ASSIGNMENT OF FUNCTION AND HOMOLOGY MODELLING OF SERK AND LEC PROTEINS IN COCOA

S. Nithya, M. K. Rajesh and George V. Thomas

Central Plantation Crops Research Institute, Kasaragod, Kerala

Introduction

Cocoa belonging to the family Malvaceae, is a major source of income for developing countries. Its seeds are used to make chocolate and cocoa powder. About 70% of cocoa trees under cultivation are unselected material propagated by seeds (Eskes, 2000). Because of the high heterozygous nature of seedlings in a given field, only 2-3% of the trees in a population of high-yielding families account for 60% of the yield (Irizarry and Rivera, 1998). Plant tissue culture techniques are an essential part of modern plant biotechnology. Regeneration *via* somatic embryogenesis (SE) offers a reliable technique for increasing yields and homogenizing cocoa production by mass propagation of quality planting materials (Alemanno *et al.,* 1997).

Since the first report of success achieved in cocoa SE (Esan, 1974), many genotypes have responded to *in vitro* SE, successful plantlet regeneration and field establishment. Despite these accomplishments, a lot of improvement has to be still made to enhance scaling-up of cocoa SE to a commercial level. Moreover, many cocoa genotypes have proven to be recalcitrant to SE. It is therefore imperative to undertake a detailed study of the events occurring during somatic embryogenesis, especially the induction phase, which is a critical step for somatic embryo initiation. Knowledge of key genes determining the somatic embryo induction phase, and the proteins they encode, has high practical utility in cocoa biotechnology, particularly for the improvement in somatic embryo production in recalcitrant genotypes.

Somatic embryogenesis involves a set of molecular events such as differential gene expression and various signal transduction pathways for activating/repressing numerous gene sets; many are yet to be identified and characterized (Chugh and Khurana, 2002). Among them, the SERK (SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE) gene, one of the first SE-specific genes identified, is reported to mark embryogenesis-competent cells. The ectopic expression of SERK has been shown to increase embryogenic potential in seedling-derived cultures (Hecht *et al.,* 2001). Genes of regulatory function are critical for developmental processes, and consistent with this transcription factors were found to be the most frequently represented among transcripts specific for SE-induction, including BABY BOOM (BBM, Boutilier *et al.,* 2002), WUSCHEL (WUS, Zuo *et al.,* 2002), AGAMOUS-LIKE15 (AGL15, Harding *et al.,* 2003), and LEC2 (Stone *et al.,* 2001).

SERK genes have since been linked to somatic embryogenesis in a number of species including *Dactylis glomerata*, *Arabidopsis thaliana*, *Medicago truncatula*, *Helianthus annus*, *Citrus unshiu*. The SERK protein may act as a transmembrane receptor of signals present in the culture medium and trigger somatic embryogenesis. This suggests that the use of SERK genes could be used to improve regeneration frequency.

The LEC transcription factor is known to activate a wide range of embryo-specific pathways in higher plants. It acts as a central regulator of the early and late phases of both zygotic and somatic embryogenesis (Stone *et al.,* 2001). During early embryogenesis, LEC genes are required to specify suspensor cell fate and cotyledon identity and for the expression of many maturation-specific genes (Harada, 2001). Somatic embryogenesis receptor kinase (SERK) and leafy cotyledon-like protein (LEC) are two proteins which are known to be involved in the early phase of triggering somatic embryo-induction in cocoa (Alemanno *et al.,* 2008).

Structural genomics is expected to elucidate many experimentally determined proteins structures. Since solving a protein structure by NMR or crystallography remains a long and expensive effort, constructing 3-D models based on structures of homologous proteins is an alternative approach. The present work attempts to unravel the functions controlled by SERK and LEC proteins in cocoa *via* homology modelling.

Methodology

Datasets

Sequences of SERK (Genbank AAU03482.1) and LEC (Genbank CAM35799.1) of cocoa were retrieved from NCBI.

Similarity search and pattern recognition

Blast searches for the SwissProt /TrEMBL curated databases were done using WU BLAST 2.0 algorithm to make out the local alignments for the cocoa SERK and LEC proteins. Further analysis was performed by the ExPasy server tools (www.expasy.org/tools). Sequence pattern recognition and determination of the modular architecture were done by the INTERPRO analysis, based on PROSITE (www.expasy.org/prosite) and Pfam databases (www.pfam.sanger.uk). Physicochemical properties of the selected proteins were determined using the PROTPARAM tools (www.expasy.ch/tools/protparam). Hydropathy plots of the deduced proteins were determined using the Kyte and Doolittle values (Kyte and Doolittle, 1982).

Homology modelling and protein function prediction

The secondary structure analysis was performed in PREDATOR available at NPS@ server (Network Protein Sequence Analysis). The tertiary structure of the protein sequences were predicted by Swiss Model server. The modeled structure was verified in PROCHECK. Function assignments were made based on the structural homologues and similar homologues identified for the test protein.

Results and Discussion

The WU blast analyses results of SERK and LEC proteins are shown in the Tables 1 and 2 respectively. SERK from cocoa showed homology to SERKs from *Carica papaya, Dimocarpus longan, Populus trichocarpa, Solanum peruvianum* etc., while LEC from cocoa showed similarity to LEC proteins from *Arachis hypogaea, Glycine max, Glycine latifolia etc.*

Table 1. WU-BLAST results for cocoa SERK protein

Species showing similarity to SERK protein from cocoa	Accession number	E-value	Identity (%)
Carica papaya	A7L4A8	1.3E-196	99.0
Dimocarpus longan	B5TTV0	1.5E-195	99.0
Populus trichocarpa	B9IQM9	8.2E-195	100.0
Solanum peruvianum	A6N8J2	7.5E-192	99.0
Citrus sinensis	C3V9W0	2.3E-190	99.0
Rosa canina	E2J573	1.8E-188	100.0
Cyclamen persicum	A7L5U3	4.5E-183	97.0
Ageratina adenophora	D5KXW3	6.6E-169	93.0

Table 2. WU-BLAST results for cocoa LEC protein

Species showing similarity to LEC protein from cocoa	Accession number	E-value	Identity (%)
Arachis hypogaea	D6C529	3.8E-65	65.0
Glycine max	B5KMS9	4.8E-65	80.0
Glycine latifolia	B5KMT1	1.9E-63	77.0
Phaseolus coccineus	Q8LJQ5	5.2E-61	75.0
Dimocarpus longan	D4P8F5	2.2E-60	66.0
Vitis vinifera	A5AJA9	1.4E-58	84.0
Selaginella sinensis	A9P3Y3	7.3E-46	85.0
Pistacia chinensis	E0ADH2	4.2E-43	72.0

The PROTPARAM results for both the proteins are shown in Table 3. The amino acid composition of SERK protein revealed that it is highly rich in leucine followed by serine and proline. Cocoa LEC protein is rich in glycine, glutamine and alanine.

Table 3. PROTPARAM results for SERK and LEC protein

Protein name	Molecular weight	Theoretical PI	No. of positively charged aminoacids	No. of negatively charged aminoacids	Aliphatic index	GRAVY value
SERK	50945.4	8.12	42	44	95.62	-0.108
LEC	23682.5	6.41	28	25	60.99	-0.639

The INTERPRO analysis of SERK shows that this protein belongs to protein kinases signature. The consensus pattern [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-LIVMFYWCS TAR]-[AIVP]-[LIVMFAGCKR]-K. SERK protein contain leucine rich repeat N-terminal domain. It acts as protein binding region. The protein carries leucine rich repeat (LRR) domains exposed at the cell membrane that should be responsible for the interaction with other signal molecules, external signal perceptions and signal transduction during somatic embryogenesis induction. This protein belongs to the receptor-like kinase-LRR (RLK) family with a serine-threonine kinase site. Receptor-like kinases are characterized by an extracellular domain, a transmembrane domain and an intracellular kinase domain and therefore play a role in the transduction of extracellular signals. The SERK protein may act as a transmembrane receptor of signals present in the culture medium and trigger somatic embryogenesis. Competent cells may contain an inactive receptor, activated by the presence of the proper ligand to switch on the embryogenic programme (Hecht *et al.,* 2001).

The INTERPRO analysis of LEC protein revealed that it belongs to NF-YB/HAP3 subunit signature. The consensus pattern for this signature is C-[VA]-[ST]-E-x-I-S-F-[LIVM]-T-[SGC]-E-A-[SCN]-[DE]-[KRQ]-C. These proteins contain CCAAT-binding factor (CBF) or NF-Y. CBF is a heteromeric transcription factor that consists of two different components both needed for DNA-binding. It consists of two subunits CBF-A and CBF-B. LEC1 and LEC2 are considered to be transcriptional regulators capable of establishing a cellular environment sufficient to initiate embryo development (Feher *et al.,* 2003). The hydropathy plots reveal that in both the proteins, majority of residues fall below zero shows that the proteins are hydrophilic in nature (Fig. 1, 2).

Secondary structure prediction for LEC and SERK protein sequence shows more coil regions compared to helices. Secondary structure prediction of SERK shows 24.89% helix and 66.09% random coil (Fig.3). Secondary structure prediction of LEC protein sequence shows 33.33% helix and 62.44% random coil (Fig. 4).

The template for SERK protein is 2qkwB (3.2 Å) and the sequence identity is 43.275 (Fig.5). The template chosen for tertiary structure prediction of LEC protein with PDB id 1n1jA (1.67Å) shows 61.628 sequence identity with the LEC protein from cocoa (Fig. 6). These modeled structures form the starting point for future docking studies to prediction potential ligands that can bind to cocoa SERK and LEC proteins as well as critical residues that stabilize ligands.

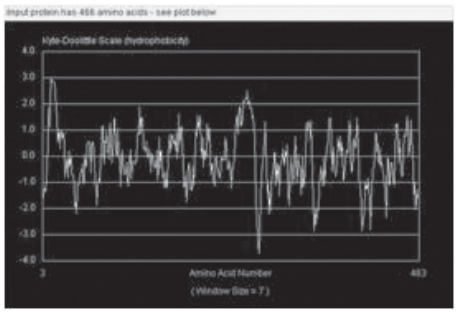

Fig. 1. Hydropathy plot of SERK protein

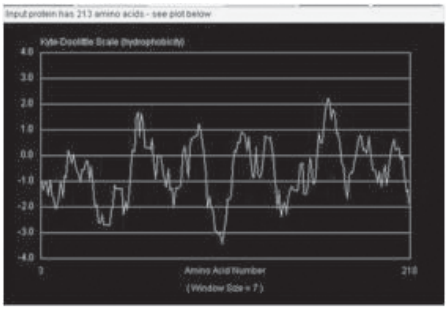

Fig. 2. Hydropathy plot of LEC protein

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MEGGSKKVKSLALVWLIFALLHPLRLISANVEGDALHSLRTNLNDPNNVLQSWDPTLVNPCTWFHVTCNN
ccccccchhhhhhhhhhhhccccceeecccccchhhhhhcccccceeecccccccccceeeeeccc
DNSVIRVDLGNAALSGQLVPQLGLLKNLQYLELYSNNISGQIPSDLGNLTSLVSLDLYLNSFSGPIPESL
cceeeeecccccccccccccccchhhhhhhccccccccccccccccceeeecccccccccccccch
GRLSKLRFLRLNNNSLSGPIPMSLTNITSLQVLDLSNNRLSGEVPDNGSFSLFTPISFANNLDLCGPVTG
hhhhhhhhhhccccccccccccccccccceeecccccccccccccccccccccccccccccccccc
RPCPGSPFSPPPPFVPPPPISSPSGNSVTGAIAGGVAAGAALLFAAPAIAFAWWRRKPQEFFFDVPAEE
ccccccccccccccccccccccccccceeeecccccchhhhhhcchhhhhhhhhcccccccccccccc
DPEVHLGQLKRFSLRELQVATDSFSNKNILGRGGFGKVYKGRLADGSLVAVKRLKEERTPGGELQFQTEV
ccccccccccchhhhhhhhhcccccccccccccccceeeecccccccceehhhhhhhcccccchhhhhhh
EMISMAVHRNLLRLRGFCMTPTERLLVYPYMANGSVASCLRERPPSQPPLDWPTRKRIALGSARGLSYLH
hhhhhhhhhhhhhhhccccccccceeeeecccccccccccccccccccccccccchhhhhhcccceeecc
DHCDPKIIHRDVKAANILLDEEFEAVVGDFGLAKLMDYKDTHVTNH
ccccccccccccchhhhhhhhhhhhhhcccchhhhhhcccccccccc
```

Fig. 3. Prediction of secondary structure of SERK protein by PREDATOR. Most likely structural feature at each residue indicated by h (-----helices), e (β-sheets) or c (random coils).

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MERGGFHGYRKLPNNTTSGLKVPEMNMRMAEVNHTNNHSNSDDNECTVREQDRFMPIANVIRIMRKILPP
ccccccccccccccccccccchhhhhhhhhhhhcccccccccccccccceeecccccccchhhhhhhhhhcccc
HAKISDDAKETIQECVSEYISFITGEANERCQREQRKTTTAEDVLWAMSKLGFDDYIEPLTVYLHRYREL
ccccccchhhhhhhhhhhhccceeecccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccchhhhhhhhh
EGERGSIRGEPLGKRAVDYGALSVAAFAPAFHMGHHHGFFGGGGAMGGYLKDASNAGSSQAAVANGEPFG
hhccccccccccccccccccccccccchhhhhhhhccccccccccccccccccccccccccceeecccccccc
HHK
ccc
```

Fig. 4. Prediction of secondary structure of LEC protein by PREDATOR. Most likely structural feature at each residue indicated by h (-----helices), e (β-sheets) or c (random coils).

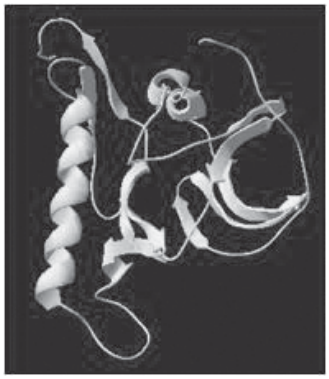

Fig. 5. Modeled structure of SERK protein

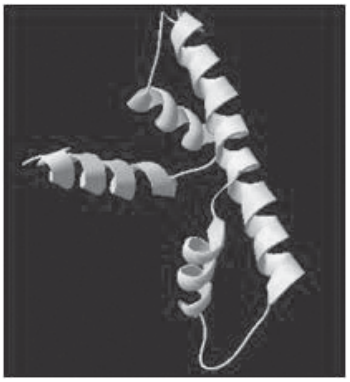

Fig. 6. Modeled structure of LEC protein

Acknowledgement

This work was supported by a grant from Department of Biotechnology (BTISnet), New Delhi, India.

REFERENCES

1. Alemanno, L., Berthouly, M., and Michaux-Ferrière, N. 1997. A comparison between *Theobroma cacao* L. zygotic and somatic embryogenesis from floral explants. *In Vitro Cell Dev. Biol. Plant* 33:163-172.
2. Alemanno, L., Devic, M., Niemenak, N., Sanier, C., Guilleminot, J., Rio, M., Verdeil, J.L. and Montoro, P. 2008. Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L. *Planta* 227:853-66.
3. Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C-M., van Lammeren, A.A.M., Miki, B.L.A., Custers, J.B.M. and van Lookeren Compagne, M.M. 2002. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14:1737-1749.
4. Chugh, A. and Khurana, P. 2002. Gene expression during somatic embryogenesis-recent advances. *Current Science* 83: 715-730.
5. Esan, E.B. 1974. Development of adventive embryos from immature embryos of *Theobroma cacao* L. culture *in vitro. 10th Annual Conference of Agricultural Society Nigeria*. 43 p.
6. Eskes, B. 2000. Introductory notes. International Workshop on New Technologies and Cocoa Breeding. Kota Kinabalu, Malaysia. pp. 8-11.
7. Feher, A., Pasternak, T.P. and Dudits, D. 2003. Transition of somatic plant cells to an embryogenic state. *Plant Cell Tiss. Org. Cult.* 74: 201-228.
8. Harada, J.J. 2001. Role of *Arabidopsis* LEAFY COTYLEDON genes in seed development. *J. Plant Physiol.* 158: 405-409.
9. Harding, E.W., Tang, W., Nichols, K.W., Fernandez, D.E. and Perry, S.E. 2003. Expression and maintenance of embryogenic potential is enhanced through constitutive expression of *AGAMOUS*-like15. *Plant Physiol.* 133: 653-663.
10. Hecht, V., Vielle-Calzada, J-P., Hartog, M. V., Schmidt, D.L., Boutilier, K., Grossniklaus, U. and de Vries, S.C. 2001. The Arabidopsis Somatic Embryogenesis Receptor Kinase 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiol.* 127: 803-816.
11. Irizarry, H. and Rivera, E, 1998. Early yield of five cacao families at three locations in Puerto Rico. *J. Agric.* 82: 163-171.
12. Kyte, J., and Doolittle, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157: 105-132.
13. 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 B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci. USA.* 98: 11806-11811.
14. Zuo, J., Niu, Q.W., Frugis, G. and Chua, N.H. 2002. The WUSCHEL gene promotes vegetative-to-embryonic transition in *Arabidopsis. Plant J.* 30: 349-359.