Analysis of common bean (*Phaseolus vulgaris L.*, genotype BAT93) calmodulin cDNA using computational tools

Kassim Amelia[1,2], Jasvin Singh[2], Farida Habib Shah[1,3], Subhash J. Bhore[1,2]

[1]Department of Molecular Biology, Melaka Institute of Biotechnology, Lot 7, Melaka International Trade Centre City, 75450 Ayer Keroh, Melaka, [2]Department of Biotechnology, Faculty of Applied Sciences, AIMST University, Semeling 08100, Kedah, [3]Department of Research and Development, Novel Plants Sdn. Bhd., 27C Jln Petaling Utama 12, 7.5 Miles Old Klang Road, 46000 Petaling Jaya, Malaysia

Submitted: 14-06-2014 Revised: 12-08-2014 Published: 02-02-2015

ABSTRACT

Background: Common bean (*Phaseolus vulgaris L.*) is an important part of the human diet and serves as a source of natural products. Identification and understanding of genes in *P. vulgaris* is important for its improvement. Characterization of expressed sequence tags (ESTs) is one of the approaches in understanding the expressed genes. For the understanding of genes expression in *P. vulgaris* pod-tissue, research work of ESTs generation was initiated by constructing cDNA libraries using 5-day and 20-day old bean-pod-tissues. Altogether, 5972 cDNA clones were isolated to have ESTs. While processing ESTs, we found a transcript for calmodulin (CaM) gene. It is an important gene that encodes for a calcium-binding protein and known to express in all eukaryotic cells. Hence, this study was undertaken to analyse and annotate it. **Objective:** The objective of this study was to analyze and annotate *P. vulgaris* CaM (*Pv*CaM) gene cDNA and its deduced protein (amino acids) sequence. **Materials and Methods:** Both strands of *Pv*CaM cDNA clone were sequenced using M13 forward and reverse primer to elucidate the nucleotide sequence. The cDNA sequence and deduced protein sequence were analyzed and annotated using bioinformatics tools available online. The secondary structures and three-dimensional (3D) structure of *Pv*CaM protein were predicted using the Phyre automatic fold recognition server. **Results:** Results showed that *Pv*CaM cDNA is 818 bp in length. The cDNA analysis results showed that it contains an open reading frame that encodes for 149 amino acid residues. The deduced protein sequence analysis results showed the presence of conserved domains required for CaM function. The predicted secondary structures and 3D structure are analogous to the *Solanum tuberosum* CaM. **Conclusions:** This study analyzed and annotated *Pv*CaM cDNA and protein. However, in order to obtain a complete understanding of *Pv*CaM protein, further study on its expression, structure and regulation is essential.

Access this article online

Website: www.phcogres.com

DOI: 10.4103/0974-8490.150536

Quick Response Code:

Key words: cDNA, expressed sequence tags, nutrition, phaseomics, proteins

INTRODUCTION

The animal products such as eggs, meat and milk are sources of dietary protein. But, legumes are a well-known and widely used as a source of dietary proteins, particularly by poor people.[1] By understanding the importance of common beans (*Phaseolus vulgaris* L.) in food supply chain, the Phaseomics international consortium was developed to establish the necessary framework of knowledge and materials for the advancement of bean genomics, transcriptomics, and proteomics. The major goal of this Phaseomics international consortium is to help in generating new common bean varieties suitable and desired by farmers and consumers.[2] As a part of this consortium, research work was undertaken for the generation of *P. vulgaris* expressed sequence tags (ESTs).[3]

So far, 5972 ESTs has been generated, and while processing and analyzing generated ESTs, calmodulin (CaM) EST was identified. CaM is known to play an important regulatory role in a bimodular mechanism of calcium

Address for correspondence:
Dr. Bhore Subhash J, Department of Biotechnology, Faculty of Applied Sciences, AIMST University, Bedong-Semeling Road, Semeling 08100, Kedah, Malaysia.
E-mail: subhash@aimst.edu.my

control in eukaryotes.[4-6] Therefore, to elucidate the *P. vulgaris* CaM (*Pv*CaM) cDNA clone sequence, it was fully sequenced, and cDNA and deduced amino acid (protein) sequences were analyzed and annotated in this study using computational tools. The *Pv*CaM gene cDNA sequence, deduced protein sequence, predicted secondary structures and three-dimensional (3D) structure is reported in this paper.

MATERIALS AND METHODS

Phaseolus vulgaris L. (genotype BAT93) seeds were provided by Patricia Lariguet, Laboratoire de Biologie Moléculaire des Plantes Supérieures, Department of Plant Biology, University of Geneva, Geneva, Switzerland. Seed germination and seedlings maintenance was done as stated by Bhore *et al.*[3]

The *Pv*CaM cDNA clone was identified from the ESTs generated from 5-day-old (days after anthesis) bean-pod-tissue cDNA library. The cDNA library was constructed (unpublished data) using "CloneMiner cDNA library construction kit" obtained from Invitrogen Corporation.

Escherichia coli cells harboring recombinant plasmid with *Pv*CaM cDNA were cultivated in 10 ml Luria-Bertani (LB) medium supplemented with 40 μg/ml Kanamycin. Plasmid DNA was isolated and purified using Wizard® Plus SV Minipreps DNA purification system (Promega). Plus and minus strand of *Pv*CaM cDNA clone was sequenced using M13 (forward) and M13 (reverse) primer.[3]

The basic alignment analysis of cDNA sequence was carried out using blast (bl2seq) program available online at National Center for Biotechnology Information and the finalized cDNA sequence was analyzed using online bioinformatics tools. The similarity search was performed using blast programs (blastN and blastP). Bioinformatics tools available at JustBio (http://www.justbio.com/) were used to deduce the protein sequence, and to find out the general features of *Pv*CaM cDNA and deduced protein sequence.

The deduced *Pv*CaM protein sequence was used as a blast (blastP) input to find the most analogous protein sequence and or structure in protein data bank.[7] However, for the prediction of secondary structures and the 3D structure of *Pv*CaM, Phyre2, a free web-based service for protein structure prediction was used.[8]

RESULTS

Full-length *Pv*CaM cDNA clone was isolated from 5-day old bean-pod-tissue cDNA library. Plus (+) and minus (−) strands sequence was aligned and after elimination of the adaptor sequence, cDNA sequence was finalized. Analysis of the results showed that *Pv*CaM cDNA is 818 bp in length. Identity of cDNA was confirmed by analyzing nucleotide (cDNA) and deduced amino acid sequences. Annotated nucleotide and deduced protein sequence of *Pv*CaM is deposited in GenBank/DDBJ/EMBL under the accession number JX869966. The basic annotated features of cDNA nucleotide and deduced protein sequence are summarized in Table 1, and cDNA nucleotide sequence along with its deduced amino acid sequence is shown in Figure 1.

The amino acid sequence analysis results showed that *Pv*CaM protein is rich in glutamic acid (13.4%) and aspartic acid (12.1%). Results also showed that *Pv*CaM

Table 1: The basic features of *Pv*CaM cDNA and its deduced protein sequence

General features	*Pv*CaM
cDNA sequence	
Size, bp	818
Molecular weight (daltons)	253,227
5' UTR, bp	66
CDS	450
3' UTR, bp	302
Stop codon	TGA (UGA)
G+C content %	43
Protein sequence	
Length, amino acids	149
Molecular weight (dalton)	16845.73
Isoelectric point (theoretical)	4.12

*Pv*CaM=*Phaseolus vulgaris* L. calmodulin; CDS=Coding sequence; cDNA= complementary DNA; TGA/UGA= codon; UTR= Untranslated Region

```
  1 - cctcaatttccttctctttcgttttccctcaccttctcgcttcccaaacgccttctctct - 60
 61 - gctccaATGGCCGACCAGCTCACCGACGAACAGATCGCCGAGTTCAAGGAGGCCTTCAGC - 120
  1 -       M  A  D  Q  L  T  D  E  Q  I  A  E  F  K  E  A  F  S   - 18

121 - CTCTTCGACAAGGATGGCGATGGTTGCATTACTACCAAGGAACTTGGGACTGTAATGCGG - 180
 19 - L  F  D  K  D  G  D  G  C  I  T  T  K  E  L  G  T  V  M  R   - 38

181 - TCACTTGGGCAAAATCCAACTGAGGCAGAGCTGCAGGACATGATAAATGAGGTTGATGCT - 240
 39 - S  L  G  Q  N  P  T  E  A  E  L  Q  D  M  I  N  E  V  D  A   - 58

241 - GATGGCAACGGAACTATCGATTTCCCAGAGTTCCTGAACCTGATGGCTCGCAAGATGAAA - 300
 59 - D  G  N  G  T  I  D  F  P  E  F  L  N  L  M  A  R  K  M  K   - 78

301 - GACACTGATTCGGAGGAGGAGCTTAAGGAAGCTTTCCGTGTGTTCGACAAGGATCAGAAT - 360
 79 - D  T  D  S  E  E  E  L  K  E  A  F  R  V  F  D  K  D  Q  N   - 98

361 - GGTTTTATCTCTGCTGCTGAGCTCCGCCATGTTATGACAAATCTGGGCGAGAAGCTGACC - 420
 99 - G  F  I  S  A  A  E  L  R  H  V  M  T  N  L  G  E  K  L  T   - 118

421 - GACGAGGAGGTTGATGAGATGATTCGCGAGGCTGATGTTGATGGCGATGGGCAGATCAAC - 480
119 - D  E  E  V  D  E  M  I  R  E  A  D  V  D  G  D  G  Q  I  N   - 138

481 - TATGAGGAGTTTGTCAAAGTGATGATGGCCAAGTGAgggtaatgaagcataactaaaatt - 540
139 - Y  E  E  F  V  K  V  M  M  A  K  *                           - 149

541 - tgaaaagcatacaagaggacagggcagataagttttgttgggaaatctatttcattaagt - 600
601 - ccatcttgattggggttattaacttttgttagttgagtggccttttttctggtgcttaat - 660
661 - gctactacgcccctaatatccactggttgcttgtcctggtcctaataaccgttgtcggct - 720
721 - catcctggtttataacttccgttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa - 780
781 - aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa - 818
```

Figure 1: Nucleotide and deduced amino acid sequence of *Phaseolus vulgaris* L. calmodulin cDNA (GenBank Accession No: JX869966). An open reading frame (ORF) and noncoding regions are shown in capital and small letters, respectively. The deduced amino acid sequence is given below the nucleotide sequence, which is numbered at both ends of each sequence line. The ORF encodes for a protein containing 149 amino acid residues. Initiation and termination codons are shown in green and red colour, respectively; *represents stop codon

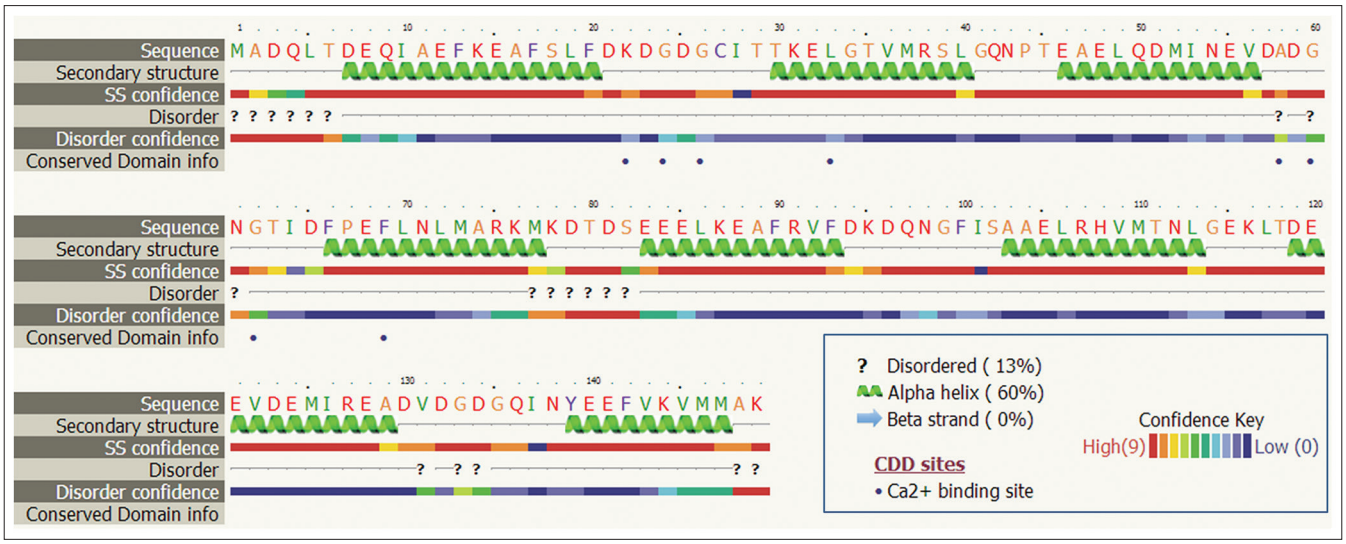

Figure 2: The predicted secondary structures of *Phaseolus vulgaris* L. calmodulin (*PvCaM*) protein

does contain less (<1%) cysteine, histidine and tyrosine. Interestingly, there was not a single residue of the tryptophan in *Pv*CaM protein. BlastP (domain enhanced lookup time accelerated basic local alignment search tool) results showed the presence of putative conserved domains in *Pv*CaM protein.

The topology of *Pv*CaM protein to show secondary structures is depicted in Figure 2. The predicted 3D structure produced for *Pv*CaM protein by homology modeling using Phyre2 is shown in Figure 3.

DISCUSSION

The full-length gene or its cDNA is required for the over-expression of gene in order to increase either the production of a desired important protein or natural products.[9] To understand the secondary and tertiary structural features of the proteins, molecular modeling is commonly used.[10-12] The main goal of this brief-study was to annotate *Pv*CaM gene cDNA and its deduced protein (amino acid) sequence. The *Pv*CaM cDNA clone was isolated from 5-day-old-pod tissue cDNA library; hence, it reflects that it is expressed in bean's 5-day-old developing-pod-tissue. However, its level of expression and its pattern of expression are not clear at this moment as we have not characterised its expression.

The guanine-cytosine (GC) content in *Pv*CaM cDNA is 43%. It is close to, but significantly higher than the GC content (39.4%) reported in nuclear DNA of broad bean.[13] BlastP results showed the presence of EFh (EF-hand) domains as specific hits. This EFh, calcium binding motif is a diverse superfamily of calcium sensors and calcium

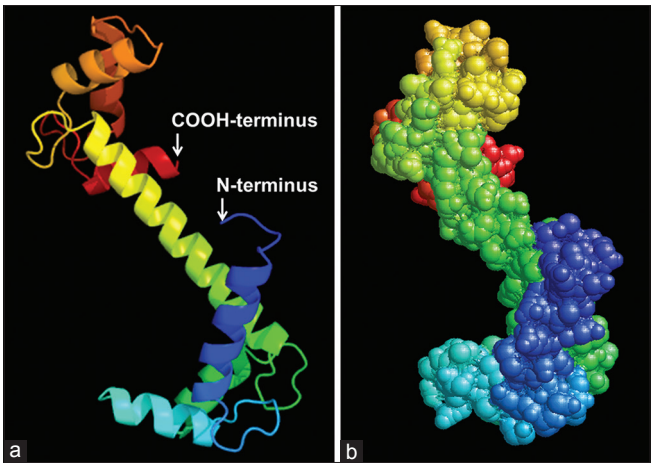

Figure 3: The predicted three-dimensional (3D) structure of *Phaseolus vulgaris* L. calmodulin (*PvCaM*) protein; (a) protein ribbon 3D structure model; (b) molecular surface 3D structure of model shown in (a)

signal modulators. Ca2+ binding induces a conformational change in the EF-hand motif, leading to the activation or inactivation of target proteins.[14-16]

Results suggest that the secondary structures of *Pv*CaM protein are mainly alpha helices (60%) [Figure 2]. The predicted 3D structure of the *Pv*CaM protein is based on the best template, 1RFJ. This template is of potato (*Solanum tuberosum*) CaM protein which shows the highest (97%) similarity (figure not shown) with *Pv*CaM protein when compared with other templates available in a database of protein.[17] The reported, potato CaM structure was determined using X-ray diffraction data in the resolution range 8.0–2.0 A.[18] Therefore, we strongly believe that the 3D structure predicted for *Pv*CaM protein in this study should be closer to its real structure. However, we suggest the further wet lab experimental work to validate the predicted structure.

Calmodulin is an important protein; because it decodes Ca^{2+} -dependent and-independent signals and could be helpful in in-depth understanding of the biological molecular mechanisms and signals.[19-22]

However, very little is known about *Pv*CaM. Therefore, further research is required to understand more about *Pv*CaM protein.

CONCLUSION

This brief-study has annotated the basic features of *Pv*CaM gene cDNA and deduced protein. Comparative molecular modelling suggests that the deduced *Pv*CaM protein is analogous to Potato CaM protein. However, in order to have a comprehensive understanding of *Pv*CaM protein further studies are required to validate the predicted 3D structure, and to understand its expression and regulation in beans.

REFERENCES

1. Massey LK. Dietary animal and plant protein and human bone health: A whole foods approach. J Nutr 2003;133:862S-5.
2. Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J. Beans (*Phaseolus* spp.) – Model food legumes. Plant Soil 2003;252:55-128.
3. Bhore SJ, Amelia K, Wang E, Priyadharsini S, Shah FH. Computational analysis of common bean (*Phaseolus vulgaris* L. genotype BAT93) lycopene ß-cyclase and ß-carotene hydroxylase gene's cDNA. Bioinformation 2013;9:197-206.
4. Tidow H, Poulsen LR, Andreeva A, Knudsen M, Hein KL, Wiuf C, *et al.* A bimodular mechanism of calcium control in eukaryotes. Nature 2012;491:468-72.
5. Strehler EE, Filoteo AG, Penniston JT, Caride AJ. Plasma-membrane Ca(2+) pumps: Structural diversity as the basis for functional versatility. Biochem Soc Trans 2007;35:919-22.
6. Ishida H, Vogel HJ. The solution structure of a plant calmodulin and the CaM-binding domain of the vacuolar calcium-ATPase BCA1 reveals a new binding and activation mechanism. J Biol Chem 2010;285:38502-10.
7. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al.* The Protein Data Bank. Nucleic Acids Res 2000;28:235-42.
8. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: A case study using the Phyre server. Nat Protoc 2009;4:363-71.
9. Kudo T, Makita N, Kojima M, Tokunaga H, Sakakibara H. Cytokinin activity of cis-zeatin and phenotypic alterations induced by overexpression of putative cis-Zeatin-O-glucosyltransferase in rice. Plant Physiol 2012;160:319-31.
10. Kountouris P, Agathocleous M, Promponas VJ, Christodoulou G, Hadjicostas S, Vassiliades V, *et al.* A comparative study on filtering protein secondary structure prediction. IEEE/ACM Trans Comput Biol Bioinform 2012;9:731-9.
11. Wang E, Chinni S, Bhore SJ. Three-dimensional (3D) structure prediction of the American and African oil-palms ß-ketoacyl-[ACP] synthase-II protein by comparative modelling. Bioinformation 2014;10:130-7.
12. Sehar U, Mehmood MA, Nawaz S, Nadeem S, Hussain K, Sohail I, *et al.* Three dimensional (3D) structure prediction and substrate-protein interaction study of the chitin binding protein CBP24 from B. thuringiensis. Bioinformation 2013;9:725-9.
13. Baxter R, Kirk JT. Base composition of DNA from chloroplasts and nuclei of *Phaseolus vulgaris*. Nature 1969;222:272-3.
14. Kretsinger RH. EF-hands embrace. Nat Struct Biol 1997;4:514-6.
15. Ikura M. Calcium binding and conformational response in EF-hand proteins. Trends Biochem Sci 1996;21:14-7.
16. Akke M, Forsén S, Chazin WJ. Solution structure of (Cd2+) 1-calbindin D9k reveals details of the stepwise structural changes along the Apo- and gt;(Ca2+) II1- and gt;(Ca2+) I, II2 binding pathway. J Mol Biol 1995;252:102-21.
17. Fox NK, Brenner SE, Chandonia JM. SCOPe: Structural Classification of Proteins – Extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Res 2014;42:D304-9.
18. Yun CH, Bai J, Sun DY, Cui DF, Chang WR, Liang DC. Structure of potato calmodulin PCM6: The first report of the three-dimensional structure of a plant calmodulin. Acta Crystallogr D Biol Crystallogr 2004;60:1214-9.
19. Villarroel A, Taglialatela M, Bernardo-Seisdedos G, Alaimo A, Agirre J, Alberdi A, *et al.* The ever changing moods of calmodulin: How structural plasticity entails transductional adaptability. J Mol Biol 2014;426:2717-35.
20. Jia L, Chu H, Wu D, Feng M, Zhao L. Role of calmodulin in thermotolerance. Plant Signal Behav 2014;9: pii: e28887. [Epub ahead of print]
21 Tang XH, Chen J, Yang XL, Yan LF, Gao J. Preservation on calcium homeostasis is involved in mitochondrial protection of *Limonium sinense* against liver damage in mice. Pharmacogn Mag 2010;6:191-7.
22 Saini AS, Taliyan R, Sharma PL. Protective effect and mechanism of *Ginkgo biloba* extract-EGb 761 on STZ-induced diabetic cardiomyopathy in rats. Pharmacogn Mag 2014;10:172-8.

Cite this article as: Amelia K, Singh J, Shah FH, Bhore SJ. Analysis of common bean (*Phaseolus vulgaris L.*, genotype BAT93) calmodulin cDNA using computational tools. Phcog Res 2015;7:209-12.

Source of Support: Nil, **Conflict of Interest:** None declared.