Journal of Information and Data Management, 2024, 15:1, doi: 10.5753/jidm.2024.2673

Computational Biology Laboratory - Combi-Lab

Karina dos Santos Machado [**Universidade Federal do Rio Grande** | ***karinaecomp@gmail.com***]
Adriano Velasque Werhli [**Universidade Federal do Rio Grande** | ***werhli@gmail.com***]

Centro de Ciências Computacionais, Universidade Federal do Rio Grande - FURG, Rio Grande, RS, Brazil.

Received: 10 May 2022 • **Published:** 16 February 2024

Abstract This article presents the Computational Biology - Combi-Lab research group at the Universidade Federal do Rio Grande (FURG) which started its activities in 2011. The main objective of the group is to bring together researchers and students who are interested in all aspects of Computational Biology. Specifically, the group aims to develop, improve and use sophisticated statistical, computational, and mathematical methods to contribute to the advancement of this research area. This article provides an overview of the Combi-Lab timeline from its founding to the actual days, highlighting various articles and discussing about the future of the group. More importantly, joint projects and collaborators are presented, and their contribution to the development of the Bioinformatics is explained. In conclusion, as we look to the past and face the challenges of the future, we hold fast to our goal of becoming a solid and leading reference in Computational Biology at our university and community, and giving back to the society the maximum that we can.

Keywords: Research Group, Computational Biology, Bioinformatics, History

1 Introduction

The current possibilities of taking measurements and analyzing quantities of interest of living organisms, from cells to highly complex life forms, is generating unprecedented amounts of data. Great scientific developments in food safety, biodiversity, human and animal disease prevention are expected as result of the analysis and interpretation of this deluge of data. Many traditional disciplines such as Mathematics, Statistics, Physics, Biology, Medicine, Computer Science, etc., are collaborating to transform these data into knowledge. This multidisciplinary work is known as Bioinformatics or Computational Biology and the interest in this field has grown steadily in recent decades.

The establishment of the Computational Biology Laboratory (Combi-Lab) was the result of a sequence of events, depicted in Figure 1. In around 2008 significant financial investments were directed to the expansion of the universities in Brazil. As a result of this increase in budget, the undergraduate course of Computer Engineering was the seed for the creation of the *Centro de Ciências Computacionais* (C3) at the *Universidade Federal do Rio Grande - FURG*, the southernmost federal university in Brazil. From 2009 onward, three undergraduate courses are offered in C3: Computer Engineering, Automation Engineering and Information Systems. In 2011, various research groups were created in the computation department (C3) including the research group in Computational Biology (Combi-Lab). Shortly after, in 2012, the Masters course in Computer Engineering was launched having Bioinformatics as one of its research lines.

Our group was created in the Computer Science department alongside other more traditional Computer Science groups, or Computer Science "core" groups. However, this presents a small but important issue. Brazilian funding agencies do not recognize Bioinformatics as a discipline within Computer Science, which means that some of our scientific production can not be evaluated as thoroughly it should be or are not considered relevant in the Computer Science core disciplines. Considering that FURG has a strong scientific background in Oceanology, Biological Sciences, and Medicine, it is only natural that a significant part of our research is done in collaboration with colleagues from these departments. These studies are indeed very valuable and have resulted in a good number of scientific articles, supervision of master and PhD students as well as the registration of a patent. As a result of our multidisciplinary approach and collaborations with colleagues from different departments our group produces high-quality scientific research. Unfortunately, our work is often undervalued because we are inserted in the Computer Science area.

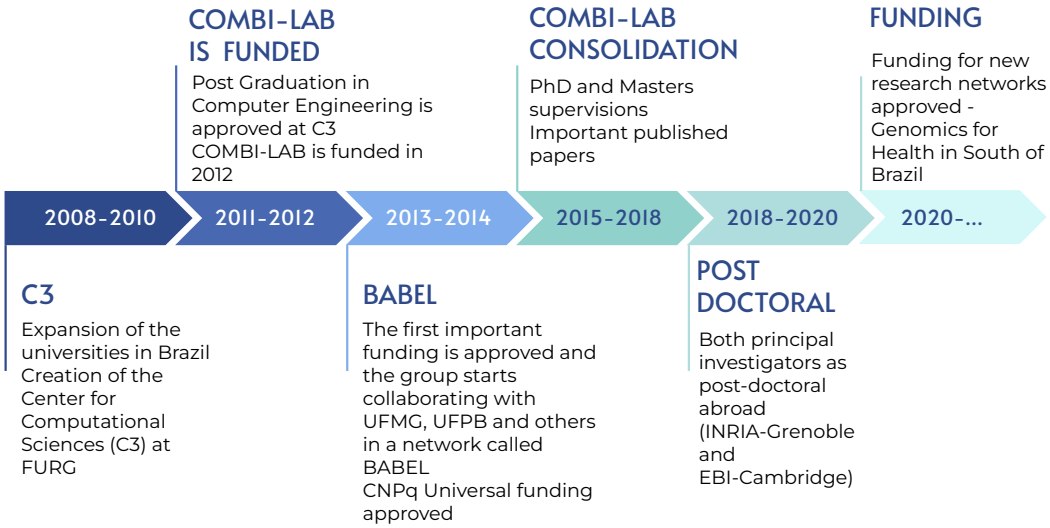
Despite this challenge and having explained how the group was formed, where the group is inserted, and how relevant is the Bioinformatics area, it is important to emphasize that the main interest of our group is to develop, improve and apply the most sophisticated statistical, computational and mathematical methods to advance the research area. In the following sections of this article, we present the current studies underway in our group and we outline the open questions we aim to address in the coming years.

The article is organized as follows: In section 2 the group structure and the profile of current members is presented. In section 3 the activities of the group are summarized and in section 4 a vision for the future of the group is considered.

2 Group Profile

The group is formed by two principal investigators, collaborations, undergraduate, master and PhD students. A brief description of their profiles is presented below

Figure 1. Timeline showing the important events that lead to the establishment of the Computational Biology Laboratory.



- **Principal investigators:**
  - *Adriano V. Werhli* obtained his undergraduate degree in Physics and an MSc in Applied Computing from Universidade do Vale do Rio Dos Sinos, Unisinos. In 2007 he obtained a PhD in Informatics from the University of Edinburgh. During the MSc Adriano worked with evolutionary computing in the protein folding problem. In his PhD thesis he investigated various Machine Learning algorithms for unveiling the structure of Biological networks from data. Moreover, new Bayesian Network methods for integration of prior knowledge were proposed. The current interests are Machine Learning and Data Science applied to Genomics and Proteomics.
  - *Karina dos S. Machado* obtained her undergraduate degree in Computer Engineering in Universidade Federal do Rio Grande - FURG in 2005. From 2005 to 2011 she obtained MSc and PhD in Computer Science from Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS. In her MSc she proposed a scientific workflow for automate molecular docking simulations with flexible receptor. During her PhD Thesis she applied many Machine Learning algorithms to select the most promising receptor snapshots from a molecular dynamics simulation to be used as a flexible receptor model in molecular docking Machado [2011]. Since 2010 she is Associate professor at FURG where she supervise undergraduate and graduate students in Health and Computer Science. Her main interests are in data science and structural bioinformatics.
- **Collaborators:** In Combi-Lab many fruitful collaborations were established over the years. Figure 2 shows the collaborations over a map and bellow we list some of the people we have collaborations with. This list is non exhaustive.
  - Pedro Eduardo Almeida da Silva (Faculty of Medicine FAMED-FURG), Coordinator of the research group Núcleo de Pesquisa em Microbiologia Médica (NUPEMM) and of the Diagnostic Area of the Brazilian Tuberculosis Research Network (REDE-TB). Alson from NUPEMM we collaborate with Andrea von Groll (FAMED-FURG);
  - José María Monserrat, Associate Professor at the Institute of Biological Sciences (ICB-FURG) working mainly with Aquatic Toxicology, Nanotoxicology, Oxidative Stress applied to Aquaculture.
  - Luis Fernando Marins (ICB-FURG), coordinator of the research group of Genetic Engineering and Biotechnology;
  - Diana F. Adamatti (Center for Computational Sciences C3-FURG), coordinator of the Laboratório de Simulação Social e Ambiental - LAMSA;
  - Vania Rodrigues de Lima (EQA -FURG), Coordinator of the Grupo de Investigação de Interações Moleculares em Membranas (GIIMM);
  - Mariana Recamonde Mendoza (INF/UFRGS), Adjunct professor at Applied Informatics Institute in UFRGS and Coordinator of the Bioinformatics Center of Hospital de Clinicas (HCPA) in Porto Alegre, RS, Brazil;
  - Raquel C. de Melo Minardi (UFMG), associate professor at Department of Computer Science in UFMG, member of the Brazilian Academy of Sciences (2019-2023) and sub-coordinator of the Graduate Program in Bioinformatics at UFMG. She is also the coordinator of BABEL [1].
  - Other collaborators of BABEL: Gerd Bruno da Rocha (UFPB), Leonardo Lima (UFSJ);
- **Alumni**: as our projects are interdisciplinary we supervise undergraduate, master and PhD students of computers engineering, computer science, pharmacy, phys-

[1] https://babel.dcc.ufmg.br/

iotherapy, physics, mathematics, medicine, biochemistry, etc.

- **International collaborations**: (i) Dr. Sergei Grudinin, CNRS research associate (CR) scientist in the Nano-D CNRS-Inria group at the Inria Grenoble – Rhone-Alpes Research Center and (ii) Dra. Evangelia Petsalaki, group leader at European Bioinformatics Institute.

3 Main research lines

Combi-Lab activities are strongly driven by the background of the two principal investigators and the biological data generation capacity installed at FURG. In Figure 3 we present an overview of the areas and applications in which we have been working.

FURG has a strong scientific record in biological sciences, particularly in the areas associated with Marine Biology and Medicine. During these years of the Combi-lab existence, many collaborative studies have been conducted, mainly in the area of Structural Bioinformatics where Molecular Docking and Molecular Dynamics simulations are applied to corroborate wet bench results or to guide experiments resulting in significant time and cost savings. These fruitful partnerships have led to not only Bioinformatics studies, but also in the fields of in Data Science and Machine Learning. These collaborative efforts have produced a wide range of groundbreaking research findings.

In the following article we present the studies that were carried out in our group. It is important to emphasise that biological topics emerge as biological problems that we work on in many different studies. One of these topics is toxicology. Toxicology researches at FURG have produced many important results in recent years, and Bioinformatics has played a key role in some of these works. Another relevant topic being investigated at FURG is Tuberculosis (TB). This research expands in various levels, ranging from genomes and molecules to disease dynamics and clinical therapeutics.

3.1 Molecular Docking

The pharmaceutical industry has been applying *in-silico* methods in the first stages of the development of new drugs, an efficient and economical process defined as Rational Drug Design Kuntz [1992]; Meng *et al*. [2011]. The *Structure-based drug design* (SBDD) is one of the types of RDD that is based on the 3D structure of the drug candidates and how they interacts with a target protein Batool *et al*. [2019]. Among the relevant computational techniques, molecular docking, virtual screening and, molecular dynamics (MD) are the most commonly used methods in SBDD.

Molecular docking is a technique for predicting the best conformation and orientation of a small molecule (ligand) in the binding site of a target receptor Lybrand [1995]; Meng *et al*. [2011]. Docking simulations sample hundreds of thousands of orientations and conformations of a ligand inside the protein binding site and evaluated them using an estimated free energy of binding (FEB). Molecular docking simulations are not simple and many enthalpic and entropic factors should be take in consideration. While most docking algorithms address ligand flexibility, receptor flexibility remains a computational and algorithmic challenge due to the numerous degrees of freedom involved Alonso *et al*. [2006]; dos Santos Machado and Grudinin [2020].

In docking software, FEB scores are computed by scoring functions. Many scoring functions are available and can be classified into four categories depending on the methods used to predict the protein-ligand binding score: (i) *Physics-based* which uses force fields; (ii) *Empirical* that calculates the score by a sum of individual terms representing important energetic factors; (iii) *Knowledge-based*, that obtains the score from summing pairwise statistical potentials and (iv) *Machine Learning* methods that trains scoring functions using features (from receptors, ligands, and complexes) obtained from known protein-ligand binding experiments Arrua [2020]; Lopes [2021]; Kadukova *et al*. [2021].

In this research line of molecular docking we have worked on both the development of tools/algorithms and the application of this technique to numerous problems. First, we are going to describe our main contributions related to developed tools and algorithms as well as scoring functions. Next, we detailed the main applications for Tuberculosis, Cancer, Nanotoxicology and so on that we have been working in the past years. It is important to mention that many of our developed tools emerges from necessities that we found during collaborations in the usage of docking to different problems.

3.1.1 Tools and algorithms

In this section we are going to briefly describe some tools and algorithms that Combi-Lab have been working in the last few years.

In 2013/2014 aiming at contributing to ligands selection for virtual screening we proposed data warehouse (DW) schemes to integrate molecular descriptors (molecular weight, logP, number of hydrogen bonds, number of rotatable bonds, etc.) from different public ligands databases, VS experimental data and docking results Perazzo *et al*. [2013]. Having this infrastructure to store all these data, in Seus *et al*. [2014] we performed a VS case studies using AutoDock4 Morris *et al*. [2009] considering as target receptor HIV-protease. The DW allowed us to generate appropriate input data mining files for inferring decision trees (DT) models to relate molecular descriptors with FEB values. From these DT models we extracted rules related to good FEB values to effectively select ligands from ZINC database Irwin *et al*. [2012]. The results showed a reduction of ZINC from 25% to 0.21% of its total of ligands according to the applied rule. From the results of this article we conclude that our proposed methodology is able to produce suitable inputs for mining this kind of data, selection promising ligands for the next steps of VS.

In Seus *et al*. [2016] we developed a framework for Virtual Screening (VS) where the user configures a VS experiment in a Web-based platform filling the path in his/her computer of the input receptor and ligand file(s), the path for the output and the size, center, and variation of the binding site (s). Then, the framework generates the appropriate python script

Figure 2. Principal collaborations at Combi-Lab

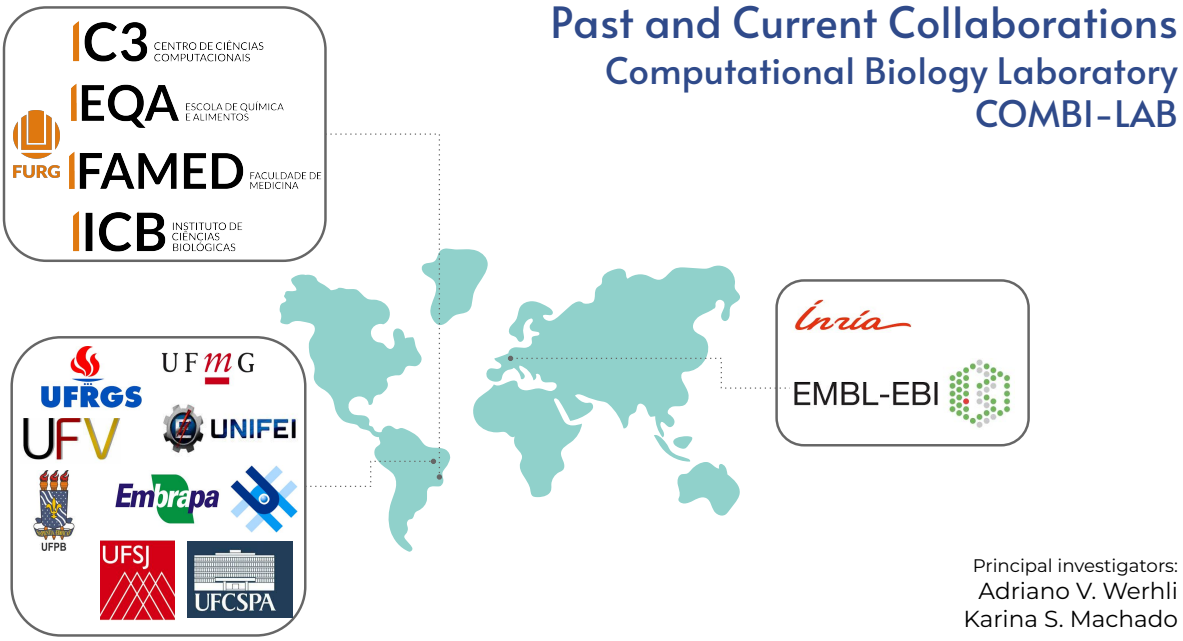
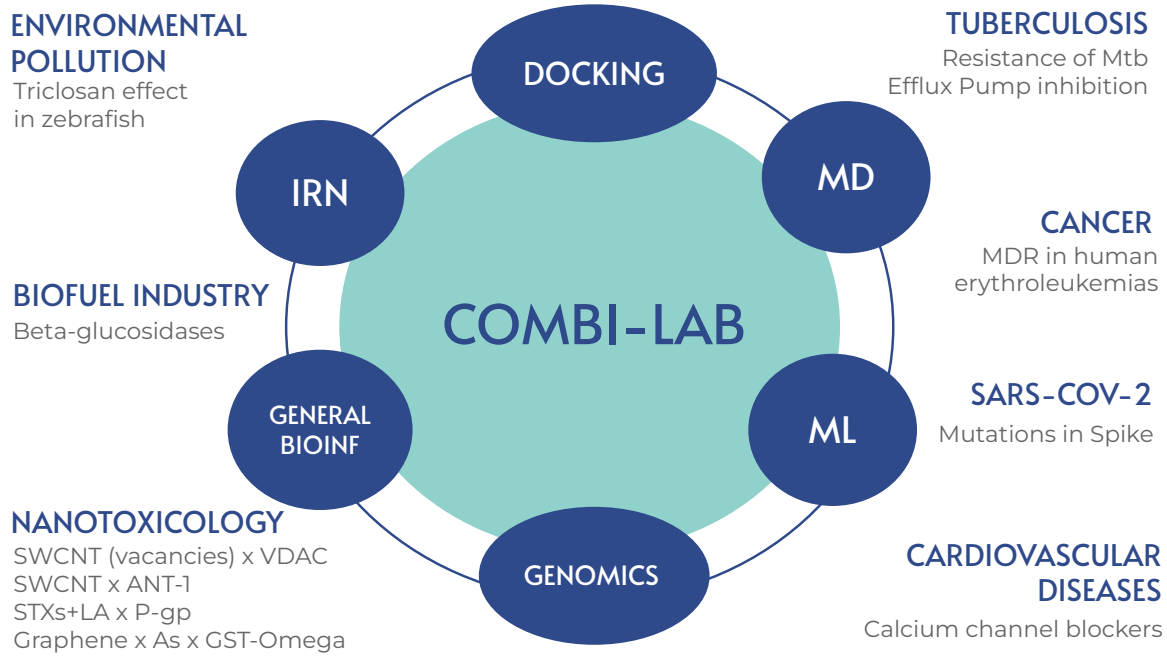


Figure 3. Main bioinformatics techniques and problems studied in Combi-Lab. The main techniques are (not limited to): Molecular docking, MD (Molecular Dynamics Simulations), ML (Machine Learning), Genomics, General Bioinf (General Bioinformatics) and IRN (Inference of Regulatory Networks).



to be performed in the user´s personal computer or cluster using Vina Trott and Olson [2010]. To perform more realistic docking simulations, the user can take into account the receptor flexibility including a set of receptor structures (obtained from PDB or from a molecular dynamics simulation, for example). To validate this framework we performed five case studies with different characteristics: one receptor with n ligands, n receptor structures with one and n ligands, one receptor with one ligand varying the grid box and combining all. At this moment we are developing a new version of this framework more robust, including new functionalities like the possibility to consider more scoring functions, and also for the new version of Vina Eberhardt *et al*. [2021].

In Silva *et al*. [2017] we propose a method to find the best FEB for large proteins using molecular docking. This method was defined as *Slice Docking* and consists in positioning many searching grids along the receptor structure and calculating the FEB for each grid configuration. In order to automate this process, we implemented this functionality in the framework for VS developed in Seus *et al*. [2016].

Since 2018 we have been working on the development of new scoring functions (SF) for molecular docking. Due to our background in ML and also because ML scoring functions are promising and present superior results Shen *et al*. [2020] we chose to work with this type of scoring function in two Master Thesis: Arrua [2020] and Lopes [2021]. In Arrua [2020] we evaluate different combinations of training datasets (PDBbind refined set, PDBbind general set, CSAR-NRC HiQ and, Decoys CSAR-NRC HiQ), molecular descriptors (more than 700 related to the receptors, ligands, and complexes) and ML algorithms (Random Forest and Gaussian Process) to design a score function to predict protein-ligand binding affinities. We also evaluated Lasso as a feature selection algorithm and Hyperparameter Optimization using GridSearch. The performance of the trained models was measured by applying available benchmarks defined in CASF 2016 Su *et al*. [2018]. According to the results, the best combination of attributes, ML algorithm and, parameters presented good scoring, docking, screening and ranking power scores, competitive compared with the others SFs.

In Lopes [2021] we propose a methodology to combine different ML techniques for selecting attributes in an Ensemble approach using the DeltaVinaRF20 scoring function. We also evaluated the proposal using CASF 2016 where the obtained scoring power was ranked as second better compared with more than 20 other SF Su *et al*. [2018]. Although the proposed ML SF presented very promising results, there are some aspects that require further investigation. Thus, in a article not published yet, we evaluate the impact of feature selection techniques such as Variance (ANOVA), Principal Component Analysis (PCA) and (iii) Random Forests in the scoring power of new proposed SF.

During the post-doc of Karina Machado in NANO-D laboratory located in Inria-Grenoble (France) in 2018/2019 Kadukova *et al*. [2021] was proposed. In this article, a novel, minimalist, and fast sidechain-free knowledge-based SF was implemented. This novel coarse-grained potential was defined by a 3D joint probability distribution function that only depends on the pairwise orientation and position between protein backbone and ligand atoms Kadukova *et al*. [2021]. The results showed that KORP obtained excellent docking and screening power for CASF 2013 and CASF 2016 benchmarks and a outstanding results in the DUD-E VS benchmark.

3.1.2 Applications

We have been applying molecular docking simulations for different purposes like nanotoxicology, studies about efflux pump and resistance in Tuberculosis, MDR in human erythroleukemias and, so on.

Nanomaterials have been used for different applications due to their optical, electrical, chemical and, biological properties. In this context, carbon nanotubes (CNTs) have potential application in nanomedicine because of their flexible structures and chemical functionalization Foldvari and Bagonluri [2008]. An important challenge for the safe use of these CNTs is the absence of data about their potential nanotoxicity. One alternative is to use chemoinformatics methods and tools for predicting structure-property relationships for these molecules. In view of this, we started in 2014 a collaboration with Dr. Jose Monserrat for applying molecular docking in nanotoxicology.

The first publication about this subject was Gonzalez-Durruthy *et al*. [2016]. In this article, molecular docking was used to evaluate the interaction between single walled carbon nanotubes (SWCNT) and a mitochondrial ADP/ATP carrier (ANT-1). As ligands, we used different SWCNTs varying the n, m Hamada indexes for armchair, zig-zag and chiral SWCNT types and oxidation. The receptor was ANT-1, a protein that catalyses the electrogenic ADP^{3-} and ATP^{4-} exchange across the inner mitochondrial membrane. In addition to *in-silico* step, experimental measures of oxygen consumption of isolated mitochondria from shrimp *Litopenaeus vannamei* exposed to carbon nanotubes were performed by the collaborators in this article. The *in-silico* results showed that FEB was statistically more negative ($p < 0.05$), following the order SWCNT-COOH $>$ SWCNT-OH $>$ SWCNT, suggesting that polar groups favor the anchorage to ANT-1. In addition, a Perturbation Theory-Nano-Quantitative Structure-Binding Relationship (PT-NQSBR) model was proposed. This model was able to distinguish between strong and weak FEB. It is important to mention that the key amino acids of ANT-1 related to the ADP transport were conserved for different species including shrimp *Litopenaeus vannamei* and fish *Danio rerio* commonly employed in ecotoxicology. Finally, the presented results indicated that the proposed methodology can be used in VS for decision in nanotoxicology.

Also related to nanotoxicology, in González-Durruthy *et al*. [2017] we evaluated the *in-silico* interactions between mitochondrial voltage-dependent anion channel (VDAC) protein of three species (*Mus musculus*, *Homo sapiens* and *Danio rerio*) with SWCNT in order to evaluate their toxicity. VDAC is an essential component of the induced-structure of mitochondrial permeability transition pore, a multi-protein complex involved in apoptosis-based mitochondrial dysfunction. SWCNTs are flexible and versatile molecules due to their physic-chemical properties and can be potentially applied for nanomedicine purposes. In addition

to the docking simulations, this work also generated predictive Nano-Quantitative Structure-Binding Relationship models (NQSBR) relating strong and weak Free Energy of Binding (FEB) obtained from the dockings simulations with geometric and physic-chemical SWCNT nanodescriptors using Perturbation Theory, regression and classification models. This study about VDAC inhibition by SWCNT could be an interesting strategy to induce mitotoxicity based on VDAC-modulation González-Durruthy *et al*. [2017].

This research was continued in Durruthy *et al*. [2019] where we presented an approach for theoretically evaluating and predicting the influence of a broad spectrum of topological vacancies in SWCNTs in the interactions with VDAC protein. It was considered a family of SWCNTs with a defined pattern of topological vacancies (v=1 to 16) obtained by removing atoms from their surface. Such vacancies can significantly affect the geometric and electronic properties of SWCNTs, increasing their reactivity and nanotoxicological potential. It was performed docking simulations and a vacancy quantitative structure-binding relationship (V-QSBR) model was proposed. The results showed a high correlation between the vacancy parameters and FEB values where FEB was more favourable with an increase in the number of SWCNTs vacancies. It demonstrates the influence of topological vacancies in the interaction mechanism that governs the mitochondrial channel nanotoxicity of CNTs.

Continuing in applications to nanotoxicology, in Ramos *et al*. [2018] we worked in a new collaboration with Jose Monserrat research group evaluating the interaction of Saxitoxins (STXs) with lipoic acid (LA) in P-glycoprotein (P-gp) using *ab initio* and molecular docking simulations and bioassays using the cell line HT-22. STXs are neurotoxins that can generate reactive oxygen species and alter antioxidant levels, promoting oxidative stress through the blocking of voltage-gated channels in neurons. The usage of antioxidant LA can be an alternative to minimize the deleterious effects induced by STXs. P-gp is an ATP-binding cassette (ABC) transporter important in the extrusion of toxic substances. As results , the *ab inition* simulations showed that the interactions of STXs with LA occurs by physisorption; molecular docking indicated that STXs can be substrate of P-gp. And finally from experimental analysis it is suggested that LA treatment minimizes STX cytotoxicity.

Continuing the research with STXs, in Ramos *et al*. [2022] we evaluated the effect of açaí as chemoprotection since it has a high concentration of antioxidant molecules. It was considered as target receptor the isoform mu of the glutathione-S-transferase (GST) enzyme of shrimp *L. vannamei*. The docking results indicated that all variants of STXs can be a substrate of GST enzyme (similar final docked positions and FEB) and from experimental results, it was observed that the use of açaí supplements induce antioxidant responses when combined with the chemical stressor STX. So, mu GST isoform could play in the detoxification of STX, opening new research possibilities to improve the chemoprotection of these organisms to the toxin by using supplements that promote the expression of this isoform Ramos *et al*. [2022].

In a different project in applications of molecular docking in nanotoxicology, we studied the interaction between arsenic (As), graphene oxide (GO) with the Glutathione S-transferase (GST)-omega protein Josende *et al*. [2020]. Arsenic is one of the most widespread contaminants, naturally occurring metalloids in the environment. Even its toxic effects have been studied, the interaction of As with other contaminants is relatively unknown, mainly when this interaction occurs with contaminants like nanomaterials. Graphene oxide (GO) is one graphene derivative, a carbon nanomaterial with a promising future. The target receptor considered in this study was GST-omega, an enzyme involved in oxidation and reduction reactions. The main contribution of this article is demonstrate that GO modulated the Arsenic toxic effect decreasing GST-Omega activity. The docking results showed the capacity of GO to interact with GST-Omega.

Our group has established a very fruitful collaboration with the research group NUPEMM coordinated by Pedro Almeida that studies Mycobacteria and more specifically *Mycobacterium tuberculosis*. Tuberculosis (TB), the disease caused by *M. tuberculosis*, is a major cause of mortality and morbidity in developing countries. This subject is present in Combi-Lab before its foundation since TB was the main application of the methodologies developed during the PhD of Karina Machado Machado [2011].

Multidrug and extensive drug resistance is an emerging health public problem and an important issue in the control of Tuberculosis. Moreover, evolution of antibiotic resistance is highly affected by efflux mechanisms of *M. tuberculosis*. Thus, substances that are capable of inhibiting efflux could be used as adjuvants to antimicrobials. In Silva *et al*. [2017] we performed *in-silico* and *in-vitro* experiments to evaluate tetrahydropyridine derivatives as efflux inhibitors and to understand their mechanism of action. Minimum inhibitory concentration (MIC) determination, fluorometric methods and docking simulations were used. As receptor we considered AcrB protein of *Escherichia coli* and as ligands for docking ethidium bromide (EtBr), NUNL02, chlorpromazine (CPZ) and tetracycline (TET). The results (in-silico and experimental) suggested that these compounds are potential efflux inhibitors, highlighting the NUNL02 molecule. Using *slice docking*, our results showed that the mechanism of action could be by competition with the substrate for binding sites and protein residues.

Besides applications in Nanotoxicology and Tuberculosis we have been using docking other projects and collaborations. In the study Guidony *et al*. [2021] we collaborate performing molecular docking simulations to investigate the interaction between triclosan and ABC efflux pump proteins of zebrafish. This molecule triclosan is present in daily products as soap, deodorante, etc and consequently it has an ubiquitous presence in the environment. Thus it is important to evaluate how this compound can interact with animals living in this environment. The *in-silico* and *in-vitro* results of this study showed that triclosan can be cytotoxic to ZF-L. The docking simulations applied the Slice Docking strategy proposed by our group in Silva *et al*. [2017] where we varies the grid box of docking along the receptor structure.

Another popular application of molecular docking is for studies related to Cancer therapy. In this sense we collaborate in the investigation of the effect of C-phycocyanin (C-PC) in three human erythroleukemia cell lines with or

without the multidrug resistance (MDR) phenotype: K562 (non-MDR), K562-Lucena (MDR; overexpression of ATP-binding cassette, sub-family B/ABCB1), and FEPS (MDR; overexpression of ABCB1 and ATP-binding cassette, sub-family C/ABCC1) e Silva *et al*. [2018]. C-PC is a protein with anti-cancer properties through the induction of cytotoxicity and inhibition of cell proliferation. First, it was performed protein-protein docking to evaluate the interaction between C-PC and ABCC1/ABCB1. Then, we performed protein-ligand molecular docking to estimated FEB for C-PC associated with chemotherapy (VCR and DNR) where the results had showed that these drugs were located within the C-PC cavity and presented similar results. Finally, protein-protein docking, protein-ligand docking and experimental results showed that C-PC is a possible chemotherapeutic agent for cells with the MDR phenotype, both alone in K562-Lucena cells (resistance due to ABCB1), or in combination with other drugs for cells similar to FEPS (resistance due to ABCC1) e Silva *et al*. [2018].

Continuing in this subject, in Salgado *et al*. [2021] it was characterized the relationship between the COX2 and ALOX5 genes and their link with the MDR phenotype in sensitive K562 and MDR (K562-Lucena and FEPS) erythroleukemia cells. Among different experiments, cytometry data showed that there was an increase in ABCB1 protein expression after exposure to ASA. Thus, our collaboration in this article was to perform docking simulations considering as ligands ASA and as target the ABCB1 protein. The results showed that ASA may be a substrate for the efflux pump ABCB1.

Another application of docking was in the study described in Santa-Helena *et al*. [2020] where we collaborate verifying if 43 new fatty dihydropyridines (DHPs) have calcium channel affinity. Calcium channels are associated with numerous functions in the cardiovascular system and DHPs are considered the most commonly prescribed calcium channel blockers for the treatment of arterial hypertension. Nifedipine (NIF) is the prototype DHP known to interact with the alpha subunit of the voltage-dependent calcium channel. In docking, the receptor was the model of the pore region of the voltage-dependent L-type calcium channel subunit alpha-1S built using homology modeling. The 43 DHPs proposed and NIF structures were submitted to energy minimization, docking and protein-ligand contacts analysis. The ranking of ligands was defined based on the number of contacts shared by NIF and DHPs with the receptor. The results showed that two DHPs significantly reduced blood pressure and heart rate in rats as well as the levels of reactive oxygen species, with performance equal to or greater than NIF in most analyses performed Santa-Helena *et al*. [2020].

It is important to mention that in all these applications our background in computer science allow us to develop tools, scripts, algorithms when necessary to better analyse all data generated by these interdisciplinary projects.

3.2 Molecular Dynamics

Molecular Dynamics (MD) is a computational technique developed for studying positional and conformational dynamics of atoms and molecules Borhani and Shaw [2012]; Rodriguez-Bussey *et al*. [2016]. MD simulations apply Newtonian dynamics to simulate the movement of all atoms of a molecular systems (proteins, lipids, solvent) considering the temperature variation caused by the velocity of the atoms. It takes into consideration bounded and unbounded forces, obtained from a force field Borhani and Shaw [2012]. A force field is empirically obtained from experimentation in real systems and provides the functional form and parameters to the simulation of forces that act in atoms of a macromolecular system Lindorff-Larsen *et al*. [2012]; Piana *et al*. [2011].

MD simulations have been extensively used in rational drug design. This type of simulation considers both target proteins (receptors) and drugs (ligands) as flexible structures, therefore, it provides a wider and more realistic view of the affinities between proteins and drugs. Borhani and Shaw [2012]; Rodriguez-Bussey *et al*. [2016]. The observation of a protein and its atoms moving in a solvent is of fundamental importance for structural quality evaluation Figueiredo *et al*. [2014]; Piana *et al*. [2011].

MD simulations are very computationally expensive and this cost grows with the number of atoms present in the molecular system under study. A plethora of computer programs and computer architectures are employed in MD simulations each with their own advantages and drawbacks. Among the many available possibilities, in our group we use the set of computer programs known as Gromacs Berendsen *et al*. [1995]. Gromacs is an open source set of computer programs designed to computationally simulate biochemical molecules, i.e., proteins, lipids and nucleic acids. In other terms, it is a MD simulation package providing many functionalities among which the estimation of intramolecular forces in complex molecules and its solvent. Abraham *et al*. [2015]; Páll *et al*. [2014]; Pronk *et al*. [2013].

In addition for using molecular docking to study drug resistance and efflux mechanisms of Mtb, we also have been applying molecular modelling and MD simulations. Among multidrug and extensive drug resistance mechanisms of Mtb, Tap efflux pump is associated with resistance to isoniazid, rifampicine and ofloxacin and with multidrug resistance.

In Scaini *et al*. [2019] we built a reliable molecular model of Tap efflux pump and tested the possible competitive inhibition between efflux inhibitors and antibiotics in the optimized structure. From twenty five Tap proposed models one was selected and went through to a 100 ns molecular dynamics simulation in a lipid bilayer, and the resulting optimized structure was used in docking studies to test if the used efflux inhibitors may act via competitive inhibition on antibiotics . RSMD analysis revealed the model is stable, where the predicted binding site stabilized between 15 and 20 ns, maintaining the RMSD at around 0.35 Å throughout the MD simulation in a lipid bilayer. Therefore this model is reliable and can also be used for further studies. The docking studies showed a possibility of competitive inhibition by NUNL02 on ofloxacin and bedaquiline, and by verapamil on ofloxacin and rifampicin. This presents the possibility that NUNL02 and verapamil are possible inhibitors of Tap efflux and highlights the importance of including efflux inhibitors as adjuvants to the tuberculosis therapy, as it indicates a possible extrusion of ofloxacin, rifampicin and bedaquilin by Tap.

In a subsequent study, Dos Santos *et al*. [2021], the effect of mefloquine (MFL) association with two first-line anti-TB drugs and six fluoroquinolones was evaluated against *Mycobacterium tuberculosis* drug resistant strains. In this study, two approaches have been explored to investigate the synergistic interaction of MFL with isoniazid, pyrazinamide, and several fluoroquinolones. The first approach is the inspection of spectroscopic responses attributed to the effect of MFL on physicochemical properties related to a liposomal membrane model composed of soybean asolectin. The second approach is the molecular dynamics (MD) simulation considering MFL interaction with a membrane model based on PIM2, a lipid constituent of the mycobacterial cell wall. The results of both approaches agree. FTIR and NMR data showed that MFL greatly affects the region between the phosphate and the first methylene groups of soybean asolectin membranes, disordering these regions. MD simulations results detected high MFL density in the glycolipid interface and showed that the drug increases the membrane lateral diffusion, enhancing its permeability. The obtained results suggest that synergistic activities related to MFL are attributed to its effect of lipid disorder and membrane permeability enhancement.

In a diverse from M. tuberculosis study, Rocha *et al*. [2021], we use metadynamics simulations analysis to probe the dynamic, pharmacophoric and catalytic environment differences between the monomers of main-protease (Mpro) from SARS-CoV-2 (MCoV1pro) and the 12 mutations SARS-CoV-2 ((MCoV2pro). We explore how much intrinsic distinctions are preserved in the functional dimer of MCoV2pro, as well as its implications for ligand accessibility and optimized drug screening. The results point to the importance of taking into account the protein conformational multiplicity for new promising anti MCoV2pro ligands. We hope these results will be useful in prospecting, re-purposing and/or designing new anti SARS-CoV-2 drugs.

3.3 Genomics

Genomics is the area that studies the DNA and genes in an organism. In our group we have developed important collaborations in this area and recently have approved an important project for the creation of a centre for genomic surveillance in the south of Brazil.

In Lettnin *et al*. [2019] we evaluated how OCT4-PG1 pseudogene can affect OCT-4 expression and mechanisms related to the multidrug resistance (MDR) phenotype in FEPS cells. OCT-4 protein is a transcription factor that regulates expression of ABC transporters, level of gene expression, activity of ABC proteins and cell sensitivity to chemotherapy were evaluated after OCT4-PG1 silencing. To corroborate and evaluate the wet lab results, a STRING [2] network was created. The network structure demonstrated that OCT4-PG1 protein interacts directly with OCT-4, SOX2, and NANOG and indirectly with ABC transporters. The conclusion was that OCT4-PG1 pseudogene plays a key role in the regulation OCT-4 transcription factor, which alters MDR phenotype in the FEPS cell line.

[2]https://string-db.org/

In another study, Recamonde-Mendoza *et al*. [2019], we investigated the Transcription Factor and miRNAs related to cardiac hypertrophy (CH) using a systems biology approach. We constructed a comprehensive TF-miRNA co-regulatory network in CH and systematically characterize its structure, from node- to systems-level properties. We integrated expression profiles for miRNAs and messenger RNA (mRNA) from an in vitro model of CH were integrated with experimentally validated interactions from seven curated databases to build the CH-related TF-miRNA regulatory network.

A completely unsupervised approach was proposed to identify core regulatory elements, based on Borda count aggregation of distinct network-based properties. Results suggested that there are distinct roles of TFs and miRNAs, which tend to act mostly as network bottlenecks and hubs, respectively. The constructed CH-related regulatory network has the potential to provide new insights about key regulators, molecular mechanisms, and the interplay between miRNAs and TFs in the pathogenesis of CH, which after proper experimental validation may contribute to the search for new therapeutic approaches.

In Genomics domain we have also studied the evolutionary process of tuberculosis drug resistance, do Carmo Guimarães *et al*. [2021]. The microevolution of drug resistance in serial isolates from six previously treated patients is investigated through phenotypic methods, followed by genotypic approaches. Using whole-genome sequencing has enabled the identification of mutations in the katG, rpsL and rpoB genes associated with drug resistance, including the detection of rare mutations in katG and mixed populations of strains. Molecular docking simulation studies of the impact of observed mutations on isoniazid binding were also performed. Sequence technologies can detect rare mutations related to drug resistance, identify subpopulations of resistant strains, and identify diverse populations of strains due to exogenous reinfection, thus improving tuberculosis control by guiding early implementation of appropriate clinical and therapeutic interventions.

3.4 General Bioinformatics, Applications, Tools and Machine Learning

Here we present studies/tools that are either methodological or related with biological problems that applies machine learning techniques.

In Agostinho *et al*. [2015] we propose an alternative sampling scheme for the inference of regulatory networks (IRN) using Markov Chain Monte Carlo (MCMC) and Bayesian networks. Usually, Bayesian networks are sampled with a Markov Chain Monte Carlo (MCMC) sampler in the structure space. Unfortunately, conventional MCMC sampling schemes are often slow in mixing and convergence. To improve MCMC convergence, an alternative method is proposed and tested with different sets of data. Moreover, the proposed method is compared with the traditional MCMC sampling scheme.

The method consists of integrating in the MCMC a simpler and faster method for the IRN, Graphical Gaussian Models (GGMs), trough a Hierarchical Bayesian model. The conclusions is that the proposed method is a viable alternative

to improve mixing and convergence of traditional MCMC schemes. It allows the use of Bayesian networks with an MCMC sampler with less iterations. The proposed method has always converged earlier than the traditional MCMC scheme. An extension of this work is described in Barreto *et al*. [2017] where we propose to perform the IRN with MCMC Sampler Guided by Mutual Information (MI).

β-glucosidases are important enzymes for the biofuel industry since they catalyse the hydrolysis of oligosaccharides and disaccharides such as cellobiose playing an important role in cellulose-degrading. The challenge is that most of the characterised β-glucosidases can be inhibited by their product, glucose. Thus, glucose-tolerante β-glucosidases have been investigated and proposed to improve the production of second-generation biofuels. The study about β-glucosidases structures and mutations related to glucose-tolerance is one of the main objectives of the BABEL project.

In Mariano *et al*. [2017] we collaborate with BABEL to perform a systematic literature review (SLR) about this subject, collecting protein structures and constructing a database of glucose-tolerant β-glucosidases. Furthermore, different bioinformatics analyses were performed: homology models of the glucose-tolerant β-glucosidases without experimental structures available; characterisation of the catalytic pocket; molecular docking with cellobiose, and so on. As a result of this SLR, 11 residues were identified as highly conserved and their presence is essential for these β-glucosidases. Besides, in this article is proposed a metric based on the number of contacts of each important residue in the catalytic pocket to suggest mutations to modify the glucose tolerance of these proteins. In the same context, in Mariano *et al*. [2019] is presented a method, called structural signature variation (SSV), to propose mutations for improving enzymes' activity. To evaluate SSV, a case study for suggesting mutations in β-glucosidases was performed. As result, 15 mutations were identified as beneficial from where three have been related in literature to the mechanism of glucose tolerance. SSV might be useful for the engineering of enzymes used in biofuel production or other industrial applications.

The prediction of upon point mutations effect in protein structure stability is useful in a wide field of applications having great scientific and commercial interest. One way of performing such predictions is to compute the $\Delta\Delta G$ (variation in the variation of the free energy) of protein structure before and after the point mutation. There are many tools available for this purpose but they frequently disagree in their results. Since 2015, we have been studying Ensemble Learning Algorithms to aggregate the results from different stability prediction tools. In the Master Thesis Camargo [2017] we propose EN-Mutate, a web tool for integrating the results of different tools and also combining them using ensemble learning techniques. In Freitas *et al*. [2021] we investigated the usage of Stacking, Bagging, and Boosting as combiner functions. The results showed that our proposed ensemble approach is appropriate to predict the effect of point mutations on protein stability showing more reliable results than the individual tools improving overall accuracy, precision, and/or recall.

Also in the application of ML in Bioinformatics, in da Silva *et al*. [2019] we propose predictive models to relate Fluorescent protein color to its three-dimensional structure characteristics. Fluorescent proteins have the ability of absorbing light and re-emitting it at a longer wavelength. They have been applied in a wide variety of fields ranging from basic science to industrial applications. Apart from the naturally occurring fluorescent proteins (Green Fluorescent Protein (GFP) and Discosoma Red fluorescent protein (DsRed)), there is a growing interest in genetically modified variants that emit light in a specific wavelength possibility of investigating the behavior of cells, to mark tissues, etc. However, genetically modifying a protein is not an easy task, especially the challenge of maintaining protein stability. It is known that there is a relationship between the structure of the fluorescent protein and the wavelengths (color) emitted where amino acids close to the choromophore play an important role. Hence, we prepared an input dataset composed of 109 fluorescent proteins where predictive attributes corresponds to the 20 types of amino acids and their minimum distance for the choromophore and as target attribute the color emitted. We compare four classical supervised learning algorithms: artificial neural networks (ANNs), decision trees (DTs), support vector machines (SVMs) and random forests (Rfs). The results comparing the algorithms showed that DT, SVM and RF were significantly better than ANNs, and RF was the best method in all the scenarios. However, the interpretability of DTs is highly relevant and can provide important clues about the mechanisms involved in protein color emission. The results are promising and indicate that the use of in silico methods can greatly reduce the time and cost of the in vitro experiments.

4 Perspectives and Future directions

The future of Combi-Lab is very much in harmony with the aspiration of its members as well as in line with the funding opportunities that are available. Nowadays here in Brazil, perhaps in the world, when the Sars-Cov2 pandemics appear to be approaching its end, Bioinformatics research funding is more than ever being directed to genomic studies. Before we talk about the future possibilities, we present the most important grants and alliances that we have participated and made our existence a reality.

- **Edital Universal CNPQ:** Both principal investigators of this research group had approved proposals in Edital Universal CNPQ. These projects were in the areas of structural Bioinformatics (docking and Virtual Screening) and inference of regulatory networks.
- **Edital Biologia Computacional CAPES:** In 2013 we participated in the call of proposals *Biologia Computacional* from CAPES. We teamed up with UFMG and UFPB and wrote a project proposing two research themes. This originates the BABEL. Our project was classified in second out of the 30 participating. In this project we worked mainly in the problem of optimizing the production of Ethanol from Sugar cane.
- **CABANA EMBL-EBI:** Adriano V. Werhli was selected to participate in the EMBL-EBI project to build a Latin American capacity in Bioinformatics. During

its 6 months visiting to EBI, Adriano has worked in the inference of phosphoproteomics networks.

- **Visiting professor in INRIA-Grenoble:** during 2018-2019, Karina S. Machado worked as visiting professor in NANO-D laboratory in INRIA Grenoble receiving a six months scholarship from CAPES and three months from the University of Grenoble. During this year, Karina worked developing new scoring functions for molecular docking.

Having built our foundations based on the grants and alliances before mentioned we now present some future perspectives. Combi-Lab in partnership with various research institutions from the federation state of Rio Grande do Sul has recently being awarded a grant to implement the *Centro de Ciências de Dados e Bioinformática (CCDB)* (Bioinformatics and Data Sciences Center). The grant is part of a bigger project which aims to create the *Rede Gaúcha de genômica aplicada à saúde* (Gaucha's network of Genomics applied in Health). Within this network, CCDB is responsible for implementing and maintaining the computing infrastructure, which is capable of providing the computation resources for collaborators that are also part of the network.

The CCDB will provide computer and methodological support for two main research lines: i) genomic analysis of patient data, and ii) zoonotic surveillance. These two research lines will demand great scientific development in numerous areas, e.g., systems biology for data integration, *in silico* models for experiments planning, data storage and handling, distributed computing, among others.

Moreover, the proposed network will produce a deluge of genomic data. The transformation of this myriad of data in knowledge is highly dependent of the computer capacity in storing, manipulating, and visualising these data. CCDB will be responsible for providing easy access as well as to provide basic analysis of such data, thus incrementing the economic and social impact produced by the research outcomes.

Apart from the aforementioned project, that is approved and it is starting, Combi-Lab is part of many other very important initiatives. They will not be listed here for confidentiality reason, however, the majority are also in the area of genomics, and have a great intersection with CCDB project.

Our group has experienced significant progress thanks to the numerous grants, alliances, and partnerships we've established with colleagues and institutions both in Brazil and abroad. Bioinformatics, being a relatively new discipline, is not yet well-defined and is often associated with either biological or computer science departments. At our university, the Combi-Lab is recognized for its success in obtaining grants and collaborations, as well as for its production of high-quality research and the development of skilled researchers. As a result, it is primarily situated within the computer science department. Despite the challenges of finding a place in academia for a developing discipline, we remain committed to advancing the field and making meaningful contributions to both the scientific community and society at large.

In conclusion, we are committed to maintaining our steady progress and striving to become a reference in the field of Computational Biology both within our university and the broader community. We recognize the importance of giving back to society and are determined to do so to the best of our abilities, as a way of showing our appreciation for the investments made in our work.

Acknowledgements

The authors acknowledge the financial support given by CAPES, CNPq and FAPERGS.

Funding

The authors acknowledge the financial support given by CAPES, CNPq and FAPERGS processes 22/2551-0000385-0 and 22/2551-0000390-7.

Authors' Contributions

KSM and AVW contributed equally to the conception and writing of this article. All authors read and approved the final manuscript.

Competing interests

The authors declare that they do not have competing interests.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25.

Agostinho, N., Machado, K. S., and Werhli, A. V. (2015). Inference of regulatory networks with a convergence improved MCMC sampler. *BMC Bioinformatics*, 16:306. DOI: 10.1186/s12859-015-0734-6.

Alonso, H., Bliznyuk, A. A., and Gready, J. E. (2006). Combining docking, molecular dynamic simulations in drug design. *Med. Res. Rev.*, 26:531–568.

Arrua, O. (2020). Função de escore baseada em machine learning para docagem molecular proteína-ligante . Master's thesis, PPGComp, Engenharia de Computação, Universidade Federal do Rio Grande (FURG), Rio Grande, Brasil.

Barreto, N. M., Machado, K. S., and Werhli, A. V. (2017). Inference of regulatory networks with mcmc sampler guided by mutual information. In *Proceedings of the Symposium on Applied Computing*, SAC '17, pages 18–23, New York, NY, USA. ACM. DOI: 10.1145/3019612.3022189.

Batool, M., Ahmad, B., and Choi, S. (2019). A structure-based drug discovery paradigm. *International journal of molecular sciences*, 20(11):2783.

Berendsen, H. J., van der Spoel, D., and van Drunen, R. (1995). Gromacs: A message-passing parallel molecular dynamics implementation. *Computer physics communications*, 91(1-3):43–56.

Borhani, D. W. and Shaw, D. E. (2012). The future of molecular dynamics simulations in drug discovery. *Journal of computer-aided molecular design*, 26(1):15–26.

Camargo, A. D. (2017). EN-MUTATE : predição do impacto de mutações pontuais em proteínas utilizando Ensemble Learning. Master's thesis, PPGComp, Engenharia de Computação, Universidade Federal do Rio Grande (FURG), Rio Grande, Brasil.

da Silva, R. S., Marins, L. F., Almeida, D. V., dos Santos Machado, K., and Werhli, A. V. (2019). A comparison of classifiers for predicting the class color of fluorescent proteins. *Computational Biology and Chemistry*, 83:107089. DOI: https://doi.org/10.1016/j.compbiolchem.2019.107089.

do Carmo Guimarães, J. d. L., von Groll, A., Unis, G., Dalla-Costa, E. R., Rossetti, M. L. R., Vianna, J. S., Ramos, D. F., Reis, A. J., Halicki, P. C. B., Scaini, J. L. R., *et al*. (2021). Whole-genome sequencing as a tool for studying the microevolution of drug-resistant serial mycobacterium tuberculosis isolates. *Tuberculosis*, 131:102137.

Dos Santos, M. C., Scaini, J. L. R., Lopes, M. V. C., Rodrigues, B. G., Silva, N. O., Borges, C. R. L., Dos Santos, S. C., dos Santos Machado, K., Werhli, A. V., da Silva, P. E. A., *et al*. (2021). Mefloquine synergism with antituberculosis drugs and correlation to membrane effects: Biologic, spectroscopic and molecular dynamics simulations studies. *Bioorganic Chemistry*, 110:104786.

dos Santos Machado, K. and Grudinin, S. (2020). On recent methods for incorporating receptor flexibility in molecular docking. In *Proceedings of the XXI Congrès do Groupe de Graphisme et Modélisation Moléculaire (GGMM),*, pages 65–65. ACM.

Durruthy, M. G., Monserrat, J. M., de Oliveira, P. V., Fagan, S. B., Werhli, A. V., Machado, K., Melo, A., González-Díaz, H., Concu, R., and Cordeiro, M. N. D. D. S. (2019). Computational mitotarget-scanning based on topological vacancies of single-walled carbon nanotubes with human mitochondrial hvdac1 channel. *Chemical Research in Toxicology*, 32(4):566–577.

e Silva, E. F., Figueira, F., Cañedo, A., Machado, K., Salgado, M., Silva, T., Wagner, E., Mattozo, F., Lima, É., Sales-Neto, J., *et al*. (2018). C-phycocyanin to overcome the multidrug resistance phenotype in human erythroleukemias with or without interaction with abc transporters. *Biomedicine & Pharmacotherapy*, 106:532–542.

Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021). Autodock vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898.

Figueiredo, D. F., Antunes, D. A., Rigo, M. M., Mendes, M. F., Silva, J. P., Mayer, F. Q., Matte, U., Giugliani, R., Vieira, G. F., and Sinigaglia, M. (2014). Lessons from molecular modeling human α-l-iduronidase. *Journal of Molecular Graphics and Modelling*, 54:107–113.

Foldvari, M. and Bagonluri, M. (2008). Carbon nanotubes as functional excipients for nanomedicines: Ii. drug delivery and biocompatibility issues. *Nanomedicine: Nanotechnology, Biology and Medicine*, 4(3):183–200.

Freitas, E. K. H. d., Camargo, A. D., Balboni, M., Werhli, A. V., and Santos Machado, K. d. (2021). Ensemble of protein stability upon point mutation predictors. In *Brazilian Conference on Intelligent Systems*, pages 73–88. Springer.

Gonzalez-Durruthy, M., Werhli, A. V., Cornetet, L., Machado, K. S., Gonzalez-Diaz, H., Wasiliesky, W., Ruas, C. P., Gelesky, M. A., and Monserrat, J. M. (2016). Predicting the binding properties of single walled carbon nanotubes (SWCNT) with an ADP/ATP mitochondrial carrier using molecular docking, chemoinformatics, and nano-QSBR perturbation theory. *RSC Adv.*, 6:58680–58693. DOI: 10.1039/C6RA08883J.

González-Durruthy, M., Werhli, A. V., Seus, V., Machado, K. S., Pazos, A., Munteanu, C. R., González-Díaz, H., and Monserrat, J. M. (2017). Decrypting strong and weak single-walled carbon nanotubes interactions with mitochondrial voltage-dependent anion channels using molecular docking and perturbation theory. *Scientific reports*, 7(1):1–19.

Guidony, N. S., Scaini, J. L. R., Oliveira, M. W. B., Machado, K. S., Bastos, C., Escarrone, A. L., and Souza, M. M. (2021). Abc proteins activity and cytotoxicity in zebrafish hepatocytes exposed to triclosan. *Environmental Pollution*, 271:116368.

Irwin, J., Sterling, T., Mysinger, M., Bolstad, E., and Coleman, R. (2012). ZINC: A free tool to discovery chemistry for biology. *Journal of Computational Chemistry*, pages 1757–1768.

Josende, M. E., Nunes, S. M., de Oliveira Lobato, R., González-Durruthy, M., Kist, L. W., Bogo, M. R., Wasielesky, W., Sahoo, S., Nascimento, J. P., Furtado, C. A., *et al*. (2020). Graphene oxide and gst-omega enzyme: An interaction that affects arsenic metabolism in the shrimp litopenaeus vannamei. *Science of The Total Environment*, 716:136893.

Kadukova, M., Machado, K. d. S., Chacón, P., and Grudinin, S. (2021). Korp-pl: a coarse-grained knowledge-based scoring function for protein–ligand interactions. *Bioinformatics*, 37(7):943–950.

Kuntz, I. D. (1992). Structure-based Strategies for Drug Design and Discovery. *Science*, 257:1078–1082.

Lettnin, A. P., Wagner, E. F., Carrett-Dias, M., dos Santos Machado, K., Werhli, A., Cañedo, A. D., Trindade, G. S., and de Souza Votto, A. P. (2019). Silencing the oct4-pg1 pseudogene reduces oct-4 protein levels and changes characteristics of the multidrug resistance phenotype in chronic myeloid leukemia. *Molecular biology reports*, 46(2):1873–1884.

Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S., and Shaw, D. E. (2012). Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *Journal of the American Chemical Society*, 134(8):3787–3791.

Lopes, P. P. (2021). Uma abordagem Ensemble Learning para aprimorar a predição de energia livre de ligação entre complexos proteína-ligante. Master's thesis, PPGComp, Engenharia de Computação, Universidade Federal do Rio Grande (FURG), Rio Grande, Brasil.

Lybrand, T. (1995). Ligand-Protein Docking and Rational Drug Design. *Curr. Opin. Struct. Biol.*, 5:224–228.

Machado, K. S. (2011). *Seleção eficiente de conformações de receptor flexível em simulações de docagem molecu-*

lar. Tese de Doutorado. Programa de Pós Graduação em Ciência da Computação PPGCC. Pontifícia Universidade Católica do Rio Grande do Sul.

Mariano, D., Leite, C., Santos, L., Marins, L., Machado, K.S. ; Werhli, A., Lima, L., and De Melo-Minardi, R. (2017). Characterization of glucose-tolerant beta-glucosidases used in biofuel production under the bioinformatics perspective: a systematic review. *GENETICS AND MOLECULAR RESEARCH*, 16:1–10.

Mariano, D. C. B., Santos, L. H., Machado, K. d. S., Werhli, A. V., de Lima, L. H. F., and de Melo-Minardi, R. C. (2019). A computational method to propose mutations in enzymes based on structural signature variation (ssv). *International journal of molecular sciences*, 20(2):333.

Meng, X. Y. Y., Zhang, H. X. X., and Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7:146–157.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *Computational Chemistry*, 16:85–91.

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., and Lindahl, E. (2014). Tackling exascale software challenges in molecular dynamics simulations with gromacs. In *International conference on exascale applications and software*, pages 3–27. Springer.

Perazzo, G. X., Winck, A. T., and Machado, K. S. (2013). A data warehouse as an infrastructure to mine molecular descriptors for virtual screening. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 1335–1336, Coimbra, Portugal.

Piana, S., Lindorff-Larsen, K., and Shaw, D. E. (2011). How robust are protein folding simulations with respect to force field parameterization? *Biophysical journal*, 100(9):L47–L49.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., *et al*. (2013). Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854.

Ramos, P., Schmitz, M., Gama, S., Portantiolo, A., Durruthy, M. G., de Souza Votto, A. P., Cornetet, L. R., dos Santos Machado, K., Werhli, A., Tonel, M. Z., *et al*. (2018). Cytoprotection of lipoic acid against toxicity induced by saxitoxin in hippocampal cell line ht-22 through in silico modeling and in vitro assays. *Toxicology*, 393:171–184.

Ramos, P. B., Colombo, G. M., Schmitz, M. J., Simião, C. S., dos Santos Machado, K., Werhli, A. V., Costa, L. D. F., Yunes, J. S., Prentice, C., Wasielesky, W., *et al*. (2022). Chemoprotection mediated by açaí berry (euterpe oleracea) in white shrimp litopenaeus vannamei exposed to the cyanotoxin saxitoxin analyzed by in vivo assays and docking modeling. *Aquatic Toxicology*, 246:106148.

Recamonde-Mendoza, M., Werhli, A. V., and Biolo, A. (2019). Systems biology approach identifies key regulators and the interplay between mirnas and transcription factors for pathological cardiac hypertrophy. *Gene*, 698:157–169.

Rocha, R. E., Chaves, E. J., Fischer, P. H., Costa, L. S., Grillo, I. B., da Cruz, L. E., Guedes, F. C., da Silveira, C. H., Scotti, M. T., Camargo, A. D., *et al*. (2021). A higher flexibility at the sars-cov-2 main protease active site compared to sars-cov and its potentialities for new inhibitor virtual screening targeting multi-conformers. *Journal of Biomolecular Structure and Dynamics*, pages 1–21.

Rodriguez-Bussey, I. G., Doshi, U., and Hamelberg, D. (2016). Enhanced molecular dynamics sampling of drug target conformations. *Biopolymers*, 105(1):35–42.

Salgado, M. T. S. F., Lopes, A. C., e Silva, E. F., Cardoso, J. Q., Vidal, R. S., Cavalcante-Silva, L. H. A., Carvalho, D. C. M., dos Santos Machado, K., Rodrigues-Mascarenhas, S., Rumjanek, V. M., *et al*. (2021). Relation between abcb1 overexpression and cox2 and alox5 genes in human erythroleukemia cell lines. *Prostaglandins & other lipid mediators*, 155:106553.

Santa-Helena, E., da Costa Cabrera, D., D'Oca, M. G. M., Scaini, J. L. R., de Oliveira, M. W. B., Werhli, A. V., dos Santos Machado, K., Gonçalves, C. A. N., and Nery, L. E. M. (2020). Long-chain fatty dihydropyridines: Docking calcium channel studies and antihypertensive activity. *Life Sciences*, 259:118210.

Scaini, J. L. R., Camargo, A. D., Seus, V. R., von Groll, A., Werhli, A. V., da Silva, P. E. A., and dos Santos Machado, K. (2019). Molecular modelling and competitive inhibition of a mycobacterium tuberculosis multidrug-resistance efflux pump. *Journal of Molecular Graphics and Modelling*, 87:98–108.

Seus, V. R., Perazzo, G. X., Winck, A. T., Werhli, A. V., and Machado, K. S. (2014). An infrastructure to mine molecular descriptors for ligand selection on virtual screening. *BioMed Research International*, 2014.

Seus, V. R., Silva, Jr., L., Gomes, J., da Silva, P. E. A., Werhli, A. V., Prates, N., Zanatta, N., and Machado, K. S. (2016). A framework for virtual screening. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, SAC '16, pages 31–36, New York, NY, USA. ACM.

Shen, C., Ding, J., Wang, Z., Cao, D., Ding, X., and Hou, T. (2020). From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(1):e1429.

Silva, L., Carrion, L. L., von Groll, A., Costa, S. S., Junqueira, E., Ramos, D. F., Cantos, J., Seus, V. R., Couto, I., da Silva Fernandes, L., *et al*. (2017). In vitro and in silico analysis of the efficiency of tetrahydropyridines as drug efflux inhibitors in escherichia coli. *International journal of antimicrobial agents*, 49(3):308–314.

Su, M., Yang, Q., Du, Y., Feng, G., Liu, Z., Li, Y., and Wang, R. (2018). Comparative assessment of scoring functions: the casf-2016 update. *Journal of chemical information and modeling*, 59(2):895–913.

Trott, O. and Olson, A. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, 31:455–461.