MHOLline 2.0: WORKFLOW FOR AUTOMATIC LARGE-SCALE MODELING AND ANALYSIS OF PROTEINS

MHOLline 2.0: WORKFLOW PARA MODELAGEM E ANÁLISE AUTOMÁTICA DE PROTEÍNAS EM LARGA ESCALA

Artur Duque Rossi
Pedro Henrique Eveling Oliveira
Dimitri Guglinski Siqueira
Victor Crisóstomo Cruz Reis
Laurent Emmanuel Dardenne
Priscila Vanessa Zabala Capriles Goliatt

Abstract: Genomes and their proteins can be analyzed by different perspectives, with different goals and applicability of results in several areas of research. In bioinformatics and computational biology, it is common the use of multiple combinations of programs and databases to extract information from the raw data. Choosing the proper program, setting its parameters, filtering the input files, extracting information from output files, and creating scripts for automating tasks can be challenging. This work describes the MHOLline, an online scientific workflow that provides a set of modules that enables a broad large-scale analysis of proteins (e.g. whole genomes) in a few hours. The version 1.0 (released in 2010), was wholly reformulated, and the new version 2.0 is available at www.mholline2.lncc.br. This version presents new features, modules, interface, code optimization, runtime reduction, more security, a new results visualization interface, and an automatic and user-friendly refinement page.

Keywords: Protein sequence analysis. Scientific workflow. Bioinformatics. Computational biology.

1Mestre em Modelagem Computacional, Universidade Federal de Juiz de Fora, arturossi10@gmail.com.
2Graduando em Engenharia Computacional, Universidade Federal de Juiz de Fora, pedro.eveling@engenharia.ufjf.br.
3Graduando em Ciência da Computação, Universidade Federal de Juiz de Fora, dimitrikta@hotmail.com.
4Graduado em Ciência da Computação, Universidade Federal de Juiz de Fora, Brazil, victorreis@ice.ufjf.br.
5Doutor em Ciências Biológicas, Laboratório Nacional de Computação Científica, dardenne@lncc.br.
6Doutora em Modelagem Computacional, Universidade Federal de Juiz de Fora, priscila.capriles@ufjf.edu.br.
Resumo: Genomas e suas proteínas podem ser analizados através de diferentes perspectivas, com diferentes objetivos e com aplicação de resultados em várias áreas da pesquisa. Nas pesquisas em bioinformática e biologia computacional é comum o uso de múltiplos programas e bases de dados para se extrair informações. Escolher os programas corretos, ajustar seus parâmetros, filtrar os arquivos de entrada, extrair informações dos arquivos de saída e criar scripts para automatizar tarefas pode ser desafiador. Este trabalho descreve o MHOLLine, um workflow científico online que dispõe um conjuntos de módulos que permitem análises de grandes conjuntos de dados (ex: genomas completos) em algumas horas. A versão 1.0 (lançada em 2010) foi totalmente reformulada, e a nova versão 2.0 está disponível no link www.mholline2.lncc.br. Nesta versão são apresentadas novas funcionalidades, módulos, interface, otimização de código, redução de tempo de execução, aumento na segurança, nova interface de visualização de resultados e uma interface amigável para refinamento de resultados de forma automática.

Palavras-chave: Análise de sequência de proteínas. Workflow científico. Bioinformática. Biologia Computacional.
1 INTRODUCTION

Genome analysis may provide information about each protein’s role in organisms, which proteins play similar or analogous biological functions, and identify proteins intrinsic to a particular specie. It also enables the search for new molecular targets for the treatment of diseases and assists in the structure-based drug design (SBDD).

The blind screening method used in drug development is responsible for creating the majority of drugs available. It consists of testing a library of compounds against in vivo and in vitro systems. The costs to purify, characterize, and synthesize these compounds associated with the need to have many compounds in a library are hard-working and highly expensive, spending billions of dollars (ADAMS; BRATNER, 2006; KOLA; LANDIS, 2004; DIMASI; GRABOWSKI; HANSEN, 2016).

Therefore, computational methods have been proposed to decrease costs and speed up the drug design process (e.g., the SBDD) (LOUNNAS et al., 2013). The more is known about the three-dimensional (3D) structure of proteins, the faster the drug discovery process becomes (CONGREVE; MURRAY; BLUNDELL, 2005; MANDAL; MOUDGIL; MANDAL, 2009). In this context, the MHOlline workflow was created to automate (i) the analysis of structural properties of proteins, (ii) the construction of 3D models of proteins aiming to achieve a large number of protein models in an easy way and in less time, accelerating the SBDD studies, and (iii) the validation of 3D models (CAPRILES et al., 2010).

Over the past ten years, MHOlline has contributed to dozens of researches in the different areas of Computational Modeling (JAISWAL et al., 2017; JAMAL et al., 2017; HASSAN et al., 2018; VASCONCELOS; CAMPOS; REZENDE, 2018; ALMEIDA et al., 2019; GONÇALVES et al., 2019). In this paper, we present the second version of this workflow, with the addition of new softwares, proprietary tools for protein modeling analysis and refinement, and administration environment.

MHOlline is a scientific workflow developed to aid researchers from Bioinformatics, Biophysics, Computational Chemistry, and Computational Biology. In 2010, the partnership between Universidade Federal do Rio de Janeiro (UFRJ) and Laboratório Nacional de Computação Científica (LNCC) resulted in the first web version of MHOlline, available in the electronic address www.mholline.lncc.br (CAPRILES et al., 2010; GUIMARÃES et al., 2013).

This workflow has been defined as a computing environment divided into three parts (Figure 1): (i) MHOllcore, that generates, processes and maintains all data; (ii) MHOllweb, the graphical user interface that displays the progress of jobs and their results; and (iii) MHOlldb,
the database used to store and query all the system data. MHOLline was developed for Linux OS. MHOLcore uses Perl, ShellScript, C, and Python2/3, MHOLdb database uses MySQL, and MHOLweb uses PHP, JavaScript, HTML, and CSS.

![Figure 1: The structure of MHOLline 2.0.](image)

There are two primary access types: with or without login. The user without login can submit new jobs, but it is limited to 50 amino acid sequences per file in FASTA format. The user with a login could be classified into three types of logged user: (i) Registered, can submit and manage its jobs without the restriction of protein sequences per file; (ii) Manager, has some administrative privileges over Registered user to help the system management; (iii) Administrator, has all privileges to maintenance of server, database, statistical analysis of MHOLline usage and query base tools that keep MHOLline updated and working

1.1 What’s New in MHOLline 2.0

The first version of MHOLline (CAPRILES et al., 2010) presented the following modules: (i) the BLAST program (ALTSCHUL et al., 1990), used to perform the local alignment of the submitted sequences against the Protein Data Bank (PDB) (BERMAN et al., 2000); (ii) the BATS tool, developed to classify the alignments in four groups (Table 1) based the E-value and identity from BLAST, and Length Variation Index (LVI) which is the MHOLline’s concept of coverage (LVI ≤ 0.1 is equivalent to coverage ≥ 90%); (iii) the FILTERS tool, created to sort G2 proteins (proteins that can be modeled by comparative modeling technique) into seven quality groups (Table 2); (iv) the ECNGet tool, developed to capture at least one Enzyme Commission (EC) number for each sequence in G2 group, whose reference protein has at least one known enzymatic function; (v) the MODELLER software (WEBB; SALI, 2016) that constructs the 3D models; (vi) the PROCHECK software (LASKOWSKI et al., 1993), used to produce Ramachandran plots; and (vii) the HMMTOP software (TUSNÁDY; SIMON, 2001), used to identify transmembrane regions in proteins.

In this new version, MHOLline has undergone the following modifications: (i) updated modules to latest versions; (ii) addition of new softwares; (iii) code refactoring to make
Table 1: The relation between G0-G3 groups with their requirements.

| Group | Requirement                                      |
|-------|--------------------------------------------------|
| G0    | Not aligned sequence                             |
| G1    | E-value > 10e-5 OR Identity < 0.15               |
| G2    | E-value ≤ 10e-5 AND Identity ≥ 0.25 AND LVI* ≤ 0.7|
| G3    | E-value ≤ 10e-5 AND (0.15 ≤ Identity < 0.25 OR LVI* > 0.7) |

*LVI: Length Variation Index (CAPRILES et al., 2010).

Table 2: Criteria used to classify the model’s quality.

| Quality           | Identity | LVI* |
|-------------------|----------|------|
| 1.Very High       | ≤ 75%    | ≤ 0.1|
| 2.High            | ≤ 50% AND <75% | ≤ 0.1|
| 3.Good            | ≤ 50%    | > 0.1 AND ≤ 0.3|
| 4.Medium to Good  | ≤ 35% AND <50% | ≤ 0.3|
| 5.Medium to Low   | ≤ 25% AND <35% | ≤ 0.3|
| 6.Low             | ≤ 25%    | > 0.3 AND ≤ 0.5|
| 7.Very Low        | ≤ 25%    | > 0.5 AND ≤ 0.7|

*LVI: Length Variation Index (CAPRILES et al., 2010).

workflow more readable and faster; (iv) security improvements; (v) creation of an online view mode where user can analyze the job results; (vi) development of a refinement system where the logged user can edit the alignment and resubmit it to the workflow also using some structural restrictions in the 3D model construction; (vii) development of a re-upload tool for data reuse, allowing data re-analysis and refinement without having to pass through the workflow processing again, saving the user time; and (viii) creation of a new interface, taking into account the layout and operational compatibility with main browsers, according to the concept of human-computer interaction.

In the system administration area, we incorporated a tool to user and system activity data collection, which are used to generate graphs, tables, and reports, providing the administrator with information to design future modifications and keep records of workflow activities. All this data is saved into the database, being possible to retrieve information from previous periods and compare it to the current scenario. Additionally, we developed functions to backup, restore, and update modules and database that runs directly from the system interface via the administrator environment. These services facilitated remote system administration and maintenance without access via a remote network server.

Revista Mundi, Engenharia e Gestão, Paranaguá, PR, v. 5, n. 6, p. 283-01, 283-14, 2020
DOI: 10.21575/25254782rmetag2020vol5n61325
2 RESULTS AND DISCUSSION

MHOLline received new softwares such as (i) MOLPROBITY (CHEN et al., 2010) to validate the stereochemistry of 3D protein models; (ii) SIGNALP (PETERSEN et al., 2011) to identify signal peptide regions in protein sequences; (iii) TMHMM (KROGH et al., 2001) to identify transmembrane helices; (iv) PSIPRED (MCGUFFIN; BRYSON; JONES, 2000) to predict the secondary structure of proteins; and (v) DSSP (KABSCH; SANDER, 1983) to interpret the secondary structure of modeled proteins. Figure 2 presents the flowchart of the new MHOLline workflow.

![Figure 2: The work process flowchart of MHOLline 2.0.](image)

Additionally, logged users can improve or reconstruct 3D models and download the original MHOLline’s results or the last two modified results through the refinement process. Allowed improvements are: (i) change the template chosen by default or add up to two more structures; (ii) cleavage of the protein in positions that were predicted by the SignalP program; (iii) use of secondary structures constraints during model construction via MODELLER; and (iv) optimize loops with MODELLER. It is important to note that when the user changes or add new templates if only one structure is selected, the workflow executes the salign function from MODELLER; otherwise, executes CLUSTALΩ software (SIEVERS et al., 2011).

The logged user can also download the alignment, perform manual adjustments in this

---

Revista Mundi, Engenharia e Gestão. Paranaguá, PR, v. 5, n. 6, p. 283-01, 283-14, 2020. DOI: 10.21575/25254782rmetag2020vol5n61325
file, and then re-upload this data to be used in the refinement tool. The interface provides information regarding the protein model and templates (Figure 3), such as the secondary structure of the sequence model predicted with PSIPRED and the template files assigned by DSSP and transmembrane regions found in the sequence model using TMHMM.

![Figure 3: Sample of information provided in the alignment section of the refinement interface. In the figure, the characters over the sequences means (c) coil, (h) alpha-helix, and (e) beta structure.](image)

During the re-submission, the logged user might choose either Manual or Automatic refinement. The Automatic mode will select the secondary structures, based on PSIPRED prediction, which occurs in template gap regions, using different combinations of size and position, in an attempt to increase the variety of 3D models. This process may generate up to 50 models. If the number of possibilities exceeds 50 combinations, the system will randomly choose 50 and then model the selected ones. If there are no gaps nor alpha/beta secondary structure prediction in gapped regions, the system will optimize loops with MODELLER generating 20 models.

Aiming to minimize the requirements to analyze a large amount of data generated in the Automatic refinement, we implemented a library written in python3 called Quality-Model Clustering (QMC) Tool to cluster proteins in groups based on its energy and stereochemical qualities (CARVALHO; ROSSI; GOLIATT, 2020). This tool takes into account the MODELLER scores (molpdf, DOPE, DOPE-HR, GA341, and N-DOPE), and Molprobity results to group the structures by similarity, returning to the logged user the clusters with their representative structures. The QMC tool is freely available to download in the git repository.
(github.com/ruanmedina/Quality-Model-Clustering).

The Manual refinement will perform the re-modeling based on user inputs (Figure 4), generating 20 models. As in the Automatic process, the input data should also consider modifications in the alignment, secondary structure constraints, and template change. In this case, the MHOL.line 2.0 refinement tool will not run the protein cluster library (QMC) since the user chooses the refinement.

Figure 4: Sample of information provided in the alignment section from the refinement interface. In the figure, the characters over the sequences means (c) coil, (h) alpha-helix, and (e) beta structure.

To validate both the Manual and Automatic refinement processes, we performed three tests: the Automatic refinement, the Manual refinement, and without any restriction. The protein used in this validation process was the ATP-diphosphohydrolase 2 (Uniprot code: A1BXT9) of Schistosoma mansoni, also called SmNTPDase2 (LEVANO-GARCIA et al., 2007). This sequence was submitted to the MHOL.line 2.0 workflow, which selected as the template the structure of nucleoside triphosphate diphosphohydrolase 1 (NTPDase1) of Rattus norvegicus (PDB code: 3ZX3) (ZEBISCH et al., 2012), presenting 33% (68/204) of identity and LVI of 0.57. To evaluate the Automatic and the Manual refinements within gap regions, we changed the template by the NTPDase1 of Legionella pneumophila subsp. pneumophila (strain Philadelphia 1 / ATCC 33152 / DSM 7513) (PDB code: 4BRP - chain A) (ZEBISCH et al., 2013), with 28% of identity and LVI of 0.53. The objective was to obtain alignment with lower coverage and identity to be able to apply structural restrictions. The system chose the
restrictions applied in the automatic approach, and Table 3 presents the constraints used in the manual procedure. In both tests, we decided to cleave the region detected as a signal peptide, and not use contact and beta-sheet restrictions, nor loop optimizations.

Table 3: Sample of restrictions applied to MODELLER in the refinement area of the logged user in MHOLine 2.0.

| Start residue | End residue | Start residue | End residue |
|---------------|-------------|---------------|-------------|
| 97            | 106         | 36            | 45          |
| 109           | 112         | 50            | 58          |
| 255           | 270         | 182           | 183         |
| 433           | 448         | 242           | 251         |
| 470           | 478         | 456           | 458         |
| 465           | 466         |               |             |

The QMC Tool filtered each set (Automatic, Manual, and No Restriction), performing a Principal Component Analysis (PCA) over MODELLER and Molprobity attributes. The dimensionality of the data is reduced from seven to two components, enough to explain more than 85% of the original data (86.8% of Automatic, 87.8% of Manual, and 92.0% of No Restriction). We selected the best model of each set by the visual inspection of each cluster’s representative model. Figure 5 presents the final three best models (one from each set) aligned with the control structure, the 3D model of SmNTPDase2, previously published by (SOUZA et al., 2014). Table 4 presents the backbone Root Mean Square Deviation (RMSD) calculated using the RMSD plugin of PyMol (Schrödinger, LLC, 2015).

As observed in Table 4 the Manual refinement achieved a smaller RMSD value, followed by the Automatic approach indicating that the use of any refinement tool might lead to better results than not applying structural improvements. Figure 5 shows that Automatic refinement (Fig. 6b), even not being as good as the Manual (Fig. 6c), has been capable of optimizing the protein structure.

Table 4: RMSD values of the backbone alignment, SmNTPDase2, Manual, Automatic, and No Restriction proteins.

| Molecule            | RMS deviations |
|---------------------|----------------|
| SmNTPDase2          | 0.0            |
| Manual Refinement   | 13.525         |
| Automatic Refinement| 14.147         |
| No Restriction      | 15.315         |
The DOPE profile has been computed for each structure (models, template, and reference) and plotted in Figure 6 to illustrate the differences between the DOPE energy for each residue. It is possible to notice that in regions where restraints were applied, the models’ energy profile was similar to the SmNTPDase2, for example, in the alpha-helix constraints between residues 97 - 106 and 109 - 112, and beta-strand constraints between residues 182 - 183 and 242 - 251. The C-terminal coils present different energy values due to their high flexibility, allowing the structure to assume different conformations in this region. This problem can be minimized by using contact restrictions.

A single FASTA file submission with a large number of sequences (e.g., entire genomes), can spend a lot of processing time and generates large files as a result. For this reason, the jobs are automatically deleted 15 days after finishing its processing in the main workflow, saving hard disk space. In case that users need more than 15 days to process their results, a data-reuse tool was created, and logged users can re-upload their original .tar.gz file (previously downloaded) to the system (SIQUEIRA, 2018).

The user re-upload the previous job file, and MHOLline checks the encrypted key to verify data integrity. Then the user can automatically compare if the re-uploaded data are the same as they were before the deletion and check if the re-uploaded files represent the same data of finished jobs in the system (whether the user has not previously deleted the job). Finally, the
tool inserts files in its original folders, restores the data in MHOLdb, and sets again as active. Thus the user will have more 15 days to continue the analyzes.

3 CONCLUSIONS

The new version of MHOLline features both system improvements and analysis tools. We refactored the system, adding features and standardizing the code and documentation and removing vulnerabilities. Unified Modeling Language (UML) diagrams have been generated to help the knowledge of each step of the workflow, from MHOLcore to MHOLweb. The web interface of MHOLline has been completely rewritten, and a new layout for the result page was developed to make the user experience even better. The main difference of new MHOLweb is the refinement page in which the user can manually or automatically modify and optimize the 3D model constructed by MHOLline.

The option to change or add more templates proved to be a significant improvement in the workflow. Sometimes, the best BLAST hit may not be the best modeling template due to its original function or organism. Now, users have the option to change the default template to best suit their needs. Additionally, the ability to indicate to MODELLER where to set secondary
structure constraints based on PSIPRED predictions and user knowledge is also a significant improvement since it gives more user autonomy to model the protein as desired.

The objective of creating new tools and enhancing existing ones in the workflow was to ensure the correct functioning of the process, renewing and innovating based on current programming practices, project management approaches, computational modeling, and information security. The latest version of MHOLLline can be accessed at www.mholline2.lncc.br.

4 PERSPECTIVES

As the next steps, we are planning to implement the second phase of MHOLLline code optimization. The modules will have their docker containers, allowing administrator users to have more control over all processes running in the MHOLLline server. Other computational biology softwares will be implemented to expand the capability of MHOLLline 2.0 to provide information about proteins, such as residue-residue contact prediction, which can also be used to improve the refinement tool.

ACKNOWLEDGMENTS

We would like to thank National Scientific Computing Laboratory and the Federal University of Juiz de Fora for making this work possible. FAPEMIG supported this work under grant no 11651/2017.

REFERENCES

ADAMS, C. P.; BRATNER, W. Estimating the cost of new drug development: is it really $802 million? Health Tracking, v. 25, n. 2, p. 420–428, 2006.

ALMEIDA, J. M. et al. Biochemical characterization and application of a new lipase and its cognate foldase obtained from a metagenomic library derived from fat-contaminated soil. International Journal of Biological Macromolecules, v. 135, n. 2, p. 442–454, September 2019.

ALTSCHUL, S. F. et al. Basic local alignment search tool. Journal of Molecular Biology, v. 215, p. 403–410, 1990.

BERMAN, H. M. et al. The protein data bank. Nucleic Acids Research, v. 28, p. 235–242, 2000.

CAPRILES, P. V. S. Z. et al. Structural modelling and comparative analysis of homologous, analogous and specific proteins from trypanosoma cruzi versus homo sapiens: putative drug targets for chagas disease treatment. BMC Genomics, v. 11, p. 610, 2010.
CARVALHO, R. M.; ROSSI, A. D.; GOLIATT, P. V. Z. C. Quality-model clustering tool: A module for clustering protein models based on quality attributes. *Revista Mundi, Engenharia e Gestão*, v. 5, n. 2, p. 219–01, 219–15, 2020.

CHEN, V. B. *et al.* Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica*, D66, p. 12–21, 2010.

CONGREVE, M.; MURRAY, C. W.; BLUNDELL, T. L. Structural biology and drug discovery. *Drug Discovery Today*, v. 10, n. 13, p. 895–907, 2005.

DIMASI, J. A.; GRABOWSKI, H. G.; HANSEN, R. W. Innovation in the pharmaceutical industry: New estimates of r&d costs. *Journal of Health Economics*, v. 47, p. 20–33, 2016.

GONÇALVES, L. M. *et al.* Trehalose synthesis inhibitor: A molecular in silico drug design. *J Cell Biochem*, p. 1–12, 2019.

GUIMARÃES, A. C. R. *et al.* Chapter XV: High-Throughput Genome Analysis for Structure-based Rational Drug Design: Comparative Genome Analysis and Protein Modeling. *Genomics II - Bacteria, Viruses and Metabolic Pathways*. [S.l.]: Hodder Arnold, 2013. ISBN: 978-1-480254-14-5.

HASSAN, S. S. *et al.* The druggable pocketome of corynebacterium diphtheriae: A new approach for in silico putative druggable targets. *Front Genet.*, v. 9, p. 44, 2018.

JAISWAL, A. K. *et al.* An in silico identification of common putative vaccine candidates against treponema pallidum: A reverse vaccinology and subtractive genomics based approach. *Int J Mol Sci.*, v. 18, n. 2, p. 402, 2017.

JAMAL, S. B. *et al.* An integrative in-silico approach for therapeutic target identification in the human pathogen corynebacterium diphtheriae. *PLoS ONE*, v. 12, n. 10, p. e0186401, 2017.

KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, v. 22, p. 2577–2637, 1983.

KOLA, I.; LANDIS, J. Can the pharmaceutical industry reduce attrition rates? *Nature Reviews*, v. 3, p. 711–715, 2004.

KROGH, A. *et al.* Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, v. 305, n. 3, p. 567–580, 2001.

LASKOWSKI, R. A. *et al.* Procheck - a program to check the stereochemical quality of protein structures. *J. App. Cryst*, v. 26, p. 283–291, 1993.

LEVANO-GARCIA, J. *et al.* Characterization of schistosoma mansoni atpdase2 gene, a novel apyrase family member. *Biochem. Biophys. Res. Commun.*, v. 352, p. 384–389, 2007.

LOUNNAS, V. *et al.* Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Computational and Structural Biotechnology Journal*, v. 5, n. 6, p. 1–14, 2013.
MANDAL, S.; MOUDGIL, M.; MANDAL, S. Rational drug design. European Journal of Pharmacology, v. 625, p. 90–100, 2009.

MCGUFFIN, L. J.; BRYSON, K.; JONES, D. T. The psipred protein structure prediction server. Bioinformatics, v. 4, p. 401–405, 2000.

PETESEN, T. N. et al. Signalp 4.0: discriminating signal peptides from transmembrane regions. Nature Methods, v. 8, p. 785–786, 2011.

Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. PyMOL - The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC. 2015.

SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. Molecular Systems Biology, v. 7, p. 539, 2011.

SIQUEIRA, D. G. Desenvolvimento de sistema de reuso de dados para o workflow MHOLine. Juiz de Fora, Brazil: [s.n.], 2018. Undergraduate Thesis.

SOUZA, V. et al. Structural comparative analysis of secreted ntpdase models of Schistosoma mansoni and Homo sapiens. Springer International Publishing, p. 91–98, 2014.

TUSNÁDY, G. E.; SIMON, I. The hmmmtp transmembrane topology prediction server. Bioinformatics, v. 17, p. 849–850, 2001.

VASCONCELOS, C. R. S.; CAMPOS, T. d. L.; REZENDE, A. M. Building protein-protein interaction networks for leishmania species through protein structural information. BMC Bioinformatics, v. 6;19, n. 1, p. 85, 2018.

WEBB, B.; SALI, A. Comparative protein structure modeling using modeller. Current Protocols in Bioinformatics, v. 54, n. 1, p. 5.6.1–5.6.37, 2016.

ZEBISCH, M. et al. Crystallographic evidence for a domain motion in rat nucleoside triphosphate diphosphohydrolase (ntpdase) 1. J Mol Biol., v. 13;415, n. 2, p. 288–306, 2012.

ZEBISCH, M. et al. Crystallographic snapshots along the reaction pathway of nucleoside triphosphate diphosphohydrolases. Structure, v. 6;21, n. 8, p. 5.6.1–5.6.37, 2013.