

Nucleotide sequence of *Phaseolus vulgaris* L. alcohol dehydrogenase encoding cDNA and three-dimensional structure prediction of the deduced protein

Kassim Amelia^{1,2}, Chin Yin Khor², Farida Habib Shah³, Subhash J. Bhore^{1,2}

¹Department of Molecular Biology, Melaka Institute of Biotechnology, Lot 7, Melaka International Trade Centre City, 75450 Ayer Keroh, Melaka, Malaysia, ²Department of Biotechnology, Faculty of Applied Sciences, AIMST University, Bedong-Semeling Road, Semeling 08100, Kedah,

³Department 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: 10-08-2014

Published: 02-02-2015

ABSTRACT

Background: Common beans (*Phaseolus vulgaris* L.) are widely consumed as a source of proteins and natural products. However, its yield needs to be increased. In line with the agenda of Phaseomics (an international consortium), work of expressed sequence tags (ESTs) generation from bean pods was initiated. Altogether, 5972 ESTs have been isolated. Alcohol dehydrogenase (AD) encoding gene cDNA was a noticeable transcript among the generated ESTs. This AD is an important enzyme; therefore, to understand more about it this study was undertaken. **Objective:** The objective of this study was to elucidate *P. vulgaris* L. AD (*PvAD*) gene cDNA sequence and to predict the three-dimensional (3D) structure of deduced protein. **Materials and Methods:** positive and negative strands of the *PvAD* cDNA clone were sequenced using M13 forward and M13 reverse primers to elucidate the nucleotide sequence. Deduced *PvAD* cDNA and protein sequence was analyzed for their basic features using online bioinformatics tools. Sequence comparison was carried out using *bl2seq* program, and tree-view program was used to construct a phylogenetic tree. The secondary structures and 3D structure of *PvAD* protein were predicted by using the PHYRE automatic fold recognition server. **Results:** The sequencing results analysis showed that *PvAD* cDNA is 1294 bp in length. It's open reading frame encodes for a protein that contains 371 amino acids. Deduced protein sequence analysis showed the presence of putative substrate binding, catalytic Zn binding, and NAD binding sites. Results indicate that the predicted 3D structure of *PvAD* protein is analogous to the experimentally determined crystal structure of s-nitrosogluthathione reductase from an *Arabidopsis* species. **Conclusions:** The 1294 bp long *PvAD* cDNA encodes for 371 amino acid long protein that contains conserved domains required for biological functions of AD. The predicted deduced *PvAD* protein's 3D structure reflects the analogy with the crystal structure of *Arabidopsis thaliana* s-nitrosogluthathione reductase. Further study is required to validate the predicted structure.

Key words: BAT93, common bean, homology modeling, molecular modeling, phaseomics, protein, protein structure prediction

INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is an important commodity in the food supply chain. It is consumed widely as it serves as a rich source of proteins, vitamins

and minerals important in balanced human diet.^[1,2] In the Asian countries, in Latin America and Africa the production of the beans is in the greater amount to meet the demand and consumption by the increasing population. Common bean is serving as a very important source of proteins for financially weak group of people and used as a model food legume.^[3] We must develop new varieties of common bean that are desired by farmers and consumers. To speed up this process, an international consortium (phaseomics) was established.^[3] As a part of this consortium, research work of generation and characterization of expressed

Access this article online

Website:

www.phcogres.com

DOI: 10.4103/0974-8490.150532

Quick Response Code:



Address for correspondence:

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

sequence tags (ESTs) for bean was initiated at Melaka Institute of Biotechnology, Malaysia.^[4] While processing ESTs, we found a cDNA clone for *P. vulgaris* L. alcohol dehydrogenase (*PvAD*).

Alcohol dehydrogenase (AD) encoding genes are found in all species of archaea, bacteria, fungi, plants and animals.^[5] AD belongs to AD family, a group of dehydrogenases that facilitate the interconversion of alcohols and aldehydes or ketones.^[6] This enzyme plays an important role in physiological processes such as alcohol and alkane metabolism, cell defense toward exogenous alcohols and aldehydes.^[7] AD is studied in some flowering plants and in few legumes.^[8] However, we do not know much about AD in *P. vulgaris*. Therefore, to elucidate the *PvAD* cDNA clone sequence, it was fully sequenced, and cDNA and deduced protein sequence was analyzed and annotated in this study using computational tools. The *PvAD* gene cDNA sequence, its deduced protein sequence, predicted secondary structures and three-dimensional (3D) structure is reported in this paper.

MATERIALS AND METHODS

Common bean (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 maintenance of seedlings was done as reported by Bhore *et al.*^[4]

The cDNA clone of *PvAD* was isolated and identified from the ESTs generated from 20 days old (days after anthesis) bean-pod-tissue cDNA library. This cDNA library was constructed (our unpublished data) using “CloneMiner cDNA library construction kit” obtained from Invitrogen Corporation.

Escherichia coli cells harboring recombinant plasmid with *PvAD* cDNA were cultivated in 10 ml LB medium supplemented with 40 µg/ml Kanamycin. Plasmid DNA was isolated and purified using Wizard® Plus SV Minipreps DNA purification system procured from Promega. Sense and antisense strand of *PvAD* cDNA clone were sequenced using M13 Forward and M13 Reverse primer.^[4]

The comparative analysis of cDNA sequence was performed using online BLASTN (bl2seq) program available at NCBI. The finalized cDNA sequence was analyzed using online bioinformatics tools. The similarity search was performed using BLASTN and BLASTP programs. Bioinformatics tools available at JustBio (<http://www.justbio.com/>) were used to deduce the protein sequence, and to find out the

general features of *PvAD* cDNA and deduced protein sequence. The tree-view program was used to construct a phylogenetic tree.

The deduced *PvAD* protein sequence was used as a BLASTP input to find the most analogous protein sequence and or structure in protein data bank (PDB).^[9] However, for the prediction of secondary structures and the 3D structure of *PvAD*, Phyre2, a free web-based service for protein structure prediction was used.^[10]

RESULTS

The *PvAD* cDNA clone isolated from 20 days old bean-pod-tissue cDNA library was sequenced for both strands. Sequence of sense (+) and antisense (−) strand was aligned and after elimination of the adaptor sequence, cDNA sequence was finalized. Our results indicate that isolated and sequenced *PvAD* cDNA is 1294 bp in length. The identity of cDNA sequence was confirmed by analyzing its nucleotide and deduced amino acid sequence. Annotated nucleotide and deduced protein sequence for *PvAD* is deposited in GenBank/DDBJ/EMBL under the accession number KF569659. The basic annotated features of cDNA and deduced protein sequence are summarized in Table 1, and cDNA sequence along with its deduced amino acid sequence is depicted in Figure 1.

The comparative analysis of *PvAD* protein shows 75–80% similarity with its counterparts from other plant species. A summary is shown in Table 2.

Analysis of the deduced protein sequence suggests that *PvAD* protein is rich in Glycine (9.7%) and Valine (9.16%). However, Glutamine, Methionine, Tryptophan, and Tyrosine amino acid content was <2%. Blastp (domain enhanced lookup time accelerated basic local alignment search tool) results showed the presence

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

General features	<i>PvAD</i>
cDNA sequence	
Size, bp	1294
Molecular weight (daltons)	401,353
5'UTR, bp	0
Coding sequence	1116
3'UTR, bp	178
Stop codon	TGA (UGA)
G+C content (%)	44
Protein sequence	
Length, amino acids	371
Molecular weight (dalton)	40,402.69
Isoelectric point (theoretical)	6.24

PvAD=*Phaseolus vulgaris* alcohol dehydrogenase; UTR=Untranslated region

1 -	ATGCGTGCAGCTGCGGTGGCATGGGAAGCAGGGAAGCCACTGTCAATAGAGAACATAGAAGTGGCAACCACGAGG	- 75
1 -	M R A A A V A W E A G K P L S I E N I E V A P P G	- 25
76 -	AAAGGAGAAGTGCCTTTGAAAATCCTCTTCAACTCCCTTTGCCGTACTGATGTTTACTGGTGGGATGTTAAGGGT	- 150
26 -	K G E V R L K I L F N S L C R T D V Y W W D V K G	- 50
151 -	CAGAATCCGCTGTTTCTCGAATATTAGGTCATGAAGCTTCAGGGATTGTGGAGAGCGTGGGCGAGGGAGTGACC	- 225
51 -	Q N P L F P R I L G H E A S G I V E S V G E G V T	- 75
226 -	CATCTGAAACAGGAGATCATGCACTCCCAATATTCAGTGGGAGTGTGGGAATGCAAACTGCAAGTCAGAA	- 300
76 -	H L K P G D H A L P I F T G E C G E C K Y C K S E	- 100
301 -	GAGAGCAACCTGTGTGAAGTGTAAAGATCAACAGTGACAGAGGAGTTATGCTGAGTGATGGCAAAACAAGGTTT	- 375
101 -	E S N L C E V L R I N S D R G V M L S D G K T R F	- 125
376 -	TCCAAAATGGACAACCCATTTACCACCTTTCTGGGAACCTCTACATTCAGTGAATACACTGTTCTTCATGCTGGC	- 450
126 -	S K N G Q P I Y H F L G T S T F S E Y T V L H A G	- 150
451 -	TGTGTTGCCAAGATCAACCTGCTGCTCCACTTGACAAAGTTTGTGTTGTCAGTTGTGGATTTGCACAGGTTT	- 525
151 -	C V A K I N P A A P L D K V C V V S C G F C T G F	- 175
526 -	GGTGTACTGTAAATGTTGCAAAACCAAGCCTAATGATACTGTTGCTGCTTTGGTTTAGGAGCTGTTGGTCTT	- 600
176 -	G A T V N V A K P K P N D T V A V F G L G A V G L	- 200
601 -	GCTGCCTGTGAAGGAGCAAGGGTTTCTGGTGCATCGAGAATCATTTGGGGTTGATTACTTCCTAGTCGATTGAA	- 675
201 -	A A C E G A R V S G A S R I I G V D L L P S R F E	- 225
676 -	CAAGCCAAACAATTGGGGTCACTGAATTTGTGAACCCCAAGAGCATAACAAGCCAGTTCAAGAGGTAATTGCT	- 750
226 -	Q A K Q F G V T E F V N P K E H N K P V Q E V I A	- 250
751 -	GAATGACTGATGGAGGAGTAGATCGTGCTATTGAATGCAGTGGAAACATCCAAGCTTCGATCTCAGCACTTGAA	- 825
251 -	E M T D G G V D R A I E C T G N I Q A S I S A L E	- 275
826 -	TGCACTCATGATGGTTGGGGTGTGCTGTACTTGTAGTGTCCAAAGAAGGATGCTGAATTCAAAACCTCATCT	- 900
276 -	C T H D G W G V A V L V S V P K K D A E F K T H P	- 300
901 -	ATGAAGTTTCATGAAGGAAGAAGTCTCAAGGGTACCTTCTATGGCCACTACAAACCCGCACTGATATTCCTTCT	- 975
301 -	M K F M E G R T L K G T F Y G H Y K P R T D I P S	- 325
976 -	GTTGTGGAGAAATACCTCAACAAGGAAGTGGAACTGGACAAATTTATCACTCACACTGTGCCATTTTCAGAGATC	- 1050
326 -	V V E K Y L N K E L E L D K F I T H T V P F S E I	- 350
1051 -	AACAAGGCATTTGACCTTATGCTGAGAGGGGAGGTATAAGGTGTTTCATCTGCATGGAAGAGTGAatcaattat	- 1125
351 -	N K A F D L M L R G E G I R C F I C M E E *	- 371
1126 -	gttttttgcatagagaacagtggtttgtgtatgtgataaaagagagtgaagataaattggcaogtaattc	- 1200
1201 -	tcttggtaactgtgcatatgagtggtgtaaaagagtgatgatgacttcttggatattgtcctgccaataa	- 1275
1276 -	aaaaaaaaaaaaaaaaaaaa	- 1294

Figure 1: Nucleotide and deduced amino acid sequence of *Phaseolus vulgaris* L. alcohol dehydrogenase cDNA clone (Accession No: KF569659). Open reading frame (ORF) and 3'non-coding region of the cDNA is shown in capital and small letters, respectively. Amino acids are represented by a single letter codes; both, nucleotide and deduced amino acid sequence is numbered at both ends of each sequence line. The ORF encodes for 371 amino acid residues. Amino acid residues are numbered beginning with the initial Methionine (M) until the last Glutamic acid (E) residue. Initiation and termination codons are shown in green and red color, respectively. Poly (A) signal (sequence) is highlighted in red color; *denote the termination codon

of putative conserved domains in *PvAD* protein. The *PvAD* protein based phylogenetic analysis results are shown in Figure 2.

The topology of *PvAD* protein to show the predicted secondary structures is shown in Figure 3. Whereas, the predicted 3D structure produced for *PvAD* protein by comparative molecular modeling using Phyre2 is shown in Figure 4.

DISCUSSION

The full length gene or its cDNA is essential for the over-expression of the gene of interest in order to increase either the production of a desired protein or natural products in the plants by using genetic engineering techniques.^[11] However, for the basic understanding of the gene (or its cDNA) structure, and secondary and

tertiary structural features of the inferred proteins various computational tools and molecular modeling is commonly used.^[12-14] In this study, the main goal was to annotate *PvAD* gene cDNA and its deduced protein sequence. The *PvAD* cDNA clone was isolated from 20 days old-pod tissue cDNA library, an indication of its expression in bean's 20 days old developing-pod-tissue. However, its level of expression and its expression regulation is not understood in beans (genotype BAT93) as we have not characterized its expression.

The GC content in *PvAD* cDNA is 44%. This much GC content is close to, but significantly higher than the GC content (39.4%) reported in nuclear DNA of broad bean.^[15] The isolated *PvAD* cDNA is truncated; hence, 5' untranslated region is missing from its sequence [Table 1 and Figure 1]. Protein analysis results showed the presence of NAD binding site (chemical binding), catalytic

Zn binding site (ion binding), and substrate binding site (chemical binding) those are essential for the biological functions of the AD.^[16-18]

The results of phylogenetic analysis indicate that *PvAD* protein is closely (80%) related to *Phaseolus acutifolius*, *Lotus corniculatus*, *Lotus japonicus* and *Rosa rugosa* AD protein. On the contrary, AD from *Mangifera indica* showed less (75%) similarity with *PvAD* protein [Table 2 and Figure 2].

Deduced protein sequence analysis results also suggest that *PvAD* protein contains 12 (27%) alpha helices and 18 (29%) beta strands [Figure 3]. The predicted secondary structures and 3D structure of the *PvAD* protein is based on the best template, 3uko. This template is of *Arabidopsis thaliana* s-nitrosoglutathione reductase (protein), which showed

the highest (56%) identity (figure not shown) with *PvAD* protein.^[19] The reported *A. thaliana* s-nitrosoglutathione reductase structure was determined by using X-ray diffraction method (resolution: 1.40Å) (DOI: 10.2210/PDB3uko/PDB). Of 371 amino acids, 369 residues (99%) have been modeled with 100% confidence by using the single highest scoring template (3uko). There was no protein structure in PDB which shows more than 56% identity with *PvAD* protein; though *PvAD* protein's phylogenetic analysis shows maximum (80%) similarity with AD protein from *P. acutifolius*, *L. corniculatus*, *L. japonicus* and *R. rugosa* AD the protein.^[20] However, we strongly believe that the 3D structure predicted for *PvAD* protein in this study could be closer to its real structure based on the confidence level (key) of the prediction [Figures 3 and 4].^[10] Yet, we suggest the further wet-lab experimental work is essential to

Table 2: Comparison of *PvAD* cDNA nucleotide and deduced amino acid sequence with its counterparts from other plant species

Species	Gene bank accession number	Length compared		Similarity (%)	
		Nucleotide (bp)	Amino acid	Nucleotide (bp)	Amino acid
<i>Arabis alpina</i>	AF110429	1624	379	71	76
<i>Capsella rubella</i>	XM_006302340	2860	379	74	76
<i>Citrus clementina</i>	XM_006434320	1674	376	76	77
<i>Cucumis melo</i>	DQ288986	1502	379	76	77
<i>Dianthus caryophyllus</i>	AY263389	1522	380	74	77
<i>Diospyros kaki</i>	JF357957	1459	379	74	78
<i>Eutrema salsugineum</i>	XM_006390060	1550	379	72	76
<i>Fourraea alpina</i>	AF110451	1789	379	71	76
<i>Gossypium hirsutum</i>	U53701	2626	379	75	79
<i>Lotus corniculatus</i>	AY227202	1429	380	81	80
<i>Lotus japonicus</i>	JN165714	1143	380	81	80
<i>Mangifera indica</i>	GU233767	1329	382	77	75
<i>Medicago truncatula</i>	XM_003602081	1143	380	80	78
<i>Nicotiana tabacum</i>	AY619947	1439	380	75	77
<i>Phaseolus acutifolius</i>	Z23170	1399	380	82	80
<i>Populus trichocarpa</i>	XM_002302159	1556	380	76	77
<i>Prunus persica</i>	XM_007200991	1378	379	78	79
<i>Pseudotsurritus turrita</i>	AF110457	1704	379	71	77
<i>Pyrus communis</i>	HQ912034	2290	380	76	77
<i>Quercus suber</i>	KF704745	1146	381	75	77
<i>Rosa rugosa</i>	KF724973	1337	389	78	80
<i>Theobroma cacao</i>	XM_007019324	1883	379	76	79
<i>Zea mays</i>	AF123535	160480	379	72	76

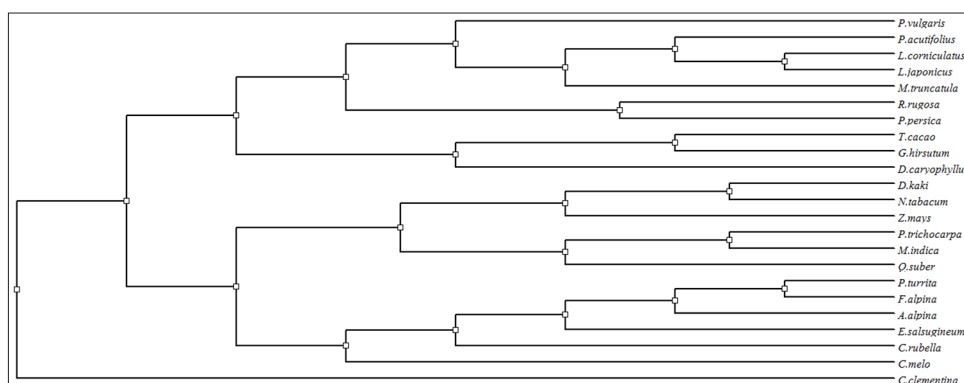


Figure 2: Phylogenetic relationship of *Phaseolus vulgaris* L. alcohol dehydrogenase protein with its counterparts from other plant species



Figure 3: Predicted secondary structures of *Phaseolus vulgaris* L. alcohol dehydrogenase protein

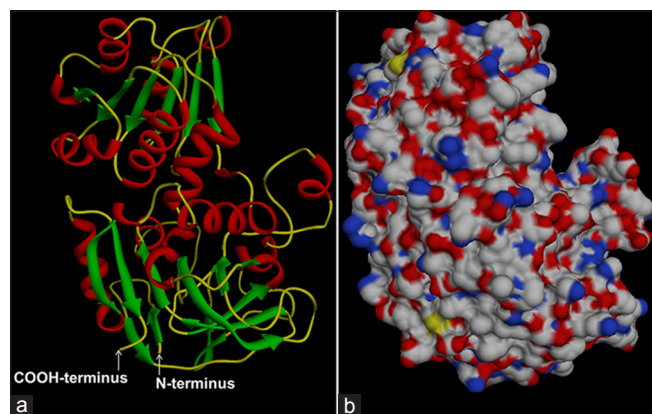


Figure 4: Predicted three-dimensional (3D) structure of *Phaseolus vulgaris* L. alcohol dehydrogenase protein; (a) protein ribbon 3D structure model; red, green and yellow color shows the helices, strands and coils (loops) of protein, respectively; (b) molecular surface 3D structure of model shown in (a)

validate the predicted structure. Therefore, further research is necessary to understand more about *PvAD* protein.

CONCLUSION

The basic features of *PvAD* gene cDNA and deduced protein are successfully elucidated in this study. Comparative

molecular modeling suggests that the deduced *PvAD* protein is analogous to *A. thaliana* s-nitrosogluthathione reductase protein. But, in order to have a full understanding of *PvAD* protein, further research to validate the predicted 3D structure, and to understand its expression and regulation in beans is required.

ACKNOWLEDGMENT

Authors are grateful to the Ministry of Science, Technology and Innovation (MOSTI), Malaysia for research funding [Research Grant Code Number: BSP (M)/BTK/004 (3)]; and to Patricia Lariguet, Department of Plant Biology, University of Geneva, Geneva, Switzerland for supplying seeds of bean, genotype BAT93.

REFERENCES

1. Neithercott T. The balanced diet: What it means and why it's so important. *Diabetes Forecast* 2011;64:36-45, 57.
2. 5 tips to build muscle strength. Proper sleep, balanced diet, and strengthening exercises can boost endurance and ensure independence. *Harv Health Lett* 2014;39:6.
3. Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J. Beans (*Phaseolus* spp.) – Model food legumes. *Plant Soil* 2003;252:55-128.

4. Bhore SJ, Amelia K, Wang E, Priyadharsini S, Shah FH. Computational analysis of common bean (*Phaseolus vulgaris* L. genotype BAT93) lycopene β -cyclase and β -carotene hydroxylase gene's cDNA. *Bioinformation* 2013;9:197-206.
5. Littlechild JA, Guy JE, Isupov MN. Hyperthermophilic dehydrogenase enzymes. *Biochem Soc Trans* 2004;32:255-8.
6. Natarajan SS, Pastor-Corrales MA, Khan FH, Garrett WM. Proteomic analysis of common bean (*Phaseolus vulgaris* L.) by two-dimensional gel electrophoresis and mass spectrometry. *J Basic Appl Sci* 2013;9:424-37.
7. Burdette DS, Jung SH, Shen GJ, Hollingsworth RI, Zeikus JG. Physiological function of alcohol dehydrogenases and long-chain (C (30)) fatty acids in alcohol tolerance of *Thermoanaerobacter ethanolicus*. *Appl Environ Microbiol* 2002;68:1914-8.
8. Fukuda T, Yokoyama J, Nakamura T, Song IJ, Ito T, Ochiai T, et al. Molecular phylogeny and evolution of alcohol dehydrogenase (Adh) genes in legumes. *BMC Plant Biol* 2005;5:6.
9. 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.
10. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: A case study using the Phyre server. *Nat Protoc* 2009;4:363-71.
11. 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.
12. 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.
13. Wang E, Chinni S, Bhore SJ. Three-dimensional (3D) structure prediction of the American and African oil-palms β -ketoacyl-[ACP] synthase-II protein by comparative modelling. *Bioinformation* 2014;10:130-7.
14. 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.
15. Baxter R, Kirk JT. Base composition of DNA from chloroplasts and nuclei of *Phaseolus vulgaris*. *Nature* 1969;222:272-3.
16. Kang C, Hayes R, Sanchez EJ, Webb BN, Li Q, Hooper T, et al. Furfural reduction mechanism of a zinc-dependent alcohol dehydrogenase from *Cupriavidus necator* JMP134. *Mol Microbiol* 2012;83:85-95.
17. Cheng S, Fan C, Sinha S, Bobik TA. The PduQ enzyme is an alcohol dehydrogenase used to recycle NAD⁺ internally within the Pdu microcompartment of *Salmonella enterica*. *PLoS One* 2012;7:e47144.
18. Bomati EK, Noel JP. Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell* 2005;17:1598-611.
19. 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.
20. Zambre M, Van Montagu M, Angenon G, Terryn N, Tepary bean (*Phaseolus acutifolius*). *Methods Mol Biol* 2006;343:407-14.

Cite this article as: Amelia K, Khor CY, Shah FH, Bhore SJ. Nucleotide sequence of *Phaseolus vulgaris* L. alcohol dehydrogenase encoding cDNA and three-dimensional structure prediction of the deduced protein. *Phcog Res* 2015;7:203-8.

Source of Support: Ministry of Science, Technology and Innovation (MOSTI), Malaysia for research funding [Research Grant Code Number: BSP (M)/BTK/004 (3)]. **Conflict of Interest:** None declared.