Kasetsart J. (Nat. Sci.) 40 : 123 - 128 (2006)

Nucleotide and Derived Amino Acid Sequences of the Cyanogenic Beta-Glucosidase (Linamarase) from Cassava (*Manihot esculenta* Crantz)

Prachumporn Toonkool* and Nusra Tongtubtim

ABSTRACT

Many isozymes of cassava cyanogenic β-glucosidase (linamarase) exist, but only one cDNA sequence (pCAS5) has been reported thus far. In order to study the structure-function relationships in this enzyme, the cDNA of cassava linamarase was cloned and sequenced. In this report, six different cDNA sequences of linamarase from cassava were cloned by reverse transcription-polymerase chain reaction (RT-PCR) using primers designed from the sequence of pCAS5. Nucleotide sequences of all six clones showed 98-99% sequence identity to the nucleotide sequence of pCAS5. Derived amino acid sequences from four cDNA clones showed 98-99% sequence identity to that of pCAS5, while the other two cDNA clones contained nucleotide sequences that led to premature termination.

Key words: beta-glucosidase, linamarase, cassava, nucleotide sequence, amino acid sequence

INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is an important crop due to the large storage roots which are a staple carbohydrate source for many populations worldwide. Young cassava leaves are also consumed by humans or used as animal feeds (Yeoh, 1989). However, both leaves and roots of cassava are cyanogenic, posing a potential health hazard to consumers. Cyanogenesis in cassava involves two cyanogenic glucosides (linamarin and lotaustralin) and sequential action of two enzymes (β-glucosidase and hydroxynitrile lyase). Normally, the glucoside substrates and the enzymes are kept in separate compartments. Exposure of these two cyanogenic glucosides to β-glucosidase, following tissue damage, results in their hydrolysis to yield glucose and an aglycone, which is acetone cyanohydrin or butanone cyanohydrin. These cyanohydrins may be broken down spontaneously or enzymatically by α-hydroxynitrile lyase to produce hydrogen cyanide (Eksittikul and Chulavatanatol, 1988; Hughes *et al*., 1992). The cyanogenic β-glucosidase, or linamarase, which is responsible for the first step of cassava cyanogenesis, is present in many parts of the plant, including petioles, stem, leaves, peels and tuber cortex. Physiologically, the enzyme plays an important role as a host defense against biological predators. Furthermore, an *in vitro* study has discovered the unique ability of cassava linamarase in alkyl glucoside synthesis by transferring glucose to a large variety of alcohol acceptors, while β-glucosidases from Thai rosewood and almond could not (Svasti *et al*., 2003). This difference in enzymatic action may

Department of Biochemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.
* Corresponding author, email: fscippt@ku.ac.th

lie in slight variations in the amino acid sequences of β-glucosidases from various sources. So, it is of scientific interest to understand the molecular basis of enzymatic catalysis in cassava linamarase. To achieve this, cassava linamarase must be cloned and expressed as a recombinant protein, its nucleotide sequence mutated, and the effects of mutation characterized. These procedures will help to identify amino acid residues responsible for transglucosylation activity.

A number of studies have shown the existence of multiple forms of cassava linamarase (Eksittikul and Chulavatanatol, 1988; Yeoh, 1989), however, only a single cDNA clone of linamarase (pCAS5) from cassava has been isolated thus far (Hughes *et al*., 1992). The identity of other cassava linamarase isozymes has not been reported. In this study, we reported the isolation of six different cDNA clones of cassava linamarase, and the amino acid sequences derived from four full-length cDNA sequences. The predicted polypeptides were very similar to that of the previously reported cassava linamarase, so they can be reasonably assumed to be catalytically active in cassava.

MATERIALS AND METHODS

Total RNA was isolated from young cassava leaves (local varieties) using RNeasy extraction kit (QIAGEN, USA) and concentration estimated by absorbance at 260 nm. Total RNA (0.2-5 μg) was converted into single stranded first strand cDNA fragment using 200 U SuperScript II reverse transcriptase (Invitrogen, USA) in the reaction containing 0.5 μg oligo$(dT)_{12\text{-}18}$ primer, 10 nmol dNTP mix, and 10 mM DTT at 42 °C for 50 min and at 70 °C for 15 min. The RNA strand of the RNA:cDNA duplex was degraded by 2 U RNase H at 37 °C for 20 min. The 1 μL aliquot of the product was used as a template for PCR with 1 μM each of CVLMF2 (5' TTC TTC AGC TAT CAG GGA TGC 3') and CVLMR2 (5' TGC TAG ATC ATT GGA GCT TCA 3') primers, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X-100, 0.2 mM of each dNTP and 2.5 U Expand HiFi DNA polymerase (Boehringer, Germany) in a total volume of 50 μL. All PCR reactions were done using a 94 °C initial denaturation for 2 min, 29 cycles of 94 °C for 60 s, 45 °C for 30 s, and 72 °C for 2 min, and a 72 °C final elongation for 8.5 min. The primer sequences correspond to bases 7-27 (sense) and bases 1677-1697 (antisense), respectively, of the cassava linamarase cDNA clone pCAS5 reported previously (Hughes *et al*., 1992). The resulting RT-PCR product was gel-purified, ligated into pGEM-Teasy vector (Promega, USA) and transformed into *E. coli* DH5α.

The nucleotide sequence of cDNA clones were determined by automated DNA sequencing (ABI 373A DNA sequencer). The nucleotide sequences obtained were translated into amino acid sequences using EMBOSS Transeq program. Alignment of nucleotide and amino acid sequences were done using ClustalW version 1.82. The physical and chemical parameters of the derived amino acid sequences were performed using ProtParam program.

RESULTS AND DISCUSSION

Cloning of cDNA of cassava linamarase

Total RNA was extracted from young cassava leaves. Approximately 500 ng total RNA was used to synthesized the first strand cDNA. PCR amplification of the first strand cDNA yielded 1.7 kb PCR product (Figure 1), which was then ligated into pGEM-Teasy vector and transformed into *E. coli* DH5α. The nucleotide sequences of cloned RT-PCR products were determined by automatic sequencing, yielding six different cDNA clones, namely 2A.29, 2A.32, 2A.33, 2A.52, 2A.55 and 2A.62. Alignment of the nucleotide sequences from all six cDNA clones showed 98-99% sequence identity to the previously reported nucleotide sequence of the cassava linamarase

cDNA clone pCAS5 (Hughes *et al*., 1992), and to each other (results not shown).

Alignment of nucleotide and derived amino acid sequences of cassava linamarase

The nucleotide sequences from 6 cDNA clones were translated and aligned. A thymine deletion in clone 2A.29, corresponding to position 570 in pCAS5, resulted in a frame-shift mutation from that position onward and a pre-mature termination of 7 amino acids later. Also in clone 2A.33, an adenine-to-thymine substitution, corresponding to position 225 in pCAS5, resulted in a non-sense mutation at that position. So, only four cDNA clones, namely 2A.32, 2A.52, 2A.55 and 2A.62, encoded the full-length amino acid sequence of cassava linamarase. Alignment of the amino acid sequences derived from these four cDNA clones showed 98-99% sequence identity to the previously reported amino acid sequence of the cassava linamarase derived from pCAS5 (Hughes *et al*., 1992), and to each other (Figure 2). All four predicted polypeptides contained the highly conserved TFNEP and VTENG motifs bearing the catalytic acid/base (Glu198) and catalytic nucleophile (Glu413), respectively. The five putative N-glycosylation motifs, NX(S/T), were also conserved. The polypeptides derived from cDNA clones were analyzed for their physical and chemical parameters (Table 1).

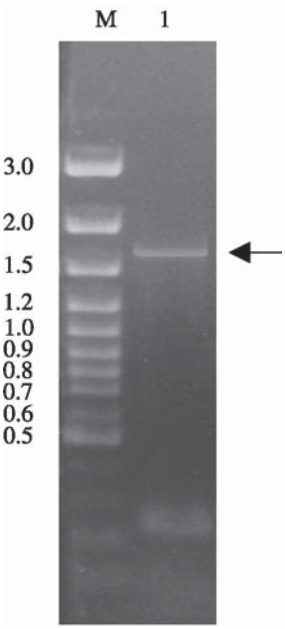

Figure 1 The cDNA of cassava linamarase (analyzed on 1% (w/v) agarose and stained with ethidium bromide). Lane 1, the 1.7 kb RT-PCR product (indicated by an arrow); M, DNA size marker (kb).

The cDNA sequences reported here were synthesized from young cassava leaves, which had been shown to be the site of cassava linamarase synthesis (Poncoro and Hughes, 1992). This is likely to produce different linamarase transcripts from the previously reported cassava linamarase cDNA sequence, pCAS5, which was extracted from yellow cotyledons (Hughes *et al*., 1992). The four polypeptides predicted from the four cDNA clones were very similar to the previously reported cassava linamarase, in terms of sequence identity, physico-chemical properties as well as the presence of the key motifs essential for catalytic activity. So, these four predicted polypeptides can

Table 1 Physical and chemical parameters of the polypeptides derived from four cDNA clones.

cDNA	Number of amino acids	Molecular weight	Theoretical p*I*
pCAS5	531	61,372.9	5.52
2A.32	531	61,461.0	5.44
2A.52	531	61,488.1	5.62
2A.55	531	61,323.8	5.44
2A.62	531	61,370.8	5.35

```
pCAS5           MLVLFISLLALTRPAMGTDDDDDNIPDDFSRKYFPDDFIFGTATSAYQIEGEATAKGRAP 60
2A.32           MLVLFISLLVLTRPAMGTDDDDDNIPDDFSRKYFPDDFIFGTATSAYQIEGEATAKGRAP 60
2A.52           MLVLFISLLVLTRPAMGTDDDDDNIPDDFSRKYFPDDFIFGTATSAYQIKGEATAKGRAP 60
2A.55           MLVLFISLLALTRPAMGTDDDDDNIPDDFSRKYFPDDFIFGTATSAYQIEGEATAKGRAP 60
2A.62           MLVLFISLLALTKPAMGTDDDDDDIPGDFNRNYFPDDFIFGTATSAYQIEGEATAKGRAP 60
                *********.**:**********:**.**.*:*****************:**********

pCAS5           SVWDIFSKETPDRILDGSNGDVAVDFYNRYIQDIKNVKKMGFNAFRMSISWSRVIPSGRR 120
2A.32           SVWDIFSKETPDRILDGSNGDVAVDFYNRYIQDIKNVKKMGFNAFRMSISWSRVIPSGRK 120
2A.52           SVWDIFSKETPDRILDGSNGDVAVDFYNRYIQDIKNVKKMGFNAFRMSISWSRVIPSGRR 120
2A.55           SVWDIFSKETPDRILDGSNGDVAVDFYNRYIQDIKNVKKMGFNAFRMSISWSRVIPSGRR 120
2A.62           SVWDIFSKETPDRILDGSNGDVVVDFYNRYIQDIKNVKKMGFNAFRMSISWSRVIPSGRR 120
                **********************.************************************:

pCAS5           REGVNEEGIQFYNDVINEIISNGLEPFVTIFHWDTPQALQDKYGGFLSRDIVYDYLQYAD 180
2A.32           REGVNEEGIQFYNDVINEIISNGLEPFVTIFHWDTPQALQDKYGGFLSRDIVYDYLQYAD 180
2A.52           REGVNEEGIQFYNDVINEIISNGLEPFVTIFHWDTPQALQDKYGGFLSRDIVYDYLQYAD 180
2A.55           REGVNEEGIQFYNDVINEIISNGLEPFVTIFHWDTPQALQDKYGGFLSRDIVCDYLQYAD 180
2A.62           REGVNEEGIQFYNDVINEIISNGLEPFVTIFHWDTPQALQDKYGGFLSRDIVYDYLQYAD 180
                **************************************************_********

pCAS5           LLFERFGDRVKPWMTFNEPSAYVGFAHDDGVFAPGRCSSWVNRQCLAGDSATEPYIVAHN 240
2A.32           LLFERFGDRVKRWMTFNEPSAYVGFAHDDGVFAPGRCSSWVNRQCLAGDSATEPYIVAHN 240
2A.52           LLFERFGDRVKRWMTFNEPSAYVGFAHDDGVFAPGRCSSWVNRQCLAGDSATEPYIVAHN 240
2A.55           LLFERFGDRVKRWMTFNEPSAYVGFAHDDGVFAPGRCSSWVNRQCLAGDSATEPYIVAHN 240
2A.62           LLFERFGDRVKRWMTFNEPSAYVGFAHDDGVFAPGRCSSWVNRQCLAGDSATEPYIVAHN 240
                ***********_**************************************************

pCAS5           LLLSHAAAVHQYRKYYQGTQKGKIGITLFTFWYEPLSDSKVDVQAAKTALDFMFGLWMDP 300
2A.32           LLLSHAAAVHQYRKYYQGTQKGKIGITLFTFWYEPLSDSEVDVQAAKTALDFMFGLWMDP 300
2A.52           LLLSHAAAVHQYRKYYQGTQKGKIGITLFTFWYEPLSDSEVDVQAAKTALDFMFGLWMDP 300
2A.55           LLLSHAAAVHQYRKYYQGTQKGKIGITLFTFWYEPLSDSEVDVQAAKTALDFMFGLWMDP 300
2A.62           LLLSHAAAVHQYRKYYQGTQKGKIGITLFTFWYEPLSDSEVDVQAAKTALDFMFGLWMDP 300
                ****************************************:*********************

pCAS5           MTYGRYPRTMVDLAGDKLIGFTDEESQLLRGSYDFVGLQYYTAYYAEPIPPVDPKFRRYK 360
2A.32           MTYGRYPRTMVDLAGDKLIGFTDEESQLLRGSYDFVGLQYYTAYYAEPIPPVDPKFRRYK 360
2A.52           MTYGRYPRTMVDLAGDKLIGFTDEESQLLRGSYDFVGLQYYTAYYAEPIPPVDPRFRRYK 360
2A.55           MTYGRYPRTMVDLAGDKLIGFTDEESQLLRGSYDFVGLQYYTAYYAEPIPPVDPKFRRNK 360
2A.62           MTYGRYPRTMVDLAGDKLIGFTDEESQLLRGSYDFVGLQYYTAYYAEPIPPVDPKFRRYK 360
                **************************************************:***_*

pCAS5           TDSGVNATPYDLNGNLIGPQAYSSWFYIFPKGIRHFLNYTKDTYNDPVIYVTENGVDNYN 420
2A.32           TDSGVNATPYDLNGNLIGPQAYSSWFYIFPKGIRHFLNYTKDTYNDPVIYVTENGVDNYN 420
2A.52           TDSGVNATPYDLNGNLIGPQAYSSWFYIFPKGIRHFLNYTKDTYNDPVIYVTENGVDNYN 420
2A.55           TDSGVNATPYDLNGNLIGPQAYSSWFYIFPKGIRHFLNYTKDTYNDPVIYVTENGVDNYN 420
2A.62           TDSGVNATPYDLNGNLIGPQAYSSWFYIFPKGIRHFLNYTKDTYNDPVIYVTENGVDNYN 420
                ************************************************************

pCAS5           NESQPIEEALQDDFRISYYKKHMWNALGSLKNYGVKLKGYFAWSYLDNFEWNIGYTSRFG 480
2A.32           NESQPIEEALQDDFRISYYKKHMWNALGSLKNYGVKLKGYFAWSYLDNFEWNIGYTSRFG 480
2A.52           NESQPIEEALQDDFRISYYKKHMWNALGSLKNYGVKLKGYFAWSYLDNFEWNIGYTSRFG 480
2A.55           NESQPIEEALQDDFRISYYKKHMWNALGSLKNYGVKLKGYFAWSYLDNFEWNIGYTSRFG 480
2A.62           NESQPIEEALQDDFRISYYKKHIWNALGSLKNYGVKLKGYFAWSYLDNFEWNIGYTSRFG 480
                **********************:************************************

pCAS5           LYYVDYKNNLTRYPKKSAHWFTKFLNISVNANNIYELTSKDSRKVGKFYVM 531    % ID
2A.32           LYYVDYKNNLTRYPKKSAHWFTKFLNISVNANNIYELTSKDSRRVGKFYVM 531      99
2A.52           LYYVDYKNNLTRYPKKSAHWFTKFLNISVNANNIYELTSKDSRKVGKFYVM 531      99
2A.55           LYYVDYKNNLTRYPKKSAHWFTKFLNISVNANNIYELTSKDSRKVGKFYVM 531      99
2A.62           LYYVDYKNNLTRYPKKSAHWFTKFLNISVNANNIYELTSKDSRKVGKFYVM 531      98
                ******************************************:*******
```

Figure 2 Multiple sequence alignment of derived amino acid sequences from four cDNA clones and pCAS5. The highly conserved TFNEP and VTENG motifs, containing the catalytic acid/base (Glu198) and catalytic nucleophile (Glu413), respectively, are underlined. The putative N-glycosylation sites are boxed. % ID indicates percent sequence identity to the amino acid sequence derived from pCAS5. *****, single, fully conserved residue; **:**, conservation of strong group; **.**, conservation of weak group; -, no consensus.

be reasonably assumed to be catalytically active in cassava. Previously, three isozymes of cassava linamarase, with p*I* values of 4.3, 3.3 and 2.9 and a single molecular weight of 63,000, were identified by chromatofocusing column chromatography (Eksittikul and Chulavatanatol, 1988). However, the chromatofocusing procedure was done in pH ranged between 5-1, which did not cover the p*I* values of the four polypeptides derived in this study and that of the polypeptide derived from pCAS5, which exhibited p*I* values between 5.3-5.6. There may be many other isoforms of linamarase, which are expressed at different times or in different parts of the plant. So, the four cDNA sequences isolated in this report (2A.32, 2A.52, 2A.55 and 2A.62) would represent only a subset of a large number of glycosidase sequences presented in the cassava genome. The two other cDNA sequences (2A.29 and 2A.33) that contained mutations and hence premature terminations could represent "pseudogenes", or "dead genes". Pseudogenes are defined as genes of similar sequences to normal genes but contain disablements, such as frameshifts or stop codons in the middle of the coding region, such that their translation would result in non-functional products. So, it is noteworthy that pseudogenes can be transcribed, but the functions of some transcripts are not known (Snyder and Gerstein, 2003).

It is not uncommon to observe in plant genomes the existence of multigene family of family 1 β-glycosidase enzymes, of which cyanogenic β-glucosidases are a member. Indeed, at least 48 sequences (including 8 probable pseudogenes) of glycosidase have been identified in the genome of *Arabidopsis thaliana*. There may be 50 members of these enzymes in the rice genome, whereas at least 16 members were recognized in the maize genome (Xu *et al*., 2004). The presence of multiple forms of enzymes may reflect functional redundancy among the gene products. On the other hand, plant may require a variety of β-glycosidase enzymes, with distinct aglycone specificities or expression patterns, for catabolism of diverse β-glycosidic metabolites for their normal growth and development (Xu *et al*., 2004).

In summary, six different cDNA clones of linamarase from cassava have been isolated by RT-PCR. Only four clones encode full-length polypeptides of linamarase. The nucleotide and amino acid sequences of these four clones revealed high sequence identity to the previously reported sequence of cassava linamarase. The full-length cDNA sequences obtained from this study will be used for further recombinant protein expression, site-directed mutagenesis and characterization in order to elucidate the structure-function relationships in this enzyme.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial supports form Kasertsart University Research and Development Institute and Thailand Research Fund, which have made this project possible.

LITERATURE CITED

Eksittikul, T. and M. Chulavatanatol. 1988. Characterization of cyanogenic β-glucosidase (linamarase) from cassava (*Manihot esculenta* Crantz). **Arch. Biochem. Biophys.** 266: 263-269.

Hughes, M.A., K. Brown, A. Pancoro, B.S. Murray, E. Oxtoby and J. Hughes. 1992. A molecular and biochemical analysis of the structure of the cyanogenic β-glucosidase (linamarase) from cassava (*Manihot esculenta* Crantz). **Arch. Biochem. Biophys.** 295: 273-279.

Poncoro, A. and M.A. Hughes. 1992. *In-situ* localization of cyanogenic β-glucosidase (linamarase) gene expression in leaves of

cassava (*Manihot esculenta* Crantz) using non-isotopic riboprobes. **Plant J.** 2: 821-827.

Snyder, M. and M. Gerstein. 2003. Defining genes in the genomics era. **Science** 300: 258-260.

Svasti, J., T. Phongsak and R. Sarnthima. 2003. Transglucosylation of tertiary alcochols using cassava β-glucosidase. **Biochem. Biophys. Res. Comm.** 305: 470-475.

Xu, Z., L.L. Escamilla-Trevino and L. Zeng. 2004. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. **Plant Mol. Biol.** 55: 343-367.

Yeoh, H.-H. 1989. Kinetic properties of β-glucosidase from cassava. **Phytochem.** 28: 721-724.