Microbial Pathogenesis 49 (2010) 311–314

Contents lists available at ScienceDirect

Microbial Pathogenesis

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

Review

Roles of PE_PGRS family in *Mycobacterium tuberculosis* pathogenesis and novel measures against tuberculosis

Chen Tian, Xie Jian-ping*

Institute of Modern Biopharmaceuticals, Key Laboratory of Ministry of Education Eco-Environment of the Three Gorges Reservoir Region, School of Life Sciences, Southwest University, Chongqing 400715, China

ARTICLE INFO

Article history:
Received 18 April 2010
Received in revised form
6 July 2010
Accepted 10 July 2010
Available online 16 July 2010

Keywords:
Mycobacterium tuberculosis
PE_PGRS
Reactivation
Persistence
Antigenic variation

ABSTRACT

Tuberculosis remains a serious threat to global public health. As one of the most successful pathogens, *Mycobacterium tuberculosis* is an adept in evasion the host immune response. *M. tuberculosis* PE_PGRS family has been widely proposed as molecular mantra to deflect host immunity. The feature of this family is the conserved N-terminal and variable C-terminal. PE_PGRS proteins, with functions largely unknown, are only found among mycobacteria and located in the mycobacterial cell wall and cell membrane. Most PE_PGRS proteins have antigenicity. Polymorphic PGRS domain might play a role in neutralize immune response. Several members of PE_PGRS family have been heterologously expressed for further function characterization. PE_PGRS gene expressing stage-specifically, especially during *M. tuberculosis* persistence, is promising intervention target to prevent reactivation. The origin, physiological role and spatiotemporal regulation feature of this family remain elusive. The understanding of these questions will shed light on the pathogenesis of *M. tuberculosis* and facilitate the discovery of new effective antibacterial intervention.

Tuberculosis (TB) is an infectious disease with global significance [1]. About one third of the world's population has been latently infected with *Mycobacterium tuberculosis*. Each year three million people die of active TB. *M. tuberculosis* is one of the most successful intracellular pathogens, which can resist a variety of antimicrobial mechanisms of monocytes and macrophages [2]. The sequencing of the *M. tuberculosis* genome reveals a unique multigenic PE_PGRS gene family encoding proteins with two major domains: a highly conserved PE domain linked to a variable PGRS domain. The name of PE is derived from N-terminal Pro(P)-Glu(E) sequence, and the name of PGRS means polymorphic GC-rich repetitive sequence. This family members have been found so far only in the genomes of mycobacteria (Fig. 1) and restricted largely to pathogenic mycobacteria [3], *M. tuberculosis*, *Mycobacterium bovis*, *Mycobacterium ulcerans*, *Mycobacterium marinum* and *Mycobacterium kansasii*. Therefore, it is tempting to speculate this family plays an important role in the pathogenesis and persistence of mycobacteria. Some genes of this family, such as Rv1759c, are proposed as candidate for subunit vaccine. Further studies on this family will benefit the understanding of TB pathogenesis and novel prevention and therapy.

1. The structure of PE_PGRS protein

The *M. tuberculosis* PE family accounts for about 4% of whole-genome content, with a total of 99 members [4]. Based on whether C-terminal of PE-encoding proteins bearing Gly-Gly-Ala or Gly-Gly-Asn multiple tandem repeat structure, it is divided into two subfamilies: PE_ and PE_PGRS, with 38 and 61 members, respectively. PE domain is highly conserved, containing 110 amino acid residues. C-terminal fragment is a PGRS domain encoded by polymorphic G + C-rich sequence, which varies in size, sequence and repeat copy numbers [5]. Glycine and alanine are frequently found in GGAGGX repeat sequence, where X represents any amino acid. Highly polymorphic PGRS domain might contribute to antigenic variation and immune evasion.

PGRS domain of PE_PGRS protein contains glycine-rich motif GGXGXD/NXUX, where X represents any amino acid and U denotes C-terminal nonpolar/large hydrophobic residues. The sequence repeat forms a Ca^{2+}-binding structure called a parallel β-roll or parallel β-helix structure, with typical features of calcium-binding protein. Fifty-six members of PE_PGRS protein family bear this motif except Rv0742, Rv0832, Rv0978c, Rv3652, and Rv3812. Rv0832 (137aa) may be frame-shifted in *M. tuberculosis* H37Rv to be fused with a 749aa encoded by Rv0833; Rv0978c (331aa) contains an unusually short (78aa) PGRS domain; while Rv3652(104aa) may also be a frame-shifted PE_PGRS protein. The maximum repeats are

* Corresponding author. Tel.: +86 23 68367108; fax: +86 23 68252365.
E-mail address: jianpingxiefudan@gmail.com (X. Jian-ping).

0882-4010/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.micpath.2010.07.004

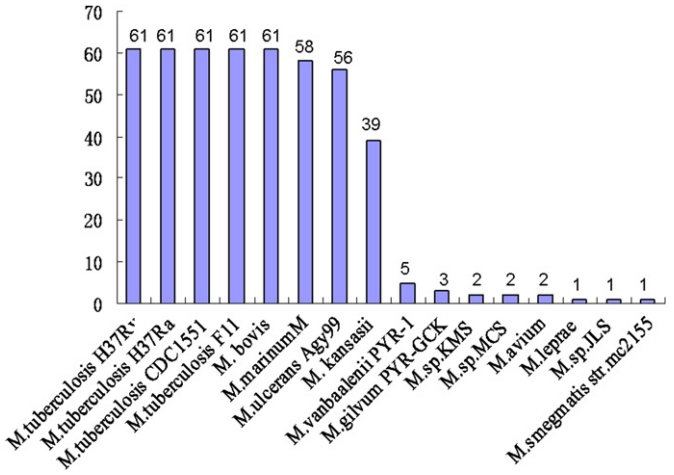

Fig. 1. Distribution of PE_PGRS family.

found in Rv3345c, with 77 copies of this calcium-binding parallel β-helix or β-roll motif GGXGXD/NXUX [6]. PE_PGRS proteins mediating Ca^{2+}-dependent interactions between *M. tuberculosis* and host cells are crucial for *M. tuberculosis* survival and pathogenesis: The initial non-specific adhesion between host cells and *M. tuberculosis* via PE_PGRS proteins ensued a dramatic decrease of macrophage intracellular Ca^{2+} concentration, thereby blocking phagolysosome maturation to the advantage of *M. tuberculosis* persistence.

2. The subcellular localization and expression of PE_PGRS

The subcellular localization of PE_PGRS protein is vital for their function. Many PE_PGRS proteins locate at *Mycobacterium* cell wall and membrane. This implicates that they might mediate the interaction between the bacteria and macrophages [7–9]. The hydrophobic stretches forming transmembrane α-helix is only predicted in Rv0151 and Rv1430 [7]. The very fact that none of proteins contains N-terminal signal peptide typical of Gram-positive bacteria indicates that special secretion mechanisms may underlie PE_PGRS family transfer. ESX-5 secretion system has been experimentally proved to be the most likely candidate for the transport of PE_PGRS proteins [10]. ESX (early secreted antigen 6 kDa [Esat-6] secretion system) secretion system is a novel type of protein secretion systems [11–15], whose substrates include T-cell antigens ESAT-6 and CFP-10 [11,16–19]. This secretion system consists of five members, namely ESX-1(Rv3871-Rv3879c), ESX-2 (Rv3884c-Rv3895c), ESX-3(Rv0282-0292), ESX-4(Rv3444c-Rv3450c) and ESX-5(Rv1782-Rv1798)[20]. ESX-1-encoding genes locate within the chromosomal region of difference 1(RD1) [21], whose deletion would account for the attenuation of the BCG vaccine strain. This suggests that ESX-1 might be a major virulence determinant of pathogenic mycobacteria [22]. The distribution of ESX-5 system limits to the slow-growing mycobacteria, especially most pathogenic species [15,23]. The established function of ESX-5 suggests that this system is largely dedicated to the transport and secretion of PE_PGRS proteins.

Some PE_PGRS family genes are involved in the *M. tuberculosis* persistence. The expression pattern of PE_PGRS family varies greatly with each member and conditions. Studies on the expression of approximately one-third of the PE_PGRS genes under various *in vitro* conditions such as hypoxia, nutrient depletion, low pH and infection of primary macrophages or mice have clearly manifested this diversity [24]. 7 genes are constitutively expressed, including pe_pgrs14, pe_pgrs24, pe_pgrs30, pe_pgrs33, pe_pgrs34, pe_pgrs35 and pe_pgrs45. The expression of another 7 genes is inducible, including pe_pgrs1, pe_pgrs16, pe_pgrs18, pe_pgrs26, pe_pgrs44, pe_pgrs51 and pe_pgrs55. No expression of pe_pgrs27 and pe_pgrs50 can be detected under tested circumstances. Our BLASTing found above detectable 14 genes exist only in pathogenic Mycobacteria. Expression of pe_pgrs44 and pe_pgrs51 is significant only in log-phase, nutrient-rich culture, and hardly detectable under other conditions. Expression of pe_pgrs1, pe_pgrs26, and pe_pgrs55 is decreased within macrophage, while that of pe_pgrs16 is significantly increased. Most genes of PE_PGRS family are expressed under all examined conditions, whereas expression of some genes is just detectable under limited environments. Some unknown regulatory strategy must underlie the differential expression of this family.

3. The function of PE_PGRS protein

3.1. PE_PGRS proteins alter the cellular structure and the colony morphology

Some cell wall associated PE_PGRS proteins can change the cell structure such as shape and size, as well as colony formation [25]. *Mycobacterium smegmatis* overexpressing Rv1818c is elongated, but replicates normally. PE domain alone can facilitate the colony spread on solid media, thereby forming larger colonies. PE domain also contains signal for subcellular localization. PGRS domain of Rv1818c can change the colony phenotype. Intracellular accumulation of PE_PGRS protein Rv1818c might contribute to the above phenotype. However, the mechanisms underlying this aggregation remain elusive.

3.2. PE_PGRS proteins function as lipase to supply energy for persistent mycobacteria

PE_PGRS gene family contains genes necessary for *M. tuberculosis* multiplication and persistence within macrophages. Inactivation of PE_PGRS62 (Rv3812) and PE_PGRS30 (Rv1651c) would decrease the replication in macrophages and survival in granulomas [26]. PE_PGRS63 (LipY, encoded by Rv3097) is a hormone-sensitive lipase (HSL) bearing this family's hallmark conserved active site motif GDSAG. LipY is about 45 kDa with a pI of 4.5. LipY specifically hydrolyzes long-chain triacylglycerol (TG) to provide energy for *M. tuberculosis* survival and persistence [27].

3.3. PE_PGRS proteins are major source of antigenic variation

Some PE_PGRS proteins are well-established antigens, and actively engaged in the interaction with host. The same antibody can cross-react with a variety of PE_PGRS antigens, indicating there are shared epitopes among diverse PE_PGRS proteins which might lie in their conserved sequence [9]. Richness of GC arising from the tandem copies of the 9 bp CGGCGGCAA repeat might be the hot spot for recombination, strand slippage during replication, thereby the source of the variation of PE_PGRS [7]. Two mechanisms underlie antigenic variation: the differential regulation and the mutation of structural genes encoding the protein [28]. The resultant sequences may be a blend of point mutation, insertion, deletion and frameshift mutations.

Another tempting but controversial source of antigen variation involving PE_PGRS might be the DNA polymerase. *M. tuberculosis* contains two Y-family DNA polymerase homologs belonging to the DinB subfamily, DinB1 (Rv1537) and DinB2 (Rv3056) [29], which have been documented participation in the gene variation, especially those genes encoding surface antigen, probably including