

## Navigating Blindly Through a Proteomics Dataset: Sponge Proteomics Study

Caroline da Silva de Araújo<sup>1</sup>, Fábio César Sousa Nogueira<sup>1</sup>, Alessandra Leda Valverde<sup>2</sup>, Magno Rodrigues Junqueira<sup>1</sup>

<sup>1</sup> UFRJ, Universidade Federal do Rio de Janeiro, Universidade Federal do Rio de Janeiro, Instituto de Química. CT - Centro de Tecnologia, Cidade Universitária 21941909 - Rio de Janeiro, RJ - Brasil ;

<sup>2</sup> UFF, Universidade Federal Fluminense, Universidade Federal Fluminense, Centro de Estudos Gerais. Campus do Valonguinho Centro 24020-150 - Niteroi, RJ - Brasil;

### **Navigating Blindly Through a Proteomics Dataset: Sponge Proteomics Study**

Sponges are sessile marine invertebrates living in symbiosis with microorganisms, with over 8,600 species globally, including around 515 in Brazil. The genus *Topsentia*, known for its cytotoxic activity against cancer cell lines and pathogens, is the focus of this study. Despite their biological significance, genomic and proteomic data for *Topsentia* sponges are scarce, with only 28 entries in the UniProtKB database for the genus. This scarcity complicates proteomic studies, necessitating alternatives like *de novo* sequencing and sequence similarity searches.

A *Topsentia* sponge sample collected in 1998 underwent a quantitative bottom-up metaproteomic analysis. The methodology involved mass spectrometry combined with *de novo* sequencing using Novor.Cloud software and protein quantification and identification with the Perseus program. The sponge sample was lysed with liquid nitrogen and subjected to a Tris HCl + DTT extraction buffer. Extracted proteins, measured at 3.89 µg/µL using a fluorimetric assay, were reduced, alkylated, digested with trypsin, cleaned with Poros R2 and C18 mesh, resuspended with 0.1% formic acid and the peptides were measured, obtaining a concentration of 16.72 µg/µL. Mass spectrometry analysis, performed in duplicate, revealed 15 identified proteins and 21 peptide groups.

Comparison with the UniProt database showed no matches to *Topsentia* proteins; instead, 11 proteins matched *Amphimedon queenslandica*, 3 matched *Suberites domuncula*, and 1 matched *Geodia cydonium*. *De novo* peptide sequencing with Novor.Cloud identified 612 protein groups, including 148 hypothetical proteins and 1 contaminant, resulting in 463 candidate proteins, all aligning with bacteria or fungi rather than sponges.

These findings highlight the challenges and extensive research required in navigating proteomic datasets for marine sponges.

***Agradecimentos:***