# ALGORITHM FOR PROTEIN SEQUENCING FROM MASS SPECTROMETRY DATA

Celso Vítor Alves Queiroz Calomeno[1], Michel Batista[1], Paulo Costa Carvalho [1]

[1.] ICC, FIOCRUZ, CARLOS CHAGAS INSTITUTE, FIOCRUZ, RUA PROF. ALGAGYR MUNHOZ MADER, 3775 - CIC 81350-010, CURITIBA/PR, BRASIL;

<p>Protein sequencing is essential for biotechnology, enabling advances in research and the development of therapies. This project aims to develop a computational tool for the analysis of mass spectrometry data, meeting the demand of the mass spectrometry platform at the Carlos Chagas Institute, Fiocruz Paran&aacute;. We used protein samples, including BSA, a protein provided by a company, and the antibody trastuzumab. The samples were subjected to enzymatic digestions with trypsin, Arg-C, Glu-C, and pepsin. Each digest was analyzed by tandem mass spectrometry using the Orbitrap Fusion Lumos. We developed an algorithm for protein sequencing, integrating peptide spectrum matching (psm) and de novo sequencing data generated by Novor Cloud and Patternlab V. The tool assembles alignments from overlapping sequenced peptides with reference sequences, generating an informative graphical interface. The SequenceAssembler tool allows summarizing psm and de novo sequencing data, optimizing the analysis process. The intuitive graphical interface facilitates the interpretation of results, offering the ability to filter relevant sequences for subsequent research steps. The tool\'s effectiveness was evaluated using BSA as a proof of concept, demonstrating its ability to handle proteins by generating their alignments, achieving a coverage percentage, and obtaining a final consensus sequence. Our tool\'s differential is the alignment of proteomic identifications against a multiple alignment of various template proteins. We developed an effective tool for protein sequencing, facilitating the visualization and analysis of alignments, generating relevant information such as the achieved alignment coverage and the generation of the final consensus sequence, thus benefiting the platforms of the Carlos Chagas Institute, Fiocruz, and offering potential use in therapeutic protein and biosimilar development projects. We hope that the tool will promote significant advances in scientific research.</p>

***Agradecimentos:*** FIOCRUZ, CARLOS CHAGAS INSTITUTE, CNPQ, PEM, LPEC