Structural and functional analyses of a saturated acyl ACP thioesterase, type B from immature seed tissue of Jatropha curcas

K. G. Srikanta Dani1,2, K. S. Hatti1, P. Ravikumar1 & A. Kush1
1 Vittal Mallya Scientific Research Foundation, Bangalore, India
2 Present address: Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol, UK

ABSTRACT

The distinguishing structural and functional domains of plant acyl-acyl carrier protein (ACP) thioesterases and their complex interaction with the ACP-linked fatty acid substrate complex have remained elusive. E. coli based heterologous expression and characterisation of many plant thioesterases reported so far have not been extended and linked to in silico modelling studies to explain the diversity in plant thioesterase substrate specificities. In this study, a thioesterase cDNA isolated from immature seed tissues of Jatropha curcas was found to be type B and specific to stearoyl acyl ACP when expressed in E. coli K27fadD88, a lipid utilisation mutant. Homology modelling and molecular docking of a selected region of the isolated JcFatB protein predicted that it had high affinity towards both stearate (18:0) and palmitate (16:0). Structural analysis of the sequence confirmed the presence of a transit peptide that is processed in multiple steps. The enzyme is localised in the chloroplasts and has an N-terminal inner chloroplast transmembrane domain characteristic of type B plant thioesterases. Docking of ligands with JcFatB and its comparison with a modelled Jatropha thioesterase type A provided further evidence for native substrate preferences of Jatropha thioesterases. This study provides essential clues to develop future methods for large-scale bacterial production of free fatty acids and for design of strategies to modulate the seed oil composition in this important non-edible, seed oil plant.

INTRODUCTION

Engineering fuel properties of seed oils has involved an incessant search for enzymes with unique substrate specificities in plant lipid metabolic pathways (Durrett et al. 2008). Jatropha curcas is a non-edible oil-yielding plant species with 35–38% total seed oil, making it an economically promising renewable energy source. Although seed oil of Jatropha is blended with diesel and used in modern combustion engines, its high viscosity and reduced cold temperature performance needs to be improved to make it a real and viable alternative to petroleum and diesel. To achieve this, we need a thorough understanding of the triacylglycerol (TAG) assembly enzyme specificities in Jatropha.

Seed oil of Jatropha contains a relatively small pool (~20%) of saturated fatty acids such as palmitate (16:0) and stearate (18:0). Their accumulation in non-photosynthetic tissues such as seeds (oil bodies) of higher plants is not easily explained (Salas & Ohlrogge 2002). Fatty acid composition of seed oils is determined largely by expression of highly specific fatty acid chain terminating enzymes, which are active in the plastids (Shine et al. 1976) namely, fatty acid acyl ACP thioesterases (FAT).

Plant thioesterases are prime candidate enzymes that could be used to tailor plant seed oils with biofuel potential for economic and environmental benefits. There are two functionally different types of FAT genes in plants. Type A (FATA) is almost exclusively involved in terminating (18:1Δ9) unsaturated, long-chain fatty acid esters. Type B (FATB), which in many species is shown to have special preference for substrates, and is also argued to be evolutionarily the older of the two (Jones et al. 1995), mainly determines the chain length of saturated fatty acid esters (16:0 and 18:0). It has been shown in many plant systems that an acyl ACP thioesterase specific to medium-chain saturated fatty acids is predominantly active in seeds. Characterisation of a thioesterase from Jatropha would help in designing novel strategies to customise its oil profile for better combustibility and fluidity at subzero temperatures.

De novo fatty acid synthesis occurs mainly in the plastids through a pathway referred to as the ´prokaryotic´ pathway (Ohlrogge et al. 1979). In non-photosynthetic tissues such as seeds, the accumulation of fatty acids in the cytosolic oil bodies involves an alternate pathway. This pathway, known as the ´eukaryotic´ pathway, involves export of acyl-ACP thioesters from the stroma to the outer membrane of the chloroplast and subsequent processing at the ER (Ohlrogge & Browse 1995). Although it is inferred that thioesterase activity is confined to the stroma (Shine et al. 1976), the existence of thioesterases on the inner walls of the chloroplast inner membrane is still not fully established.

Multiple sequence alignment between various published plant FATA and FATB cDNA sequences reveals that FATA
Factors influencing substrate specificity of plant thioesterases

Srikanta Dani, Hatti, Ravikumar & Kush

are significantly (150 bp) shorter than FATBs at the 5′-end (P < 0.01). Plant FATBs are shown to possess a unique N-terminal sequence, which is absent in FATAs. However, there is no consensus on the role of the N-terminal region in determining substrate preference of the thioesterase (Facciotti & Yuan 1998; Salas & Ohlrogge 2002; Mayer & Shanklin 2005). The distinguishing structural and functional domains of a thioesterase and its complex interaction with the acyl carrier protein (ACP) substrate complex have remained elusive. A combinatorial approach that involves both in silico analyses and in vivo experiments could help in understanding the distinct substrate affinities of these enzymes. In the current investigation, we have isolated and partially characterised a fatty acid acyl ACP thioesterase B from immature seed tissue of *Jatropha curcas*. The supporting modelling and docking evidence points towards the substrate specificity determining factors that, until now, have not been explored.

**MATERIAL AND METHODS**

Plant tissue and chemical source

*Jatropha* plants were grown from the seed stock of a high oil yielding genotype of *Jatropha curcas*, L., obtained from a forestry plantation network in Chikkaballapura, Karnataka, India. Plants were maintained in natural conditions until flowering and fruit set. Immature seeds, harvested at various developmental stages, were frozen in liquid nitrogen and stored at −70 °C. Restriction enzymes were procured from Promega (Madison, WI, USA), molecular biology-grade chemicals were obtained from both Merck (Darmstadt, Germany) and Sigma-Aldrich (St Louis, MO, USA). The seed oil quantification was done after extraction using the Soxhlet extraction method. The fatty acids in the seed oil were estimated by GC-MS analysis of their respective methyl esters as described in Li et al. (2006).

Isolation, cloning and sequencing of *Jc*FATB cDNA

The following full-length cDNAs of acyl ACP thioesterases from various plants in the NCBI (Accession numbers: X73849, U17098, AY078350, DQ856315, DQ847275, AJ242916, MFU65642, AF076353, AF036565, CAHOACPTB) gene repository were subjected to multiple sequence alignment using Clustal W (version 1.83). Conserved, consensus sequences were identified and a reliable end-specific primer pair was procured from Sigma-Aldrich. The forward primer was 5′ ATG GTA CCA TGG TTG CTA CTG CTA C 3′ and the reverse primer was 5′ ATA GTA CTT TAG GCA CTT TCA ACT GGA ATC 3′. RNA was isolated from immature seed tissue at various developmental stages using the TRI Reagent (a mixture of gua- nine thiocyanate and phenol; Sigma Aldrich, St Louis, MO, USA). The reverse transcription reaction was carried out with total RNA using the Quantitect RT reaction kit (Qiagen, Valencia, CA, USA). This was followed by PCR (initial incubation at 94 °C for 3 min, 30 cycles of 94 °C for 30 s, 50 °C for 1.5 min, 72 °C for 1.5 min, and a final incubation at 72 °C for 15 min). Amplification was performed in a 25-µl reaction volume that contained 5 µl of the single-stranded cDNA pool, 2.5 µl 10× PCR buffer (TAPS, pH 8.8: 1.5 mM MgCl2; 50 mM KCl), 2.5 mM each of dATP, dCTP, dTTP, dGTP, 10 µM each of forward and reverse primers and one unit of Taq polymerase.

A consistent ~1200 bp cDNA was obtained, which was ligated into pBS (plasmid BlueScript) and used to transform *E. coli* DH5α electro-competent cells. Plasmid minipreps were made from putative clones grown on selection medium. Presence of the insert was confirmed by both restriction digestion and PCR analyses. The confirmed cDNA clone was sequenced (Applied Biosystems, Carlsbad, CA, USA) and analysed using the online tool NEB Cutter 2.0 (New England Biolabs, Hitchin, UK). We named the clone JFB3 (putative *Jatropha* thioesterase B, clone number 3). In this paper, the JFB3 cDNA is referred to as *Jc*FATB, and its protein form as *Jc*FatB.

**Sequence analysis**

The BLAST analysis confirmed the sequence identity to FATB cDNAs. A reported FATB cDNA sequence from *Jatropha* (GenBank ID: EU106891.1; Wu et al. 2009) had 100% identity with our probe. Moreover, a host of confirmed FATB-type thioesterases from various plant species displayed high sequence similarity with *Jc*FATB.

The N-terminal sequence of 120 amino acids was screened for the presence of a transit peptide using TargetP1.1 server. This was followed by predictions using ChloroP1.1, which is a neural network-based method for predicting chloroplast transit peptides and their cleavage sites (Emanuelsson et al. 1999, 2007). The complete protein sequence was also put through transmembrane topology and cellular localisation prediction using TopPred (von Heijne 1992).

**Modelling and molecular dynamics of *Jc*FatB**

The *Jc*FatB cDNA (GI: 248570278) was translated using the ExpASy translation tool to obtain a protein sequence of 418 amino acids. A conserved domain analysis was also performed on the sequence. The I-TASSER server was used to develop a 3D model for the complete protein (Zhang 2008; Roy et al. 2010). The *Jc*FatB amino acid sequence (ACT09366.1) was compared with *Arabidopsis thaliana* AtFatB (NP_172327.1), AtFatA (NP_189147.1) and also with the *Jc*FatA (ABX82799.3) sequence from *Jatropha curcas*.

As a PDB template was not available for the entire length of the protein (418 amino acids), a complete model could not be developed for the original protein. Alternatively, we selected and successfully modelled a region of 274 amino acids (from 131 to 404) based on the crystal structure of oleoyl thioesterase from *Lactobacillus plantarum* (PDBID: 2OWN) chain B, which had 24% identity with *Jc*FatB. The same template was also used to model *Jc*FatA (23% identity) for a comparative study. The unaligned regions of both modelled proteins were cropped, and the structure from 131 to 404 of *Jc*FatB and 77 to 369 of *Jc*FatA was considered for molecular dynamics (MD) refinement. A simulation cell was created in a cubic periodic box, with a minimum distance of 10 Å between the protein and the box walls. The protein was bathed with TIP4P water molecules together with appropriate number of sodium ions to neutralise the system. Energy minimisation was performed using 5000 steps of the steepest descent method or until the maximum force was below 100 kJ mol⁻¹ nm⁻¹⁻¹.

Long-range electrostatic interactions were calculated using the particle-mesh Ewald method (PME), with a distance
cutoff of 1 nm. Short-range repulsive and attractive dispersion interactions were described using a Lennard–Jones potential, which was cut off at 1 nm. The LINCS algorithm was used to constrain bonds, allowing a time step of 1 fs. Neighbour searching was carried out every 10 steps. A Parrinello–Rahman barostat pressure of 1 bar was used, with a coupling constant of Tau_P = 0.5 ps and compressibility of 4.5e-5 (bar^-1). Water and protein molecules were coupled separately to the thermal bath at 300 K, using a v-rescale coupling constant, Tau_T = 0.1 ps. An initial preparatory run for 50 ps was carried out to allow the randomisation of water molecules around the protein, followed by 10 ns isobaric–isothermal ensemble simulation. All MD simulations were performed using GROMACS version 4.07 compiled in single-precision mode (Spoel et al. 2005; Hess et al. 2008).

Substrate specificity analysis through molecular docking

Receptor preparation
The frame having least RMSD to the average structure of all the frames between 6 and 10 ns of MD was energy minimised for 1000 steps using the steepest descent algorithm and was used for the docking procedure. All hydrogen atoms and charges were added and the protein defined using Amber-99SB force-field parameters (Chu et al. 2009). DMS of the receptor was calculated after ripping hydrogen atoms. The program ‘sphgen’ was used to generate 4 Å spheres around the receptor.

Ligand preparation
The fatty acid-bound 4′-phosphopantetheine (4PPT) group of holo-ACP was constructed based on earlier structural studies of ACPs (Roujineikova et al. 2002; Zornetzer et al. 2006). The 4′-phosphopantetheine-linked stearate (18:0) was cut and taken directly from solution structure of the spinach acyl carrier protein (PDBID: 2FVA). The stearate moiety was replaced with palmitate (16:0) and oleate (18:1(n-9)). All the three structures viz. palmitate, stearate and oleate, covalently bound to 4′-phosphopantetheine (4PPT) were geometrically optimised and energy minimised using the Dundee PRODRG2 Server. AM1-BCC partial charges were added to ligands using ANTECHAMBER provided in the Chimera program.

Blind docking of fatty acid
The command Showbox was used to create space sufficient to cover the entire protein and a grid box was built with minimum of 50 Å from protein to box edges. GRID was run at a grid spacing of 0.3 Å with distance cut-off of 10 Å. The coefficient of the dielectric of 4.0 and the exponent and repulsive of attractive Lennard–Jones term for the van der Waals potential of 6 and 12 were used, respectively. The ‘anchor-and-grow’ flexible ligand-docking algorithm in DOCK 6 (Lang et al. 2009) was employed to blindly dock the fatty acids to fatty acyl thioesterases (FAT). The grid-based score, which is based on the non-bonded terms of the molecular mechanic force field, was employed to screen 50,000 orientations clustered at 2.0 Å RMSD, and the best 100 were retained, all showing large negative interaction energies. The top 100 conformers were manually examined and those that bound to the putative active site explained by Mayer & Shanklin (2005, 2007) were considered for the Amber GB/SA scoring.

Amber GB/SA scoring function
The protonation state of the receptor was calculated using the H++ server version 2.0 (Gordon et al. 2005; Anandakrishnan & Onufriev 2008) and ligands were prepared using the script ‘prepare_amber.pl’ provided with the Dock 6 program (Lang et al. 2009). While the ligands defined by GAFF (Wang et al. 2004) parameters were kept rigid, all atoms of receptors defined by AMBER99SB with a distance cut-off of 10 Å were assigned flexibility. The 100 steps of minimisation followed by the 3000 steps of molecular dynamics simulation at a constant temperature of 300 K, and another 100 steps of minimisation were carried out.

Protein–protein docking
All the hetero-atoms from the solution structure of the acyl carrier protein or ACP (PDBID: 2 fva) were removed and side chains reconstructed using the Dunbrack rotamers library (Dunbrack 2002). This was docked against JcFatA and JcFatB using the protein–protein docking program ClusPro (Comeau et al. 2004). The top 10 balanced models were manually examined to determine the ACP binding site onto JcFatA and JcFatB.

Expression of JcFATB in the E. coli strain K27fadD88 mutant
Escherichia coli K27fadD88 was obtained from the Coli Genetic Resource Center (CGSC) at Yale University (USA). The confirmed original pBS clone (with blue white selection and confirmed translational fusion) of JcFATB (JFB3) was used to transform electro-competent cells of the K27 mutant. The control and transformed K27 cells were grown initially for about 8 h, until they reached a cell density of 0.55 (measured at 600 nm) and were induced with 200 µM IPTG for 6–8 h. Total lipids from the culture media were extracted through chloroform–methanol treatment, as given by Voelker & Davies (1994). The total cell pellets from all treatment groups were also collected and independently lysed through sonication. The lysed pellets were used for total lipid extraction. A standard fatty acid methyl ester mixture (FAME) obtained from Sigma (Cat no: F.A.M.E Mix C4-C24, 18919) was used for mass and retention time reference in GC-MS. GC-MS was carried out with a Shimadzu-QP2010. Samples were injected at a split ratio of 1:4 (1 ml). Column (Restek, 30 m × 0.25 mm i.d. × 1 µm thickness, Rtx-5; Restek Corporation, Bellefonte, PA, USA) conditions were maintained at 80 °C for 2 min, increased to 200 °C at a ramp rate of 20 °C per minute and held for 1 min. Then it was further increased to 280 °C at a rate of 2 °C per minute and held for 15 min. Helium was used as carrier gas in a total run time of 45 min. MS was performed at an injector temperature of 225 °C and detector temperature of 280 °C. The ionisation energy was maintained at 70 eV, with a scan range of 40–500 amu.

RESULTS

Sequence analysis
A 1254 bp full-length cDNA of Jatropha fatty acyl ACP thioesterase was isolated from developing seed tissues of Jatropha and the sequence uploaded to GenBank (ID: GQ226036.1).
The length of the cDNA, BLAST analysis and multiple sequence alignment confirmed its identity as a type-B thioesterase (Fig. 1A).

The N-terminal sequence of JcFatB (residues 1–125) belonged to acyl-thio_N super family (pfam12590). Amino acids from 136 to 388 belonged to the FATA domain (COG3884), within which two hot-dog domains (4HBT; cd00586) containing the catalytic residues were observed, each from the 139th to 225th and 305th to 399th amino acid.

Factors influencing substrate specificity of plant thioesterases
Srikanta Dani, Hatti, Ravikumar & Kush

Fig. 1. A: Multiple sequence alignment between the thioesterases of Arabidopsis thaliana (At) and Jatropha curcas (Jc). Conserved residues are underlined, distinct amino acid residues that were previously reported to impart substrate specificity to FATA and FATB are marked with filled circles, asterisks denote the CPDL-identified putative specificity determining positions, and the residues comprising the catalytic triad of C, H and N are denoted by +. Potential re-entrant loop is marked with a thick line above the JcFatB sequence. B: Hydrophobicity plots for JcFatA and JcFatB from the TopPred server. Potential membrane-spanning domains are marked with a red vertical line (with explanation). The plot is aligned to represent the highly conserved C-terminal domain of both JcFatA and JcFatB.
plastids. TargetP1.1 (Score = 0.959 at >95% specificity) and ChloroP1.1 confirmed the presence of a 58 amino acid-long cTP (chloroplast transit peptide) at the N-terminal end of the JcFatB. The region from 82 to 107 residues had significantly high hydrophobicity, confirming its transmembrane nature. Orientation of the transmembrane helix, because it is associated with the inner membrane of the chloroplast, is preferentially from the outside (cytoplasm) to the inside (stroma). A short N-terminal sequence of 25 amino acids, which is part of the mature protein, is predicted to be in the chloroplast inter-membranal space. An additional stretch of 20 amino acids (357–376), which is downstream of the active site (Fig. 1A) at its C-terminal, had a potential to be membrane-bound. Although TopPred had predicted that this region marginally missed the upper threshold for not being a significant transmembrane region (Fig. 1B), during protein molecular dynamics, we observed that the same sequence was highly unstable (Fig. 2B). This was not observed in JcFatA. This supported the view that JcFatB could potentially have a membrane re-entrant loop. All these factors indicate that JcFatB is bound to the inner membrane of the chloroplast in at least one, and may be more than one place (excluding the cTP domain).

Homology model evaluation

The I-Tasser modelling server yielded five models for both JcFatA and JcFatB. The JcFatA model, with least C-score of −2.14, TM-score of 0.6755 and aligned region RMSD of 0.75 A˚, and JcFatB with least C-score of −2.54, TM-score of 0.6004 and aligned region RMSD of 0.49 A˚ were considered for further studies.

The modelled JcFatA and JcFatB are nearly superimposable and are in line with the modelled structure of ArFatB in Mayer & Shanklin (2005) (Fig. 3). After removing unaligned regions, both the modelled proteins were subjected to molecular dynamics refinement for 10 ns. It was observed that the RMSD of both structures attained a plateau state after 6 ns of simulation (Fig. 2). This region between 6 and 10 ns was considered for further analysis. The RMS deviation of the α carbons of amino acids ranged between 0.05 and 0.5 nm, except in the unstable loop (357–376) in JcFatB, which deviated up to 1.2 nm during the simulation (Fig. 2).

Comparative docking

Out of the three fatty acid ligands considered for docking (viz. 4PPT-16:0, 4PPT-18:0 and 4PPT-18:1\(^{\Delta9}\)), more than 50% of all the ligands bound to the active catalytic triad of JcFatA. 4PPT-18:1\(^{\Delta9}\) had the least Amber GB/SA score compared to 4PPT-16:0 and 4PPT-18:0. As none of the 116 possible models of 4PPT-18:1\(^{\Delta9}\) predicted by Dock 6 bound to the active site of JcFatB, 4PPT-18:1\(^{\Delta9}\) was not considered for Amber GB/SA scoring. However, the top models of 4PPT-16:0 and 4PPT-18:0 that docked to the active site were considered for Amber scoring, which yielded reliable negative energies (Fig. 4 and Figure S2).

Out of the top 10 models obtained for protein–protein docking between ACP and JcFatB, five models showed binding of two proteins near the catalytic triad. For JcFatA, nine out of 10 models were consistent (Figure S3). Although 4PPT was not covalently bound to ACP during protein–protein docking, it was observed that the Ser38 residue of ACP, which carries 4PPT-bound fatty acid, was oriented towards the thioesterase active site. The results indicate that the protein–protein interaction between ACP and thioesterase was sufficient to explain the spatial alignment of the crucial residues Ser38 of ACP with the thioesterase (Fig. 3).

Lipid profiles of E. coli K27/fadD88 expressing a FATB cDNA from J. curcas

It was rationalised that over-expression of JcFATB through an established bacterial expression vector may potentially be lethal to the strain (Serrano-Vega et al. 2003). The use of a translational fusion construct in the plasmid Blue Script ensured a low and controlled expression level of the foreign protein. Since this mutant is shown to develop temperature sensitivity when transformed with a vector comprising foreign thioesterase genes, cells were incubated at lower than normal growth temperatures after transformation. E. coli fadD88 cells showed extremely slow growth, even at sub-optimal temperature of
After testing a range of growth temperatures, the best growth rate for the transgenic mutant was observed at 28 °C. Fatty acid profiles were estimated from the cells maintained under the optimised conditions (28 °C, 200 rpm). The fatty acids were quantified in the medium and cell pellet. Palmitate (16:0) was the most abundant (30–40%) fatty acid in all temperature and induction treatment groups. *Jatropha curcas* fatty acyl ACP thioesterase B (*JcFatB*) activity in *E. coli* led to a 50% and 90% increase in stearate in the medium and pellet, respectively (Fig. 5). Since the mutant cannot utilise free fatty acids available in the medium (Dörmann *et al.* 1995), the observed increase in palmitate in the culture supernatant (Fig. 5) indicates a possible affinity of *JcFatB* to palmitoyl (16:0) ACP. However, the total lipid from the culture pellet showed a decrease in palmitate. The palmitoleate (16:1) content decreased in the transformed cells.

**DISCUSSION**

A qualitative profiling of *Jatropha* seed oil showed that nearly 40% of total seed oil was made up of oleate (18:1\(^{\Delta9}\)) and 35% was linoleate (18:2\(^{\Delta9,\Delta12}\)). Stearate (18:0) content was <7% and palmitate (16:0) had a share of 17% (see Fig. S1). Since stearate (18:0) is a minor component of the seed oil, *JcFatB* may have multiple substrate affinity, especially to the palmitoyl or steroyl-ACP in the source. We
hypothesise that in *Jatropha* a mechanism is possibly operating that leads to accumulation of low amounts of stearic acid, even in the presence of a stearoyl-specific thioesterase. It is also possible that *JcFatB* is active only in a small window of expression during early stages of seed development. A semi-quantitative RT-PCR with seed RNA at two different developmental stages indicated a decrease in *JcFatB* expression as the seed matures (data not shown). This indicates strong temporal regulation of thioesterase B in *Jatropha*.

The detailed analysis of the protein sequence and its predicted structural affiliations will be of immense help in understanding the dynamics of plastid targeting of thioesterases and the subsequent enzyme–substrate interactions. It is shown that chloroplast-targeted nuclear-encoded proteins have distinctive subgroup motifs in their transit sequence, which are generally 1–60 amino acids at the N-terminal of the protein (Lee et al. 2008). ChloroP1.1 consistently identified significant potential cTP cleavage sites between positions 49–59 for most of the thioesterases (Accession numbers: Q9FQX9, AY078350, O24419, Q9SQ3Q, DQ856315, DQ847275; data not shown). There is evidence to show that the transit peptides have specific sub-motifs that are processed in multiple steps after the protein is imported into the chloroplast (Bruce 2000; Lee et al. 2006; Agne & Kessler 2009). Although the interactions between the subgroup motifs are perceived to be very complicated, the indication that *JcFatB* has three potential cTP cleavage sites at positions 19, 49 and 59 suggests that its transit peptide would be processed in multiple steps. This could have a bearing on the functional positioning of *JcFatB* in the plastid. Most transmembrane helices of any membrane-bound protein have a length of 18–22 amino acids (Somerville et al. 2004). It is interesting to note that the average periplasmic space between the outer and inner chloroplast membranes is between 6 and 7 nm, which is equal to the length of a linear stretch of around 25 amino acids.

Several reports show that nuclear-encoded chloroplast inner membrane spanning proteins (belonging to the Tic class) are transported to the stroma as pre-proteins (with cTP) and form a soluble intermediate in the stroma (Lubeck et al. 1997; Li & Schnell 2006). The hydrophobic 18 amino acid region of the transit sequence (N-terminal) is shown to be necessary for anchorage with the lipid bilayer of the chloroplast (Pilon et al. 1995). Type B thioesterases such as *JcFatB* could be first targeted to the stroma and then re-inserted into the inner membrane after cTP cleavage. Unlike previous predictions (Facciotti & Yuan 1998), for the protein to be localised in the stroma, the direction of the transmembrane helical domain from amino acid 85 to 107 has to be from cytoplasm towards chloroplasm. We believe that *JcFatB* and, in general, all type B thioesterases are assembled in the chloroplast, such that they have a short sequence of 25 amino acids in the N-terminal region that freely hangs in the space between the outer and inner chloroplast membranes. In plants, the transport flux density of saturated free fatty acid (16:0) across chloroplast membranes is very low compared to mono-unsaturated (18:1\(^{9\alpha}\)) fatty acids (Somerville et al. 2004). Our prediction that *JcFatB* has a short putative N-terminal periplasmic domain supports the view that a thioesterase could be in direct contact with a hypothetical FFA (free fatty acid) transporter involved in the ‘eukaryotic’ pathway (Koo et al. 2004).

The modelled region, where most of the previous reports have identified the existence of a papain-like active site (Ghosh et al. 2007; Mayer & Shanklin 2007) is inside the chloroplast. A short unstable loop from amino acid 357 to 376 is identified by most of the predictors as an inner transmembrane domain, which could also be a re-entrant loop.

It is now shown that seed-specific expression of *JcFatB* increases palmitate (16:0) content in transgenic *Arabidopsis* by three-fold, and the increase in stearate (18:0) is marginal (Wu et al. 2009). In an earlier work that involved overexpression of *Arabidopsis* thioesterase B in its own seeds (Dörmann et al. 2000), it was observed that palmitate content increased from 10% to 39%. However, both results have to be seen in the context that higher plants normally possess higher original amounts of palmitate than stearate in their seeds. Oleate (18:1\(^{6\alpha}\)) is possibly the most abundant fatty acid found in the seeds of the majority of angiosperms. Its synthesis requires desaturation of stearate (18:0), and as a result the relative pool of stearate will always be less than its actual synthesised amount at any given time. Any marginal increase in stearate observed by Wu et al. (2009) could be because of either hydrolysis by FATB or chain elongation of palmitate after it is transported from plastid to cytosol (Dörmann et al. 2000). In most plants, the presence of a thioesterase (type A) with an almost exclusive specificity for oleate and occasional, limited activity towards stearate shows that type B thioesterases are likely to possess varied affinities towards stearate and other saturated fatty acyl ACPs. A combination of a high amount of stearate in seeds along with a rare stearate-specific type A thioesterase is an exception rather than the rule in plants (Hawkins & Kridl 1998).

Availability of substrates for thioesterase action determines the observed abundances in both *E. coli* and *Arabidopsis*. Previous reports on *Arabidopsis* type B cDNA (AFATB) expression in *E. coli* had shown an increase in myristate (Dörmann et al. 1995) and the same gene when expressed in *Arabidopsis* showed a minute increase in myristate (14:0) (Dörmann et al. 2000). However, the same logic does not apply to palmitate (16:0), which is the most abundant and crucial fatty acid for membrane functional dynamics in *E. coli*. When the cloned *JcFatB* was expressed in the *E. coli fadD88* mutant, it is highly unlikely that the plant thioesterase had a substrate limitation with respect to palmitate (16:0) evident from GC-MS lipid profiles of *E. coli* cells. Heterologous expression of *Jatropha* thioesterase B in the *E. coli* mutant shows that *JcFatB* could have greater affinity towards 18:0 than 16:0 and this is difficult to gauge in planta. Comparative docking clearly showed that the potential active site of *JcFatB* did not have any reliable binding affinity with the unsaturated fatty acid ligand (Table S1). Modelled *JcFatA* showed affinity to both unsaturated and saturated fatty acids, with a preference for unsaturated fatty acids. Some reports indicate that stearate is preferred to other shorter chain length saturated fatty acids for chain elongation (Dehesh et al. 1996; Hawkins & Kridl 1998), reducing the relative observed abundance of stearate during plant expression studies. In addition, when in *E. coli*, *JcFatB* is most likely to be a free enzyme, which could significantly alter its physical configuration *in vivo*. We believe that this could be the most plausible combination of factors that can explain the difference in substrate preferences of *JcFatB* when over-expressed in plant and bacterial hosts.
Although plant thioesterase substrate specificity studies carried out in vitro with protein extracts from seeds (crude) and bacterial expression systems (pure) can act as indicators of their natural specificity in vivo, the purification and preparation procedures for extracts are shown to strongly influence enzyme behaviour and substrate specificity (Sánchez-García et al. 2010). Bacterial fermentation is shown to be a promising tool for large-scale production of free fatty acids from plant origin (Lu et al. 2008), making it very important to establish the behaviour of fatty acid mutants such as E. coli fadD88 under the expression of plant lipid biosynthetic genes. Since plant thioesterases and other nuclear-encoded plastid-targeted proteins are highly likely to have originally come from the ancestral prokaryotic plastid genome, and such assumptions have been useful in deriving phylogenetic trees and gene transfers, their characterisation in E. coli is necessary (Jones et al. 1995; Reyes-Prieto & Bhattacharya 2007).

ACKNOWLEDGEMENTS

This work was supported by a grant from the Department of Biotechnology, Government of India. We thank Mr. A.C. Karunakara for his support during GC-MS profiling of lipids. Visualisation and publication quality figures were generated using various tools, including VMD, Sirius and Chimera. We are grateful to Prof. Ram Samudrala (University of Washington, Seattle, USA) for his critical comments on the methods employed in this paper and Mr. K. G. Avinash (IISc, India) for his help in the literature survey. We would like to thank the anonymous reviewers for their valuable suggestions in improving the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. GC-MS lipid profiles.

Figure S2. Top predictions of 4PPT–fatty acid–ACP interactions are represented along with 4-hydroxybenzol coenzyme A (red), 4PPT-16:0 (cyan), 4PPT-18:0 (black) and 4PPT-18:1A9 (blue), which are bound to the active site of (C) JcFatA (green) with ACP (orange). (D) JcFatB with ACP (yellow).

Figure S3. Top 10 protein–protein docking of acyl carrier protein (ACP) with (A) JcFatA in green and (B) JcFatB in blue. Bacterial substrate 4-hydroxybenzol coenzyme A (red) is also given for reference.

Table S1. The binding energy of 4PPT-16:0, 4PPT-18:0 and 4PPT-18:1A9 compounds to JcFatA and JcFatB, scored by grid and Amber with Dock6.4 (kcal mol⁻¹).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

REFERENCES


Dehesh K., Jones A., Knutzon D.S., Voelker T.A. (1996) Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by over-expression of ChFATB2, a thioesterase cDNA from Cuphea hookeriana. The Plant Journal, 9, 167–192.


Factors influencing substrate specificity of plant thioesterases
Supplement 1  GC-MS Lipid Profiles

A

Standard (C4-C24) shown till 25 min out of 50 min

Jatropha Seed Oil Profile

E.coli Control

E.coli JFB3

Retention Time
Table: Retention time of the standard fatty acid methyl esters

<table>
<thead>
<tr>
<th>S No</th>
<th>Average Retention Time (in mins)</th>
<th>Fatty Acid Methyl Ester</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>05.25</td>
<td>Octanoic Acid ( 8:0)</td>
</tr>
<tr>
<td>2</td>
<td>09:18</td>
<td>Lauric Acid (12:0)</td>
</tr>
<tr>
<td>3</td>
<td>12.65</td>
<td>Myristic Acid (14:0)</td>
</tr>
<tr>
<td>4</td>
<td>17.30</td>
<td>Palmitoleic Acid (16:1)</td>
</tr>
<tr>
<td>5</td>
<td>17.80</td>
<td>Palmitic Acid (16:0)</td>
</tr>
<tr>
<td>6</td>
<td>23.16</td>
<td>Linoleic Acid (18:2)</td>
</tr>
<tr>
<td>7</td>
<td>23.28</td>
<td>Oleic Acid (18:1)</td>
</tr>
<tr>
<td>8</td>
<td>24.40</td>
<td>Stearic Acid (18:0)</td>
</tr>
<tr>
<td>9</td>
<td>27.70</td>
<td>Linolenic acid (18:3)</td>
</tr>
<tr>
<td>10</td>
<td>37.80</td>
<td>Erucic Acid (22:1)</td>
</tr>
</tbody>
</table>
Supplement 2: Top predictions of 4PPT--fatty acid--ACP inter-actions are represented along with 4-hydroxybenzol coenzyme A (red), 4PPT-16:0 (cyan), 4PPT-18:0 (black) and 4PPT-18:1Δ9(blue), which are bound to the active site of (C) JcFatA (green) with ACP (orange). (D) JcFatB with ACP (yellow).
Supplement 3: Top 10 protein–protein docking of acyl carrier protein (ACP) with (A) JcFatA in green and (B) JcFatB in blue. Bacterial substrate 4-hydroxybenzol coenzyme A (red) is also given for reference.
Table S1

Table S1: The binding energy of 4PPT-16:0, 4PPT-18:0 and 4PPT-18:1<sup>†</sup> compounds to \( \text{JcFatA} \) and \( \text{JcFatB} \) scored by Grid and Amber with Dock6.4 (kcal/mol)

<table>
<thead>
<tr>
<th>Model Number</th>
<th>( \text{JcFatA} )</th>
<th>( \text{JcFatB} )</th>
<th>( \text{JcFatA} )</th>
<th>( \text{JcFatB} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grid Score</td>
<td>Amber Score</td>
<td>Grid Score</td>
<td>Amber Score</td>
</tr>
<tr>
<td>1</td>
<td>-99.478386</td>
<td>-124.304161</td>
<td>-104.065681</td>
<td>-116.347893</td>
</tr>
<tr>
<td>2</td>
<td>-99.274185</td>
<td>-78.777115</td>
<td>-103.904198</td>
<td>-118.185188</td>
</tr>
<tr>
<td>3</td>
<td>-98.985565</td>
<td>-150.150299</td>
<td>-103.805847</td>
<td>-133.489792</td>
</tr>
<tr>
<td>4</td>
<td>-98.755936</td>
<td>-67.120712</td>
<td>-103.712158</td>
<td>-138.220215</td>
</tr>
<tr>
<td>5</td>
<td>-98.600243</td>
<td>-136.337952</td>
<td>-103.591530</td>
<td>-133.489792</td>
</tr>
<tr>
<td>6</td>
<td>-98.160767</td>
<td>-155.507370</td>
<td>-103.567459</td>
<td>-140.234207</td>
</tr>
<tr>
<td>7</td>
<td>-98.057381</td>
<td>-143.887173</td>
<td>-103.241844</td>
<td>-140.187347</td>
</tr>
<tr>
<td>8</td>
<td>-97.733666</td>
<td>-93.674973</td>
<td>-103.206421</td>
<td>-122.743271</td>
</tr>
<tr>
<td>9</td>
<td>-97.837807</td>
<td>-106.273186</td>
<td>-103.074242</td>
<td>-111.322182</td>
</tr>
<tr>
<td>10</td>
<td>-97.847820</td>
<td>-123.588264</td>
<td>-102.949867</td>
<td>-155.479141</td>
</tr>
</tbody>
</table>

The five models with their scores highlighted in **bold**, are selected because they have the most favorable conformations with high binding affinities ¥. As only 3 models from previous blind docking matched with binding site, only those were considered for Amber Scoring function. † As none of the models from previous blind docking matched with the binding site, this ligand was not considered for Amber scoring with \( \text{JcFatB} \).