Journal of Information and Data Management, 2024, 15:1, doi: 10.5753/jidm.2024.2607

Biophysical Chemistry of Macromolecules Research Group at the State University of Maringá

Diego de Souza Lima [**Universidade Estadual de Maringá** | ***diegodslima92@gmail.com***]
Gisele Strieder Philippsen [**Universidade Federal do Paraná** | ***gistrieder@ufpr.br***]
Elisangela Andrade Ângelo [**Instituto Federal de Educação, Ciência e Tecnologia do Paraná** | ***elisangela.angelo@ifpr.edu.br***]
Maria Aparecida Fernandez [**Universidade Estadual de Maringá** | ***mafernandez@uem.br***]
Flavio Augusto Vicente Seixas [**Universidade Estadual de Maringá** | ***favseixas@uem.br***]

Department of Technology, Universidade Estadual de Maringá, Av. Ângelo Moreira da Fonseca, 1800 - Parque Danielle, Umuarama, PR, 87506-370, Brazil.

Received: 31 March 2022 • **Published:** 21 February 2024

Abstract The interdisciplinary field of Biophysical Chemistry, which applies concepts from Physical Chemistry to describe biological phenomena, is essential for modern molecular biology advancements. This approach enables the description of biological systems in terms of their constituent parts, such as atoms and molecules, facilitating a structural understanding of their characteristics. Nonetheless, to describe such large systems, computational methods are needed. The Biophysical Chemistry of Macromolecules research group at the State University of Maringá is dedicated to investigating such systems, mainly protein-ligand complexes, through bioinformatics approaches combined with experimental techniques to validate *in silico* results. The main purpose of the research projects is to develop applications for drug discovery in the context of antimicrobial, antiviral, antifungal, and antihyperglycemic agents, with the aim of advancing the field of bioinformatics in Brazil.

Keywords: Bioinformatics; Machine learning; Molecular docking; Molecular dynamics; Virtual screening

1 Introduction

1.1 History

The Biophysical Chemistry of Macromolecules research group (BCM) located at the State University of Maringá (Universidade Estadual de Maringá - UEM) dedicates itself to investigating the biophysical phenomena of protein-ligand interactions through bioinformatics tools. The main goal is to prospect novel drugs for human or veterinary uses, but other phenomena, derived or not from protein-ligand interactions, are also investigated. The results obtained using computational methods are later validated *in vitro* and/or *in vivo* through biochemical assays, such as gene expression studies, protein binding, cell culture, and animal model assays. Our group follows up the *in silico* studies with experiments that aim to validate the generated data through simulations when possible, or describe biological phenomena observed in experiments done by our group or from collaborators.

At the Structural Biochemistry Laboratory (SBL) on the UEM campus in Umuarama, Paraná, Brazil, the BCM group conducts its activities. The construction of the building that houses the laboratory began in 2009 with funding from an extension project involving private enterprises. The SBL building covers an area of 110 m^2 and comprises three sectors: bioinformatics, protein extraction and purification, and biophysical chemistry of macromolecules. Projects approved by the development agency of the state of Paraná (Fundação Araucária), the Secretariat of Science and Technology of the state of Paraná (SETI), and the federal government through FINEP, provided financial support for the establishment of these sectors. The maintenance of equipment and consumables are financed by CAPES, CNPq, Fundação Araucária, and through revenue from a project providing services in physical-chemical and microbiological analysis of raw food materials and industrialized products. After the laboratory construction in 2010, the BCM group began to structure itself, with Prof. Dr. Flavio Augusto Vincente Seixas as the leader, two laboratory technicians, and undergraduate students. The first works focused on the biophysical chemistry of macromolecules and protein crystallography [Seixas et al., 2011], protein structure modeling, and molecular docking [Homem et al., 2013].

The visibility of these works promoted collaboration with other research groups from UEM and other institutions in Brazil. Participation in the postgraduate program in Cell Biology at UEM began in 2012. Since then, the number of researchers and students in the group has increased, as well as the number of publications and patent deposits, reaching an average of 11 productions per year. Currently, the composition of the BCM researchers is quite diversified due to the multidisciplinary nature of the research performed, and is composed by chemists, biochemists, veterinarians, pharmacists, physicists, and biologists, with the participation of undergraduate, master, doctoral, and post-doctoral students. The research collaborations of the BCM group happens in partnership with researchers from UEM, Brazil, and abroad, mainly with Argentina, Austria, and the United Kingdom.

1.2 Structural Biochemistry Laboratory

Among the SBL sectors, bioinformatics is the one that stands out in terms of scientific production. Equipped with five high-performance AMD octa-core computers with CUDA accelerators, this sector is well-equipped to handle various tasks, such as docking simulations, virtual screening, molecular dynamics simulations, binding energy calculations, rendering, and image processing. In addition, BCM researchers have ongoing projects at CENAPAD/SP (proj520 and proj870) and LNCC/SDumont, allowing access to computer clusters with the highest performance in Brazil.

The protein extraction and purification sector is equipped with a refrigerated centrifuge, liquid chromatographs (Akta Prime and Akta Pure M), an ultra freezer at -80 °C, and several other small types of equipment. In this sector, experiments are carried out to extract and purify proteins of microbial, animal, and plant origin, which are used in biophysical validation tests carried out at the SBL itself or by the laboratories of collaborating groups.

The biophysical chemistry of macromolecules sector is equipped with UV-Vis and Near-IR scanning spectrophotometers, HPLC and GC chromatographs, a complete 2D electrophoresis system, photo documentation, and several other smaller analytical benchtop equipment. This structure allows the performance of enzyme kinetic assays, protein binding, and structural characterization of macromolecules in solution. Other biophysical assays are also performed at the UEM Research Support Center Complex (COMCAP) using equipment such as circular dichroism, Q-TOF, dynamic light scattering (DLS), differential scanning calorimetry (DSC), GC-MS, HPLC-MS/MS, and nuclear magnetic resonance (NMR). The laboratory also maintains collaborations with biophysics laboratories and institutions in Brazil and abroad, occasionally utilizing different testing equipment to complement its own.

To validate simulations, the BCM research group employs molecular cloning to obtain highly pure recombinant proteins for biophysical assays. This technique enables the production of proteins in adequate concentrations for various experiments, such as enzymatic kinetics and enzyme inhibitor interactions, providing critical experimental support for *in silico* results.

There are several ways to perform molecular cloning, depending on the protein of interest and available resources. In the BCM group, gene expression is conducted in strains of *E. coli*, which accounts for more than 80% of recombinant protein expression studies. One of the key steps in molecular cloning is the plasmid design containing the gene of interest. It must be compatible with the *E. coli* strain and contain markers that allow the selection of recombinant strains, as well as tags (markers) that facilitate the purification process. The choice of label for purification must take into account the protein characteristics and the purification columns used in later steps. Among the commonly used tags, the histidine tag stands out, which adds a sequence with approximately six histidines to the protein of interest. This small histidine tail interacts with the nickel present in the stationary phase during liquid chromatography purification, thus isolating the protein of interest.

Since the SBL does not have a bioterium or a microbiology laboratory, experimental validation assays involving cell culture and animal models are carried out in collaboration with our partners.

2 Computational analysis protocols

2.1 Docking and virtual screening

The bioinformatics research from BCM has the experimental validation of the *in silico* results as one of its differentials. Nonetheless, to identify a valid ligand, the virtual screening process follows a strict prospecting protocol. It begins with the choice of the target protein, giving preference to high-resolution crystallographic structures in the presence of ligands as substrates, substrate analogs, or inhibitors. The last criterion is adopted because proteins in the Apo form (without ligands) may not be in the catalytically active conformation. After all, significant conformational changes may occur during the binding process, avoiding the correct docking of the ligand candidate.

Docking programs use different search and ranking methods that score electrostatic and/or hydrophobic interactions differently [Fan et al., 2019]. Therefore, selecting the optimal docking program depends on the nature of interactions between the protein and the crystallographic ligand. In our protocol, this choice is made by redocking the crystallographic ligand using different programs. We consider a program validated and suitable for the virtual screening phase if it can recover the pose of the crystallographic ligand with a low root mean square deviation (RMSD) in at least five repetitions. Thus, we assume that the program ”understands” the nature of the protein-ligand interactions in the investigated system and can be applied to a library of unknown compounds. SBL uses AutoDock [Morris et al., 2012], AutoDock Vina [Trott and Olson, 2010], Molegro Virtual Docker [Thomsen and Christensen, 2006], CLC Drug Discovery Workbench, and GOLD [Verdonk et al., 2003] as docking programs. At least two different programs must be validated at this stage to allow the average relative score evaluation used for ligand ranking and selection. If two or more different programs, using different search and ranking methods, led to a common result, the likelihood that the ligand found may be a true ligand increases.

Obtaining the crystallographic structure of the target protein from the Protein Data Bank (PDB) [Burley et al., 2018] is not always feasible, which requires the structure to be modeled in such cases. In this process, the homologous ligand-bound crystallographic structure with the highest sequence identity is employed as a template. One issue is that the modeling programs do not optimize the amino acid residues interactions with the ligand; thus, it is necessary to perform structure minimization to maximize these interactions. This is performed with both, the modeled protein and the ligand crystallographic structure when its redocking fails in all the evaluated programs. The minimization process is performed with the NAMD2/VMD

[Phillips et al., 2005] packages using the Charmm36 force field [Brooks et al., 2009]. Ligand force field parameters are generated by the SwissParam server [Zoete et al., 2011], where the atom partial charges are replaced by the CHELPG partial charges calculated by the B3LYP 6-311G method in an SPC solvation model from the Orca program [Neese, 2012]. The minimization process is performed in steps: in the first step, the structure of the complex is immersed in a periodic box containing water molecules and sufficient amounts of counter ions Na+ or Cl- to neutralize the system charges, which is subjected to approximately 20,000 steps of conjugate gradient (CG) minimization. In this first step, the ligand atoms are fixed in space so that the experimental crystallographic information regarding their position is preserved and the protein adapts to their presence. In the second step, the whole system is subjected to another 10,000 CG steps, this time with the ligand atoms also free. Most of the time, the structure resulting from this procedure is sufficient to obtain a low redocking RMSD. When this does not happen, the complex is subjected to a third step where the atoms of the protein and ligand are fixed in space and the waters and salts (including crystallographic waters) are subjected to 60 ps of equilibrium molecular dynamics under NPT conditions (constant Number of atoms, Pressure: 1 atm, and Temperature: 300K). In the fourth step, the complex resulting from the previous step is subjected to another 20,000 CG steps with all atoms free. The structure resulting from this step is then analyzed by the ligand redocking procedure in different programs, each with its own particularity (size of the box or search radius, choice of the best search and ranking algorithms, efficiency, presence/absence of structural waters, etc).

After target preparation and validation of the docking protocols, virtual screening is performed using the crystallographic/modeled ligand score as a reference score (selection threshold), and also for mapping the main protein-ligand interactions. Only the screened ligands that obtain higher scores than the reference score (best results), with more interactions with the protein, or stronger electrostatic interactions, are selected for the next step. Then, the best results are re-evaluated by the same program, and only those that ranked higher than the reference in all three additional simulations, go on to the next step, which is an evaluation by another program previously validated by redocking. With the new program, the same criterion of evaluating the best results by means of four repetitions is used, and now only the ligands that were repeated in all eight simulations (four repetitions with each program) move on to the experimental validation step. If three or more programs were validated by redocking, the best results were evaluated by all programs by four repetitions each. It is important to compare the best poses for each ligand, obtained by each program. If they are very similar (low RMSD), the probability that it is a true ligand is higher. However, if they are random (high RMSD between poses from different programs), this indicates that no interaction pattern was found, implying a lower probability of being a true ligand.

Since each program provides scores with different dimensions, it is difficult to compare the results and select the best ligand from a shortlist. Therefore, to rank the compounds, the average scores of the compounds obtained with the repetitions by each program are normalized by the min-max method (Equation 1), where x is the average docking score of ligand i obtained by the program j, x^{max} is the maximum score obtained, and x^{min} is the minimum score:

$$x'_{i,j} = \frac{x_{i,j} - x_j^{min}}{x_j^{max} - x_j^{min}} \quad (1)$$

Then, the arithmetic mean of the relative scores for each ligand ($\overline{x}'_i$) is calculated (Equation 2), which is used to rank the compounds in descending order:

$$\overline{x}'_i = \frac{\sum_{j=1}^{n} x'_{i,j}}{n} \quad (2)$$

2.2 Molecular dynamics simulations

Molecular dynamics (MD) simulations performed by the BCM group have two major purposes. Regarding the first one, the equilibrium MD aims to check whether ligands identified by virtual screening tend to be retained at the binding site under equilibrium conditions over time, since ligands identified by false-positive results tend to undock from the protein during the simulation. The program used for these simulations is NAMD2 with a Charmm36 force field. The second purpose relates to equilibrium MD simulations that aims to evaluate the conformational stability of nucleic acids, especially certain types of noncoding RNAs (ncRNAs) [Duarte Junior et al., 2019]. For these simulations, Gromacs [Abraham et al., 2015] with Amber force field [Wang et al., 2004] is used, because Amber force field is considered one of the best resources for nucleic acid simulations.

MD for free energy calculation, on the other hand, serves to estimate the $\Delta G_{binding}$ of a group of ligands and identify the one with the highest affinity, to select it for validation studies *in vitro* and *in vivo*. In these calculations, the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method [Genheden and Ryde, 2015] is used. The program employed for these simulations is Gromacs with force field Gromos96.

The laboratory has been working with collaborators to perform additional experiments to validate the predictions of the simulations. Beyond that, BCM group has provided support to collaborators by proposing, through simulations, the likely metabolic targets from the effect observed in their biochemical assays, as well as in the description of the protein-ligand interactions at the molecular level.

3 Main works and publications

The protocols described above have been applied in several works developed by BCM research group. The applications described below refer to the most impactful papers published.

3.1 Applications in drug discovery and development

3.1.1 Antihyperglycemic drugs

The enzymes salivary α-amylase and pancreatic α-amylase catalyze the hydrolysis of oligosaccharides and polysaccharides, such as starch and glycogen, with glucose as the end product after additional enzymatic reactions. Although this constitutes an important energy-obtaining mechanism for living organisms, the inhibition of these enzymes has the effect of reducing the glycemic index, leading to a therapeutic approach for metabolic disorders such as diabetes and obesity, and establishing an alternative to pharmacological strategies of hormonal regulation of blood glucose [Kato et al., 2017]. Thus, the structures of human salivary α-amylase and porcine pancreatic α-amylase were modeled in the presence of the substrate amylose tetrasaccharide and minimized by molecular dynamics to led the enzymes in a catalytically favorable conformation. These structures were used as targets in virtual screening with the inhibitor acarbose and the substrate amylose tetrasaccharide as reference ligands for docking scores, as well as a guide for analyzing intermolecular interactions. Thus, different libraries were tested, such as that of hydrolyzed and condensed tannins [Bueno et al., 2019a], the library of compounds characterized from Merlot grape pomace extracts [Kato-Schwartz et al., 2020a], the library of compounds from *Poncianella pluviosa* extract [Kato-Schwartz et al., 2020b], and compounds from *Camellia sinensis* [da Silva et al., 2021]. In these studies, docking simulations indicated the compounds from each extract that most likely bind to the amylases and cause a reduction in their catalytic activity, as observed in *in vitro* kinetic studies, and the possible binding modes for these inhibitors. As a result, these works suggested natural compounds present in the extracts that could help in the control of postprandial glycemia.

3.1.2 Acetylcholinesterase and butyrylcholinesterase inhibitors

Molecular docking approaches have also found applications in addressing issues related to the esterase class of enzymes. Acetylcholine (ACh), a neurotransmitter involved in the process of conducting electrical impulses from one neuron to another, is rapidly hydrolyzed by the enzymes acetylcholinesterase (AChE) [Anand and Singh, 2013] and butyrylcholinesterase (BuChE) [Cerbai et al., 2007; Johnson and Moore, 2012], leading to reduced ACh levels in Alzheimer's disease. As a result, there is a need to identify agents that can inhibit these enzymes, thereby increasing ACh levels in the central nervous system and controlling the disease [Anand and Singh, 2013; Yamazaki et al., 2021].

The inhibitory potential of biopeptides derived from hydrolyzed fish proteins, targeting the AChE enzyme, was demonstrated through *in vitro* enzyme inhibition studies [Moreira et al., 2022]. In collaboration with the BCM group, we conducted docking simulations and identified that the amino acid L-arginine was the component most likely to bind to AChE, providing a possible molecular description of the phenomenon observed *in vitro*.

In the work of Yamazaki et al. [Yamazaki et al., 2021], a series of arylcarbamate-N-acylhydrazone-derived compounds were synthesized, which showed potential for inhibition of the BuChE enzyme *in vitro*, particularly by compound 10c. Docking simulations of the synthesized compounds were performed with BuChE, which also pointed to compound 10c as the most likely to bind to the enzyme. By mapping the interactions of the compound with BuChE, it was possible to design and test ligand modifications *in silico*, allowing the guided synthesis of new compounds derived from 10c (research in progress).

3.1.3 Identification of new antifungal candidates

Paracoccidioidomycosis (PCM) is a systemic mycosis caused by fungi of the genus *Paracoccidioides spp.* [Shikanai-Yasuda et al., 2017]. It is considered a neglected disease with occurrence in Latin America, mainly among rural workers who had contact with contaminated soil. The pharmacological arsenal for treating mycoses is very limited compared to that available to combat bacteria. Moreover, the lack of a specific drug for the treatment of PCM forces the use of broad-spectrum antifungals for a period of 12 to 24 months, which usually leads to interruption of the treatment due to the costs and time involved. This motivated the search for compounds that could act as antifungal agents by inhibiting different metabolic targets.

The enzyme homoserine dehydrogenase (HSD) is involved in the synthesis of the amino acids methionine, threonine, and isoleucine in fungi and is absent in mammals, which makes it an interesting drug target. In the study conducted by BCM research group [Bueno et al., 2019b], the structure of HSD from *P. brasiliensis* (*Pb*HSD) was modeled in the presence of the cofactor NAD+ and the substrate L-homoserine to be used in virtual screenings with the library of natural products from Zinc database (112,572 molecules). After applying our standard protocol for virtual screening and equilibrium molecular dynamics simulations, as described earlier, three compounds were selected for minimum inhibitory concentration (MIC) and minimum fungicidal concentration (MFC) assays. The most promising compound, named HS9, showed a fungicidal profile with MIC and MFC values of 8 $\mu g \cdot mL^{-1}$ and low mammalian cell toxicity, which motivated the filing of a patent application at the National Institute of Industrial Property (INPI) (process BR102018012453-6). The results of this work identified a promising compound for studies in an animal model of PCM treatment. Given the experience of the BCM group with binding and inhibition assays using recombinant *Pb*HSD, we were invited to participate in collaborative work to test 4-methoxy-naphthalene derivative compounds as inhibitors of this enzyme [Bagatin et al., 2019]. The contribution of all this bioinformatics work, coupled with experimental validations, has reinforced the potential of HSD as a target for novel antifungal drugs.

Another target for antifungal application explored by the BCM group is the enzyme chorismate synthase (CS), which is also present in fungi but absent in humans. This enzyme catalyzes the seventh step of the shikimate

pathway to produce chorismate, which is the branch point for the biosynthesis of aromatic metabolites important for the survival of the fungus, such as amino acids, folate, naphthoquinones, and menaquinones. The CS enzyme of *P. brasiliensis* (*Pb*CS) was modeled in the presence of the cofactor FMNH2 and the substrate 5-enolpyruvylshikimate-3-phosphate (EPSP) and used in virtual screening of different compound libraries, which generated different publications. In the first work, the natural product library from the Zinc Database plus the library of synthetic compounds with 80% similarity to known CS inhibitors were used, totaling 206,955 compounds. Three compounds were selected after the screening of this library, with docking scores higher than EPSP. These compounds proved to be stable in MD simulations performed with the respective protein-ligand complex and were purchased for *in vitro* assays. The binding and inhibition studies of the recombinant enzyme were conducted by Dr. Peter Macheroux's group at the Graz University of Technology in Austria since his group is the only one in the world with the technology to perform CS kinetic assays. The compound named CP1 showed an IC_{50} of $47\mu M$ and K_D of $64\ \mu M$, values considered excellent for this enzyme. The MIC and MFC studies showed a fungicidal profile with values of 2 and $4\ mg \cdot L^{-1}$, respectively. Studies in a murine model of infection *in vivo* demonstrated that CP1 was able to decrease tissue damage caused by systemic *P. brasiliensis* infection more efficiently than the reference drug, itraconazole [Rodrigues-Vendramini et al., 2019]. Given the efficiency of CP1 for PCM treatment, another patent application was filed at INPI (BR102016021164-6) as a candidate for systemic-acting fungicide.

In the second work using CS as a target, the structure of this enzyme from the fungus *Candida albicans* (*Ca*CS) was modeled using the same protocol adopted for *Pb*CS. However, in this case, the library used was assembled from compounds with a Tanimoto index of 70% to the known CS inhibitors as described in the Binding Database [Gilson et al., 2016], plus the library of compounds from Sigma-Aldrich, all purchasable, totaling 196,099 molecules. After applying the virtual screening protocol, two molecules were selected for the *in vitro* tests, named CS8 and CaCS2. Although the *in silico* simulations were quite encouraging, CaCS2 did not show *in vitro* activity against *Candida spp*, but had excellent activity against *P. brasilienses*, with MIC and MFC values of $32\ \mu g \cdot mL^{-1}$, indicating a fungicidal profile. However, in combination with amphotericin-B, the reference drug against PCM, it showed a synergistic effect with MIC and MFC values of $4\ \mu g \cdot mL^{-1}$. The binding and inhibition studies of recombinant *Pb*CS conducted by Dr. Peter Macheroux's group (Austria) indicated K_D of $20\ \mu M$ and IC_{50} of $29\ \mu M$, the best result ever described for this enzyme up to the time of publication [Bueno et al., 2019c]. Together, these results validated CS as a potential antifungal drug target against *P. brasiliensis*.

3.1.4 Computational biochemistry in antibacterial development

Tuberculosis is an infectious disease caused by the microorganism *Mycobacterium tuberculosis* and represents a worldwide public health problem, with an estimated 10 million people infected and 1.5 million deaths per year [Abubakar et al., 2016]. According to the World Health Organization (WHO), resistance to currently used drugs is a relevant issue in the context of this pathogen [World Health Organization, 2014]. KatG is a bifunctional enzyme with both catalase and broad-spectrum peroxidase activity that oxidizes various electron donors including NADPH [Johnsson et al., 1997; Singh et al., 2008]. It protects *M. tuberculosis* against toxic reactive oxygen species (ROS) including hydrogen peroxide and organic peroxides, thus contributing to its survival within host macrophages by countering the phagocyte oxidative burst [Sherman et al., 1996; Ng et al., 2004]. Point mutations in KatG have been associated with isoniazid (INH) drug resistance [Cardoso et al., 2004]. To investigate the effects of these mutations, our research group performed a series of equilibrium molecular dynamics simulations, according to a previously described protocol, with nine point mutations of KatG detected in resistant clinical isolates in comparison to the native enzyme [Pimentel et al., 2017]. The results showed that the mutations did not cause any structural impairment in the enzyme but led to a decrease in the volume of the KatG active site that expelled INH or hindered its interaction with the iron atom of the catalytic site [Pimentel et al., 2017]. Because hydrogen peroxide has a smaller size than INH, its interaction with KatG was not affected by the decrease in the active site volume caused by the mutations, which allowed the variants unaffected by INH to continue exerting their metabolic role in the bacteria. This work illustrates the relevance of molecular dynamics simulations to describe, at the molecular level, the effects of point mutations in drug target proteins that may lead to drug resistance in pathogens.

Still, in the context of antimicrobial studies, the work published by Trevisan et al. [Trevisan et al., 2020] demonstrated the potential of bioinformatics tools in the study and inference of biochemical mechanisms involved in the inhibition of a pathogen by unknown components of a plant extract. In this work, the aqueous and ethyl acetate fractions of *Stryphnodendron adstringens* plant extracts, known as barbatimão, were evaluated as growth inhibitors of *Staphylococcus aureus* bacteria, from reference strain and foodborne isolates. *In vitro* assays demonstrated the antibacterial activity of the extract, with MIC values between 125 and $250\ \mu g \cdot mL^{-1}$, without toxic effects on Vero and HaCaT cells [Trevisan et al., 2020]. To identify the probable targets in bacterium, the 30 compounds described in the *S. adstringens* extract were submitted to the SEA web server [Keiser et al., 2007], which performs a search for inhibitors with chemical similarity to the query molecule. If there is a structural and statistical similarity between the query and the ligand of a given enzyme, there is a given probability that the query can also binds to this enzyme. The results showed that some of the compounds in the extract could bind to the enzymes FabG, FabZ, and FabI, which are involved

in essential fatty acid biosynthesis in bacteria (FAS II), as they differ from the enzymes present in the mammalian fatty acid biosynthesis system (FAS I). The three bacteria enzymes were then modeled and subjected to conventional docking with the molecules present in the extract, which allowed the inference of the compounds epigallocatechin 3-O-(3-methoxy-4-hydroxy) benzoate (CID 44257119) and epigallocatechin 3-O-gallate (CID 65064) as most likely to bind to the evaluated enzymes of the FAS II route. In addition to the contribution to the field of research on antibacterial agents applied to *S. aureus* species, this work illustrated the potential of bioinformatics tools in the study and identification of biochemical mechanisms involved in pathogen inhibition.

3.1.5 Computational biochemistry in antiviral development

Brazil is one of the largest producers of silk in the world [Giacomina et al., 2017]. Silk culture revolves around the cultivation of the silkworm (*Bombyx mori*), whose production is affected by chemical agents (insecticides and agrochemicals) and biological agents such as fungi and viruses, the latter being the *Bombyx mori* nuclear polyhedrosis virus (*Bm*NPV). This pathogen causes the disease known as grasserie, responsible for losses of approximately 20% of the worldwide silkworm crop, being able to reach 70% to 100% at the local producer level [Ribeiro et al., 2009]. Among the virus infection and dispersion mechanisms, cathepsin (V-Cath) is a protease of a broad spectrum that is essential for horizontal transmission of viral infection [Hom and Volkman, 2000]. In a work conducted by the BCM group [Bueno et al., 2019d], and in collaboration with the silk production company BRATAC S/A, the structure of cathepsin V of *Bm*NPV (*Bm*NPV-Cath) was modeled in the presence of the broad-spectrum protease inhibitor MYP and was used in virtual screening. A library of purchasable natural products from Zinc database, containing 111,500 molecules, was screened. Four compounds were selected and evaluated by molecular dynamics with $\Delta G_{binding}$ calculation using the MM-PBSA method [Genheden and Ryde, 2015]. None of the compounds showed a better $\Delta G_{binding}$ than the reference inhibitor MYP; however, the compounds Bm2 and Bm5 had the highest affinities *in silico*. In studies with *B. mori* caterpillars infected with *Bm*NPV, Bm5 proved to be the best performing compound *in vivo* with the lowest mortality rate of caterpillars and pupae. The results of this project validated *Bm*NPV-Cath as an antiviral drug target against *Bm*NPV and represented a direct biotechnological contribution to the improvement of Brazilian sericulture.

3.2 Genomics applications

3.2.1 Non-coding RNA sequence classification with deep learning

The BCM group also developed computational methods for nucleotide sequence analysis, such as non-coding RNAs (ncRNAs) classification. Specifically, the field of research encompasses short ncRNAs, which are RNA molecules of approximately 200 nucleotides that are not translated into proteins, but are crucial functional elements in prokaryotes and eukaryotes [Esteller, 2011]. One problem that arises when studying ncRNAs regards their classification into different classes. ncRNAs are experimentally classified based on sequence, structure, and function. Nonetheless, such methods are time-consuming and expensive. Therefore, computer-aided methods are required to speed up the process and help researchers prospect novel transcripts. For this reason, we developed an alignment-free method, referred to as NCYPred (**N**on-**C**oding/**Y** RNA **Pred**iction) [De Souza Lima et al., 2021] (Figure 1), using a recurrent neural network (RNN) trained to predict sequences into 13 short non-coding RNA (sncRNA) classes, extracting features directly from sequence data. To train and evaluate our model, a dataset consisting of 45,447 sncRNA sequences from Rfam 14.3 [Kalvari et al., 2021] was built, belonging to the classes: 5.8S rRNA, 5S rRNA, Box C/D (CD-box), Box H/ACA (HACA-box), Group I catalytic intron (Intron-gp-I), Group II intron (Intron-gp-II), Leader, microRNA (miRNA), Riboswitch, Ribozyme, transfer RNA (tRNA), Y RNA, Y RNA-like and stem-bulge RNA (sbRNA - nematodes and insect). Since Y RNAs and sbRNAs are functional homologs [Boria et al., 2010], we grouped both classes into a single Y RNA class.

In summary, our method consists of extracting features from an input sequence with a type of RNN, called a bidirectional long short-term memory (biLSTM) network with an attention mechanism, which assigns importance to each sequence position and encodes relevant information into a context vector, used to classify the input to its respective class. Visualization of the computed context vectors, using dimensional reduction techniques such as t-SNE, allows users to analyze and compare different transcripts. NCYPred obtained results comparable with state-of-the-art models described in the literature, while also being the first to classify Y RNAs. In comparison with alignment-based methods, such as *Nhmmer* [Wheeler and Eddy, 2013], NCYPred achieved similar results, while being less computationally expensive, and capable of correctly classifying sequences that *Nhmmer* was not able to find significant hits. After prospecting for novel ncRNA transcripts, secondary and tertiary structures can be predicted using software such as RNAFold [Langdon et al., 2018] and RNAComposer [Purzycka et al., 2015], allowing for the analysis using molecular dynamics simulations, as described earlier. NCYPred is freely available as a web server (https://www.gpea.uem.br/ncypred/), where users can easily upload sncRNA sequences in FASTA format and download the predictions from our model.

4 Conclusion

The BCM group at the State University of Maringá (UEM) is dedicated to investigating biophysical phenomena through bioinformatics tools. The main goal of the group is to discover novel drugs for human or veterinary uses. Results obtained using computational methods are also validated *in vitro* and/or *in vivo* through biochemical assays.

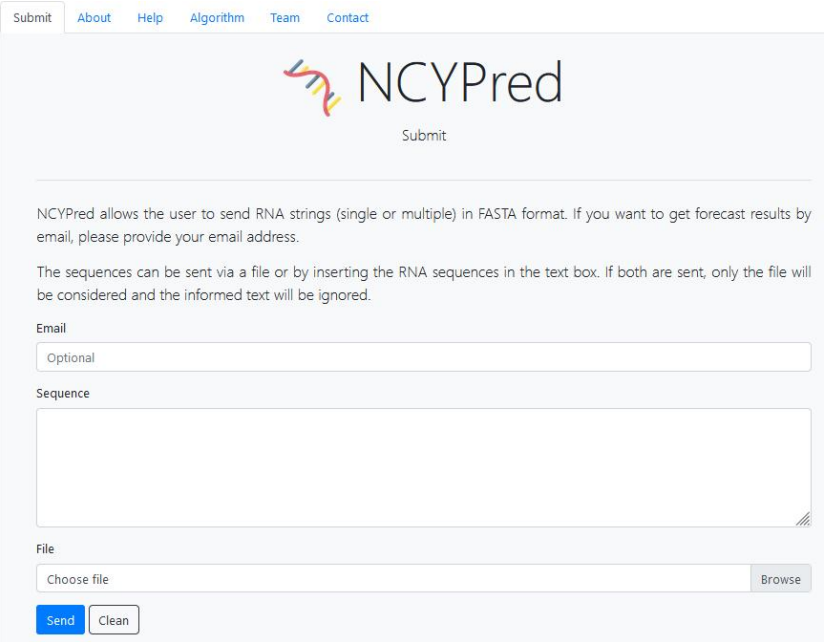

Figure 1. NCYPred web server home page at https://www.gpea.uem.br/ncypred/.

Our group has ongoing research involving protein ligands, antimicrobials, and antivirals. We are also developing new lines of research involving machine learning techniques, to further collaborate in the advancement of bioinformatics research in Brazil.

Acknowledgements

The authors would like to thank Capes, CNPq, FINEP, and Fundação Araucária for the financial support, LNCC, and CENAPAD/SP for providing computational facilities, and Fiação de Seda Bratac S/A for their support.

Competing interests

The authors declare no competing interests.

References

Seixas F. A., Silva M. R. da, Murakami M. T., Tosqui P. and Colombo M. F. The water effect on the kinetic of the bovine liver catalase. *Protein & Peptide Letters* 18 (9): 879–885, 2011.

Homem, D. P., Flores Jr., R., Tosqui, P., Rozada, T. C., Basso, E. A., Gasparotto Junior, A., and Seixas, F. A. V. Homology modelling of dihydrofolate reductase from *T. gondii* bonded to antagonists: molecular docking and molecular dynamics simulations. *Molecular BioSystems* 9 (6): 1308–1315, 2013.

Fan, J., Fu, A., and Zhang, L. Progress in molecular docking. *Quantitative Biology* 7 (2): 83–89, 2019.

Morris, G. M., Goodsell, D. S., Pique, M. E., Lindstrom, W. L., Huey, R., Forli, S., Hart, W. E., Halliday, S., Belew, R., and Olson, A. J. AutoDock Version 4.2. *User guide:* 1–66, 2012.

Trott, O., and Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31 (2): 455–461, 2010.

Thomsen, R., and Christensen, M. H. MolDock: A New Technique for High-Accuracy Molecular Docking. *Journal of Medicinal Chemistry* 49 (11): 3315–3321, 2006.

Verdonk M. L., Cole J. C., Hartshorn M. J., Murray C. W., and Taylor R. D. Improved protein-ligand docking using GOLD. *Proteins* 52 (4): 609–623, 2003.

Burley, S. K., Berman, H. M., Christie, C., Duarte, J. M., Feng, Z., Westbrook, J., Young, J., and Zardecki, C. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Science* 27 (1): 316–330, 2018.

Phillips, J. C., Braun, R.,Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26 (16): 1781–1802, 2005.

Brooks, B. R., Brooks III, C. L., Mackerell Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* 30 (10): 1545–1614, 2005.

Zoete, V., Cuendet, M. A., Grosdidier, A., and Michielin,

O. SwissParam: A fast force field generation tool for small organic molecules. *Journal of Computational Chemistry* 32 (11): 2359–2368, 2011.

Neese, F. The ORCA program system. *WIREs Computational Molecular Science* 2 (1): 73–78, 2012.

Duarte Junior, F. F., Bueno, P. S. A., Pedersen, S. L., Rando, F. S., Pattaro Júnior, J. R., Caligari, D., Ramos, A. C., Polizelli, L. G., Lima, A. F. dos S., Lima Neto, Q. A. de, Krude, T, Seixas, F. A. V., and Fernandez, M. A. Identification and characterization of stem-bulge RNAs in *Drosophila melanogaster*. *RNA Biology* 16 (3): 330–339, 2019.

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2: 19–25, 2015.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* 25 (9): 1157–1174, 2004.

Genheden, S., and Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* 10 (5): 449–461, 2015.

Kato, C. G., Gonçalves, G. de A., Peralta, R. A., Seixas, F. A. V., de Sá-Nakanishi, A. B., Bracht, L., Comar, J. F., Bracht, A., and Peralta, R. M. Inhibition of α-Amylases by Condensed and Hydrolysable Tannins: Focus on Kinetics and Hypoglycemic Actions. *Enzyme Research* 2017: 2090, 2017.

Bueno, P. S. A., Kato-Schwartz, C. G., de Souza Lima, D., Bracht, A., Peralta, R. M., and Seixas, F. A. V. In silico evaluation of condensed and hydrolysable tannins as inhibitors of pancreatic α-amylase. *Journal of Molecular Modeling* 25 (9): 275, 2019.

Kato-Schwartz, C. G., Corrêa, R. C. G., de Souza Lima, D., de Sá-Nakanishi, A. B., de Almeida Gonçalves, G., Seixas, F. A. V., Haminiuk, C. W. I., Barros, L., Ferreira, I. C. F. R., Bracht, A., and Peralta, R. M. Potential anti-diabetic properties of Merlot grape pomace extract: An *in vitro*, *in silico* and *in vivo* study of α-amylase and α-glucosidase inhibition. *Food Research International* 137: 109462, 2020.

Kato-Schwartz, C. G., de Sá-Nakanishi, A. B., Guidi, A. C., Gonçalves, G. de A., Bueno, F. G., Zani, B. P. M., de Mello, J. C. P., Bueno, P. S. A., Seixas, F. A. V., Bracht, A., and Peralta, R. M. Carbohydrate digestive enzymes are inhibited by *Poincianella pluviosa* stem bark extract: relevance on type 2 diabetes treatment. *Clinical Phytoscience* 6 (1): 31, 2020.

da Silva, T. B. V., Castilho, P. A., de Sá-Nakanishi, A. B., Seixas, F. A. V., Dias, M. I., Barros, L., Ferreira, I. C. F. R., Bracht, A., and Peralta, R. M. The inhibitory action of purple tea on in vivo starch digestion compared to other *Camellia sinensis* teas. *Food Research International* 150: 110781, 2021.

Anand, P., and Singh, B. A review on cholinesterase inhibitors for Alzheimer's disease. *Archives of Pharmacal Research* 36 (4): 375–399, 2013.

Cerbai, F., Giovannini, M. G., Melani, C., Enz, A., and Pepeu, G. N1phenethyl-norcymserine, a selective butyrylcholinesterase inhibitor, increases acetylcholine release in rat cerebral cortex: a comparison with donepezil and rivastigmine. *European Journal Pharmacology* 572 (2-3): 142–150, 2007.

Johnson, G., and Moore, S. W. Why has butyrylcholinesterase been retained? Structural and functional diversification in a duplicated gene. *Neurochemistry Internacional* 61 (5): 783–797, 2012.

Yamazaki, D. A. S., Rozada, A. M. F., Baréa, P., Reis, E. C., Basso, E. A., Sarragiotto, M. H., Seixas, F. A. V., and Gauze, G. F. Novel arylcarbamate-N-acylhydrazones derivatives as promising BuChE inhibitors: Design, synthesis, molecular modeling and biological evaluation. *Bioorganic Medicinal Chemistry* 32: 115991, 2021.

Moreira, T. F. M., Pessoa, L. G. A., Seixas, F. A. V., Ineu, R. P., Gonçalves, O. H., Leimann, F. V., and Ribeiro, R. P. Chemometric evaluation of enzymatic hydrolysis in the production of fish protein hydrolysates with acetylcholinesterase inhibitory activity. *Food Chemistry* 367: 130728, 2022.

Shikanai-Yasuda, M. A., Mendes, R. P., Colombo, A. L., Queiroz-Telles, F. de, Kono, A. S. G., Paniago, A. M. M., Nathan, A., Valle, A. C. F. do, Bagagli, E., Benard, G., Ferreira, M. S., Teixeira, M. de M., Silva-Vergara, M. L., Pereira, R. M., Cavalcante, R. de S., Hahn, R., Durlacher, R. R., Khoury, Z., Camargo, Z. P. de, Moretti, M. L., and Martinez, R. Brazilian guidelines for the clinical management of paracoccidioidomycosis. *Revista da Sociedade Brasileira de Medicina Tropical* 50 (5): 715–740, 2017.

Bueno, P. S. A., Rodrigues, F. A. V., Santos, J. L., Canduri, F., Biavatti, D. C., Pimentel, A. L., Bagatin, M. C., Kioshima, E. S., Gauze, G. de F., and Seixas, F. A. V. New inhibitors of homoserine dehydrogenase from *Paracoccidioides brasiliensis* presenting antifungal activity. *Journal of Molecular Modeling* 25 (325), 2019.

Bagatin, M. C., Rozada, A. M. F., Rodrigues, F. A. V., Bueno, P. S. A., Santos, J. L., Canduri, F., Kioshima, E. S., Seixas, F. A. V., Basso, E. A., and Gauze, G. F. New 4-methoxy-naphthalene derivatives as promisor antifungal agents for paracoccidioidomycosis treatment. *Future Microbiology* 14 (3): 235–245, 2019.

Rodrigues-Vendramini, F. A. V., Marschalk, C., Toplak, M., Macheroux, P., Bonfim-Mendonça, P. de S., Svidzinski, T. I. E., Seixas, F. A. V., and Kioshima, E. S. Promising New Antifungal Treatment Targeting Chorismate Synthase from *Paracoccidioides brasiliensis*. *Antimicrobial Agents and Chemotherapy* 63 (1), 2019.

Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research* 44 (D1): D1045–D1053, 2016.

Bueno, P. S. A., Rodrigues-Vendramini, F. A. V., Toplak, M., Macheroux, P., Kioshima, E. S., and Seixas, F. A. V. New inhibitors of chorismate synthase present antifungal activity against *Paracoccidioides brasiliensis*. *Future*

Microbiology 14 (11): 969–980, 2019.

Abubakar, I., Lipman, M., McHugh, T. D., and Fletcher, H. Uniting to end the TB epidemic: advances in disease control from prevention to better diagnosis and treatment. *BMC Medicine* 14 (47), 2016.

World Health Organization, World Health Organization Global Tuberculosis Report 2014. Geneva: World Health Organization, 2014. *http://www.who.int/tb/publications/global_report/en/*, 2018.

Johnsson, K., Froland, W. A., and Schultz, P. G. Overexpression, purification, and characterization of the catalase-peroxidase KatG from *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* 272 (5): 2834–2840, 1997.

Singh, R., Wiseman, B., Deemagarn, T., Jha, V., Switala, J., and Loewen, P. C. Comparative study of catalase-peroxidases (KatGs). *Archives of Biochemistry and Biophysics* 471 (2): 207–214, 2008.

Sherman, D. R., Mdluli, K., Hickey, M. J., Arain, T. M., Morris, S. L., Barry, C. E., and Stover, C. K. Compensatory ahpC Gene Expression in Isoniazid-Resistant *Mycobacterium tuberculosis*. *Science* 272 (5268): 1641–1643, 1996.

Ng, V. H., Cox, J. S., Sousa, A. O., MacMicking, J. D., and McKinney, J. D. Role of KatG catalase-peroxidase in mycobacterial pathogenesis: countering the phagocyte oxidative burst. *Molecular Microbiology* 52 (5): 1291–1302, 2004.

Cardoso, R. F., Cooksey, R. C., Morlock, G. P., Barco, P., Cecon, L., Forestiero, F., Leite, C. Q., Sato, D. N., Shikama, M. de L., Mamizuka, E. M., Hirata, R. D., and Hirata, M. H. Screening and characterization of mutations in isoniazid-resistant *Mycobacterium tuberculosis* isolates obtained in Brazil. *Antimicrobial Agents and Chemotherapy* 48 (9): 3373–3381, 2004.

Pimentel, A. L., de Lima Scodro, R. B., Caleffi-Ferracioli, K. R., Siqueira, V. L., Campanerut-Sá, P. A., Lopes, L. D., de Almeida, A. L., Cardoso, R. F., and Seixas, F. A. V. Mutations in catalase-peroxidase KatG from isoniazid resistant *Mycobacterium tuberculosis* clinical isolates: insights from molecular dynamics simulations. *Journal of Molecular Modeling* 23 (4): 121, 2017.

Trevisan, D. A. C., da Silva, P. V., Farias, A. B. P., Campanerut-Sá, P. A. Z., Ribeiro, T. D. V. R., Faria, D. R., de Mendonça, P. S. B., de Mello, J. C. P., Seixas, F. A. V., and Mikcha, J. M. G. Antibacterial activity of Barbatimão (*Stryphnodendron adstringens*) against *Staphylococcus aureus*: *in vitro* and *in silico* studies. *Letters in Applied Microbiology* 71 (3): 259–271, 2020.

Keiser, M. J., Roth, B. L., Armbruster, B. N., Ernsberger, P., Irwin, J. J., and Shoichet, B. K. Relating protein pharmacology by ligand chemistry. *Nature Biotechnology* 25: 197–206, 2007.

Giacomina, A. M., Garcia Jr., J. B., Zonattic, W. F., Silva-Santos, M. C., Laktima, M. C., and Baruque-Ramos, J. Brazilian silk production: economic and sustainability aspects. *Procedia Engineering* 200: 89–95, 2017.

Ribeiro, L. de F. C., Brancalhão, R. M. C., Torquato, E. F. B., and Fernandez, M. A. Susceptibility of the *Bombyx mori* cardia cells to Nucleopolyhedrovirus, multiple subgroup, BmMNPV. *Journal of Invertebrate Pathology* 100 (3): 195–198, 2009.

Hom, L. G., and Volkman, L. E. Autographa californica M nucleopolyhedrovirus chiA is required for processing of V-CATH. *Virology* 277 (1): 178–183, 2000.

Bueno, P. S. A., Biavatti, D. C., Chiarello, A. S., Fassina, V. A., Fernandez, M. A., and Seixas, F. A. V. The structure of viral cathepsin from *Bombyx mori* Nuclear Polyhedrosis Virus as a target against grasserie: docking and molecular dynamics simulations. *Journal of Biomolecular Structure and Dynamics* 37 (14): 3607–3615, 2019.

Esteller, M. Non-coding RNAs in human disease. *Nature Review Genetics* 12: 861–874, 2011.

De Souza Lima, D., Amichi, L. J. A., Constantino, A. A., Fernandez, M. A., and Seixas, F. A. V. NCYPred: A Bidirectional LSTM Network with Attention for Y RNA and short non-coding RNA classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021.

Kalvari, I., Nawrocki, E. P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., Rivas, E., Eddy, S. R., Finn, R. D., Bateman, A., and Petrov, A. I. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research* 49 (D1): D192–D200, 2021.

Boria, I., Gruber, A. R., Tanzer, A., Bernhart, S. H., Lorenz, R., Mueller, M. M., Hofacker, I. L., and Stadler, P. F. Nematode sbRNAs: homologs of vertebrate Y RNAs. *Journal of Molecular Evolution* 70 (4): 346–358, 2010.

Wheeler, T. J., and Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29 (19): 2487–2489, 2013.

Langdon, W. B., Petke, J., and Lorenz, R. Evolving Better RNAfold Structure Prediction. Lecture Notes in Computer Science, vol. 10781. Springer, pp. 220–236, 2018.

Purzycka, K. J., Popenda, M., Szachniuk, M., Antczak, M., Lukasiak, P., Blazewicz, J., and Adamiak, R. W. Automated 3D RNA Structure Prediction Using the RNAComposer Method for Riboswitches. *Methods in Enzymology* 553: 3–34, 2015.