Key amino acids in understanding evolutionary characterization of Mn/Fe-Superoxide Dismutase: A phylogenetic and structural analysis of proteins from *Corynebacterium* and hosts

Alberto Oliveira[1], Pammella Teixeira[1], Debmalya Barh[2], Preetam Ghosh[3], Vasco Azevedo[1]

[1]*Universidade Federal de Minas Gerais,* [2]*Institute of Integrative Omics and Appiled Biotechnology,* [3] *Virginia Commonwealth University*

Species from genus *Corynebacterium* can survive in a hostile environment, for example, inside a bacterial phagosome within immune system cells, such as macrophages, probably due to the production of superoxide dismutase. Recently, some studies showed this enzyme could protect bacterial cells against ROS produced by biochemical mechanisms. *In addition*, there are indications that in some pathogens, including some species of *Corynebacterium,* Mn/Fe-SOD may have an additional function in infection and colonization of the host. Here, we intend to conduct the coevolution analysis of amino acids from Mn/Fe-Superoxide Dismutase of *Corynebacterium* and its host, with the aim to understand the conservation and correlation among the enzymes from these organisms and consequently understand the evolutionary stages of this protein. A multiple sequence alignment of the SOD protein family (Pfam code: PF00081) was conducted by from the Pfam database and then subjected to three filtering procedures. The residue-position pairs were considered to be correlated if they passed the following thresholds: the correlation score absolute value was higher than 10 (i.e., the p-value associated with the shift in frequency is lower than 10^{-10}). The maximum likelihood method from PhyML software and multiple sequence alignment from ClustalX were used in order to understand about the phylogenetic diversion between the sequences. The comparative molecular modeling of proteins from *C. pseudotuberculosis* was performed with the software Modeller. The Validations of models were done by the PROCHECK tool that use the Ramachandran plot, the Discrete Optimized Protein Energy (DOPE) score, a statistical potential able to provide an energetic validation and by the RMSD. The tool DoGSiteScorer was used to find pockets and sub-pockets in the protein structure. Five amino acid sets were found, wherein seven residues were present in higher than 80% frequencies namely, Thr^{24}, Asn^{69}, Pro^{149}, Gly^{72}, Gly^{73}, $Met^{25,}$ and Gln^{146}. Two pockets were identified on the protein structure near the active site, which contain some residues observed in the sets of correlated residues. Pocket 1 is composed of Phe^{63}, Asn^{149}, Gln^{146}, Gly^{72}, Asn^{76} and Val^{128}. Pocket 2 is composed of Ile^{24}, Met^{25} and Trp^{81}. Analyzing the multiple alignments of Mn-SOD, it was possible to understand some divergences between bacteria and mammals in which were possible identify some key amino acids between bacterial and mammalian sequences. These amino acids were found in four loops and constitute the pocket regions that were found in our results.