DEVELOPMENT OF SYNTHETIC TRANSCRIPTION FACTORS FOR SOMATIC EMBRYOGENESIS ENHANCEMENT IN CACAO

Rahman M. Z. A.[1]*, Jainuddin H.[1], Abdullah M. F.[1], Daud A. B.[1], Fisal A.[2], Mohtar N. S.[3], Lea J.[1], Norasekin T.[1], Ishak Z.[2] and Alias A. D.[1]

[1]Malaysian Cocoa Board, Centre for Cocoa Biotechnology Research, Commercial Zone 1, Norowot Road, South KKIP, 88460 Kota Kinabalu, Sabah, Malaysia

[2]Malaysian Cocoa Board, Cocoa Innovative and Technology Centre, Lot 12621, Nilai Industrial Area,71800 Nilai, Negeri Sembilan, Malaysia

[3]Universiti Teknologi MARA Sabah, Faculty of Applied Science, Locked Bag 71, 88997 Kota Kinabalu, Sabah.

**Corresponding author: hilmi@koko.gov.my*

Malaysian Cocoa J. 16: 125-132 (2024)

ABSTRACT - *Somatic embryogenesis is a crucial biotechnological approach for mass propagation of cacao (Theobroma cacao) plants. However, achieving high efficiency and reproducibility in somatic embryogenesis remains a challenge and limits its full industrial application for commercial scale propagation. Attempts have been made to enhance somatic embryogenesis in cacao by the overexpression of the LEAFY COTYLEDON (LEC2) gene in transgenic cacao tissue by promoting the transition of somatic cacao cells from the vegetative to embryonic state. The works on these transcription factors have shown very promising results and provide confirmation that transcription factors can be used to enhance SE without compromising plant development. With the emergent technology of bio-mimicry and protein engineering, a new synthetic transcription factor associated with cell reprogramming could be developed to induce somatic embryogenesis in cacao. In this study, we have demonstrated that the DNA binding domain of peptide-based synthetic transcription factors ranging from 13-100 amino acids could be engineered and have the ability to bind at the targeted DNA binding site as a natural counterpart of the LEC2 transcription factor. The newly developed synthetic transcription factor would offer additional advantages as opposed to the conventional counterpart in terms of its robustness, cost of production and ease of cell delivery without undergoing "transgenic-event" and eliminate the risk of GMO generation.*

Keywords: LEAFY COTYLEDON 2, artificial transcription factor, molecular dynamics simulation, protein modelling

INTRODUCTION

Seeds are the most common method of plant propagation in the world. However, seeds have several disadvantages despite their reliability as some plants lack seed dormancy, which means embryos cannot be halted during development, and therefore their seeds are not viable for a short period of time. Consequently, seed-based propagation imparts high production costs and introduces high levels of genetic heterogeneity, as many of these species have self-incompatible flowers that require two genotypes to produce seeds. In crop production, high genetic heterogeneity is undesirable as it is hard to preserve desired traits and loss of these traits can result in significant reductions in productivity. As a result of this issue, different techniques have been developed to produce plants asexually rather than through seeds. Maximova *et al.* (2002) have identified somatic embryogenesis (SE) as a preferred method for cacao propagation. It is still a significant limitation for cacao SE to be propagated on a commercial scale due to its time-consuming nature, low efficiency, genotype dependency, and low conversion rate. The processes that induce a cell to change its fate and enter an embryonic program of development are not known, nor is it known whether a common pathway is used to initiate embryonic development in all of these diverse cell types. LEAFY COTYLEDON 2 (LEC2) regulates a wide range of embryonic processes (Meinke *et al*., 1994; Stone *et al*., 2001) and encodes a transcription factor with a DNA binding domain called the B3 domain, which has been observed only in plant proteins (D. Gaj, 2001; Santos Mendoza *et al*., 2005; Braybrook *et al*., 2006), and binds to the 5′ flanking regions of LEC2-induced genes (Giraudat *et al*., 1992). During the early development and maturation phases of developing zygotic embryos, LEC2 is mainly expressed (D. Gaj, 2001). In addition to developing and maintaining suspensors and cotyledons, it is essential for acquiring desiccation tolerance and inhibiting premature germination. It has been shown that two transcription factors closely related to LEC2, ABA INSENSITIVE3 (ABI3) and FUSCA3 (FUS3), also play critical roles in embryogenesis (D. Gaj, 2001). As part of embryonic development, the LEC genes also regulate fatty acid biosynthesis and storage lipid deposition. In maize and canola, overexpression of ZmLEC1 and BnLEC1 resulted in a 35% and 20% increase in seed oil contents,

respectively (Shen *et al*., 2010; Tan *et al*., 2011). The ectopic expression of AtLEC2 in Arabidopsis leaves led to an accumulation of seed-specific fatty acids (C20:0 and C20:1) and increased oleosin mRNA levels (Stone *et al*., 2001). In addition, AtWRI1, a downstream target of AtLEC2, interacts with key genes upstream in the fatty acid metabolism pathway (Maeo *et al*., 2009).

A number of strategies have been attempted to enhance somatic embryogenesis in cacao by overexpressing the BABYBOOM (BBM) gene and LEC2 gene in transgenic cacao tissues to facilitate the transition of somatic cacao cells from vegetative to embryonic (Zhang *et al*., 2014). Comparatively to non-embryogenic calli, dedifferentiated cells capable of somatic embryogenesis (embryogenic calli) expressed higher levels of TcLEC2. An ectopic expression of TcLEC2 in cacao leaves changed the expression levels of several seed-related genes. Transient overexpression of TcLEC2 in immature zygotic embryos altered gene expression profiles and fatty acid composition. TcLEC2 overexpression in cacao explants greatly increased the frequency of somatic embryo regeneration. When cultured on a hormone-free medium, TcLEC2 overexpressing cotyledon explants demonstrated an iterative embryogenic chain reaction (Shires *et al*., 2017) with an exceptionally high level of embryogenic competency. In the first commercially relevant plant, *T. cacao*, a chimeric transcription factor activated by dexamethasone was used to demonstrate titratable control over somatic embryo formation. This four-fold increase in embryo production rate was achieved by fused glucocorticoid receptors with an embryogenic transcription factor LEAFY COTYLEDON 2 (Shires *et al*., 2017). Previously, T. cacao somatic embryogenesis was restricted to flower parts, but the construct confers a unique embryogenic potential on leaves.

Gene regulation plays a pivotal role in numerous biological processes, and the ability to precisely control gene expression offers valuable insights into fundamental cellular mechanisms and potential therapeutic applications. Traditional approaches for manipulating gene expression, such as RNA interference and small molecule inhibitors, have certain limitations in terms of specificity and efficacy. The advent of synthetic transcription factors (STFs) has revolutionized the field by providing a versatile toolset for targeted gene regulation. Synthetic transcription factors (STFs) are engineered proteins designed to regulate gene expression in a precise and targeted manner. The design of STFs involves selecting or engineering a DNA-binding domain that can specifically recognize and bind to a target DNA sequence, coupled with an effector domain that modulates transcriptional activity. A common strategy in STF design is to utilize modular approaches, where DNA-binding domains from naturally occurring transcription factors are fused with custom effector domains to either activate or repress target genes. These modular STFs can be fine-tuned to enhance specificity and function, making them powerful tools for manipulating plant developmental processes, including somatic embryogenesis (SE).

Previous research has demonstrated the potential of STFs in various plant species. For example, the LEAFY COTYLEDON (LEC) genes, particularly LEC2, have been extensively studied for their role in promoting somatic embryogenesis in plants such as *Arabidopsis thaliana* and *Theobroma cacao*. Studies by Stone et al. (2001) and Gaj (2001) showed that overexpression of LEC2 in *Arabidopsis* can induce the formation of somatic embryos from vegetative tissues, indicating its key role in embryogenesis. Similarly, the BABY BOOM (BBM) transcription factor, which has been shown to trigger somatic embryogenesis in *Brassica napus* and other species, has been used to design synthetic versions aimed at improving SE efficiency (Boutilier *et al*., 2002).

In trees and other plants, synthetic transcription factors have been applied to improve regeneration protocols. For instance, in *Populus* species, STFs designed to modulate the expression of genes involved in stress response and development have been explored to enhance regeneration and transformation efficiencies (Mahfouz et al., 2012). Moreover, the use of CRISPR-based transcriptional activators, which are a form of STFs, has been employed to activate endogenous genes involved in somatic embryogenesis in various crops, including maize and rice, demonstrating the versatility of STFs across different plant systems (Lowder *et al*., 2015).
These studies highlight the potential of synthetic transcription factors to overcome the limitations associated with natural transcription factors, such as lack of specificity and undesired off-target effects. By employing rational design strategies, STFs can be tailored to specific developmental pathways, providing a robust and scalable solution for enhancing somatic embryogenesis and other critical processes in plant biotechnology.

The works on these transcription factors have shown very promising results and provide confirmation that transcription factors can be used to enhance SE without compromising plant development. Here, we demonstrated that the DNA binding domain of peptide-based synthetic transcription factors ranging from 13-100 amino acids could be engineered and exhibit the DNA

binding capability of native LEC2 while having similar affinity toward the targeted DNA binding site.

MATERIALS AND METHODS

Sequence Analysis

The TcLEC2 protein sequence was retrieved from Genome criollo V2 at http://cocoa-genome-hub.southgreen.fr/. The B3 domain of LEC2 protein was identified based on consensus sequence via Multiple Sequence Alignment (MSA) of closely related plant LEC2 proteins (*Arabidopsis thaliana*, *Brassica napus* and Malvaceae) using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

TcLEC2 STFs Protein Structure Modeling

A comparative modeling approach and *de novo* method were conducted to model the structure of STFs. L1 STF was modelled using comparative modeling technique via YASARA software version 12.5.7 (YASARA Biosciences GmbH, Austria), whereas, L7 STF was modelled using *de novo* protein prediction, PEP-FOLD server *(https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/)*. The modelled protein structures were evaluated and validated using structure validation server, SAVESv6.0 (*https://saves.mbi.ucla.edu*) and the overall quality scores were compared between the generated models and the best model with the highest score will be used as the reference model for the subsequent experiment.

Modeling DNA Structure of Direct Target of STFs Binding Site

TcAGL15 was identified as the direct target of TcLEC2 and the DNA sequence was retrieved from Genome criollo V2 at *http://cocoa-genome-hub.southgreen.fr/*. The stretch of DNA region containing LEC2 recognition site (CATGCA) was modelled from TcAGL15 into three-dimensional model of B-DNA structure using the web-based tool, 3DNA 2.0 (*http://web.x3dna.org*).

Molecular Docking of STFs with TcAGL15 DNA

Molecular docking experiments were performed using YASARA software version 12.5.7 (YASARA Biosciences GmbH, Austria) with AutoDock plugin. Protein structures of L1 and L7 were treated as ligand and TcAGL15 DNA structure of the region of interest was used as the receptor for the docking experiments. A total of 25 docking runs were conducted and the results were clustered to identify the distinct complex conformations and contacting residues.

Gene synthesis of TcLEC2 STFs

The genes for TcLEC2 STFs (L1, L2, L6, L7 and L8) were sent for gene synthesis at GeneArt and were delivered as recombinant plasmids pMX.

Molecular cloning of TcLEC2 STFs

The synthetic constructs of L1, L2, L6, L7 and L8 were subcloned into pET100/D-TOPO and transformed into *E. coli* strain KRX for prokaryotic expression. The positive transformants were verified by Sanger sequencing.

Protein expression of TcLEC2 STFs using prokaryotic expression system

The TcLEC2 STFs (L1, L2, L6, L7 and L8) proteins were expressed by 0.1mM IPTG and 0.1% rhamnose induction in 2000ml *E. coli* KRX batch culture under optimal conditions at 37°C for 16 hours. The bacterial cultures were pelleted down at 10,000 x *g* for 15 min at 4°C.

Crude protein samples preparation of Synthetic TcLEC2 STFs

The recombinant KRX cells were then subjected to cell lysis using Q500 Sonicator (QSonica, USA) with sonication pulse rate at 15 seconds ON, 45 seconds OFF for the total cycle duration of 10 minutes. The soluble fractions of the expressed fusion proteins were isolated by centrifuging the lysed samples at 10,000 x *g* for 20 min at 4°C.

Protein Purification of L1, L2, L6, L7 and L8 TcLEC2 STFs

The soluble fractions of the expressed fusion proteins were subjected to affinity chromatography using 1ml HisTrap HP columns and were run using ÄKTA pure protein purification system. Optimization of binding and elution buffer's composition for his-tagged affinity chromatography were conducted using varying NaCl concentration from 0.5M to 2M. The eluted purified fusion proteins were analyzed by using 12% SDS-PAGE.

Characterization of TcLEC2 STFs using Electrophoretic Mobility Shift Assay (EMSA)

LightShift™ Chemiluminescent EMSA Kit (Thermofisher, USA) was used to conduct EMSA assay. The reaction mixture was prepared by combining the purified protein of TcLEC2 STFs (L1, L2, L6, L7 and L8) biotin labeled DNA probe with the binding buffer according to manufacturer's instruction (Thermofisher, USA). The reaction mixture samples were incubated at 25°C for 30 min to allow sufficient interaction between the

transcription factor peptide and the biotin labeled DNA probe. The reaction mixture samples then subjected to 12% native polyacrylamide gel electrophoresis (PAGE) at a constant voltage of 100V for 50 min or until satisfactory separation of DNA-protein complexes is achieved by considering the expected size of the complexes. The DNA-protein complexes PAGE gel was transferred to a positive nylon membrane, UV crosslinked, probed with streptavidin-HRP conjugate and incubated with the substrate to develop the membrane for visualization of DNA-protein complexes band.

RESULTS AND DISCUSSIONS

Modelled Protein Structures of L1 and L7 STFs

Our understanding of DNA binding proteins depends on our understanding and in-depth knowledge of their protein structures (Hegyi and Gerstein, 1999). A protein crystal structure forms the basis for structural biology and protein chemistry (Wiencek, 1999). It has been reported that only a few thousand protein structures have been solved despite significant efforts. To date, no crystal structure has been reported for TcLE2. Research in computational biology has made it possible for researchers to model protein structures despite the fact that crystal structure counterparts are not available for experimentation. There are two approaches to modeling protein: comparative homology modeling and *de novo* design. In order to perform comparative homology modeling, proteins need to share high sequence homology and have their structure solved via X-ray crystallography or NMR. *De novo* design, however, relies on algorithms and structural libraries to predict the structure of a protein rather than starting templates.

L1 STF consists of 100 amino acids, which include a DNA binding domain from TcLEC2, but a regulatory domain was not included. In this study, the three-dimensional structure of L1 STF was modeled using an automated comparative modeling approach using a protein template (PDB ID: 4LDV) from Arabidopsis thaliana Auxin Response Factor 1 (ARF). Figure 1 shows the predicted structure of L1 STF derived from homology modeling. In order to evaluate and validate the protein structure, SAVES v6.0 was used. It was obtained a ERRAT score of 90.2174 and passed the minimum required score (>= 0.2) for VERIFY3D. Ramachandran plot (Figure 2), however, indicates that the structure requires some refinement since 2.2% of residues are in disallowed regions.

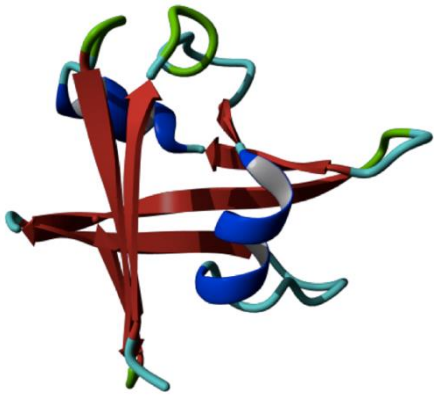

Figure 1: L1 protein structure generated from comparative modeling approach using YASARA. The L1 B3 domain folds in a seven-stranded open ß barrel structure.

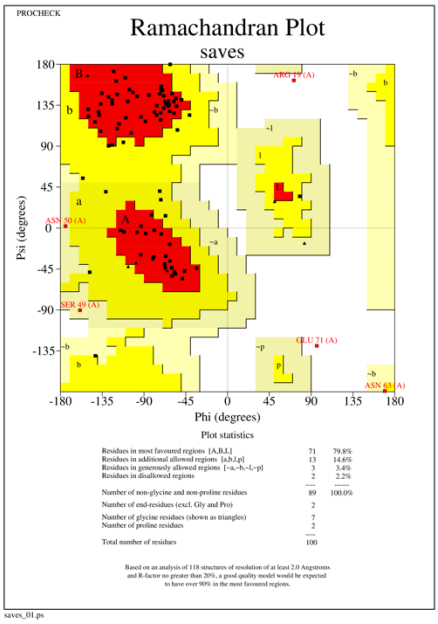

Figure 2: Protein structure quality evaluation and validation. Ramachandran plot analysis of L1 generated model by YASARA. 79.8% amino acid residues located in the most favoured regions.

L7 STF is a hairpin peptide rationally designed based on binding molecules and composed of 13 amino acids (CFYHFPGHFFYAK). The three-dimensional structure of the L7 STF was modeled using PEP-FOLD. The PEP-FOLD method models 3D conformations of peptides between 9-25 amino acids in aqueous solution using a hidden Markov model-derived structural alphabet. Figure 3 shows the predicted structure of L7 STF based on PEP-FOLD *de novo* experiment.

Figure 3: L7 peptide structure generated from de novo modeling approach using PEPFOLD. The L7 STF forms hairpin protein structure.

Protein-DNA interaction of L1 and L7 STFs

As an aid in understanding how L1 and L7 STFs bind DNA, molecular docking was performed on a stretch of DNA from TcAGL15 with the RY motif as the receptor and the predicted structures of the L1 and L7 STFs as ligands. Each protein was docked 25 times globally. According to L1-DNA docking experiments (Figure 4), the docked protein-DNA structure complex, B3 domain, recognizes the DNA largely in its major groove. The B3 ß barrel is positioned laterally to the DNA with its axis almost parallel to the axis of the DNA double helix (Figure 4). There are two parallel sugar-phosphate backbones that are parallel to two adjacent ß strands (ß5 and ß6). These loops (S61-R66), which are located on either side of the barrel, penetrate the major groove further and form interactions that contribute to the recognition of specific DNA bases (Figure 5). Besides these base contacts, DNA binding is also influenced by interactions with residues like E8, K10, A13, L34, F59, and L90 on the DNA backbone. Through hydrogen bonding and electrostatic interactions, L7 hairpin peptide interacts with DNA recognition sites through docking experiments (Figure 6). In the future, molecular dynamics simulations should be used to unravel the details of DNA binding interactions.

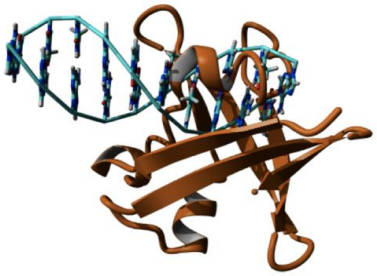

Figure 4: Protein structure of L1 STF/AGL15 DNA complex. The B3 ß barrel is positioned laterally to the DNA with the axis of the barrel almost parallel to the axis of the DNA double helix.

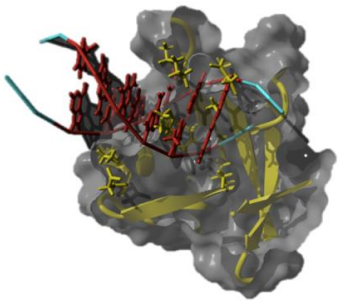

Figure 5: The cross section of complex L1 STF molecular surface. L1 binds and make contact at the major groove of the targeted DNA region.

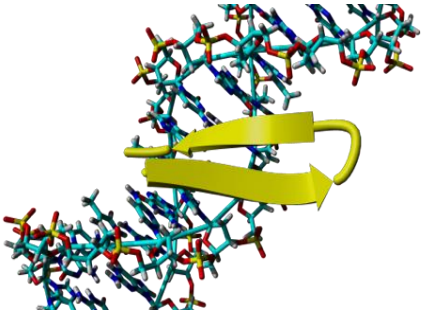

Figure 6: Docked structure of L7/TcAGL15 DNA complex.

Sub-cloning of Synthetic TcLEC2 STF Constructs into Prokaryotic Expression Vector

The host harboring the recombinant pMX plasmids for TcLEC2 STF constructs were grown in LB broth containing ampicillin (50 µg /mL) overnight at 37°C, 150 rpm. Following the manufacturer's instructions, QIAprep Spin Miniprep kit (Qiagen, Germany) was used to extract the recombinant plasmid. We amplified the genes encoding TcLEC2 STF constructs and cloned them into pET100/D-TOPO (Thermo Fisher, USA) as directed. L1, L2, L6, L7, and L8 recombinant plasmids were transformed into E. coli KRX. Plasmid isolation and sanger sequencing confirmed positive transformants (Figure 7).

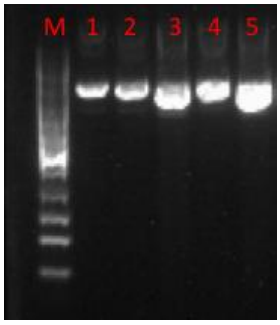

Figure 7. Recombinant plasmid pET100/D-TOPO for L1, L2, L6, L7, L8 constructs isolation. Lane M : 1kb DNA marker. Lane 1 : pET100/D-TOPO-L1 plasmid. Lane 2 : pET100/D-TOPO-L2 plasmid. Lane 3 : pET100/D-TOPO-L6 plasmid. Lane 4 : pET100/D-TOPO-L7 plasmid. Lane 5 : pET100/D-TOPO-L8 plasmid.

Protein expression of TcLEC2 STFs using prokaryotic expression system

TcLEC2 STF proteins were expressed in *E. coli* strain KRX because it has a number of attributes that make it an excellent protein expression host. The ompT- and ompP-mutations eliminate one source of proteolysis of overexpressed proteins in *E. coli*. The T7 RNA polymerase gene is incorporated into KRX by using the rhamnose promoter (rhaPBAD) BAD is one of the most widely used protein expression systems because it has a well-defined promoter independent of *E. coli* RNA polymerase promoters, as well as the rapid elongation rate of T7 polymerase (Golomb and Chamberlin, 1974). The promoter is catabolitely repressed by glucose and activated by adding recombinant rhamnose BAD protein. The isomerase (RhaA), kinase (RhaB) and aldolase (RhaA) are not consumed during growth due to a regulatory cascade of BAD rhamose activating transcription from rhaPBAD.

As part of our protein expression experiment, we induced the cultures with 0.1mM IPTG and 0.1% rhamnose for 16 hours when they reached an O.D.600 of 0.5–0.6 to help them produce soluble proteins. We found that induction with 0.1% rhamnose increased protein yields by up to 80% compared to culture without rhamnose when compared with induction with 0.1% rhamnose. Therefore, even though IPTG alone is capable of driving expression of the STF proteins via T7 promoter, rhamnose is crucial to driving the rhamnose promoter (rhaPBAD) for optimal protein expression.

Protein Purification of L1, L2, L6, L7 and L8 TcLEC2 STFs

Using conventional His-tagged affinity chromatography, few attempts have been made to purify DNA-free TcLEC2 STFs in the past. As a result, the purified proteins remained tightly bound to the host chromosome, making downstream applications like ChIP-seq difficult. With the aim of a single step DNA removal technique, an efficient method for removing contaminated bound DNA from protein samples was explored.

As reported by Vingadassalon *et al*. (2016), contaminated DNA can be removed using either heparin columns, ion exchange columns or DNases such as benzonase. Pre- or post-treatment steps were required for all of the described techniques, which resulted in lost protein yields. Therefore, in this study, we attempt to develop an efficient and simple step-by-step method for removing bound DNA and recovering apo-DNA binding protein (Figure 8). As part of our optimized affinity chromatography protocol for purifying DNA-binding proteins, we introduced three washing steps. As soon as the targeted protein binds to a nickel sepharose column, the first washing step is conducted using a low salt concentration washing buffer (20mM phosphate, 0.5M NaCl, 50mM imidazole, pH 7.4) for five column volumes. During the first washing step, unspecific proteins were removed from the affinity column along with contaminated host chromosomal DNA still tightly bound to the DNA binding protein. The column was washed for five column volumes with a high concentration of salt (2M NaCl) added to the second wash buffer (20mM phosphate, 2M NaCl, 50mM Imidazole, pH 7.4). By interrupting the electrostatic interactions between DNA impurities and the protein of interest, high salt decreases non-specific binding. After washing the column for five column volumes, we used the same buffer composition as the first washing step, which is low salt wash buffer (20mM phosphate, 0.5M NaCl, 50mM Imidazole, pH 7.4). The final washing step is important for equilibration of the column with appropriate salt concentrations before the proteins are eluted from it. The purified proteins for L1, L2, L6, L7 and L8 TcLEC2 STFs were analyzed on SDS PAGE (Figure 9).

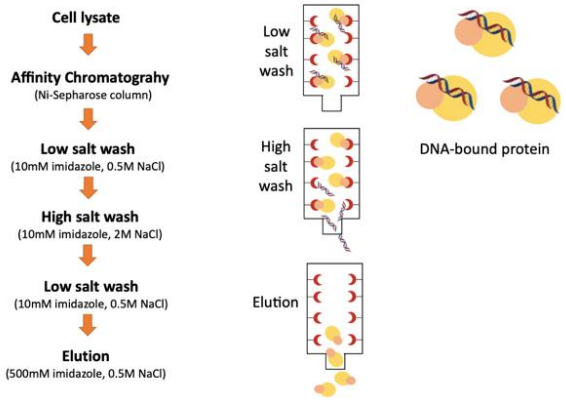

Figure 8. Overview and schematic diagram of optimized single step affinity chromatography for DNA-binding protein purification protocol.

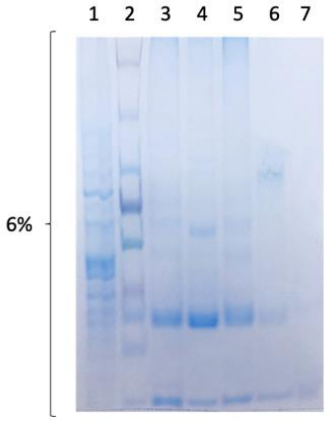

Figure 9. SDS PAGE (6%) of purified L1, L2, L6, L7 and L8 TcLEC2 STFs. Lane 1 : Low range protein marker. Lane 2 : Broad range protein marker. Lane 3 : DNA-free purified L1 protein. Lane 4 : DNA-free purified L2 protein. Lane 5 : DNA-free purified L6 protein. Lane 6: DNA-free purified L7 protein. Lane 7: DNA-free purified L8 protein.

Protein Characterization of TcLEC2 STFs

Electrophoretic mobility shift assay (EMSA) was conducted for five TcLEC2 artificial transcription factors (STFs) constructs namely L1, L2, L6, L7 and L8 and the assay results showed that only L1, L2 and L7 STFs have shifted DNA bands (Figure 10). Based on EMSA principle, the shifted DNA band indicates that there is DNA-protein interaction present, thus effecting the mobility of the detected DNA on the gel. Key domains and residues responsible for DNA binding domain were identified. In order to characterize further the DNA-protein interactions between corresponding STFs, DNA Immunoprecipitation sequencing (DIP-Seq) will be employed to determine the specificity of DNA binding sequence.

Figure 10. EMSA/gel shift assay. Lane 1 : O3 (biotin-labelled DNA probe). Lane 2 : O3 + bovine serum albumin (BSA). Lane 3 : O3 + purified L1 protein. Lane 4 : O3 + purified L2 protein. Lane 5 : O3 + purified L6 protein. Lane 6 : O3 + purified L7 protein. Lane 7 : O3 + purified L8 protein. Arrows indicate positions of mobility shift.

CONCLUSIONS

Although somatic embryogenesis is a biotechnological method for propagating cacao plants in mass quantities, it has some challenges regarding efficiency and reproducibility. In this study, STFs were used to induce somatic embryogenesis in cacao plants as a way of overcoming these limitations.

Using sequence analysis and protein structure modeling, we have successfully modeled the structure of two synthetic transcription factors, L1 and L7 and molecular docking experiments were performed to understand the interaction between these synthetic transcription factors and the targeted DNA binding site. The results showed that L1 and L7 synthetic transcription factors exhibited DNA binding capabilities and interacted with the DNA in a manner similar to the native LEC2 transcription factor. The synthetic transcription factors were successfully expressed in prokaryotic expression system and an optimized purification protocol to obtain DNA-free synthetic transcription factors have developed. Electrophoretic mobility shift assay (EMSA) was conducted to validate the DNA-protein interaction, and the results confirmed the binding of L1, L2, and L7 synthetic transcription factors to the targeted DNA.

Overall, this study demonstrates the potential of synthetic transcription factors in enhancing somatic embryogenesis in cacao. The

engineered synthetic transcription factors showed DNA binding capabilities similar to the native LEC2 transcription factor, offering advantages such as robustness, cost-effectiveness, and ease of cell delivery. Further research, including DNA immunoprecipitation sequencing (DIP-Seq) to determine the specificity of DNA binding sequences, can provide deeper insights into the DNA-protein interactions. The development of synthetic transcription factors could contribute to the industrial-scale propagation of cacao plants through somatic embryogenesis, addressing the limitations of traditional seed-based propagation methods.

ACKNOWLEDGEMENTS

This work is supported by Developmental Fund (P20001001210005 - Development of Cocoa Agri-Biotechnology Innovation Program) from Ministry of Finance under Eleventh Malaysia Plan and Temporary Research Fund (TRF) (02-03-TRF0009 - Development of artificial transcription factors for cell reprogramming and genome editing) from Malaysian Cocoa Board.

REFERENCES

Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2006). 'Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis.', *Proceedings of the National Academy of Sciences of the United States of America* **103:** 3468–3473.

D. Gaj, M. (2001). 'Direct somatic embryogenesis as a rapid and efficient system for in vitro regeneration of Arabidopsis thaliana', *Plant Cell, Tissue and Organ Culture* **64**: 39–46.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. and Goodman, H.M. (1992). 'Isolation of the Arabidopsis ABI3 Gene by Positional Cloning', *The Plant Cell Online* **4:** 1251–1261.

Golomb, M. and Chamberlin, M. (1974). Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis. *Journal of Biological Chemistry* **249**: 2858–2863.

Maeo, K., Tokuda, T., Ayame, A., Mitsui, N., Kawai, T., Tsukagoshi, H., Ishiguro, S. and Nakamura, K. (2009). 'An AP2-type transcription factor, WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis', *The Plant Journal* **60:** 476–487.

Maximova, S. N. *et al*. (2002) 'Efficiency, genotypic variability, and cellular origin of primary and secondary somatic embryogenesis of Theobroma cacao L.', *In Vitro Cellular & Developmental Biology - Plant* **38**: 252–259.

Meinke, D.W., Franzmann, L.H., Nickle, T.C. and Yeung, E.C. (1994). 'Leafy Cotyledon Mutants of Arabidopsis.', *The Plant cell* **6:** 1049–1064.

Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M. and Lepiniec, L. (2005). 'LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves', *FEBS Letters* **579:** 4666–4670.

Shen, B., Allen, W.B., Zheng, P., Li, C., Glassman, K., Ranch, J., Nubel, D. and Tarczynski, M.C. (2010). 'Expression of ZmLEC1 and ZmWRI1 Increases Seed Oil Production in Maize', *Plant physiology* **153:** 980–987.

Shires, M. E., Florez, S.L., Lai, T.S. and Curtis, W.R. (2017). 'Inducible somatic embryogenesis in Theobroma cacao achieved using the DEX-activatable transcription factor-glucocorticoid receptor fusion', *Biotechnology Letters* **39:** 1747–1755.

Stone, S. L., Kwong, L.W., Yee, K.M., Pelletier, J., Lepiniec, L., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2001). 'LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development', *Proceedings of the National Academy of Sciences* **98:** 11806–11811.

Tan, H., Yang, X., Zhang, F., Zheng, X., Qu, C., Mu, J., Fu, F., Li, J., Guan, R., Zhang, H., Wang, G. and Zuo, J. (2011). 'Enhanced Seed Oil Production in Canola by Conditional Expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in Developing Seeds', *Plant Physiology* **156:** 1577–1588.

Vingadassalon, A., Bouloc, P. and Rimsky, S. (2016). Removing nucleic acids from nucleoid-associated proteins purified by affinity column. *Journal of Biological Methods* **3**:e35.

Zhang, Y., Clemens, A., Maximova, S.N., Guiltinan, M.J. (2014). 'The Theobroma cacao B3 domain transcription factor TcLEC2 plays a duel role in control of embryo development and maturation'. *BMC Plant Biology* **14:** 106.