MOLECULAR DOCKING STUDIES ON DNA-BINDING PROCESS OF LEC2 SYNTHETIC TRANSCRIPTION FACTORS

Rahman M. Z. A.*, Jainuddin H., Mohammed A. and Kasran R.
Biotechnology Division, Centre for Cocoa Biotechnology Research, Malaysian Cocoa Board, Commercial Zone 1, Norowot Road, South KKIP, 88460, Kota Kinabalu, Sabah, Malaysia.
**Corresponding author: hilmi@koko.gov.my*

Malaysian Cocoa J. 14: 44-48 (2022)

ABSTRACT - *Transcription factors (TFs) are essential for the expression of all proteins, including those involved in the plant regeneration and totipotency. LEAFY COTYLEDON 2 (LEC2) exerts significant impacts on determining embryogenic potential and various metabolic processes through a complicated genetic regulatory network and is sufficient to induce somatic embryo development in vegetative cells. Previously, electrophoretic mobility shift assay (EMSA) was conducted for five TcLEC2 artificial transcription factors (ATFs) constructs namely L1, L2, L6, L7 and L8 and the assay results showed that only L1, L2 and L7 ATFs have shifted DNA bands. 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. In order to validate further the DNA-protein interactions between corresponding ATFs, molecular dynamics studies were conducted to investigate the interaction of the ATFs with the targeted DNA sequence. The protein structure of ATFs models were built using either comparative modeling approach (L1) or de novo approach via PEP-FOLD server (L7) Molecular docking experiments were conducted to depict the protein-DNA interaction and to delineate the mechanism of DNA binding for TcLEC2 ATFs. The details of DNA binding activity of the ATFs provided here will be helpful to better understand the regulation of LEC2 in plants.*

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

INTRODUCTION

LEAFY COTYLEDON 2 (LEC2) regulates many distinct aspects of embryogenesis (Meinke *et al.*, 1994)(Stone *et al.*, 2001) and encodes a transcription factor with a B3 domain, a DNA binding region found thus far only in plant proteins (D. Gaj, 2001; Santos Mendoza *et al.*, 2005; Braybrook *et al.*, 2006) which binds specifically to the RY motifs in the 5′ flanking regions of LEC2-induced genes (Giraudat *et al.*, 1992). LEC2 is exclusively expressed in developing zygotic embryos during both the early development and maturation phases. It is required for development and maintenance of suspensors and cotyledons and for the acquisition of desiccation tolerance and inhibition of premature germination (D. Gaj, 2001). The LEC genes are also involved in regulation of fatty acid biosynthesis and storage lipid deposition during embryo development. The seed specific overexpression of ZmLEC1 and BnLEC1 led to 35% and 20% increase in seed oil contents in maize and canola, respectively (Shen *et al.*, 2010; Tan *et al.*, 2011). Ectopic expression of AtLEC2 in Arabidopsis leaves resulted in the accumulation of seed specific fatty acids (C20:0 and C20:1) and increased the mRNA level of oleosin (Stone *et al.*, 2001). Furthermore, a direct downstream target of AtLEC2, AtWRI1 is known to control fatty acid metabolism through interactions with key genes upstream in the pathway (Maeo *et al.*, 2009).

The overexpression of TcLEC2 in cacao explants greatly increased the frequency of regeneration of stably transformed somatic embryos. TcLEC2 overexpressing cotyledon explants exhibited a very high level of embryogenic competency and when cultured on hormone free medium, exhibited an iterative embryogenic chain-reaction (Shires *et al.*, 2017). The first demonstration of a titratable control over somatic embryo formation in a commercially relevant plant, *T. ca*cao, was achieved using a dexamethasone activatable chimeric transcription factor. This four-fold enhancement in embryo production rate utilized a glucocorticoid receptor fused to an embryogenic transcription factor LEAFY COTYLEDON 2 (Shires *et al.*, 2017). Where previous *T. cacao* somatic embryogenesis has been restricted to dissected flower parts, this construct confers an unprecedented embryogenic potential to leaves.

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. In our previous work, we had successfully designed five new TcLEC2 artificial transcription factors (ATFs) which labelled as L1, L2, L6, L7 and L8 with distinct designs and amino acids composition. Mobility shift assay (EMSA) was

conducted for these ATFs and the assay results showed that only L1, L2 and L7 ATFs have shifted DNA bands. 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. In order to validate further the DNA-protein interactions between corresponding ATFs, molecular docking studies were conducted to investigate the interaction of the ATFs with the targeted DNA sequence.

MATERIALS AND METHODS

TcLEC2 ATFs Protein Structure Modeling

Comparative modeling approach and *de novo* method were conducted to model the structure of ATFs. L1 ATF was modelled using comparative modeling technique via YASARA software version 12.5.7 (YASARA Biosciences GmbH, Austria), whereas, L7 ATF 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 ATFs 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 ATFs 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.

RESULTS AND DISCUSSIONS

Modelled Protein Structures of L1 and L7 ATFs

To understand how a DNA binding protein works, we must have a good understanding and in-depth knowledge of its protein structure (Hegyi and Gerstein, 1999). Protein crystal structure is the key element in the field of structural biology and protein chemistry (Wiencek, 1999). Although significant effort has been put into deciphering protein structures, only a few thousand protein structures have been solved. To date, there is no crystal structure of TcLE2 protein reported yet. Advancement in computational biology had made it possible for researchers to model protein structure despite the unavailability of their experimented crystal structure counterpart. Protein modeling can be modelled using either comparative homology modeling approach or *de novo* design approach. Comparative homology modeling require a protein template that shares high sequence homology that have been solved their structure from X-ray crystallography or NMR experiments. However, *de novo* design approach didn't require any starting templates and rather rely on algorithm and structural libraries to predict the structure of the protein of interest.

L1 ATF consist of 100 amino acids which include binding DNA domain from TcLEC2 while the regulatory domain was excluded. The three-dimensional structure of the L1 ATF was modelled using YASARA via automated comparative modeling approach and protein template (PDB ID: 4LDV) from *Arabidopsis thaliana* Auxin Response Factor 1 (ARF) was chosen as the protein template. The predicted structure of L1 ATF generated from the homology modeling experiment was shown in Figure 1. To evaluate and validate the predicted protein structure, SAVES v6.0 was used. The ERRAT score of 90.2174 was obtained and passed the VERIFY3D minimum required score (>= 0.2). However, the Ramachandran plot (Figure 2) result indicates that the structure needs some refinement because 2.2% residues were in disallowed region.

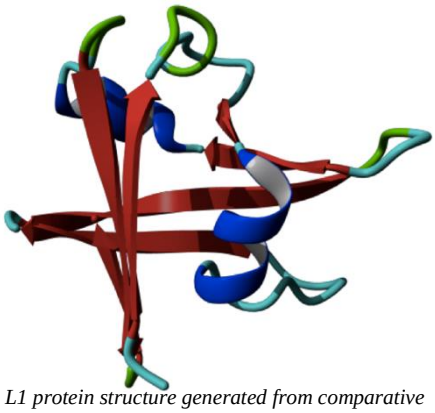

Figure 1: L1 protein structure generated from comparative modeling approach using YASARA. The L1 B3 domain folds in a seven-stranded open ß barrel structure.

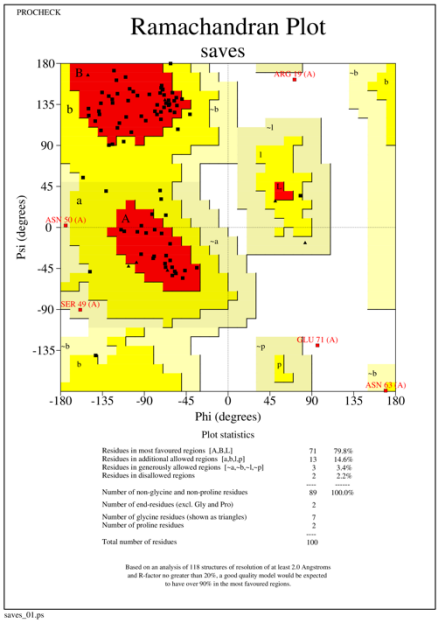

Figure 1: Protein structure quality evaluation and validation. Ramachandran plot analysis of L1 generated model by YASARA. 79.8% amino acid residues located in the most favoured regions.

L7 ATF is a hairpin peptide rationally designed based of polyamide DNA binding molecules and composed of 13 amino acids (CFYHFPGHFFYAK). The three-dimensional structure of the L7 ATF was modelled using PEP-FOLD server. PEP-FOLD uses a hidden Markov model-derived structural alphabet for de novo modeling of 3D conformations of peptides between 9-25 amino acids in aqueous solution. The predicted structure of L7 ATF generated from PEP-FOLD *de novo* experiment was shown in Figure 3.

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

Protein-DNA interaction of L1 and L7 ATFs

To gain insight into the mechanism of DNA binding by L1 and L7 ATFs, molecular docking was conducted by treating predicted the structures of L1 and L7 ATFs as ligand and a stretch of DNA structure from TcAGL15 with RY motif as the receptor. 25 global docking runs were conducted for each proteins. Based on L1-DNA docking experiments, the docked protein-DNA structure complex, B3 domain recognizes the DNA largely at the major groove of CATGCA elements (Figure 4). 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 4). Two adjacent ß strands (ß5 and ß6) run over the major groove, parallel to the two sugar-phosphate backbones. The loop connecting these strands (S61-R66), located on either side of the barrel, further penetrate the major groove and make interactions that contribute to specific DNA base recognition (Figure 5). In addition to these base contacts, DNA binding involves interactions of the DNA backbone with residues E8, K10, A13, L34, F59 and L90. For L7-DNA docking experiments, L7 hairpin peptide interacts with the recognition site of DNA region via hydrogen bonding and electrostatic interaction (Figure 6). Additional experiments should be conducted in the future to unravel the detail interaction that govern the mechanism of DNA binding using molecular dynamics simulation approach.

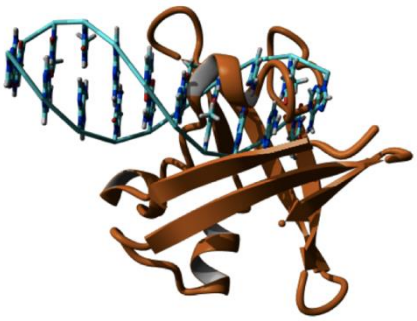

Figure 4: Protein structure of L1 ATF/AGL15 DNA complex. The B3 ß barrel is positioned laterally to the DNA with the axis of the barrel almost parallel to the axis of the DNA double helix

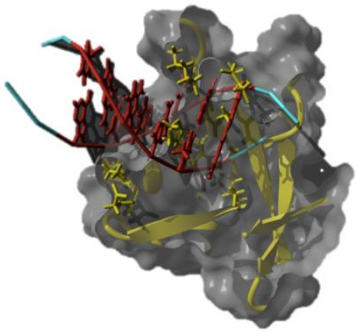

Figure 5: The cross section of complex L1 ATF molecular surface. L1 binds and make contact at the major groove of the targeted DNA region.

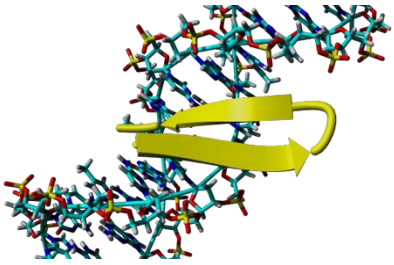

Figure 6: Docked structure of L7/TcAGL15 DNA complex.

CONCLUSIONS

Structural insight of L1 and L7 ATFs-DNA complex were successfully depicted via protein modeling and molecular docking to unravel the protein-DNA interaction and mechanism of DNA binding for LEC2 (L1) and LEC2-like (L7) proteins. This study will provide us a valuable insight for designing an artificial transcription factor for biotechnology application.

ACKNOWLEDGEMENTS

This work is supported by Developmental Fund (P20001001210005 - Development of Cocoa Agri-Biotechnology Innovation Program) from Ministry of Finance under Eleventh Malaysia Plan.

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.

Hegyi, H. and Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *Journal Molecular Biology*. **288:** 147–164.

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.

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.

Wiencek, J.M. (1999) New strategies for protein crystal growth. *Annual Review Biomedical Engineering*. **1:** 505–534.