BIO_15 - Tool development for analysis of scFv NGS data in databases related to 3D structures: Evolution of antibodies *In silico*

Mathias Coelho Batista[1]*; Cássio Pinheiro Oliveira[1]; Rhonaldo Parente Frota[1]; Andréa Queiroz Maranhão[2]; Marcelo Macedo Brigido[2]; Marcos Roberto Lourenzoni[1].
[1]Fiocruz Ceará;
[2]UnB - Universidade de Brasília.

Introduction: The Protein Engineering and Healthcare Solutions Group (GEPeSS) develops antibodies (Ab) and fragments with potential use in cancer diagnostics and therapies. CAR-T cell is a new cancer therapy and uses the single-chain fragment variable (scFv) to recognize tumor cell markers. The engineering of new scFv is a strategy for obtaining optimized Ab for different diseases (humanization, thermostability, affinity). GEPeSS uses structural bioinformatics combined with directed evolution techniques to optimize new therapeutic scFvs. These tools are used to find phage display selected antibodies, allowing the identification of enriched VH and VL sequences, comparing NGS data before and after selection. The tool also gives additional information on CDRs (Complementary Determining Regions) 3D structure, inferred through the amino acid sequence.

Objective: Develop a tool capable of analyzing NGS data from scFv or Fab libraries selected by Phage Display. The tool must provide the user (client) with the results of the analysis, allowing the feeding of the sequence database (DB) in the GEPeSS server.

Methodology: A graphical interface was developed for the client to make guided entries of the parameters and to display the results of the NGS data analysis. Data processing is performed on the customer's computer. The amino acid sequences are sent to the GEPeSS server. An API was developed to receive data, perform analysis on the server and integrate the sequence data and 3D structures of the VH and VL domains obtained in public databases.

Results: The graphical interface facilitated the work of the user (client). Processing has become faster on the customer's computer, with the inclusion of multiprocessing. The API proved to be able to continue the analysis on the client's machine and to integrate different databases. The structural database already has almost 5,000 scFv structure data, allowing access to CDR frequency and residues information by position with different filters. The relationship between DBs of sequences and structures allows their use to obtain massive 3D structures of scFv.

Conclusion: The tool allows the customer to quickly access information of interest. Sequence and structure DBs are relational. As the tool is used, the DB of VH and VL sequences is increased and becomes robust to be linked to Artificial Intelligence algorithms to generate scFvs. It is being used to find antibodies against different antigens, including SARS-CoV-2, and those used to build CAR-T cells.

Keywords: scFv engineering; Phage Display; Bioinformatics