Immunobiology 228 (2023) 152321

Contents lists available at ScienceDirect

Immunobiology

journal homepage: www.elsevier.com/locate/imbio

The PE-PPE Family of *Mycobacterium tuberculosis*: Proteins in Disguise

Christopher D'Souza [a], Uday Kishore [b], Anthony G. Tsolaki [a,*]

[a] Biosciences, Department of Life Sciences, College of Health, Medicine and Life Sciences, Brunel University London, Uxbridge UB8 3PH, United Kingdom
[b] Department of Veterinary Medicine, United Arab Emirates University, Al Ain, United Arab Emirates

ARTICLE INFO

Keywords:
Tuberculosis
Mycobacterium tuberculosis
PE
PPE
Structure
Secretion
Function

ABSTRACT

Mycobacterium tuberculosis has thrived in parallel with humans for millennia, and despite our efforts, *M. tuberculosis* continues to plague us, currently infecting a third of the world's population. The success of *M. tuberculosis* has recently been attributed, in part, to the PE-PPE family; a unique collection of 168 proteins fundamentally involved in the pathogenesis of *M. tuberculosis*. The PE-PPE family proteins have been at the forefront of intense research efforts since their discovery in 1998 and whilst our knowledge and understanding has significantly advanced over the last two decades, many important questions remain to be elucidated.

This review consolidates and examines the vast body of existing literature regarding the PE-PPE family proteins, with respect to the latest developments in elucidating their evolution, structure, subcellular localisation, function, and immunogenicity. This review also highlights significant inconsistencies and contradictions within the field. Additionally, possible explanations for these knowledge gaps are explored. Lastly, this review poses many important questions, which need to be addressed to complete our understanding of the PE-PPE family, as well as highlighting the challenges associated with studying this enigmatic family of proteins.

Further research into the PE-PPE family, together with technological advancements in genomics and proteomics, will undoubtedly improve our understanding of the pathogenesis of *M. tuberculosis*, as well as identify key targets/candidates for the development of novel drugs, diagnostics, and vaccines.

1. Introduction

Tuberculosis (TB) is a disease caused by *Mycobacterium tuberculosis* (*M. tuberculosis*), which currently infects a third of the world's population. According to the World Health Organisation's (WHO) Global Tuberculosis Report 2022, ~10.6 million people became newly infected with TB in 2021 (~9.9 million in 2020). The WHO also reported ~1.4 million TB deaths (~1.3 million in 2020) and an additional ~187,000 deaths in those co-infected with human immunodeficiency virus 1 (HIV-1) (~214,000 in 2020), thereby bringing figures back up to those seen in 2017. The COVID-19 pandemic reduced the capacity of healthcare services to provide care due to the overwhelming number of COVID-19 cases, as well as the capacity of patients to seek care due to lockdowns and the risk of COVID-19 infection. As a result, the COVID-19 pandemic has reversed years of progress in reducing the disease burden of TB (World Health Organization, 2022). The United Nations Sustainable Development Goals and WHO EndTB Strategy targets for 2030 include an 80 % reduction in TB incidence and a 90 % reduction in TB deaths compared to 2015, as well as a reduction in catastrophic financial costs (World Health Organisation, 2015). However, the current numbers suggest that these targets are unlikely to be achieved without significant innovation (World Health Organization, 2022). This should include a treatment regimen that is simpler and shorter, a vaccine that reduces the risk of infection and reactivation of the disease, and a diagnostic test that is rapid, accurate and suitable for use at the point of care. In addition, addressing non-specific issues such as poverty will also help achieve the targets for 2030.

The success and complexity of *M. tuberculosis* lie within its genome, which was first published in 1998 and found to contain 4,411,529 base pairs and ~4000 genes (Cole et al., 1998). Amongst the ~4000 genes, a

* Corresponding author.
E-mail addresses: anthony.tsolaki@brunel.ac.uk, agtsolaki@gmail.com (A.G. Tsolaki).

https://doi.org/10.1016/j.imbio.2022.152321
Received 28 June 2022; Received in revised form 22 December 2022; Accepted 27 December 2022
Available online 30 December 2022

novel collection of 168 genes were identified and designated the PE-PPE family. These genes were found to account for ~7–10 % of the *M. tuberculosis* genome coding capacity, thereby highlighting their significance and a potential source of targets/candidates for the development of novel drugs, diagnostics, and vaccines (Cole et al., 1998; Fishbein et al., 2015). PE-PPE family members have captivated researchers since their discovery and whilst our knowledge of this family has evolved and advanced tremendously, many important questions remain unanswered regarding their biological significance.

2. General characteristics of the PE-PPE family

Members of the PE-PPE family were designated based on the presence of either a conserved proline-glutamic acid (PE) or proline-proline-glutamic acid (PPE) motif, within the highly conserved N-terminal domain of the protein, which is ~100 or ~180 residues long, respectively. Whilst the N-terminal domains are conserved within each family, they are not homologous. In contrast, the C-terminal domains of these proteins are highly variable, both in terms of their sequence and length (Cole et al., 1998).

The N-terminal domain of PE proteins is ~100 amino acids in length and adopts a helix-turn-helix conformation. The PE residues are located within the N-terminal domain. The YXXXD/E type VII secretion signal is also located within the N-terminal domain, which is required for recognition by the type VII secretion system and its secretion over the inner mycobacterial membrane. These features are exclusive to PE proteins (Fig. 1a) (Strong et al., 2006; Daleke et al., 2012a).

The N-terminal region of PPE proteins is ~180 amino acids in length and adopts a helical bundle-like conformation. The PPE residues are located within the N-terminal region. The WxG motif is located within the second and third α-helix of the N-terminal region. In addition, the hydrophobic tip or hh motif is located within the 4th and 5th α-helix of the N-terminal region. These features are exclusive to PPE proteins (Fig. 1b) (Strong et al., 2006; Poulsen et al., 2014).

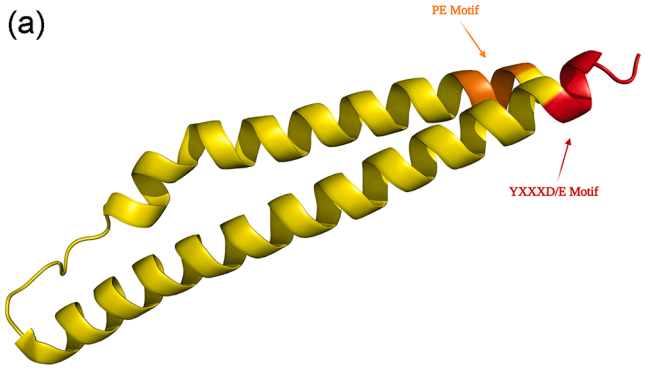

Fig. 1a. General Structure of the PE Family Members. The N-terminal regions of proline-glutamic acid (PE) proteins are ~100 amino acids in length and adopt an antiparallel helix-turn-helix conformation, as shown in yellow. The PE residues are shown in orange. The YXXXD/E type VII secretion signal is shown in red (PDB = 4KXR; Daleke et al., 2012a; Korotkova et al., 2014). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

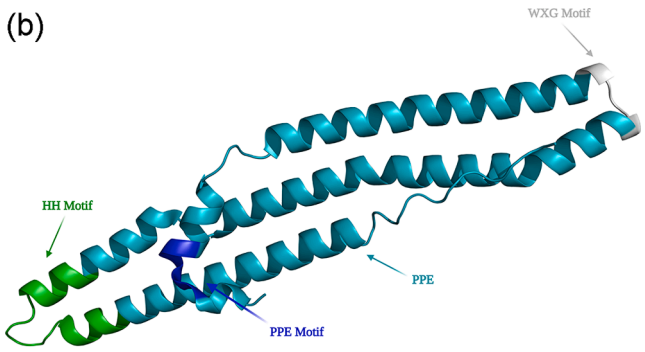

Fig. 1b. General Structure of the PPE Family Members. The N-terminal regions of proline-proline-glutamic acid (PPE) proteins are ~180 amino acids in length and adopt a helical bundle-like conformation consisting of 5 α-helices, as shown in cyan. The PPE residues are shown in blue. The WxG motif is located within the 2nd and 3rd α-helices, as shown in light grey. The hydrophobic tip or hh motif is located within the 4th and 5th α-helices, as shown in green (PDB = 4KXR; Poulsen et al., 2014; Korotkova et al., 2014). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Classification of the PE-PPE family

Given the significant amount of diversity within each family, members of each family can be further classified into sub-families based on their highly unique and variable C-terminal domains (Fig. 2) (Cole et al., 1998; Gey Van Pittius et al., 2006). The PE family can first be classified into the PE and PE-Polymorphic GC-Rich (PE-PGRS) sub-families. The PE-PGRS sub-family is the largest and accounts for 65 out of the 99 PE members. These members have a guanine and cytosine (GC) content of ~80 % and are rich in repeats of Gly-Gly-Ala/Gly-Gly-Asn. These members are also large in size, with some members reaching up to 1400 amino acids in length. Given these characteristics, the PE-PGRS members are prone to recombination events and are highly variable (Poulet and Cole, 1995; Cole et al., 1998; Gey Van Pittius et al., 2006). The PE family can also be classified into 5 sub-lineages based on phylogeny (Gey Van Pittius et al., 2006). Sub-lineages I and II are the smallest sub-lineages and consist of the ancestral PE members, with PE34 and PE35 belonging to I and PE5 and PE15 belonging to II. Sub-lineage III consists of PE22, PE25 and PE36. Sub-lineage IV contains the PE members secreted via the ESX-5 type VII secretion system. Sub-lineage V includes the PE-PGRS sub-family, as well as several other PE members that have unique C-terminal domains, such as lipases (Daleke et al., 2011).

The PPE family is classified into the PPE and PPE-Major Polymorphic Tandem Repeat (PPE-MPTR) sub-families. Members of the PPE-MPTR sub-family contain many repeats of an Asn-X-Gly-X-Gly-Asn-X-Gly motif. PPE-MPTR members are also significantly large, especially for secreted proteins, with lengths of over 3000 amino acids (Hermans et al., 1992; Cole et al., 1998; Gey Van Pittius et al., 2006). The PPE family can also be classified into 5 sub-lineages based on phylogeny (Gey Van Pittius et al., 2006). Sub-lineage I consists of PPE68. Sub-lineage II contains the 10 PPE-PPW members, characterized by the presence of a conserved PxxPxxW motif within their C-terminal domains. Sub-lineage III consists of PPE36, PPE41, PPE57, PPE58, PPE59 and PPE69. Sub-lineage IV is the largest sub-group, which contains 26 PPE-SVP members; these members are designated based on the presence of a conserved

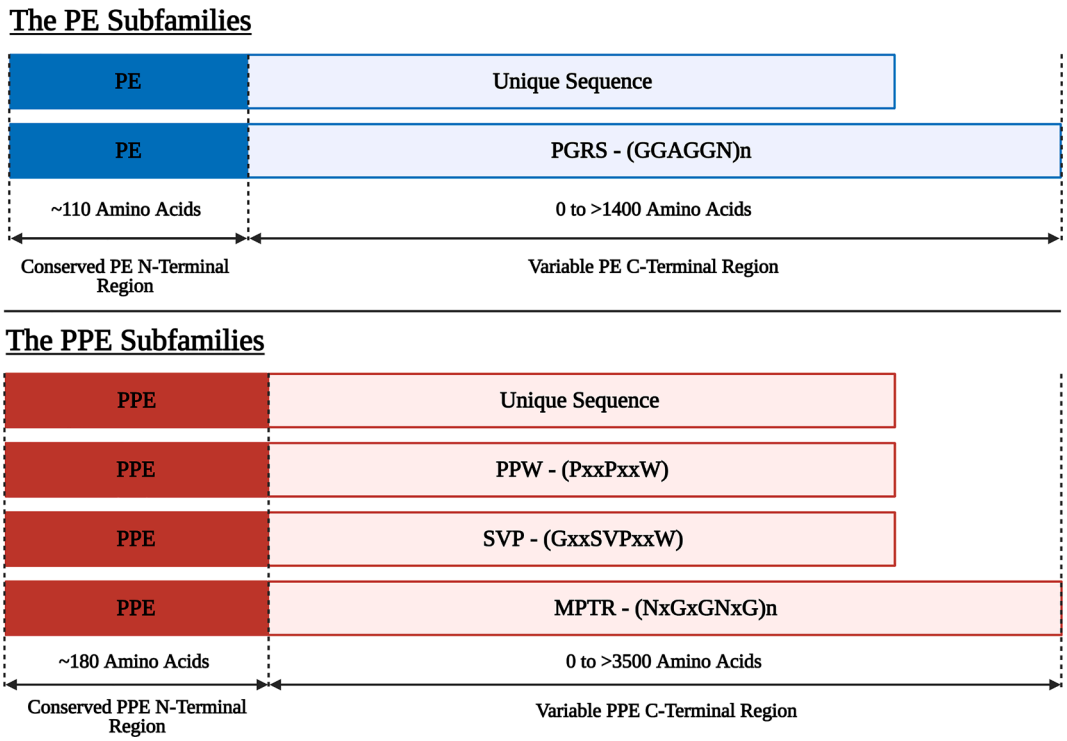

Fig. 2. Classification of the PE-PPE Family. A schematic representation of the members belonging to each of the proline-glutamic acid (PE)-proline-proline-glutamic acid (PPE) sub-families, including their conserved N-terminal regions and distinct C-terminal regions/motifs, as well as their respective positions and lengths (Gey Van Pittius et al., 2006).

GxxSVPxxW motif within their C-terminal domains. Sub-lineage V contains the PPE-MPTR sub-family members.

4. Evolution of the PE-PPE family

M. tuberculosis has undergone reductive evolution in order to remove genes that are not essential for survival, whilst preserving and expanding genes that are essential for its niche. The pressure reflected by reductive evolution has allowed *M. tuberculosis* to refine its genome without compromising its virulence (Ahmed et al., 2008). However, the PE-PPE family has greatly expanded and is therefore a unique exception to this (Gey Van Pittius et al., 2006; Singh et al., 2014). Additionally, since the PE-PPE family accounts for ~7–10 % of the *M. tuberculosis* genome coding capacity, it suggests that these genes have an important role in the evolution of the pathogenesis of *M. tuberculosis* (Cole et al., 1998; Fishbein et al., 2015).

PE-PPE genes are exclusive to the *Mycobacteria* genus. Genes with some degree of homology to the PE-PPE family have been identified in closely related bacteria, such as *Nocardia farcinia* and *Streptomyces avermitilis;* however, this was due to the non-specific alignment of repetitive regions. The identified genes also lack the conserved N-terminal PE-PPE domains, as well as the distinguishable PE-PPE motifs (Ikeda et al., 2003; Ishikawa et al., 2004; Gey Van Pittius et al., 2006). Whilst PE-PPE genes have been found in both pathogenic and saprophytic *Mycobacteria*, comparative genomic analysis has shown that PE-PPE genes are mainly present within the pathogenic and slow-growing species of *Mycobacteria*. The highest number of PE-PPE genes are found within the *M. tuberculosis* complex, as well as other *Mycobacteria* such as *M. leprae*, *M. marinum*, and *M. avium* (Gey Van Pittius et al., 2006; McGuire et al., 2012). This indicates a favourable evolutionary selection for these genes specifically within the pathogenic species of *Mycobacteria*. This also indicates that the expansion of the PE-PPE family within the pathogenic species of *Mycobacteria* is involved in the adaptation toward a pathogenic and intracellular lifestyle (Cole et al., 1998; Fishbein et al., 2015). However, the presence of PE-PPE genes within both pathogenic and saprophytic *Mycobacteria* suggests that the PE-PPE family are also important for general survival and growth (Gey Van Pittius et al., 2006; Fishbein et al., 2015). Interestingly, the pathogenic and obligate species *M. leprae,* which has undergone massive reductive evolution, was found to contain a very limited number of PE-PPE genes. The PE-PPE genes belonging to both *M. tuberculosis* and *M. leprae* are PE15-PPE20 (ML_0538-9), PE13-PPE18 (ML_1053 and ML_1182), PE5 (ML_2534c), PPE1 (ML_1991), PPE2 (ML_1828c) and PPE68 (ML_0051c). The PE-PPE genes that endured the extensive reductive evolution of *M. leprae* may be a set of essential genes required for survival within humans (Cole et al., 2001; Gey Van Pittius et al., 2006; McGuire et al., 2012).

The PE-PPE gene families have been closely associated with the ESAT-6 cluster regions (ESX), which encode the type VII secretion systems (Gey Van Pittius et al., 2001). The ESX regions are considered to be derived from a precursor plasmid found in fast-growing *Mycobacteria*, which contained virulence factors capable of interacting with the host macrophage (Dumas et al., 2016; Newton Foot et al., 2016). Phylogenetic evidence suggests that the PE-PPE gene families expanded by multiple gene duplication events of the closely associated ESX regions. As the ESX regions were duplicated, the closely linked PE-PPE genes were also co-duplicated (Fig. 3). As PE-PPE genes have co-evolved with the ESX regions, it is also likely that these genes are functionally intertwined. There are 5 ESX regions in *M. tuberculosis;* however, there are variable numbers of ESX regions in other *Mycobacteria* such as *M. smegmatis*, which only has 3 ESX regions (ESX-4, 1 and 3) (Gey Van Pittius et al., 2006; Karboul et al., 2006). The ancestral ESX cluster region is ESX-4 (Rv3444c-Rv3450c), which was duplicated to create ESX-1 (Rv3866-Rv3883c), ESX-3 (Rv0282-Rv0292), ESX-2 (Rv3884c-Rv3895c), and lastly ESX-5 (Rv1782-Rv1798). Interestingly, the ancestral ESX cluster region ESX-4 does not contain any PE-PPE genes, and therefore, these genes must have first integrated into ESX-1 and then expanded during the subsequent gene duplication events (Gey Van Pittius et al., 2001; Gey Van Pittius et al., 2006). The PE-PPE genes belonging to ESX-1, PE35 (Rv3872) and PPE68 (Rv3873), are therefore

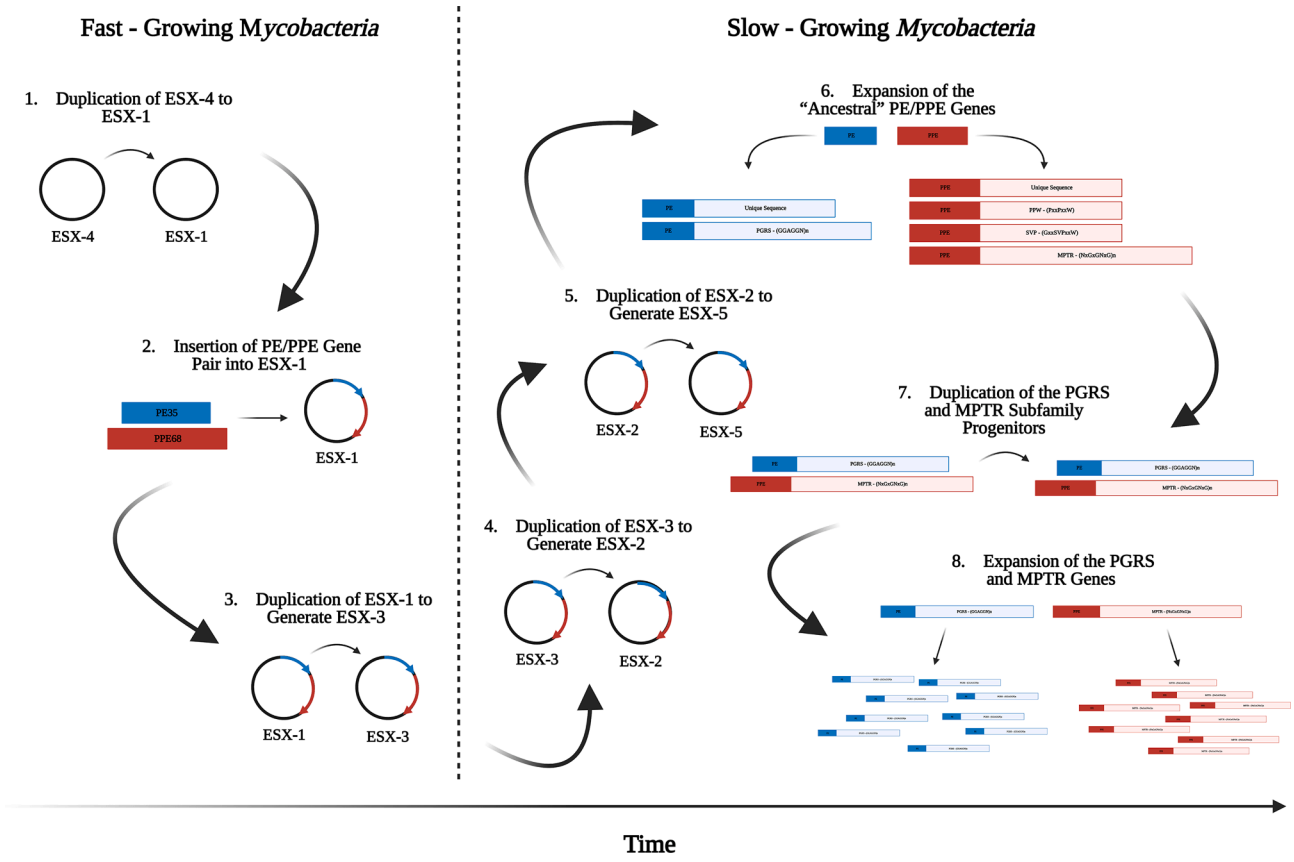

Fig. 3. Evolutionary History of the PE-PPE Family. A diagrammatic representation of the evolutionary history of the proline-glutamic acid (PE)-proline-proline-glutamic acid (PPE) family. The ancestral ESAT-6 cluster region (ESX)-4 (Rv3444c-Rv3450c) was first duplicated to create ESX-1 (Rv3866-Rv3883c). As the ancestral ESX cluster region ESX-4 does not contain any PE-PPE genes, they must have first integrated into ESX-1. The PE-PPE genes belonging to ESX-1 (PE35 [Rv3872] and PPE68 [Rv3873]) are therefore the progenitors of all PE-PPE genes. The PE-PPE genes were then further expanded during the subsequent ESX cluster region duplication events to create ESX-3 (Rv0282-Rv0292), ESX-2 (Rv3884c-Rv3895c), and lastly ESX-5 (Rv1782-Rv1798). After the primary duplication events involving the entire ESX cluster regions, secondary duplication events then took place involving PE-PPE genes from the newly duplicated ESX cluster region, in order to further expand the family (Gey Van Pittius et al., 2001; Gey Van Pittius et al., 2006). Abbreviations: polymorphic GC-rich (PE-PGRS); major polymorphic tandem repeat (MPTR).

the progenitors of all PE-PPE genes (Gey Van Pittius et al., 2001). Whilst these progenitor PE-PPE proteins do contain the conserved N-terminal PE and PPE domains, they do not contain any long, unique, or repetitive C-terminal domains, as found in PE-PPE proteins belonging to the more recent ESX cluster regions, such as ESX-5 (Gey Van Pittius et al., 2006). After the primary duplication events involving the entire ESX cluster region, secondary duplication events seem to have taken place involving single or paired PE-PPE genes from the newly duplicated ESX cluster region, in order to further expand the family. This occurrence has been largely observed in ESX-5, which gave rise to the PPE-SVP sub-family of genes that were independently duplicated from the rest of the genome (Gey Van Pittius et al., 2006). Additional events such as single nucleotide polymorphisms (SNPs), insertions, deletions, homologous recombination, and transposable insertion elements also likely contributed to the expansion and evolution of the PE-PPE genes, as well as generating variation and the emergence of unique PE-PPE sub-families (Karboul et al., 2008; McEvoy et al., 2009; Pérez-Lago et al., 2011; McEvoy et al., 2012; Phelan et al., 2016). It should be noted that whilst PE-PPE genes appear to be scattered throughout the genome, their arrangement is not completely random. Most of the PE-PPE genes are organized into operons, consisting of a PE gene followed by a PPE gene. This is believed to allow for strategic and orchestrated regulation (Tundup et al., 2006).

5. Structure of the PE-PPE family proteins

The first structural insight into PE-PPE proteins was provided by

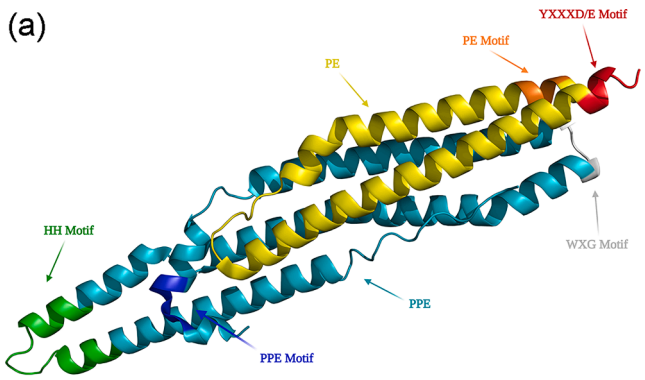

Fig. 4a. Structure of the PE25-PPE41 Heterodimer. The two-helix bundle of the proline-glutamic acid (PE) protein (yellow) interacts with alpha-helices 2 and 3 of the proline-proline-glutamic acid (PPE) protein (cyan) to form a four-helix bundle and establish a stable heterodimer conformation. This brings the conserved WxG (light grey) and YxxxD/E (red) motifs in close proximity, thus forming the potential composite recognition structure for ESAT-6 cluster region (ESX) type VII secretion (PDB = 4KXR; Daleke et al., 2012a; Poulsen et al., 2014; Korotkova et al., 2014). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Strong et al. (2006) who solved the crystal structure of the PE-PPE protein pair, PE25-PPE41 (Fig. 4a). This revealed the individual structures of a PE and PPE protein, which interact via their conserved N-terminal domains in order to form a PE-PPE heterodimer. The PE motif of the 99-residue long PE25 is located at residues 8–9; residues 8–37 and 45–84 each form an α-helix, which are antiparallel and connected via a loop formed by residues 38–44. PPE41 is 194 amino acids in length. The PPE motif is located at residues 7–9. PPE41 consists of 5 α-helices. α-helix 2 (residues 21–53) and 3 (residues 58–103) are antiparallel and responsible for interacting and stabilising the two-helix bundle belonging to its cognate PE protein PE25, via several hydrophobic and steric interactions. PE-PPE proteins appear to form a heterodimer consisting of one PE protein and one PPE protein. The PE protein consists of a two-helix bundle, and the PPE protein consists of 5 α-helices. The two helix-bundle of the PE protein interacts with α-helices 2 and 3 of the PPE protein to form a four-helix bundle and establish a stable heterodimeric conformation, which also brings together their WxG and YxxxD/E motifs, thus likely forming a composite recognition structure to allow for ESX type VII secretion (Daleke et al., 2012a; Poulsen et al., 2014). Due to the homology, sequence similarities and conservation of the N-terminal domains of PE-PPE proteins, it is expected that most PE-PPE pairs will adopt this structural arrangement (Strong et al., 2006; Ekiert and Cox, 2014; Korotkova et al., 2014; Chen et al., 2017). It should be noted that the PE-PPE heterodimer structure is highly reminiscent of the heterodimers formed by Esx proteins, such as EsxA-EsxB, otherwise known as ESAT-6-CPF-10 (Renshaw et al., 2005). ESX-1 secretion-associated proteins (Esp) such as EspB, are also similar to the PE-PPE heterodimer (Solomonson et al., 2015). Thus, the helix-bundle and composite recognition structure are shared features amongst ESX type VII secretion substrates (Ates et al., 2015a).

Interestingly, the PE25-PPE41 heterodimer was later shown to interact with the cytosolic chaperone, ESX secretion-associated protein G (EspG)-5 (Fig. 4b) (Ekiert and Cox, 2014; Korotkova et al., 2014). EspG5 was shown to interact exclusively with PPE41, specifically with its hydrophobic tip located between α-helices 4 and 5. This hydrophobic tip, otherwise known as the hh motif, is conserved amongst all PPE proteins. EspG prevents PE-PPE heterodimer aggregation by binding and shielding the PPE hydrophobic tip. EspG is also required for the correct folding of PE-PPE heterodimers, as well as for stabilising and keeping the complex in a secretion-competent state. Given the conservation of the hydrophobic tip in PPE proteins, the PE-PPE-EspG heterotrimer is currently considered the most physiologically relevant structure, as supported by the recently solved crystal structure of the PE8-PPE15-EspG5 heterotrimer (Chen et al., 2017). As a follow-up of the PE25-PPE41-EspG5 and PE8-PPE15-EspG5 crystal structures belonging to ESX-5, Williamson and colleagues recently solved the crystal structure of the PE5-PPE4-EspG3 heterotrimer, which belongs to ESX-3 (Williamson et al., 2020). In the three-dimensional structure, EspG3 seemed to bind exclusively to the hydrophobic tip of PPE4, similarly to that of the PE-PPE-EspG5 structures. However, EspG3 interacted with PPE4 at a significantly different angle compared to the EspG5 interactions observed with PPE15 or PPE41. Therefore, every ESX secretion system likely has a unique shape complementarity between its EspG chaperone and cognate PE-PPE heterodimer. The crystal structure also revealed that the hydrophobic tip of PPE4 was more extended, which may be an exclusive feature of ESX-3-associated PPE proteins, thereby aiding in distinguishing between cognate and non-cognate PE-PPE heterodimers. A comparison between the EspG crystal structures has revealed that the C-terminal helical domain of EspG3 is dynamic and capable of opening and closing like a hinge, which accommodates slight variations in the PPE hydrophobic tip and/or releases the PE-PPE heterodimer at the ESX type VII secretion system, with energy input potentially from the ESX-component and ATPase, EccA (Ekiert and Cox, 2014).

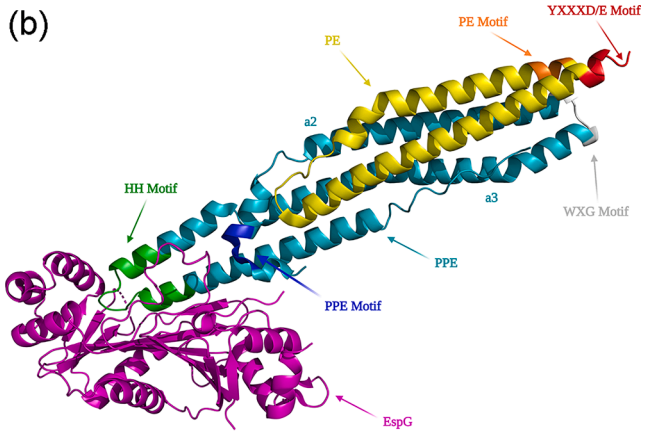

Fig. 4b. Structure of the PE25-PPE41-EspG Heterotrimer. The two-helix bundle of the proline-glutamic acid (PE) protein (yellow) interacts with alpha-helices 2 and 3 of the proline-proline-glutamic acid (PPE) protein (cyan) to form a four-helix bundle and establish a stable heterodimer conformation. This brings the conserved WxG (light grey) and YxxxD/E (red) motifs in close proximity, thus forming the potential composite recognition structure for ESAT-6 cluster region (ESX) type VII secretion. The PPE protein region shown in green is the hydrophobic tip or hh motif, which interacts with the cytosolic chaperone, ESX secretion-associated protein G (EspG) shown in magenta, in order to stabilise the complex in a secretion-competent state (PDB = 4KXR; Daleke et al., 2012a; Poulsen et al., 2014; Korotkova et al., 2014). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Interestingly, there are examples of other PE-PPE pairings that do not conform to the conclusions drawn from the crystal structures. For example, the co-operonic pair PE9 and PE10 have been shown to interact with one another and localise to the cell surface together (Tiwari et al., 2015). This example raises many questions: i) why are these two genes arranged as a PE-PE operon instead of a PE-PPE operon? ii) how do the two interact, given that they are both PE proteins? iii) how are they secreted given they are not able to form the composite recognition structure? and iv) how are they stable given that they are not able to bind to the cytosolic chaperone EspG? It is therefore likely that there are other PE-PPE protein arrangements other than the PE-PPE-EspG heterotrimer, that some PE-PPE proteins are stable without a partner and/or the cytosolic chaperone EspG, and that there are other possible routes of secretion for PE-PPE proteins. The heterotrimeric arrangement may also be imperfect given the large discrepancy in the number of PE (99) and PPE (69) proteins, as it would be impossible for every PE protein to have an exclusive cognate PPE protein (Cole et al., 1998). Therefore, there are likely to be numerous possible combinations of PE-PPE proteins, perhaps depending on the stage of the infection and/or the immune response mounted by the host (De Martino et al., 2019). This notion seems compatible with the reductive evolution of *M. tuberculosis*, as this strategy would allow *M. tuberculosis* to efficiently utilize novel PE-PPE pairs to increase its functional repertoire from a reduced selection of proteins (Veyrier et al., 2011). The disordered regions observed within members of the PE-PPE family may also increase its functional repertoire, thus further offsetting the potentially negative effects of reductive evolution. Disordered regions may achieve this by accommodating various partners with slight structural differences, or by giving multiple functions to a single protein, otherwise known as moonlighting (Ahmad et al., 2019). Whilst it is believed that sub-lineage IV PPE-SVP proteins are co-transcribed along with a sub-lineage IV PE protein, it is possible that there are transient PE partners instead, which are able to bind to a specific collection of PPE proteins (Gey Van Pittius et al., 2006). For example, PPE51 has been recently shown to interact with PE19; however, their genomic locations would not allow for the pair to be co-transcribed (Gey Van Pittius et al., 2006; Wang et al., 2020). Given that the PE domain is often responsible for cell surface localisation, it is

likely that the PE protein is able to deliver PPE proteins, or even other PE proteins, to the cell surface and then dissociate, allowing them to perform independent functions (Cascioferro et al., 2007; Ramakrishnan et al., 2015), as reported in the case of heterodimer ESAT-6-CPF-10 at low pH (de Jonge et al., 2007; Ma et al., 2015).

It will be interesting to assess the structures of PE-PPE proteins with longer and unique C-terminal domains. For example, the crystal structure of PE25-PPE41 includes a 99 amino acid PE protein and a 194 amino acid PPE protein, which are not longer than the highly conserved N-terminal PE-PPE domains (Strong et al., 2006; Ekiert and Cox, 2014; Korotkova et al., 2014). Solving the structures of PE-PPE proteins belonging to the most recent sub-lineage V (PE-PGRS and PPE-MPTR) will also be interesting, as they may offer further insights into the evolution of the PE-PPE family (Gey Van Pittius et al., 2006). A couple of questions also need to be answered: Are these proteins or protein complexes able to form higher-order structures? Does the variability within PE and PPE proteins deliberately impact their structure–function relationships? (Phelan et al., 2016; Hakim and Yang, 2021).

6. Subcellular localisations and secretion of the PE-PPE family proteins

PE-PPE proteins have been assigned a diverse range of functions and whilst some of their functions require them to be surface-bound, others require them to be secreted (Sampson, 2011). Therefore, understanding the subcellular localisation of a given PE-PPE protein may provide insights into its function (Fig. 5). Regardless of their final subcellular localisation, PE-PPE proteins first need to transit through the inner mycobacterial membrane via their cognate ESX type VII secretion system (Abdallah et al., 2006; Abdallah et al., 2007). ESX-1 has been shown to secrete PPE68 (Sani et al., 2010). ESX-3 secretes the PPE-PPW members, PPE4 and PPE20, as well as their respective cognate partners PE5 and PE15 (Tufariello et al., 2016). Given the homologous nature of the sub-family, perhaps all PPE-PPW proteins and their cognate PE partners rely on ESX-3 secretion (Siegrist et al., 2009). ESX-5 has been shown to secrete PE18, PE19, PPE25, PPE26 and PPE27, as well as the sub-lineage IV and V PE-PPE proteins (Abdallah et al., 2009; Bottai et al., 2012; Ates et al., 2015b).

The cytosolic chaperone EspG binds to the PE-PPE heterodimer via the hydrophobic tip or hh motif of the PPE protein in order to stabilise the complex in a secretion-competent state. Since there are unique EspG homologs for each ESX secretion system, the ESX-specific EspG chaperone proteins are also responsible for delivering the correct PE-PPE heterodimer to its cognate ESX secretion system (Daleke et al., 2012b). Interestingly, it is the hydrophobic tip or hh motif of the PPE protein that confers secretion system specificity. Replacing the hydrophobic tip or hh motif of PPE68 with that of PPE18 was sufficient to reroute the protein from ESX-1 to ESX-5 (Phan et al., 2017). ESX secretion system specificity is also determined by unique shape complementarity between the EspG chaperone protein and its cognate hydrophobic tip or hh motif, which may also be structurally different in PPE proteins belonging to different ESX systems (Williamson et al., 2020). Therefore, whilst ESX-specific EspG chaperone proteins are responsible for delivering the PE-PPE heterodimer to their cognate ESX secretion system, it is the PPE protein, specifically the hydrophobic tip or hh motif, which ultimately confers secretion system specificity. Thus, the sequence of the hydrophobic tip or hh motif of a PPE protein could be used to predict its cognate ESX secretion system. A recent study has suggested that the

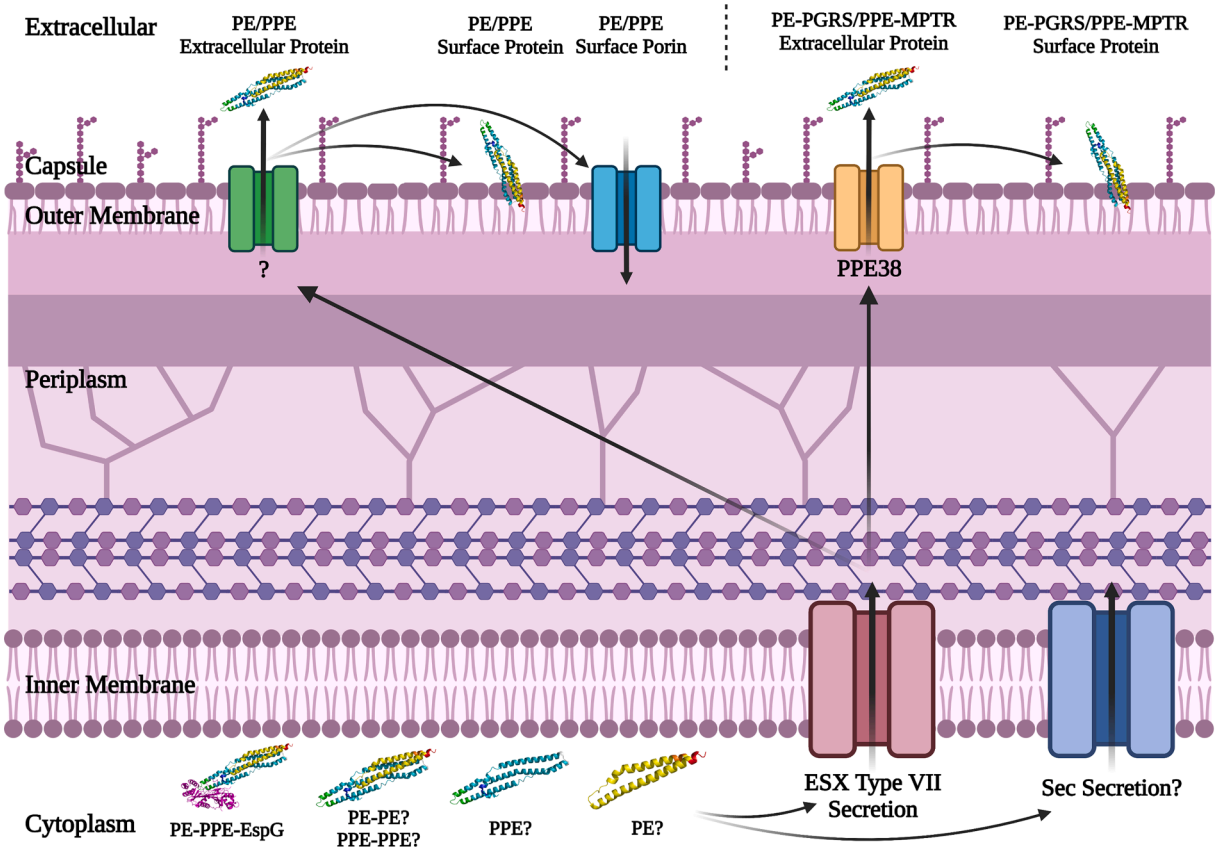

Fig. 5. Secretion and Subcellular Localisations of the PE-PPE Family. A diagrammatic representation of a cross-section of *M. tuberculosis* cell wall, depicting the secretion and subcellular localisation of the proline-glutamic acid (PE)-proline-proline-glutamic acid (PPE) family proteins. PE-PPE proteins first need to transit through the inner mycobacterial membrane via their cognate ESAT-6 cluster region (ESX) type VII secretion system. The classical *Sec* pathway may also be involved in the secretion of PE-PPE proteins. The arrows indicate the movement of PE-PPE family proteins through the mycobacterial strata, towards their final positions. Abbreviations: polymorphic GC-rich (PE-PGRS); major polymorphic tandem repeat (MPTR); ESX secretion-associated protein G (EspG).

linker 2 domain of the secretory system component EccC5, also mediates the secretion of specific PE and PPE proteins via ESX-5 (Bunduc et al., 2020).

Interestingly, some ESX-5-dependent PE-PPE proteins also depend on other PE-PPE proteins for their secretion. The PPE-SVP protein, PPE38, is essential for the secretion of all PPE-MPTR and PE-PGRS proteins (Ates et al., 2018a; Ates et al., 2018b). Since PPE-MPTR and PE-PGRS proteins are part of sub-lineage V, it is worthwhile to investigate whether PPE38 is also essential for other ESX-5 dependent PE-PPE proteins, such as the sub-lineage IV PE-PPE proteins (Abdallah et al., 2009; Ates et al., 2015b), and whether other PE-PPE proteins play similar roles to PPE38 for the other ESX secretion systems. The loss of PPE-MPTR and PE-PGRS secretion due to a defective PPE38 has been associated with a hypervirulent phenotype of *M. tuberculosis* (Ates et al., 2018a; Ates et al., 2018b). A disrupted EspG5 in *M. marinum* also resulted in a hypervirulent phenotype in an adult zebrafish infection model (Weerdenburg et al., 2012). Since loss of ESX-5 secretion is fatal for pathogenic strains (Di Luca et al., 2012; Ates et al., 2015b), the increased virulence observed suggests that PPE-MTPR and PE-PGRS proteins are virulence attenuating and not essential for *M. tuberculosis*, whilst the remaining ESX-5 PE-PPE proteins are essential for *M. tuberculosis*. In addition, the EspG5 disruption data suggests that the essential ESX-5 PE-PPE proteins may not require the cytosolic chaperone protein EspG and that EspG is not essential for all PE-PPE proteins (Daleke et al., 2012b).

In addition to the ESX secretion systems, the classical *Sec* pathway may also be involved in the secretion of PE-PPE proteins. *Sec* pathway signal sequences have been found to be cleaved in several PE-PPE proteins, indicating that they may be secreted via the classical Sec pathway as an alternative secretion pathway (De Souza et al., 2011).

7. Functions of the PE-PPE family proteins

Initially, the PE-PPE family due to its polymorphic and repetitive nature was considered a source of genetic and antigenic variation in *M. tuberculosis* (Cole et al., 1998). However, PE-PPE proteins have since been assigned a wide range of diverse roles, which are discussed below.

7.1. Interaction with and modulation of immune cells

Toll-like receptors (TLRs) are pattern recognition receptors (PRRs) of the innate immune system and recognise pathogen-associated molecular patterns (PAMPs) of various pathogens, including *M. tuberculosis.* TLR recognition of PAMPs results in the activation of pro-inflammatory signalling pathways (Vijay, 2018). PE-PPE family members have been observed to modulate the host immune response by interacting with TLRs.

PE-PGRS11 and PE-PGRS17 bind to TLR2, resulting in the maturation of dendritic cells (DCs), as well as an enhancement in their ability to stimulate the proliferation and secretion of cytokines from cluster of differentiation (CD)4^+ T-cells (Bansal et al., 2010a). PPE34 also stimulates the maturation of DCs in a TLR2-dependent manner. PPE34-matured DCs secrete interleukin (IL)-10 in addition to promoting the secretion of IL-4, IL-5, and IL-10 from $CD4^+$ T-cells, thereby skewing the immune response in favour of T_h2 cellular immunity. PPE34-induced T_h2 cellular immunity is associated with increased expression of cyclooxygenase-2 (COX-2) in DCs (Bansal et al., 2010b). PPE26 binds to TLR2, resulting in macrophage activation, as well as enhanced expression of co-stimulatory molecules (CD80 and CD86) and major histocompatibility complex (MHC) molecules (class I and II). Additionally, macrophages activated by PPE26 showed upregulated expression of C-X-C chemokine receptor 3 (CXCR3), as well as the secretion of interferon-gamma (IFN-ɣ) and IL-2 in naïve $CD4^+$ T-cells, indicating PPE26-induced T_h1 polarisation (Su et al., 2015). PPE57 binds to TLR2 resulting in macrophage activation and enhanced expression of the cell surface markers CD40, CD80, CD86 and MHC class II molecules. Furthermore, PPE57-stimulated macrophages also polarise naïve $CD4^+$ T-cells to upregulate CXCR3 expression, as well as increase IFN-ɣ and IL-2 secretion, indicating a T_h1 polarisation (Xu et al., 2015). PPE60 binds to TLR2 resulting in DC maturation and enhanced expression of CD80, CD86 and MHC I and II. PPE60-stimulated DCs polarise naïve $CD4^+$ T-cells to increase the secretion of IFN-ɣ, IL-2 and IL-17A, thus potentiating the T_h1/ T_h17 immune responses (Su et al., 2018). Similarly, PE27 has been shown to induce DC maturation via mitogen-activated protein kinase (MAPK) and nuclear factor kappa B (NF-κB) signalling pathways, eventually leading to T_h1 polarisation. PE27 also induced memory T-cell responses that produced IFN-ɣ in *M. tuberculosis*-infected mice (Kim et al., 2016). PPE39 appears to enhance DC maturation via TLR4-mediated MAPK and NF-κB signalling, in addition to promoting T_h1 polarisation that coincided with increased T-box protein expression in T-cells (T-BET) (Choi et al., 2019).

M. tuberculosis is able to utilise the immunogenic PE-PPE family proteins to modulate both the innate and adaptive immune responses. Interestingly, the PE-PPE family proteins are able to elicit both T_h1 and T_h2 responses through various interactions and pathways. T_h1 responses are typically essential for intracellular pathogens whereas T_h2 responses are beneficial against extracellular pathogens. However, a balance between the two is essential for the successful clearance of a pathogen and to restrict excessive T_h1-induced tissue damage. Given that *M. tuberculosis* is an intracellular pathogen, T_h1 responses are critical for its clearance, and therefore, it would seem counterintuitive for *M. tuberculosis* to use PE-PPE family proteins to elicit the T_h1 response. Therefore, it could be hypothesised that the PE-PPE family proteins that induce a T_h1 response function as an immunological "smoke screen". During the initial stage of infection, *M. tuberculosis* may express these immunogenic proteins in order to elicit a T_h1 immune response against these "smoke screen" proteins. During the subsequent stage of infection, *M. tuberculosis* may then cease the expression of these immunogenic proteins and express other important functional proteins. As the host immune response will be primed against the "smoke screen" proteins, which are no longer expressed, it allows *M. tuberculosis* and its important functional proteins to establish infection within the host safely. The "smoke screen" proteins, therefore, hide *M. tuberculosis* in plain sight from the host immune response, thereby allowing for the successful establishment of an infection within the host. Once this stage has passed, PE-PPE proteins that induce a T_h2 response or dampen the T_h1 response may then play a role in maintaining immune evasion.

7.2. Modulation of inflammatory responses

The inflammatory response is an initial immune response against invading pathogens such as *M. tuberculosis* and is also involved in the subsequent activation and shaping of the adaptive immune response. Cytokines such as IL-1, IL-6, and tumour necrosis factor-alpha (TNF-α) are pro-inflammatory, whereas anti-inflammatory cytokines such as IL-10 and transforming growth factor-beta (TGF-β) regulate or dampen the inflammatory response (Mogensen, 2009). A consequence of PE-PPE-induced TLR recognition is modulation of the inflammatory response, as outlined below.

7.2.1. Pro-inflammatory response triggered by PE-PPE family members

M. smegmatis heterologously expressing PPE44 has been shown to increase the expression and secretion of pro-inflammatory cytokines IL-6 and IL-12p40 in THP-1 macrophages, via the p38, extracellular signal-regulated kinase 1/2 (ERK) and NF-κB signalling axis (Yu et al., 2017). Both purified PE-PGRS33 and *M. smegmatis* heterologously expressing PE-PGRS33 increase TNF-α production in RAW 264.7 macrophages, via the apoptosis signal-regulating kinase 1 (ASK1), p38 and c-Jun N-terminal kinase (JNK) signalling pathways (Basu et al., 2007). Purified PPE26 and PPE57 increase IL-6, IL-12p40 and TNF-α release in RAW264.7 macrophages, via the MAPK and NF-κB signalling pathways (Su et al., 2015; Xu et al., 2015). Purified PPE60 increases IL-1β, IL-6, IL-

12p70, IL-23p19 and TNF-α production in DCs, via activation of p38, JNK and NF-κB signalling pathways (Su et al., 2018). *M. smegmatis* heterologously expressing PE13 increases IL-6 and IL-1β via the p38, ERK and NF-κB signalling axis in THP-1 macrophages. In addition, PE13 was also found to decrease the secretion of the suppressor of cytokine signalling 3 (SOCS3), which has an important role in the regulation of pro-inflammatory response (Li et al., 2016). An *M. marinum* PPE38 mutant showed significantly lower levels of IL-6 and TNF-α in THP-1 macrophages compared to the wild-type. *M. smegmatis* heterologously expressing PPE38 induced higher levels of IL-6 and TNF-α, thereby validating the pro-inflammatory effects of PPE38 (Dong et al., 2011). The PGRS domain of PE-PGRS5 induces higher levels of IL-12 and TNF-α in THP-1 macrophages treated with purified protein or recombinant *M. smegmatis* (Sharma et al., 2020; Sharma et al., 2021b). Purified PE27 promotes the secretion of IL-1β, IL-6, IL-12p70 and TNF-α in DCs via MAPK and NF-κB signalling (Kim et al., 2016). PE4 increased levels of IL-2, IL-6, and TNF-α in splenocytes of mice immunised with purified PE4 compared to PBS-injected mice. (Singh et al., 2012). Purified PE6 seems to increase the secretion of IL-6, IL-12, and TNF-α in RAW264.7 macrophages via TLR4 interaction and NF-κB signalling (Sharma et al., 2021a).

7.2.2. Anti-inflammatory responses triggered by PE-PPE family members

M. smegmatis heterologously expressing PE5 and PE15 appears to enhance the expression and secretion of anti-inflammatory cytokine IL-10 whilst also downregulating the secretion of pro-inflammatory cytokine IL-12 in THP-1 macrophages (Tiwari et al., 2012). PE31 heterologously expressed in *M. smegmatis* increases anti-inflammatory cytokine IL-10, whilst also decreasing pro-inflammatory cytokines IL-6 and IL-12p40 in THP-1 macrophages via the NF-κB signalling pathway (Ali et al., 2020). THP-1 macrophages stimulated with purified PE35-PPE68 showed increased IL-10 secretion as a result of TLR2 interaction, as well as p38 and EKR1/2 MAPK activation (Tiwari et al., 2014). Interestingly, the pair were also shown to increase monocyte chemo-attractant protein 1 (MCP-1), which is a chemo-attractant for monocytes/lymphocytes and therefore, essential for the granulomatous response (Lu et al., 1998). *M. smegmatis* heterologously expressing PPE37 decreases IL-1β, IL-6, IL-12p70, and TNF-α production in peritoneal exudate macrophages, via lower activation of NF-κB, MAPK, ERK and p38 signalling pathways (Daim et al., 2011). PE-PGRS41 increased anti-inflammatory cytokine IL-10 and decreased pro-inflammatory cytokines IL-1β, IL-6, and TNF-α in recombinant *M. smegmatis* infected THP-1 macrophages (Deng et al., 2017). RAW264.7 macrophages treated with purified PE32-PPE65 increased IL-10 and decreased IL-6 and TNF-α levels, whilst also reducing the number of IFN-ɣ and IL-2 secreting $CD4^+$ and $CD8^+$ T-cells, thereby inhibiting the T_h1 response (Khubaib et al., 2016). *M. smegmatis* heterologously expressing PE-PGRS30 reduces the production of IL-6, IL-12 and TNF-α in THP-1 macrophages however, the mechanisms and signalling pathways remain to be identified (Chatrath et al., 2016). THP-1 macrophages incubated with purified PPE18 promote the release of IL-10 via TLR2-mediated activation of p38 MAPK (Nair et al., 2009). Interestingly, PPE18 was later shown to inhibit the production of IL-12p40 and TNF-α by blocking the nuclear translocation of transcription factors p50, p60, c-rel, and NF-κB (Nair et al., 2011). THP-1 macrophages infected by *M. smegmatis* heterologously expressing PE17 showed increased transcriptional levels of IL-10 and decreased transcriptional levels of IL-1β, IL-6 and TNF-α via JNK signalling (Abo-Kadoum et al., 2021).

Interestingly, purified PPE32 enhances both pro-inflammatory (IL-6 and TNF-α) and anti-inflammatory (IL-10) cytokine production in U937 macrophages, via the NF-κB and MAPK signalling pathways (Deng et al., 2014; Deng et al., 2016), suggesting that the PE-PPE family proteins have contrasting roles during host interaction. *M. tuberculosis* may use these proteins as a molecular switch by expressing a specific combination in order to induce or fine-tune a specific immune response, perhaps depending on the stage of infection. It is therefore essential to elucidate the stage of infection at which these proteins are expressed and utilised by *M. tuberculosis*. Therefore, a comprehensive timeline of PE-PPE family protein expression and function during infection will be of great value. Overall, modulation of the host's inflammatory response by PE-PPE family proteins ultimately favours *M. tuberculosis* survival within the host.

7.3. Ca^{2+} binding

Ca^{2+} is an essential signalling molecule involved in controlling cellular activities and maintaining cellular homeostasis. In addition, Ca^{2+} is also known to be involved in regulating immune cells and processes during infections with pathogens such as *M. tuberculosis* (Sharma and Meena, 2016). The PE-PPE family proteins have recently been implicated in altering Ca^{2+} signalling.

Fifty-six out of 61 PE-PGRS proteins contain repeats of the glycine-rich motif GGXGXD/NXUX; a nonapeptide sequence predicted to bind Ca^{2+}. PE-PGRS33 and PE-PGRS61 bind Ca^{2+} using the GGXGXD/NXUX motif in order to elicit their immunomodulatory effects (Yeruva et al., 2016). Similarly, the functions of PE-PGRS5, which are TLR4-dependent, also seem to be Ca^{2+}-dependent, as Ca^{2+} stabilises the binding between PE-PGRS5 and TLR4 (Sharma et al., 2021b). In addition, Ca^{2+}-dependent PE-PPE proteins such as PE-PGRS5, PE-PGRS33, and PE-PGRS61 have been associated with altering the Ca^{2+} concentration within the host cell, which results in compromised Ca^{2+} signalling, leading to inhibited phagosome acidification and maturation, thus promoting *M. tuberculosis* intracellular survival (Sharma and Meena, 2016).

7.4. Inhibition of phagolysosome fusion/maturation

Upon infection with *M. tuberculosis,* the fusion of phagosomes containing *M. tuberculosis* with lysosomes is a defence strategy used by host macrophages to eradicate the intracellular pathogen. Acidification of the phagolysosome not only inflicts acid stress on *M. tuberculosis,* but also increases the efficacy of hydrolases, nitric oxides, and reactive oxygen species against *M. tuberculosis* (Stallings and Glickman, 2010). However, PE-PPE family proteins subvert this process using various mechanisms.

PE-PGRS30 inhibits phagosome-lysosome fusion within macrophages, as evidenced by increased lysosomal-associated membrane protein 1 (LAMP-1) expression in THP-1 macrophages infected with PE-PGRS30-deficient *M. tuberculosis* (Iantomasi et al., 2011). PPE10 is implicated in the control of phagosome acidification; however, the underlying mechanisms remain to be elucidated (Stewart et al., 2005). *M. smegmatis* heterologously expressing PE-PGRS62 inhibits phagosome maturation in J774A macrophages by blocking the acquisition of Rab7 and LAMP-1 (Thi et al., 2012). PE-PGRS5 has recently been shown to inhibit phagosomal acidification by downregulating the expression of late endosomal markers Rab7 and cathepsin D (CTSD) in THP-1 cells infected with recombinant *M. smegmatis* (Sharma et al., 2020; Sharma et al., 2021b).

By utilising these PE-PPE proteins, *M. tuberculosis* is able to subvert this initial host macrophage defence, which would typically eradicate intracellular pathogens, in order to establish and persist within its preferred intracellular niche. In addition, the PE-PPE-induced subversion of phagolysosome fusion/maturation has further implications for *M. tuberculosis* persistence.

7.5. Inhibition of antigen presentation

Antigens from intracellular pathogens such as *M. tuberculosis* are degraded and subsequently loaded onto MHC class II molecules, which allows $CD4^+$ T-cells to recognise and generate effector and memory T-cell responses against the intracellular pathogen (Harding and Boom, 2010). However, PE-PPE proteins inhibit antigen presentation by

preventing the process required to produce antigens, as described below.

The PGRS domain of PE-PGRS17 reduces antigen presentation by protecting the protein from ubiquitin-dependent proteasome proteolysis and processing (Koh et al., 2009a). Purified PPE18 was also recently shown to inhibit antigen presentation in peritoneal macrophages by preventing peptide degradation through disrupting phagolysosome acidification (Dolasia et al., 2020). Interestingly, an *M. tuberculosis* PE-PGRS47 mutant showed enhanced MHC class II antigen presentation in infected mice, indicating that PE-PGRS47 also inhibits antigen presentation in macrophages (Saini et al., 2016). Phagolysosomes and ubiquitin-dependent proteolysis are both essential in the degradation and processing of peptides required for antigen presentation (Harding and Boom, 2010). Therefore, *M. tuberculosis* is able to use these PE-PPE family proteins in order to inhibit such processes, thereby dampening potential T-cell responses, which contribute to *M. tuberculosis* persistence.

7.6. Stress resistance

M. tuberculosis is an intracellular pathogen that must endure an onslaught of stress responses from the host in order to persist within its preferred intracellular niche. Upon infection, *M. tuberculosis* is confronted by stress such as low pH, hypoxia, hydrolases, nitric oxides, reactive oxygen species, nutrient starvation, and antimicrobial drugs (Stallings and Glickman, 2010). The PE-PPE family proteins have been reported to provide resistance against many of these stresses through various mechanisms, as discussed below.

7.6.1. Resisting host-induced stress

PE4 expression in *M. tuberculosis* is induced by both acidic and hypoxic conditions, highlighting a potential role in resisting host-induced stress (Singh et al., 2012). Heterologously expressed PE11 alters cell wall lipid content and colony morphology in *M. smegmatis*, which increases its resistance against sodium dodecyl-sulfate (SDS), lysozyme, hydrogen peroxide (H_2O_2), and low pH (Singh et al., 2016). *M. smegmatis* heterologously expressing PE5 and PE15 suppressed inducible nitric oxide synthase (iNOS) transcript levels in both J774.1 and THP-1 infected macrophages (Tiwari et al., 2012). PPE2 also inhibits nitric oxide production by suppressing iNOS transcription. Green fluorescent protein (GFP)-tagged PPE2 was found to achieve this by translocating to the nucleus of RAW264.7 macrophages using its nuclear localisation signal and interacting with the iNOS promoter via its leucine zipper DNA binding motif, evidenced by the binding of purified PPE2 with oligonucleotides representing the iNOS promoter region (Bhat et al., 2017). Surprisingly, PPE2 also inhibits reactive oxygen species production in transfected RAW264.7 macrophages via its SRC homology 3 (SH3) domain, which binds to the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex component p67phox, thereby preventing the proper assembly of the NADPH oxidase complex, and thus, inhibiting reactive oxygen species production (Srivastava et al., 2019). Interestingly, J774A macrophages infected with *M. smegmatis* heterologously expressing PE-PGRS62 showed similar levels of iNOS transcripts, yet reduced levels of iNOS protein, suggesting that PE-PGRS62 also inhibits nitric oxide production, likely at the post-transcriptional level, possibly through iNOS translational dysregulation or iNOS protein degradation (Thi et al., 2012). Transcriptional analysis showed that hypoxia-induced PE-PGRS11 protects against oxidative stress by triggering TLR2-dependent expression of both B-Cell lymphoma 2 (BCL-2) and COX-2 (Chaturvedi et al., 2010). PE13, a cell wall-associated protein that significantly alters colony size and morphology, has been shown to increase the survival of recombinant *M. smegmatis* against low pH, diamides, H_2O_2, and SDS (Li et al., 2016). Similarly, cell wall-associated protein, PPE60, which also has roles in cell wall integrity and permeability, increases the survival of recombinant *M. smegmatis* against low pH, H_2O_2, SDS, and diamides (Gong et al., 2019). Surprisingly, *M. smegmatis* heterologously expressing PE-PGRS5 was recently shown to promote stress by inducing nitric oxide production in RAW264.7 macrophages, in a Ca^{2+}- and TLR4-dependent manner (Grover et al., 2018; Sharma et al., 2020; Sharma et al., 2021b).

Members of the PE-PPE family are evidently upregulated and utilised by *M. tuberculosis* in order to acquire resistance against a variety of stresses encountered within the intracellular milieu of the host cell. It is also interesting to note that a single member of the PE-PPE family is able to provide resistance against two different types of stress via distinct mechanisms, such as PPE2 against nitric oxide and reactive oxygen species (Srivastava et al., 2019). This observation aligns with the reductive evolution of *M. tuberculosis*, as giving multiple functions to a single protein, otherwise known as "moonlighting", would allow *M. tuberculosis* to retain its functional repertoire from a reduced number of proteins (Veyrier et al., 2011; Ahmad et al., 2019). It is also interesting to observe that different members of the PE-PPE family are able to confer resistance against the same stress by using different mechanisms, such as PPE2 and PE-PGRS62 against nitric oxide (Thi et al., 2012; Bhat et al., 2017). Whilst this observation goes against the reductive evolution of *M. tuberculosis,* this is perhaps to ensure that *M. tuberculosis* is able to confer resistance and ensure its survival irrespective of deleterious mutations or blocked pathways.

7.6.2. Resisting nutrient starvation stress

PPE36 and PPE62 are considered essential for haem-iron acquisition. In addition, these cell surface proteins have also been shown to bind, uptake and utilise haemoglobin, evidenced by the inability of *M. tuberculosis PPE36* and *PPE62* deletion mutants to grow in Hartmans de Bond (HdB) minimal media with human haemoglobin or heme as the sole iron source (Mitra et al., 2017; Mitra et al., 2019). *M. tuberculosis PPE37* deletion mutants highlighted PPE37 as an essential protein for haem-iron acquisition; however, its exact role is yet to be elucidated (Tullius et al., 2019). *M. tuberculosis PPE51* mutants showed PPE51 to be essential for the acquisition of sugars, including glucose, glycerol, and thio-glycoside, as well as other small molecules such as propionamide (Korycka-Machała et al., 2020; Wang et al., 2020). Wang and colleagues have shown that PE20 and PPE31 are up-regulated and required for growth under magnesium-limited conditions, and therefore, essential for magnesium acquisition (Wang et al., 2020).

Nutrient acquisition PE-PPE proteins, such as those previously mentioned, are secreted and localised to the cell surface, indicating that *M. tuberculosis* may utilise these proteins to form pore-like nutrient-selective channels in order to transport nutrients over the resilient cell wall of *M. tuberculosis,* which is concordant with the earlier findings of Ates and colleagues (Ates et al., 2015b). Since the cell wall of *M. tuberculosis* lacks classical porins for transport, PE-PPE family proteins appear to provide an effective mechanism to overcome such limitations and acquire the nutrients required to survive within the nutrient-limited intracellular niche. This concept could also apply to other ESX-dependent substrates with similar structures to the PE-PPE family proteins. Whilst the evidence suggests that the PE-PPE family proteins would allow for the import of nutrients, it could also be hypothesised that the pore-like channels formed by PE-PPE family proteins may also allow for the export of other molecules.

7.6.3. Resisting antibiotic and antimicrobial drug stress

PE11, a cell wall protein with esterase activity, has been shown to alter the cell wall lipid content, which increases surface hydrophobicity, thereby conferring resistance to several antibiotics and anti-microbial drugs including rifampicin, isoniazid, ethambutol, ampicillin, and vancomycin, by preventing penetration through the cell wall (Singh et al., 2016). Interestingly, gene pairs associated with drug resistance typically consist of a known drug target and a PE-PPE family gene. *katG-PPE54* and *rpob-PPE54* gene pairs/interactions were identified and associated with isoniazid and rifampicin resistance, respectively. Furthermore, *embA-PPE68* and *embB-PPE54* gene pairs/interactions were associated with ethambutol resistance (Cui et al., 2016). Single nucleotide

polymorphisms (SNPs), insertions and deletions were found to be enriched in the PE-PGRS genes belonging to 37 extensively drug-resistant (XDR) *M. tuberculosis* strains. Mutations were notably found in PE_PGRS3, PE_PGRS6, PE_PGRS9, PE_PGRS10, PE_PGRS 19, PE_PGRS 33 and PE_PGRS 49, which highlight their potential role in drug resistance (Kanji et al., 2015). Similarly, genetic variation in PPE genes PPE18, PPE19, PPE46, and PPE47, were associated with the expansion of isoniazid resistance in *M. tuberculosis* (Hang et al., 2019). Homoplastic SNPs, which are associated with conferring drug resistance, were recently characterised in 1170 *M. tuberculosis* strains; ~36% of all homoplastic SNPs were found in PE-PPE genes, further highlighting their potential role in drug resistance. A novel mutation in PE-PGRS7 was also found to confer streptomycin resistance; however, the underlying mechanism of resistance requires further investigations (Tantivitayakul et al., 2020).

It is evident that PE-PPE family members are involved in conferring resistance against various antimicrobial drugs. However, further research is required to elucidate the underlying resistance mechanisms, which may also highlight novel drug targets. By extrapolating the pore-like molecular/nutrient-selective channel functions of the PE-PPE family, it may hint toward a possible mechanism of drug resistance; it is possible that the PE-PPE family may also function as efflux proteins or, as part of an export assembly.

7.7. Modulation of host cell death pathways

Apoptosis, necrosis, autophagy, and pyroptosis are different forms of cell death, which have been observed during infection with *M. tuberculosis* (Chai et al., 2020). PE-PPE family proteins have been observed to modulate these forms of cell death by various mechanisms.

7.7.1. Apoptosis

Apoptosis, a form of programmed cell death, is an innate immune defence strategy utilised to contain and eradicate infection (Reed, 2000). Purified PPE32 induces apoptosis in THP-1 macrophages in a caspase-3- and caspase-9-dependent manner. PPE32 also promotes endoplasmic reticulum stress-related gene expression, which may be responsible for PPE32-induced apoptosis (Deng et al., 2016). The purified PGRS domain of PE-PGRS5 induces apoptosis in RAW264.7 macrophages in a TLR4- and Ca^{2+}-dependent manner, by localising to the endoplasmic reticulum and promoting endoplasmic reticulum stress, resulting in caspase-8 activation (Grover et al., 2018; Sharma et al., 2020; Sharma et al., 2021b). *M. smegmatis* heterologously expressing PE9-PE10 promotes apoptosis in THP-1 macrophages by interacting with TLR4 and inducing interferon regulatory transcription factor 3 (IRF-3) signalling, which leads to the upregulation of pro-apoptotic genes *bax, bim* and *bid,* as well as caspase-3 cleavage (Tiwari et al., 2015). Purified PE6 induces apoptosis in RAW264.7 macrophages by increasing the production of pro-apoptotic molecules Bax, Cytochrome *C*, and pcMyc. PE6 achieves this by activating both caspase-3 and caspase-9, as well as the endoplasmic reticulum-associated unfolding protein response (UPR) pathway, which resulted in increased production of activating transcription factor 6 (ATF6), C/EBP homologous protein (CHOP), binding immunoglobulin protein (BiP), eukaryotic translation initiation factor 2 alpha (eIF2α), inositol-requiring enzyme 1 alpha (IRE1α), and calnexin (CANX) (Sharma et al., 2021a). Interestingly, PE17-induced apoptosis was recently associated with regulating host cell gene expression by reducing H3K9me3 chromatin occupancy, which resulted in the activation of pro-apoptotic signalling pathways in recombinant *M. smegmatis* infected THP-1 macrophages (Abo-Kadoum et al., 2021). Conversely, PE-PPE proteins such as *M. smegmatis* heterologously expressed PE31 inhibits apoptosis in THP-1 macrophages by activating the NF-κB signalling pathway, which induces IL-10 and guanylate binding protein 1 (GBP-1) production, thereby inhibiting the activation of caspase-3 (Ali et al., 2020). PE-PGRS41 heterologously expressed in *M. smegmatis* showed decreased cleavage of caspase-3 and caspase-9 in THP-1 macrophages, thereby inhibiting *M. smegmatis* infection-induced apoptosis (Deng et al., 2017). *M. smegmatis* heterologously expressing PPE10 inhibits apoptosis in THP-1 macrophages by decreasing transcriptional levels of the pro-apoptotic gene *Bax*, as well as decreasing expression levels of caspase-3, caspase-7, and caspase-8 through the linear ubiquitin chain assembly complex (LUBAC), HOIL-1-interacting protein (HOIP), and NF-κB signalling axis (Asaad et al., 2021). PE-PGRS62 decreases endoplasmic reticulum stress, as well as caspase-3 and caspase-9 cleavage, leading to the attenuation of apoptosis in recombinant *M. smegmatis* infected THP-1 macrophages (Long et al., 2019).

7.7.2. Necrosis

Necrosis, a form of passive cell death, is typically favourable for the pathogen and aids in the dissemination of infection (Dobos et al., 2000). PE-PGRS33 induces necrosis in macrophages, as evidenced by the increased levels of lactate dehydrogenase and nucleosomes in the supernatant of murine bone marrow macrophage culture infected with recombinant *M. smegmatis* (Dheenadhayalan et al., 2006). Other PE-PPE proteins such as PE11, PPE26, PPE68, and PE25/PPE41, have also been associated with the induction of necrosis; however, the underlying mechanisms are yet to be fully elucidated (Tundup et al., 2014; Deng et al., 2015; Danelishvili et al., 2015; Mi et al., 2017).

7.7.3. Autophagy

Autophagy is an innate immune defence strategy that serves to degrade and process intracellular pathogens within autophagosomes for MHC class II presentation (Gutierrez et al., 2004). *M. smegmatis* heterologously expressing PE-PGRS41 inhibits autophagy in THP-1 macrophages by blocking autophagy-related protein 8 (ATG-8), which is responsible for controlling the expansion and size of precursor autophagosomes. PE-PGRS41 also suppresses the conversion of microtubule-associated protein light chain 3 (LC3)-I to LC3-II, which is also involved in autophagosome formation (Deng et al., 2017). Recently, RAW264.7 macrophages transfected to express PE-PGRS20 and PE-PGRS47 were shown to inhibit autophagy initiation by interacting with and activating Rab1A, which leads to mammalian target of rapamycin (mTOR) activation and autophagy inhibition (Strong et al., 2021). Purified PE6 suppresses the initiation of autophagy in RAW264.7 macrophages by promoting inhibitory phosphorylation of Unc-51 like autophagy activating kinase 1 (ULK1), which is responsible for initiating autophagophore biogenesis. PE6 achieves this by promoting activating phosphorylation of mammalian target of rapamycin complex 1 (mTORC1), which is responsible for the inhibitory phosphorylation of ULK1. PE6-inhibited autophagy initiation was also evidenced by reduced conversion of microtubule-associated protein light chain 3B (LC3B)-I to LC3B-II, and increased accumulation of SQSTM1/p62, which is a ubiquitin-binding scaffold protein degraded by autophagy (Sharma et al., 2021a). Surprisingly, a PE-PGRS29-deficient *M. tuberculosis* strain showed that PE-PGRS29, a cell surface protein containing a eukaryotic-like ubiquitin-associated domain, binds ubiquitin, thus triggering ubiquitin-mediated xenophagy and subsequent clearance of *M. tuberculosis* in murine bone marrow-derived macrophages. Whilst seeming counterintuitive, this strategy may be used by *M. tuberculosis* to optimise intracellular bacterial loads whilst also restricting host inflammatory responses (Chai et al., 2019).

7.7.4. Pyroptosis

Pyroptosis is a rapid and highly inflammatory form of cell death, which is associated with plasma-membrane rupture, as well as the release of inflammatory factors and bacilli (Bergsbaken, et al., 2009). *M. smegmatis* heterologously expressing PPE60 promotes pyroptosis in THP-1 macrophages in a caspase-, NOD- LRR- and pyrin domain-containing protein 3 (NLRP3)-, and gasdermin (GSDM)-dependent manner (Gong et al., 2019). More recently, PPE13 was shown to activate the NLRP3 inflammasome, which induces caspase-1 activation, and

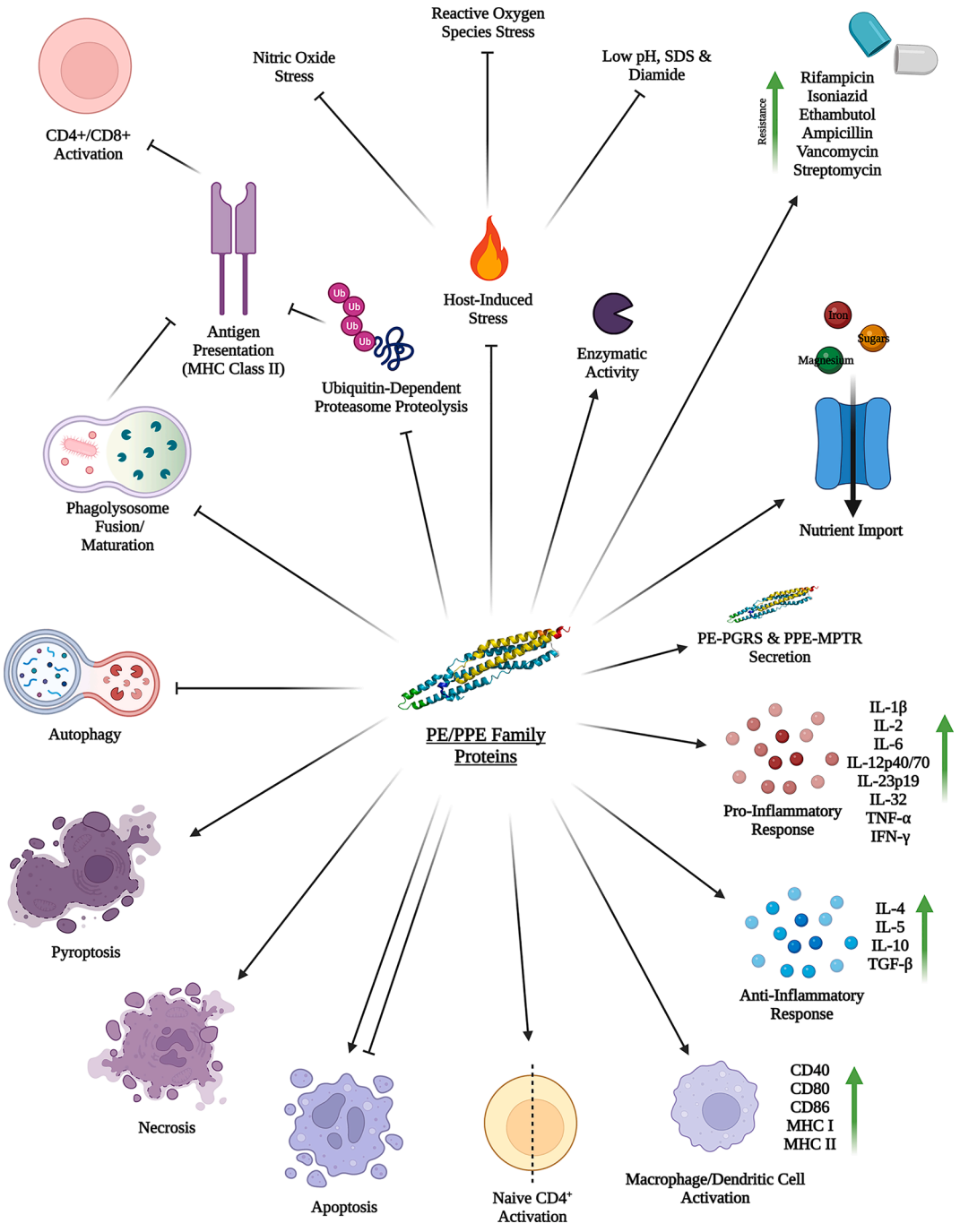

Fig. 6. Functions of the PE-PPE Family. A diagrammatic representation of the functions elicited by the proline-glutamic acid (PE)-proline-proline-glutamic acid (PPE) family proteins. Initially, the PE-PPE family due to its polymorphic and repetitive nature was considered a source of genetic and antigenic variation in *M. tuberculosis* (Cole et al., 1998). However, PE-PPE proteins have since been assigned a wide range of diverse roles, which are outlined in the figure. The functions of the PE-PPE family are extensive, including modulating host immune responses, subverting host defence strategies, resisting diverse stresses of the host, and manipulating host cell fates, ultimately to ensure the survival of *M. tuberculosis*. The arrows indicate the PE-PPE-induced effects whereas the blunt-end arrows indicate the PE-PPE-inhibited effects. Chains of arrows and blunt-end arrows also highlight the functional pathways and consequential effects of PE-PPE-elicited functions. Abbreviations: cluster of differentiation (CD); major histocompatibility complex (MHC); interleukin (IL); interferon-gamma (IFN-ɣ); transforming growth factor-beta (TGF-β); tumour necrosis factor-alpha (TNF-α); sodium dodecyl-sulfate (SDS); polymorphic GC-rich (PE-PGRS); major polymorphic tandem repeat (MPTR).

GSDM-mediated pyroptosis in THP-1, J774A.1, and murine bone marrow-derived macrophages, infected with recombinant *M. smegmatis* (Yang et al., 2020).

M. tuberculosis may use specific PE-PPE family proteins to inhibit apoptosis during the early stages of infection, in order to prevent eradication and establish infection. Conversely, *M. tuberculosis* may use other PE-PPE family proteins to promote apoptosis during the later stages of infection, in order to allow *M. tuberculosis* to effectively disseminate without eliciting a robust inflammatory response. However, this does not explain why *M. tuberculosis* would promote cell death pathways such as pyroptosis, which is associated with extensive inflammation. Perhaps, this further emphasises how crucial it is to elucidate the stage of infection at which these proteins are expressed and utilised by *M. tuberculosis*.

7.8. Additional functions

PE-PPE family proteins have been recently shown to have contrasting functions assigned to their N- and C-terminal domains. For example, full-length PPE37 is cleaved into N- and C-terminal domains under low iron conditions (Ahmad et al., 2018). The purified N-terminal domain of PPE37 promotes the proliferation and differentiation of monocytic THP-1 cells into CD11c, dendritic cell-specific intercellular adhesion molecule 3 grabbing non-integrin (DC-SIGN) positive semi-mature dendritic cells. In contrast, the C-terminal domain of PPE37 induces apoptosis in a caspase 3-dependent manner, in transfected THP-1 macrophages (Ahmad et al., 2018). Given the reductive evolution of *M. tuberculosis,* this study further highlights how *M. tuberculosis* is able to increase its functional repertoire from a reduced number of genes/proteins.

PE-PPE family proteins possess a variety of enzymatic domains. PE11, or lipX, exhibits lipase/esterase activity and preferentially hydrolyses short to intermediate *p*-nitrophenyl esters (Singh et al., 2016), as does PE16 (Sultana et al., 2013). PE-PGRS63, or lipY, shows lipase activity and is involved in the hydrolysation of triacylglycerols (Deb, 2006). Phosphoglycerate mutase activity identified in *M. tuberculosis* is attributed to PE-PGRS11 (Chaturvedi et al., 2010). The different enzymatic features belonging to PE-PPE family proteins allow *M. tuberculosis* to process a variety of substrates for survival. It is also interesting to find that similar enzymatic/functional proteins such as PE11/lipX and PE-PGRS63/LipY, have dissimilar purposes. For example, an upregulation of PE-PGRS63/lipY during starvation points toward a role in fatty acid acquisition for energy, whereas the cell wall localisation and cell wall composition changes associated with PE11/LipX points toward a role in altering cell wall composition for stress resistance (Deb, 2006; Singh et al., 2016).

Purified PE-PGRS60 binds to fibronectin, which would allow *M. tuberculosis* to attach to the host's extracellular matrix, resulting in enhanced adhesion and invasion (Meena and Meena, 2015). Using a series of functional deletion mutants, the PGRS domain of PE-PGRS33 was shown to facilitate entry into J774 macrophages in a TLR2-dependent manner. It was also hypothesised that the interaction of PE-PGRS33 with TLR2 activates the pro-adhesive pathway via phosphatidylinositol 3-kinase (PI3K) activity, which enhances complement receptor 3 (CR3) avidity for *M. tuberculosis* (Palucci et al., 2016).

PPE38 is essential for the secretion of other PE-PPE family members, specifically the PE-PGRS and PPE-MPTR sub-family proteins. Interestingly, the loss of PE-PGRS and PPE-MPTR secretion, due to a deleted or mutated PPE38, increases the virulence of *M. tuberculosis*, thereby highlighting the virulence attenuating nature of the PE-PGRS and PPE-MPTR sub-family proteins (Ates et al., 2018a; Ates et al., 2018b).

PE6 was recently shown to possess a nucleus/nucleolus-targeting sequence, the presence of an Spt5 C-terminal nonapeptide repeat sequence responsible for binding Spt4, and the ability to bind with DNA. Further studies into these features may reveal unique functions for the PE-PPE family proteins, such as compromising host cell machinery by manipulating RNA polymerase II (Sharma et al., 2021a).

The PE-PPE family proteins are crucial for the successful establishment, maintenance, and dissemination of *M. tuberculosis* infection within the host. The functions of the PE-PPE family are extensive, including modulating host immune responses, subverting host defence strategies, resisting diverse stresses of the host, and manipulating host cell fates, ultimately to ensure the survival of *M. tuberculosis* (Fig. 6). Given that their functions are incredibly diverse, it is important to elucidate the collective role of these proteins, rather than just their individual functions. As previously mentioned, elucidating the time of utilisation for these proteins is important since these proteins have implications in all stages of infection. To make matters more complex, the co-operonic nature, or as previously speculated, the transient nature of PE-PPE protein pairing may further increase their functional diversity and perhaps, allow *M. tuberculosis* to express specific combinations of PE-PPE proteins depending on the current stage of infection or the pressures exerted by the host.

8. Diagnostic potential of the PE-PPE family

The early and accurate diagnosis of TB will greatly help in the battle against TB. However, the diagnosis of TB primarily relies upon microscopy and culturing techniques. Both techniques have their limitations, such as microscopy having a low sensitivity of ~50–60 % and culturing being highly time-consuming (Siddiqi et al., 2003). In addition, newer diagnostic techniques such as polymerase chain reactions (PCR) that are highly sensitive and rapid, are not always appropriate due to socio-economic factors such as cost and expertise (Boehme et al., 2010). Given the limitations of existing diagnostic techniques, serological diagnostic techniques, which are simple, rapid, and inexpensive, were developed for the diagnosis of TB; by using techniques such as enzyme-linked immunosorbent assays (ELISA) to measure antibody responses against TB-specific immunogenic antigens (Abebe et al., 2007; Abebe and Bjune, 2009). With their ability to elicit robust antibody responses, PE-PPE proteins are among the promising immunogenic antigens to be used for TB sero-diagnostics. The potential of PE proteins such as PE11 (Narayana et al., 2007), PE35 (Mukherjee et al., 2007), PE-PGRS62 (Koh et al., 2009b), and PPE proteins such as PPE2 (Abraham et al., 2014), PPE41 (Choudhary et al., 2003), PPE57 (Zhang et al., 2007) as immunogenic antigens for TB sero-diagnostics has been examined. Whilst PE-PPE proteins have been used to detect TB successfully, many of these proteins did not reach the necessary sensitivity to be used as an effective TB sero-diagnostic immunogenic antigen. However, it should be noted that the sensitivity of a PE or PPE protein may be increased when in complex with its cognate partner protein. For example, PPE41 was shown to have a sensitivity of 45 % in detecting TB; however, when PPE41 was in complex with its cognate partner protein PE25, its sensitivity increased to 75 % (Tundup et al., 2008). This principle may be applied to other PE-PPE proteins with already high sensitivities, in order to improve their sensitivity further.

A notable example of a highly sensitive PE-PPE protein is PPE17. Using ELISA, PPE17 was able to distinguish between *M. bovis* Bacillus Calmette-Guerin (BCG) vaccinated healthy individuals and TB-infected individuals, including those with active, latent, and extrapulmonary TB infections, with high sensitivity. In addition, PPE17 was also shown to have stronger reactivity against sera from TB patients compared to other PE-PPE proteins such as PPE2, and even other important antigens such as ESAT-6 (Khan et al., 2008; Abraham et al., 2016; Abraham et al., 2018). Interestingly, the N-terminal domain of PPE17 was found to induce a higher antibody response in TB patients compared to the C-terminal domain of PPE17. Despite the conserved and homologous N-terminal domains of the PE-PPE family, antibodies raised against the N-terminal domain of PPE17 from TB patients did not significantly cross-react with the N-terminal domains of other PPE proteins such as PPE18, PPE44 and PPE65. This is likely due to unique PPE17 N-terminal domain amino acids, which result in slight yet unique variations of the PPE17 N-terminal domain structure (Abraham et al., 2017). Whilst PE-PPE proteins such as PPE17 are excellent for distinguishing between BCG-vaccinated healthy individuals and TB-infected individuals using ELISA, these proteins are not able to accurately distinguish between active or latent TB infections. Interestingly, IL-2 responses stimulated by antigens such as PE35 and PPE68 were recently shown to distinguish between active and latent TB infections. Effector T-cells, which are more associated with active infections, secrete IFN-ɣ, whereas memory T-cells, which are more associated with latent infections, secrete IL-2, or IFN-ɣ and IL-2 (Sargentini et al., 2009). Therefore, by measuring the IL-2 response following stimulation with PE35 and PPE68, it is possible to distinguish between active and latent TB infections (Pourakbari et al., 2015).

Aside from TB sero-diagnostics, PE-PPE genes such as *PPE8* are also being used in PCR testing. Amplification of a 1291 base pair fragment from *PPE18* was able to detect *M. tuberculosis* from sputum, pleural, and

cerebral spinal fluid samples with impressively high sensitivity compared to other diagnostic techniques such as culturing (PCR = 90.4 %, 77.7 % and 70 %, respectively, vs Culture = 62.9 %, 24.4 % and 10 %, respectively) (Srivastava et al., 2006). In addition, PE-PPE proteins such as PE5 have also been recently used as diagnostic antigens in intradermal tests for bovine TB (Melo et al., 2015).

9. Vaccine potential of the PE-PPE family

PE-PPE proteins are rich in immunogenic epitopes, and therefore, they are promising candidates for the development of novel TB vaccines (Chaitra et al., 2005; Vordermeier et al., 2012; Copin et al., 2014; Stylianou et al., 2018). Promising PE-PPE proteins have been implemented into novel recombinant, attenuated/whole-cell, and subunit vaccine candidates.

PE-PPE proteins have been used in novel recombinant vaccine candidates. For example, the PE N-terminal domain has been shown to be responsible for cell wall localisation. Therefore, this N-terminal domain may be used as an N-terminal fusion protein to generate novel recombinant *Mycobacteria* isolates expressing heterologous antigens on their surface (Cascioferro et al., 2007). The PE domain of PE-PGRS33 was added to the MPT65 antigen of *M. bovis BCG* as an N-terminal fusion protein. The recombinant *M. bovis BCG* strain was then able to express the MPT65 antigen on its surface, resulting in enhanced protection against *M. tuberculosis* infection in mice, compared to the parental *M. bovis BGC* strain (Sali et al., 2010).

PE-PPE proteins have also played an important role in the development of attenuated/whole-cell vaccine candidates. For example, the preclinical *M. tuberculosis* Δppe25-pe19 vaccine candidate is attenuated via a partial deletion of the ESX-5 genetic locus, involving three PPE genes (PPE25, PPE26 and PPE27) and two PE genes (PE18 and PE19) (Bottai et al., 2012). The *M. tuberculosis* Δppe25-pe19 vaccine candidate was shown to provide enhanced protection compared to *M. bovis* BCG in murine efficacy studies. This enhanced protection is likely due to the *M. tuberculosis* Δppe25-pe19 strain being able to elicit a robust $CD4^+$ T-cell response against key antigens such as ESAT-6-CFP-10, which would not be possible using the *M. bovis* BCG vaccine due to the attenuating RD1 deletion (Hsu et al., 2003; Sayes et al., 2012). Furthermore, whilst this deletion results in the loss of ESX-5-associated PE-PPE proteins, the *M. tuberculosis* Δppe25-pe19 strain is still able to elicit a robust $CD4^+$ T-cell response against other ESX-associated PE-PPE proteins. Given the homologous nature of the PE-PPE family, this may allow for the cross-recognition of the deleted ESX-5-associated PE-PPE proteins, thereby minimizing the loss of the antigenic repertoire of the *M. tuberculosis* Δppe25-pe19 strain (Sayes et al., 2016).

PE-PPE proteins have also been used in the development of novel subunit vaccine candidates. A notable example is GSK's M72/$AS01_E$ vaccine, which utilises a recombinant fusion protein derived from *M. tuberculosis* antigens, MTB32A (PepA) and MTB39A (PPE18), in combination with an $AS01_E$ adjuvant (Skeiky et al., 2004). A recent phase IIb trial tested the vaccine's efficacy against pulmonary TB in adults with latent TB infections (Van Der Meeren et al., 2018). The results demonstrated a reduction in the incidence of TB, as well as an overall vaccine efficacy of 50%; however, this is still unsatisfactory. Individuals who received the vaccine also showed higher levels of M72/$AS01_E$ -specific antibodies and $CD4^+$ T-cells, throughout a 3-year period (Tait et al., 2019).

PE-PPE proteins have been used in other vaccine strategies. A fusion protein including both acute- (PPE44 and EsxV) and latent-phase (HspX) antigens, was recently encapsulated in liposomes containing dimethyldioctadecylammonium (DDA)/trehalose-6,6′-dibehenate (TDB) and shown to induce a robust T_h1 response, as well as enhance the efficacy of BCG when administered in combination (Mansury et al., 2019). Interestingly, PPE44 was previously used in a plasmid DNA vaccine alongside the BCG, which resulted in stronger T- and B-cell responses compared to the BCG alone. This concept may be altered by replacing PPE44 with other PE-PPE antigen(s) and/or by using the plasmid DNA vaccine in combination with vaccines other than the BCG, in order to improve immune responses (Romano et al., 2008; Bruffaerts et al., 2014).

The development of novel vaccine candidates primarily focuses on eliciting a T_h1 cellular response against one or, ideally, more immunogenic antigens. This approach aims to promote the recruitment of protective T-cells and the release of IFN-ɣ upon challenge with *M. tuberculosis* (Delogu et al., 2014). However, recent evidence suggests that the T_h1 response alone may not be sufficient to protect against TB. The neglected B-cell response may also be required to fully prime the host immune response against *M. tuberculosis* (Chan et al., 2014). Therefore, novel vaccine candidates should focus on eliciting a B-cell response, specifically against the surface-associated or secreted immunogenic antigens of *M. tuberculosis*. PE-PPE proteins are candidates for eliciting B-Cell responses. Several PE-PPE proteins have been found to elicit humoral immune responses including PE4 (Singh et al., 2012), PE11, PE-PGRS17 and PE-PGRS33, (Narayana et al., 2007), PE-PGRS62 (Koh et al., 2009b), PPE2 (Abraham et al., 2014), PPE17 (Khan et al., 2008), PPE41 (Choudhary et al., 2003), PPE42 (Chakhaiyar et al., 2004), and PE25-PPE41 (Tundup et al., 2008). However, members of the PE-PGRS sub-family are of particular interest, as antibodies targeting the PGRS domain have been detected in both TB-infected humans and animals (Delogu and Brennan, 2001). PE-PGRS33 is currently of great interest as a potential component for a novel subunit vaccine candidate and has been shown to elicit robust humoral responses (Cohen et al., 2014; Gastelum Aviña et al., 2015).

Whilst the PE-PPE family possesses promising candidates for the development of novel vaccines, there are also potential consequences to their use. Firstly, PE-PPE proteins have been shown to modulate the host immune response. For example, PE-PPE proteins have been shown to suppress pro-inflammatory cytokines and promote anti-inflammatory cytokines, ultimately dampening the T_h1 response and supporting intracellular survival (Khubaib et al., 2016). Therefore, careful consideration should be taken when selecting immunogenic antigens to prime the host's immune response. Secondly, PE-PPE proteins such as the PE-PGRS and PPE-MPTR sub-family proteins are not present in every *M. tuberculosis* strain. For example, hypervirulent *M. tuberculosis* strains with PPE38 mutations are not able to secrete the PE-PGRS and PPE-MPTR sub-family proteins (Ates et al., 2018a; Ates et al., 2018b). Therefore, novel vaccine candidates that prime the host immune response against PE-PGRS and PPE-MPTR proteins may not effectively recognise and protect against the PPE38-deficient *M. tuberculosis* strains that lack these proteins, thereby allowing other more hypervirulent strains to circulate (Hanekom et al., 2011). Lastly, given the polymorphic nature of PE-PPE proteins, it is important to consider the genetic variability of the protein when designing a novel vaccine candidate (McEvoy et al., 2012; Phelan et al., 2016). An example of this involves the previously mentioned GSK M72/$AS01_E$ subunit vaccine. A study by Homolka et al. (2016) investigated the variability of the PPE18 amongst 71 MTBC species and strains. The subunit vaccine component unveiled 96 single nucleotide polymorphisms (68 non-synonymous vs 28 synonymous), as well as three insertions and one deletion. It is important to consider genetic variability when designing a novel vaccine candidate, as genetic variability may result in structural variability, which may generate unrecognizable antigens compared to the vaccine antigen, thus reducing the vaccine's efficacy (Hakim and Yang, 2021). Therefore, the variation observed in PPE18 may explain the limited 50 % vaccine efficacy observed in the recent phase IIb trial (Van Der Meeren et al., 2018; Tait et al., 2019). In addition, non-synonymous homoplastic SNPs were recently identified in PPE18, specifically in isolates belonging to lineage 1, suggesting that the H37Rv (lineage 4) derived vaccine candidate, may have variable efficacy in populations exposed to lineage 1 isolates (Tantivitayakul et al., 2020). To prevent such issues, bioinformatics could first be used to assess the variability of antigens, and experimental tests could then be used to measure the protection of immunogenic antigens against divergent strains of *M. tuberculosis*. Furthermore, a

reference collection of MTBC species and isolates representing global MTBC phylogeny should be established and used universally to assess the protective potential of novel vaccine candidates.

10. Research challenges for studying the PE-PPE family

Whilst significant amounts of research have been conducted into understanding these unique proteins, progress is difficult due to the many challenges raised by these unusual proteins. It has proven difficult to obtain soluble and stable recombinant PE-PPE proteins. As a result, our knowledge of their structure remains limited, with only a few PE-PPE protein crystal structures having been solved to date. This may be due to the requirement of partner and/or chaperone protein, as well as their high structural disorder (Strong et al., 2006; Ekiert and Cox, 2014; Korotkova et al., 2014; Chen et al., 2017; Ahmad et al., 2019; Williamson et al., 2020). Many commonly used techniques, such as mass spectrometry, are also hindered due to the highly homologous nature of PE-PPE proteins, as well as their localization and lack of trypsin cleavage sites (Banu et al., 2002).

PE-PPE genes contain a GC content of up to ~80 %, making tasks such as sequencing, aligning, and cloning difficult. In addition, multiple gene duplications, as well as the highly repetitive nature of PE-PPE genes, exacerbate such challenges (Hermans et al., 1992; Poulet and Cole, 1995; Cole et al., 1998; Gey Van Pittius et al., 2006; McEvoy et al., 2012). For these reasons, PE-PPE genes are often omitted from bioinformatic studies (Meehan et al., 2019). However, modern longer-read sequencing techniques such as MinION have been recently demonstrated to sequence the GC-rich and highly repetitive PE-PPE family genes at a higher resolution compared to shorter-read sequencing techniques such as Illumina. For example, twice as many SNPs were found in PE-PGRS genes using MinION compared to Illumina (Quick et al., 2016; Bainomugisa et al., 2018). In addition to offering higher resolution sequencing, modern sequencing techniques such as MinION also offer a simpler and cheaper technique that may be more suitable in less socio-economically developed settings, where *M. tuberculosis* and drug-resistant isolates are more prevalent.

PE-PPE family members are often described as redundant and therefore, knock-out and overexpression experiments focussing on single PE-PPE genes may not be useful to elucidate function. However, as previously mentioned, whilst PE-PPE members such as PPE2 and PE-PGRS62 both confer resistance against nitric oxide, they achieve this using two distinct mechanisms (Thi et al., 2012; Bhat et al., 2017). Therefore, it is important to research these proteins further and elucidate the exact mechanism behind every function observed. This also further highlights the importance of elucidating the collective function of PE-PPE members, perhaps those belonging to the same sub-family and/or those regulated in parallel, rather than just individual functions.

11. Concluding remarks

This review highlights the significant progress that has been made over the last two decades in our understanding of the PE-PPE family. There is a growing body of evidence showing that the PE-PPE family is crucial in the establishment, maintenance, and dissemination of *M. tuberculosis* infection. Furthermore, evidence shows that the PE-PPE family proteins allow *M. tuberculosis* to infect the host with a strategic approach, by elegantly controlling and adapting to the host, which reflects the co-evolution of *M. tuberculosis* with humans. This review also highlights dogma inconsistencies, knowledge gaps, as well as numerous important questions to be answered. Whilst explanations have been explored, further research is greatly required to provide clarity and further understanding of the PE-PPE family, which will also likely inspire the development of novel drugs, diagnostics, and vaccines against *M. tuberculosis*. As technology and methodologies continue to advance, it is still of the utmost importance to consider the challenges associated with studying the PE-PPE family.

The PE-PPE family is a lucrative area of research, with immeasurable breakthroughs yet to be made. With our current knowledge of the PE-PPE family and our technological advancements, this decade will likely be a time of exponential progress.

CRediT Authorship Contribution Statement

Christopher D'Souza: Writing – original draft. **Uday Kishore:** Writing – review & editing. **Anthony G. Tsolaki:** Conceptualization, Writing – review & editing.

Declaration of Funding

The authors declare no funding was used for the work reported in this paper.

Data Availability

The authors declare no data was used for the work reported in this paper.

Dedication

Anthony G. Tsolaki would like to dedicate this article to the memory of Professor Robert B. Sim. Bob was a mentor and friend who greatly influenced my interest in studying the innate immunity of tuberculosis. Not only was Bob incredibly generous with his time and resources, but he also provided invaluable guidance and support on the studies that we worked on together. Bob was also delightful company, and I will always remember our lunches where we would discuss not only science, but also a range of topics with humour and memorable anecdotes. Bob made a significant contribution to my scientific career, as well as the careers of countless others in the field of innate immunity. He will be greatly missed.

Declaration of Competing Interest

The authors declare no competing financial or personal interests influenced the work reported in this paper.

Acknowledgements

Figs. 1–6 and the graphical abstract were created using BioRender and PyMol.

References

Abdallah, A.M., Verboom, T., Hannes, F., Safi, M., Strong, M., Eisenberg, D., Musters, R.J. P., Vandenbroucke Grauls, C.M.J.E., Appelmelk, B.J., Luirink, J., Bitter, W., 2006. A specific secretion system mediates PPE41 transport in pathogenic *Mycobacteria*. Mol. Microbiol. 62 (3), 667–679.

Abdallah, A.M., Gey Van Pittius, N.C., DiGiuseppe Champion, P.A., Cox, J., Luirink, J., Vandenbroucke Grauls, C.M.J.E., Appelmelk, B.J., Bitter, W., 2007. Type VII secretion — *Mycobacteria* show the way. Nat. Rev. Microbiol. 5 (11), 883–891.

Abdallah, A.M., Verboom, T., Weerdenburg, E.M., Gey Van Pittius, N.C., Mahasha, P.W., Jiménez, C., Parra, M., Cadieux, N., Brennan, M.J., Appelmelk, B.J., Bitter, W., 2009. PPE and PE_PGRS proteins of *Mycobacteria* marinum are transported via the type VII secretion system ESX-5. Mol. Microbiol. 73 (3), 329–340.

Abebe, F., Bjune, G., 2009. The protective role of antibody responses during *M. tuberculosis* infection. Clin. Exp. Immunol. 157 (2), 235–243.

Abebe, F., Holm Hansen, C., Wiker, H.G. and Bjune, G. (2007). Progress in serodiagnosis of *M. tuberculosis* infection. *Scand. J. Immunol.*, 66(2-3), 176–191.

Abo-Kadoum, M.A., Assad, M., Ali, M.K., Uae, M., Nzaou, S.A.E., Gong, Z., Moaaz, A., Lambert, N., Eltoukhy, A., Xie, J., 2021. *M. tuberculosis* PE17 (Rv1646) promotes host cell apoptosis via host chromatin remodeling mediated by reduced H3K9me3 occupancy. Microb. Pathog. 159, 105147.

Abraham, P.R., Pathak, N., Pradhan, G., Sumanlatha, G. and Mukhopadhyay, S. (2017). The N-terminal domain of *M. tuberculosis* PPE17 (Rv1168c) protein plays a dominant role in inducing antibody responses in active TB patients. *PLoS ONE*, 12(6), p.0179965.

Abraham, P.R., Devalraju, K.P., Jha, V., Valluri, V.L. and Mukhopadhyay, S. (2018). PPE17 (Rv1168c) Protein of *M. tuberculosis* detects individuals with latent TB infection. *PLoS ONE*, 13(11), p.0207787.

Abraham, P.R., Latha, G.S., Valluri, V.L., Mukhopadhyay, S., 2014. *M. tuberculosis* PPE protein Rv0256c induces strong B cell response in tuberculosis patients. Infect. Genet. Evol. 22, 244–249.

Abraham, P.R., Udgata, A., Latha, G.S., Mukhopadhyay, S., 2016. The *M. tuberculosis* PPE protein Rv1168c induces stronger B cell response than Rv0256c in active TB patients. Infect. Genet. Evol. 40, 339–345.

Ahmad, J., Farhana, A., Pancsa, R., Arora, S.K., Srinivasan, A., Tyagi, A.K., Babu, M.M., Ehtesham, N.Z. and Hasnain, S.E. (2018). Contrasting function of structured N-terminal and unstructured C-terminal segments of *M. tuberculosis* PPE37 protein. *mBio*, 9(1).

Ahmad, J., Khubaib, M., Sheikh, J.A., Pancsa, R., Kumar, S., Srinivasan, A., Babu, M.M., Hasnain, S.E., Ehtesham, N.Z., 2019. Disorder-to-order transition in PE–PPE proteins of *M. tuberculosis* augments the pro-pathogen immune response. FEBS Open Bio 10 (1), 70–85.

Ahmed, N., Dobrindt, U., Hacker, J., Hasnain, S.E., 2008. Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. *Nat. Rev. Microbiol.* 6 (5), 387–394.

Ali, M.K., Zhen, G., Nzungize, L., Stojkoska, A., Duan, X., Li, C., Duan, W., Xu, J., Xie, J., 2020. *M. tuberculosis* PE31 (Rv3477) attenuates host cell apoptosis and promotes recombinant M. smegmatis intracellular survival via up-regulating GTPase guanylate binding protein-1. Front. Cell. Infect. Microbiol. 10.

Asaad, M., Kaisar Ali, M., Abo Kadoum, M.A., Lambert, N., Gong, Z., Wang, H., Uae, M., Nazou, S.A.E., Kuang, Z., Xie, J., 2021. *M. tuberculosis* PPE10 (Rv0442c) alters host cell apoptosis and cytokine profile via linear ubiquitin chain assembly complex HOIP-NF-κB signaling axis. Int. Immunopharmacol. 94, 107363.

Ates, L.S., Houben, E., Bitter, W., 2015a. Type VII secretion: A highly versatile secretion system. Microbiol. Spectrum 4 (1).

Ates, L.S., Ummels, R., Commandeur, S., Van Der Weerd, R., Sparrius, M., Weerdenburg, E., Alber, M., Kalscheuer, R., Piersma, S.R., Abdallah, A.M., Abd El Ghany, M., Abdel Haleem, A.M., Pain, A., Jiménez, C.R., Bitter, W., Houben, E.N.G., 2015b. Essential role of the ESX-5 secretion system in outer membrane permeability of pathogenic *Mycobacteria*. PLoS Genet. 11 (5), 1005190.

Ates, L.S., Dippenaar, A., Ummels, R., Piersma, S.R., Van Der Woude, A.D., Van Der Kuij, K., Le Chevalier, F., Mata Espinosa, D., Barrios Payán, J., Marquina Castillo, B., Guapillo, C., Jiménez, C.R., Pain, A., Houben, E.N.G., Warren, R.M., Brosch, R., Hernández Pando, R., Bitter, W., 2018a. Mutations in PPE38 Block PE_PGRS secretion and increase virulence of *M. tuberculosis*. Nat. Microbiol. 3 (2), 181–188.

Ates, L.S., Sayes, F., Frigui, W., Ummels, R., Damen, M.P.M., Bottai, D., Behr, M.A., Van Heijst, J.W.J., Bitter, W., Majlessi, L., Brosch, R., 2018b. RD5-mediated lack of PE_PGRS and PPE-MPTR export in BCG vaccine strains results in strong reduction of antigenic repertoire but little impact on protection. PLoS Pathog. 14 (6), 1007139.

Bainomugisa, A., Duarte, T., Lavu, E., Pandey, S., Coulter, C., Marais, B.J. and Coin, L.M. (2018). A complete high-quality MinION nanopore assembly of an extensively drug-resistant *M. tuberculosis* Beijing lineage strain identifies novel variation in repetitive PE-PPE gene regions. *Microb. Genom.*, 4(7).

Bansal, K., Elluru, S.R., Narayana, Y., Chaturvedi, R., Patil, S.A., Kaveri, S.V., Bayry, J., Balaji, K.N., 2010a. PE_PGRS antigens of *M. tuberculosis* induce maturation and activation of human dendritic cells. J. Immunol. 184 (7), 3495–3504.

Bansal, K., Sinha, A.Y., Ghorpade, D.S., Togarsimalemath, S.K., Patil, S.A., Kaveri, S.V., Balaji, K.N., Bayry, J., 2010b. Src homology 3-interacting domain of Rv1917c of *M. tuberculosis* induces selective maturation of human dendritic cells by regulating PI3K-MAPK-NF-κB signaling and drives Th2 immune responses. J. Biol. Chem. 285 (47), 36511–36522.

Banu, S., Honoré, N., Saint Joanis, B., Philpott, D., Prévost, M.C., Cole, S.T., 2002. Are the PE-PGRS proteins of *M. tuberculosis* variable surface antigens? Mol. Microbiol. 44 (1), 9–19.

Basu, S., Pathak, S.K., Banerjee, A., Pathak, S., Bhattacharyya, A., Yang, Z., Talarico, S., Kundu, M., Basu, J., 2007. Execution of macrophage apoptosis by PE_PGRS33 of *M. tuberculosis* is mediated by toll-like receptor 2-dependent release of tumor necrosis factor-α. J. Biol. Chem. 282 (2), 1039–1050.

Bergsbaken, T., Fink, S.L., Cookson, B.T., 2009. Pyroptosis: host cell death and inflammation. Nat. Rev. Microbiol. 7 (2), 99–109.

Bhat, K.H., Srivastava, S., Kotturu, S.K., Ghosh, S., Mukhopadhyay, S., 2017. The PPE2 protein of *M. tuberculosis* translocates to host nucleus and inhibits nitric oxide production. Sci. Rep. 7 (1).

Boehme, C.C., Nabeta, P., Hillemann, D., Nicol, M.P., Shenai, S., Krapp, F., Allen, J., Tahirli, R., Blakemore, R., Rustomjee, R., Milovic, A., Jones, M., O'Brien, S.M., Persing, D.H., Ruesch Gerdes, S., Gotuzzo, E., Rodrigues, C., Alland, D., Perkins, M. D., 2010. Rapid molecular detection of tuberculosis and rifampin resistance. N. Engl. J. Med. 363 (11), 1005–1015.

Bottai, D., Di Luca, M., Majlessi, L., Frigui, W., Simeone, R., Sayes, F., Bitter, W., Brennan, M.J., Leclerc, C., Batoni, G., Campa, M., Brosch, R., Esin, S., 2012. Disruption of the ESX-5 system of *M. tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol. Microbiol. 83 (6), 1195–1209.

Bruffaerts, N., Romano, M., Denis, O., Jurion, F., Huygen, K., 2014. Increasing the vaccine potential of live M. bovis BCG by co-administration with plasmid DNA encoding a tuberculosis prototype antigen. Vaccines 2 (1), 181–195.

Bunduc, C.M., Ummels, R., Bitter, W., Houben, E.N.G., 2020. Species-specific secretion of ESX-5 type VII substrates is determined by the linker 2 of EccC5. Mol. Microbiol. 114 (1), 66–76.

Cascioferro, A., Delogu, G., Colone, M., Sali, M., Stringaro, A., Arancia, G., Fadda, G., Palù, G., Manganelli, R., 2007. PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. Mol. Microbiol. 66 (6), 1536–1547.

Chai, Q., Wang, X., Qiang, L., Zhang, Y., Ge, P., Lu, Z., Zhong, Y., Li, B., Wang, J., Zhang, L., Zhou, D., Li, W., Dong, W., Pang, Y., Gao, G.F., Liu, C.H., 2019. A *M. tuberculosis* surface protein recruits ubiquitin to trigger host xenophagy. Nat. Commun. 10 (1).

Chai, Q., Wang, L., Liu, C.H., Ge, B., 2020. New insights into the evasion of host innate immunity by *M. tuberculosis*. Cell. Mol. Immunol. 17 (9), 901–913.

Chaitra, M.G., Hariharaputran, S., Chandra, N.R., Shaila, M.S., Nayak, R., 2005. Defining putative T cell epitopes from PE and PPE families of proteins of *M. tuberculosis* with vaccine potential. Vaccine 23 (10), 1265–1272.

Chakhaiyar, P., Nagalakshmi, Y., Aruna, B., Murthy, K.J.R., Katoch, V.M., Hasnain, S.E., 2004. Regions of high antigenicity within the hypothetical PPE major polymorphic tandem repeat open-reading frame, Rv2608, show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. J Infect Dis 190 (7), 1237–1244.

Chan, J., Mehta, S., Bharrhan, S., Chen, Y., Achkar, J.M., Casadevall, A., Flynn, J., 2014. The role of B cells and humoral immunity in *M. tuberculosis* infection. Semin. Immunol. 26 (6), 588–600.

Chatrath, S., Gupta, V.K., Dixit, A., Garg, L.C., 2016. PE_PGRS30 of *M. tuberculosis* mediates suppression of proinflammatory immune response in macrophages through its PGRS and PE domains. Microbes Infect. 18 (9), 536–542.

Chaturvedi, R., Bansal, K., Narayana, Y., Kapoor, N., Sukumar, N., Togarsimalemath, S. K., Chandra, N., Mishra, S., Ajitkumar, P., Joshi, B., Katoch, V.M., Patil, S.A., Balaji, K.N., 2010. The multifunctional PE_PGRS11 protein from *M. tuberculosis* plays a role in regulating resistance to oxidative stress. J. Biol. Chem. 285 (40), 30389–30403.

Chen, X., Cheng, H., Zhou, J., Chan, C., Lau, K., Tsui, S.K.W., Au, S.W.N., 2017. Structural basis of the PE–PPE protein interaction in *M. tuberculosis*. J. Biol. Chem. 292 (41), 16880–16890.

Choi, H.H., Kwon, K.W., Han, S.J., Kang, S.M., Choi, E., Kim, A., Cho, S.N., Shin, S.J., 2019. A novel PPE39 from *M. tuberculosis* strain Beijing/K induces Th1 polarization via dendritic cell maturation. J. Cell Sci. 132 (17).

Choudhary, R.K., Mukhopadhyay, S., Chakhaiyar, P., Sharma, N., Murthy, K.J.R., Katoch, V.M., Hasnain, S.E., 2003. PPE antigen Rv2430c of *M. tuberculosis* induces a strong B-cell response. Infect. Immun. 71 (11), 6338–6343.

Cohen, I., Parada, C., Acosta Gao, E., Espitia, C., 2014. The PGRS domain from PE_PGRS33 of *M. tuberculosis* is target of humoral immune response in mice and humans. Front. Immunol. 5.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.D., 1998. Deciphering the biology of *M. tuberculosis* from the complete genome sequence. Nature 393 (6685), 537–544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., 2001. Massive gene decay in the leprosy Bacillus. Nature 409 (6823), 1007–1011.

Copin, R., Coscollá, M., Seiffert, S.N., Bothamley, G., Sutherland, J., Mbayo, G., Gagneux, S., Ernst, J.D., 2014. Sequence diversity in the PE_PGRS genes of *M. tuberculosis* is independent of human T cell recognition. *mBio* 5 (1).

Cui, Z.J., Yang, Q.Y., Zhang, H.Y., Zhu, Q., Zhang, Q.Y., 2016. Bioinformatics identification of drug resistance-associated gene pairs in *M. tuberculosis*. Int. J. Mol. Sci. 17 (9), 1417.

Daim, S., Kawamura, I., Tsuchiya, K., Hara, H., Kurenuma, T., Shen, Y., Dewamitta, S.R., Sakai, S., Nomura, T., Qu, H., Mitsuyama, M., 2011. Expression of the *M. tuberculosis* PPE37 protein in *Mycobacteria* smegmatis induces low tumour necrosis factor alpha and interleukin 6 production in murine macrophages. J. Med. Microbiol. 60 (5), 582–591.

Daleke, M.H., Cascioferro, A., De Punder, K., Ummels, R., Abdallah, A.M., Van Der Wel, N., Peters, P.J., Luirink, J., Manganelli, R., Bitter, W., 2011. Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) protein domains target LipY lipases of pathogenic *Mycobacteria* to the cell surface via the ESX-5 pathway. J. Biol. Chem. 286 (21), 19024–19034.

Daleke, M.H., Ummels, R., Bawono, P., Heringa, J., Vandenbroucke Grauls, C.M.J.E., Luirink, J., Bitter, W., 2012a. General secretion signal for the mycobacterial type VII secretion pathway. Proc. Natl. Acad. Sci. 109 (28), 11342–11347.

Daleke, M.H., Van Der Woude, A.D., Parret, A.H.A., Ummels, R., ed Groot, A.Marit., Watson, D., Piersma, S.R., Jiménez, C.R., Luirink, J., Bitter, W. and Houben, E.N.G. (2012b). Specific chaperones for the type VII protein secretion pathway. *J. Biol. Chem.*, 287(38), 31939–31947.

Danelishvili, L., Everman, J., Bermudez, L.E., 2015. *M. tuberculosis* PPE68 and Rv2626c genes contribute to the host cell necrosis and bacterial escape from macrophages. Virulence 7 (1), 23–32.

de Jonge, M.I., Pehau Arnaudet, G., Fretz, M.M., Romain, F., Bottai, D., Brodin, P., Honoré, N., Marchal, G., Jiskoot, W., England, P., Cole, S.T., Brosch, R., 2007. ESAT-6 from *M. tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. J. Bacteriol. 189 (16), 6028–6034.

De Martino, M., Lodi, L., Galli, L., Chiappini, E., 2019. Immune response to *M. tuberculosis*: a narrative review. Front. Pediatr. 7, 350.

De Souza, G.A., Leversen, N.A., Målen, H., Wiker, H.G., 2011. Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. J. Proteomics 75 (2), 502–510.

Deb, C., 2006. A novel lipase belonging to the hormone-sensitive lipase family induced under starvation to utilize stored triacylglycerol in *M. tuberculosis*. J. Biol. Chem. 281 (7), 3866–3875.
Delogu, G., Brennan, M.J., 2001. Comparative immune response to PE and PE_PGRS antigens of *M. tuberculosis*. Infect. Immun. 69 (9), 5606–5611.
Delogu, G., Manganelli, R., Brennan, M.J., 2014. Critical research concepts in tuberculosis vaccine development. Clin. Microbiol. Infect. 20 (5), 59–65.
Deng, W., Li, W., Zeng, J., Zhao, Q., Li, C., Zhao, Y., Xie, J., 2014. *M. tuberculosis* PPE family protein Rv1808 manipulates cytokines profile via co-activation of MAPK and NF-κB signaling pathways. Cell. Physiol. Biochem. 33 (2), 273–288.
Deng, W., Zeng, J., Xiang, X., Li, P., Xie, J., 2015. PE11 (Rv1169c) selectively alters fatty acid components of *Mycobacteria* smegmatis and host Cell IL-6 level accompanied with cell death. Front. Microbiol. 6.
Deng, W., Yang, W., Zeng, J., Abdalla, A.E., Xie, J., 2016. *M. tuberculosis* PPE32 promotes cytokines production and host cell apoptosis through caspase cascade accompanying with enhanced ER stress response. Oncotarget 7 (41), 67347–67359.
Deng, W., Long, Q., Zeng, J., Li, P., Yang, W., Chen, X., Xie, J., 2017. *M. tuberculosis* PE_PGRS41 enhances the intracellular survival of *M.* smegmatis within macrophages via blocking innate immunity and inhibition of host defense. Sci. Rep. 7 (1).
Dheenadhayalan, V., Delogu, G., Brennan, M.J., 2006. Expression of the PE_PGRS 33 protein in *Mycobacteria* smegmatis triggers necrosis in macrophages and enhanced mycobacterial survival. Microbes Infect. 8 (1), 262–272.
Di Luca, M., Bottai, D., Batoni, G., Orgeur, M., Aulicino, A., Counoupas, C., Campa, M., Brosch, R., Esin, S., 2012. The ESX-5 associated eccB5-eccC5 locus is essential for *M. tuberculosis* viability. PLoS One 7 (12), 52059.
Dobos, K.M., Spotts, E.A., Quinn, F.D., King, C.H., 2000. Necrosis of lung epithelial cells during infection with *M. tuberculosis* is preceded by cell permeation. Infect. Immun. 68 (11), 6300–6310.
Dolasia, K., Nazar, F., Mukhopadhyay, S., 2020. *M. tuberculosis* PPE18 protein inhibits MHC class II antigen presentation and B cell response in mice. Eur. J. Immunol. 51 (3), 603–619.
Dong, D., Wang, D., Li, M., Wang, H., Yu, J., Wang, C., Liu, J., Gao, Q., 2011. PPE38 modulates the innate immune response and is required for *Mycobacteria* marinum virulence. Infect. Immun. 80 (1), 43–54.
Dumas, E., Christina Boritsch, E., Vandenbogaert, M., Rodríguez De La Vega, R.C., Thiberge, J.M., Caro, V., Gaillard, J.-L., Heym, B., Girard Misguich, F., Brosch, R., Sapriel, G., 2016. Mycobacterial pan-genome analysis suggests important role of plasmids in the radiation of type VII secretion systems. Genome Biol. Evol. 8 (2), 387–402.
Ekiert, D.C., Cox, J.S., 2014. Structure of a PE–PPE–EspG complex from *M. tuberculosis* reveals molecular specificity of ESX protein secretion. PNAS 111 (41), 14758–14763.
Fishbein, S., Van Wyk, N., Warren, R.M., Sampson, S.L., 2015. Phylogeny to function: PE-PPE protein evolution and impact on *M. tuberculosis* pathogenicity. Mol. Microbiol. 96 (5), 901–916.
Gastelum Aviña, P., Velazquez, C., Espitia, C., Lares-Villa, F., Garibay Escobar, A., 2015. A PE_PGRS33 protein of *M. tuberculosis*: an ideal target for future tuberculosis vaccine design. Expert Rev. Vaccines 14 (5), 699–711.
Gey Van Pittius, N.C., Gamieldien, J., Hide, W., Brown, G.D., Siezen, R.J., Beyers, A.D., 2001. The ESAT-6 gene cluster of *M. tuberculosis* and other high G+C Gram-positive bacteria. Genome Biol. 2.
Gey Van Pittius, N.C., Sampson, S.L., Lee, H., Kim, Y., Van Helden, P.D., Warren, R.M., 2006. Evolution and expansion of the *M. tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (ESX) gene cluster regions. BMC Evol. Biol. 6, 95.
Gong, Z., Kuang, Z., Li, H., Li, C., Ali, M.K., Huang, F., Li, P., Li, Q., Huang, X., Ren, S., Li, J., Xie, J., 2019. Regulation of host cell pyroptosis and cytokines production by *M. tuberculosis* effector PPE60 requires LUBAC mediated NF-κB signalling. Cell. Immunol. 335, 41–50.
Grover, S., Sharma, T., Singh, Y., Kohli, S., P., M., Singh, A., Semmler, T., Wieler, L.H., Tedin, K., Ehtesham, N.Z., Hasnain, S.E. (2018). The PGRS domain of *M. tuberculosis* PE_PGRS protein Rv0297 Is involved in endoplasmic reticulum stress-mediated apoptosis through toll-like receptor 4. *mBio*, 9(3).
Gutierrez, M.G., Master, S.S., Singh, S.B., Taylor, G.A., Colombo, M.I., Deretic, V., 2004. Autophagy is a defense mechanism inhibiting BCG and *M. tuberculosis* survival in infected macrophages. Cell 119 (6), 753–766.
Hakim, J.M.C., Yang, Z., 2021. Predicted structural variability of *M. tuberculosis* PPE18 protein with immunological implications among clinical strains. Front. Microbiol. 11.
Hanekom, M., Gey Van Pittius, N.C., McEvoy, C., Victor, T.C., Van Helden, P.D., Warren, R.M., 2011. *M. tuberculosis* Beijing Genotype: a template for success. Tuberculosis 91 (6), 510–523.
Hang, N.T.L., Hijikata, M., Maeda, S., Thuong, P.H., Ohashi, J., Van Huan, H., Hoang, N. P., Miyabayashi, A., Cuong, V.C., Seto, S., Van Hung, N., Keicho, N., 2019. Whole genome sequencing, analyses of drug resistance-conferring mutations, and correlation with transmission of *M. tuberculosis* carrying katG-S315T in Hanoi, Vietnam. *Sci. Reports* 9 (1).
Harding, C.V., Boom, W.H., 2010. Regulation of antigen presentation by *M. tuberculosis*: A role for toll-like receptors. Nat. Rev. Microbiol. 8 (4), 296–307.
Hermans, P.W., van Soolingen, D., Van Embden, J.D., 1992. Characterization of a Major Polymorphic Tandem Repeat in *M. tuberculosis* and Its Potential Use in the Epidemiology of *Mycobacteria* Kansasii and *Mycobacteria* gordonae. J. Bacteriol. 174 (12), 4157–4165.
Homolka, S., Ubben, T., Niemann, S., 2016. High sequence variability of the PPE18 gene of clinical *M. tuberculosis* complex strains potentially impacts effectivity of vaccine candidate M72/AS01E. PLoS One 11 (3), 0152200.
Hsu, T., Hingley Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M., Eisenberg, D., Russell, R.G., Derrick, S.C., Collins, F.M., Morris, S.L., King, C.H., Jacobs, W.R. (2003). The primary mechanism of attenuation of Bacillus Calmette–Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc. Natl. Acad. Sci. U. S. A.*, 100 (21), 12420–12425.
Iantomasi, R., Sali, M., Cascioferro, A., Palucci, I., Zumbo, A., Soldini, S., Rocca, S., Greco, E., Maulucci, G., De Spirito, M., Fraziano, M., Fadda, G., Manganelli, R., Delogu, G., 2011. PE_PGRS30 is required for the full virulence of *M. tuberculosis*. Cell. Microbiol. 14 (3), 356–367.
Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Ōmura, S., 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat. Biotechnol. 21 (5), 526–531.
Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., Shiba, T., Hattori, M., 2004. The complete genomic sequence of *Nocardia farcinica* IFM 10152. Proc. Natl. Acad. Sci. 101 (41), 14925–14930.
Kanji, A., Hasan, Z., Ali, A., McNerney, R., Mallard, K., Coll, F., Hill Cawthorne, G., Nair, M., Clark, T.G., Zaver, A., Jafri, S., Hasan, R., 2015. Characterization of genomic variations in SNPs of PE_PGRS genes reveals deletions and insertions in extensively drug resistant (XDR) *M. tuberculosis* strains from Pakistan. Int. J. Mycobacteriol. 4 (1), 73–79.
Karboul, A., Gey Van Pittius, N.C., Namouchi, A., Vincent, V., Sola, C., Rastogi, N., Suffys, P., Fabre, M., Cataldi, A., Huard, R.C., Kurepina, N., Kreiswirth, B., Ho, J.L., Gutierrez, M.C., Mardassi, H., 2006. Insights into the evolutionary history of Tubercle bacilli as disclosed by genetic rearrangements within a PE_PGRS duplicated gene pair. BMC Evol. Biol. 6 (1).
Karboul, A., Mazza, A., Gey Van Pittius, N.C., Ho, J.L., Brousseau, R., Mardassi, H., 2008. Frequent homologous recombination events in *M. tuberculosis* PE-PPE multigene families: potential role in antigenic variability. J. Bacteriol. 190 (23), 7838–7846.
Khan, N., Alam, K., Nair, S., Valluri, V.L., Murthy, K.J.R., Mukhopadhyay, S., 2008. Association of strong immune responses to PPE protein Rv1168c with active tuberculosis. Clin. Vaccine Immunol. 15 (6), 974–980.
Khubaib, M., Sheikh, J.A., Pandey, S., Srikanth, B., Bhuwan, M., Khan, N., Hasnain, S.E., Ehtesham, N.Z., 2016. *M. tuberculosis* co-operonic PE32/PPE65 proteins alter host immune responses by hampering Th1 response. Front. Microbiol. 7.
Kim, W.S., Kim, J.S., Cha, S.B., Kim, S.J., Kim, H., Kwon, K.W., Han, S.J., Choi, S.Y., Shin, S.J., 2016. *M. tuberculosis* PE27 activates dendritic cells and contributes to Th1-polarized memory immune responses during in vivo infection. Immunobiology 221 (3), 440–453.
Koh, K.W., Lehming, N., Seah, G.T., 2009a. Degradation-resistant protein domains limit host cell processing and immune detection of *Mycobacteria*. Mol. Immunol. 46 (7), 1312–1318.
Koh, K.W., Soh, S.E., Seah, G.T., 2009b. Strong antibody responses to *M. tuberculosis* PE-PGRS62 protein are associated with latent and active tuberculosis. Infect. Immun. 77 (8), 3337–3343.
Korotkova, N., Freire, D., Phan, T.H., Ummels, R., Creekmore, C.C., Evans, T.J., Wilmanns, M., Bitter, W., Parret, A.H.A., Houben, E.N.G., Korotkov, K.V., 2014. Structure of the *M. tuberculosis* type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer. Mol. Microbiol. 94 (2), 367–382.
Korycka-Machała, M., Pawełczyk, J., Borówka, P., Dziadek, B., Brzostek, A., Kawka, M., Bekier, A., Rykowski, S., Olejniczak, A.B., Strapagiel, D., Witczak, Z., Dziadek, J., 2020. PPE51 is involved in the uptake of disaccharides by *M. tuberculosis*. Cells 9 (3), 603.
Li, H., Li, Q., Yu, Z., Zhou, M., Xie, J., 2016. *M. tuberculosis* PE13 (Rv1195) manipulates the host cell fate via p38-ERK-NF-κB axis and apoptosis. Apoptosis 21 (7), 795–808.
Long, Q., Xiang, X., Yin, Q., Li, S., Yang, W., Sun, H., Liu, Q., Xie, J., Deng, W., 2019. PE_PGRS62 promotes the survival of *Mycobacteria* smegmatis within macrophages via disrupting ER stress-mediated apoptosis. J. Cell. Physiol. 234 (11), 19774–19784.
Lu, B., Rutledge, B.J., Gu, L., Fiorillo, J., Lukacs, N.W., Kunkel, S.L., North, R., Gerard, C., Rollins, B.J., 1998. Abnormalities in monocyte recruitment and cytokine expression in monocyte chemoattractant protein 1-deficient mice. J. Exp. Med. 187 (4), 601–608.
Ma, Y., Keil, V., Sun, J., 2015. Characterization of *M. tuberculosis* EsxA membrane insertion. J. Biol. Chem. 290 (11), 7314–7322.
Mansury, D., Ghazvini, K., Amel Jamehdar, S., Badiee, A., Tafaghodi, M., Nikpoor, A.R., Amini, Y., Jaafari, M.R., 2019. Enhancement of the effect of BCG vaccine against tuberculosis using DDA/TDB liposomes containing a fusion protein of HspX, PPE44, and EsxV. Artif. Cells Nanomed. Biotechnol. 47 (1), 370–377.
McEvoy, C.R.E., Warren, R.M., Van Helden, P.D., Gey Van Pittius, N.C., 2009. Multiple, independent, identical IS6110 insertions in *M. tuberculosis* PPE genes. Tuberculosis 89 (6), 439–442.
McEvoy, C.R.E., Cloete, R., Müller, B., Schürch, A.C., Van Helden, P.D., Gagneux, S., Warren, R.M., Gey Van Pittius, N.C., 2012. Comparative analysis of *M. tuberculosis* PE and PPE genes reveals high sequence variation and an apparent absence of selective constraints. PLoS One 7 (4), 30593.
McGuire, A., Weiner, B., Park, S., Wapinski, I., Raman, S., Dolganov, G., Peterson, M., Riley, R., Zucker, J., Abeel, T., White, J., Sisk, P., Stolte, C., Koehrsen, M., Yamamoto, R.T., Iacobelli Martinez, M., Kidd, M.J., Maer, A.M., Schoolnik, G.K., Regev, A., 2012. Comparative analysis of *Mycobacteria* and related actinomycetes yields insight into the evolution of *M. tuberculosis* pathogenesis. BMC Genomics 13 (1), 120.
Meehan, C.J., Goig, G.A., Kohl, T.A., Verboven, L., Dippenaar, A., Ezewudo, M., Farhat, M.R., Guthrie, J.L., Laukens, K., Miotto, P., Ofori Anyinam, B., Dreyer, V., Supply, P., Suresh, A., Utpatel, C., Van Soolingen, D., Zhou, Y., Ashton, P.M.,

Brites, D., Cabibbe, A.M., 2019. Whole Genome Sequencing of *M. tuberculosis*: current standards and open issues. Nat. Rev. Microbiol. 17 (9), 533–545.
Meena, L.S., Meena, J., 2015. Cloning and characterization of a novel PE_PGRS60 protein (Rv3652) of *M. tuberculosis* H37Rv exhibit fibronectin-binding property. Biotechnol. Appl. Biochem. 63 (4), 525–531.
Melo, E., Souza, I., Ramos, C., Osório, A., Verbisck, N., Araújo, F., 2015. Evaluation of the use of recombinant proteins of *Mycobacteria* bovis as antigens in intradermal tests for diagnosis of bovine tuberculosis. Archivos De Medicina Veterinaria 47 (3), 273–280.
Mi, Y., Bao, L., Gu, D., Luo, T., Sun, C., Yang, G., 2017. *M. tuberculosis* PPE25 and PPE26 proteins expressed in *Mycobacteria* smegmatis modulate cytokine secretion in mouse macrophages and enhance mycobacterial survival. Res. Microbiol. 168 (3), 234–243.
Mitra, A., Speer, A., Lin, K., Ehrt, S., Niederweis, M., 2017. PPE surface proteins are required for heme utilization by *M. tuberculosis*. *mBio* 8 (1).
Mitra, A., Ko, Y.H., Cingolani, G., Niederweis, M., 2019. Heme and hemoglobin utilization by *M. tuberculosis*. Nature Commun. 10 (1).
Mogensen, T.H., 2009. Pathogen recognition and inflammatory signaling in innate immune defenses. Clin. Microbiol. Rev. 22 (2), 240–273.
Mukherjee, P., Dutta, M., Datta, P., Dasgupta, A., Pradhan, R., Pradhan, M., Kundu, M., Basu, J., Chakrabarti, P., 2007. The RD1-encoded antigen Rv3872 of *M. tuberculosis* as a potential candidate for serodiagnosis of tuberculosis. Clin. Microbiol. Infect. 13 (2), 146–152.
Nair, S., Ramaswamy, P.A., Ghosh, S., Joshi, D.C., Pathak, N., Siddiqui, I., Sharma, P., Hasnain, S.E., Mande, S.C., Mukhopadhyay, S., 2009. The PPE18 of *M. tuberculosis* interacts with TLR2 and activates IL-10 induction in macrophage. J. Immunol. 183 (10), 6269–6281.
Nair, S., Pandey, A.D., Mukhopadhyay, S., 2011. The PPE18 protein of *M. tuberculosis* inhibits NF-κB/rel-mediated proinflammatory cytokine production by upregulating and phosphorylating suppressor of cytokine signaling 3 protein. J. Immunol. 186 (9), 5413–5424.
Narayana, Y., Joshi, B., Katoch, V.M., Mishra, K.C., Balaji, K.N., 2007. Differential B-cell responses are induced by *M. tuberculosis* PE antigens Rv1169c, Rv0978c, and Rv1818c. Clin. Vacc. Immunol. 14 (10), 1334–1341.
Newton Foot, M., Warren, R.M., Sampson, S.L., Van Helden, P.D., Gey Van Pittius, N.C., 2016. The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems. BMC Evol. Biol. 16 (1).
Palucci, I., Camassa, S., Cascioferro, A., Sali, M., Anoosheh, S., Zumbo, A., Minerva, M., Iantomasi, R., De Maio, F., Di Sante, G., Ria, F., Sanguinetti, M., Palù, G., Brennan, M.J., Manganelli, R., Delogu, G., 2016. PE_PGRS33 contributes to *M. tuberculosis* Entry in macrophages through interaction with TLR2. PLoS One 11 (3), e0150800.
Pérez-Lago, L., Herranz, M., Martínez Lirola, M., Bouza, E., García De Viedma, D., 2011. Characterization of microevolution events in *M. tuberculosis* strains involved in recent transmission clusters. J. Clin. Microbiol. 49 (11), 3771–3776.
Phan, T.H., Ummels, R., Bitter, W., Houben, E.N.G., 2017. Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. Sci. Rep. 7 (1).
Phelan, J.E., Coll, F., Bergval, I., Anthony, R.M., Warren, R., Sampson, S.L., Gey Van Pittius, N.C., Glynn, J.R., Crampin, A.C., Alves, A., Bessa, T.B., Campino, S., Dheda, K., Grandjean, L., Hasan, R., Hasan, Z., Miranda, A., Moore, D., Panaiotov, S., Perdigao, J., 2016. Recombination in PE-PPE genes contributes to genetic variation in *M. tuberculosis* lineages. BMC Genomics 17 (1).
Poulet, S., Cole, S.T., 1995. Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *M. tuberculosis*. Arch. Microbiol. 163 (2), 87–95.
Poulsen, C., Panjikar, S., Holton, S.J., Wilmanns, M., Song, Y.H., 2014. WXG100 protein superfamily consists of three subfamilies and exhibits an α-helical C-terminal conserved residue pattern. PLoS One 9 (2), 89313.
Pourakbari, B., Mamishi, S., Marjani, M., Rasulinejad, M., Mariotti, S., Mahmoudi, S., 2015. Novel T-cell assays for the discrimination of active and latent tuberculosis infection: the diagnostic value of PPE family. Mol. Diagn. Ther. 19 (5), 309–316.
Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H., Becker-Ziaja, B., Boettcher, J.-P., Cabeza Cabrerizo, M., Camino Sanchez, A., Carter, L.L., Doerrbecker, J., 2016. Real-time, portable genome sequencing for ebola surveillance. Nature 530 (7589), 228–232.
Ramakrishnan, P., Aagesen, A.M., McKinney, J.D., Tischler, A.D., 2015. *M. tuberculosis* resists stress by regulating PE19 expression. Infect. Immun. 84 (3), 735–746.
Reed, J.C., 2000. Mechanisms of apoptosis. Am. J. Pathol. 157 (5), 1415–1430.
Renshaw, P.S., Lightbody, K.L., Veverka, V., Muskett, F.W., Kelly, G., Frenkiel, T.A., Gordon, S.V., Hewinson, R.G., Burke, B., Norman, J., Williamson, R.A., Carr, M.D., 2005. Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. EMBO J. 24 (14), 2491–2498.
Romano, M., Rindi, L., Korf, H., Bonanni, D., Adnet, P.Y., Jurion, F., Garzelli, C., Huygen, K., 2008. Immunogenicity and protective efficacy of tuberculosis subunit vaccines expressing PPE44 (Rv2770c). Vaccine 26 (48), 6053–6063.
Saini, N.K., Baena, A., Ng, T.W., Venkataswamy, M.M., Kennedy, S.C., Kunnath Velayudhan, S., Carreño, L.J., Xu, J., Chan, J., Larsen, M.H., Jacobs, W.R., Porcelli, S. A., 2016. Suppression of autophagy and antigen presentation by *M. tuberculosis* PE_PGRS47. Nature Microbiol. 1 (9).
Sali, M., Di Sante, G., Cascioferro, A., Zumbo, A., Nicolò, C., Donà, V., Rocca, S., Procoli, A., Morandi, M., Ria, F., Palù, G., Fadda, G., Manganelli, R., Delogu, G., 2010. Surface expression of MPT64 as a fusion with the PE domain of PE_PGRS33 enhances *Mycobacteria* bovis BCG protective activity against *M. tuberculosis* in mice. Infect. Immun. 78 (12), 5202–5213.
Sampson, S.L., 2011. Mycobacterial PE-PPE proteins at the host-pathogen interface. Clin. Dev. Immunol. 2011, 1–11.
Sani, M., Houben, E.N.G., Geurtsen, J., Pierson, J., De Punder, K., Van Zon, M., Wever, B., Piersma, S.R., Jiménez, C.R., Daffé, M., Appelmelk, B.J., Bitter, W., Van Der Wel, N., Peters, P.J., 2010. Direct visualization by cryo-EM of the mycobacterial capsular layer: A labile structure containing ESX-1-secreted proteins. PLoS Pathog. 6 (3), 1000794.
Sargentini, V., Mariotti, S., Carrara, S., Gagliardi, M.C., Teloni, R., Goletti, D., Nisini, R., 2009. Cytometric detection of antigen-specific IFN-γ/IL-2 secreting cells in the diagnosis of tuberculosis. BMC Infect. Dis. 9 (1).
Sayes, F., Sun, L., Di Luca, M., Simeone, R., Degaiffier, N., Fiette, L., Esin, S., Brosch, R., Bottai, D., Leclerc, C., Majlessi, L., 2012. Strong immunogenicity and cross-reactivity of *M. tuberculosis* ESX-5 type VII secretion-encoded PE-PPE proteins predicts vaccine potential. Cell Host Microbe 11 (4), 352–363.
Sayes, F., Pawlik, A., Frigui, W., Gröschel, M.I., Crommelynck, S., Fayolle, C., Cia, F., Bancroft, G.J., Bottai, D., Leclerc, C., Brosch, R., Majlessi, L., 2016. CD4+ T cells recognizing PE-PPE antigens directly or via cross reactivity are protective against pulmonary *M. tuberculosis* infection. PLoS Pathog. 12 (7), 1005770.
Sharma, T., Grover, S., Arora, N., Manjunath, P., Ehtesham, N.Z., Hasnain, S.E., 2020. PGRS domain of Rv0297 of *M. tuberculosis* is involved in modulation of macrophage functions to favor bacterial persistence. Front. Cell. Infect. Microbiol. 10.
Sharma, S., Meena, L.S., 2016. Potential of Ca2+ in *M. tuberculosis* H37Rv pathogenesis and survival. Appl. Biochem. Biotechnol. 181 (2), 762–771.
Sharma, N., Shariq, M., Quadir, N., Singh, J., Sheikh, J.A., Hasnain, S.E., Ehtesham, N.Z., 2021a. *M. tuberculosis* protein PE6 (Rv0335c), a novel TLR4 agonist, evokes an inflammatory response and modulates the cell death pathways in macrophages to enhance intracellular survival. Front. Immunol. 12.
Sharma, T., Singh, J., Grover, S., Manjunath, P., Firdos, F., Alam, A., Ehtesham, N.Z., Hasnain, S.E., 2021b. PGRS domain of Rv0297 of *M. tuberculosis* functions in a calcium dependent manner. Int. J. Mol. Sci. 22 (17), 9390.
Siddiqi, K., Lambert, M.L., Walley, J., 2003. Clinical diagnosis of smear-negative pulmonary tuberculosis in low-income countries: the current evidence. Lancet Infect. Dis. 3 (5), 288–296.
Siegrist, M.S., Unnikrishnan, M., McConnell, M.J., Borowsky, M., Cheng, T.Y., Siddiqi, N., Fortune, S.M., Moody, D.B., Rubin, E.J., 2009. Mycobacterial ESX-3 is required for mycobactin-mediated iron acquisition. Proc. Natl. Acad. Sci. 106 (44), 18792–18797.
Singh, Y., Kohli, S., Sowpati, D.T., Rahman, S.A., Tyagi, A.K., Hasnain, S.E., 2014. Gene co-option in *Mycobacteria* and search for virulence attributes: comparative proteomic analyses of *M. tuberculosis*, *Mycobacteria* Indicus Pranii and Other *Mycobacteria*. Int. J. Med. Microbiol. 304 (5–6), 742–748.
Singh, P., Rao, R.N., Reddy, J.R.C., Prasad, R., Kotturu, S.K., Ghosh, S., Mukhopadhyay, S., 2016. PE11, a PE-PPE family protein of *M. tuberculosis* is involved in cell wall remodelling and virulence. Sci. Rep. 6 (1).
Singh, S.K., Tripathi, D.K., Singh, P.K., Sharma, S., Srivastava, K.K., 2012. Protective and survival efficacies of Rv0160c protein in murine model of *M. tuberculosis*. Appl. Microbiol. Biotechnol. 97 (13), 5825–5837.
Skeiky, Y.A.W., Alderson, M.R., Ovendale, P.J., Guderian, J.A., Brandt, L., Dillon, D.C., Campos Neto, A., Lobet, Y., Dalemans, W., Orme, I.M., Reed, S.G., 2004. Differential immune responses and protective efficacy induced by components of a tuberculosis polyprotein vaccine, Mtb72F, delivered as naked DNA or recombinant protein. J. Immunol. 172 (12), 7618–7628.
Solomonson, M., Setiaputra, D., Makepeace, K., Lameignere, E., Petrotchenko, E., Conrady, D., Bergeron, J., Vuckovic, M., DiMaio, F., Borchers, C., Yip, C., Strynadka, N., 2015. Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism. Structure 23 (3), 571–583.
Srivastava, S., Battu, M.B., Khan, M.Z., Nandicoori, V.K., Mukhopadhyay, S., 2019. *M. tuberculosis* PPE2 protein interacts with p67phox and inhibits reactive oxygen species production. J. Immunol. 203 (5), 1218–1229.
Srivastava, R., Kumar, D., Waskar, N.M., Sharma, M., Katoch, V.M., Srivastava, B.S., 2006. Identification of a repetitive sequence belonging to a PPE gene of *M. tuberculosis* and its use in diagnosis of tuberculosis. J. Med. Microbiol. 55 (8), 1071–1077.
Stallings, C.L., Glickman, M.S., 2010. Is *M. tuberculosis* stressed out? A critical assessment of the genetic evidence. Microbes Infect. 12 (14–15), 1091–1101.
Stewart, G.R., Patel, J., Robertson, B.D., Rae, A., Young, D.B., 2005. Mycobacterial mutants with defective control of phagosomal acidification. PLoS Pathog. 1 (3), e33.
Strong, E.J., Ng, T.W., Porcelli, S.A., Lee, S., 2021. *M. tuberculosis* PE_PGRS20 and PE_PGRS47 proteins inhibit autophagy by interaction with Rab1A. mSphere 6 (4).
Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., Eisenberg, D., 2006. Toward the structural genomics of complexes: crystal structure of a PE-PPE protein complex from *M. tuberculosis*. Proc. Natl. Acad. Sci. 103 (21), 8060–8065.
Stylianou, E., Harrington Kandt, R., Beglov, J., Bull, N., Pinpathomrat, N., Swarbrick, G. M., Lewinsohn, D.A., Lewinsohn, D.M., McShane, H., 2018. Identification and evaluation of novel protective antigens for the development of a candidate tuberculosis subunit vaccine. Infect. Immun. 86 (7), e00014.
Su, H., Kong, C., Zhu, L., Huang, Q., Luo, L., Wang, H., Xu, Y., 2015. PPE26 induces TLR2-dependent activation of macrophages and drives Th1-Type T-cell immunity by triggering the cross-talk of multiple pathways involved in the host response. Oncotarget 6 (36), 38517–38537.
Su, H., Zhang, Z., Liu, Z., Peng, B., Kong, C., Wang, H., Zhang, Z., Xu, Y., 2018. *M. tuberculosis* PPE60 antigen drives Th1/Th17 responses via toll-like receptor 2-dependent maturation of dendritic cells. J. Biol. Chem. 293 (26), 10287–10302.
Sultana, R., Vemula, M.H., Banerjee, S., Guruprasad, L., 2013. The PE16 (Rv1430) of *M. tuberculosis* is an esterase belonging to serine hydrolase superfamily of proteins. PLoS One 8 (2), e55320.
Tait, D.R., Hatherill, M., Van Der Meeren, O., Ginsberg, A.M., Van Brakel, E., Salaun, B., Scriba, T.J., Akite, E.J., Ayles, H.M., Bollaerts, A., Demoitié, M.-A., Diacon, A.,

Evans, T.G., Gillard, P., Hellström, E., Innes, J.C., Lempicki, M., Malahleha, M., Martinson, N., Mesia Vela, D., 2019. Final analysis of a trial of M72/AS01E vaccine to prevent tuberculosis. N. Engl. J. Med.

Tantivitayakul, P., Ruangchai, W., Juthayothin, T., Smittipat, N., Disratthakit, A., Mahasirimongkol, S., Viratyosin, W., Tokunaga, K., Palittapongarnpim, P., 2020. Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in *M. tuberculosis*. Sci. Rep. 10 (1).

Thi, E.P., Hong, C.J.H., Sanghera, G., Reiner, N.E., 2012. Identification of the *M. tuberculosis* protein PE-PGRS62 as a novel effector that functions to block phagosome maturation and inhibit iNOS expression. Cell. Microbiol. 15 (5), 795–808.

Tiwari, B.M., Kannan, N., Vemu, L., Raghunand, T.R., 2012. The *M. tuberculosis* PE proteins Rv0285 and Rv1386 modulate innate immunity and mediate bacillary survival in macrophages. PLoS One 7 (12), 51686.

Tiwari, B., Soory, A., Raghunand, T.R., 2014. An immunomodulatory Role for the *M. tuberculosis* region of difference 1 locus proteins PE35 (Rv3872) and PPE68 (Rv3873). FEBS J. 281 (6), 1556–1570.

Tiwari, B., Ramakrishnan, U.M., Raghunand, T.R., 2015. The *M. tuberculosis* protein pair PE9 (Rv1088) - PE10 (Rv1089) forms heterodimers and induces macrophage apoptosis through toll-like receptor 4. Cell. Microbiol. 17 (11), 1653–1669.

Tufariello, J.M., Chapman, J.R., Kerantzas, C.A., Wong, K.-W., Vilchèze, C., Jones, C.M., Cole, L.E., Tinaztepe, E., Thompson, V., Fenyö, D., Niederweis, M., Ueberheide, B., Philips, J.A., Jacobs, W.R., 2016. Separable Roles for *M. tuberculosis* ESX-3 Effectors in Iron Acquisition and Virulence. Proc. Natl. Acad. Sci. 113 (3), 348–357.

Tullius, M.V., Nava, S., Horwitz, M.A. 2019. PPE37 is essential for *M. tuberculosis* heme-iron acquisition (HIA), and a defective PPE37 in *Mycobacteria* bovis BCG prevents HIA. *Infect. Immun.*, 87(2).

Tundup, S., Akhter, Y., Thiagarajan, D., Hasnain, S.E. 2006. Clusters of PE and PPE genes of *M. tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. *FEBS Lett.*, 580(5), 1285–1293.

Tundup, S., Pathak, N., Ramanadham, M., Mukhopadhyay, S., Murthy, K.J.R., Ehtesham, N.Z., Hasnain, S.E., 2008. The co-operonic PE25/PPE41 protein complex of *M. tuberculosis* elicits increased humoral and cell mediated immune response. PLoS One 3 (10), 3586.

Tundup, S., Mohareer, K., Hasnain, S.E., 2014. *M. tuberculosis* PE25/PPE41 protein complex induces necrosis in macrophages: role in virulence and disease reactivation? FEBS Open Bio 4 (1), 822–828.

Van Der Meeren, O., Hatherill, M., Nduba, V., Wilkinson, R.J., Muyoyeta, M., Van Brakel, E., Ayles, H.M., Henostroza, G., Thienemann, F., Scriba, T.J., Diacon, A., Blatner, G.L., Demoitié, M.-A., Tameris, M., Malahleha, M., Innes, J.C., Hellström, E., Martinson, N., Singh, T., Akite, E.J., 2018. Phase 2b controlled trial of M72/AS01E vaccine to prevent tuberculosis. N. Engl. J. Med. 379 (17), 1621–1634.

Veyrier, F.J., Dufort, A., Behr, M.A., 2011. The rise and fall of the *M. tuberculosis* genome. Trends Microbiol. 19 (4), 156–161.

Vijay, K., 2018. Toll-like receptors in immunity and inflammatory diseases: Past, present, and future. Int. Immunopharmacol. 59, 391–412.

Vordermeier, H.M., Hewinson, R.G., Wilkinson, R.J., Wilkinson, K.A., Gideon, H.P., Young, D.B., Sampson, S.L., 2012. Conserved immune recognition hierarchy of mycobacterial PE-PPE proteins during infection in natural hosts. PLoS One 7 (8), 40890.

Wang, Q., Boshoff, H.I.M., Harrison, J.R., Ray, P.C., Green, S.R., Wyatt, P.G., Barry, C.E., 2020. PE-PPE proteins mediate nutrient transport across the outer membrane of *M. tuberculosis*. Science 367 (6482), 1147–1151.

Weerdenburg, E.M., Abdallah, A.M., Mitra, S., De Punder, K., Van Der Wel, N.N., Bird, S., Appelmelk, B.J., Bitter, W., Van Der Sar, A.M., 2012. ESX-5-deficient *Mycobacteria* marinum is hypervirulent in adult zebrafish. Cell. Microbiol. 14 (5), 728–739.

Williamson, Z.A., Chaton, C.T., Ciocca, W.A., Korotkova, N., Korotkov, K.V., 2020. PE5-PPE4-EspG3 heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems. J. Biol. Chem. 295 (36).

World Health Organisation, 2015. The End TB Strategy. World Health Organisation.

World Health Organization, 2022. Global Tuberculosis Report 2022. World Health Organization.

Xu, Y., Yang, E., Huang, Q., Ni, W., Kong, C., Liu, G., Li, G., Su, H., Wang, H., 2015. PPE57 induces activation of macrophages and drives Th1-type immune responses through TLR2. J. Mol. Med. 93 (6), 645–662.

Yang, Y., Xu, P., He, P., Shi, F., Tang, Y., Guan, C., Zeng, H., Zhou, Y., Song, Q., Zhou, B., Jiang, S., Shao, C., Sun, J., Yang, Y., Wang, X., Song, H., 2020. Mycobacterial PPE13 activates inflammasome by interacting with the NATCH and LRR domains of NLRP3. FASEB J. 34 (9), 12820–12833.

Yeruva, V.C., Kulkarni, A., Khandelwal, R., Sharma, Y., Raghunand, T.R., 2016. The PE_PGRS proteins of *M. tuberculosis* Are Ca2+ binding mediators of host-pathogen interaction. Biochemistry 55 (33), 4675–4687.

Yu, Z., Zhang, C., Zhou, M., Li, Q., Li, H., Duan, W., Li, X., Feng, Y., Xie, J., 2017. *M. tuberculosis* PPE44 (Rv2770c) is involved in response to multiple stresses and promotes the macrophage expression of IL-12 p40 and IL-6 via the p38, ERK, and NF-κB signalling axis. Int. Immunopharmacol. 50, 319–329.

Zhang, H., Wang, J., Lei, J., Zhang, M., Yang, Y., Chen, Y., Wang, H., 2007. PPE protein (Rv3425) from DNA segment RD11 of *M. tuberculosis*: a potential B-cell antigen used for serological diagnosis to distinguish vaccinated controls from tuberculosis patients. Clin. Microbiol. Infect. 13 (2), 139–145.