

PE/PPE Protein Families in Mycobacteriophages

July 28, 2020

Prepared by: Vaishnavi Mahalingam

The PE/PPE protein family is a set of proteins that is associated with *Mycobacterium tuberculosis* genomes. Their precise functions are not known, but they are associated with the bacterial cell wall and exposed to the extracellular environment. The PE protein family is characterized by a 110-amino acid N-terminal domain with a proline (P) residue at position 9 and a glutamic acid (E) residue at position 10. The PPE protein family is characterized by a 180-amino acid N-terminal domain with two proline (P) residues at positions 9 and 10, and a glutamic acid residue (E) at position 11.

These proteins can be identified using HHPred, looking for hits to a serine alpha-beta hydrolase. The protein will also align to a conserved triad of lipases, cutinases, and esterases, perhaps as separate hits but also as part of one hit.

Additionally, these proteins can be identified by looking for conserved motifs in the residue sequence. At the N-terminus, there is a pentapeptide sequence motif of GxSxG/S (Figure 1) as well as a conserved serine, aspartic acid, and histamine, or SDH (Figure 1).

For further information regarding the PE/PPE protein family, please refer to the following paper by Sultana et al.: "The PE-PPE Domain in Mycobacterium Reveals a Serine α/β Hydrolase Fold and Function: An *In-Silico* Analysis."

```
>KayaCho gp88
MKTLLTTAACAATVAALVTPPTALAAAKTDVVFVIGGTGTGMKLTVEVFGTGGSPGLTDPFVTDTRDQLEVPVYDYGAPWANPQASVPQVAA
VVAQTDDPAVVI GLSKGAQVTHGVEAQDTRTDTRYVVIGDPD SDHGISRAFGLSPAQTVKTHDWEEIVAERDAVGDFPDRPWNLATA
ESLASWAVVHPQYGNNGADDPLTRLDEAEVTVTKNPNGTFTTKRVI PVRELSILKPMRDTERTLTGNDRLTDEVERHLKPAVDAGWSR
NDRKPTTKTDADDKPAKDATTKETKADRAEARDTKKESKTDD
```

Figure 1. The conserved GxSxG/S and SDH motifs are highlighted in the amino acid sequence of gene 88 of B6 phage KayaCho.