“CBDynamics”: a Mathematical Modeling Study about the Synthetic Production of Cannabidiol.

Team leaders: Araujo, G.N., Redígolo, J.L.S.

Team members: Contelli, A.P., Oliveira, B.H., Nicanor, B.M., Fenner, C.K., Valdejão, G.G.G., Santos, G.C.F., Pinheiro, L.F., Oliveira, M.B., Andraus. M.Q., Lopes, V.T., Pereira, W.A.R.A.

PIs: Lacerda, T.M., Segato, F.*

*e-mail: symb-lab@usp.br

ABSTRACT

Cannabidiol (CBD) is a valuable medical compound known for its numerous therapeutic benefits, particularly in the treatment of patients with chronic diseases such as epilepsy. However, in Brazil, its acquisition is shrouded in bureaucratic and legal complexities, primarily stemming from its association with *Cannabis sativa.* Therefore, this study aims to produce cannabidiolic acid (CBDA), the direct precursor to CBD, using biosynthesis and recombinant gene technology with *Saccharomyces cerevisiae* yeast. This production involves two essential pathways: the geranyl diphosphate (GPP) one, already present in the yeast metabolism, and the olivetolic acid rote. Consequently, both precursors will be combined in a second reaction to form cannabigerolic acid, which will then be converted into cannabidiolic acid, which can be purified and converted into CBD. Finally, this innovative alternative method seeks to provide a more affordable and less bureaucratic way for patients to obtain CBD, thereby significantly enhancing patient access to this valuable therapeutic compound.

Keywords: Canabidiol, *Cannabis sativa*, olivetolic acid, therapeutic, chronic diseases.

1 INTRODUCTION

Cannabis sativa has a long history of medicinal use, dating back to its cultivation in 4000 BC in China (Li, 1974). In India, associated with religiosity, it was used both medicinally and recreationally (Touw, 1981). In the 19th century, it was introduced into Western medicine, but its study declined in the 20th century due to increased recreational use and legal restrictions (Kalant, 2001; Bonfá *et al.*, 2008).

Contributing to its medicinal history, the compound Cannabidiol, which comes from the *Cannabis sativa* plant, has been widely used to treat various diseases. It acts on the central nervous system and is mainly used to treat epilepsy, anxiety, Parkinson's disease, Alzheimer's disease and chronic pain (Alves, 2020).

The demand for cannabidiol-based medicines has been increasing in Brazil, due to the search for alternatives to aggressive and sometimes ineffective drugs. However, the scientific use of Cannabis in Brazil is stigmatized due to legislation that classifies marijuana as an illicit drug and prohibits its use. Brazilian Law No. 11,343/2006 criminalizes the use, cultivation, and trafficking of drugs, including Cannabis, but this law also provides exceptions to unlawfulness when it comes to the use of some illicit drugs for medicinal and scientific purposes. The regulation of the medicinal and scientific use of Cannabis in

Brazil is overseen by the National Health Surveillance Agency (ANVISA), but it is a bureaucratic, time-consuming process with a low approval rate. (ANVISA) (Marinho; Neves, 2022).

Therefore, efforts towards pharmacological medical research on Cannabis presupposes a shift in the substance's status, making it worthy of scientific consideration and breaking free from the cultural prejudices. This reevaluation, linked to its promising potential in treating neurological diseases, aligns with rules set by government health institutions at both national and supranational levels. (Apolinário *et al.,* 2022).

To transform Cannabis therapy more available and accessible, thus making its use more democratic, is to align with the values of CBDynamics. Added to project specific objectives of carrying out structural modeling, kinetic modeling and performance of experimental tests of transformation and expression of the metabolic pathways of biosynthesis of the compound of interest, it prioritizes meeting the planned interests.

2 MATERIALS AND METHODS

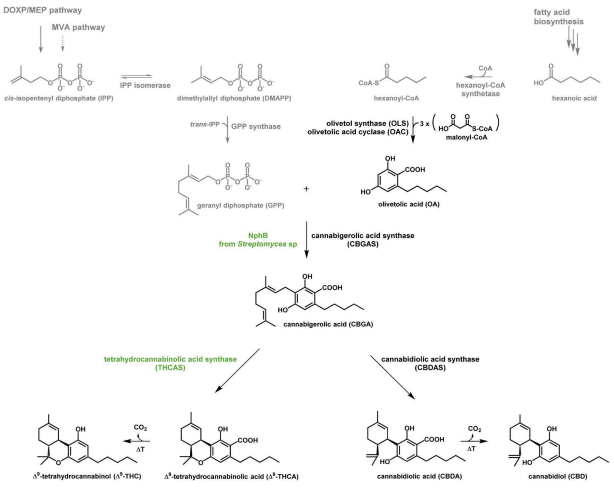
2.1 Structural modeling and construction of biological circuits

2.1.1 Obtaining coding sequences for cannabidiolate pathway genes

The synthesis of olivetolic acid begins with the activation of hexanoic acid by Hexanoyl-CoA synthase, an acyltransferase that attaches the Coenzyme-A group to the carboxylic end of hexanoic acid, preparing it for subsequent reactions in the metabolic pathway. In the second step, the enzyme olivetol synthase, a transferase, catalyzes the formation of olivetol, utilizing the hexanoyl-CoA generated in the previous stage and malonyl-CoA as substrates. In the third step, the resulting olivetol-CoA is converted to olivetolic acid by the action of the enzyme olivetolic acid synthase, which cyclizes olivetol-CoA, releasing the CoA group and producing olivetolic acid.

The fourth step in the total synthetic metabolic pathway of cannabidiol (CBDA) involves the prenylation of the olivetolic acid produced in the previous stage. This reaction, catalyzed by the enzyme cannabigerolic acid synthase in plants of the Cannabis genus, utilizes olivetolic acid and geranyl diphosphate (GPP) as substrates, resulting in the production of cannabigerolic acid, with the release of a diphosphate group. The final step in CBDA biosynthesis involves the conversion of cannabigerolic acid to CBDA, catalyzed by an oxidoreductase. The CBDA synthase, responsible for this stage in Cannabis plants, utilizes cannabigerolic acid as a substrate, generating CBDA and hydrogen peroxide as products. The resulting CBDA can be purified from transformed and cultivated cells and subsequently converted to the final product, cannabidiol, through thermal decarboxylation.

Figure 1: Metabolic pathway for CBD and THC.



Source: 2017, Zirpel *et al.*

In order to facilitate the transformation of microorganisms for the production of CBDA, the chassis to be used will be the yeast *Saccharomyces cerevisiae* SC9721, genotype MATα his3-Δ200 URA3-52 leu 2Δ1 lys2Δ202 tryp 1Δ63, which already has the GPP pathway naturally, requiring only the insertion of the olivetolic acid pathway and the final synthesis pathway of CBDA. For this construction, it will be necessary to induce the synthesis of five enzymes encoded by five genes that must be cloned into expression vectors suitable for the chosen chassis. The choice of *S. cerevisiae* as the chassis is due to the fact that it is a yeast that has been well characterized in the literature and is easy to control in cultivation. Furthermore, as CBDA is a highly complex compound, a prokaryotic organism would not have the necessary machinery for expression.

2.1.3 T2A Sequences

Aiming to ensure that the five enzymes are synthesized in the biological circuit and have the expected function, a possible route of action was found: the use of 2A sequences. They are short, high-efficiency sequences for the synthesis of multiple individual proteins from a single mRNA molecule.

The technique consists of exploiting the inability of the eukaryotic ribosome to insert a peptide bond between the last proline and glycine residues, but without detaching itself from the mRNA molecule. This failure leads to a jump between amino acid residues without stopping protein synthesis, which generates multiple individual proteins from a single mRNA. The T2A and P2A sequences are the most efficient in this function and their sequences can be easily obtained from bioinformatics banks. Despite the ease of working with 2A sequences, the functions of the resulting proteins must be evaluated, since in the N- and C-terminal regions of the separated proteins there will be fragments corresponding to the cleavage of the 2A sequence used (Mansouri, 2014).

In this project, it was needed to evaluate 3 different positions in each enzyme to implement this technique. Using the following T2A sequence: **EGRGSLLTCGDVEENPG P**

Consequently, there were three possible situations for our enzymes: 1. the protein gets the (EGRGSLLTCGDVEENPG) sequence on its C-terminal; 2. he initial (P) on its N-terminal and the final (EGRGSLLTCGDVEENPG) on its C-terminal or; 3. only the single proline in the N-terminal for each protein expressed.

For that reason, new models were built, based on the first and evaluated ones, that could express the new size and conformation of those enzymes, observing the interferences at the molecular docking, the main point of our metabolic pathway.

Evaluating all of the models obtained with the proposed modifications (15 total models), the subsequent key points were considered: how good the model of the native protein is and its size, both of which directly imply the correspondence of protein functionality interference. Also, those evaluation methods were repeated to obtain the best possible protein models.

2.1.4 Biological circuits

The biological circuits for the genes of interest used can be assembled in any expression vector that has a replication origin for *E. coli* and yeast, as well as a strong inducible promoter for yeast. Using *Saccharomyces cerevisiae* as a

chassis, the olivetolic acid and CBDA synthesis pathways require the insertion of its synthesis pathway by means of biological circuits compatible with the expression system.

To ensure that the five enzymes are synthesized with the expected function in the circuit, two possibilities were tested: the use of fusion proteins and the use of 2A sequences. The activity of the possible proteins was assessed by structural and kinetic modeling in order to find the proteins with the best efficiency in producing CBDA

The genes for CBDA synthesis were cloned into two separate biological circuits, so that the resulting inserts and expression vectors are of a reasonable size and, when translated, are more efficient at converting substrates and producing CBDA.

Using the fusion proteins, pairs of constructs containing two or three genes, in different combinations, were evaluated for the biosynthesis of two fusion proteins. The genes were fused into an insert, spaced with flexible linkers of 10 amino acids to maintain the activity and structure of the individual proteins. Each insert was cloned into pRS vectors.

For the use of 2A sequences, two inserts were constructed, of two and three genes, spaced by the T2a sequence (GAGGGCAGGCAGTCTGCTGACATGCGGTGACGTGGAAGAGAATCCCGGCCCT). The genes were combined until they formed and generated two inserts of approximately the same size.

SWISS-MODEL and AlphaFold2 servers were used for structural modeling of the native proteins with T2A appendages and fusion proteins. The three-dimensional structures and energy parameters of the proteins were evaluated using SWISS-MODEL, the PyMOL software, and the QMean server.

Structural and kinetic modeling of the proteins separated using 2A sequences or fusion proteins allowed the construction of expression cassettes already suitable for the chassis, which will be cloned into the pRS425 and pRS426 vectors using the Gibson Assembly method. The biological circuits were constructed using the cassettes developed previously, with Gal1 promoters, efficient ADH1 terminators for the chassis, the insert developed with genes intercalated by 2A sequences, and selection markers. The expression cassettes were cloned into the pRS vectors mentioned beforehand. Several biological circuits were tested before coming to the conclusion of which ones would be best for the project. In this case, the plasmids pRS423, pRS424, pRS425, pRS426, pYES2, and pET26b were used for testing.

These vectors required expression cassettes to be made, consisting of a promoter - CDS - terminator. The following sequences were therefore tested for the assembly of each cassette:

Table 1: Sequences tested for the cassettes's assembly

Promoter	CDS			Terminator
GAL1	CSHCS	OS	OAC	AHD1
GAL1	CSHCS	OS	CsaPT4	AHD1
GAL1	CBDAS	OS	OAC	AHD1
GAL1	CBDAS	OS	CsaPT4	AHD1
GAL1	PT4	CBDAS	-	AHD1
GAL1	PT4	OAC	-	AHD1

Source: 2023, Authors.

The pYES2 vectors and plasmids for *E. coli* do not need cassettes, since they already possess a promoter-terminator pair. Finally, the assembled plasmids were visualized using the

SnapGene Viewer software and, after obtaining these plasmids, it was possible to build a total of 72 different circuits with different combinations between the enzymes' coding sequences. With this, all of them were evaluated and optimized and two biological circuits were chosen, using the pRS425 and the pRS426 plasmid.

2.1.5 Structural modeling of enzymes

Structural modeling of enzymes was carried out in order to review the function of enzymes and their properties. Starting with a bibliographical review of the enzyme and its function, moving on to its kinetic properties, FASTA sequence , evaluation of signal peptide, and structural parameters and FASTA sequence: CsHCS1 (Genbank: AFD33345.1), OLS (GenBank: BAG14339.1), OAC (Uniprot: I6WU39, CsPT4 (UniProt: A0A455ZJC3), CsBDAS (NCBI: NP_001384865.1). The first native protein models were generated via homology models, checking possibilities, improved stability and angles of amino acids that would appear to be less stable. Finally, it was verified whether the models could be used correctly.

After the gathering of the FASTA sequences, the existence of signal peptides was evaluated for protein expression using the SignalP 6.0 website (Teufel *et al.*, 2022). Signal was found in CSBDAS and transit peptide in CsPT4. The sequences were then submitted to the Swiss Model server (Waterhouse *et al.*, 2018), and the models were chosen based on the highest identity and query possible, and we prioritized the more accurate templates, ideally obtained from previously crystallographic studies. Then, the Ramachandran plots (Hollingsworth; Karplus, 2010) and the MolProbity (Williams *et al.*, 2018) parameters were also evaluated for the first model validation. Additionally, According to the NCBI description, there are glycosylations in only one of our enzymes: the cannabidiolic acid synthase (CBDAS), with seven glycosylation sites predicted at the amino acid residues 45, 65, 168, 296, 304, 328 and 498 (Onofri, De Meijer, Mandolino, 2015).

Based on the files generated by SwissModel, evaluations were carried out using other softwares, such as ProCheck from UCLA-DOE LAB SAVESv6.0 (Lakowski *et al*, 1993), to visualize the results in summarized and graphical form. Energy refinement was also tested using the Charmm-Gui server (Jo *et al*., 2008; Park *et al.,* 2023), using ProChe'ck to verify QMEAN (Benkert, Biasini, Schwede; 2011) and evaluating the generated Ramachandran plots, that represents the psi (ψ) and phi (φ) bond angles of the amino acids present in a peptide. It is used to determine "permitted" (i.e. energetically and geometrically favorable) areas for backbone or spine dihedral points of the amino acid (Coumar, 2021). In other hands, QMEAN stands for Qualitative Model Energy Analysis. It describes the major geometrical aspects of protein structures using five different structural descriptors. The local geometry is analyzed by a new kind of torsion angle potential over three consecutive amino acids. Finally, the PDB file was inserted at the MolProbity (Williams *et al.*, 2018) and hydrogenated.

2.2 Molecular docking

2.2.1 Docking

In this study, molecular docking was employed to identify optimal sites in enzymes for

substrate positioning with the aim of optimizing the system. Various software tools, such as ClusPro and HDOCK, were utilized to analyze the enzymes in their natural state and after the addition of a T2A gene sequence for result comparison. ClusPro served as a web server for protein-protein docking and HDOCK employed a hybrid algorithm for protein-protein and protein-RNA/DNA docking. After evaluation, HDOCK was chosen as the primary docking tool. The initial analysis involved molecular docking between enzymes and substrates in the metabolic pathway. Hexanoyl-CoA synthase, identified as an acyltransferase, was found to bind the Coenzyme-A group to the carboxylic end of a hexanoic acid molecule, playing a crucial role in the metabolic pathway. A practical approximation approach was adopted for the analysis of the chemical cavity of Hexanoyl-CoA synthase, given its impracticality to scrutinize each atomic bond, due to the transmembrane nature of the enzyme.

The subsequent step involved the analysis of molecular enzyme-substrate docking for proteins containing the T2A gene sequence, yielding optimal positions consistent with the analysis of the protein in its natural state. CavityPlus, a software designed for detecting cavities and potential binding sites in proteins' 3D models ("CavityPlus," n.d.), was employed to analyze the five enzymes. Subsequently, the molecular docking plugin for Pymol facilitated a more in-depth examination of these binding sites (Seeliger & De Groot, 2010). Each enzyme generated approximately 10 models for each ligand, featuring a range of 6 to 15 cavities, subsequently combined and compared in the context of phylogenetic analyses.

2.2.2 Phylogenetic analysis

Simultaneously, a protein blast was conducted by the team using NCBI to retrieve FASTA sequences for similar proteins corresponding to each enzyme. These sequences were subsequently compared using the T-COFFEE MUSCLE alignment site to assess their alignment. For a comprehensive understanding of the essential regions in these alignments, the software Jalview was employed. With the sequences' alignment visualized on Jalview, extensive information from the literature regarding the binding sites, active sites, and catalytic pockets of each enzyme was gathered and analyzed.

2.3 Gene Expression Modeling

In order to represent gene expression and metabolic pathway action in biological circuits, mixed-effects models were used to describe individual responses in a population, adopting the simplified dynamic case of gene expression described by Llamosi *et al.* (2016). This gene expression model considers transcriptional and translational parameters, and through experimental analysis, they determine constants for mRNA and protein production.

After describing the profile considered for gene expression, Boolean logic based on biomolecules, as shown in Miyamoto *et al.* (2013), was applied to model logic gates in the metabolic pathway. Substrates like galactose and hexanoic acid are considered, with excess substrates deemed non-determinants. For this analysis 2 situations were evaluated, the first considering the transformation of both plasmids on a cell from the yeast population, and the second each circuit individually.

3 RESULTS

3.1 Overview

The development of our modeling approach for this project was far from trivial. The majority of our enzymes remained poorly characterized, with some lacking a crystallized structure in databases. Each enzyme presented a unique challenge— either being too small, dimeric, glycosylated, or embedded in membranes. This necessitated a thorough evaluation of proteins from both structural and dynamic perspectives. The process involved identifying the most closely related crystal structure and predicting optimal protein stability in vials. Understanding the nature of our enzymes and devising appropriate approaches became imperative throughout this comprehensive endeavor.

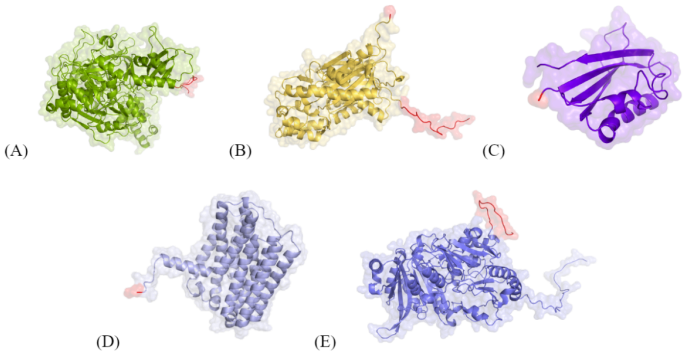
Although the project presented in this competition revolves around computational development, it is crucial to note that the wet lab evaluation is set for the next steps, contingent upon the acquisition of approvals related to cannabidiol production. In this case, the computational phase concluded in 2023, revealing promising results regarding the functionality and applicability of simulated biological circuits. Since the release requests were recently submitted, the project is prepared and ready for the beginning of the wet lab phase.

3.2 Structural Modeling

After the native protein models were prepared, attempts to integrate T2A sequences using SWISS-MODEL consistently reproduced the original proteins. Subsequent investigation uncovered the automatic exclusion of T2A sequences.

Challenges arose with our modifications, resulting in identical models, due to SWISS-MODEL's homology approach . Transitioning to the MODELLER extension in the PyMOL Molecular Graphics System software, a new homology, or more precisely, comparative modeling, was constructed based on the native-protein model as a template. The structures below were visualized in a surface-cartoon view, with the red portion in each representing the T2A sequence, where respectively A, B, C, D and E represents CsHCS1, OS, OAC, CsPT4 and CBDAS T2A-modified.

Figure 2: Models from our protein with T2A sequence.

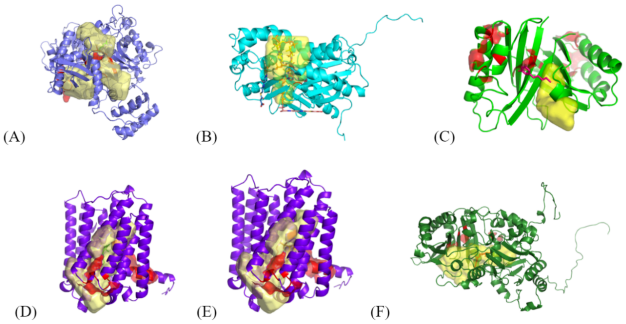


Source:2023, Authors.

3.3 Molecular Docking

As seen above, firstly, we docked at HDOCK server each enzyme with its substrates described by the biological path. However, intended to compare these sites docked and the probabilistic catalytic pockets from these proteins, we made a phylogenetic study of each enzyme studied. Using the alignment between at least 5 similar-proteins from each search, we could obtain the site-information and, consequently, compare with both our previous docking results and cavity analysis, as we can see at figure below.

Figure 3: Proteins with both phylogenetic and cavity analysis docked with their respectives substrates from HDOCK server.



Source: 2023, Authors.

All results present a yellow-translucent-surface as their best cavities obtained from CavityPlus, while the red parts are the result obtained from phylogenetic analysis. Respectively, “A” indicates CsHCS with hexanoic acid; “B” the OS with both hexanoyl co-a and 3 malonyl co-a; “C” which represents the docking between OAC and olivetol; both “D” and “E” showing the CsPT4 with olivetolic acid and geranyl diphosphate and, finally, the “F” one showing the CBDAS docked with CBGA.

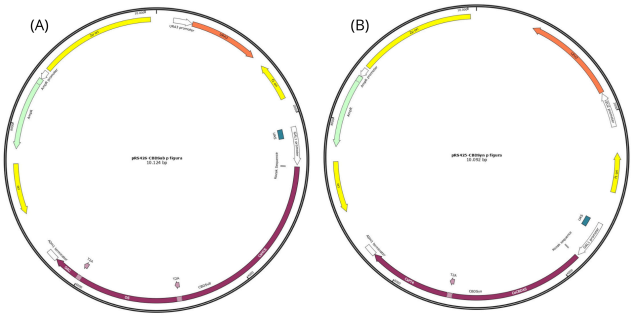
Evaluating these results, we had a direct correlation of what phylogenetics indicated as the main amino acids of the catalytic site and the cavities found. Almost all images represented also related the substrate over these regions. However, the docking between OAC and hexanoic acid showed a certain difference, which implied a second approach: redock. Using the information from the OAC crystal and olivetol and comparing this docking with that performed, we identified a strong correlation, thus validating this position.

3.4 Biological Circuits

After all the cassettes had been built and the sequence, characteristics and genetic constructions of the assembled plasmids had been analyzed, the best candidates were chosen .

It was possible to build a total of 72 different circuits with different combinations of enzyme coding sequences. With this, it was feasible to evaluate all of them, optimize and choose two biological circuits, using the plasmids pRS425 and pRS426 as the final choice, as shown in the following figures.

Figure 4: pRS426 biological circuit - CBDSub and pRS425 biological circuit - CBDSyn, respectively.



Source: 2023, Authors.

3.5 Gene Expression Modeling

The first logic circuit involves transforming both plasmids in a yeast cell population as seen on Figure 5. The only added substrate, aside from galactose, is hexanoic acid ("Hexanoic acid A" in the circuit). Excess cofactors like GPP, malonyl-CoA, and Mg2+ ions are assumed in the yeast intracellular medium.

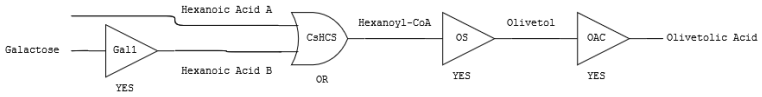
Figure 5: Complete Boolean-based logic circuit for CBDA metabolic pathway.

Source: 2023, Authors (made with VisualParadigm).

The second logic circuit examined the individual circuits. Two yeast populations, "*S. cerevisiae* A" transformed with pRS426-CBDSub

and "*S. cerevisiae* B" transformed with pRS425-CBDSyn, are considered. For the first of these same discussed substrates are added to the first population and the goal is to extract Olivetolic Acid (OA) after successive reactions, as illustrated in Figure 6.

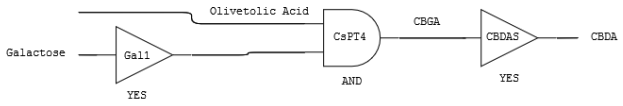
Figure 6: Boolean-based logic circuit for Olivetolic Acid metabolic pathway.



Source: 2023, Authors (made with VisualParadigm).

On the other hand, cells containing the pRS425-CBDSyn plasmid exhibit a distinction. Instead of introducing hexanoic acid as one of the substrates, the input substrate becomes the Olivetolic Acid obtained from the preceding population, as depicted in the subsequent figure.

Figure 7: Boolean-based logic circuit from Olivetolic Acid to CBDA pathway.



Source: 2023, Authors (made with VisualParadigm).

5 HUMAN CENTERED DESIGN

The success of the project will depend on the collaboration between all stakeholders - Health, Government, Industry, Academia and Local Community - with a firm commitment to ethics, safety and accessibility. In order to offer the use of CBD, an innovative alternative, since it has a favorable and very tolerable safety profile. Since the team's mission is related to offering the hope of access to high-quality medicines at more affordable prices, ensuring patients, who suffer limitations and side effects associated with the various traditional pharmaceutical products, a better quality of life, aiming at maintaining the comfort and well-being of the individual.

4 CONCLUSION

The results obtained during the research carried out in the project revealed promising applications in the field of synthetic biology. An innovative approach was employed to manipulate cannabidiol expression, resulting in relevant advances in understanding of protein control and modeling mechanisms, also providing a good basis for future wet-lab research steps.

All results that were found through research provide valuable information about its promising practical applications. The ability to build the metabolic pathway for CBD synthesis, study of biological circuits and proteins for its expression guarantee promising perspectives in areas such as the pharmaceutical industry and medicine, through the appearance of alternative therapies and treatment of numerous diseases.

The project represents a significant step towards substantial contributions to the scientific community and society as a whole by promoting advances in several fields of application.

6 REFERENCES

A. Šali and T. L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815, 1993.

Alves, Francisco Eduardo Ferreira. A utilização medicinal do canabidiol como recurso terapêutico: revisão bibliográfica. *Revista Interfaces: Saúde, Humanas e Tecnologia*, v. 8, n. 2, p. 581-590, 2020.

Apolinário, J. A. F. *et al.* O uso medicinal de cannabis: tabus morais, conflitos éticos e legais. *Ciência ET Praxis,* v. 15, n. 29, p. 58–72, 28 ago. 2022.

Biological circuits designed with SnapGene software (www.snapgene.com)

BLAST: Basic Local Alignment Search Tool. (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi

Coumar, M. S. (2021). Molecular Docking for Computer-Aided Drug Design. *Academic Press.*

G.C.P van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries and A.M.J.J. Bonvin (2016). "The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes." *J. Mol. Biol.*, 428, 720-725 (2015).

Guex, N., Peitsch, M.C., Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 30, S162-S173 (2009).

Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol Concepts*. 2010 Oct;1(3-4):271-283. doi: 10.1515/BMC.2010.022. PMID: 21436958; PMCID: PMC3061398.

HADDOCK Web Server. (n.d.). Retrieved from https://wenmr.science.uu.nl/haddock2.4/

Kalant, Harold. Medicinal use of cannabis: history and current status. *Pain Research and Managemen*t, v. 6, n. 2, p. 80-91, 2001.

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., . . . Vajda, S. (2017). The ClusPro web server for protein–protein docking. *Nature Protocols*, 12(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst., 26, 283-291.

Li, Hui-Lin. An archaeological and historical account of Cannabis in China. *Economic botany*, v. 28, n. 4, p. 437-448, 1974.

M., Versari, C., Cinquemani, E., Ferrari-Trecate, G., Hersen, P., & Batt, G. (2016). What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast. *PLoS computational biology*, 12(2), e1004706.

Mansouri, M.; Berger, P. Strategies for multigene expression in eukaryotic cells. *PlasmidAcademic Press Inc.*, , 2014.

Marinho, C. A. G.; Neves, I. F. Regulamentação do uso medicinal e científico da cannabis no Brasil. *Revista Ibero-Americana de Humanidades, Ciências e Educação,* v. 8, n. 11, p. 1264–1283, 30 nov. 2022.

Miyamoto, T., Razavi, S., DeRose, R., & Inoue, T. (2013). Synthesizing biomolecule-based Boolean logic gates. *ACS synthetic biology,* 2(2), 72-82.

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. *Journal of Molecular Biology*, 302(1), 205–217. https://doi.org/10.1006/jmbi.2000.4042

Onofri, C., de Meijer, E. P., & Mandolino, G. (2015). Sequence heterogeneity of cannabidiolic-and tetrahydrocannabinolic acid-synthase in *Cannabis sativa* L. and its relationship with chemical phenotype. *Phytochemistry,* 116, 57-68.

R.V. Honorato, P.I. Koukos, B. Jimenez-Garcia, A. Tsaregorodtsev, M. Verlato, A. Giachetti, A. Rosato and A.M.J.J. Bonvin (2021). "Structural biology in the clouds: The WeNMR-EOSC Ecosystem." *Frontiers Mol. Biosci.*, 8, fmolb.2021.729513.

Seeliger, D., & De Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of Computer-aided Molecular Design,* 24(5), 417–422. https://doi.org/10.1007/s10822-010-9352-6

S. Jo, T. Kim, V.G. Iyer, and W. Im (2008). CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* 29:1859-1865.

S.J. Park, N.R. Kern, T. Brown, J. Lee, and W. Im (2023). CHARMM-GUI PDB Manipulator: Various PDB Structural Modifications for Biomolecular Modeling and Simulation (2023) *J. Mol. Biol.* 167995.

Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, G., Haas, J., Schwede, T. QMEANDisCo - distance constraints applied on model quality estimation. *Bioinformatics* 36, 1765-1771 (2020).

Teufel, F., Almagro Armenteros, J.J., Johansen, A.R. *et al*. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40, 1023–1025 (2022). https://doi.org/10.1038/s41587-021-01156-3

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Touw, Mia. The religious and medicinal uses of *Cannabis* in China, India and Tibet. *Journal of psychoactive drugs*, v. 13, n. 1, p. 23-34, 1981.

Visual Paradigm. © 1999-2023 by Visual Paradigm. All rights reserved.

Wang, S., Xie, J., Pei, J., & Lai, L. (2023). CavityPlus 2022 Update: An Integrated Platform for Comprehensive Protein Cavity Detection and Property Analyses with User-friendly Tools and Cavity Databases. Journal of Molecular Biology, 435(14), 168141. https://doi.org/10.1016/j.jmb.2023.168141

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009 May 1;25(9):1189-91. doi: 10.1093/bioinformatics/btp033.

Yan, Y., Zhang, D., Zhou, P., Li, B., & Huang, S. (2017). HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research,* 45(W1), W365–W373.

Zirpel, B. et al. Engineering yeasts as platform organisms for cannabinoid biosynthesis. *Journal of Biotechnology*, v. 259, p. 204–212, 10 out. 2017.