PROTEIN ENGINEERING AND BIOTECHNOLOGY: TOWARDS SUSTAINABLE PRODUCTION OF RARE SUGARS

Prado, C. D[1]*, Fernandes, C. C[1], Lemos, E G D M[1]

[1] *Department of Agricultural and Environmental Biotechnology, Laboratory of Microorganism and Plant Biochemistry, São Paulo State University, Faculty of Agricultural and Veterinary Sciences - UNESP-FCAV, Jaboticabal, São Paulo, Brazil.*
** Corresponding author's email address: prado.unespfcav@gmail.com*

ABSTRACT

The diversity of enzymatic activities in plant biomass offers potential for biotechnological applications, driving interest in green chemistry and biomass biorefinery for high-value products. The production of rare sugars like tagatose and psicose is particularly attractive due to their sweetening and pharmacological properties, despite being scarce in nature due to complex and costly synthesis. A study involving the sequencing of a bacterial consortium isolated from sugarcane bagasse identified a partial genome of the genus *Chitinophaga* sp. (CB10), with 456 ORFs (Open Reading Frame) related to lignocellulosic biomass processing. A database of carbohydrases was developed from this sequencing data aiming to use these enzymes in rare sugar biorefinery from biomass. *In silico* studies of candidate enzymes for rare sugar biosynthesis were conducted, followed by cloning and expression of heterologous proteins. D-psicose 3-epimerase (DPEase) was selected as a potential enzyme for biomass degradation towards rare sugar production. Expression and purification of DPEase showed promising results after prospecting, cloning, and protein induction steps, indicating potential for efficient biomass conversion into rare sugars. These findings highlight the biotechnological feasibility of these strategies for sustainable production of high-value compounds from biomass.

Keywords: Enzymes, Biorefinery, Biotechnology, Rare Sugars

INTRODUCTION

The plant biomass offers valuable metabolites such as sugars and phenolic compounds, driving the production of second-generation ethanol (2G) and the establishment of biomass biorefineries. The metabolism of lignocellulosic biomass requires the action of various enzymes, made possible by biochemical cooperation among bacterial communities. The Laboratory of Microbiology and Plants (LBMP) at FCAV - UNESP has been researching alternative routes for rare sugar production, using studies in microbiology and plants to obtain industrial molecules like enzymes and enzyme cocktails for catalyzing fermentative processes. Based on metagenomic data, the laboratory successfully identified biotechnologically relevant enzymes for rare sugar production using recombinant protein engineering. A bacterial consortium from soil and dry sugarcane straw demonstrated lignocellulolytic potential. Metagenomic analyses revealed species diversity and abundance, with modifications in the appearance and composition of sugarcane fiber. The work of LBMP is crucial for exploring alternative routes for rare sugar production and understanding bacterial community interactions in biomass metabolism, contributing to advances in sustainable bioproducts and efficient fermentative processes. From an academic and biotechnological standpoint, this work employed bioinformatics tools to prospect an enzyme of interest for rare sugar production, aiming to produce it through recombinant protein engineering for application in enzyme cocktails. Preliminary results showed excellent protein expression in the soluble fraction after 6 hours of fermentation. Furthermore, all stages of prospecting, *in silico* analyses, recombinant protein engineering, expression, and protein extraction were validated, emphasizing the importance of these tools in industrial enzymology.

1 MATERIAL & METHODS

Prospection and *In Silico* Analyses

The enzyme prospecting process involved a detailed analysis of re-annotated metagenomic data from the bacterial consortium associated with lignocellulosic decomposition, stored on NCBI (National Center for Biotechnology Information) and analyzed using the bioinformatics tool K-Base. Each bacterial species' genomic sequence was analyzed in K-Base using Prokka functionality to re-annotate the genomes. Enzymes identified underwent thorough evaluations of their DNA sequences and translated proteins, utilizing the BLAST Protein tool provided by NCBI with Swiss-Prot UniProt and Pataa functions. Following a detailed search for enzymes in the LBMP bacterial consortium metagenomic data and careful selection, primers were designed to target specific gene regions responsible for enzyme synthesis. Enzymes like D-psicose-3-epimerase were identified in the bacterial consortium sequence. Primer sets were developed initially to amplify these enzyme-encoding regions, and comprehensive *in silico* analyses were conducted to ensure primer accuracy and functionality prior to synthesis, simulating the genetic engineering process for obtaining recombinant proteins. Additionally, the Signal P6 and Phobius tools were employed to analyze amino acid sequences for signal peptides, aiming to identify their presence or absence and providing the base pair count for peptide removal from the target gene nucleotide sequence during primer design, particularly at the sequence ends if a signal peptide was present.

D-psicose-3-epimerase

For these enzymes, we performed *in silico* analysis of primer quality based on nucleotide sequence, following the same procedure described earlier. However, in designing the primers, we considered the use of cloning genetic engineering for the PET-SUMO

(Thermo Scientific)) expression vector. It is important to highlight that this vector does not have restriction sites; therefore, restriction enzymes were not used in primer design. The Forward and Reverse primers were composed of a specific number of nucleotides only, as determined by analysis of their respective sequences using Primer Stats software, followed by alignment of primer sequences with the target enzyme sequences in Bioedit software.

Fusion Protein

To finalize primer construction, a computational simulation replicates their use under real experimental conditions, such as *in silico* application in laboratory procedures like PCR and cloning for heterologous expression. The process began by downloading nucleotide sequences of the PET-SUMO vector. Using SnapGene software, fusion protein analysis determined the fusion protein sequence's exact position, restriction sites, and other crucial elements. Virtual assembly of the PCR product fragment was then performed, using DNA from the protein obtained through metagenomic sequencing of the LBMP bacterial consortium. This simulated the complete fragment construction, evaluating its viability and efficacy, and simulating cloning and insertion into the PET-SUMO plasmid. The software generated a detailed plasmid design with the cloned fragment and provided tools for deeper analysis. This step is crucial for detecting potential frame misalignments. Subsequently, the transcript sequence obtained from SnapGene was submitted to NCBI's Blast Protein tool to confirm its relationship with the target protein, and physical-chemical parameters related to the transcript were obtained using Protparam software.

Standardization of PCR Reaction, amplification and purification of PCR fragment

Upon receipt of the synthesized primers from Sigma®, they were diluted to a stock concentration of 100 pmol and aliquots of 50 µL were prepared, resulting in a final concentration of 10 pmol, in a sterilized environment. To assess primer efficiency, extracted DNA (bacterial consortium) was used for amplification via PCR (Reaction: 95 °C (2 min) + [95 °C (45 sec), Tm°C (45 sec), and 72 °C (3 min)] x 30 cycles + 72 °C (10 min)), using the synthesized oligonucleotides. PCR reactions were performed in a total volume of 25 µL, containing 10X Buffer (2 µL); dNTPs (0.6 µL); Synthesized Forward and Reverse Oligonucleotides (0.5 µL each); Taq polymerase (1.6 µL); RNase-free water (11.16 µL); and 3.0 µL of DNA (50 ng/µL). The integrity of the obtained amplicons was evaluated on a 1% agarose gel with 1X TBE buffer. Samples were prepared with 1 µL of Dye Loading + 5 µL of sample. To confirm the fragment size, the gel included 1 µL of molecular weight marker (1Kb Plus Invitrogen) + 5 µL of Dye Loading. Samples were subjected to horizontal electrophoresis for approximately one hour at a constant voltage of 90 V. The image was visualized using the Dual LED Blue/White Light Transilluminator – KASVI agarose gel visualization system and subsequently documented using a BioRad photodocumenter. The quality of the extracted genetic material was assessed on a NanoDrop ND-1000 spectrophotometer (Thermo Scientific) at wavelengths of 260 nm and 280 nm, with the 260/280 ratio used to estimate DNA purity. The fragments were purified from agarose gel using the Zymoclean™ Gel DNA Recovery kit (Zymo Research), following the manufacturer's instructions. The image was visualized using the KASVI Dual LED Blue/White Light Transilluminator system and subsequently documented with a BioRad photodocumenter. The quality of the extracted DNA was assessed using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific) at wavelengths of 260 nm and 280 nm, with the 260/280 ratio used to determine the genetic material purity.

Cloning and Expression of Heterologous Proteins: Amplification, Purification, and Ligation of Target Fragment + PET-SUMO Plasmid

For cloning using the PET-SUMO vector, PCR was performed using a Taq polymerase recommended by the expression vector manufacturer's protocol, followed by purification of the fragment. The PCR reaction was carried out in a total volume of 50 µL, containing Taq Platinum Invitrogen (0.5 µL); 10X Buffer (5 µL); MgCl2 (1.5 µL); dNTPs (0.5 µL); Synthesized Forward and Reverse Oligonucleotides (2 µL each); RNase-free water (38 µL); and 0.5 µL of DNA (50 ng/µL). The extracted DNA (bacterial consortium) was used for amplification via PCR (Reaction: 95 °C (2 min) + [95 °C (45 sec), Tm°C (45 sec), and 68 °C (3 min)] x 30 cycles + 68 °C (30 min)), using the synthesized oligonucleotides. Following these steps, the integrity of the fragments was assessed on a 1.5% agarose gel, where samples (25 µL of sample + 5 µL of dye) and marker (5 µL of marker + 1 µL of dye) were applied to the gel and subjected to electrophoresis at 90 V for 60 minutes. Subsequently, the fragments were visualized, and the target fragment was excised and transferred to a 1.5 mL Eppendorf tube. Purification was initiated using the Zymoclean™ Gel DNA Recovery kit (Zymo Research), following the manufacturer's protocol. The quality of the extracted DNA was assessed using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific) at wavelengths of 260 nm and 280 nm, with the 260/280 ratio used to determine the genetic material purity. Subsequently, the ligation reaction of the fragment + PET-SUMO vector was performed using 3 µL of purified DNA following the protocol established by the expression vector manufacturer.

Transformation and Selection of Clones

In this stage, competent *E. coli* ER2266 cells from the LBMP competent cell bank were used, along with competent *E. coli* BL21 cells from the PET-SUMO expression vector kit. The PUC19 vector included in the plasmid kit was also used as a positive control for this step. For the transformation, 2 µL of the ligation reaction was pipetted directly into tubes containing the competent *E. coli* ER2266 and *E. coli* BL21 cells, gently mixed without pipette agitation. Subsequently, a positive control was prepared by adding circular PUC19 vector to new tubes of *E. coli* ER2266 and *E. coli* BL21. The tubes were incubated on ice for 5 to 30 minutes followed by a heat shock at 42°C for 30 seconds in a dry bath without agitation, then immediately transferred back to ice. An aliquot of 250 µL of SOC medium (Super Optimal Broth with Catabolite repression) at room temperature was added to the tubes

and incubated in a shaker at 37°C and 200 rpm for 1 hour. Afterwards, 100-200 µL of each transformation was pipetted onto petri dishes containing LB agar medium (10.0 g/L Tryptone; 10.0 g/L NaCl; 5.0 g/L Yeast Extract) supplemented with Kanamycin (20 µg/mL). Two different cell volumes were used at this stage to ensure spaced/isolated colonies. The plates were then incubated at 37°C for approximately 12 hours. The remaining ligation reaction was used for a 1.5% agarose gel electrophoresis to observe the presence of the fragment ligated to the plasmid. After incubation, randomly selected colonies (10 colonies) were subjected to a colony PCR protocol to confirm the transformation and cloning of the recombinant protein. Half of a colony was removed from the plate with a sterilized toothpick and transferred to a 1.5 mL Eppendorf tube containing 15 µL of colony PCR lysis buffer, and then incubated at 99°C for 10 minutes in a dry bath to lyse the cells. After cell lysis, 3 µL of the lysate was added to a PCR tube containing other reaction components (Taq Polymerase (1.6 µL); 10X Buffer (2 µL); MgCl2 (0.64 µL); dNTPs (0.60 µL); Synthesized Forward and Reverse Oligonucleotides (0.5 µL each); RNase-free water (11.16 µL)) for a final volume of 20 µL. The extracted DNA was used for amplification via PCR (Reaction: 95 °C (2 min) + [95 °C (45 sec), Tm°C (45 sec), and 72 °C (3 min)] x 30 cycles + 72 °C (30 min)), using the synthesized oligonucleotides.To check for the presence of positive clones, the samples were analyzed on a 1% (w/v) agarose gel. Colonies that showed the expected fragment size corresponding to the predicted protein from *in silico* analyses were used to prepare glycerol stocks (30% (v/v)) and stored at -80°C.

Sanger Sequencing for Positive Clones

The recombinant plasmids were extracted and purified using cells from the growth of positive clones in LB medium (10.0 g/L Tryptone; 10.0 g/L NaCl; 5.0 g/L Yeast Extract) supplemented with Kanamycin (20 µg/mL) (5 mL LB + 5 µL Kan) at 37°C overnight, 225 rpm. The extraction and purification protocol used was the Wizard Plus SV Minipreps DNA Purification System following the manufacturer's instructions. The purified samples were quantified, and an aliquot was used for agarose gel electrophoresis analysis and also for a PCR reaction with the primers used in the Sanger sequencing step to observe specific sizes of the heterologous protein in both frames. Subsequently, the samples were submitted for Sanger sequencing following the protocols and methodologies established by the research group.

Expression of Recombinant DPEase (D-psicose-3-epimerase)

Recombinant *E. coli* ER2265 cells for DPEase expression were cultivated with agitation at 200 rpm in a 500 ml flask containing 200 ml of LB culture medium (10.0 g/L Tryptone; 10.0 g/L NaCl; 5.0 g/L Yeast Extract) at 37°C with 20 µg/ml Kanamycin. When the optical density at 600 nm of the bacterial culture reached 0.4 ~ 0.6, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM to induce enzyme expression. The culture was incubated at 16°C with agitation at 150 rpm for 24 hours to express the enzyme. The expression level of recombinant proteins was evaluated using SDS-PAGE. Cell lysates were analyzed by SDS-PAGE, using a discontinuous buffer system[1]. The polypeptide bands were visualized on a 12.5% gel using Coomassie Brilliant Blue R-250 staining. The identity of recombinant D-psicose-3-epimerase protein was analyzed by Western blotting using an anti-polyhistidine monoclonal antibody (Sigma, St. Louis, MO). The polypeptides were electro-transferred from the gel to a PVDF membrane (Bio-Rad). The reaction with the anti-polyhistidine monoclonal antibody (Sigma) was performed at a 1:1000 dilution to detect the recombinant protein produced in *E. coli* ER2566. The immunoenzymatic peroxidase conjugate w;2ith goat anti-mouse IgG antibody (Bethyl Laboratories Inc. – Montomery, TX, USA) was used at a 1:500 dilution in PBS-LPD 5% for revealing the recombinant protein-antibody complex. After 1 hour of incubation, the membrane was washed with PBST and PBS, and the color development reaction proceeded after adding the chromogen-substrate mixture (DAB; Sigma)[2].

Homology analysis of epimerase sequences

The amino acid sequence of the enzyme under study was used to build the three-dimensional model of the protein using the Swiss-Model tool. Subsequently, the structure was subjected to comparative analysis with sequences of other sugar phosphate epimerases using the ESPript 3.0 tool.

2 .RESULTS & DISCUSSION

Prospection, *in silico* analyses and primer design

From the LBMP bacterial consortium metagenomic data, 52 bacterial sequences were identified and analyzed to find enzymes related to rare sugar production. These sequences were reannotated in the K-base database, revealing proteins with biotechnological potential. Target enzymes were located within protein contigs using the Prokka tool. An Excel spreadsheet organized relevant information about each enzyme. To broaden enzyme prospecting, protein sequences from various genomes were explored beyond the LBMP consortium, including genomes from soil bacterial isolates like *Burkholderia* sp. D15. Identified enzymes underwent detailed evaluation of nucleotide and translated protein sequences using the BLAST Protein tool, focusing on non-redundant and patent-free sequences. These refined analyses directed the study towards promising enzyme sequences. fter prospecting and selecting enzymes of interest, the D-psicose 3-epimerase enzyme was chosen for study, and an analysis was conducted to determine the possible presence of a signal peptide in the protein sequence. The protein sequence was replicated in the SignalP 6 and Phobius programs (Figure 1 a e b) and subjected to detailed analysis. The predictive analysis performed on these platforms did not identify any potential signal peptides in the amino acid sequence, indicating that this sequence does not possess transmembrane domains in the protein. The primer design phase was conducted using the Primer Stats software, as described in the methodology. Each primer sequence was entered into the designated field and subjected to

analysis. The outcome of this assessment generated a detailed report (Figure 1 c e d) with the characteristics of the constructed primer sequences. Selection of the most suitable primer was based on the efficiency parameters obtained, ensuring the choice of the best option for subsequent experiments. The results for the design of Forward and Reverse primers for the gene encoding D-psicose-3-epimerase are shown in the figure below. Certain parameters such as the amount of C and G in the sequences are relevant, as these bases not only contribute to strength but also increase the melting temperature of the primers. Another important parameter is hairpin formation and self-annealing of the primer. As observed in the figure, these parameters are described as "pass," indicating that the primer design met these criteria. The ideal amount of C and G in the sequence should be between 40% and 60%, provided that the temperature difference between the primers does not exceed 5°C. These conditions can influence PCR reactions and the obtaining of the target fragment; however, it is not always possible to achieve a perfect design with all parameters 100% adjusted. In an *in silico* quality analysis, the sequences may not always have significant amounts of C and G to meet parameter adjustments, so primer design ends up being a combination of various factors, including the target gene sequence, expression vector, and restriction enzyme sequence, among others. Therefore, the primers were designed for nucleotide sequences that best suited the project's objectives, respecting the quality of this *in silico* analysis. Thus, the sequences of the oligonucleotides Forward 5´ATGAAACACGGCATCTATTATTCTTACT3´ and Reverse 3´TCAGCCGCCCAAGGACGAAG5´ were obtained for amplifying the region encoding D-psicose-3-epimerase. The annealing temperatures were calculated using the online Thermo calculator tool from Thermofisher.

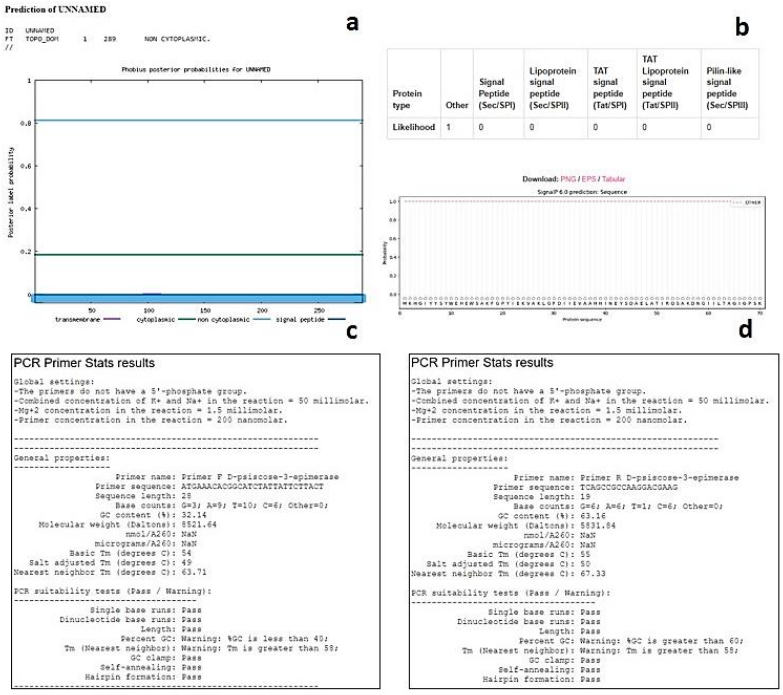

Figure 1 Results (a e b) of Protein Sequence Analysis of D-psicose 3-epimerase Enzyme using the SignalP 6.0 Online Software from the Department of Health and Technology (DTU, Denmark). Report obtained through the PCR Primer Test tool for the design of primers for the gene encoding D-psicose-3-epimerase, specifically primer Forward (c) and primer Reverse (d).

Fusion Protein, Standardization of PCR Reaction, amplification and purification of PCR fragment

Using SnapGene software, it was possible to construct, simulate the cloning of the targeted sequences, and evaluate *in silico* the entire gene construction and expression vector, resulting in the recombinant protein. The selected expression vector sequences and the genes encoding the proteins to be cloned were utilized in this process. Data from this analysis, including the translated protein sequence, were subjected to BLAST against curated databases to assess the similarity of the recombinant protein with available sequences. Another critical parameter was verifying the correct frame of the target gene post-cloning. Below in Figure 2, an example of the *in silico* recombination and engineering for D-psicose-3-epimerase protein inserted into the PET-SUMO expression vector is depicted. To conclude, the online software ProtParam, accessible at https://web.expasy.org/protparam/, was employed to examine the physicochemical and structural properties of a protein. For the recombinant D-psicose-3-epimerase protein, important information was obtained through this analysis, including the molecular weight (33 kDa), isoelectric point (5.43), and number of amino acids (294).As illustrated in the figure below (figura 3), successful amplification of the amplicons was achieved using the primer set designed for the D-psicose-3-epimerase enzymes. The gene encoding D-psicose-3-epimerase was amplified by testing the bacterial consortium DNA (lanes 2, 3, and 4), resulting in specific amplification of the predicted 885 bp fragment for D-psicose-3-epimerase as per *in silico* analysis. To eliminate any non-specific amplification detected on the gel, the amplicon was purified following the standardized PCR reaction, as described in the materials and methods section.

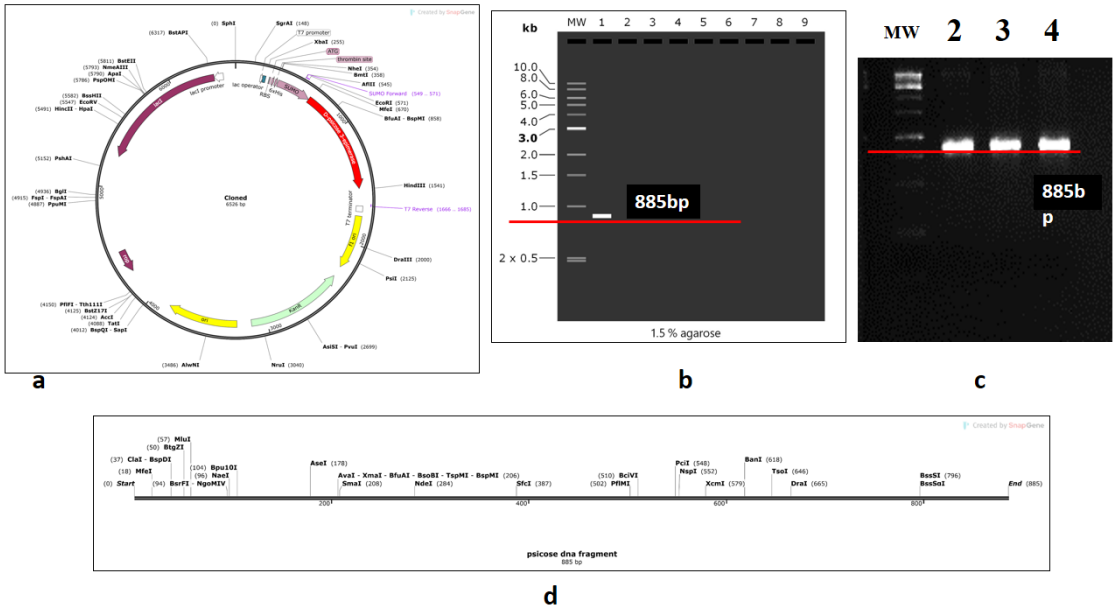

Figure 2 Representation of the plasmid (PET-SUMO) and the gene encoding D-psicose-3-epimerase enzyme inserted into the cloning site. Image obtained using SnapGene software Amplification of the coding regions for D-psicose-3-epimerase enzyme (lanes 2, 3, and 4), lane 1 (Molecular weight marker)

Cloning and Expression of Heterologous Proteins: Amplification, Purification, and Ligation of Target Fragment + PET-SUMO Plasmid

According to the results of amplification of the genes encoding D-psicose-3-epimerase and Ribose-5-phosphate isomerase enzymes, the process of cloning and heterologous expression was initiated. Initially, the entire cloning and expression engineering procedure for D-psicose-3-epimerase enzyme was carried out using the PET-SUMO expression vector, as detailed in the protocol described in the Materials and Methods section. To confirm the cloning step of the target gene, colony DNA was subjected to electrophoresis after transformation, resulting in 16 positive clones (Figure 3a) with gene amplification at the expected size (885 bp). Plasmid DNA from the positive clones was extracted and purified, showing amplification at a size corresponding to vector + insert (6526 bp), suggesting correct insertion of the target gene into the expression vector (Figure 3b). Additionally, plasmid DNA from positive clones was subjected to PCR reactions using Sanger sequencing primers, as well as respective primers for target gene amplification at different sizes, depending on the primer combination used. For this test, clone 1 was selected, which exhibited amplification at the expected sizes (Figure 3c). The objective at this stage was to obtain amplified fragments at specific sizes to confirm the presence of the recombinant protein sequence in both frames, thereby validating this entire cloning and recombinant protein engineering step.

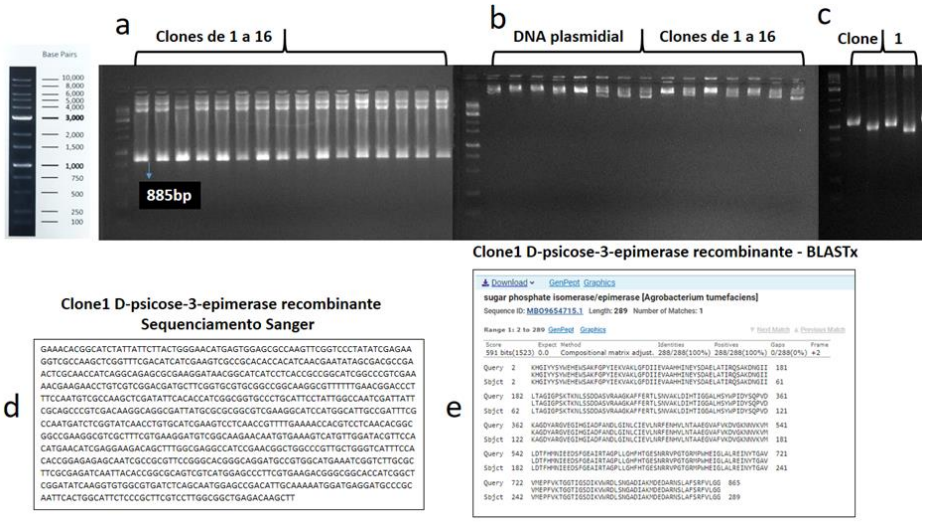

Figure 3 Results and confirmation of the cloning stage for D-psicose-3-epimerase: a) Confirmation of positive clones post-extraction, b) Plasmid DNA amplification, c) PCR amplification to validate the target gene in the correct frame, d) Nucleotide sequence after sequencing of the target gene, e) Similarity analysis and frame of the sequenced target gene sequence.

The plasmid DNA underwent sequencing followed by bioinformatic analysis as per LBMP/IPBEN standard protocol. Results showed accurate cloning of the target gene into the correct frame as a recombinant form within *E. coli* ER2265 competent cells for clones 1, 4, 6, 8, 9, 15, and 16. In Figure 3d, sequencing data for clone 1 is presented, followed by BLASTX analysis (Figure 3e) where nucleotide sequence is translated into protein. For clone 1, a 100% similarity was observed with proteins from the isomerase/epimerase family acting on sugars, confirming the validity and precision of this stage of the work. Similar analyses were conducted for the other clones, yielding comparable results.

Expression of Recombinant DPEase (D-psicose-3-epimerase)

Initially, protein expression was evaluated over a 24-hour fermentation period (Figure 4a and 4b). Subsequently, the extract from the 6-hour time point was used to assess protein expression in the soluble fraction. For the production of recombinant D-psicose-3-epimerase protein, the method involving induction of the T7 RNA polymerase promoter with 0.1 mM IPTG was employed (Park et al., 2016). The results of SDS-PAGE analysis demonstrated the presence of a band with an approximate molecular weight of 33 kDa throughout the 24-hour fermentation and also in the soluble fraction extract at 6 hours (Figure 4a and 4c - SDS PAGE).The same observations were made using Western blotting analysis with the same protein sample. The monoclonal anti-polyhistidine antibody revealed a strongly labeled band, approximately 33 kDa, corresponding to the protein of interest during the 24-hour enzyme production period and also in the soluble fraction at 6 hours (Figure 5b and 5d - Western Blotting). However, the monoclonal anti-polyhistidine antibody did not detect any significant bands at t=0, without the addition of IPTG. These results are consistent with those reported by Park *et al*. (2016)[3], where higher expression levels were observed after 24 hours of cultivation, and the protein exhibited a molecular weight of 33 kDa, consistent with the findings of MU *et al*., 2011[4].

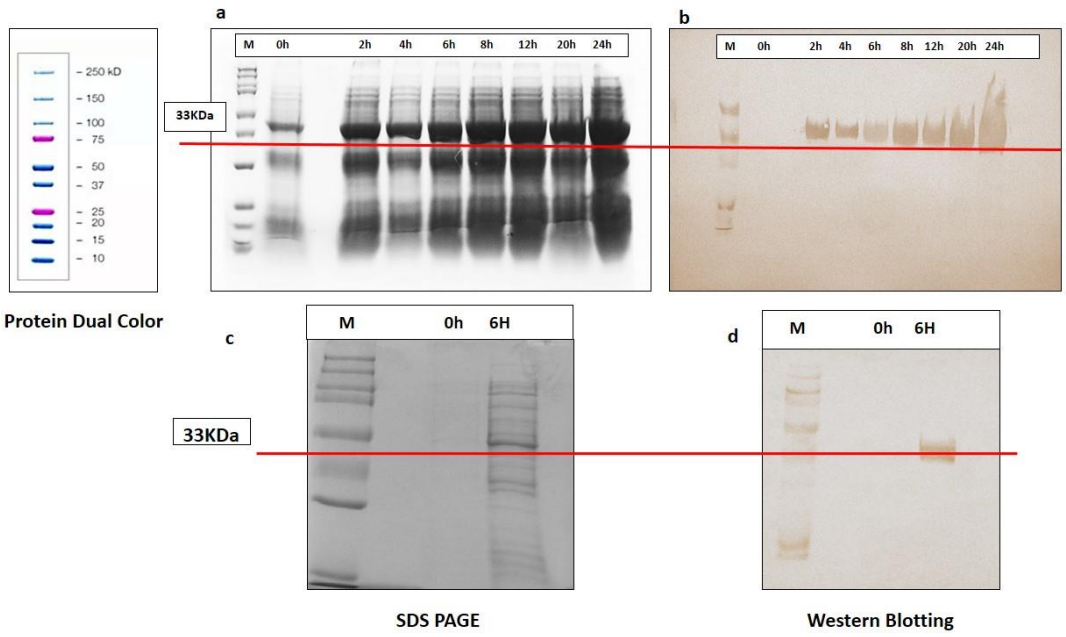

Figure 4 SDS-PAGE and Western Blotting of recombinant DPEase extraction

Homology Analysis of A0A4Q5LXW (D-Psicose 3- Epimerase)

The multiple sequence alignment of A0A4Q5LXW_A against other epimerases (Figure 5) reveals the conservation and functional importance of specific amino acid residues in the sugarphosphate epimerase family. Conserved regions, highlighted in white letters on a red background, indicate crucial areas for maintaining protein structure and function across different epimerase variants. The conserved catalytic residues, such as Glutamate 141, Aspartic acid 173, Histidine 199, and Glutamate 234, play key roles in coordinating essential metal ions for catalytic activity. Additionally, substrate-binding residues, marked with bold circles, likely contribute to substrate recognition and binding specificity. The presence of identical amino acids highlighted in red and moderately conserved regions underscores the functional significance of these positions within the sugarphosphate epimerase family. These findings suggest a high degree of evolutionary conservation in critical functional residues across different epimerase homologs, highlighting their importance in enzymatic activity and substrate specificity.

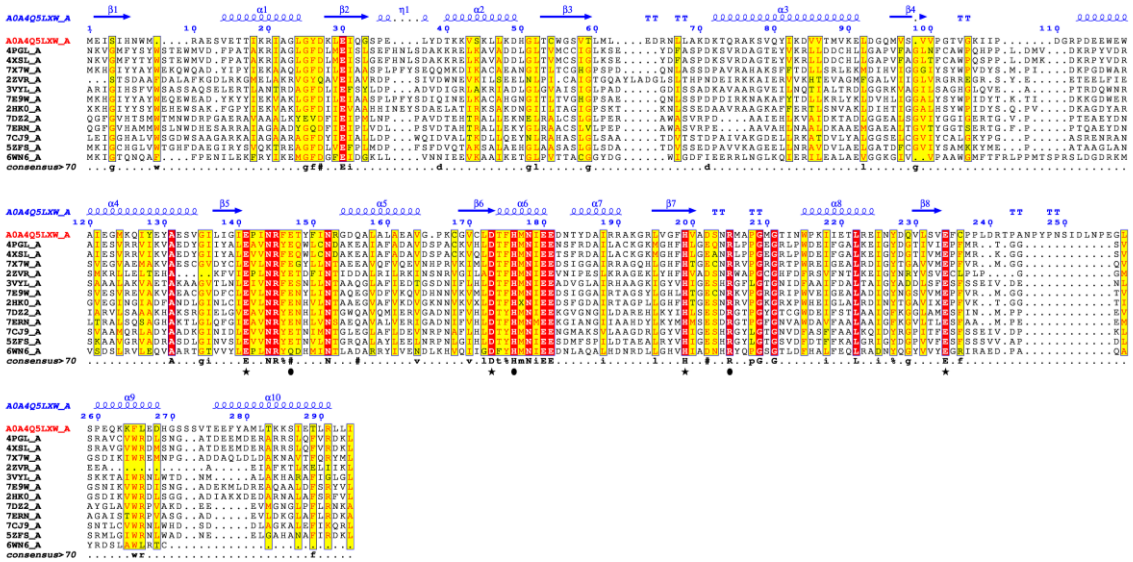

Figure 5 A multiple sequence alignment was performed using the amino acid sequences of A0A4Q5LXW_A against other epimerases belonging to the sugar phosphate epimerase family. Conserved regions were identified and marked with white letters on a red background, moderately conserved regions with red letters on a yellow background, and identical amino acids were highlighted in red. The conserved catalytic residues responsible for metal binding (Glutamate 141, Aspartic acid 173, Histidine 199, and Glutamate 234) were indicated with bold stars, while substrate-binding residues were marked with bold circles. The alignment included sequences from epimerases of Pseudomonas cichorii (4PGL and 4XSL), Clostridia bacterium (7X7W), Thermotoga maritima (2ZVR), Mesorhizobium loti (3VYL), Ruminiclostridium cellulolyticum H10 (7E9W), Agrobacterium tumefaciens (2HKO), Sinorhizobium fredii CCBAU 83666 (7DZ2), Agrobacterium sp. SUL3 (7ERN), Methylomonas sp. DH-1 (7CJ9), Arthrobacter globiformis (5ZFS), and Escherichia coli (6WN6). The alignment was generated using the Swiss-Model tool[5] and assembled in ESPript 3.0[6].

3 CONCLUSION

So far, the results have shown excellent expression of the enzyme psicose-3-epimerase over 24 hours, with a significant increase over time. Additionally, it was possible to identify the presence of this enzyme in the soluble fraction after the extraction step. These findings validate all the *in silico* analysis steps, as well as the recombinant protein engineering stage. Furthermore, these results demonstrate the potential for producing biotechnologically relevant enzymes using bioinformatics tools combined with recombinant protein engineering techniques, thereby directly contributing to more sustainable processes aimed at high-yield enzymatic pathways.

REFERENCES

1 LAEMMLI, Ulrich K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. nature, v. 227, n. 5259, p. 680-685, 1970.

2 TOWBIN, Harry; STAEHELIN, Theophil; GORDON, Julian. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proceedings of the national academy of sciences, v. 76, n. 9, p. 4350-4354, 1979.PARK, Chang-Su et al. Production of d-psicose from d-fructose by whole recombinant cells with high-level expression of d-psicose 3-epimerase from Agrobacterium tumefaciens. Journal of bioscience and bioengineering, v. 121, n. 2, p. 186-190, 2016.

3 PARK, Chang-Su et al. Production of d-psicose from d-fructose by whole recombinant cells with high-level expression of d-psicose 3-epimerase from Agrobacterium tumefaciens. Journal of bioscience and bioengineering, v. 121, n. 2, p. 186-190, 2016

4 MU, Wanmeng et al. Cloning, expression, and characterization of a D-psicose 3-epimerase from Clostridium cellulolyticum H10. Journal of agricultural and food chemistry, v. 59, n. 14, p. 7785-7792, 2011.

5 WATERHOUSE, Andrew et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research, v. 46, n. W1, p. W296-W303, 2018.

6 ROBERT, Xavier; GOUET, Patrice. Deciphering key features in protein structures with the new ENDscript server. Nucleic acids research, v. 42, n. W1, p. W320-W324, 2014.

ACKNOWLEDGEMENTS

Grant 2022/13149-1, São Paulo Research Foundation (FAPESP) and Faculty of Agricultural and Veterinary Sciences - UNESP-FCAV, Jaboticabal, São Paulo, Brazil.