DNA preparation of pGEMT/PGRPS; pET-41/PGRP-S

A sample of an overnight culture of *E. coli* XL1-Blue transformed with pGEM-T Easy/PGRPS or BL21 (DE3) (pLysS) transformed with pET-41/PGRP-S was centrifuged at 16 000 x g for 1 minute, and the cell pellet resuspended in 100 μl of

Alkaline lysis solution I. Solution II was then added and the suspension incubated on ice. After 2 minutes, 150 µl of solution III was added and the suspension was mixed gently by inversion and incubated on ice for a further 5 minutes. The sample was extracted twice, first with phenol: chloroform: isoamyl (25:24:1) followed by chloroform. The aqueous layer was transferred into a fresh microcentrifuge tube, followed by DNA precipitation with ethanol acetate. The mixture was incubated for 30 minutes at –70^{o}C. The DNA was pelleted by centrifugation at 12, 000xg for 10 minutes, washed with 70% ethanol and air dried at room temperature for 10 minutes. DNA was reconstituted with appropriate volume of sterile distilled water.

2.2.2.2.6 Overexpression and Purification of the GST-tagged fusion protein

The optimal inducer concentration and induction time was determined by an induction study to determine the optimal conditions for protein induction.
A single colony of *E. coli* BL21 (DE3) pLysS transformed with the PGRP-S/pET-41 construct was inoculated into LB broth and grown for 16 hours. This culture was diluted 1: 25 and grown to an optical density at 600nm of 0.6 (mid-log phase). The culture was induced with 2mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were allowed to grow for 16 hours before being harvested by centrifugation at 5000xg for 10 minutes. The cell pellet was resuspended in 1M Cell lysis buffer containing 1mM aprotonin, leupeptin, phenyl methyl sulfonyl fluoride (PMSF) and EDTA. The cell suspension was sonicated for 30 seconds and insoluble protein was removed by centrifugation at 12 000rpm for 20 minutes. The supernatant was retained for purification of the Glutathione S-transferase (GST)-tagged protein.

Soluble protein extract isolated from *E. coli* BL21 (DE3) pLysS twice with 10-column volumes GST wash/bind buffer. The fusion protein was eluted from the matrix with 3x volumes of GST elution buffer. All collected fractions were electrophoresed on a 12% polyacrylamide gel according to the method of Laemmli (1970). The analysis of the gels were performed as described in section 2.2.2.2.8

2.2.2.2.7 Interaction Studies using Pull-Down Assay

The ability of the fusion protein to interact with soluble protein from *Drosophila melanogaster* was to be determined. The total fly protein fraction protein was prepared by homogenising 15 flies in 1.5ml of radioimmunoprecipation assay (RIPA) cell lysis buffer containing 1mM aproptinin, leupeptin, PMSF and EDTA. The homogenate was centrifuged for 30 minutes at 13 000xg and the supernatant retained for the interaction study. The success of the extraction was determined by SDS-PAGE analysis as described in section 2.2.2.2.8.

An overnight culture of *M. luteus* was harvested by centrifugation at 5000xg for 10 minutes. The pelleted cells were resuspended in 1M Cell Lysis buffer, and sonicated for 30 seconds. The insoluble protein was removed by centrifugation at 13 000xg for 10 minutes. The insoluble fraction containing the cell wall component and subsequently the peptidoglycan fraction was resuspended in phosphate buffered saline containing 1mM aproptinin, leupeptin, PMSF and EDTA and sonicated for 10 seconds. This was retained as a positive control for interaction with PGRP-S, as peptidoglycan is the natural ligand for PGRP. The soluble fraction was retained as negative control, as no known ligands for PGRPS should be found in this fraction.

E. coli BL21 (DE3) pLysS cells transformed with pET-41a without any insert were prepared as described in section 2.2.2.4. These transformed cells were then induced to overexpress as described in section 2.2.2.6 in this way GST was purified. This was used as a negative control.

PGRP-S/GST fusion protein was immobilised onto four separate GST matrix columns. Soluble fly protein was added to one GST matrix and allowed to interact with the immobilised fusion protein for 18 hours at 4^{o}C. Similarly, insoluble bacterial protein, soluble bacterial protein and pure GST were allowed to interact with the fusion protein. The column was washed with 8 volumes of GST-wash solution from

the GST Bind Kit® (Novagen). Protein complexes were eluted with 3 volumes of GST Elution solution from the GST bind kit® (Novagen).

2.2.2.2.8 SDS PAGE analysis

The sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) method described by Laemmli (1970) was used to separate protein for analysis. SDS-PAGE makers (Fermentas) were used to determine protein size. The gel was electrophoresed at 120V for 120 minutes, Stained in coomassie brilliant blue G (Sigma) staining solution and destained overnight. The electrophoretograms were visualised using a GS-800 calibrated densitometer (Biorad)

2.2.2.2.9 Radial diffusion Assays

An overnight culture of bacteria (*S. aureus*, *E. coli*, *M. luteus*) was inoculated into LB Broth and grown with shaking at 37^{o}C for 18 hours. Midlogarithmic phase organisms were obtained by re-inoculating the overnight culture in a 50 000X dilution of fresh LB broth and incubated at 37^{o}C for 3 hours. The bacteria were harvested by centrifugation at 900xg at 4^{o}C for 10 minutes. The cells were washed with cold 10mM Sodium phosphate buffer pH 7.4. The cells were resuspended in three different types of 10mM Sodium phosphate buffer pH 7.4; Sodium phosphate without a salt supplement, phosphate buffer with 5mM $CaCl_2$ and phosphate buffer with 10µM $ZnCl_2$. The optical density was measured at 620nm and used to calculate a volume of cells to use in order to obtain a volume containing 4 x 10^6 colony forming units and this was added to an underlay agar of LB broth, 1% agarose and 0.02% Tween 20. This mixture was used to prepare a pour plate in which holes were punched and 1µg/µl fusion protein was added. As a positive control 12µg/ml tetracycline was used and as a negative control the buffer in which the fusion protein was suspended. This was incubated at 37^{o}C for three hours. An overlay agar layer

consisting of LB broth and 1% agarose was poured over this preparation after the incubation and the plates were incubated for 18 hours. The clearings were visualised after staining with Coomasie brilliant blue R-250 for 18 hours. The plates were destained with an aqueous solution of 10% acetic acid and 2% dimethylsulphoxide for ten minutes.

CHAPTER THREE

RESULTS

3.1 Genetic Analysis

3.1.1 Population distribution of experimental samples

Single nucleotide polymorphisms in the PGRP domain of PGRPs of *Drosophila*, particularly the short forms, can result in the immunocompromise of the mutant flies (Lazzaro et al, 2004). The aim of the genetic analysis portion of this study was to determine whether increased levels of SNPs could be associated with the HIV and HIV/TB positive population. The HIV positive population constitutes a pool of immunocompromised humans to screen. Tuberculosis is a disease commonly associated with HIV positive patients. *Mycobacterium tuberculosis* is an acid fast Gram positive bacterium (Madigan et al, 1997) and should be recognised by PGRP-S, it may hint at dysfunctional PGRPs.

The establishment of databases detailing variation within the human genome has been vitally important since the release of the first draft genome. The establishment of a "reference genome" is a difficult and contentious task. For example, in the sequence of only one individual, James Watson, two million polymorphisms, most of which have already been noted (Egholm, 2007). It is because of the variant nature of genomes that several variation databases exist. The largest and most referenced is dbSNP at NCBI (Sherry et al, 2001). This database has 9 million variations noted to date (Egholm, 2007). Other databases include Ensembl (Hubbard et al, 2007), the Japanese SNP database jSNP (Hirakawa et al, 2002), the UCSC genome browser (Kent et al, 2002), and SNPer (Riva and Koane, 2004). Many of these databases reference dbSNP.

Sequences for these populations commonly come from either cell lines, predominantly sourced from Caucasians, or from several specific population groups. These include the Han Chinese from Beijing (HCB), Japanese from Tokyo (JPT), Utah residents with West European ancestry (CEU), and Yoruba from Ibadan in Nigeria (YRI) (Hubbard et al, 2007). The YRI population and several sequenced African American individuals represent the data from Africa. The genetic information from sub-Saharan Africa is therefore grossly underrepresented. This study therefore presents the first genetic survey of population variance of the PGRP-S gene in individuals from the region. The study groups were divided into groups of healthy controls, HIV patients on antiretroviral treatment (HIV) group, and HIV patients on antiretroviral treatment and Tuberculosis medication (HIV/TB). The population distribution of these three groups are represented in Table 3. The populations are based on language groups, as genetic homogeneity cannot be ensured. As no South African populations were isolated, and intermarriage between different language groups occurred.

Table 3: Population distribution of experimental groups

Sample Group	Population	Number of Inviduals
Healthy Controls	Zulu	7
	Sotho	3
	Xhosa	4
	Other	2
HIV positive	Zulu	6
	Sotho	6
	Tswana	2
	Tsonga	1
	Xhosa	1
	Other	1
HIV/TB positive	Zulu	7
	Sotho	5
	Ndebele	1
	Tsonga	2
	Venda	1
	Other	1

3.1.2 Amplification of coding regions

This study is a pilot study on variation in the Southern African Sub-Saharan African population and as such did not consist of a large survey group. The Yorubi of Nigeria is only other group in Africa examined for variation in this gene. Furthermore, only the coding regions were amplified because the study by Lazzaro showed the significant variation was variation within the regions coding for the PGRP domain (Lazzaro et al, 2004). Variations in the introns were therefore not considered. The introns were only included in the study to a small extent as the forward primer was designed to bind upstream of the 5' untranslated region and the reverse downstream of the end of the 3' region of each exon. In that way the full length of the coding

region would be amplified and the amount of ambiguous sequencing data at the beginning and end of the sequencing reaction would be minimised.

As Exon 2 and 3 were separated by 60 base pairs, both regions were amplified as a single product, referred to as Exon 2. The amplification of Exon 1 of patients 1-14 is depicted in figure 7. The amplification of Exon 2 of patients 15 to 19 is depicted in figure 4. Amplicons of both Exon 1 and 2 were obtained for 48 patients. The remaining two patients' gene failed to amplify.

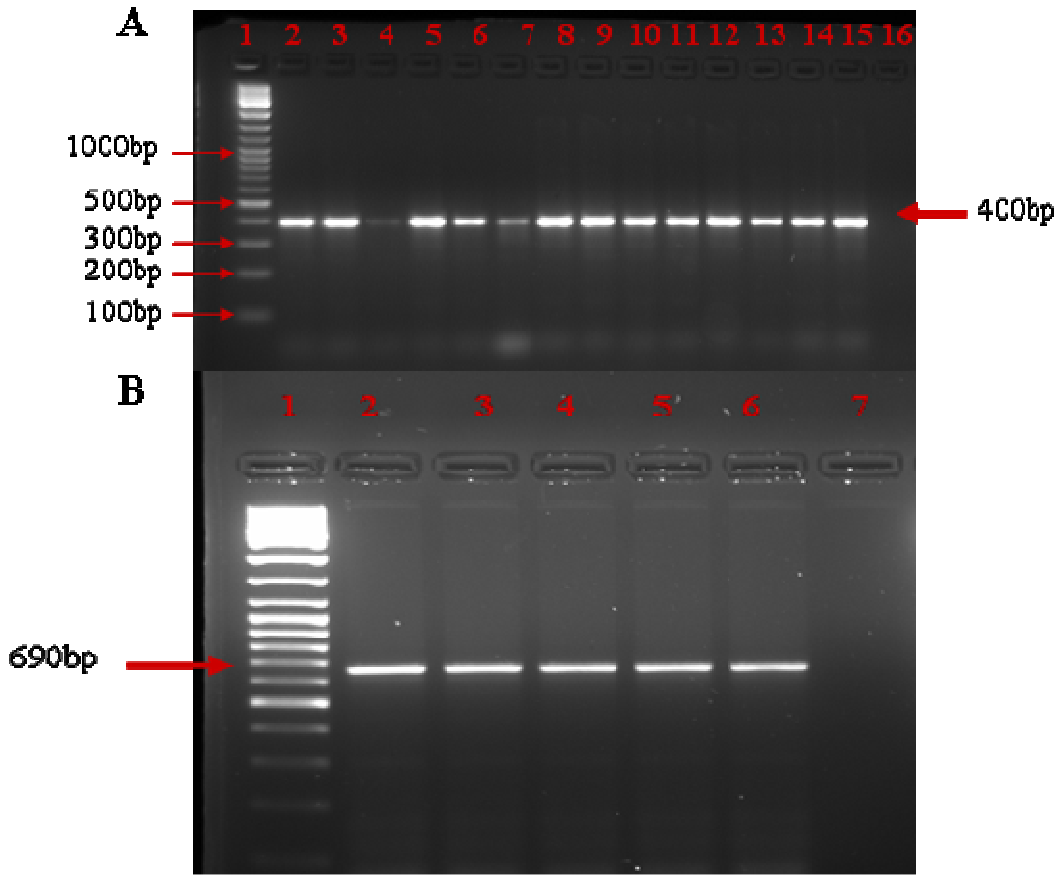

Figure 4: Routine Amplification of PGRP-S. Exon 1 in Block A and Exon 2 in Block B. A 1% electrophoretogram demonstrating the 400 base pair (bp) Exon 1 amplicon and 600bp Exon 2 amplicon. These amplicons were subsequently used in sequencing reactions. Lanes 2-15 of block A represent samples of the first exons

from Patients 1 to 14 respectively. A negative control using *Drosophila* DNA as a template is represented in lane 16. Lanes 2-4 of Block B represent the second exon of Patients 15 to 19. A negative control using *Drosophila* DNA as template is represented in lane 7. In both diagrams the Fermentas 100bp plus DNA ladder is used as a molecular weight marker.

3.1.3 Sequencing of PGRP-S

The amplicons of the patients in the study were electrophoresed and gel purified before automated DNA sequencing. The BigDye® terminator system was used for the direct sequencing of the amplicons, as detailed in section 2.2.1.5. To ensure that mutations were not introduced by exposure of the agarose gel to ultraviolet light, the intercalating agent Ethidium bromide was replaced with SyBr Gold (Invitrogen), which required blue light and therefore would not induce mutations.

3.1.3.1 Confirmation of sequencing data

Two assemblies exist for the genomic sequence of Chromosome 19. NT 011109.15 is referred to as the reference sequence and NW 927217.1 as the alternate assembly. The mRNA transcript, AF076483, originally described by 2001, is referred to as the reference mRNA sequence (Liu et al, 2001). In this project, all primers were designed using this sequence.

To determine whether the obtained sequences were PGRP-S, the Basic Local Alignment Sequence Tool (BLAST) was used to determine whether the sequences obtained aligned with the reference sequence at the NCBI database. Aligning the sequences obtained resulted in two hits in the genome, both on chromosome 19. All the sequences matched the sequence for PGRP-S on both the reference and alternate assembly of the chromosome only, confirming that the sequences obtained were PGRP-S.

3.1.3.2 Determination of Single Nucleotide Polymorphisms

In this study, only the coding regions were examined as the primary aim was to determine whether novel protein variants of PGRP-S existed. As described in section 2.2.1.3, non-coding regions of the gene were included in the amplification and subsequent sequencing of the gene. This was to ensure that the area of interest would be fall well into the sequencing reaction and thus avoiding the poor quality sequence data associated with the terminal ends of the reaction. To obtain a full sequence of transcript from the sequences obtained for each sample both Exon 1 and 2 for each reactions were aligned to the mRNA. In this manner, all non-coding regions could be trimmed off and as a result, only the sequence for the transcript would remain.

Table 4: Summary of known non-coding variation in the PGRP-S gene from the dBSNP, Ensembl, HapMap, jSNP, SNPer and UCSC genome browser databases

SNPid	reference contig position	SNP	Ancestral allele	function	population frequency data
rs61513108	8793133	A/G	unknown	intron	none
rs61497458	18792085	A/G	unknown	intron	none
rs61403305	18794085	G/T	unknown	intron	none
rs60612918	18792689	C/T	unknown	intron	none
rs58866037	18791600..18791601	-/TT	unknown	intron	none
rs58829899	18793324	A/G	unknown	intron	none
rs58401596	18788946	C/G	unknown	intron	none
rs57714530	18792514	A/T	unknown	intron	none
rs11477830	18791591	-/T	unknown	intron	none
rs8106183	18792825	A/G	A	intron	G allele freq 0.0034 in YRI
rs3786822	18793324	A/G	G	intron	none
rs2302788	18789744	A/G	unknown	intron	HAP map allele
rs2072561	18794085	G/T	T	intron	HAP map allele
rs2041993	18793185	A/G	G	intron	HAP map allele
rs2041992	18792689	C/T	C	intron	HAP map allele
rs2041991	18792514	A/T	A	intron	Asian/Caucasian data
rs12461346	18789290	A/G	unknown	intron	none
rs11669048	18788946	C/G	unknown	intron	none
rs11083793	18789163	A/T	unknown	intron	none
rs57027496	18794866	A/G	unknown	5′ near gene	none
rs55645419	18794693	A/C	unknown	5′ near gene	none

SNPid	reference contig position	SNP	Ancestral allele	function	population frequency data
rs11880830	18794626	C/T	C	5' near gene	No T allele found in samples of CEU, HCB, JPT, YRI
rs7245826	18794744	A/G	unknown	5' near gene	none
rs7245473	18794671	C/T	unknown	5' near gene	none
rs2341638	18794777	C/T	unknown	5' near gene	none
rs2072563	18794866	A/G	unknown	5' near gene	HAP map allele
rs59433803	18794671	C/T	unknown	5' near gene	none
rs61742389	18788724	C/G	unknown	3' near gene	none
rs35952458	18790254	-/C	unknown	3' near gene	none
rs2072562	18794516	G/T	G	5'UTR	none
rs60974605	18794516	G/T	unknown	mRNA (26)	none
rs1054882	18790024	C/T	unknown	3' untranslated	none
rs1054881	18789994	C/T	unknown	3' untranslated	CEPH

The cells highlighted in yellow represent a duplication in the dbSNP database, where the same polymorphism was submitted twice. In the above table, as well as the table below, previously recorded polymorphisms found in this study are highlighted in green.

Table 5: Summary of known variation within the coding region of PGRP-S from the dBSNP, Ensembl, HapMap, jSNP, SNPer and UCSC genome browser databases

SNPid	reference contig position	SNP	Ancestral allele	function	population frequency data
rs59269540	18789565	C/T	unknown	synonmous	none
rs34478490	18794265^18794266	-/CG	unknown	mRNA (227)	
rs34180629	18794397	A/C	unknown	missense, mRNA(145)G-V	AA population: C allele 0.026
rs28722714	18794222	A/G	G	synonmous D(92)D	AA population: A allele 0.038
rs13343537	18791051	G/T	G	missense, mRNA(404)H-Q	No T allele found in samples of CEU, HCB, JPT, YRI
rs34323192	18789815	-/C	C	Frameshift	none

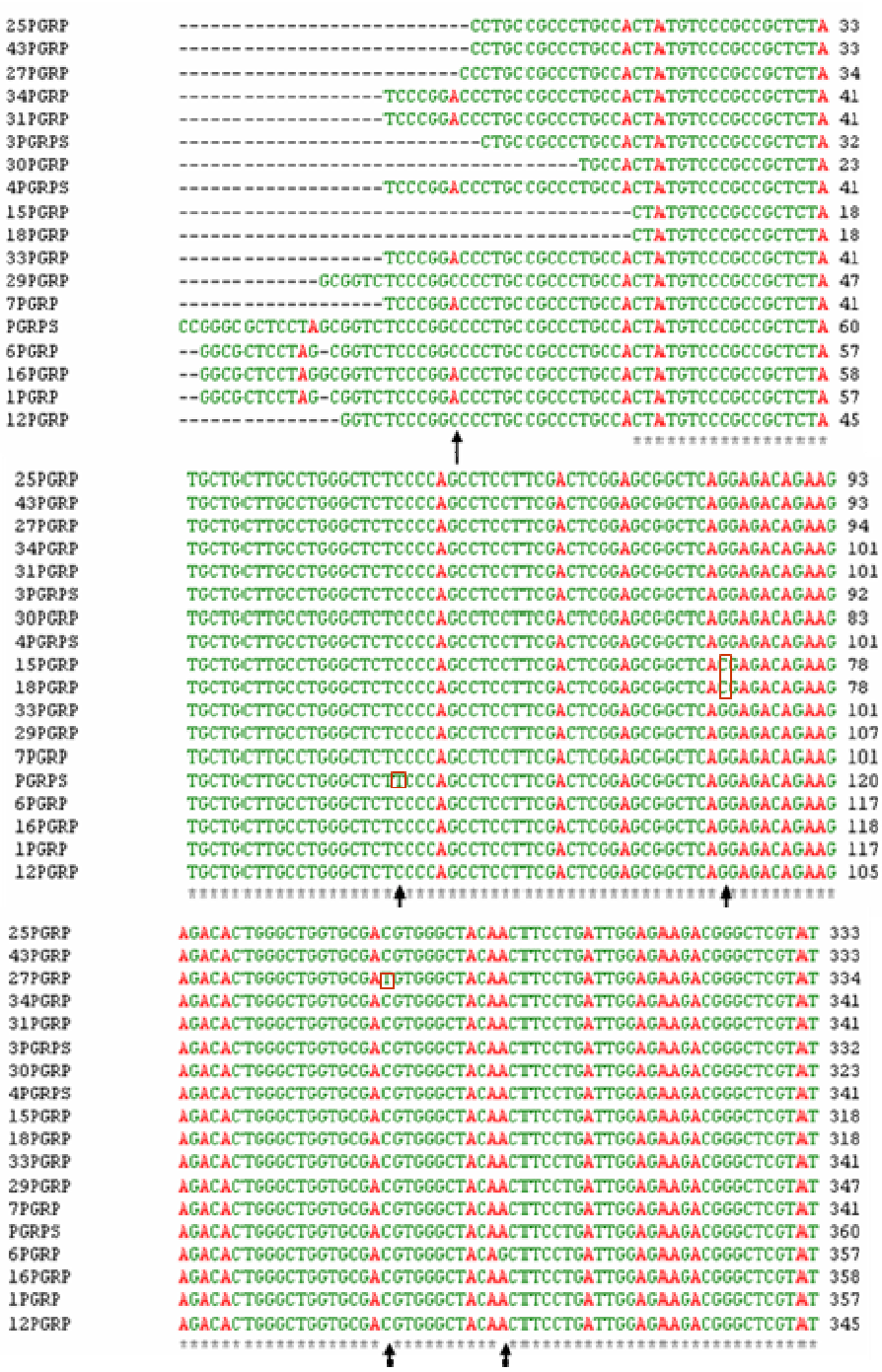

Figure 5: Truncated Clustal W multiple alignment of samples highlighting the variations observed by direct sequencing.

In the figure above, PGRPS represents the reference mRNA sequence, and consensus transcript sequences assembled as described in section 3.1.3.2. The number prefixing the term refers to sample number

Every sequenced sample differed from the mRNA transcript at position 80. The allele in the transcript is a T while all the genomic sequences were homozygous for a C at that position. The substitution does not change the amino acid coded for, as both forms of the codon in question, CCC and CCT codes for a proline residue. Nonetheless, it is striking that all the genomic sequences differ from the mRNA transcript. It is not recorded as a DNA polymorphism and all PGRP-S reference genomic sequences examined (GenBank, the UCSC genome browser and Ensembl) have a C at that particular position. In mRNA sequences from all of the above databases, a T allele is found in this position.

Further examination of alignments of the sequenced samples showed that the individuals sequenced did not show significant variation from the standard mRNA sequence. Only 5 variations were found with 2 polymorphisms previously observed and 3being potentially novel. The polymorphisms are highlighted in the Clustal W alignment presented in Figure 5. The genotypes of all the polymorphisms were determined by direct examination of the chromatograms, are summarized in table 6).

The first polymorphism observed corresponds to mRNA position 26. This has been previously observed, and is referred to as rs2072562. The polymorphism lies in the 5' untranslated region and therefore has no bearing on the gene sequence of the gene product. In this study group the C-allele occurs with a higher frequency (0.53) than the A-allele (0.47). NCBI states that the C-allele is the ancestral allele. At present there is no recorded frequency data for this polymorphism at either NCBI or Ensembl.

Another previously described polymorphism, a C/T substation at mRNA position 320 called rs28722714, occurs in 6 patients, 3 HIV/TB patients, 2 HIV patients and 1 healthy control. All six these patients are heterozygous for the polymorphism as observed in figure 6.

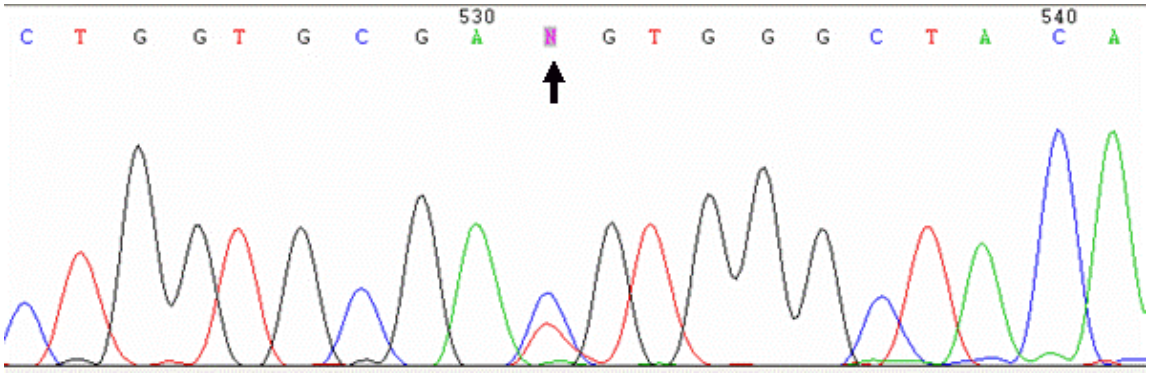

Figure 6: Partial sequence analysis of Exon 1 of Patient 27. Patient 27 is has a polymorphic site at position 320 of the mRNA. The patient is heterozygous for C and a T allele at this position.

Both codons encode Aspartic acid. This previously noted polymorphism is observed in 6 patients, 3 HIV/TB patients, 2 HIV patients and 1 control. The calculated frequency of heterozygosity listed at NCBI is 0.074 for the African American population, and therefore much lower than the frequency of 0.3 found in this sample population group. However, this is likely to be a function of the small sample size. No TT homozygotes were observed in the sample group and this is echoed by NCBI, which also does not record any TT homozygotes in the population sampled.

A variation at position 651 was observed in sample 12. This was not observed in any other patients and examination of the chromatogram showed that this C/T (Y) ambiguity may be more a function of poor sequencing and can only be confirmed by further rounds of sequencing.

Sample 15 contained an A-allele at position 110 instead of a G. Examination of the chromatogram showed that the polymorphism (representing a synonymous

glutamine-glutamine substitution). Although not as neatly demonstrated as the polymorphism at 320, the R ambiguity does occur four times. Further samples and more sequencing can determine whether this polymorphism which has not been reported at NCBI, Ensembl, JSNP, SNPer or the UCSC genome browser is indeed a novel polymorphism.

PGRP-S contains what is described by Ensembl as residue splice junctions. Two residues, asparagine 75 (96 on the immature peptide) and aspartic acid 116 (136 on immature peptide) are coded for by codons which span Exon 1 and 2 and 2 and 3 respectively. The CAA codon for asparagines falls at the end of Exon 1 (CA) and the beginning of Exon 2 (A). A polymorphism is observed in sample 6 at mRNA position 312 where the patient is homozygous for G and not the middle A of the codon. This polymorphism is unobserved in any other sample and is not recorded at any of the databases used to confirm polymorphisms. If confirmed by further rounds of sequencing and an altered PCR strategy to be discussed later, not only does this represent a potentially novel polymorphism, but also a variation leading to an amino acid change that is previously unreported.

Table 6: Frequency polymorphic genotypes distributed among control, HIV positive and HIV/TB positive groups.

Genotype	**mRNA Position**	**Exon**	***n***	**Frequency**
CC	27	1	6	0.3
CA	27	1	13	0.65
AA	27	1	1	0.05
GG	110	1	16	0.8
GA	110	1	4	0.2
CC	320	1	12	0.6
CT	320	1	8	0.4
CC	651	3	18	0.9
CT	651	3	2	0.1

Of the polymorphisms examined, only rs2072562 had all both homozygotes (CC/AA) and therefore this was the only sample that could be examined for Hardy-Weinberg equilibrium.

As the frequency for the C-allele is 25/40 and the A-allele is 15/40, the Hardy-Weinberg equation would look as follows:

$p^2+2(pq)+q^2 = 1$
$(25/40)^2 +2(25/40)(15/40) + (15/40)^2$
= (0.625)2 + 2(0.625)(0.375) + (0.375)2

0.390625+ 0.46875+0.140625= 1.

Therefore, for this SNP, the alleles are in Hardy-Weinberg equilibrium. To test this, a Chi-squared test was performed, where the null hypothesis is that the alleles are in equilibrium.

$\chi 2 = \sum(\text{observed-expected})^2/\text{expected}$

$= [(\underline{6-7.8125})^2 + (\underline{13-9.3750})^2 + (\underline{1-2.8125})^2]$
$\quad 7.8125 \qquad 9.3750 \qquad 2.8125$

= 0.421/7.8125 + 1.41/9.3750 + 0.421/2.8125

= 2.25.

As the $\chi 2$ test has one degree of freedom (df) and a 5% confidence interval with 1df is 3.84, the alleles are in equilibrium.

3.2 Bioinformatics analysis

3.2.1 Protein alignment

The sequencing results demonstrated both previously reported and potentially novel polymorphisms. The discovery of previously noted polymorphisms give validity to the quality of the sequencing, and indicate that a further sampling may confirm the existence of the novel polymorphisms found. To determine whether the polymorphisms resulted in amino acid changes or frameshift mutations, the sequences for the coding regions were translated.

Patient 6 was the only patient with a non synonymous change as shown in Figure 24. To date, only three polymorphisms have been reported in the open reading frame of PGRP-S. Of those three reported polymorphisms (rs34180629, rs13343537 and rs28722714), only two were missense mutations, and the remaining SNP a synonymous one. As only these polymorphisms have been reported since the discovery of the gene, a new protein variant is therefore an interesting new avenue for exploration. Although the existence of variants of the protein has been reported, there is currently no published literature on the effect of variant amino acid residues on the structure and functionality, particularly with respect to the ability of the protein to bind peptidoglycan. Thus, as no body of literature exists on the subject, it is therefore very important to examine the nature of the variant protein to determine how differently it would behave compared to the wild type PGRP-S.

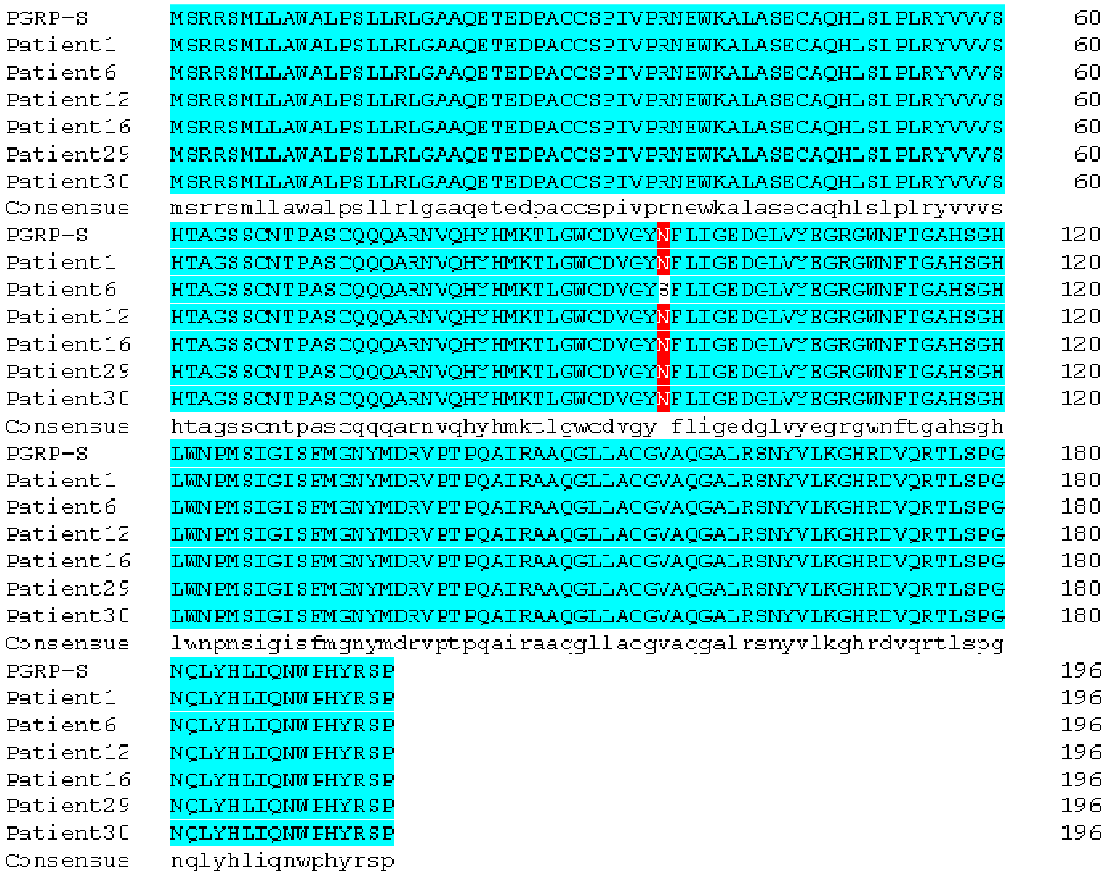

Figure 7: Multiple alignment of the translated mRNA sequences of the HIV/TB positive group with PGRP-S standard amino acid sequence

A non-synonymous change is found at amino acid position 75 of the mature peptide of patient 6. Asparagine is substituted with Serine. This is a conserved amino acid change in terms of charge, as both amino acids are neutral and polar, but the bulkiness of the side chain has been greatly reduced. Patient 6 is the only patient with a non-synonymous polymorphism.

3.2.2 Secondary structure predictions

3.2.2.1 Secondary structure predictions of PGRP-S and variant proteins

The variant protein of Patient 6 differs by one amino acid only. However, the amino acid is located in a secondary structural motif (a strand). The change could potentially also affect the protein's secondary structure. Secondary structure prediction was performed on the mutant protein using the PsiPred secondary structure prediction tool (Bryson et al, 2005; Jones et al, 1995). As the crystal structure for Wild-type PGRP-S has been elucidated, a secondary structure prediction on the wild type amino acid sequence has a twofold purpose. It can serve as a visual reference to compare the mutant secondary structures too and can serve as a form of quality control on the accuracy of the secondary structure prediction tool.

Figure 8 shows a comparison of the secondary structure prediction for the wild type PGRP-S protein, as well as a structure prediction for the variant protein of patient 6. According to the prediction, the single amino acid change results in large changes in the secondary strucure of the protein. The accuracy and validity of the prediction will be discussed further and in greater detail in section 4.5.

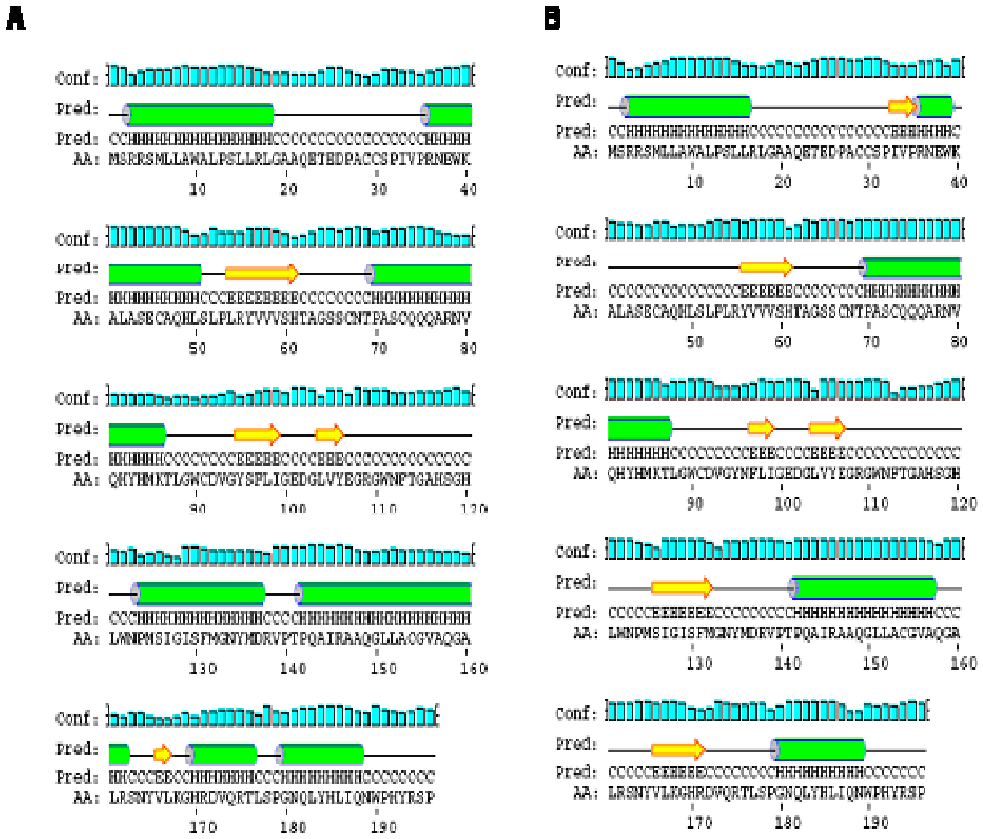

Figure 8: Secondary structure predictions of mutant samples from Patient 6(A) and Wild type PGRP-S (B) using the PsiPred Secondary structure prediction tool.

In the above representation the helices are represented by green barrels and strands by yellow area. A comparison of the predicted structure with the crystal structure deposited at the Protein Databank (1YCK) revealed that the prediction was not highly accurate in predicting the amount of secondary structural motifs. Of the structures correctly predicted, the lengths of the motifs were at most off target by two residues. The incorrect residues were always predicted with low confidence. The secondary structure prediction of Patient 6 (Structure A) differs from the prediction given for the wild type protein. This could possible be due to the fact that the amino acid change is at the beginning of a strand and could therefore have an effect on the folding dynamics. The accuracy of the prediction can be called in to question as it predicts

that a structure found in all PGRPs is missing in the mutant. The structure in question is demonstrated in Figure 9.

3.2.3.2 Conservation of structural motifs across PGRPs.

A secondary structure feature reported to be conserved in all PGRPs, both vertebrate and invertebrate, where the structure of the molecule has been elucidated is a conserved helix structure, preceded by a strand. Notably, while the helical structure is present in sample 6, the strand is missing. Both these structures are preserved in PGRPs across several species and in all in forms of the protein (short, intermediate or long), and therefore the absence of this motif is remarkable. Figure 10 highlights the conservation of this particular structural motif across species.

All secondary structure predictions performed for short PGRPs have this motif, and therefore the missing motif in sample 6 is noteworthy. This could either be a significant change due to the position of the amino acid in its particular structural motif, or a reflection of the nature of secondary structure prediction tools being unreliable. This will be discussed later in greater detail.

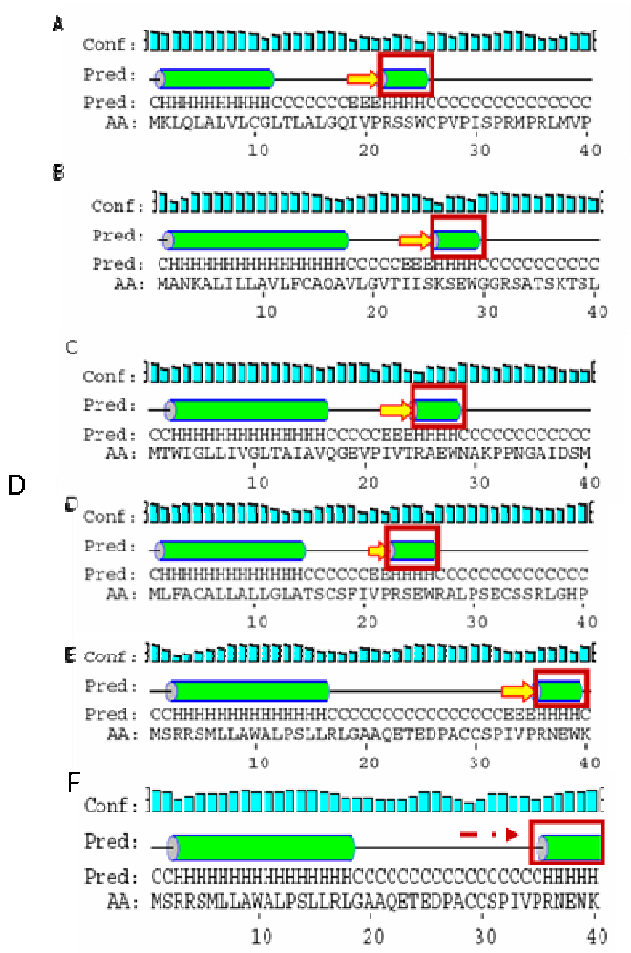

Figure 9: PsiPred predictions highlighting a conserved motif found in PGRP-S found in different organisms.

Block A represents PGRPSB2 from *Drosophila melanogaster*; Block B represents PGRP SC2, also from *Drosophila*. Block C represents PGRP SD from *Drosophila*. Block D represents the mouse cytokine Tag 7. Block E represents PGRP-S in humans. Block F represents sample 6. All secondary structure predictions done of short PGRPs have this motif, and therefore the missing motif in Patient 6 is noteworthy. This could either be a significant change, or a reflection of the nature of secondary structure prediction tools being unreliable. This could be confirmed not only by Circular dichroism spectroscopy or by in vivo studies to determine whether the amino acid substitution and the subsequent potential change results in a phenotype alteration.

3.2.3 Tertiary structure analysis

An amino acid change would not only potentially affect the secondary structure of the protein, but also the intramolecular contacts of the tertiary structure. As the tertiary structure of PGRP-S is recorded at the Protein Data Bank as 1YCK, the Swiss PDB viewer was used to mutate the wild type asparagine (N) to the serine residue of the mutant protein. The steric changes induced by change in residue are observed in Figure 10.

When examining changes in amino acid residues, it is important to note the change in polarity, hydropathy and side chain constitution and bulkiness as all these factors will play a role in the secondary structure and consequently tertiary structure and function of the protein. Steric hindrance is of particular importance for protein interacting with a substrate, as the obstruction of the binding of the protein to the target will have an effect on the function of the protein. The location of the residue change is also therefore important. Changes in active sites or near residues that interact with substrates are particularly significant. The location of the variant residue in the protein of PGRP-S is therefore of interest. Currently, no crystal structures exist for peptidoglycan docked in PGRP-S. However, homology assisted docking of a PGRP-S analogue, muramyl tripeptide, was performed and used to determine which residues interact with peptidoglycan. These residues are refered to as “potential peptidoglycan binding residues”. One of these residues is tyrosine 74 (Guan et al, 2005). As the variant residue lies next to this significant amino acid, the changes resulting from the substitution may therefore have significant effects on the behaviour of the protein. This will be discussed in grater detail in section 4.5.

.

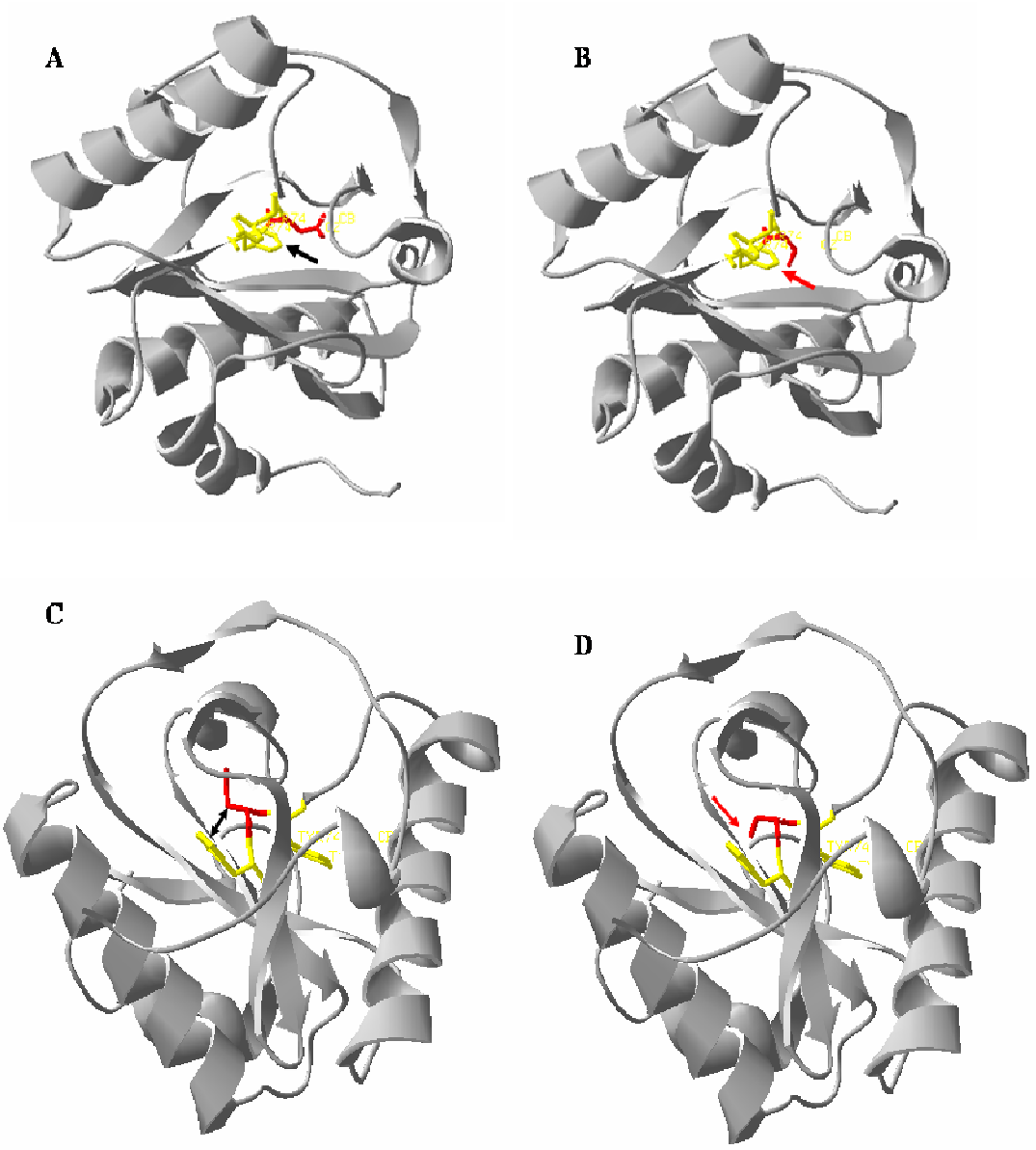

Figure 10: Structural analysis of PGRP-S (1YCK) and Patient 6 using the Swiss PDB viewer.

Blocks A and C represent the wild type protein, and blocks B and D the mutant protein. The residue mutated is represented in red and the neighbouring residue in yellow.

The Swiss PDB viewer can be used in the analysis of an elucidated protein structure deposited at the Protein Data Bank (PDB) (Guex, and Peitsch, 1997). The non-synonymous residue, Asparagine75 is next to Tyrosine 74, a potential peptidoglycan-contacting residue. The residue was mutated to the predicted Serine75 found in patient 6. The potential effect on interactions was examined. Block A is a view of the top of the wild type protein, and Block B is view of the mutant protein in the same orientation. The arrows indicate the proximity of the side chain of residue 75 with the phenol ring of the tyrosine molecule. The tyrosine molecule is a putative peptidoglycan contacting molecule, and therefore it's proximity to the target molecule is crucial (Guan et al, 2005). In Block A and B, it can be observed that the side-chain of the serine is much closer to the phenol ring of tyrosine 74. When rotating the molecule to a different view, it is still evident that the variant side chain is closer to the phenol ring of tyrosine 74. As the peptidoglycan molecule interacts with tyrosine, the decreased intermolecular distance between the residue and its' neighbouring amino acid may potentially interfere with interactions that residue may be involved in.

From the above figure, it appears that the mutation induces a change in the intramolecular distance of the mutant residue and its' neighbouring tyrosine residue. This can be confirmed by measuring the distance between the molecules using the atomic distance measurement function of the Swiss PDB viewer. The distances between the two molecules are depicted in Figure 11.

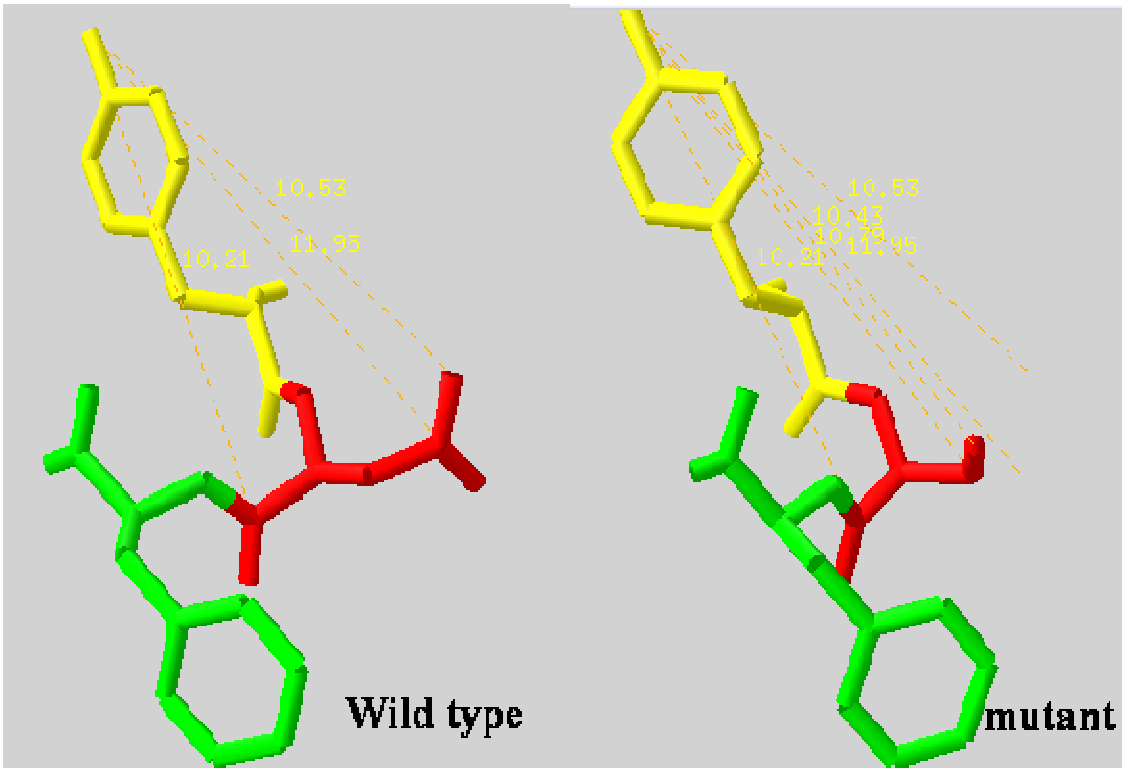

Figure 11: Measurement of distances between residues 74 and 75 of wild-type PGRP and the variant protein of Patient 6.

Tyrosine 74 is represented in yellow, residue 75 (asparagine in the wild type and serine in the mutant) is represented by in red. Phenylanlanine 76 is represented in green.

From the measurements of the extremities of the side chains, it is clear that the distance between the residues are indeed decreased, with the furthest distance decreasing from 11.95Å to 10.79Å, From the orientation of the molecules, it is possible that there is also a change in the hydrogen bonding pattern of the residues, as some hydrogen onds may be created or lost with the substitution of the variant residue. The Swiss PDB viewer can also be used to predict the location of the hydrogen bonds between the residue. This prediction is demonstrated in figure 12.

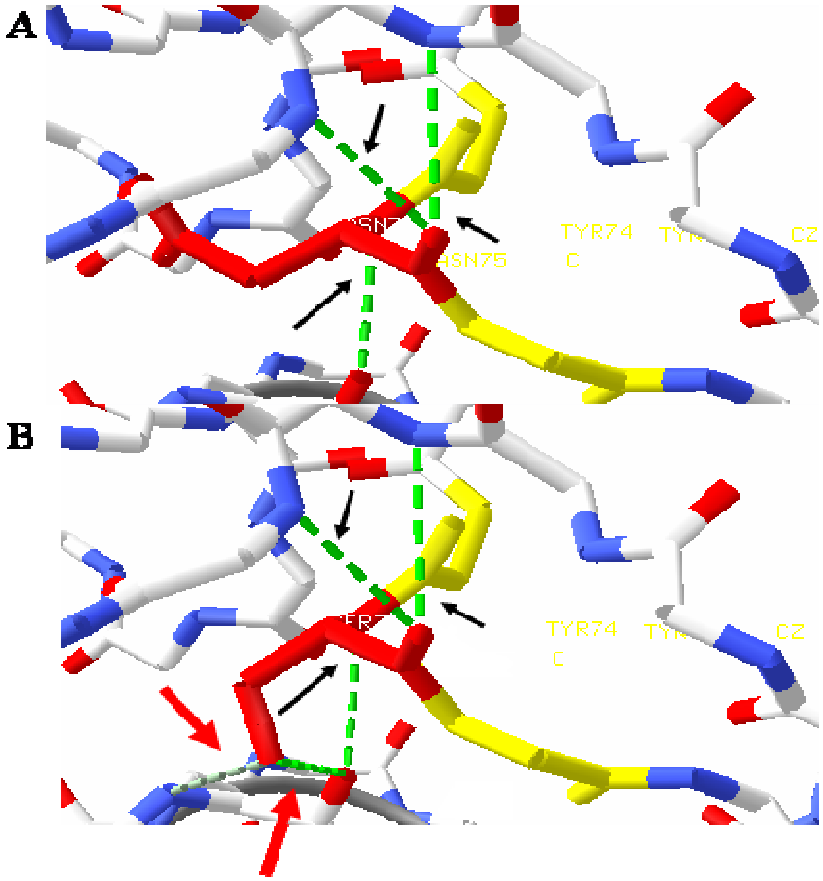

Figure 12: Hydrogen bonds between residues 74 and 75 of PGRP-S.

The orientation of the side chains in the wild type protein (Block A) differs to that of the mutant (Block B). As previously mentioned residues 74 and 75 are brought into closer proximity in the mutant protein. Although no steric clashes between the residues are predicted by the Swiss PDB programme, there is a change in the hydrogen bonding profile, indicated in the figure as a dashed green line. Although the hydrogen bonds that are present in the wild type protein are conserved (as indicated by the black arrows), the mutant protein has two additional hydrogen bonds with its neighbouring amino acids (indicated by red arrows).

3.3 Biochemical analysis

3.3.1 Heterologous Expression of a PGRP-S/GST fusion protein

As outlined in section 8 of chapter one, the biological functions of short PGRPs differ significantly between vertebrates and invertebrates. While the different types of PGRP-S' in *Drosophila* initate signalling transduction cascades and have direct enzymatic activity, current literature suggests that human PGRP-S is solely an antimicrobial peptide. This is rather unusual as in other organisms, the protein has a host of functions. In the rat sleep deprivation results in increased expression of PGRP-S in brain (Rehman et al, 2001). Recently it has been demonstated that the protein is induced by the transcription factor RelA in cases of cerebral ischemia in mice (Lang et al, 2008). Therefore the potential exists for PGRP-S to have other functions besides that of an antimicrobial protein. To be able to begin to understand what this role could possibly be, a clearer understanding of the protein-protein interactions of PGRP-S is required. For this reason, in vitro interaction studies of the protein were performed.

PGRP-S was successfully cloned, overexpressed and purified from E. *coli* BL21 cells. It was cloned from a leukocytes obtained from a fresh culture of blood from a healthy control. PGRP-S is usually cloned from bone marrow, and for activity studies usually expressed in *Drosophila* S2 cells as the protein is reported to be glycosylated. Using the protein without any posttranslational modifications would give an indication of the importance of the modifications. Figure 13 demonstrates the final elution steps of the GST affinity purification.

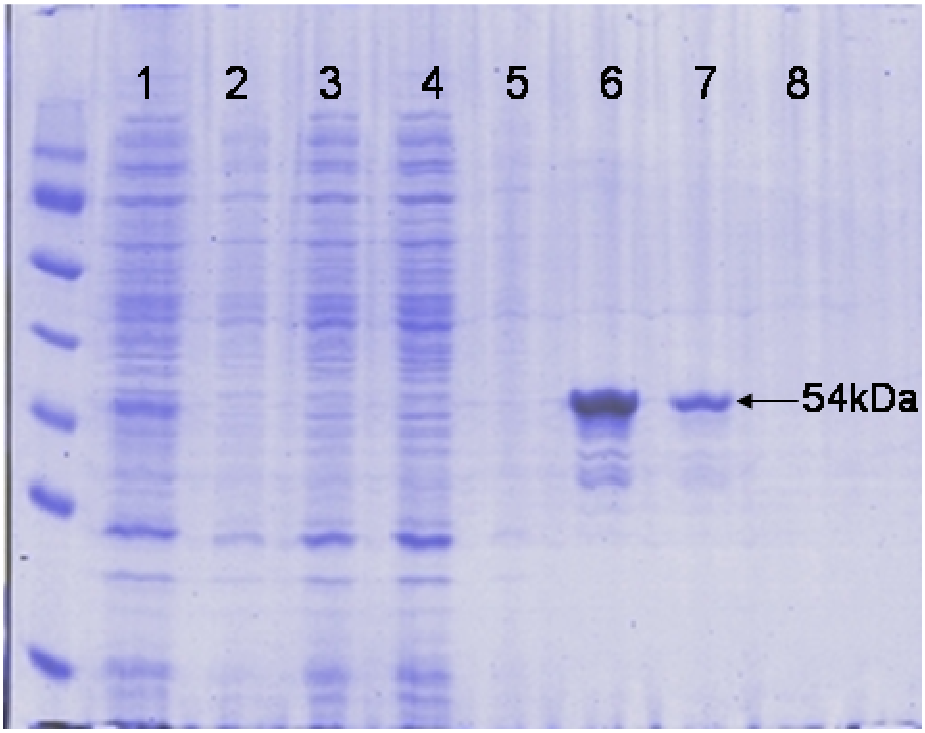

Figure 13: Purification of heterologously expressed PGRP-S/GST fusion protein. The first five lanes show the proteins washed from the column. Lanes 6-8 show the three reduced glutathione elution steps with the purified protein isolated in the second elution. The absence of protein in column 8, the third elution step, indicates that all the purified, GST-tagged protein eluted in the second glutathione elution.

3.3.2 Interaction studies of PGRP-S

Various reports have stated that the mouse homologue of PGRP-S, Tag7, forms a cytotoxic complex with Heat Shock Protein 70 (Sashchenko et al, 2005). It has also been reported that besides peptidoglycan, PGRPs can interact with lipopolysaccharides as well. The aim of the pulldown assay was to determine whether human PGRP-S could interact with protein components of *Drosophila melanogaster*. Figures 14 and 15 shows an interaction study which aimed to determine which proteins PGRP-S was capable of binding or forming an interaction with

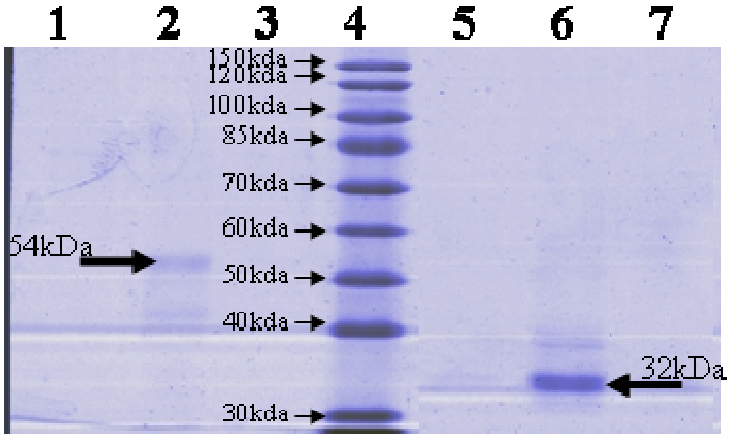

Figure 14: Pull Down assay to determine whether crude *Drosophila melanogaster* protein interacts with PGRP-S.

PGRP-S/GST fusion protein and GST were bound to a GST affinity matrix. *Drosophila* protein was added to the column and allowed to interact overnight. Upon GSH-elution, only the fusion protein was observed in the elution step (Lane 2) and no unbound or fusion protein/interactor complexes were observed in the pre and post wasjing steps (Lanes 1 and 3, respectively) No interactions were found in the column with GST (Lanes 5-7, the prewash, elution and postwash steps respectively), the negative control. This was to demonstrate that had any interacting proteins had been found for the fusion protein, it was due to an interaction with PGRP and not its GST

fusion tag. The interacting capability of the recombinant protein was further examined with an extended set of control reactions depicted in figure 15.

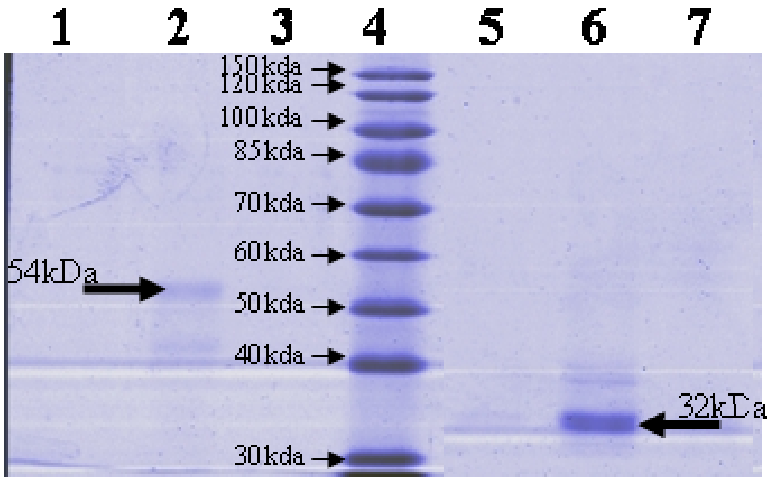

Figure 15: Pull Down Assay to determine whether human PGRP-S interacts with bacterial components.

This electrophoretogram contains another set of controls. Lanes 1-4 contain a double negative control. GST bound to a GST affinity column was allowed to interact with bacterial components. Lanes 1 and 2 represent a GST interaction with *Micrococcus luteus* soluble and insoluble protein respectively. No interaction was found between GST and *M. luteus* components Lanes 3 and 4 represent a GST interaction with *Escherichia coli* soluble and insoluble protein respectively. As with *M. luteus*, GST did not interact with *E. coli* components. Lanes 5 and 6 represent a PGRPS/GST interaction with *M. luteus* soluble and insoluble protein respectively. Lane 7 and 8 represent a PGRPS/GST interaction with *E. coli* soluble and insoluble protein respectively. The fusion protein interacted with bacterial components. These components would be very small protein components, as excluding the size of the fusion protein; the components in the complex would constitute 6 and 14kDa to the 60 and 68kDa complexes respectively.

3.3.3 Activity studies of PGRP-S

The interaction study demonstrated that the fusion protein is capable of binding bacterial components. As PGRP-S is antibacterial protein, a biochemical analysis would have to include antimicrobial assays. Radial diffusion assays were used to determine antibacterial activity as it is described as a method to obtain maximal sensitivity while using the minimal amount of protein (Lehrer et al, 1991). Furthermore, it eliminates the possibility of a false positive due to smearing of protein from an impregnated filter.

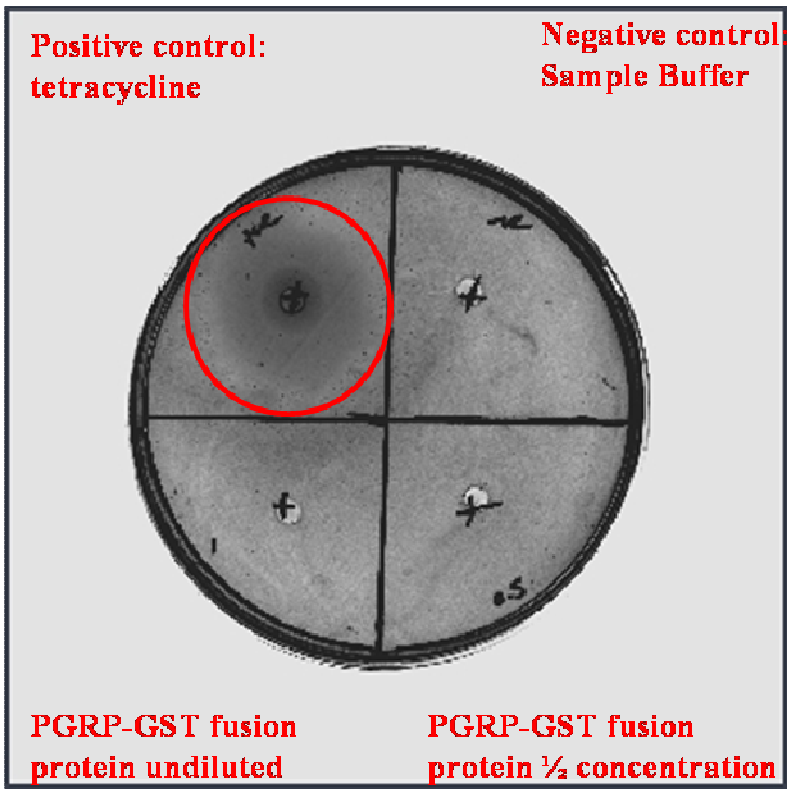

Figure 16: Determination of antimicrobial activity of PGRP-S against *Escherichia coli* by Radial Diffusion assay.

E. coli was used as a representative Gram-negative bacterium, which PGRP-S does not preferentially bind to and is therefore not highly active against. No clearings were noted. To determine whether this was due to protein inactivity, the assay was supplemented with zinc chloride, as zinc is the divalent ion required for activity against Gram negative bacteria.

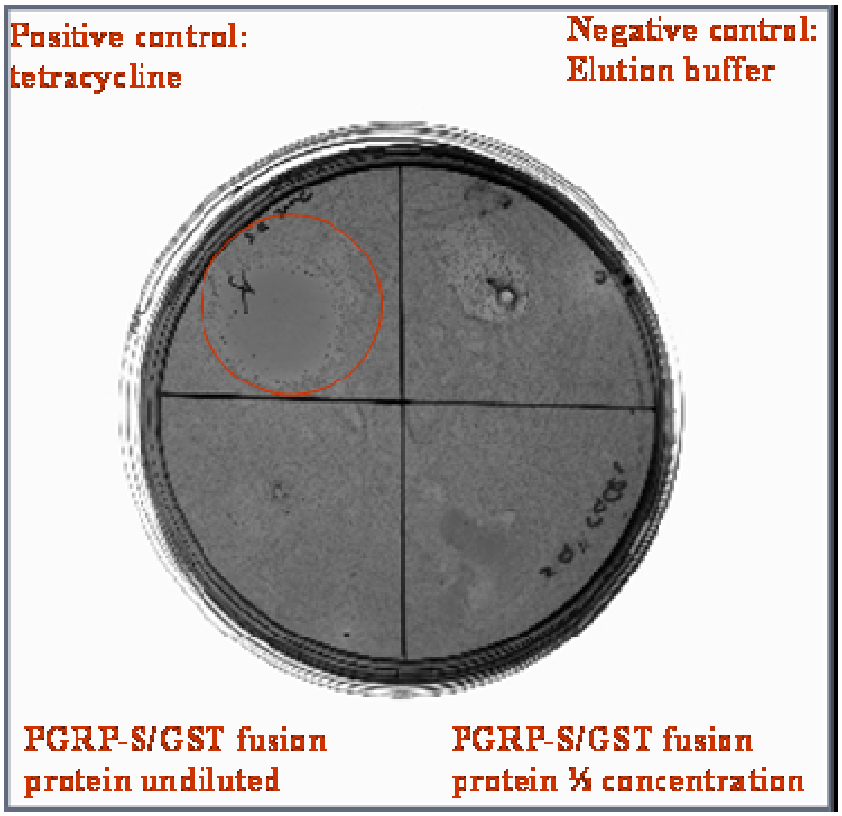

Figure 17: Determination of antimicrobial activity of PGRP-S against *Esherichia coli* by Radial Diffusion assay supplemented with zinc chloride.

As with the unsupplemented assay, there was no noted antimicrobial activity. The fusion protein does not appear to be active against Gram negative bacteria.

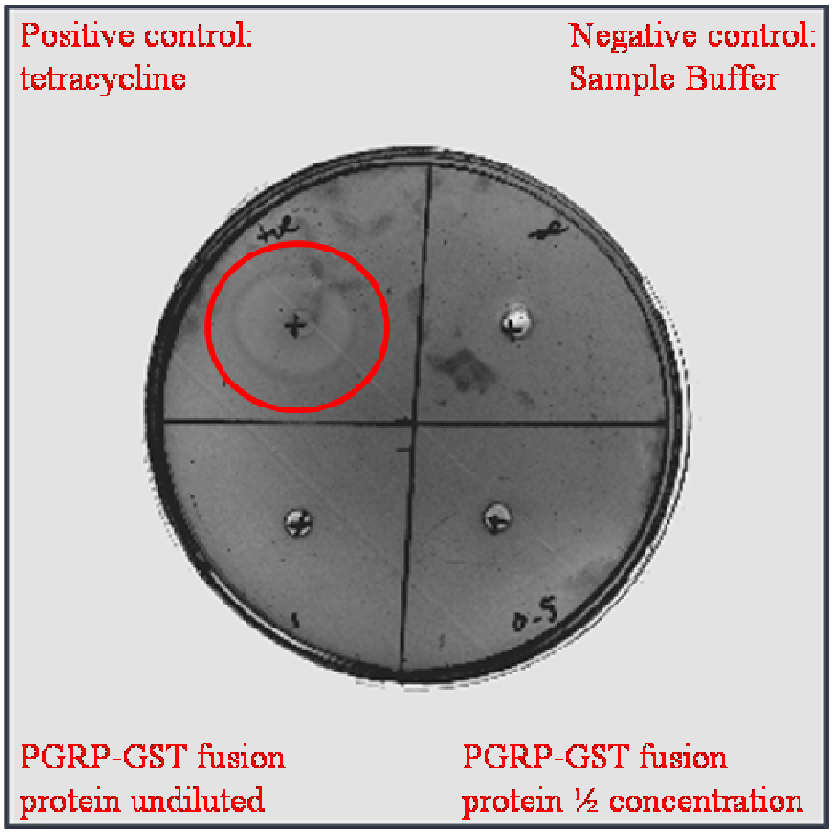

Figure 18: Determination of antimicrobial activity of PGRP-S against *Staphylococcus aureus* by Radial Diffusion assay.

S. aureus was chosen as a representative Gram positive bacterium as PGRP-S is reported to have a bactericidal effect on this culture. As with E. *coli*, no clearings were noted, and similarly, the experiment was repeated with a zinc supplement, as this is the divalent ion required for activity against Gram positive bacteria.

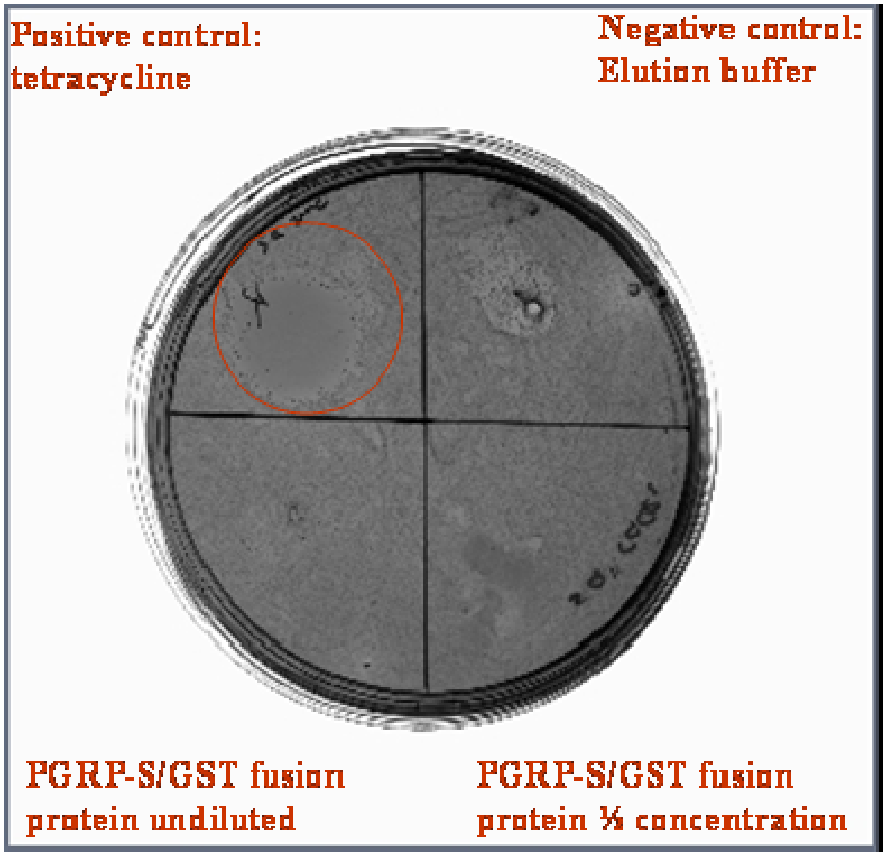

Figure 19: Determination of antimicrobial activity of PGRP-S against *Staphylococcus aureus* by Radial Diffusion assay supplemented with Zinc chloride. As with the experiment testing the antimicrobial activity against *E. coli* with a calcium supplement, no clearings were noted. The fusion protein appears to have no antimicrobial activity.

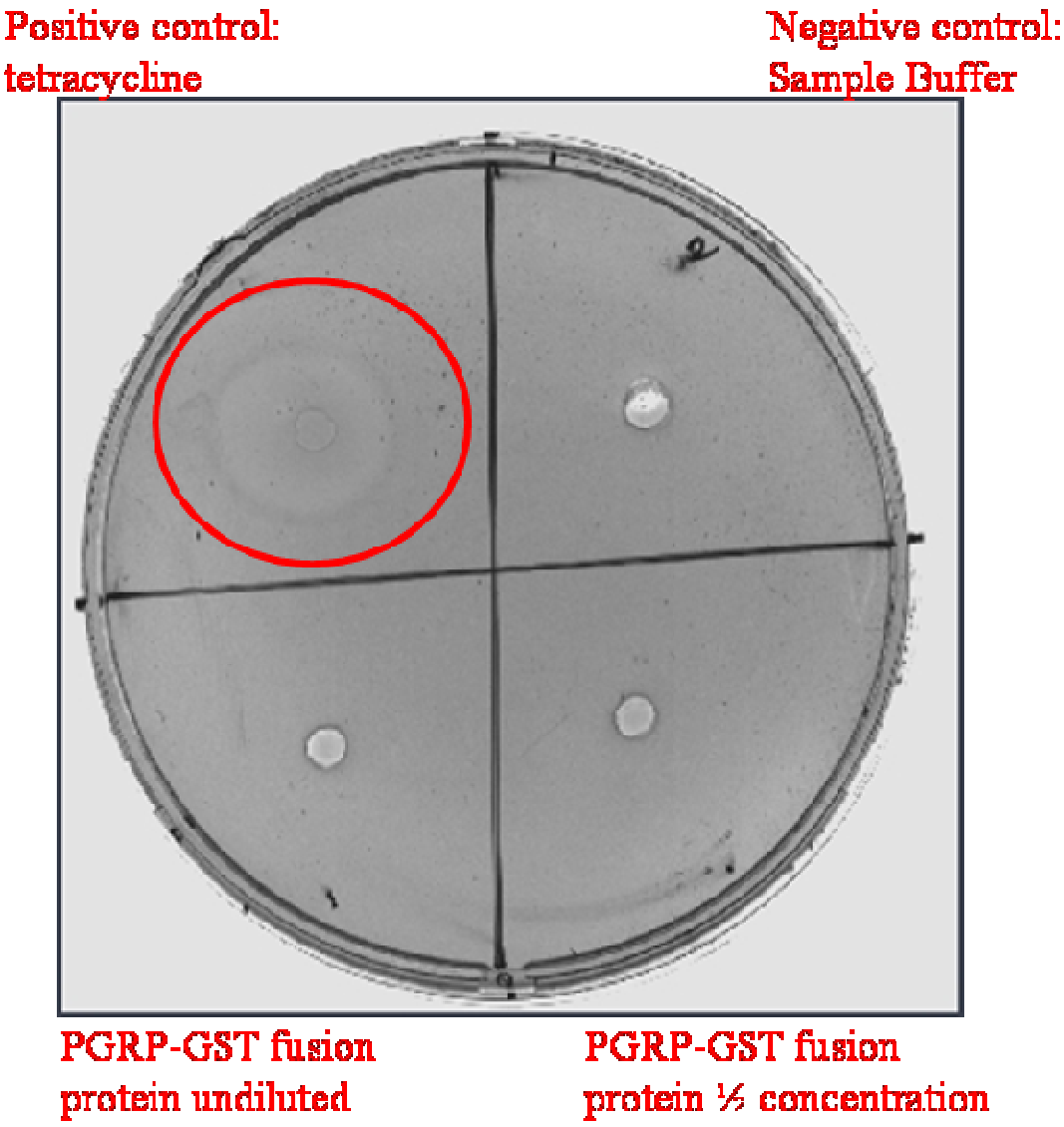

Figure 20: Determination of antimicrobial activity of PGRP-S against *Staphylococcus aureus* by Radial Diffusion assay supplemented with calcium chloride.

Calcium chloride can be used as an alternate ion source for activity against Gram positive bacteria. As with the experiment testing the antimicrobial activity against *E. coli* and *S. aureus* with a zinc supplement, no clearings were noted. The fusion protein appears to have no antimicrobial activity.

CHAPTER FOUR

DISCUSSION

4.1 Genetic substructure in sub-Saharan Africa

The black South African population can be classified on a linguistic basis. The Sotho/Tswana group, the Nguni group and the Venda group. The Nguni group, comprising of Zulu, Xhosa and Tsonga speakers constitute 45.2% of first language home speakers. The Sotho group, comprised of Northern and Southern Sotho speakers as well as Tswana speakers constitutes 25.1% of first language home speakers (Lane et al, 2002). Populations from southern Africa are poorly represented in population sequencing studies. They may be similar to a group of African Americans designated the WAFR (West African) described by Tian et al (2006).This group included the Bini, a group from the Edo state in Nigeria (Tian et al, 2006). Most black South Africans originated from a group that dispersed from the current Nigerian/Cameroonian border approximately 5000 years ago (Beleza et al, 2005). Since then, much differentiatiation has taken place, as geography, rather than language group, plays a more defining role in the African genetic landscape (Salas et al, 2002). The sub-Saharan African population is therefore a unique and relatively unexplored group as only the Yorubi of Nigeria has been examined for population variation.

The pilot study conducted by this research had a sample size too small to make any firm conclusions about the distribution of variation in the population. However, it does give direction for future study. Three of the four variant samples, Patients 12 (Zulu), 15 (Sotho) and 27 (Zulu), are all Nguni speakers. Patient 6, however, is Venda. Vendas comprise a small portion of the South African population, and constitute only 2.2% of first language mother tongue speakers. Furthermore, Vendas

are the most genetically distinct of any of the groups examined. Phylogenetic analysis places them closer to Tsongas, an Nguni group, than any other population group based on autosomal and Y-chromosome data (Lane et al, 2002). They are considered to be one of the last population groups to migrate into Southern Africa. They settled in the Limpopo province in the North of South Africa and have had very little admixture with other population groups (Lombard et al, 2006). Studies on variation in cytochrome P450 and Glutathione S-transferases (Dandara et al, 2002; Sommers et al, 1991) in the Venda population did not yield any variation from other African population. Vendas are predicted to be more susceptible to tuberculosis due to an increased frequency of the occurrence of variants in the vitamin D receptor. 37% of tuberculosis cases in the Limpopo province are in patients of Venda descent (Lombard et al, 2006). It is therefore interesting to note that the most remarkable potential mutation takes place in a member of this population.

4.2 Distribution of variation in the PGRP gene

In the PGRP-S gene, a total of 39 Single Nucleotide Polymorphisms have been reported. Of those 39, 17 occurred upstream of the coding region, one in the 5' untranslated region, 11 were intronic, and two were downstream of the coding region. Two frame shift mutations have also been noted (Hubbard et al, 2007). Of the polymorphisms in the coding region, four (rs13343537, rs28722714, rs34478490, rs34180629) have been reported, with one of the changes being a synonymous mutation at Aspartic acid 92 (Sherry et al, 2001). A recent study examined the sequence variability of bovine PGRP-S. The bovine protein is a 190 amino acid protein coded for by a 3 exon, 688 base pair transcript. The signal peptide, PGRP domain and zinc amide/N-acetylmuramoyl-L-alanine amidase acid (Ami_2) domain are conserved in humans, *Mus musculus*, *Rattus norvergicus*, *Canis familiaris*, *Bos taurus* and *Bos indicus*. The gene and protein structure is therefore conserved between cattle and humans. Only 3 SNPs were detected in the coding region, with two of them being non-synomous. The majority of variations found in bovine species

are transversions. Of the 6 SNPs observed in the coding regions of human PGRP-S, two insertions, two transitions and two transversions are noted. Overall, the published data points to the fact that, unlike *Drosophila*, variations leading to immunological consequences may not lie in the coding regions, but rather in the intronic regions. Of the variations within the coding regions, most of the variation was found in Exon 1, with one synonymous change in Exon 2 and none in Exon 3.

4.3 Discovery of previously noted polymorphisms

Polymorphism rs2072562 is a C to A substititution at mRNA position 26, reference contig position 18794516. This is a change in the 5' untranslated region. This polymorphism, at present, has no frequency data. This study therefore represents the first genotype frequency data for this particular polymorphism. Of twenty patients, the CA heterozygous genotype occurs at a frequency of 0.75, the homozygous CC with a frequency of 0.15, and the homozygous AA genotype with a frequency of 0.1. The frequencies are not in Hardy Weinberg equilibrium. This is to be expected as the Hardy Weinberg equilibrium is based on an infinite or sufficiently large population. Twenty samples are therefore not representative. However, the frequency of the occurrence of the C-allele does appear to confirm the assertion by NCBI that the C-allele is the ancestral allele as it occurs with an allele frequency of 0.53 compared to 0.48 of the A-allele. The homozygous AA phenotype was only observed twice, in a HIV positive and an HIV/TB positive patient. This SNP is found to exist in Hardy-Weinberg equilibrium.

The previously reported polymorphism rs28722714 was detected in the patients 1, 4, 6, 11, 15, 24, 27 and 29. This is a C to T substitution at mRNA position 320, reference contig position 18794222 and all the aforementioned patients are heterozygous for the polymorphism. The C-allele is the ancestral allele. Again, in confirmation, the homozygous CC genotype is found occurring with a frequency of 0.6. eight patients (frequency 0.4) were heterozygous for the CT genotype. Patient 27

is the first patient in which this polymorphism was observed. In all the heterozygous patients the allele was given an ambiguity code as the two bases could not be distinguished between each other. Both the C and T-alleles would result in the amino acid Aspartate. This is the second previously observed polymorphism found; validating the quality of the sequencing data and giving credibility to any potentially novel SNPs that may be found.

4.4 A potentially novel synonymous polymorphism.

A G to A polymorphism is observed at mRNA position 110. The first patient found to have this polymorphism was Patient 15, as this was the only patient with an A in position 110. Further investigation showed that the polymorphism was a heterozygous change. This genotype was observed 16 times, a frequency of 0.8. No homozygous AA genotypes were observed. Four GG homozygotes were observed, and this genotype was observed only in HIV or HIV/TB patients. The G allele is the reference allele from all databases. Furthermore this allele occurs with an allele frequency of 0.6. It is therefore possibly the ancestral allele. The change is a synonymous mutation (glutamine to glutamine). It is however, interesting to note that the polymorphism is at the beginning of the mature peptide. If a non-synonymous change had taken place at this position, it could potentially have implications on the processing and trafficking of the protein as it occurs at the point of splicing of the signal peptide and mature peptide.

4.5 A potentially novel non-synonymous polymorphism

Patient 6 has a homozygous A to C substitution at mRNA position 312. It is a unique change, unobserved in both the sample population or previously noted in any database. This polymorphism is interesting for several reasons. Firstly, the existence of this polymorphism needs to be confirmed with more rounds of sequencing as the residue affected is a residue overlap splice site. This residue, as well as Aspartic acid

116 is coded for by a codon that overlaps two exons, with the resultant mutant reside falling on a junction of exon 1 and 2.

Secondly, the polymorphism represents a potentially novel non-synonymous polymorphism. Only 1 synonymous and 2 non-synonymous polymorphisms have been reported on the NCBI, Ensembl, UCSC and SNPer databases. The polymorphism results in an Aspargine 75 to Serine substitution.

None of the residues affected by the reported polymorphisms, or this potentially novel substitution, are cysteines involved in disulphide bonding. This would have been significant not only to the stability and tertiary structure of the protein, but would also be likely to affect the functionality of the protein, as the protein is reported to function as disulphide linked homodimer (Lu et al, 2006).

No mutations are reported in any evolutionarily conserved residues, or residues involved in stabilizing reactions. Aspartate 71 and glycine 91 are predicted to form a bidentate salt bridge. These residues are unaffected. Tyrosine 85 forms a tight hydrophobic cluster with alanine 20 and 56 as well as tryptophan 18, a residue found in all insect and mammalian PGRPs (Guan et al, 2005). None of these residues are affected,

The mutation in Patient 6 comes into play with peptidoglycan contacting residues. Histidine 40, threonine 41, tyrosine 62, histidine 63, tyrosine 74, histidine 96 and 99, asparagines 102 and glutamine 153 are predicted to be peptidoglycan contacting residues. Although these residues are not all located together in the primary sequence, in the tertiary structure of the protein they are found in and around the peptidoglycan binding cleft (Guan et al, 2005). Although the mutation in patient 6, asparagine 75 to serine, is not one of the predicted peptidoglycan contacting residues, serine 75 is next to tyrosine 74, which is predicted to be required for peptidoglycan contact. The mutation may therefore play a significant role in the protein's functionality.

An asparagine to seine substitution is a conservative change in terms of electrostatics as both amino acids are uncharged polar amino acids. The asparagines residue is heavier than serine, 114.1D compared to 87.1D of serine. The pK value of the α carboxy terminal of the residues are similar, asparagines with a pK of 2.14 and seine 2.19. The pK values of the α amino terminal of serine is higher (9.21) than that of asparagines (8.72) (Voet and Voet, 1995). A serine substitution also represents a change in hydrophobicity as it has a hydropathy of -0.8 compared to the -3.5 of asparagines (Kyte and Doolittle, 1982). The primary structural change is therefore a conservative in terms of charge and polarity, but results in a decrease in residue mass and a slightly increased hydrophobicity.

A change in secondary structure is more difficult to predict. All secondary structure changes are all predicted by PsiPred (Bryson et al, 2005; Jones, 1999). The accuracy of this method is questionable. The crystal structure of PGRP-S at the protein databank (1YCK) shows that structure has 5 helices and 7 strands. The PsiPred prediction predicts 7 helices and 5 strands. In terms of the prediction of the number of secondary structure, that is not a poor accuracy. However, the positions of the predicted secondary structures are grossly inaccurate. Of the predicted structures, only one helix is corresponds to the actual helix in that region.
Of the seven strands, only two of the strands predicted by PsiPred are in the correct region. Not a single secondary structure position is perfectly predicted. One of the secondary structures that were predicted with a degree of accuracy is a 3_{10} helix reported to be found in all PGRP structures (Guan et al, 2005). In PGRP-S and all short PGRPs examined this helix is preceded by a strand. In the PsiPred prediction of Patient 6's protein, the preceding strand is missing, even though the mutation is not in this area. Indeed, significant secondary structural changes are predicted in Patient 6. This is questionable. The mutation is the second amino acid in a strand. Asparagine and Serine are both short polar amino acids are classed as turn and bend admirers (Malkov et al, 2005). The amino acid substitution should therefore not disrupt the

secondary structure. However, secondary structure can only be accurately assessed by circular dichroism spectroscopy.

The Swiss PDB viewer can be used to analyse the effect of alternate residues on the intramolecular contacts of a molecule, as well as to determine whether the substitution is viable (Guex and Pietsch, 1997). Asparagine 75 is next to tyrosine 74, a potentially peptidoglycan contacting residue. It is therefore also situated in the peptidoglycan binding cleft. What is overwhelmingly clear is that the intramolecular distance between residue 74 and 75 is decreased. The mutant residue may therefore be a steric hindrance in the binding of residue 74 to peptidoglycan. Furthermore, the serine residue substitution results in the formation of hydrogen bonds with isoleucine 107 and glycine 108 that were not there before. This could also have a potential negative effect by inducing rigidity in the molecule that may affect its ability to bind peptidoglycan.

The changes to the protein brought about by the mutant residue are not due the residue itself, but rather due to its interplay with its neighbouring residue. As it is only one peptidoglycan binding residue, it is possible that the overall capability of the mutant protein to bind peptidoglycan will not be negated, but the protein will be functional, albeit with a reduced capability to bind peptidoglycan.

4.6 Antimicrobial activity of PGRP-S

All PGRPS function as dimers (Lu et al, 2006). This is not completely surprising as *Drosophila* PGRPs undergo ligand induced dimerization (Mellroth et al, 2005). The nature of these dimerization reactions is not completely understood. The preferential heterodimerization of PGRP I α and β where co-expressed (Lu et al, 2006) raises the question, how similar must PGRPs be in order to form dimers? Would it be possible for human PGRP-S and short PGRPs from *Drosophila* to interact, dimerize, or even form a functional complex?

The pull down assays, which allowed crude *Drosophila* protein to interact with PGRP-S fusion protein immobilized on a GST affinity matrix, did not yield any interacting proteins. The reason for this could be threefold. Firstly, it could be that PGRP-S from humans and all the short PGRPs from *Drosophila* could simply not interact. The proteins may just be too evolutionarily distinct to form functional complexes or complex at all. Werner et al, when describing the *Drosophila* PGRP family aligned several PGRPs across different species. Several residues were conserved but the primary structures of the proteins were not strikingly similar. As the dimers are disulphide linked, it is possible that these differences in primary structure account for the inability to interact.

Secondly, the lack of interaction may be a function of the environment the proteins were allowed to interact in. The column may not be the appropriate medium and perhaps does not represent the physiological conditions that would have allowed the interaction to take place. The reaction may take place in vivo, but not in vitro. This can be tested by transfecting *Drosophila* S2 cells with a fusion tag and a eukaryotic promoter and subsequently purify any resultant complexes with affinity chromatography.

The recombinant protein in this experiment did not form inclusion bodies as it did in previous experiments. The protein may therefore be incorrectly folded. Furthermore in studies that examined the activity of PGRP-S, the protein was produced in a eukaryotic system (Cho et al, 2005, Lu et al, 2006). The non-functionality of the protein is highlighted by the complete absence of antibacterial activity, even when the reaction is supplemented with divalent cations (Wang et al, 2007). It is not inexplicable that the protein has no antibacterial activity as N-glycosylation is described as an essential post-translational modification required for activity, as glycosylase complete negates antibacterial activity (Lu et al, 2006). This study demonstrates the lack of activity of a non-glycosylated version of the protein.

The most likely scenario, however, is that the entire PGRP protein, i.e. the signal peptide and mature peptide, was cloned. As with many antibacterial proteins, such as the Cathlecidins, the proteins appear to be inactive until the protein is processed. On initiation of the project, PGRPs were not classed as antibacterial proteins, and the signal peptide was classed as a putative signal peptide. To be able to perform this experiment accurately, the following steps have to be performed. Firstly, the mature peptide only needs to cloned in a eukaryotic system. Once the mature peptide has been purified, a glycosylase treated protein preparation can be used as a control to confirm the necessity of N-glycosylation for activity. To examine the effect of the mutation observed in Patient 6, site directed mutagenesis can be used to induce the mutation, and the mutant protein can then be cloned, overexpressed and purified and used in activity studies and compared to the activity of the wild type protein.

Although PGRPs are unique because they are N-glycosylated, they are not the only glycosylated proteinacious antimicrobial agents. O-glycosylation is a post-translational mechanism that is found, usually, in proline rich peptides. Drosocin from *Drosophila,* lebocin from *Bombyx mori,* formaecin from ants, diptericin from *Anopheles gambiae* and pyrrhocoricin from *Pyrrhochoris apterus* are all O-glycosylated peptides (Otvos, 2002). It is proposed that membrane permeabilization is not the primary mode of action of glycosylated peptides as the kinetics of membrane permeabilization for pyrrhocoricin, drosocin and formaecins is slow and usually requires hours of contact between the peptide and the micro-organism (Andreu and Rivas, 1999). These AMPs are proposed to work by binding DnaK and subsequently interfering with the bacterial protein folding mechanism, as this method has been demonstrated by pyrrhocoricin (Kragol et al, 2001). Although it has been demonstrated that the murine PGRP homologue Tag7 forms a cytotoxic complex with Hsp70 (Saschenko, 2004) it is unlikely that PGRP-S may function in the same manner, as pyrrhocorin is a 20-mer peptide (Cociancich et al, 1994), and therefore able to enter the nucleus and disrupt the translation machinery, while the 42kDa PGRP-S monomer would be too large to function in a similar manner.

4.7 Alternate splicing of PGRP-S

Northern blot analysis of human PGRP expressing tissues showed that different splice variants exist as transcripts of different sizes were found in different tissues. PGRP-S expressed in the bone marrow exists as 1.4kb, 0.9kb and 0.5kb transcripts.1.4 and 0.9kb transcripts exist in peripheral blood leukocytes and a 0.9kb transcript in the foetal liver (Liu et al, 2001). No studies have been reported on the functionality of the products of the splice variants. When cloning procedures from PGRP-S has been described it is never mentioned what size transcript was used. Furthermore, although polymorphisms have been found in what is described as the promoter region, there is nothing published about human PGRP-S promoters. This may suggest that there is a degree of cross promoter usage, although this has not been examined. The field of PGRP-S regulation is therefore an unexplored one.

The porcine PGRP-L isoforms are created by alternative splicing. These isoform play a role in the regulation of β-defensin in the pig (Sang et al, 2005). It could therefore be useful to determine whether the alternate splice variants of PGRP-S result in different proteins. Furthermore no reports have been made about the regulation of splice variation. It is possible that RNA editing may be one such regulation factor.

4.8 RNA editing

All samples sequenced from genomic DNA have a C-allele at position 80. The mRNA sequence, however, has a T-allele in the same locus . This is confirmed by all the databases used to obtain genomic and mRNA sequences. This particular change could represent an RNA editing event. RNA editing has been reported in eukaryotes ranging from protozoans (Benne et al, 1986), houseflies and German cockroaches (Xu et al, 2006) to mice (Powell et al, 1987; Sommer et al, 1991) and humans (Gott,

2000). In humans, RNA editing contributes to antibody diversity (Miramatsu et al, 2007).

A basic RNA editing event is the deamidation of Cytosine resulting in a Uracil substitution in the transcript. Typically, this affects translation and may cause an amino acid substitution, the introduction of a termination codon or a new reading frame due to the introduction of a new start codon (Samuel, 2003). The CCC to TCC change is a synonymous mutation. This is not unheard of, but is not as common as substitutions which induce a change (Gott and Emmerson, 2000). This silent change could have other effects besides a translational change. It may change the RNA structure, or, more importantly for PGRP-S it may affect the RNA splicing allowing for the selection of alternative splice sites. This C to T substitution from the genomic DNA to mRNA may represent a mechanism for the production of splice variants. This is a completely unexplored field and could potentially yield important information on the functionality of PGRP-S.

4.8 Future directions

It is clear that there are still many potential areas of study in the genetic diversity and functioning of PGRP-S. It would be advised to confirm the existence of the mutation in Patient 6 with further rounds of PCR closer to the region of interest. The polymorphism at mRNA position 27 and 110 can be further genotyped to determine whether these polymorphisms are diseases associated.
Most interestingly from a genetic point of view is to examine promoter and intronic variation, as intronic variation appears to be more common than polymorphisms in the coding region. It is possible that potential PGRP dysfunction may be as a result of aberrant splicing and regulation rather than protein dysfunction as not much variation ahs been reported in the protein itself.

This study has demonstrated the importance of glycosylation for the functioning of PGRP-S. The ideal set of experiments to perform therefore would be to produce the protein in cell culture and determine proteins that interact in vivo. Furthermore, it would be useful to know which human proteins PGRP-S interacts with. Site directed mutagenesis of the crucial peptidoglycan residues would demonstrate their true value in the protein.

Although the sample size of this study is too small to draw statistical conclusions, there appears to be a trend that the protein does not display much genetic variation, and that the Sub-Saharan African population does not differ from the rest of the population in terms of this protein. The conservation of this protein may highlight the immunological significance of the protein.

4.9 Conclusion

As with bovine PGRP-S, there is a very low frequency of variation in the coding region of the protein, and even less variation in the protein. The variation seen in the Southern African population does not vary tremendously from the trends seen in Europeans and Asians. The high degree of conservation of the PGRP protein within mammals and human populations suggest that the molecule plays a critical role in normal immune functioning. At present there is not enough data to suggest am association between PGRP, HIV and tuberculosis. However, there are still many research questions about the protein that can be asked, particularly with the recent literature associating the protein with clinical diseases.

CHAPTER FIVE

REFERENCES

Akira, S. and Takeda, K. (2004).Toll-like receptor signalling *Nat Rev Immunol*.**4,** 499-511

Andreu, D. and Rivas, L. (1998). Animal antimicrobial peptides: an overview. *Biopolymers* 47**,** 415-433.

Andersen-Nissen, E., Smith, K.D., Bonneau, R., Strong, R.K., Aderem, A. (2007). A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. *J Exp Med*. **204**, 393-403.

Bachali, S., **Jager, M.**, **Hassanin, A.**, **Schoentgen, F.**, **Jollès, P.**, **Fiala-Medioni, A.**, **Deutsch, J.S**. (2002). Phylogenetic analysis of invertebrate lysozymes and the evolution of lysozyme function. *J. Mol. Evol.* **54,**652-664.

Banchereau, J. and Steinman, R.M. (1998). Dendritic cells and the control of immunity. *Nature* **392**,245-252.

Beleza, S., **Gusmão, L.**, **Amorim, A.**, **Carracedo, A.**, **Salas, A**. (2005).The genetic legacy of western Bantu migrations. *Hum Genet.* **117,** 366-375.

Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., Tromp, M.C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46,** 819-826.

Bettencourt, R., Tanji, T, Yagi,Y., Ip, Y.T. (2004). Toll and Toll-9 in *Drosophila* innate immune response. *J Endotoxin Res.* **10**, 261-268.

Beutler, B. (2004). Inferences, questions and possibilities in Toll-like receptor signalling. *Nature*, **430**, 257-263.

Bilak, H., **Tauszig-Delamasure, S.**, **Imler, J.L.** (2003). Toll and Toll-like receptors in *Drosophila. Biochem Soc Trans*. **31**,648-51.

Bischoff, V., **Vignal, C.**, **Duvic, B.**, **Boneca, I.G.**, **Hoffmann, J.A.**, **Royet, J**. (2006). Downregulation of the *Drosophila* immune response by peptidoglycan-recognition proteins SC1 and SC2. *PLoS Pathog*. **2**, e14.

Brogden, K.A., Ackermann, M., McCray, P.B. Jr, Tack, B.F. (2003) Antimicrobial peptides in animals and their role in host defences. *Int J Antimicrob Agents*. **22**,465-478. *Cell Mol Life Sci.* **58,** 978-989.

Brown, K.L. and Hancock, R.E.W. (2006). Cationic host defense (antimicrobial) peptides. *Curr. Opin. Immunol*. **18**, 24-30.

Bryson,K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., Jones D.T. (2005). Protein structure prediction servers at University College London. *Nucl. Acids Res.* **33,** W36-38.

Cerenius, L. and Söderhäll, K. (2004). The prophenoloxidase-activating system in invertebrates. *Immunol Rev.* **198,** 116-126

Chang, C.I., Pili-Floury, S., Hervé, M., Parquet, C., Chelliah, Y., Lemaitre, B., Mengin-Lecreulx, D., Deisenhofer, J. (2004). A *Drosophila* pattern recognition receptor contains a peptidoglycan docking groove and unusual L,D-carboxypeptidase activity. *PLoS Biol.* **2**, e277

Cho, J.H., Fraser, I.P., Fukase, K., Kusumoto, S., Fujimoto, Y., Stahl, G.L., Ezekowitz, R.A. (2005). Human peptidoglycan recognition protein S is an effector of neutrophil- mediated innate immunity. *Blood* **106**, 2552–2558.

Clark, R. and Kupper, T. (2005). Old meets new: the interaction between innate and adaptive immunity. *J Invest Dermatol*. **125,** 629-637.

Cociancich ,S., Dupont, A., Hegy, G., Lanot, R., Holder, F., Hetru, C., Hoffmann J.A., Bulet, P. (1994). Novel inducible antibacterial peptides from a hemipteran insect, the sap-sucking bug *Pyrrhocoris apterus*. *Biochem J.* **300,** 567-575.

Dandara, C., **Sayi, J.**, **Masimirembwa, C.M.**, **Magimba, A,**, **Kaaya, S,**, **De Sommers, K.**, **Snyman, J.R.**, **Hasler, J,A**. (2002).Genetic polymorphism of cytochrome P450 1A1 (Cyp1A1) and glutathione transferases (M1, T1 and P1) among Africans. *Clin Chem Lab Med.* **40**, 952-957.

De Smet, K. and Contreras, R. (2005). Human antimicrobial peptides: defensins, cathelicidins and histatins. *Biotechnol Lett***. 27**, 1337-1347.

Düring, K., Porsch, P., Mahn, A., Brinkmann, O., Gieffers, W. (1999). The non-enzymatic microbicidal activity of lysozymes. *FEBS Lett.* **449,** 93-100.

Dziarski, R. (2004). Peptidoglycan recognition proteins (PGRPs). *Mol Immunol.***40,** 877-886.

Dziarski, R., Platt, K.A., Gelius, E., Steiner, H., Gupta, D. (2003). Defect in neutrophil killing and increased susceptibility to infection with nonpathogenic gram-positive bacteria in peptidoglycan recognition protein-S (PGRP-S)-deficient mice. *Blood.* **102,** 689-97.

Dziarski, R. and Gupta, D.(2006). The peptidoglycan recognition proteins (PGRPs). *Genome Biol.* **7,** 232.1-232.13

Dziarski, R. and Gupta, D.(2006). Mammalian PGRPs: novel antibacterial proteins. *Cell Microbiol.* **8,** 1059-1069.

Egholm, M. (2007). The James Watson Genotype viewer version 1.69: http://jimwatsonsequence.cshl.edu/about.html, last modified: 11July 2007

Evrard, C., **Fastrez, J.**, **Declercq, J.P**. (1998). Crystal structure of the lysozyme from bacteriophage lambda and its relationship with V and C-type lysozymes. *J. Mol. Biol* **276,** 151-164

Fearon, D.T. (1999). Innate immunity and the biological relevance of the acquired immune response. *QJM* **92**, 235-257.

Fearon, D.T. and Locksley R.M. (1996). The instructive role of innate immunity in the acquired immune response. *Science* **272**, 50-53.

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol.* **3**, 710-720.

García, J.R., Jaumann, F., Schulz, S., Krause, A., Rodríguez-Jiménez, J., Forssmann, U., Adermann, K., Klüver, E., Vogelmeier, C., Becker, D., Hedrich, R., Forssmann, W.G., Bals, R. (2001). Identification of a novel, multifunctional β-defensin (hBD-3) with specific antimicrobial activity, interaction with oocyte membrane, and inducing macrophage chemoattraction. *Cell Tissue Res.* **306**, 257-264.

Gillespie,J.P., Kanost, M.R., Trenczek, T. (1997). Biological mediators of insect immunity. *Annu Rev Entomol.* **42,** 611-643

Golec, M. (2007). Cathlecidin LL-37:LPS-neutralizing, pleitropic peptide. *Ann Agric Environ Med* **14**, 1-4

Gordon, L.I., Douglas, S.D., Kay, N.E., Yamada, O., Osserman, E.F., Jacob, H.S. (1979). Modulation of neutrophil function by lysozyme. Potential negative feedback system of inflammation. *J Clin Invest*. **64,** 226-232.

Gott, J.M., and Emeson, R.B. (2000). Functions and mechanisms of RNA editing. *Annu Rev Genet.* **34,** 499-531.

Guan, R., Wang, Q., Sundberg, E.J., Mariuzza, R.A. (2005). Crystal structure of human peptidoglycan recognition protein S (PGRP-S) at 1.70 A resolution. *J Mol Biol.* **347,** 683-91.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.

Hashimoto,C. Hudson, K.L, Anderson, K.V. (1988).The Toll gene of *Drosophila*, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell*. **52**, 269-79

Hirakawa, M., Tanaka, T., Hashimoto, Y., Kuroda, M., Takagi, T., Nakamura, Y. (2002). JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res.* **30,** 158-162.

Hubbard, T.J., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coate, G., Cunningham, F., Cutts, T., Down, T., Dyer, S.C., Fitzgerald, S., Fernandez-Banet,J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Ouverdin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion,S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X.M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., Birney, E. (2007). Ensembl 2007. *Nucleic Acids Res.* **35**(Database issue):D610-7.

Janeway CA Jr. (1992). The immune system evolved to discriminate infectious nonself from noninfectious self. *Immunol Today.* **13**, 11-16.

Janeway, C.A. Jr and Medzhitov, R. (2002). Innate immune recognition. *Annu Rev Immunol.* **20,** 197-216.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292,** 195-202.

Kadowaki, N., Hoa, S., Antonenkoa, S., de Waal Malefyta, R., Kasteleina, R., Bazana, F., Liua, Y.-J. (2001). Subsets of human dendritic cell precursors express different Toll-like receptors and respond to different microbial antigens *J. Exp. Med.* **194,** 863-870

Kambris, Z., **Hoffmann, J.A.**, **Imler, J.L.**, **Capovilla, M**. (2002). Tissue and stage-specific expression of the Tolls in *Drosophila* embryos. *Gene Expr Patterns*. **2**, 311-317.

Kang, D., Liu, G., Lundström, A., Gelius, E., Steiner, H. (1998). A peptidoglycan recognition protein in innate immunity conserved from insects to humans.

Proc Natl Acad Sci U S A. **95,** 10078-10082

Kent, W.J., Sugnet, C. W., Furey, T. S., Roskin, K.M., Pringle, T. H., Zahler, A. M., Haussler, D. (2002). The Human Genome Browser at UCSC. *Genome Res.* **12**, 996-1006

Khush, R.S., and Lemaitre, B. (2000). Genes that fight infection: what the *Drosophila* genome says about animal immunity. *Trends Genet.* **16,** 442-429.

Kiselev, S.L., **Kustikova, O.S.**, **Korobko, E.V., Prokhortchouk, E.B.**, **Kabishev, A.A.**, **Lukanidin, E.M., Georgiev, G.P**. (1998). Molecular cloning and characterization of the mouse *tag7* gene encoding a novel cytokine. *J Biol Chem.* **273,** 18633-18639.

Kragol, G., Lovas, S., Varadi, G., Condie, B.A., Hoffmann, R., Otvos, L. Jr. (2001). The antibacterial peptide pyrrhocoricin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. *Biochemistry* **40,** 3016-3026.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. **157,**105-32.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* **227,** 680-685.

Lane, A.B., Soodyall, H., Arndt, S., Ratshikhopha, M.E., Jonker, E., Freeman, C., Young, L., Morar, B., Toffie, L. (2002). Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. *Am J Phys Anthropol* **119,**175-185.

Lang, M-F., Schneider, A., Krüger, C., Schmid, R., Dziarski, R., Schwaninger, M. (2008). Peptidoglycan recognition protein S (PGRP-S) is regulated by NFκB. *Neurosci Lett.* **430,** 138-141

Larin, S.S., Korobko, E.V., Kustikova, O.S., Borodulina, O.R., Raikhlin, N.T., Brisgalov, I.P., Georgiev, G.P., Kiselev, S.L. (2004). Immunotherapy with autologous tumor cells engineered to secrete Tag7/PGRP, an innate immunity recognition molecule. *J Gene Med.* **6**, 798-808.

Lavine, M.D. and Strand, M.R. (2002). Insect hemocytes and their role in immunity. *Insect Biochem Mol Biol*. **32,**1295-1309.

Lazzaro, B.P., Sceurman, B.K., Clark, A.G. (2004). Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. *Science.* **303,**1873-1876..

Lehrer, R.I., Rosenman, M., Harwig, S.S., Jackson, R., Eisenhauer, P. (1991). Ultrasensitive assays for endogenous antimicrobial polypeptides. *J Immunol Methods*. **137,** 167-173

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., Hoffmann, J.A. (1996). The dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults.*Cell* **86**, 973-983.

Levashina, E.A., Langley, E., Green, C., Gubb, D., Ashburner, M., Hoffmann, J.A., Reichhart, J.M. (1999). Constitutive activation of toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science.* **285,**1917–1919

Ligoxygakis, P., Bulet, P., Reichhart, J.M. (2002). Critical evaluation of the role of the Toll-like receptor 18-Wheeler in the host defense of Drosophila. *EMBO Rep.* **3**, 666-673.

Ligoxygakis, P., Pelte, N., Hoffmann, J.A., Reichhart, J.M. (2002). Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science.* 297,114-116.

Liu, C., Xu, Z., Gupta, D., Dziarski, R. (2001). Peptidoglycan recognition proteins: a novel family of four human innate immunity pattern recognition molecules. *J Biol Chem* **276,** 34686–34694

Lombard, Z., Dalton, D.L., Venter, P.A., Williams, R.C., Bornman, L. (2006). Association of HLA-DR, -DQ, and vitamin D receptor alleles and haplotypes with tuberculosis in the Venda of South Africa. *Hum Immunol*. 6**7**, 643-654.

Lu, X., Wang, M., Qi, J., Wang, H., Li, X., Gupta, D., Dziarski, R. (2006). Peptidoglycan recognition proteins are a new class of human bactericidal proteins. *J Biol Chem* **281,** 5895–5907

Madigan, M.T., Martinko, J.M., Parker, J. (1997). Chapter 23.3: "Mycobacterium and Tuberculosis: Pathology of tuberculosis" in Biology of Microorganisms, 8th Edition. Prentice Hall International, London.

Malkov, S., Zivkovic, M.V., Beljanski, M.V., Zaric, S.D. (2005). Correlations of amino acids with secondary structure types: connection with amino acid structure *arXiv:q-bio/0505046v1 [q-bio.BM]* URL: http://arxiv.org/abs/q-bio/0505046

Means, T.K., Golenbock, D.T., Fenton, M.J. (2000).The biology of Toll-like receptors. Cytokine Growth Factor Rev. **11,** 219-232.

Medzhitov, R. and Janeway, C.A. Jr (1997)a. Innate immunity: the virtues of a nonclonal system of recognition. *Cell.* **91,** 295-298.

Medzhitov, R. and Janeway, C.A. Jr (1997)b. Innate immunity: impact on the adaptive immune response. *Curr Opin Immunol.* **9,** 4-9

Medzhitov, R., Preston-Hurlburt, P., Janeway, C.A. Jr (1997). A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature.* **388,** 394-397

Mellroth, P., Karlsson, J., Steiner, H. (2003). A scavenger function for a *Drosophila* peptidoglycan recognition protein *J. Biol. Chem.* **278,** 7059-7064.

Mellroth, P. and Steiner, H. (2006). PGRP-SB1: an N-acetylmuramoyl L-alanine amidase with antibacterial activity *Biochem Biophys Res Commun.* **350**, 994-999

Mellroth P, **Karlsson J**, **Håkansson J**, **Schultz N**, **Goldman, W.E., Steiner, H.** (2005). Ligand-induced dimerization of *Drosophila* peptidoglycan recognition proteins in vitro. *PNAS* **102**, 6455-6460.

Michel, T., Reichhart, J.M., Hoffmann, J.A., Royet, J. (2001). *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature* **414,** 756-759

Moiseyenko, V.M., Danilov, A.O., Baldueva, I.A., Danilova, A.B., Tyukavina, N.V., Larin, S,S., Kiselev, S.L., Orlova, R.V., Anisimov, V.V., Semenova, A.I., Shchekina, L.A., Gafton, G.I., Kochnev, V.A., Barchuk, A.S., Kanaev, S.V., Hanson, K.P., Georgiev, G.P. (2005). Phase I/II trial of gene therapy with autologous tumor cells modified with tag7/PGRP-S gene in patients with disseminated solid tumors: miscellaneous tumors. *Ann Oncol.* **16,** 162-168.

Muramatsu, M., Nagaoka, H., Shinkura, R., Begum, N.A., Honjo, T. (2007). Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. *Adv Immunol.* **94,**1-36.

Muzio, M. and Mantovani, A. (2001). Toll-like receptors (TLRs) signalling and expression pattern. *J Endotoxin Res.* **7**, 297-300.

O'Neill, L. (2000). The Toll/interleukin-1 receptor domain: a molecular switch for inflammation and host defence *Biochem. Soc. Trans*. **28**, 557–563

Oppenheim, J., Biragyn, A., Kwak, L., Yang, D. (2003).Roles of antimicrobial peptides such as defensins in innate and adaptive immunity *Ann Rheum Dis*. **62**(Suppl 2): ii17–ii21.

Otvos, L. Jr. (2002). The short proline-rich antibacterial peptide family. *Cell Mol Life Sci.* **59,** 1138-1150.

Ozinsky, A., Underhill, D.M., Fontenot, J.D., Hajjar, A.M., Smith, K.D., Wilson, C.B., Schroeder, L., Aderem, A. (2000). The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *PNAS* **97**,13766-13771

Pancer, Z. and Cooper, M.D. (2006). The evolution of adaptive immunity. *Ann. Rev. immunol.* **24,** 497-518

Powell, L.M., Wallis, S.C., Pease, R.J., Edwards, Y.H., Knott, T.J., Scott, J. (1987). A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell.* **50,** 831-840.

Rehman, A., Taishi, P., Fang, J., Majde, J.A., Krueger, J.M. (2001). The cloning of a rat peptidoglycan recognition protein (PGRP) and its induction in brain by sleep deprivation. *Cytokine* **13**, 8-17.

Reiser, J.B, Teyton, L., Wilson, I.A. (2004).Crystal structure of the *Drosophila* peptidoglycan recognition protein (PGRP)-SA at 1.56 Å resolution. *J Mol Biol.* **340**, 909-917.

Riva, A. and Kohane, I.S. (2002). SNPper: retrieval and analysis of human SNPs. *Bioinformatics.* **18,** 1681-1685.

Roach, J.C., Glusman, G., Rowen, L., Kaur, A., Purcell, M.K., Smith, K.D., Hood, L.E., Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. *PNAS* **102**, 9577-9582.

Rock, F.L., Hardiman, G., Timans, J.C., Kastelein, R.A., Bazan, J.F. (1998). A family of human receptors structurally related to *Drosophila* Toll *PNAS* **95,**588-593.

Rohatgi, A., Ayers, C.R., Khera, A., McGuire, D.K., Das, S.R., Matulevicius, S., Timaran, C.H., Rosero, E.B., de Lemos, J.A. (2008). The association between peptidoglycan recognition protein-1 and coronary and peripheral atherosclerosis: Observations from the Dallas Heart Study doi: 10.1016/*J. Artherosclerosis*.2008.07.015

Rolff, J. (2007). Why did the acquired immune system of vertebrates evolve? *Dev. Comp. Immunol*. **31**, 476-482.

Rosetto, M., Engström, Y., Baldari, C.T., Telford, J.L., Hultmark, D. (1995). Signals from the IL-1 receptor homolog, Toll, can activate an immune response in a *Drosophila* hemocyte cell line. *Biochem Biophys Res Commun* **209,**111-116

Salas, A., Richards, M., De la Fe, T., Lareu, M.V, Sobrino, B., Sánchez-Diz, P., Macaulay, V., Carracedo, A. (2002). The making of the African mtDNA landscape. *Am J Hum Genet*. **71,** 1082-1111.

Samuel, C.E. (2003). RNA editing minireview series. *J Biol Chem*. **278,** 1389-90.

Sang, Y., Ramanathan, B., Ross, C.R., Blecha, F. (2005). Gene silencing and overexpression of porcine peptidoglycan recognition protein long isoforms: involvement in beta-defensin-1 expression. *Infect Immun.* **73,** 7133-7141.

Sashchenko, L.P., Dukhanina, E.A., Yashin, D.V., Shatalov, Y.V., Romanova, E.A., Korobko, E.V., Demin, A.V., Lukyanova, T.I., Kabanova, O.D., Khaidukov, S. V., Kiselev, S.L., Gabibov, A.G., Gnuchev, N.V., Georgiev, G.P. (2004). Peptidoglycan Recognition Protein Tag7 Forms a Cytotoxic Complex with Heat Shock Protein 70 in Solution and in Lymphocytes
J. Biol. Chem. **279,** 2117-2124.

Scheffers, D.J. and Pinho, M.G. (2005). Bacterial cell wall synthesis: new insights from localization studies. *Microbiol Mol Biol Rev.* **69,**585-607.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308-311

Soderhall, K. and Cerenius, L. (1998). Role of the prophenoloxidase-activating system in invertebrate immunity. *Curr Opin Immunol.* **10**, 23-28.

Sommer, B., Köhler, M., Sprengel, R., Seeburg, P.H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell.* **67,** 11-19.

Sommers, D.K., Moncrieff, J., Avenant, J.C. (1991). Absence of polymorphism of sparteine oxidation in the South African Venda. *Hum Exp Toxicol.* **10,**175-178.

Steiner, H., Hultmark, D., Engström, A., Bennich, H., Boman, H.G. (1981) Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* **292**, 246-248.

Sun, C., **Mathur, P.**, **Dupuis, J.**, **Tizard, R.**, **Ticho, B.**, **Crowell, T.**, **Gardner, H.**, **Bowcock, A.M.**, **Carulli, J**. (2006). Peptidoglycan recognition proteins Pglyrp3 and Pglyrp4 are encoded from the epidermal differentiation complex and are candidate genes for the Psors4 locus on chromosome 1q21 *Hum Genet.* **119**, 113-125

Tauszig, S., Jouanguy, E., Hoffmann, J.A., Imler, J.L. (2000).Toll-related receptors and the control of antimicrobial peptide expression in *Drosophila. PNAS* **97**, 10520-10525.

Tian C, Hinds D.A., Shigeta, R., Kittles, R., Ballinger, D.G., Seldin, M.F. (2006). A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. *Am J Hum Genet* **79,** 640-649.

Thompson, C.B. (1995). New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* **3**, 531-539.

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673-4680.

Tydell, C.C., Yuan, J., Tran, P., Selsted, M.F. (2006). Bovine peptidoglycan recognition protein-S: antimicrobial activity, localization, secretion, and binding properties *J. Immunol*. **176,** 1154-1162

UNAIDS (2002). Report on the global HIV/AIDS epidemic, July.

Vasselon, T. and Detmers, P.A. (2002). Toll receptors: a central element in innate immune responses. *Infect Immun*. **70**,1033-1041

Voet, D. and Voet J.G. (1995). Chapter 4: "Amino Acids" in Biochemistry 2nd Edition pg 58-59. John Wiley and Sons Inc, New Jersey

Wang, M., Liu, L.H., Wang, S., Li, X., Lu, X., Gupta, D., Dziarski, R. (2007). Human peptidoglycan recognition proteins require zinc to kill both gram-positive and gram-negative bacteria and are synergistic with antibacterial peptides. *J Immunol* **178,** 3116-3125.

Werner, T., Liu, G., Kang, D., Ekengren, S., Steiner, S., Hultmark, D. (2000). A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster PNAS* **97,**13772-13777.

Yan, H., Hancock, R.E. (2001). Synergistic interactions between mammalian antimicrobial defense peptides. *Antimicrob Agents Chemother.* **45,**1558-1560.

Yang, D., Chertov, O., Oppenheim, J.J. (2001). The role of mammalian antimicrobial peptides and proteins in awakening of innate host defenses and adaptive immunity. *Cell. Mol. Life. Sci.* **58,** 978-989

Xu, M., **Wang, Z.**, **Locksley, R.M.** (2004). Innate immune responses in peptidoglycan recognition protein L-deficient mice. *Mol Cell Biol.* **24**, 7949-7957

Xu. Q., Wang, H., Zhang, L., Liu, N. (2006). Sodium channel gene expression associated with pyrethroid resistant house flies and German cockroaches. *Gene*. **379**, 62-67

Yoshida, H., Kinoshita, K., Ashida, M. (1996).Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori.* *J Biol Chem.* **271,**13854-13860.

APPENDIX

I) ETHICAL CLEARANCE, SUBJECT AND PATIENT INFORMATION

Subject Information Sheet

Hello

We are Drs Ntwasa and Sarkin and we are doing a study to see why certain people are prone to infection by TB and HIV. This study may benefit medical practice in the future because it could contribute to our understanding of how we respond naturally to infections and how we can devise means to protect ourselves from germs. To do this we need a small amount of blood from you (10ml – about two teaspoons). This blood will then be examined by our team of researchers. We will only use your blood to test for the differences or variations in the peptidoglycan recognition protein or PGRP gene. These variations may affect the function of an important cell component called Toll. Toll and PGRP are present in all human beings but slight differences in these cell components may occur from person to person. We would like to investigate this because these slight variations may prove to be useful in identifying people who are prone to developing HIV and TB infections.

The results of tests done on your blood sample will be kept confidential and will only be used by this select group of researchers. The results of these test are still investigational and therefore at this stage cannot be used to alter your treatment. There will be no personal benefits to you. You are free to choose whether you wish to participate in this study. You are not under any pressure to take part. You may also withdraw at any time you wish if you feel you no longer want to be part of the study.

The removal of blood will be done during routine examination in the hospital. This means an extra tube will be taken when other samples are done as part of care in the hospital. There is no risk to this procedure, only a little pain (the prick of a needle) and perhaps a small bruise. The procedure will be performed by a qualified hospital staff member.

We will also ask you to complete a short questionnaire which asks simple questions such as your language, place of birth and that of your family.

We thank you very much for showing interest

PATIENT QUESTIONAIRE

NB: A translation is possible when required

We are studying cells in patients with infections. This may help science to find out why certain patients have these diseases. Your doctor will continue to explain the nature of your medical conditions. All information will be strictly confidential.

Patient Number _____e.g. 1-50

YOU

Place of birth Province Country
Home language

YOUR PARENTS

Your mother
Place of birth_____
Province_____Country

Your Father
Place of birth_____
Province_____Country_____

YOUR GRANDPARENTS

Your mother's mother
Place of birth_____
Province_____Country

Your Father's mother
Place of birth_____
Province_____Country_____

Your mother's father
Place of birth_____
Province_____Country

Your Father's father
Place of birth_____
Province_____Country_____

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Ntwasa et al

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER M040242**
PROJECT	Peptidoglycan recognition protein (PGRP) polymorphisms in TB and AIDS patients.
INVESTIGATORS	Dr M Ntwasa et al
DEPARTMENT	Molecular & Cell Biology
DATE CONSIDERED	04.02.27
DECISION OF THE COMMITTEE*	Approved unconditionally

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 04.04.26 **CHAIRPERSON** .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable

cc: Supervisor : Dr M Ntwasa

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10005, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

II) MATERIALS AND METHODS

Table7: Reagents for DNA techniques:

50X TAE:	2M Tris, 5.7% glacial acetic acid, 0.5M EDTA pH 8.0
Solution 1:	50mM glucose. 25mM Tris, 10mM EDTA pH 8.0
Solution 2:	0.2M NaOH, 1% SDS (w/v)
Solution 3:	3M $NaCH_3COO$
Ethanol acetate:	96% (v/v) absolute ethanol, 4% (v/v) 3M $NaCH_3COO$

Table 8: Reagents for Protein Electrophoresis

SDS-PAGE sample buffer: 0.225M Tris-Cl (pH 6.8), 50% glycerol, 5% SDS, 05% bromophenol blue, 0.25mM DTT
SDS-PAGE electrophoresis buffer: 0.5M Tris Base, 1/92M glycine, 0.5% SDS
Coomassie blue stain: 0.1% Coomasie blue R-250 (w/v), 10% acetic acid, 40% ethanol
Destain solution: 40% methanol, 10% acetic acid

Table 9: Reagents for protein extraction and purification

Cell lysis buffer:	25 mM Tris, 150 mM NaCl, 0.2% SDS, 0.5 mM PMSF (Sigma), 1 μg/ml aprotinin, 4.0 mM Pefabloc (Roche)
RIPA buffer:	1x PBS, 1% Nonidet P-40, 0.5% sodium dioxycholate, 0.1% SDS
GST purification kit:	BugBuster protein extraction reagent, and 10, 000 units benzonase
GST bind wash buffer (10x):	43 mM Na_2HPO_4, 14.7 mM KH_2PO_4, 1.37 M NaCl, 27 mM KCl pH 7.3
GST elution buffer:	500mM Tris-HCl pH 8.0100mM glutathione
GST mag agarose beads:	50% slurry, 0.05 M Na_2HPO_4 pH 7.5, 0.15 M NaCl, 0.1 M NaN_3
Protease inhibitor stocks	10 mg/ml PMSF (Sigma), 20 mg/ml Pefabloc (Roche), 1mg/ml Apportionin (Roche)

Table 10: Lists of kits and suppliers

KIT	SUPPLIER	CAT NUMBER
QIAmp DNA Blood midi Kit	Qiagen	51183
QIAmp RNA Blood mini Kit	Qiagen	52304
Acess RT-PCR kit	Promega	A1250
p-GEM-T Easy Vector system	Promega	A1360
GenElute™ Gel Extraction kit	Sigma	NA1111
SyBr® Gold nucleic acid stain	Invitrogen	S-11494
BL21(DE3)pLysS competent cells	Novagen	69451-3
GST-Bind™ Purification kit	Novagen	70794-3
GeneRuler™100 basepair plus	Fermentas	SM0321
PageRuler™ Prestained protein ladder	Fermentas	SM0671

Table 11: List of specialised equipment

GeneAmp® PCR system 2004	Applied Biosystems
Dark® reader	Clare chemical research
G: Box Chemi XL gel documentation system	Syngene
GS-800 calibrated scanning densitometer	Biorad

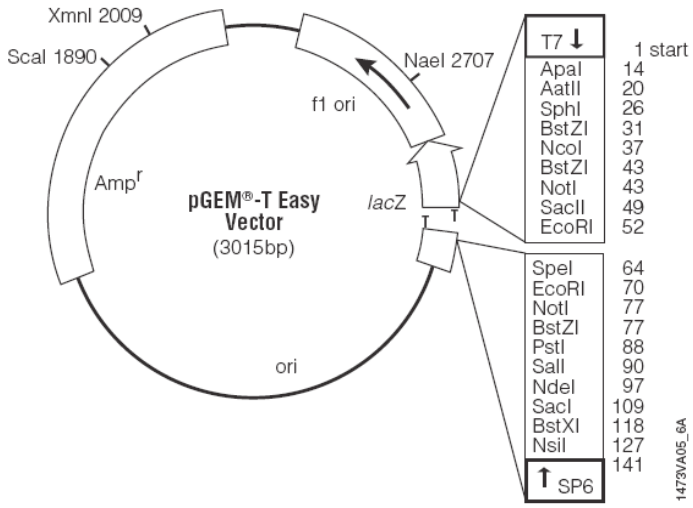

Figure 21: pGEM®-T Easy Vector circle map and sequence reference points.

T7 RNA polymerase transcription initiation site	1
multiple cloning region	10–128
SP6 RNA polymerase promoter (–17 to +3)	139–158
SP6 RNA polymerase transcription initiation site	141
pUC/M13 Reverse Sequencing Primer binding site	176–197
lacZ start codon	180
lac operator	200–216
β-lactamase coding region	1337–2197
phage f1 region	2380–2835
lac operon sequences	2836–2996, 166–395
pUC/M13 Forward Sequencing Primer binding site	2949–2972
T7 RNA polymerase promoter (–17 to +3)	2999–3

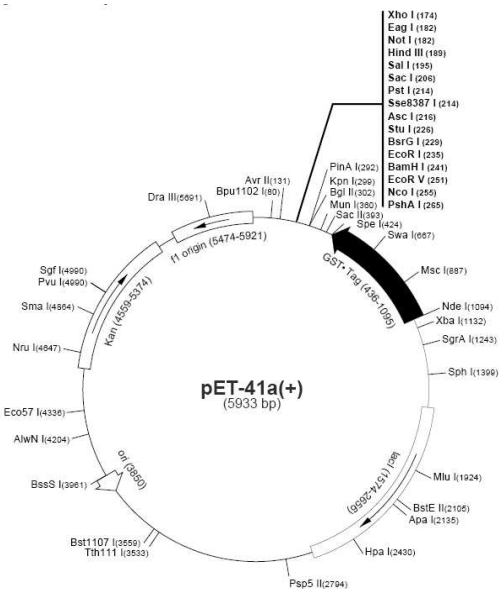

Figure 22: pET-41a Vector circle map and sequence reference points

T7 promoter	1167–1183
T7 transcription start	1166
GST•Tag coding sequence	436–1095
His•Tag coding sequence	397–414
S•Tag coding sequence	310–354
Multiple cloning sites	
(PshAI-XhoI)	174–265
His•Tag coding sequence	150–173
T7 terminator	26–72
lacI coding sequence	1574–2656
pBR322 origin	3850
Kan coding sequence	4559–5374
F1 origin 5474–5921	

URLs for databases:

Reference sequences and data mining

BLAST: http://www.ncbi.nlm.nih.gov/blast/Blast.cgi

Clustal W: http://www.ebi.ac.uk/Tools/clustalw2/index.html

SNP verification

dbSNP: http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp&cmd=search&term=

Ensembl: http://www.ensembl.org/index.html

jSNP: http://snp.ims.u-tokyo.ac.jp/

SNPer: http://snpper.chip.org/.

UCSC genome brower: http://genome.ucsc.edu/

Proteomics analysis

RCSB protein database: http://www.rcsb.org/pdb

PsiPred secondary structure prediction server: http://bioinf.cs.ucl.ac.uk/psipred/

Swiss PDB viewer: http://www.expasy.ch/spdbv/

III) Results

>ref|NT_011109.15|Hs19_11266:c18795026-18790173 Homo sapiens chromosome 19 genomic contig, reference assembly

AGGCTTGGACCTGGGGGAAGGAAGTGTTTTTAAAATGGGTAGCTTTGTTACCTCAGTCCTTGCGGGGAGA
AGATGTGCCAATGCCCCGGGCCTCTGCGATGTAAGTTCCTGTTGAGTTGTTCAATGTTCCACTGAGAGGA
GAGACAAGCAGGACCGGGATCCCGGCCCCGGGTCTTCCAGAAACGGCGGGGCGAAGCGCCCAGGGGCCTG
TGCGTCCCTCCCTGCTGAGCGAGGGGGCCTGTCATTGCCGTGGGCGTGACCCAGACCCCAACCACAGTGC
ATCCCGCCCTGGCCCAGCCAGAGAAGGAAGCTGAGTCTGGGGTCTGCTGGGCCAGCAGGAAGTCCCAGCA
GGGTGTGAAGCAAGACTTTCCGGGCCACTCCTGGAATCCCCCAGCAGATAAAGGCGGCCCCTCCAC**CGGG**
CGCTCCTAGCGGTCTCCCGGACCCTGCCGCCCTGCCACTATGTCCCGCCGCTCTATGCTGCTTGCCTGGG
CTCTCCCCAGCCTCCTTCGACTCGGAGCGGCTCAGGAGACAGAAGACCCGGCCTGCTGCAGCCCCATAGT
GCCCCGGAACGAGTGGAAGGCCCTGGCATCAGAGTGCGCCCAGCACCTGAGCCTGCCCTTACGCTATGTG
GTGGTATCGCACACGGCGGGCAGCAGCTGCAACACCCCCGCCTCGTGCCAGCAGCAGGCCCGGAATGTGC
AGCACTACCACATGAAGACACTGGGCTGGTGCGACGTGGGCTACAAGTGAGTGTGGCAGGGGGGCACGGG
CTGGACTGGGGATCCCAGAGGCAAAGGTGTGCAAGTGCCCGCCAGCGGGACCATTTCTAGGCCTGGTGCT
GAGGCCACCTGGCTGACACCCTGAGCAGCTTACTGAACCTCACAAGGTCATTTCTTCCTCGGGTCACCAC
GTGGGCCGGGGCAGGTGCTCACAGCACAAAGATGCGTGAGAAGGGGGGATGGGCTGAAATCCAAGCCCCG
CTCCCGCTCCCGGGAAGAAGGGGCAGGCTTTGCCTAACACACACCAAGGTGTCCGCGACGCCTTATGAAG
GCGCCCCTTCCTGATGCCTGGTGCCTTTGTTAGGAGGCGGAAAACCAGGCTGACTCAGTGCGGAGCTGGG
TTCCAGGGTCAGCTCATCTTATTGCTGGCTGAGGCCTGGGGCGGGTCCTCCCGGCTTCAGCTTCCTCCTC
GGCAAAGTCAATGGAATCGCGGCCCATCTGCTCCTTATATGAGCCAATACAGGCAAAGCACCTAGGAACT
GCCTGGCACACGGGGGCCCGCAAAGCCACGGGCTATTACCATTAGGCACCTACTATGTGCCAGGCAGTGT
GCCAGATGTGGCATGGAGAACGGTGCCACTCTAGCCTCCTGGGGTTGAAAGACCAGCAGGGAGAAACAAG
CCAAGGCCACAGGTGAATTCAGACGGTGATAGAGGCGATGGATAAAAGGAAAGAGGTGAGGGAGGGTGAC
CTCAGGCTGGGACCAGACCCACACAGGCTCCTTGGAAGGGTGGGGCAAGGGAAAGCCTCCCTGAGAGTGG
CTGCCTTGGCTCTGGGATGGGGGAGCAGACAGAGGCCTGCAGGTGGCAGGGGGTGCTGGGGGCAAAGCCA
ACGGGGTTGAGTGGGCGGGTCCTGTGGGAAGGACAGGGTGGCTGAACACAGCCGGGCGAGCGGAATGAAA
TGAGGAAAGGGCAGGACCAGGCAGGGCTTTGTGGACCACAGTGAGGACTGTAGCTTTTACTCCGAGATGG
GAGCCATGGGAGGCTTCCCCCCGCAGAGCAGGGACCAGGGCTAGTGTAGGTACAAGAGACTCCCTCTGGC
TGCTGTGGGGAGAACAGACTGTAAGGGGTGCCGGCTGAAGCAGGCTGAAGCAGGTCCGGGGCCACGCCTA
CCAGCTACCTCCTCCTTAGAATGGGCTGCAGTCCAGCCTGACGGTTTCCACAGGGAGGGAACTTTCAATC
CTGCAGGCTCTCAGACCCAGGTAAGCATCAGAATCTCCTGCAGAGCTTATTGGAGGAACATGCAGGTCCC
CAGAGACCACTTCAATAGACCCCACCCAGGGGACACATGTTCTCAGGACCTCCTAAGGCTGTGTCTCACA
CACACACACACACAAAACCCACCCAGAGAAGTCTAGAAGTGTCCAAAGGTATTCCCAGCCCTTCCAGTTC
AAGATTTTATCAAAGACTCCAACTTCAACATTTTGTTTCCAAGACTCAGGCGCTCTGATCCTAAGTGTTC
CCATGAGGTCACGAAGGCCCGAGCCTCGGGTCCACATCACCACATACAGAAGCAGCCGCTGTCTGGCCCT
CAGCTTTTAGGATTCCGAATCTGGTTCTAACGTTCTGAATCCCCTTTCCATGTCACGTCACCTTCTCTCT
CTGAGCCTCGTTTCTTCACCTGTAAATGGGGCTTCTGACTGTACCCGCTTGGCGATGGTGGCTGAGGTTT
CAATGAGATGCATTCATGCAACAAATGACCAGGGAGAACCAACCACCCCCGACCCCAGCACCATCATGCC
AGGCTCCGGGAATATGGCAGGGCCATAACCAGCCAGTCCCAGCAGCCAGCATCACCAGCAGGGCCATAAC
CAGCCAGTCCCAGCAGCCAGCATCACTGGGGCCAGATTCTGGGAGCTGGCTTCCTAAGCAGCAGATTCGG
GACTCAAGATTCGGAAGTTCCGACATTCTGTGCGTCATTAATGACACTGTGAGCGCGGGGGCTGAGTTGG
CACATACTGCAGCTCCAAGGCACTGATCTTCTAGTTCATTCCCATTTGCCAAGTCAGACACGCTCAATCA
GATGTGCTGAATGAATCAGCCTTAAGGATCTGCTGCCAGGCTTGGGAACCTCAAAGAGGCAGGTGCTGGG
GCTCCAGGGTCGGTATCTGGTTACGACCCCAACTTTGAGTCTGTGATCCTGTCCCCCAGTGGACCCCGAG
GGCAGGAAGCAAGCTGGGGCAAAGGCAGAGTTCTGAGAGGCAGGGGGCTCTGAGTAGCTTGTACAAATTA
TGTCGCTTTCTGAGTCTCTGCTCTCTCTGCTTAAGATGAGGATGATAAGAGTCCCTGCCTTGGCTGGGCA
CGGTGACTCATGCTTGTAATCCCAGCACTTTGGGAGGCCAAAGGCGGGCAGATCACCTGAGGTCGGGAGT
TCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAATAATACAAAAATTAGCCGGGTGTGGT
GGCTCATGCTTGTAATCCCAGGTACTCAGGAGGCTGAGGCACGAGAATCACTTGAACCCAGGAGGCAGAG
GTTGCGGTGAGCTGAGATCACGCCACTGCACTCCAGCCTGAGCAACAAGAGTGAAACCCCGTCTTAAAAA
AAAAAAAAAAAAAAAGTCTCTGCCTCAGAGGGCTGTCCTGCAAGGATTCAGTAAAACAGGACTTCCACAA
ATCCTCTAGGGAGGCAGGAGAGAGCAACGCACAAGATCACATGGGTGGGGACCATGGAGGCCTGGTTAAA
AACCCCAGCTGTGCCCCAGACAGGCTGTGTGACCCCGGGGCAGGTGATTCTACCGCTCTGATTCTGTCTA
TATAAAACGGGGGTGATAGTGACTCACAGGGTTGCTGTGAGGATGGAAAAAGGTGTATGCAGAGCTGTGC
TTGGCACATGGCAGGGGCCCAATAAATGACCACAGTCATTGTTACTATCGTGGGAAATTAACGGGTGAGA
AGGATGCTGGGGAAGTTGTGTGCACCACGGCAGTAACGGAGGCAGGGTGAGTGGAGGAAGGAGGAGGAGT
CCTCCTCGGGAAACCCTCCCTCACTCATCAAGCCCCCCACCTCCCCATGCAG**CTTCCTGATTGGAGAAG**
ACGGGCTCGTATACGAGGGCCGTGGCTGGAACTTCACGGGTGCCCACTCAGGTCACTTATGGAACCCCAT
GTCCATTGGCATCAGCTTCATGGGCAACTACATGGGTGAGTGACTGTCCAGCCAGTTGGGATGGGGGCTG
ACACGGGGCCCTACGCTATCACTTCCCCTTCGGGGAAGCCCCTGCCTGACTCCTACCTCTGCCT**CCTCCA**
GATCGGGTGCCCACACCCCAGGCCATCCGGGCAGCCCAGGGTCTACTGGCCTGCGGTGTGGCTCAGGGAG

CCCTGAGGTCCAACTATGTGCTCAAAGGACACCGGGATGTGCAGCGTACACTCTCTCCAGGCAACCAGCT
CTACCACCTCATCCAGAATTGGCCACACTACCGCTCCCCCTGAGGCCCTGCTGATCCGCACCCCATTCCT
CCCCTCCCATGGCCAAAAACCCCACTGTCTCCTTCTCCAATAAAGATGTAGCTCAAAATGTGTTCCTCAG
CAGGTCCCACCACACACCTCCCCTCCTTCTCTCCCTCCCACAAGACCAAAGCCCTCCCACCCCTTCCCCT
CCCAGCACTCTCCTCACTGGTGCTTGGAACCCAGAGGCCCCGGCCCTGAACCAAGACCCCGGAGATCCCA
GCTCGAGAGCAACATCACAGCTCAGAACCCAGAGGTCTCAGGTCAGAGCCCAGGAATCAGCACCAAGGCG
GCCGGGATGTCACCAACAGTGGGTCCTGGGGGTCCTCCTGGGCAGTGAACCACGCTGCAAGACACCTCCT
TCCCACCACACAGCCAAAGGCATCCTGCGCAGAAACTTTGCTTGGACTCCTTTAGGGACGGGCTGCTCAC
TACTCCTCCCAGGTCTGCCCTTGCACTGTTTAGAGAGTGTACAAGCTTTTCCTAGGCCTGAGTCCCAATC
TGCCTCCGAGGCACAGACACACAG

Figure 23: Genomic sequence of PGRP-S. Exons are highlighted in bold and primer binding sites indicated

AF076483 Homo sapiens peptidoglycan recognition protein precursor (PGRP) mRNA, complete cds

CCGGGCGCTCCTAGCGGTCTCCCGGCCCCTGCCGCCCTGCCACTATGTCCCGCCGCTCTATGCTGCTTGC
CTGGGCTCTTCCCAGCCTCCTTCGACTCGGAGCGGCTCAGGAGACAGAAGACCCGGCCTGCTGCAGCCCC
ATAGTGCCCCGGAACGAGTGGAAGGCCCTGGCATCAGAGTGCGCCCAGCACCTGAGCCTGCCCTTACGCT
ATGTGGTGGTATCGCACACGGCGGGCAGCAGCTGCAACACCCCCGCCTCGTGCCAGCAGCAGGCCCGGAA
TGTGCAGCACTACCACATGAAGACACTGGGCTGGTGCGACGTGGGCTACAACTTCCTGATTGGAGAAGAC
GGGCTCGTATACGAGGGCCGTGGCTGGAACTTCACGGGTGCCCACTCAGGTCACTTATGGAACCCCATGT
CCATTGGCATCAGCTTCATGGGCAACTACATGGATCGGGTGCCCACACCCCAGGCCATCCGGGCAGCCCA
GGGTCTACTGGCCTGCGGTGTGGCTCAGGGAGCCCTGAGGTCCAACTATGTGCTCAAAGGACACCGGGAT
GTGCAGCGTACACTCTCTCCAGGCAACCAGCTCTACCACCTCATCCAGAATTGGCCACACTACCGCTCCC
CCTGAGGCCCTGCTGATCCGCACCCCATTCCTCCCCTCCCATGGCCAAAAACCCCACTGT

Figure 24: mRNA sequence for PGRP-s, with areas that will demarcate into exos highlighted

DNA Sequences for mutant Patients

>Patient 6 (HIV)

```
1     GGCGCTCCTAGCGGTCTCCCGGCCCCTGCCGCCCTGCCACTATGTCCCGCCGCTCTATGC
61    TGCTTGCCTGGGCTCTCCCCAGCCTCCTTCGACTCGGAGCGGCTCAGGAGACAGAAGACC
121   CGGCCTGCTGCAGCCCCATAGTGCCCCGGAACGAGTGGAAGGCCCTGGCATCAGAGTGCG
181   CCCAGCACCTGAGCCTGCCCTTACGCTATGTGGTGGTATCGCACACGGCGGGCAGCAGCT
241   GCAACACCCCCGCCTCGTGCCAGCAGCAGGCCCGGAATGTGCAGCACTACCACATGAAGA
301   CACTGGGCTGGTGCGACGTGGGCTACAGCTTCCTGATTGGAGAAGACGGGCTCGTATACG
361   AGGGCCGTGGCTGGAACTTCACGGGTGCCCACTCAGGTCACTTATGGAACCCCATGTCCA
421   TTGGCATCAGCTTCATGGGCAACTACATGGATCGGGTGCCCACACCCCAGGCCATCCGGG
481   CAGCCCAGGGTCTACTGGCCTGCGGTGTGGCTCAGGGAGCCCTGAGGTCCAACTATGTGC
541   TCAAAGGACACCGGGATGTGCAGCGTACACTCTCTCCAGGCAACCAGCTCTACCACCTCA
601   TCCAGAATTGGCCACACTACCG
```

>Patient 12 (HIV/TB)

```
1     GGTCTCCCGGCCCCTGCCGCCCTGCCACTATGTCCCGCCGCTCTATGCTGCTTGCCTGGG
61    CTCTCCCCAGCCTCCTTCGACTCGGAGCGGCTCAGGAGACAGAAGACCCGGCCTGCTGCA
121   GCCCCATAGTGCCCCGGAACGAGTGGAAGGCCCTGGCATCAGAGTGCGCCCAGCACCTGA
181   GCCTGCCCTTACGCTATGTGGTGGTATCGCACACGGCGGGCAGCAGCTGCAACACCCCCG
241   CCTCGTGCCAGCAGCAGGCCCGGAATGTGCAGCACTACCACATGAAGACACTGGGCTGGT
301   GCGACGTGGGCTACAACTTCCTGATTGGAGAAGACGGGCTCGTATACGAGGGCCGTGGCT
361   GGAACTTCACGGGTGCCCACTCAGGTCACTTATGGAACCCCATGTCCATTGGCATCAGCT
421   TCATGGGCAACTACATGGATCGGGTGCCCACACCCCAGGCCATCCGGGCAGCCCAGGGTC
481   TACTGGCCTGCGGTGTGGCTCAGGGAGCCCTGAGGTCCAACTATGTGCTCAAAGGACACC
541   GGGATGTGCAGCGTACACTCTCTCCAGGCAACCAGCTCTACCACCTCATCCAGAATTGGT
601   CAC
```

>Patient 15 (HIV)

```
1     CTATGTCCCGCCGCTCTATGCTGCTTGCCTGGGCTCTCCCCAGCCTCCTTCGACTCGGAG
61    CGGCTCACGAGACAGAAGACCCGGCCTGCTGCAGCCCCATAGTGCCCCGGAACGAGTGGA
121   AGGCCCTGGCATCAGAGTGCGCCCAGCACCTGAGCCTGCCCTTACGCTATGTGGTGGTAT
181   CGCACACGGCGGGCAGCAGCTGCAACACCCCCGCCTCGTGCCAGCAGCAGGCCCGGAATG
241   TGCAGCACTACCACATGAAGACACTGGGCTGGTGCGACGTGGGCTACAGCTTCCTGATTG
301   GAGAAGACGGGCTCGTATACGAGGGCCGTGGCTGGAACTTCACGGGTGCCCACTCAGGTC
361   ACTTATGGAACCCCATGTCCATTGGCATCAGCTTCATGGGCAACTACATGGATCGGGTGC
421   CCACACCCCAGGCCATCCGGGCAGCCCAGGGTCTACTGGCCTGCGGTGTGGCTCAGGGAG
481   CCCTGAGGTCCAACTATGTGCTCAAAGGACACCGGGATGTGCAGCGTACACTCTCTCCAG
541   GCAACCAGCTCTACCACCTCATCCAGAATTGGCCACACTACCGCTCCCCCTGAGGCC
```

>Patient 27 (HIV)

```
1     CCCTGCCGCCCTGCCACTATGTCCCGCCGCTCTATGCTGCTTGCCTGGGCTCTCCCCAGC
61    CTCCTTCGACTCGGAGCGGCTCAGGAGACAGAAGACCCGGCCTGCTGCAGCCCCATAGTG
120   CCCCGGAACGAGTGGAAGGCCCTGGCATCAGAGTGCGCCCAGCACCTGAGCCTGCCCTTA
180   CGCTATGTGGTGGTATCGCACACGGCGGGCAGCAGCTGCAACACCCCCGCCTCGTGCCAG
240   CAGCAGGCCCGGAATGTGCAGCACTACCACATGAAGACACTGGGCTGGTGCGATGTGGGC
300   TACAACTTCCTGATTGGAGAAGACGGGCTCGTATACGAGGGCCGTGGCTGGAACTTCACG
360   GGTGCCCACTCAGGTCACTTATGGAACCCCATGTCCATTGGCATCAGCTTCATGGGCAAC
420   TACATGGATCGGGTGCCCACACCCCAGGCCATCCGGGCAGCCCAGGGTCTACTGGCCTGC
480   GGTGTGGCTCAGGGAGCCCTGAGGTCCAACTATGTGCTCAAAGGACACCGGGATGTGCAG
540   CGTACACTCTCTCCAGGCAACCAGCTCTACCACCTCATCCAGAATTGGCCACACTACCGC
600   TCCCCCTGAGGCCCTGCTGATCCGCACCCCATTCCTCCCCTCCCATGGCCAAAAACCC
```

Table 7: Table of individual patient genotypes

Sample	Status	27	80	110	320	651
Patient 1	HIV/TB	CA	CC	GG	CT	CC
Patient 3	control	CA	CC	GG	CC	CC
Patient 4	control	CC	CC	GG	CT	CC
Patient 6	HIV	CC	CC	GG	CT	CT
Patient 7	HIV	CA	CC	GA	CC	CC
Patient 10	HIV/TB	CA	CC	GG	CC	CC
Patient 11	HIV/TB	CA	CC	GG	CT	CC
Patient 12	HIV/TB	CC	CC	GG	CC	CT
Patient 15	HIV	CA	CC	GG	CT	CC
Patient 16	HIV/TB	CC	CC	GG	CC	CC
Patient 18	HIV	AA	CC	GG	CC	CC
Patient 24	HIV	CA	CC	GA	CT	CC
Patient 25	HIV	CA	CC	GG	CC	CC
Patient 27	HIV	CA	CC	GA	CT	CC
Patient 29	HIV/TB	CA	CC	GG	CT	CC
Patient 30	HIV/TB	CC	CC	GG	CC	CC
Patient 31	control	CA	CC	GG	CC	CC
Patient 34	control	CA	CC	GA	CC	CC
Patient 36	control	CA	CC	GG	CC	CC
Patient 43	HIV	CC	CC	GG	CC	CC