
Khadijah Hanim Abdul Rahman and Mohd Razip Samian

Abstract—Acetoacetyl-CoA thiolase (AACT, β-ketothiolase) is an enzyme that catalyzes the biological Claisen condensation of two molecules of acetyl-CoA to form acetoacetyl-CoA. In plants, AACT is known to be involved in the early step of mevalonate pathway which is essential in producing isoprenoid compounds such as sterols, carotenoids and growth regulators. In order to study this gene, we attempt to clone the gene from oil palm. The AACT gene was cloned from a genomic library of oil palm using a homologous probe from the AACT cDNA clone. The probe was consisted of 1.5 kb in length. A fragment of 12,470 bp was successfully sequenced and contains nine exons interrupted by eight introns. This fragment represented 866 bp in length or 59% of 1247 bp of the putative AACT mRNA sequence. Transcription start site (TSS) and its regulatory regions were predicted at the 5′-flanking region of this gene. Three TSS points, at positions -1284, -1283 and -280 from the putative start codon were identified. The binding sites for GT-1 element, SBF-1 and sequences resembling the binding sites of putative start codon were identified. The binding sites for GT-1 element, SBF-1 and sequences resembling the binding sites of other transcription factors were also identified.

Index Terms—β-ketothiolase, enzyme, mevalonate pathway, polyhydroxybutyrate (PHB), regulatory regions.

I. INTRODUCTION

In recent years, oil palm has become an important crop for oils and fats production as well as in the oleochemicals industry. Its cultivation in Malaysia has tremendously expanded and covers about two third of its agricultural area. Researches at molecular level and in genetic engineering have enabled the oil palm to produce a novel value added products such as bioplastics, high oleate and high stearate palm [1]. Among all of oil palm’s parts, it is believed that oil palm fruits can be the target site for accumulation of genetically engineered products since it has two storage tissues, mesocarp and kernel. Presently, the production of PHA in oil palm is also targeted to the palm fruit since its mesocarp contains high flux of acetyl-CoA as a precursor for PHA production [2].

In plants, there are two distinct forms of thiolases detected. Thiolase I or 3-ketoacyl-CoA thiolase (KAT, E.C. 2.3.1.16) is involved in the degradative pathways and has broad substrate specificity between four to twenty-two carbons. KAT normally takes part in the last step of β-oxidation to catalyze the cleavage of acetyl-CoA from 3-ketoacyl-CoA [3]. The second thiolase, called Thiolase II or acetoacetyl-CoA thiolase (β-ketothiolase, AACT, E.C. 2.3.1.9), is involved in the first step in many biosynthetic pathways such as mevalonate pathways and poly (3-hydroxybutyric acid) (PHB) synthesis in many microorganisms. Unlike KAT, AACT enzyme is specific for two carbons substrates and catalyzes the non-decarboxylating Claisen condensation of two acetyl-CoA molecules into acetoacetyl-CoA [4].

Basically, this research focuses on the characterization of oil palm AACT gene and identification of its regulatory regions. In microorganisms, AACT enzyme is involved in the first step of polyhydroxybutyrate (PHB) biosynthesis [5] which is the most common member of polyhydroxyalkanoates (PHA). PHA which served as carbon and energy storage polymer in microorganisms can be developed into biodegradable and eco-friendly substitute for petroleum plastics [6]. Currently, PHB is produced by costly production by the microorganisms. As an alternative, PHB can be synthesized in plants at lower cost since it does not require expensive fermentation and processing facilities [7]. The AACT enzyme was suggested to be rate-limiting enzyme for PHB production [8]. However, information available regarding AACT enzyme in plants and specifically in oil palm is still scarce. Therefore, this study was carried out to provide more knowledge and understanding of the AACT gene organization in the oil palm, specifically.

In this study, genomic DNA of oil palm was partially digested to prepare a genomic library using λDASH™II as a cloning vector. A clone was isolated from the genomic library using pOP CA90, a putative oil palm AACT cDNA clone obtained from Malaysian Palm Oil Board (MPOB) as a probe [6]. Later, the isolated clones were subcloned into pBluescript SK II (-) vector and sequenced. The fragment was then further characterized for their genomic organization.

II. MATERIALS AND METHOD

A. Preparation of DNA Probe

The DNA probe used in this study was digested out from a plasmid clone, (pOP CA90) [6] harboring a cDNA fragment of oil palm AACT gene by EcoRI-XhoI restriction digestion. The fragment was gel purified by QIAquick Gel Extraction Kit (QIAGEN). The resultant 1.5 kb fragment is termed as the CA90 probe. The DIG labeled fragment was used as a homologous probe to isolate the AACT gene from oil palm.
B. DNA Extraction

DNA was extracted from *Elaeis guineensis* Jacq. young leaves tissues according to method described by Murray and Thompson, 1980 [9]. The DNA was then quantified using spectrophotometer and checked for integrity by agarose gel electrophoresis.

C. Isolation of Oil Palm AACT Genomic Clones

An oil palm genomic library constructed in a Lambda DASH® II vector was screened with CA90 labeled probe using plaque hybridization protocol [10]. Hybridization and detection were performed according to ‘The DIG System User’s Guide for Filter Hybridization’ by Boehringer Mannheim Biochemica. Hybridization was carried out at 45°C for primary screening and eventually increased to 65°C, a high stringency condition for secondary and tertiary screening. Chemiluminescent signal was detected by exposing the membrane to standard X-ray film (FUJI Photo Film, Ltd) for 2 hours.

Potential clones were picked using wide-bore micropipette tips and preparation of plate lysates and λDNA isolation were performed according to Sambrook *et al.*, 1989 [10].

At the end of the primary screening, only 1 putative clone hybridized to probe CA90. Positively hybridized clone was subjected to secondary and tertiary screening. This recombinant lambda clone was extracted and digested with several restriction enzymes. The fragments generated were hybridized with CA90 DIG labeled probe. After hybridization, the clone was digested with NotI restriction enzyme. Approximately 12 kb in size of insert were gel purified using Mega Quick Spin (INTRON) and ligated to NotI predigested pBluescript KS II (-) vector and transformed into *E. coli* JM109. White colonies which grew on LB agar with 100 µg/ml ampicillin, supplemented with 20 µl 50 mg/ml X-gal and 20 µl 100 mM IPTG were chosen for plasmid extraction. Recombinant plasmids were subjected to PCR analysis and southern hybridization using the CA90 probe described above for insert confirmation.

D. Sequencing and Sequence Analysis

Positive pBluescript clone was sequenced using standard sequencing primers, M13F and M13R, as well as primers specific to the AACT sequence. The nucleotide sequences were determined by employing primer walking approach. The resulting nucleotide sequences were analyzed using BioEdit and database searches were performed with BLAST (Basic Local Alignment Search Tool) program [11].

E. AACT Promoter Identification

The AACT promoter region was analyzed for putative transcription start site (TSS) using the TSSP software [12] and NNPP2.2 (Neural Network Promoter Program) [13]. Meanwhile, PLACE [14], plantCARE [15] and TFSearch (Yutaka Akiyama, http://www.rwcp.or.jp/papia, TFSearch: Searching Transcription Factor Binding Sites) were the software programs used for identifying cis-acting regulatory elements and transcription factor binding sites (TFBS).

III. RESULTS AND DISCUSSION

A. Cloning of Oil Palm AACT Gene

A putative AACT clone was isolated from the oil palm genomic library. This clone was designated as λOP1 was further analyzed by Southern hybridization. Southern analysis (Fig. 1) showed the restriction digestion patterns for the clone.

The λOP1 was sequenced and revealed partial-length genomic clone (GenBank accession no: JQ606806) of 12, 470 bp long.

Alignment of the genomic sequence with that of cDNA showed that the genomic DNA consisted of nine exons of 69, 100, 171, 105, 138, 77, 66, 85 and 75 bp interrupted by eight introns of 2919, 82, 116, 1430, 683, 87, 622 and 2177 bp long. The assembled exon regions were completely identical to cDNA sequence of AACT gene previously determined by Teen *et al.*, 2008 [6]. However, the insert obtained is a partial genomic DNA since it only represented 886 bp in length of 1247 bp of the putative AACT cDNA sequence. Besides that, the insert also consist of 3,389 bp non-coding region at the 5’-flanking region upstream of the first exon (Fig. 2).

![Fig. 1. Electrophoresis profile and Southern hybridization analysis of λOP1. A(i) is the electrophoresis profile of λOP1 digested with lane 2: BglII; lane 3: NcoI and lane 4: NdeI restriction enzymes. A(ii) is the Southern hybridization analysis of λOP1.](image)

![Fig. 2. Structural organizations of the oil palm AACT gene: Intron-exon distribution. Exons are numbered from 1 to 9. Translated sequences are shown as black boxes and untranslated sequences are shown as shaded box.](image)

Based on the DNA sequence information, exon 1 contains the coding information for the first 13 amino acids of the protein. Exon 2 codes for 34, exon 3 codes for 57, exon 4 codes for 35, exon 5 codes for 46, exon 6 codes for 25, exon 7 codes for 22, exon 8 codes for 29 and exon 9 codes for 25 amino acid residues. When the exon sequences are read continuously, they code for a polypeptide of 285 amino acid residues. Most of the exon/intron boundaries followed the GT/AG rule except for exon 3(5’ splice donor), exon 5 (3’ splice acceptor) and exon 7 shown in Table I.

A gene in a genome is normally very long, consists of...
introns, exons, regulatory regions and other functional regions. In *O. sativa*, the AACT gDNA span approximately 5.5 kb long with fourteen exons and thirteen introns [16], [17] while in *Arabidopsis*, AACT gDNA span approximately 3.3 kb long with twelve exons and eleven introns (The Arabidopsis Information Resource (TAIR), (www.arabidopsis.org). According to Kano et al., 1991 [18], the human mitochondrial AACT gDNA span as long as 27 kb and contains twelve exons interrupted by eleven introns. Based on this information, the oil palm AACT gDNA is expected to be longer than *O. sativa* and *Arabidopsis* AACT gDNA and probably the same size range of human’s AACT. This is due to the insert obtained in this study is only represented 59% of the putative AACT mRNA sequence.

B. Promoter and Regulatory Elements Identification

The prediction of transcriptional start site (TSS) is important in order to discern either the location of promoter or regulatory regions of the gene of interest. In this study, three transcriptional start sites were predicted. The first TSS predicted was at nucleotide 1,284 bp (+1a) (Fig. 3) upstream of the putative AACT start codon. The TATA box sequence, which normally serves as common core promoter element was identified 29 bp upstream of the predicted TSS. The second TSS was predicted to be located near the first TSS predicted with 1 bp shift at position 1,283 bp (+1b) upstream of start codon. The third TSS was predicted to be at position 280 bp (+1c) (Fig. 3) upstream of the putative translation start site.

Both of the softwares used for TSS prediction are based on the recognition of putative RNA polymerase II promoter sequences in plant DNA [12] and based on the TATA box and initiator (inr) elements [13]. Therefore, it is difficult to determine the degree of accuracy when using these programs since their predictions are highly dependent on TATA boxes and known TFBS from the available databases. Consequently, the chances of getting false positive results are rather high since knowledge on transcription regulation of plants is still very limited [19]. Besides the TSS region, cis-acting regulatory elements are also important in serving as binding sites for transcription factors in order to initiate transcription. The cis-acting regulatory elements predicted were SBF-1, GT-1, TATA boxes, TRANSINITMONOCOTS motif and 5'UTR Py-rich stretch motif.

The significance of these motifs will be discussed below.

TFSearch predicted a few transcription factor binding sites (TFBS). One of the predicted TFBS is the SBF-1 which is a cis-element that functioned as a transcriptional silencer in chalcone synthase (CHS) gene [20]. The SBF-1 has a consensus sequence of GGTTAA (A/T)(A/T)(A/T) and involved in the organ-specific expression during plants development. AACT and chalcone synthase belong to the thiolase superfamily [21]. Generally, enzymes in this family share an overall similarity in their protein structure and function [4], [21]. Due to their relationship, there are possibilities that they share some common transcription regulators. This assumption is supported by experiments done on yeasts; proteins that have similar functions and structures tend to be co-expressed [22]. Rombauts et al., 2003 [23], in their review mentioned that co-expressed genes are likely to act similarly therefore they are expected to be co-regulated. Considering that co-regulation occurs at transcriptional level, it is expected that co-expressed genes contain similar cis-regulatory elements in their promoter region.

Within the proximity of the predicted TSS region, a GT-1 binding site was detected upstream of the TATA-box. GT-1 is a cis-regulatory element that is involved in light activation. In *Arabidopsis*, GT elements belong to a small family of transcription factors that are ubiquitously expressed [24]. In addition, GT-1 element is also a common transcription factors among housekeeping genes and is not limited to genes that are light regulated [16], [24] suggesting a possible role in the AACT gene regulation.

A TRANSINITMONOCOTS motif was also detected at position 68 bp upstream of the ATG start codon. However, this region does not agree with the putative translation start site in this gene since the putative start codon is detected downstream of this motif. According to Joshi et al., 1997 [25], the c (a/c) (A/G)(A/C)AUUGC sequence has been identified as a consensus among monocot plants.

The sequence upstream of the AUG start codon is highly associated with the efficiency of translation initiation [26].

When these bases differ from the consensus sequence, part of the 40S ribosome subunit is unable to identify the ATG codon as the translation start site thus lower the rate of polypeptide synthesis. In 2008, Teen et al., [6] did a translation efficiency test on the 5' UTR sequence of AACT and the results showed that oil palm AACT cDNA is expressed at low levels. This result corroborate that the TRANSINITMONOCOTS motif is expected not to be part of the cis-regulatory element involved in the regulation of AACT gene.

Besides these motifs, a 5UTR Py-rich motif was also detected. The 5UTR Py-rich motif was reported to be necessary for high transcription level in tomato

<table>
<thead>
<tr>
<th>Exon</th>
<th>Intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>(bp)</td>
</tr>
<tr>
<td>1</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>171</td>
</tr>
<tr>
<td>4</td>
<td>105</td>
</tr>
<tr>
<td>5</td>
<td>138</td>
</tr>
<tr>
<td>6</td>
<td>77</td>
</tr>
<tr>
<td>7</td>
<td>66</td>
</tr>
<tr>
<td>8</td>
<td>85</td>
</tr>
<tr>
<td>9</td>
<td>75</td>
</tr>
</tbody>
</table>

*TABLE I: INTRON-EXON BOUNDARIES OF THE OIL PALM AACT GENE*

<table>
<thead>
<tr>
<th>Acetoacetyl-CoA thiolase</th>
<th>5' splice donor</th>
<th>3' splice acceptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGCCCAAGGTatatataacct...</td>
<td>tatgcagAGTATGATGATAT</td>
<td></td>
</tr>
<tr>
<td>CATCCTAAgtaaataatatatat...</td>
<td>catgttgcGTCCTCTTA</td>
<td></td>
</tr>
<tr>
<td>GGGATGAAGGgtagaatatttttcac...</td>
<td>cAACAATGGAATAAGT</td>
<td></td>
</tr>
<tr>
<td>AGAAAGCAAGgttgeaatc...</td>
<td>tgtgcgAAAGGAGATC</td>
<td></td>
</tr>
<tr>
<td>GAAGAGCAAGgttaactgtat...</td>
<td>caactcctatAGGATTCTT</td>
<td></td>
</tr>
<tr>
<td>ATGGTCTCGgttagacate...</td>
<td>tattgcagGTGAAGGT</td>
<td></td>
</tr>
<tr>
<td>TGAAAAGAgttagtata...</td>
<td>tattgcagATGACTGT</td>
<td></td>
</tr>
<tr>
<td>TAGTATAAGgttggtgtgt...</td>
<td>tgtgagcaTGATGTCG</td>
<td></td>
</tr>
<tr>
<td>GCAAAGATCgtaactgtg...</td>
<td>gtaactgtgT</td>
<td></td>
</tr>
</tbody>
</table>

Acetoacetyl-CoA thiolase
3-hydroxy-3-methylglutaryl-CoA reductase 2 (HMGR2) gene [27]. HMGR and AACT enzymes are known to be involved in the isoprenoid biosynthesis. The isoprenoid compounds are known to accumulate slowly over a long period, suggesting that enzymes in this pathway are only present in low quantity [28]. However, according to Daraselia et al., 1996 [27], HMGR enzymes in plants can be highly expressed according to the necessities of their growth and development. Therefore, the transcription levels of the enzymes involved are highly dependent upon their developmental needs. Teen et al. 2008, [6] revealed that a Northern analysis of AACT mRNA transcripts are present in higher quantities in the riper (13, 17, 20 WAA) then younger fruits (6, 8 and 11 WAA) suggesting that the AACT enzyme can be highly expressed according to developmental stages. Thus, SUTR Py-rich cis-element could possibly be important in determining the high transcription level in the AACT enzyme.

Fig. 3. Nucleotide sequence of the 5' upstream region of the oil palm AACT gene. The putative transcription start sites (TSSs) are asterisked. The initiation codon (ATG) for the AACT is bolded and shown as +1. The putative TATA box sequences, the identified cis-acting regulating elements: GT-1, SBF-1, 5UTR Py-rich stretch and TRANSINITMONOCOTS motif are boxed according to their position. Nucleotide sequence of the first exon with is corresponding amino acid is presented.

Additionally, promoters are very complex structures. It is composed of many structural features and their regulatory elements are very short, which highly complicates their identification thus making the bioinformatic prediction of TSSs and promoters difficult. Consequently, functional analysis is still crucial to characterize the AACT gene promoter and to understand the underlying molecular mechanisms of the transcriptional regulation of AACT gene expression.

IV. CONCLUSION

In this study, the genomic DNA library of oil palm has been successfully constructed in the λDASH® II replacement
vector. A clone was successfully isolated in this study using labeled cDNA clone of the putative acetocacyl-CoA thiolase. A clone, designated as λOP1 was sequenced and analyzed. The λOP1 fragment is 12,470 bp in length and consists of nine exons interrupted by eight introns. At the 5’ end of the fragment, there is a 3,389 bp long non-coding region which carry the promoter and regulatory regions of this gene as predicted using bioinformatics tools. Three transcriptional start sites were predicted together with TATA boxes. The predicted binding sites for GT-1, SBF-1 and transcriptional start sites were predicted with GT-elements and GT-factors,

Subsequently, this work has provided the basis for future studies which include the characterization of the full length genomic DNA sequence for AACT and determination of the AACT promoter and regulatory regions through functional analysis.

ACKNOWLEDGMENT

Authors would like to thank Ravigadevi Sambanthamurthi of Malaysian Palm Oil Board for providing the pOP CA90 cDNA clone.

REFERENCES


Khadijah Hanim Abdul Rahman was born on August 29, 1983 in Kuala Lumpur, Malaysia. She obtained her master degree in 2011 in biotechnology from Universiti of Sains Malaysia, Penang, Malaysia and a bachelor degree in biotechnology from Universiti Putra Malaysia, Selangor, Malaysia in 2005. Her major field of study is in biotechnology and molecular biology. Currently, she serves as a lecturer at School of Bioprocess Engineering, Universiti Malaysia Perlis, Malaysia. Her research interest is in the regulation of oil palm genes, bioprotein production from agricultural wastes and also cellulases production.

Mohd Razip Samian was born on September 30, 1957. He obtained his doctorate in 1990 from Universiti of New South Wales, Australia in biotechnology. His major field of study is in molecular biology and structural biology of proteins. Currently, he serves as a professor at School of Biological Sciences, Universiti Sains Malaysia. His research interest is in the molecular genetics of bioplastic (polyhydroxyalkanoic acid, PHA) production in Pseudomonas sp/USM 4-55, structural biology and overexpression of proteins.