Fatty acid binding proteins in the fall armyworm, *Spodoptera frugiperda*

by

Xin Jiang

B.Sc., Simon Fraser University, 2009

Thesis Submitted In Partial Fulfillment of the
Requirements for the Degree of
Master of Science

In the
Department of Biological Sciences
Faculty of Science

© Xin Jiang 2013

SIMON FRASER UNIVERSITY

Summer 2013

Approval

Name: **Xin Jiang**

Degree: **Master of Science (Biological Sciences)**

Title of Thesis: **Fatty acid binding proteins in the fall armyworm,**
Spodoptera frugiperda

Examining Committee: **Chair:** Dr. David Green
Associate Professor

Dr. Norbert Haunerland
Senior Supervisor
Professor

Dr. Allison Kermode
Supervisor
Professor

Dr. William Davidson
Supervisor
Professor
Department of Molecular Biology and
Biochemistry

Dr. Julian Christians
Internal Examiner
Associate Professor

Date Defended/Approved: July 26, 2013

Partial Copyright Licence

SFU

Abstract

The expression of fatty acid binding protein (FABP) is often up-regulated during high energy demanding activities. In this study, we have confirmed the expression of two FABPs in the flight muscles of the fall armyworm, *Spodoptera frugiperda*. Based on real-time PCR results, the expression of both genes appeared to be affected by flight. The promoter of one gene was sequenced and cloned into a GFP reporter vector for a reporter assay using the SF9 cell line. The expression of this gene in the SF9 cell line does not appear to be up-regulated in response to fatty acid treatment. Therefore, alternative strategies to analyze the regulation of this gene are required.

Keywords: FABP; *Spodoptera frugiperda*; real-time PCR, GFP, SF9 cell

To my beloved parents,

John Jiang and Ruth Jiang

Acknowledgements

I would like to give thanks to everyone who has contributed in completing this thesis. Thank you all for spending time helping me during your busy hours; and I wish you all become more successful in the future.

First I would like to sincerely thank my senior supervisor, Dr. Norbert Haunerland, for his guidance, encouragement and patience in the past four years. Without his knowledge, understanding and support, I would never be able to overcome the difficulties I encountered during my research and study. I would also like to give thanks to members of my supervisory committee, Dr. William Davidson and Dr. Allison Kermode, for their guidance and valuable advice. My gratitude is also extended to my external examiner, Dr. Julian Christians, whom I also really enjoyed working with during my teaching assistantship semesters.

I would like to show my special thanks to Dr. Jutta Rickers-Haunerland for her great inspirations and support during the most difficult times in my research. Many thanks to Katrina Salvante for her assistance in sequencing, to Cesily Bernales for her contributions in cloning and cell culture technique development, and to Ahmed Adel for his generous supply of the insects he raised. I also appreciate to all present and past members of the Haunerland Lab for making this working environment an enjoyable place during my graduate training.

I would finally like to acknowledge the great support from my friends and family. Special thanks to my parents, John Jiang and Ruth Jiang, for their understanding, support, and unconditional love. You are always my heroes.

Table of Contents

List of Tables

List of Figures

x

List of Acronyms

AA	Arachidonic acid
AKH	Adipokinetic hormone
COUP-TF	Chicken ovalbumin upstream promoter transcription factor
CRABP	Cellular retinoic acid binding protein
CTE	Carboxy-terminal extension
DAG	Diacylglycerol
DBD	DNA binding domain
DR	Directed repeat
EcR	Ecdysone receptor
ER	Everted repeat
EST	Expressed sequence tag
FA	Fatty acid
FABP	Fatty acid binding protein
FAT	Fatty acid translocase
FATP	Fatty acid transport protein
FXR	Farnesoid X receptor
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescence protein
HDLp	High density lipophorin
HRE	Hormone response element
IBABP	Ileal bile acid binding protein
iLBP	Intracellular lipid binding protein
IR	Inverted repeat
JH III	Juvenile hormone III
LA	Linoleic acid
LBD	Ligand binding domain
LDLp	Low density lipophorin
NR	Nuclear receptor
NUP	Nested universal primer A
OA	Oleic acid
OpMNPV	*Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus
ORF	Open reading frame
PA	Palmitic acid
PCR	Polymerase chain reaction
PPAR	Peroxisome proliferator-activated receptor
PPRE	Peroxisome proliferator response element
RACE	Rapid amplification of cDNA ends

RAR	Retinoic acid receptor
RARE	Retinoic acid response element
RORα1	RAR-related orphan receptor alpha 1
RT	Reverse transcriptase
RXR	Retinoic X receptor
TAG	Triacylglycerol
TSS	Transcription start site
UPM	Universal primer mix A
USP	Ultraspiracle
UTR	Untranslated region
VDR	Vitamin D receptor
VLDL	Very low-density lipoprotein

1. Introduction

1.1. Fatty acid transport in mammals and insects

In animal muscle cells, a variety of molecules can be used to provide energy during contractions. When muscles undergo short-term contraction, carbohydrates are normally used as the primary source because they can be quickly mobilized and metabolized. For long-term contraction activities, such as doing a Marathon run, the amount of carbohydrates stored in cells is not sufficient to meet the energy demand. Under those circumstances, fatty acids (FAs) serve as the major energy-providing molecules, because they contain more energy per unit mass than carbohydrates.

In most animals, fatty acids are not abundant in muscle cells (but are stored in adipose tissues). Thus, in order to let the energy production rate to keep up with the continuous consumption of ATP during sustained muscle contraction periods, fatty acids are constantly transported through the circulatory system from adipose tissues to the target sites. Fatty acids are taken up by the muscle cells and sent to the mitochondria, where beta-oxidation releases the energy needed for muscle contraction. While the general process of fatty acid transport and utilization appears to be conserved in most animals, several noteworthy differences can be spotted between mammals and insects.

In mammals, fatty acids are normally stored in the form of triacylglycerol (TAG), which is synthesized and stored in adipose cells. When required, TAG is hydrolyzed by the action of a lipase, and free fatty acids are released into the bloodstream, where they bind to serum albumin and flow to target sites (Figure 1-1). Alternatively, TAG, embedded in the core of lipoproteins, can be transported to muscles. For dietary and endogenous fatty acids, the corresponding carrier lipoproteins are chylomicrons and very low-density lipoproteins (VLDLs), respectively (Van der Vusse et al., 1998). These particles consist of a core made of TAG, surrounded by phospholipids and proteins.

Once at the target sites, TAG is converted into free fatty acids and glycerol by the action of lipoprotein lipase (Figure 1-1) (Van der Vusse et al., 1998).
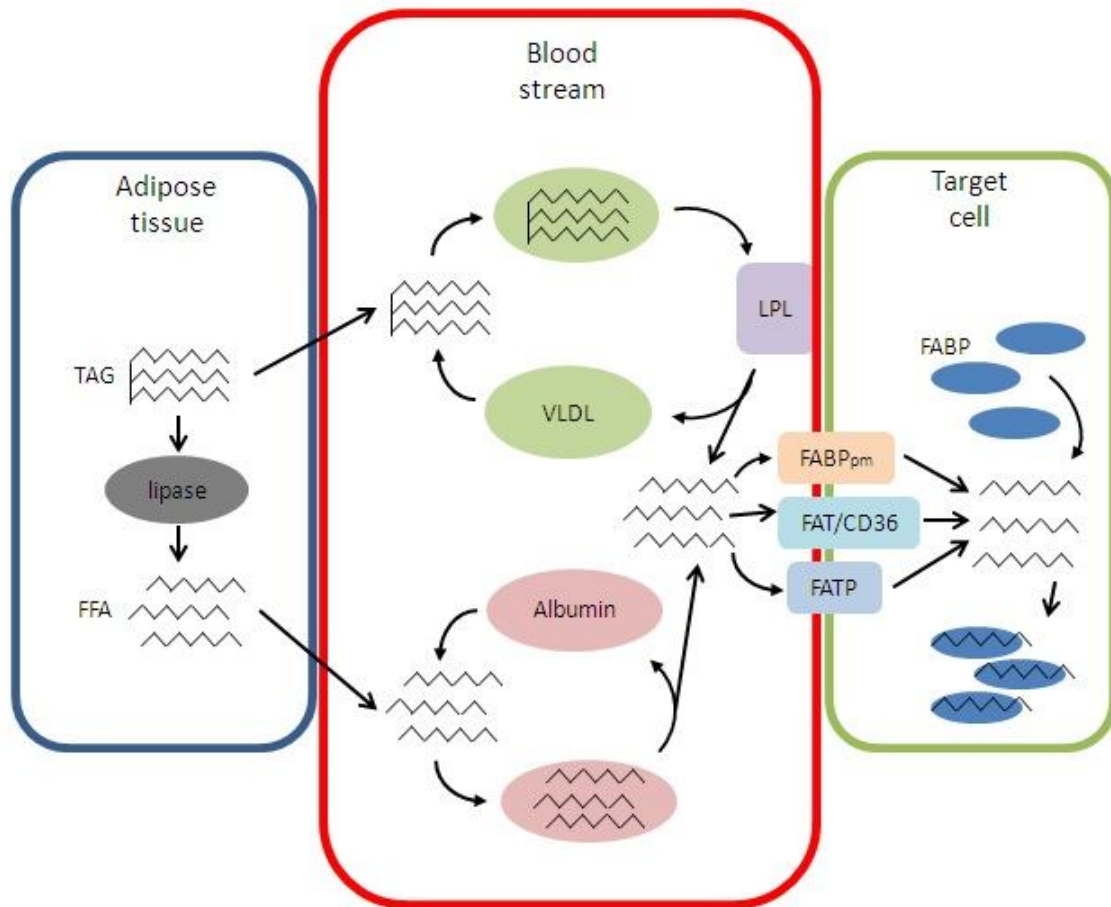
Figure 1-1   Schematic representation of fatty acid transport in mammals

TAG - triacylglycerol; FFA - free fatty acid; VLDL - very low density lipoprotein; LPL - lipoprotein lipase; FABPpm - plasmalemmal fatty acid binding protein; FAT/CD36 - fatty acid translocase; FATP - fatty acid transport protein; FABP - fatty acid binding protein

In the past, it was believed that fatty acid molecules are able to penetrate the lipid bilayer cell membrane and enter cells by passive diffusion due to their hydrophobic nature. More recent findings suggest that in addition to passive diffusion, a variety of proteins are involved in transporting fatty acids into the intracellular compartment. These transmembrane-trafficking proteins, including plasmalemmal fatty acid binding protein (FABPpm), fatty acid translocase (FAT/CD36), and fatty acid transport protein (FATP), ensure that the water-insoluble fatty acid molecules are carried efficiently across

cell membrane (Baillie et al., 1996; Bonen et al., 2003; Jia et al., 2007). Once inside the cells, cytosolic fatty acid binding protein (FABP) will bind the fatty acids and transport them from the sarcolemma to the mitochondria for oxidation. Binding to FABPs may also prevent that the detergent-like fatty acid molecules damage cellular proteins and membrane structures (Glatz et al., 1998).

Insects appear to be similar to mammals with respect to extended muscle activity, such as sustained flight, which also relies on fatty acid as the major fuel source. For example, the migratory locust is capable of sustaining flight activity for more than 10 hours. In the initial phase of the flight, the energy used is mainly supplied from carbohydrates, but when the sugar pool is depleted, the insect switches to fatty acid as the main source of ATP production (Van der Horst, 2001).

In order to meet the high energy demand in the flight muscles, insects have developed a shuttle system that efficiently transports fatty acids from storage tissues to their target sites (Figure 1-2). In insects, fatty acids are stored in the form as TAGs in fat body, a multifunctional organ involved in lipid storage and protein synthesis. In response to flight, adipokinetic hormone (AKH) released from the corpus cardiacum triggers the activation of TAG lipase, which hydrolyzes TAG into diacyglycerol (DAG), which in turn is released into the hemolymph (Van der Horst et al., 2001). In the hemolymph, DAG binds to the surface of a lipoprotein known as high density lipophorin (HDLp), which is structurally distinct from mammalian lipoproteins. To reduce exposure of the hydrophobic DAG molecules to the aqueous surroundings of the hemolymph, additional apolipoprotein molecules (apolipophorin III, apoLp-III) assemble to the surface of the complex. This lipid-enriched lipophorin has a lower density than HDL, and is therefore referred to as low density lipophorin (LDLp). When LDLp particles reach the muscle tissues, apoLp-III and the DAG is released, returning lipophorin into its HDLp state. HDLp is now ready for reloading of lipid molecules at the fat body, and the cycle of loading and unloading can continue (Weers and Ryan, 2003). Once released, DAG molecules are hydrolyzed at the flight muscle cell membrane into free fatty acids and glycerol by a lipophorin lipase, and free fatty acid is taken into the cells, where intracellular FABP facilitates transport to the mitochondria for subsequent β-oxidation (Figure 1-2).
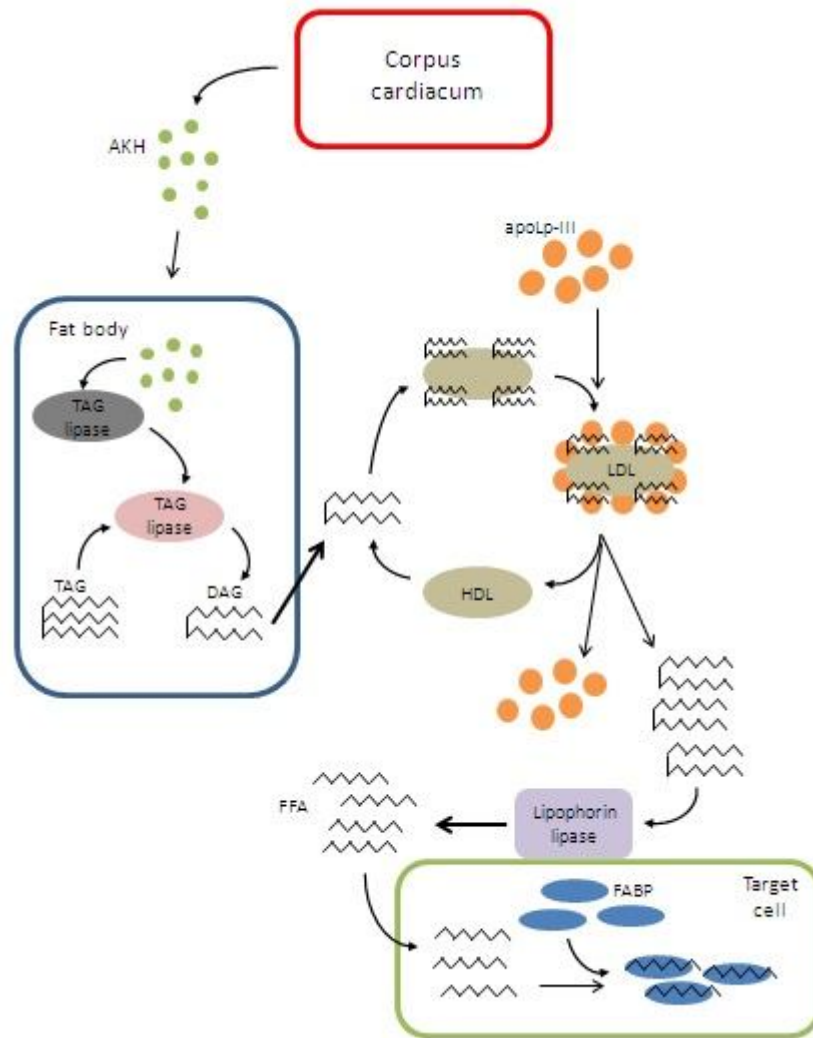
Figure 1-2   Schematic representation of fatty acid transport in insects

AKH - adipokinetic hormone; TAG - triacylglycerol; DAG - diacylglycerol; apoLp-III - apolipophorin III; LDLp - low density lipophorin; HDL - high density lipophorin; FFA - free fatty acid; FABP - fatty acid binding protein.

1.2. Cytosolic FABP

Within cells, cytosolic FABPs are responsible for transporting fatty acids to the mitochondria. The binding of fatty acids to FABPs results in a gradient of free fatty acids between the two sides of the cell membrane, thus enhancing the influx of free fatty acids (Weisiger, 1996). This results not only in a pool of intracellular fatty acid as substrates for β-oxidation, but also reduces the amount of unbound free fatty acids on both sides of

the membrane, which could be damaging to cellular membranes due to their detergent-like property (Glatz et al., 1998). Therefore, cytosolic FABPs seem to participate in both fatty acid transport and cellular protection.

Cytosolic FABPs belong to a conserved multigene family known as intracellular lipid binding proteins (iLBPs). Members of this family have been identified throughout the animal kingdom, both in vertebrates and invertebrates; however, none have been identified in plants and fungi. It is believed that this gene family evolved after the separation of the animal kingdom from other organisms, more than 1000 million years ago (Haunerland and Spencer, 2004). The members of this family originated from their common ancestor through gene duplications before the vertebrate-invertebrate split, giving rise to two major branches on the phylogenetic tree (Schleicher et al., 1995).

All members of the iLBP family are characterized by a common three-dimensional structure, which consists of a barrel made of antiparallel beta strands, and two alpha helices that form the lid of the barrel (Figure 1-3). Stabilized through combination of enthalpic and entropic forces governed by the tertiary structure of the protein, a single fatty acid molecule can be enclosed within the barrel with strong affinity (Banaszak et al., 1994).

The different types of FABPs were originally named based on the tissue from which they were discovered. However, this naming system is not very accurate because most FABPs are found expressed in tissues other than their discovery site. For example, mammalian heart FABP (H-FABP) is synthesized in abundant quantity in heart muscles, but it has also been found in lower quantities within other tissues, such as skeletal muscles, brain, stomach, and lungs (Table 1-1). In addition, often more than one type of FABP is present in a single tissue type, as in the adipose tissue, where L-FABP and I-FABP were found to coexist (Murphy, 1998).

Table 1-1   Members of the iLBP family

iLBP type	Mammalian expression site	Ligand
Subfamily I		
CRBP I, II, III, and IV	Ubiquitous in mammalian cells	Retinoid
CRABP I and II	Ubiquitous in mammalian cells	Retinoic acid
Subfamily II		
L-FABP (liver)	Liver, intestine, kidney, lung, and pancreas	Long chain fatty acids, acyl-CoA, heme
I-BABP (intestinal)	Ileum	Long chain fatty acids
Lb-FABP (liver-basic)		
Subfamily III		
I-FABP (intestinal)	Intestine, liver	Long chain fatty acids
Subfamily IV		
H-FABP (heart)	Heart, skeletal muscle, brain, kidney, lung, mammary, placenta, testis, ovary, stomach	Long chain fatty acids
A-FABP (adipocyte)	Adipose tissue, liver, and macrophages	Long chain fatty acids
E-FABP (epidermal)	Skin, adipose tissue, lung, brain, heart, skeletal muscle, testis, retina, and kidney	Long chain fatty acids
B-FABP (brain)	Brain, glia cells, and retina	Long chain fatty acids, docosahexahenoic acid
M-FABP (myelin)	Brain and Schwann cells	Long chain fatty acids
T-FABP (testis)	Testis	Long chain fatty acids

Haunerland, 2004, used with permission.
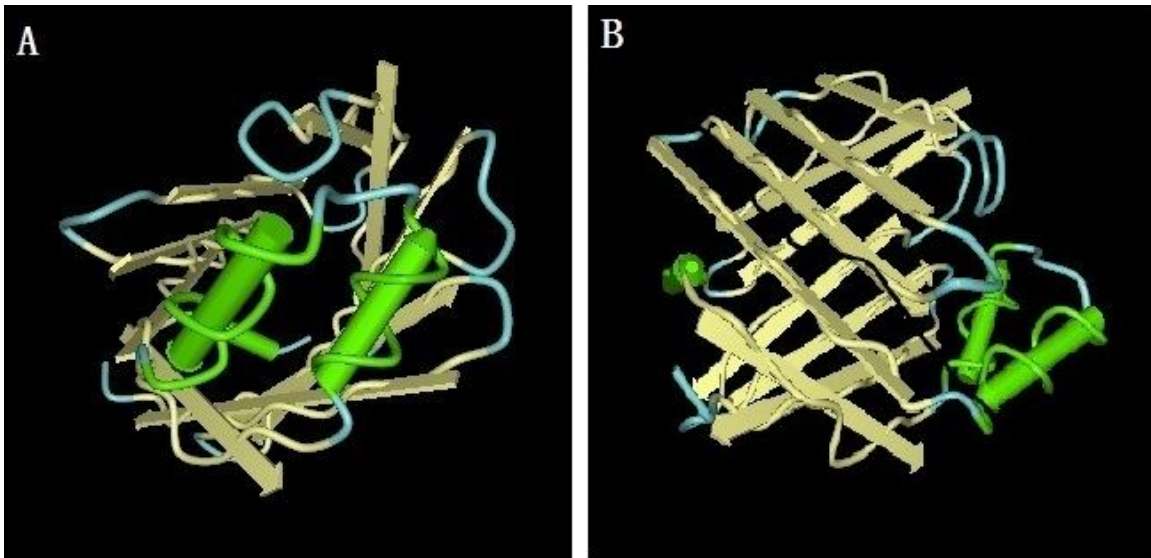
Figure 1-3   Structure representation of the iLBP family

A: top view from the opening of the beta barrel. B: side view. Yellow highlights: beta strands. Green highlights: alpha helices. Picture from Protein Data Bank (ID: 2RCT). Tarter et al, 2008, used with permission.

1.3. FABP Expression

In both mammals and insects, high metabolic activities such as repeated cycles of muscle contraction and relaxation depend on fatty acids as the primary energy source. Based on this fact, one would anticipate that the amount of cytosolic FABP produced in muscle cells is correlated to the rate of metabolism. This hypothesis was supported by evidence in which rats trained with weeks of endurance exercise possessed a ~30% higher cytosolic FABP content in the muscles than untrained rats (Clavel et al., 2002). Moreover, after the desert locust *Schistocerca gregaria*, whose metabolic activity in the muscle can increase up to 20 times during flight (Kammer and Heinrich, 1978), was subjected to endurance training by sustained flight, the expression of muscle FABP increased several folds (Haunerland, 1994). Increases in FABP expressions are also common in other tissues where lipid metabolism is prominent, such as hepatocytes, macrophages, and cardiac myocytes (Kazemi et al, 2005; Vida et al, 2013; Zheng et al, 2012). Therefore, the rate of FABP synthesis appears to have a positive correlation with lipid metabolic activity, and the amount of intracellular fatty acid. This relationship is often controlled by transcription factors such as the peroxisome proliferator-activated

receptor (PPAR) (Huang et al, 2012), which belongs to a group of ligand activated transcription factors known as nuclear receptors (NRs).

1.4. Nuclear receptors

Nuclear receptors are often involved in regulating various physiological processes, from development and reproduction, to homeostasis and metabolism within the animal kingdom. Upon binding to a ligand, nuclear receptors will respond by interacting with the regulatory regions of target genes, thus affecting gene expression (Owen and Zelent, 2000). Their ligands are usually small, hydrophobic molecules, range from steroid and thyroid hormones, to retinoic acids, vitamin D, and fatty acids (Laudet, 1997). For some members of the NR superfamily, called "orphan receptors", ligands have not been identified and/or may not exist; they in turn are thought to alter the gene expression pattern in a ligand-independent manner (Giguere, 1999).

Throughout the animal kingdom, the general structure of nuclear receptors is known to be evolutionarily conserved. Nuclear receptors share a common organization of five major domains (A/B-C-D-E-F) (Brelivet et al, 2004): near the end of the N-terminus is the transactivation AF-1 domain (A/B), whose sequence is less conserved than the other domains. Following AF-1 is the DNA binding domain (DBD), region C, which, upon binding to the DNA sequence, will enable AF-1 to activate transcription. This is also the most conserved domain with two zinc finger motifs, which play crucial roles in DNA binding and are found in all members of the superfamily, plus a carboxy-terminal extension (CTE). In some cases, two sets of DBD can exist in tandem on a single nuclear receptor (Wu et al., 2007). Following the AF-1 and DBD is the hinge region D, which participates in nuclear localization signaling and linking the C-terminal ligand binding domain (LBD). The overall architecture of the LBD is well conserved; however, variations in the amino acid sequence still allow a high degree of ligand specificity between family members. The LBD is usually linked to a second transactivation domain, known as the AF-2 domain, which is responsible for recruiting coactivators upon conformational change due to ligand binding (Renaud et al., 1995). The C-terminal F domain shows little evolutionary conservation, and is even missing in certain members; it may play a role in coactivator binding (Figure 1-4).
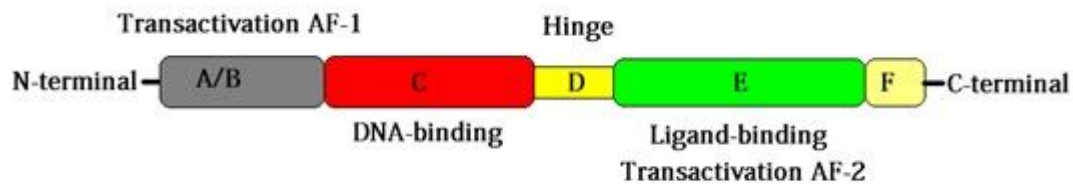
Figure 1-4   *Schematic structure of a nuclear receptor*

The DNA sequence motifs recognized by nuclear receptors are known as hormone response elements (HREs). These elements usually contain one or two of the consensus half-site sequence PuGGTCA (where Pu is a purine). These half-sites can be configured as direct repeats (DRs), inverted repeats (IRs) or everted repeats (ERs), each with a number of spacer nucleotides in between (Escriva et al., 2004). Nuclear receptors can bind to HRE sequences as monomers, homodimers, or heterodimers. For monomeric receptors, a single half-site preceded by a 5'-flanking A/T rich sequence is usually recognized (Giguere, 1999). Retinoid X receptors, on the other hand, are known to form dimers with itself (homodimers), or dimers with other receptors (heterodimers), such as PPAR (RuizLozano et al., 1997). Their binding sites usually exist in the direct repeat of an imperfect NR consensus sequence AGGTCAnAGGTCA (also known as a DR1 element).

1.5. Nuclear receptors control FABP expression

In mammals, upon ligand binding, nuclear receptors RXRα and PPARα, are known being able to dimerize and subsequently binding to peroxisome proliferator response elements (PPREs), which are located in the promoter of target genes, thereby regulating transcription (Fruchart, 2001). Transcription activity controlled by this type of heterodimers can be regulated through phosphorylation or coactivator modulations (Aubry et al., 1999). Some of the target genes controlled by the heterodimers are involved in fatty acid transport in tissues most capable of catabolizing fatty acids. These include the FATP, FAT/CD36 coding genes, and certain members of the iLBP family (Motojima et al., 1998). For instance, liver FABP expression can be up-regulated at the mRNA level upon the binding of the RXRα-PPARα to a PPRE in the promoter (Poirier et al., 1997); ileal bile acid binding protein (IBABP) expression is regulated via binding of

farnesoid X receptor (FXR) and RXRα to an inverted repeat element (Hwang et al., 2002); and the expression of cellular retinoic acid binding protein II (CRABP-II) is controlled by RXR dimerizing with retinoic acid receptor (RAR), followed by binding to a DR5 retinoic acid response element (RARE) (Hewson et al., 2002).

Although the natural ligands of some of these nuclear receptors remain unknown, there is evidence that points to various polyunsaturated long chain fatty acids, branched fatty acids, or eicosanoids (Gronemeyer et al., 2004). For example, RXRα, besides being capable of binding to 9-cis retinoic acid and methoprene (an insect juvenile hormone analog that is often used as a retinoic acid mimic), can also be activated by docosahexaenoic acid, a long-chain polyunsaturated fatty acid enriched in mammalian brain (Harmon et al., 1995). Thus, it is hypothesized that in addition of being substrates of members of the iLBP family, fatty acid molecules themselves may also play roles as signaling molecules, which interact with nuclear receptors and control gene expressions, inducing the expression of their own binding proteins. Nevertheless, nuclear receptors that control the expressions of iLBP genes have never been found in insect species.

1.6. Research goal

The first insect FABP was isolated from the flight muscle of the migratory locust *S. gregaria* in 1990 (Haunerland and Chisholm, 1990), and later found in a different locust species, *Locusta migratoria* (Maatman et al., 1994). This gene was found to be up-regulated in the flight muscle when the locust species *S. gregaria* was tethered to fly over long periods (Haunerland, 1994). Ever since its discovery, the number of FABPs isolated from insects has been increasing: in 1992, two FABPs were found to be abundantly expressed in the midgut of the tobacco hornworm *Manduca sexta* (Smith et al., 1992), in which one had its 3D structure solved (Benning et al., 1992); in 2000, a FABP was discovered from the thoracic muscle of adult triatomine bug *Dipelogaster maximus* (Cavagnari et al., 2000); in 2011, Huang et al. have also cloned a FABP gene from the midgut of the oriental leafworm moth *Spodoptera litura*, in which expression was down-regulated by starvation but up-regulated by feeding (Huang et al., 2011). Despite the fact that new FABPs have been constantly discovered and characterized in many insect species, no one has attempted to study the expression of these genes

10

during high metabolic activities such as flight after it was first performed on the locusts (Haunerland, 1994).

The goal of this study was to identify any FABP genes in the flight muscle of the fall armyworm *Spodoptera frugiperda* (Lepidoptera, Noctuidae), a holometabolous insect, and study their regulation by flight on fatty acids. This insect is native to the tropical regions of the Western hemisphere from the United States to Argentina. The life cycle of this insect consists of the egg stage, six larval stages, a pupal stage, and the adult moth stage. Similar to the migratory locusts, adult moths are also known for their migratory characteristic, which increases the range of egg dispersal, sometimes dramatically. On one occasion, it was documented that the long range movement from Mississippi to Canada was achieved within thirty hours on a low level jet stream (Johnson, 1987). Due to its migratory characteristic, we anticipated that the regulation pattern of the FABPs in its flight muscle would be similar to the locusts; and since FABP expressions in mammals are often controlled by nuclear receptors, we believed that seeking for potential nuclear receptor binding sites within the promoter sequences would provide us insights about the transcription regulations of insect FABPs.

2. Search for a muscle FABP

2.1. Introduction

The first two FABPs within the insect order Lepidoptera were found in the midgut of *Manduca sexta* larvae, which possess large amounts of lipid. When isolated, these two proteins (named FABP1 and FABP2, accession P31416.1 and P31417.3) were found to contain bound fatty acids in a 1:1 molar stoichiometric ratio (Smith et al., 1992). Analysis of the alignment between each of the two amino acid sequences with mammalian FABPs have revealed a number of shared features, including 9 conserved glycine residues, a basic residue that may participate in binding to the carboxyl end of fatty acid molecules, and conservation of many residues that are important in binding to the hydrophobic tail of the fatty acid. Immunological screens in various *Manduca sexta* tissues have demonstrated that the expression of the two genes is restricted to the midgut.

A third FABP in *M. sexta* was discovered in brain tissue in 1998 (accession AAC24317.1) (Mansfield et al., 1998). The amino acid sequence of this protein is most similar to the CRABP I sequences from bovine (sequence similarity 71%). Moreover, it was suggested that the promoter region of this gene contains a motif resembling a mammalian retinoic acid response element. Based on these evidences, Mansfield et al. concluded that retinoic acid plays a role in insect signal transduction, and named the protein *Manduca sexta* cellular retinoic acid binding protein (msCRABP). However, through competitive binding assays it was later discovered that this iLBP had negligible interactions with retinoic acid but instead bound fatty acids with far higher affinity (Folli et al., 2005). Given that the sequence is equally similar to vertebrate CRABP and heart-type FABP, the protein should probably be designated as a FABP. To avoid confusion, however, this lepidoptera member of the iLBP family will still be designated as CRABP, based on the original discovery.

In Noctuids, an additional FABP was discovered from the fat body of the fourth instar of the corn earworm *Helicoverpa zea* (1998, accession AAC25674.1) and, later in *Helicoverpa assulta* (2005, accession ABC02870.1). In this study, it was named as FABP3.

While one of the two midgut FABPs has recently been found in *Spodoptera* litura (Huang et al, 2011), no other FABPs have been identified in *Spodoptera* species. In this study, we seek to investigate the expression of the 4 iLBPs mentioned above within three types of tissues of the fall armyworm, *Spodoptera frugiperda*: sixth instar larvae fat body, adult flight muscle, and a cell line derived from pupae ovarian cells, also known as the SF9 cell line. Moreover, since the organism is known to engage in migratory flight similar to the migratory locust species *S. gregaria* (see Chapter 1.6), we have also decided to analyze the expression of any FABP(s) in the flight muscle during flight. In the locust, expression of flight muscle FABP is only present in the adult state. Its expression begins immediately after adult ecdysis and continues at high rates for almost two weeks (Haunerland, 1997). Once reaching a constant level, the locusts can utilize fatty acid as their primary energy source for endurance activity and hence are able to fly for long periods of time. During flight, muscle FABP expression was shown to continue to increase. When 20 days old adult locusts were tethered to fly, the level of FABP mRNA increased up to 12-fold. However, this increase was observed only after a delay of at least 4 hours of rest; and this phenomenon was observed for locusts that underwent flight activity for 2, 4, 8 or 12 hours (Chen and Haunerland, 1994). Flight times shorter than 1 hour had no influence in FABP expression. Since *S. frugiperda* exhibits migratory behavior similar to the desert locust, we hypothesized that the expression of *crabp* and/or *fabp3* could be regulated in similar ways during flight.

2.2. Materials and methods

2.2.1. Total RNA extraction

Total RNA extraction was carried out according to the manufactorers' protocol (QIAGEN's RNeasy® Mini Kit, Cat#: 74104). To extract TOTAL RNA from flight muscle, total thoracic muscle tissues from an adult *S. frugiperda* were excised and immediately

frozen in liquid nitrogen. The frozen tissues were ground into powder under liquid nitrogen, followed by suspension in 350 µL of buffer RLT (containing β-mercaptoethanol). The mixture was homogenized by passing through a 16 Gauge needle for 15 times. The SF9 adherent cells were cultured in 5 mL of growth medium (500 µL of fetal bovine serum (Gibco® Sera, Cat#: 12483-020), 4.5 mL of Grace's supplemented insect medium (Invitrogen, Cat#: 11605-102), 5 µL of 10 mg/mL gentamicin solution (Invitrogen, Cat#: 15710-064)) within the 25 cm² Corning cell culture flask (Fisher Scientific, Cat#: 10-126-28) at 27˚C. To keep the cells at the log phase of growth, they were sub-cultured using fresh medium when reached roughly 80%. When required for RNA extraction, roughly 1×10^6 cells at passage #20-35 were manually detached from the bottom of the flask. The cells were collected and briefly spun at 800 g, then washed twice with 1 mL of PBS. The cell pellet was re-suspended using 350 µL of buffer RLT, and homogenized by passing 10 times through a 21 Gauge needle. For extraction of RNA from fat body tissue, fresh peripheral fat body from sixth instar larvae were excised and mixed with 350 µL of buffer RLT, followed by homogenization in the same manner as the SF9 cells.

After the total content for each sample was thoroughly mixed with 350 µL of ice cold 70% ethanol by pipetting for 10 times, it was loaded onto the RNeasy Mini spin column. The homogenate was filtered through the column by spinning at 13000 rpm for 15 seconds, followed by three additional washes: first with 700 µL buffer RW1, followed by spinning for 15 seconds, and then second and third wash each with 500 µL buffer RPE, followed by spinning for 15 seconds and 2 minutes, respectively. The column was dried by spinning at 13000 rpm for 1 minute. Subsequently, bound RNA was eluted from the column with 30 µL of RNase free water, and the RNA concentration in the elute determined using the NanoDrop spectrophotometer.

2.2.2. cDNA synthesis

Total RNA (1000 ng) from muscle, SF9 cells, or fat body was pipetted and mixed with 4 µL of the 5X iScript reaction mix and 1 µL of iScript reverse transcriptase from BioRad's iScript™ cDNA Synthesis Kit (Cat#: 170-8890). The total volume of the reaction mixture was brought to 20 µL by adding nuclease free water. The thermal parameters were: 5 minutes at 25˚C, followed by 30 minutes at 42˚C, and 5 minutes at 85˚C. cDNA concentration was then diluted to 25 ng/uL for each sample.

2.2.3. PCR

In order to detect the expression of the four iLBP genes, PCR primers were designed based on EST sequences obtained using the tblastn tool of the online EST data base SPODOBASE (Negre et al, 2006). Single-stranded cDNA (100 ng) from flight muscle, SF9 cells, or fat body were mixed with 12.5 µL of Fermentas' 2X PCR Master Mix (Cat#: K0172), along with 10 µmoles of forward and reverse primer of each gene (Appendix A). Total volume of each reaction was brought to 25 µL by addition of nuclease free water. Thermal cycling parameters were: 5 minutes of initial denaturation at 98°C, followed by 35 cycles of 30 seconds at 98°C, 30 seconds at 56°C, and 30 seconds at 72°C, and a final elongation time of 7 minutes at 72°C. PCR products were analyzed on 1.0% agarose gel, and then purified with the QIAquick Gel Extraction Kit from QIAGEN (CAT#: 28704) and sent for Sanger's sequencing using the corresponding PCR primers.

2.2.4. Flight experiment

S. frugiperda moths were held at 27°C for 11 to 21 days after adult eclosion. Under these conditions, female adults usually start laying eggs 7 days after emergence, and therefore flight experiments were conducted on day 6. At that age, most male adults were able to sustain flight for a maximum of 2.5 hours; whereas females could usually fly for at least 4 hours. As experimental subjects, 16 females were divided into four test groups of equal size: three groups were allowed to fly for 4 hours, followed by 0, 2, or 4 hours of resting period, respectively; whereas the control group was neither put to fly nor rest. Before flight, each moth was anesthetized using carbon dioxide. Next, a piece of corrugated fiberboard was glued to the dorsal side of the cuticle (below the wings) of using epoxy adhesive glue. Once the glue had dried, the moth was awakened by being released from the carbon dioxide. Each fiberboard was fixed on the top of a metal ring stand, having the posterior end of the insect facing an electrical fan. To initiate flight, the fan was turned on to medium rotation level to generate wind flowing towards the posterior end of the moth, causing it to fly continuously at room temperature (Figure 2-1). After flying for a constant 4 hours period, the moth was released from the apparatus and allowed to rest at 27°C for 0, 2, or 4 hours. Once the resting period was

over, the entire thoracic muscles of each moth was excised and used for total RNA extraction and cDNA synthesis (section 2.2.1).
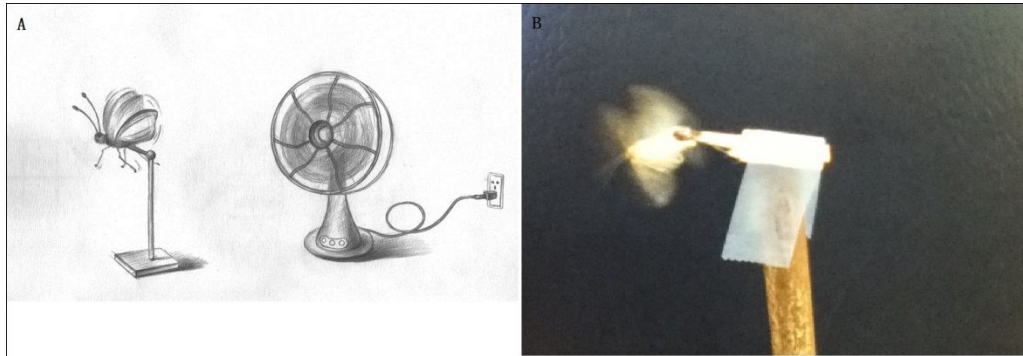
Figure 2-1   Flight experiment apparatus
A: sketch of the apparatus. B: photo of a moth flying in action.

2.2.5.   Optimization of real-time PCR

The primers targeting the mRNA of *crabp*, *fabp3*, *rpl37a*, and *gapdh* were intended to be used as real-time PCR primers. In order to test their efficiency, a regular PCR was performed for each pair of primers using the conditions outlined in section 2.2.3.   After the PCR products were purified and subjected to DNA concentration measurement, each was diluted with nuclease free water in a 10-fold dilution series to obtain seven standard solutions, ranging from 1.0 ng/μL to 1.0 fg/μL. Next, 1 μL of each standard solution was mixed with 5 μmoles of each corresponding primer, along with 5 μL of BioRad's SsoFast™ EvaGreen® Supermix (Cat#: 172-5200). Nuclease free water was added into the mixture to bring the total reaction volume to 10 μL. Three replicates were prepared for each sample.   Once prepared, all samples were placed inside BioRad's MJ Mini™ Personal Thermal Cycler (Cat#: PTC-1148C). Cycling parameters were: 30 seconds of initial denaturation at 95°C, followed by 40 cycles of 5 seconds at 95°C, 5 seconds at 57°C, and plate read. The reaction ended with a melting curve analysis, in which the temperature slowly increased from 65°C to 95°C, with increments of 0.5°C for every 2 seconds followed by a plate read. Once finished, BioRad's CFX Manager™ Software (Cat#: 184-5003) automatically generated a standard curve by plotting the Ct values against the logarithms of the starting template quantity in each

corresponding sample (Appendix B). Each standard curve was used to generate reaction efficiency for the corresponding target by using the formula:

$$\text{Efficiency} = (10^{-1/\text{slope}} - 1) \times 100 \quad \text{(Livak et al, 2001)}$$

To measure gene expression for each individual test subject, cDNA (100 ng) was used as template; and each PCR reaction was carried out in triplicate.

2.3. Results

2.3.1. iLBP expressions in different tissues

When 100 ng of cDNA from the adult flight muscle were used as template for PCR, a clear band appeared below the 250 bp band when primers targeting the *fabp3* or *crabp* gene were used (Figure 2-2, A), which matched the expected sizes of the products. The same two bands were observed when the cDNA from larval fat body was used as template (Figure 2-2, B). With cDNA from SF9 cells as template, a clear band was observed with the CRABP primers (Figure 2-2, C). These bands were purified and sent for Sanger's sequencing.
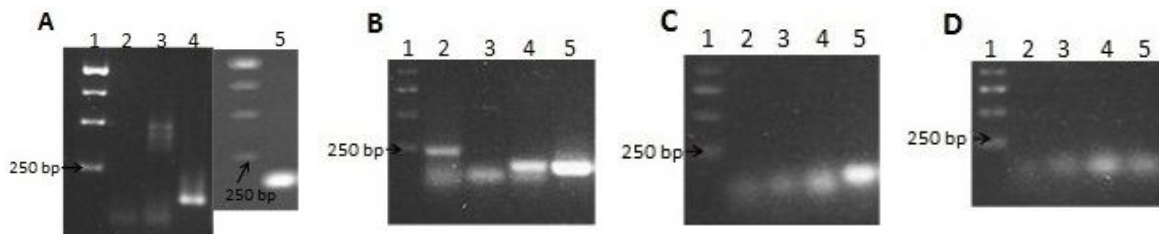
Figure 2-2   PCR detection of iLBP expression in *S. frugiperda*

PCR template: cDNA from adult flight muscle (A), larvae fat body (B), SF9 cells (C), or water (D). Lane 1: DNA ladder. Lane 2: FABP1. Lane 3: FABP2. Lane 4: FABP3. Lane 5: CRABP.

The sequences of the PCR products were aligned with the EST sequences for *fabp3* and *crabp* (Figure 2-3); the nearly perfect match confirms the specificity of the FABP3 and CRABP primers.

A

```
Sanger's_sequencing    ------------------------------------------------------------
FABP3_EST              ATGTCTTTCCTAGGCAAGGAATACAAATTTGAGAGGCAGGAGAACTTTGAGGAATTCGTG 60

Sanger's_sequencing    -----------------------------------------------CCNNNNNNNNNNNNNNN 16
FABP3_EST              AACTCTTTGGGTCTTCCCGCAGAGCAGACTCAAGGGTACCTCAACTACAACCCTGTACTG 120
                                                                      *
Sanger's_sequencing    NNGNACNCCAAGAATGGAGACACCTACACCCTGACTTCCGTCACTGCTCAGGGCACGAAG 76
FABP3_EST              AAGTACACCAAGAATGGAGACATGTACACCCTGACTTCCGTCACTGCTCAGGGCACGAAG 180
                        * ** **************   *****************************************

Sanger's_sequencing    GAAGTCTCCTTCAAGTCTGGGAGTCGCCTTCGA---------------------------- 108
FABP3_EST              GAGGTCTCCTTCAAGTCTGGGAGTCGCCTTCGATGAGACTGTTGCTGGCAAGAAGGTCAAC 240
                       ** ****************************

Sanger's_sequencing    ------------------------------------------------------------
FABP3_EST              ACCACTTACACCGTTGACGGTGACACCATCACACAGGTCCAGAAGTCTGATGATGGAGTC 300

Sanger's_sequencing    ------------------------------------------------------------
FABP3_EST              CTCACCATCACCAGGACTTTCTCTGGAAATGAGTTGGTTGTGACCCTGAAGACCAACAAA 360

Sanger's_sequencing    ---------------------------------------
FABP3_EST              TGGGACGGTGTTGCTACTAGATACTACAAAGTTGTTTAA 399
```

B

```
Sanger's_sequencing    ------------------------------------------------------------
CRABP_EST              ATGGAATTTGTCGGCAAGAAATACAAGATGGTCTCTTCGGAGAACTTCGACGAGTTCATG 60

Sanger's_sequencing    --------------------------------------------NCNNTCCGACGGTG 14
CRABP_EST              AAGGCTATCGGCGTTGGTCTCATCACCCGCAAGGCTGCTAACGCCGTCACTCCGACGGTG 120
                                                                   *  *********
Sanger's_sequencing    GAACTTCGCAAGGAGGGAGACGTTTACAACCTAGTGACGTCATCACCTTCAAGACCACN 74
CRABP_EST              GAACTTCGCAAGGAGGGAGACGTCTACAACCTAGTGACGTCATCCACCTTCAAGACCACT 180
                       *****************************  ********************************

Sanger's_sequencing    NAAGNCAAGTNNNNNNNNNNNNNNN------------------------------------ 98
CRABP_EST              GAGGTCAAGTTCAAGCCTGGTGAGGAGTTCGATGAGGACCGAGCTGATGGTGCCAAGGTG 240
                        * * *****
Sanger's_sequencing    ------------------------------------------------------------
CRABP_EST              AAGTCAGTGTGCACCTTCGAGGGTAACACCTTGAAGCAGATCCAGAAGGCTGCCGATGGC 300

Sanger's_sequencing    ------------------------------------------------------------
CRABP_EST              ATGGAAGTCAGCTACGTCAGAGAATTCGGACCCGAAGAGATGAAGGCTGTGATGACAGCT 360

Sanger's_sequencing    ---------------------------------------
CRABP_EST              AAGGATGTGACCTGCACCAGGGTGTACAAGGTCCAATAA 399
```

Figure 2-3   Alignment between PCR product and EST sequences

A: FABP3. B: CRABP.

2.3.2.   Optimization of real-time PCR

The real-time PCR standard curves created with primers targeting *crabp*, *fabp3*, *rpl37a*, and *gapdh* were used to calculate the respective efficiency values of the amplifications: 1.008, 0.837, 1.008, and 0.887 (Appendix B). Except for GAPDH, the R^2 values of CRABP, FABP3, and RPL37A were all greater than 0.99. Unlike the other three, the shape of the amplification curves of GAPDH did not reach the plateau phase; the shapes of the melting peaks suggest that, a secondary peak may be present in some

18

samples, indicating that the GAPDH primers are not specific. Thus, the GAPDH primers were discarded.

In order to determine the specificity of the remaining housekeeping gene, primers targeting *rpl37a* was used to perform a PCR reaction, using cDNA from flight muscle, larval fat body, or SF9 cells as templates. In the gel image, a clear band is apparent below the 250 bp band for all three samples (Figure 2-4). The sequence of the PCR product closely matches the EST sequence of *rpl37a* (Figure 2-5), indicating that these primers are specific for the *rpl37a* mRNA of *S. frugiperda*.
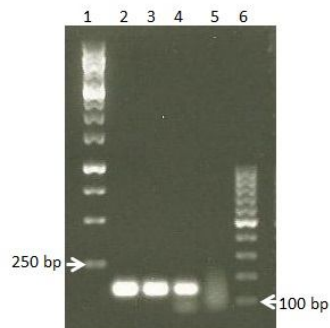
Figure 2-4   ***PCR detection of rpl37a expression in S. frugiperda***

PCR template: cDNA adult flight muscle (lane 2), larvae fat body (lane 3), SF9 cells (lane 4), or water (lane 5). Lane 1: 1.0 DNA ladder. Lane 6: 100 DNA ladder.

```
Sanger's_sequencing   ------------------------------------------------------------
L37a_EST              ATGGCCAAACGCACGAAGAAGGTCGGAATTACTGGCAAATATGGCACACGTTACGGTGCC 60

Sanger's_sequencing   --------------------------NNNNNNNNNNNNNGNACGCAAA-TACACTTGCTCG 32
L37a_EST              TCCCTCCGTAAAATGGTCAAAAAAATGGAAGTCACCCAGCACGCAAAATACACTTGCTCG 120
                                                * ******* ************

Sanger's_sequencing   TTCTGCGGTAAGGATGCCATGAAGCGTTCCTGTGTCGGCATCTGGTCATGCAAGCGCTGT 92
L37a_EST              TTCTGCGGTAAGGATGCCATGAAGCGTTCCTGTGTCGGCATCTGGTCATGCAAGCGCTGT 180
                      ************************************************************

Sanger's_sequencing   AAGAGGACCGN------------------------------------------------- 103
L37a_EST              AAGAGGACCGTAGCTGGTGGTGCGTGGGTATTCTCCACCACCGCTGCCTCCTCCTGCAGG 240
                      **********

Sanger's_sequencing   ---------------------------------
L37a_EST              TCCGCAGTCAGAAGATTGCGTGAGGTCAAGTAA 273
```

Figure 2-5   Alignment between PCR product and rpl37a EST sequences

2.3.3.   Real-time PCR data analysis

In order to determine if the expression of the housekeeping gene *rpl37a* was stable across different testing groups, the ANOVA test was used to compare the Ct values (Appendix C).   The results (F=1.778, P=0.20) indicated that no significant difference existed between test groups.

Since serial dilutions were not carried out for each cDNA sample, the corresponding standard curves could not be built, and the efficiency of each reaction could not be determined.   Instead, it was assumed that the reaction efficiency of each reaction was similar to the efficiency derived from the corresponding standard curve built from known amounts of template (Appendix B).

In order to analyze the expression of each target gene, we have developed a method to normalize its initial quantity against the reference gene:

The equation that describes the exponential amplification of PCR is known as:

$$X_n = X_0 \times (1 + E_x)^n$$

Where X_n is the number of amplicons at cycle n, X_0 is the initial number of template molecules, E_x is the efficiency of amplification, and n is the number of cycles. If we substitute n with Ct, the equation becomes:

$$X_{Ct} = X_0 \times (1 + E_x)^{Ct,x}$$

Where X_{Ct} represents the number of target molecules at the threshold cycle. A similar equation for the reference gene reaction will be:

$$R_{Ct} = R_0 \times (1 + E_R)^{Ct,R}$$

When rearranged, these two equations become:

$$X_0 = X_{Ct,x} \times (1 + E_x)^{-Ct,x} \quad \text{and} \quad R_0 = R_{Ct,R} \times (1 + E_R)^{-Ct,R}$$

To normalize X_0 against R_0, the expression becomes:

$$\frac{X_0}{R_0} = \frac{X_{Ct,x} \times (1 + E_x)^{-Ct,x}}{R_{Ct,R} \times (1 + E_R)^{-Ct,R}}$$

Since the PCR products of both the target and reference gene have similar sizes, one can assume that the amounts of amplicon of the two genes are similar at the threshold cycle. Thus, the expression becomes:

$$\frac{X_0}{R_0} = \frac{(1 + E_x)^{-Ct,x}}{(1 + E_R)^{-Ct,R}}$$

By using this equation, we were able to compute the initial quantity of CRABP and FABP3 for each cDNA sample using the Ct values (Appendix C), normalized against the reference gene *rpl37a* (Table 2-1). When the normalized expressions for both genes were compared between different test groups using ANOVA test, none of the differences appeared to be significant (CRABP: F=1.988, P=0.17. FABP3: F=3.157, P=0.064).

Table 2-1   *crabp and fabp3 expressions normalized against rpl37a*

Group	Subject ID	*crabp* expression	mean	variance	*fabp3* expression	mean	variance
#1:	01	0.17	0.23	0.03	0.18	0.29	0.02
Control	02	0.14	0.23	0.03	0.13	0.29	0.02
	03	0.08	0.23	0.03	0.49	0.29	0.02
	04	0.53	0.23	0.03	0.36	0.29	0.02
#2:	05	0.64	0.38	0.03	0.28	0.53	0.05
4 hours flight	06	0.36	0.38	0.03	0.58	0.53	0.05

Group	Subject ID	*crabp* expression	mean	variance	*fabp3* expression	mean	variance
0 hour rest	07	0.35	0.38	0.03	0.38	0.53	0.05
	08	0.17	0.38	0.03	0.87	0.53	0.05
#3: 4 hours flight 2 hours rest	09	1.72	0.83	0.29	0.38	0.16	0.02
	10	0.60	0.83	0.29	0.06	0.16	0.02
	11	0.28	0.83	0.29	0.05	0.16	0.02
	12	0.71	0.83	0.29	0.17	0.16	0.02
#4: 4 hours flight 4 hours rest	13	0.42	0.72	0.12	0.22	0.25	0.00
	14	0.66	0.72	0.12	0.25	0.25	0.00
	15	1.31	0.72	0.12	0.33	0.25	0.00
	16	0.48	0.72	0.12	0.21	0.25	0.00

2.4. Discussion

2.4.1. iLBP gene expressions

Previously, it had been demonstrated that the RNA extracted using QIAGEN's RNeasy® Mini Kit contained negligible amount of genomic DNA carryover, which alone would not result in visible PCR bands for the 35 cycles (data not shown). Thus, the PCR bands observed in figure 2-2 and figure 2-4 must originate from cDNA templates, rather than genomic DNA carryover. Based on this assumption, we can conclude that *crabp* is expressed in the flight muscle, fat body, and SF9 cells, whereas *fabp3* is expressed in both flight muscle and fat body (Figure 2-2). Expression of *fabp1* and *fabp2* appeared to be absent in all three tissue types since no clear bands were observed (Figure 2-2).

2.4.2. Reference genes analysis

The housekeeping genes *gapdh* and *rpl37a* were selected as reference genes for the real-time PCR experiments. GAPDH is known to catalyze the sixth step of glycolysis and thus serves to break down glucose for energy extraction. Other than its involvement in metabolism, it is also involved in transcription activation of certain genes such as the histone H2B coding gene (Zheng et al., 2003). It is one of the most

commonly used housekeeping genes selected for comparisons of gene expression data (Barber et al., 2005). By using the tblastn tool of the SPODOBASE EST database, we have found an EST sequence that codes a polypeptide sequence 89% identical to the GAPDH from the silkworm *Bombyx mori*. However, when the primers GAPDH-F and GAPDH-R were used to generate a real-time PCR standard curve using known quantity of diluted templates, poor linearity with an R^2 value of 0.910 was observed when Ct values were plotted against the logarithms of the starting template quantity in each standard solution. Moreover, none of the amplification curves reached the plateau phase of PCR, whereas the melting curves provided evidence that non-specific PCR products were generated during the amplification process (Appendix B). These results may be due to the fact that both primers contain several repeats of guanines or cytosine, which may allowed them to easily form primer dimers during amplification, thus reducing the efficiency of the reaction. Therefore, these primers were not used for real-time PCR experiments, and *gapdh* was not used as a housekeeping gene.

The *rpl37a* gene encodes the ribosomal protein L37a, which is a structural component of the 60S subunit of the ribosome. It is one of the mostly stably expressed genes in various mammalian tissues and cell lines such as the brain, macrophages, and granulocytes derived cell lines (Cao et al., 2012; Maess et al., 2010; Pfister et al., 2011). Moreover, it has been previously reported that in the invertebrate *Dictyocaulus viviparus* (also known as the lungworm of cattle), the expression of this gene had the least variation during different developmental stages when compared with four other common housekeeping genes (Strube et al., 2008). While we did not compare the stability of the *rpl37a* gene with other housekeeping genes, the ANOVA test result suggested that the expression of this gene is stable across different testing groups.

2.4.3.   CRABP and FABP3 expression analysis

The expression levels of *crabp* and *fabp3* in each tested subject were determined by normalizing the initial quantity of each target gene template against *rpl37a*, under the assumption that the reaction efficiency of each gene remained the same across all samples. ANOVA test results revealed that the expression of each gene from each test group was not significantly different from any other group. While this could indicate that flight does not have an effect on iLBP expression in *S. frugiperda,* it is also possible that

23

we could not detect statistically significant differences due to the small sample size. Moreover, potentially different reaction efficiencies between test subjects within the same group may also contribute to the *crabp* expression variances in group #3 and #4 (Table 2-1). Therefore, at this time, we do not have unambiguous statistical evidence to either reject or accept the hypothesis that flight and post-flight resting have an effect on the expression of *crabp* or *fabp3*.

On the other hand, a trend of increase is observed when the average expression of *crabp* from the control group is subsequently compared with group #2, #3, and #4 (Table 2-1). This trend suggests that flight and post-flight rest can have an impact on the expression of *crabp* in flight muscles. In contrast, a similar trend could not be observed for *fabp3*. While the present study does not reveal conclusive evidence, our results lend support to the hypothesis that the expression of *crabp* is affected by flight, while *fabp3* expression is largely independent.

3. Sequencing of the CRABP promoter

3.1. Introduction

In the previous chapter, we described the identification of four different iLBP genes in *S. frugiperda*. Among the four, *crabp* was found to be the only gene expressed in adult flight muscles, larvae fat body, and the SF9 cell line; while *fabp3* expression was found to be present in all tissues except in SF9 cells. Some evidence was provided to suggest that in adult flight muscles, expression of *crabp*, but not *fabp3*, may be affected by flight.

As a prerequisite for understanding the molecular mechanism involved in this gene regulation, we decided to characterize the promoter sequence of the *crabp* gene in *S. frugiperda*, and search for potential regulatory elements that may interact with nuclear receptors. By applying the rapid amplification of cDNA ends (RACE) technique, in combination with another method that uses restriction enzyme digested genomic DNA libraries as PCR templates, the 5' untranslated region (5'UTR) and promoter sequence can be obtained, for subsequent bioinformatics analysis with programs such as the JASPAR (Casamar et al., 2010).

3.1.1. 5'RACE library construction

RACE is a technique that uses polymerase chain reaction (PCR) and reverse transcriptase (RT) to rapidly obtain cDNA containing complete 5' or 3'UTR. Using cDNA as template, along with gene specific primers, the 5' or 3'UTR sequence of target genes, as well as the transcription start and end sites of the gene, can be obtained by PCR. Previously, the transcription start site (TSS) of the *mscrabp* gene was determined using combination of RACE and PCR with gene specific primers (Mansfield et al., 1998). In this study, a similar approach was used in determining the 5'UTR sequence and TSS of the *crabp* gene in *S. frugiperda*.

25

The SMARTer™ RACE cDNA Amplification Kit from Clontech employs a technology which enables the generation of full length double strand cDNA from mRNA (Zhu et al., 2001). The primer used for the first strand synthesis is the 5'-CDS primer A, which is an oligo (dT) primer with the sequence 5'-(T)$_{25}$VN-3', where N = A, C, G, or T, and V = A, G, or C. When the SMARTScribe™ Reverse Transcriptase reaches the end of an RNA template, its terminal transferase activity adds 3-5 undisclosed residues to the 3' end of the first-strand cDNA. This added sequence can then anneal to the 3' end of a SMARTer II A Oligonucleotide primer, which enables the RT to switch its template from mRNA to the complementary cDNA, and start the synthesis of the second strand (Figure 3-1). All of the resulting double-stranded cDNA have the SMARTer II A sequence on one side, which is also part of the Universal Primer A Mix (UPM) and Nested Universal Primer A (NUP), the primers used for subsequent PCR reactions.
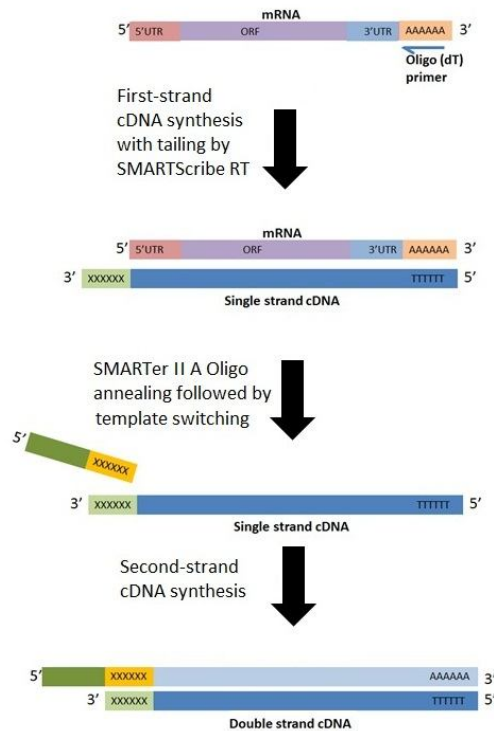
Figure 3-1   ***Schematic representation of the SMARTer™ RACE cDNA Amplification***

Picture from Zhu et al., 2001, used with permission.

26

Containing the sequence of the SMARTer II A Oligonucleotide, the UPM and NUP primers serve as the outer and inner forward primers for nested PCR, where the reverse primers contained target gene specific sequences. The resulting PCR products can be sequenced from the reverse end using a gene specific primer, in order to determine the TSS located after the SMARTer II A sequence (Figure 3-2, A). By aligning the sequence of the PCR product with the coding sequence of the gene, the entire 5'UTR sequence can be determined in this manner (Figure 3-2, B).
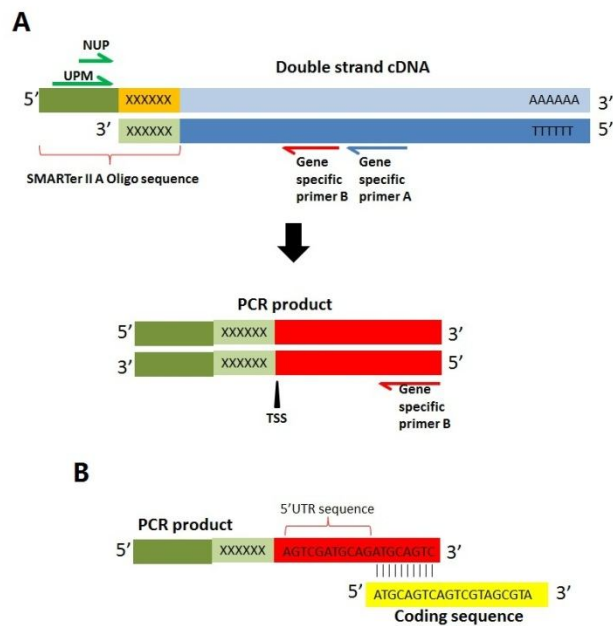
Figure 3-2   Nested PCR schematics

A: Sequencing of nested PCR product determines TSS. B: Aligning PCR products with coding sequence determines 5'UTR sequence.

Construction of a restriction DNA library is useful for determining unknown genomic DNA sequences that are in close proximity to known sequences. In this study, the strategy was to carry out a restriction digestion of intact genomic DNA, followed by ligating the digested fragments with a linearized pZsGreen1-DR vector using T4 DNA ligase. The resulting DNA library is useful as template for PCR, using a gene specific primer that annealed to the 5'UTR sequence of *crabp*, and another primer that annealed to the multiple cloning region of the vector. Ideally, any resulting PCR product would

consist of a piece of unknown sequence upstream of the 5'UTR, flanked by a portion of the vector sequence and the 5'UTR sequence. By sequencing the PCR product, the unknown sequence upstream of the 5'UTR can be determined (Figure 3-3).
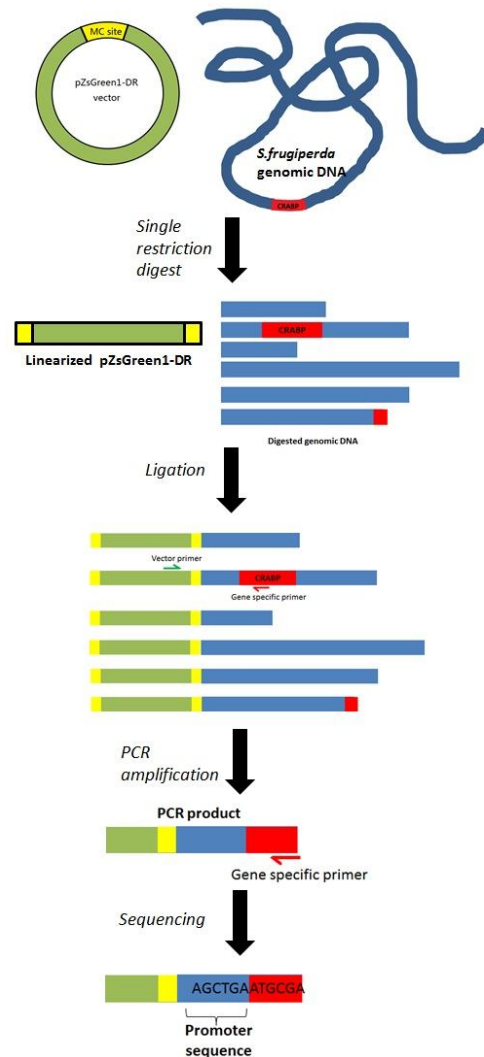
Figure 3-3   Sequencing strategy using restriction DNA library

3.1.2.   Promoter sequence analysis

The JASPAR database is a leading open-access database of matrix profiles describing the DNA binding patterns of various transcription factors and other proteins

(Casamar et al., 2010). With its aid, potential binding sites of nuclear receptors within the conserved regions of the promoter sequence can be predicted based on existing matrix profiles generated from experimental evidence.   These potential hormone response elements may be employed in reporter gene studies or gel shift assays, which may be useful in analyzing the regulation of the gene.

3.2. Materials and methods

3.2.1. mRNA extraction

The QuickPick™ SML mRNA kit from Bio-Nobile (Cat#: 51012) was used to extract total mRNA from 1.0E6 SF9 cells, with some modification of the manufactory's protocol. After the cell media was removed from the cell pellet, cells were washed with 600 µL PBS by a quick spin at 300 x g for 5 minutes. After the wash, cells were re-suspended using 400 µL lysis/binding buffer. Homogenization of the cells was achieved using a mechanical agitator, followed by 10 needle strokes using the 21 Gauge needle. Next, 4 µL of β-mercaptoethanol was added into the mixture, followed by addition of 60 µL of Oligo $(dT)_{30}$ coated magnetic particles. The mixture was then continuously mixed on a tube rotator for 5 minutes at room temperature. Next, the magnetic particles from the mixture were removed using a PickPen® tip (Bio-Nobile), and washed twice for 15 seconds with 400 µL buffer A, and once with 300 µL buffer B. After the third wash, the particles containing mRNA were eluted into 30 µL DEPC-treated water. The particles were heated at 70 °C for 5 minutes, and then removed from the water with the aid of the PickPen® tip. The concentration of the mRNA that remained dissolved in the water was determined with the NanoDrop spectrophotometer.

3.2.2. 5'RACE library

The procedure of double stranded cDNA library synthesis was performed as outlined in the protocol of the SMARTer™ RACE cDNA Amplification Kit from ClonTech (Cat#: 634923): 2.75 µL of the isolated mRNA was mixed with 1 µL of 5'-CDS Primer A, then heated at 72°C for 3 minutes, followed by 2 minutes of incubation at 42°C. Next, 1 µL of SMARTer II A Oligonucleotide primer was then added to the mixture, followed by

addition of 4 µL master mix A (2 µL 5X First-strand buffer, 1 µL 20 µM DTT, and 1 µL 10 mM dNTP mix), 0.25 µL RNase inhibitor (40 U/µL), and 1 µL SMARTScribe™ Reverse Transcriptase. The mixture was then incubated at 42˚C for 90 minutes, followed by 10 minutes of incubation at 70˚C. At the end, 100 µL of Tricine-EDTA buffer was added to the mixture. The mixture was then placed at -20˚C for short term storage.

To obtain the 5'UTR sequence of *crabp*, a hemi-nested PCR was performed. For the first run, 5 µL of 5'RACE library cDNA was mixed with 25 µL of 2X PCR master mix from Fermentas, 5 µL of 10X UPM primer, 10 µmoles of CRABP-A primer (TTGGCACCATCAGCTCGGTCC), and variable amount of water for a total reaction volume of 50 µL. Thermal cycling parameters were: initial denaturation at 94˚C for 3 minutes, followed by 40 cycles of 94˚C for 30 seconds, 55˚C for 30 seconds, and 72˚C for 3 minutes, along with a final extension at 72˚C for 7 minutes. First round PCR products were analyzed on a 1.0% agarose gel, and then used as template for a second round of PCR: 5 µL of template was mixed with 25 µL of PCR master mix, 1 µL of NUP primer, 10 µmoles of CRABP-B primer (CATCGAACTCCTCACCAGGC), and water up to a total volume of 50 µL. Thermal cycling parameters were: initial denaturation at 94˚C for 3 minutes, followed by 40 cycles of 94˚C for 30 seconds, 54˚C for 30 seconds, and 72˚C for 1 minute, along with a final extension at 72˚C for 7 minutes. PCR products were analyzed on a 1.0% agarose gel, and then purified with the QIAquick Gel Extraction Kit from QIAGEN (CAT#: 28704) and sent for Sanger's sequencing.

3.2.3.   Restriction library

Genomic DNA from *S. frugiperda* was extracted from SF9 cells using the GenElute™ Mammalian Genomic DNA Miniprep Kit from Sigma-Aldrich (Cat#: G1N10), following the manufacturers' protocol. Isolated DNA was analyzed on a 1.0% agarose gel for quality check.   DNA concentration and purity were determined based on NanoDrop readings.

To digest the genomic DNA, 100 ng was mixed with 5 µL of 10X Standard FastDigest Buffer, 5 µL of FastDigest restriction enzyme from Fermentas (SmaI, BglII, or BamHI), and water to a total reaction volume of 50 µL. To digest the pZsGreen1-DR vector, 700 ng of the vector was mixed with 2 µL of 10X buffer, and 1 µL of the same

FastDigest restriction enzyme. Incubation and enzyme inactivation procedures were performed based on the protocol of the manufacturer. Digested vectors were analyzed on a 1.0% agarose gel to confirm digestion success, and then purified with QIAGEN's QIAquick Gel Extraction Kit.

For ligation, 2 μL of the digested genomic DNA was mixed with 250 ng of vector digested with the corresponding enzyme, along with 2 μL of 10X T4 DNA ligase buffer, 1 μL of T4 DNA ligase, and water up to a total volume of 20 μL. The reaction mixture was incubated at 25°C for 2 hours, followed by 15 minutes at 65°C for enzyme inactivation. Products were kept at -20°C for long term storage.

3.2.4.   Promoter sequencing

In order to obtain the sequence upstream of the 5'UTR of *crabp*, a nested PCR strategy was applied. For the first run, 1 μL of the restriction library (SmaI, BglII, or BamHI) was used as template, which was mixed with 25 μmoles of pZs-1 (CCTGCGTTA TCCCCTGATTCTGTGGATAAC) and CRABP-1(GTATTTCTTGCCGACAAATTCCAT) primers, and 1 μL of 10 mM dNTP mix. In addition, 10 μL of 5X iProof HF Buffer and 1 μL of iProof DNA Polymerase from BioRad's iProof™ High-Fidelity DNA Polymerase kit (Cat#: 172-5300) were added into the mixture, along with nuclease free water to bring the total reaction volume to 50 μL.   Thermal cycling parameters were: initial denaturation at 98°C for 5 minutes, followed by 40 cycles of 98°C for 30 seconds, 59°C for 30 seconds, and 72°C for 10 minutes, along with a final extension at 72°C for 10 minutes. The resulting PCR products were analyzed on a 1.0% agarose gel; and then used as templates for the second run. The pipetting and cycling parameters for the second run was same as the first, with the following exceptions: the primers used were pZs-2 (AGCGCTACCGGACTCAGATCT) and CRABP-2 (TGTGTGAGTCCGTACTACC G); and the annealing temperature was 59°C. When analyzed on a gel, major bands were purified and sent for Sanger's sequencing using pZs-2 and CRABP-2 primers.

3.2.5.   Promoter amplification

To amplify the genomic sequence from upstream of the 5'UTR to the first exon of the gene, ~100 ng of intact genomic DNA was used as PCR template, which was mixed

with 25 µmoles of CRABP-3 (AACATTCCGTTGTATCGTCC) (designed based on Sanger's sequencing results) and CRABP-4 (CCGATAGCCTTCATGAACTCGTCG) primers, along with 1 µL of 10 mM dNTP mix, 10 µL of 5X iProof HF Buffer and 1 µL of iProof DNA Polymerase. Thermal cycling parameters were the same as in the nested PCR mentioned above, except that the annealing temperature was changed to 55°C, and the extension time was changed to 2 minutes. After gel analysis, the major PCR product was purified from the gel and sent for Sanger's sequencing.

3.3. Results

3.3.1. 5' RACE nested PCR

The mRNA extracted from SF9 cells had a concentration of 13.5 ng/µL. Since 2.75 µL of this mRNA solution were used for 5'RACE cDNA library synthesis, it was assumed that 37.13 ng of mRNA were converted into cDNA. When the gene-specific primer CRABP-A and the Universal Primer Mix (UPM) were used to perform PCR on the cDNA template, a clear band appeared above the 250bp ladder band, with a faint band below (Figure 3-4, A). Due to the presence of the minor band, it was difficult to obtain the major amplification product in sufficient purity for sequencing; therefore a hemi-nested PCR was performed using the gene specific primer CRABP-B and the Nested Universal Primer (NUP), whose sequence is part of UPM's. After the nested PCR, a single clear band was observed above the 250 bp ladder band (Figure 3-4, B). This band was purified and sent for Sanger's sequencing, using the CRABP-B primer.

3.3.2. Identification of 5'-UTR sequence

When the sequence of the nested PCR product was aligned with the EST sequence of CRABP (Spodobase, cluster SF9L03784), a 100% sequence match was observed at the beginning of the open reading frame (ORF) and the 26 bases upstream of the start codon. Four mismatches were present within the sequence further upstream (Figure 3-5). As expected, the 5' end of the sequenced PCR product is identical to the sequence of the UPM and NUP primers, which annealed to the SMARTer II A Oligo sequence.
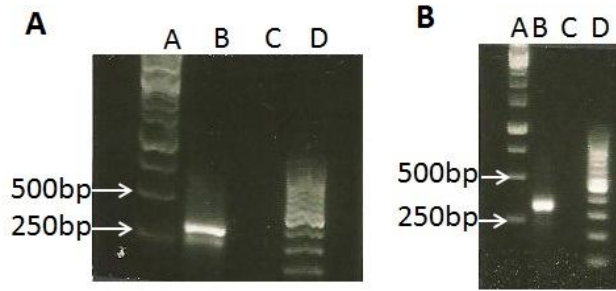
Figure 3-4   5'-RACE gene specific hemi-nested PCR

A: outer PCR. B: hemi-nested PCR. Lane A: 1.0 DNA ladder. Lane B: PCR product. Lane C: negative control using water template. Lane D: 100 DNA ladder.
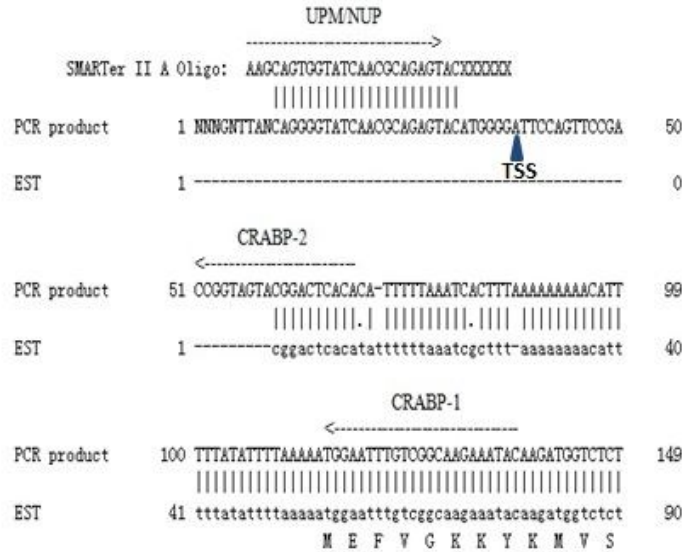
Figure 3-5   Sequence alignment for EST and nested PCR product

The alignment shown here contains the 5'end of the EST sequence and the nested PCR product. The coding sequence is indicated by the amino acid labels. Arrows indicate the positions of primers.

3.3.3. Restriction digestion

Genomic DNA extracted from SF9 cells had a concentration of 36.9 ng/µL, and an A_{260}/A_{280} absorbance ratio of 2.10. When ran on a 1.0% agarose gel, a smear

33

appearance was observed in the upper region of the gel, indicating little degradation of the DNA (Figure 3-6).

Figure 3-6   Quality of genomic DNA extracted from SF9 cells
Lane A: 1.0 DNA ladder. Lane B: Genomic DNA.

The genomic DNA obtained from SF9 cells and the pZsGreen1-DR vector was digested by the restriction enzymes SmaI, BglII, and BamHI. The digested vector samples were analyzed by agarose gel electrophoresis to confirm complete digestion. From the gel image, band shifting was observed for all three digestions when compared with undigested vector, indicated successful digestion (Figure 3-7). After purified from the gel, DNA concentration for SmaI, BglII, and BamHI digested vectors were 14.6 ng/µL, 10.0 ng/µL, and 17.2 ng/µL, respectively. For each restriction digest, digested vector was ligated with genomic DNA digested with the corresponding enzyme to produce three restriction libraries.
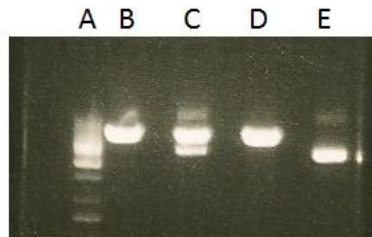
Figure 3-7   Restriction digest of pZsGreen1-DR vector
Lane A: 1.0 DNA ladder. Lane B: SmaI digestion. Lane C: BglII digestion. Lane D: BamHI digestion. Lane E: Undigested vector.

3.3.4. *Sequencing of the crabp promoter*

When PCR was performed with the vector specific primer pZs-1, the CRABP-1 primer, and the three restriction libraries as template, no band larger than 250 bp was visible (Figure 3-8, A). However, when the PCR products were re-amplified with the nested primers pZs-2 and CRABP-2, a strong band with the size of ~1.5 kb was observed for the BamHI library (Figure 3-8, B). This band was purified and sent for Sanger's sequencing, using the nested primers pZs-2 and CRABP-2.
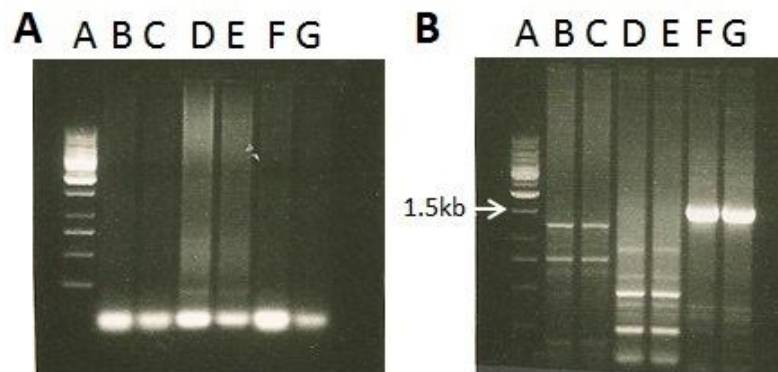
Figure 3-8   Restriction library nested PCR

A: outer PCR. B: nested PCR. Lane A: 1.0 DNA ladder. Lane B, C: Smal library. Lane D, E: BglII library. Lane F, G: BamHI library.

The sequence obtained from Sanger's sequencing contained a few unresolved nucleotides in the middle region (Figure 3-9, A). Since this sequence ended upstream of the 5'-UTR region, a PCR reaction was performed using primers CRABP-3 and CRABP-4, which annealed to the 5' end of the obtained sequence and the first exon of the gene, respectively (Figure 3-9, B). Genomic DNA was used as template.

The resultant PCR product yielded a strong band with a size of ~1.5 kb (Figure 3-10). When the product was sequenced from both ends, one of its ends was virtually identical with the 5'-UTR sequence, indicating that the 1.5 kb sequence obtained by restriction library-nested PCR was indeed the promoter sequence of the *crabp* gene. The transcription start site appeared to locate at -76 bp upstream of the start codon (Figure 3-11).

A

CRABP-3
```
------------------------------------------>
CCGCGGGCCCGGGATCCCAGAGTCTCGTAATTATATTTAATGTTTTGTACGTTGNTTAAACATTCCGTTGTATCGTCCAATAAACTCTTTATTTGATTTGCAAAT
TGAGTATTTATTTTTCTTTGCCAGAATAATTTCTTTAACACTGAATTGATAATCAATCTTTTATTACCGTACCTGGGTTTGAGAACACTCCCTGCTTGTTCATAAT
TACTTTGATTTATTTGAAGAAGTTTCAATAACACCTCGGCCTCTCCAGTGACACTACTCTTCAAATAATGTAACTTTTGAACATCACTACTCGCCGTGTTTTCAT
GAACTAATGAAATAAATAAGTCTTGAAAGGTACGGCCAATCTTCATAATTCCCACTAAATCTTGNNANTGTATCCTCGATGANCTAACGAGTTGATCATTANTC
ATATTTGAAAATGAATTATGTTTCGATGATGTGTGGGTACGGACGATGCGCTTATCGTCNACAAAAGGCCTTGTAAATCNGCTAATAACCCAANANATNNNTCT
TCATAATTATANAAATCCTCGTTNACNAANNATTTAACGTTTGNTNNTGTTCTTTCNNTGNTATCTTGACTANTTGTGGATGANATTCTCTGAATGTTGTCCAA
TATTCGTCTATNCATTTTATTCTTGCCTCGACGTANNNTTTGTCANCCTGGTTTTAGGACATTTCTTAATGTTTACTTGCGTTTTTTGTAATAAAGCAGCTGTGT
NTTCTAAAACGTTTATTAGCTCCTTAGTAGTCATTGCACCTTTTCTTGCTTCTATTCATTCATTAATGACGGACNTTCGGAATGCGATGTCGCTATTGTTCTCGTT
GGCCGGCCGATATCTTCGATAACTTGTTTTTTGCAATTAGCTCGGTTCGTAAATCCGGCTCGAAGGACCACTATGTTTTTGATATTAAGCTTGATGGTAGTTTC
GTAGGAATTGATTCAGTTAAAATGCGTGTTTGCTTTAGGAGTTTTTATTTACGACAAACGGGTGACAACTAGTGACAAATACATTTTGTTTATTTATTCTTTTTT
TACCATAAACAATATAGATATAACAGGGAATATTATTTTGGTTGGGAAAGAAGGGGATAGGAAATAACTTGTTAAATAATTTTTGCAATGGTCCTGCTGTTAC
TAGGACGAATGAATAATGGATAGAAAGTCATGGGGTTACAAATAGGTAGCTATTACTCGCTCATAATATGTTTTATGTTGTCTGTAACGTAATTTTTCTTCAAA
GGATTACAAAAATGCTTGTTTGGCTTCGTCAATCTTCCCAGATTAAGCCACAAGACATGTAATCATCTCCTTTATATGTGTATAATCTAATTTGCAAAGCAAAC
AAAATCTTAGCTCGCGAAATCTCGCTACACGCTCACACCATCCTCCGTCTCACAATTTCCCCCGCTACAGAGTATATAATCAGCGCGCGGTGTCAAGGTCAATT
CAGTTCAATAACTTCATTCGTTCCGTTCATCTCAGTCATTCGTTCGTTGACGCATGCGCAGAGCCGAGCAGTTGTACCNNNNNCGCGNNNNNNNNNNNNNNNNNN
T
```

B

Figure 3-9   Restriction library nested PCR product sequence

A: Sequence obtained from Sanger's sequencing. Arrow indicates the location of the CRABP-3 primer. B: Positions of PCR primers. Color coded areas indicate regions with known sequence; black lines indicate unknown sequence. Primers are represented by arrows.

Figure 3-10   Amplification of crabp promoter

Lane A: PCR product obtained using genomic DNA template. Lane B: 1.0 DNA ladder.

A

```
PCR      1  NNNNNNNNNNNNNNNNNNNNNATTGAGTATTTATTTTTCTTTGCGAAAATAA   50
PCR     51  TTTCTTTAAAACTGAATTGATAAATCAATCTTTTATTACCGTACCTGGTTT  100
PCR    101  TGAGAACACTCCATGCTTGTTCATAATTACTTTGAGTTATTTGAATAGGT   150
PCR    151  TTCAATAACACCTCGGCCTCTCCAGTGACACTACTCTTCAAATAGTGTAA   200
PCR    201  CTTTTGAACATCACTAATCGCCGTGTTTTCATGAACTAATGAAATAAATA   250
PCR    251  AGTCTTGAAAGGTAGGCCAATCTTCATAATTCCCACTAAATCTTGGTAGT   300
PCR    301  TGTATCCTCGGCAGCCTAACGAGTTGATCATTAGTCGTATTTGTAAATGA   350
PCR    351  ATTATTTTTCGATGATGTGGGTAGGGACGATGCGTTTATCGTCGACAAAA   400
PCR    401  GGTCTTGTAAGTCGGCTAATAACCCAAGATATAAATCTTCATAAGTATAG   450
PCR    451  AAATCCTCGTTAACGAAGTAATTTAACGTTTGTTTTTGTTCTTTCGTTGT   500
PCR    501  TATCTTGACTAATTGTTGATGAGATTCTCTGAATGTTGTCCAATATTCGT   550
PCR    551  CTATACATTTTATTCTTGCCTCGACGTAACCTTTTGTCAGCCTGGTTTTA   600
PCR    601  GGACATTTCTTAATGTTTACTTGCGTTTTTTGTAATAAAGCAGCTGTGTC   650
PCR    651  TTCTAAAACGTTTATTAGCTCCTTAGTAGTCATTGCACCTTTTCTTGCTT   700
PCR    701  CTATTCATTCATTAATGACGGACCTTCGGAATGCGATGTCGCTATTGTTC   750
PCR    751  TCGTTGGCCGGCCGGATATCTTCGATAACTTGTTTTTTGCAATTAGCTCGG  800
PCR    801  TTCGTAAATCCGGCTCGAAGGACCCACTATGTTTTTGATATTAAGCTTGAT  850
PCR    851  GGTAGTTTCGTAGGAATTGATTCAGTTAAAATGCGTGTTTGCTTTAGGAG  900
PCR    901  TTTTTATTTACGACAAACGGGTGACAACTAGTGACAAATACATTTTGTTT  950
PCR    951  ATTTATTCTTTTTTTACCATAAACAATATAGATATAACAGGGAATATTAT 1000
PCR   1001  TTTGGTTGGGAAAGAAGGGGATAGGAAATAACTTGTTAAATAATTTTTGC 1050
PCR   1051  AATGGTCCTGCTGTTACTAGGACGAATGAATAATGGATAGAAAGTCATGG 1100
PCR   1101  GGTTACAAATAGGTAGCTATTACTCGCTCATAATATGTTTTATGTTGTCT 1150
PCR   1151  GTAACGTAATTTTCTTCAAAGGATTACAAAAATGCTTGTTTGGCTTCGTC 1200
PCR   1201  AATCTTCCCAGATTAAGCCACAAGACATGTAATCATCTCCTTTATATGTG 1250
PCR   1251  TATAATCTAATTTGCAAAGCAAACAAAATCTTAGCTCGCGAAATCTCGCT 1300
PCR   1301  ACACGCTCACACCATCCTCCGTCTCACAATTTCCCCCGCTACAGAGTATA 1350
PCR   1351  TAATCAGCGCGCGGTGTCAAGGTCAATTCAGTTCAATAACTTCATTCGTT 1400
```

B

```
PCR    1401  CCGTTCATCTCAGTCATTCGTTCGTTGACGCATGCGCAGAGCCGAGCAGT  1450
                                                 ............. |
5' UTR    1  ------------------------------------NNNGNTTANCAGGG    14

                                          *
PCR    1451  TGTACCGTTTGCCGCGTTTATTCATTCCAGTTCCGACCGGTAGTACGGAC  1500
             ....|.........|.....|.... ||||||||||||||||||||||||
5' UTR   15  GTATCAACGCAGAGTACATGGGGATTCCAGTTCCGACCGGTAGTACGGAC    64
                └────────── SMARTer II A Oligo sequence ──────────┘

PCR    1501  TCACACATTTTAAATCACTTTAAAAAAAAACATTTTTATATTTTAAAAA   1550
             ||||||||||||||||||||||||||||||||||||||||||||||||||
5' UTR   65  TCACACATTTTAAATCACTTTAAAAAAAAACATTTTTATATTTTAAAAA    114

PCR    1551  TGGAATTTGTCGGCAAGAANTACAAGANGNNNNNNNNNNNNNNNNN       1594
             |||||||||||||||||||| |||||||| |
5' UTR  115  TGGAATTTGTCGGCAAGAAATACAAGATGGTCTCT----------       149
                                       M  E  F  V  G  K  K  Y  K  M  V  S
```

Figure 3-11   Promoter sequence of crabp

A: Sequence upstream of the 5'UTR. B: Alignment between the PCR product sequence and the 5'UTR sequence. Asterisk: transcription start site.

3.3.5.   Analysis of the crabp promoter sequence

When the 1.5 kb promoter sequence from *S. frugiperda* was aligned with the promoter from the closely related species *Helicoverpa zea* (obtained from unpublished previous work), high level of conservation was observed from the start codon to -289 bp

upstream of the transcription start site (Figure 3-12). By analyzing this piece of conserved region with the online program ConSite, it was found that an area between -112 and -79 contained four potential binding sites for nuclear receptors, with three on the positive strand and one on the negative. These sites were for chicken ovalbumin upstream promoter transcription factor (COUP-TF) on the negative strand, RAR-related orphan receptor alpha 1 (RORα1), retinoic X receptor dimerized with vitamin D receptor (RXR-VDR) from *Homo sapiens*, and ultraspiracle (USP) from *Drosophila melanogaster* on the positive strand (Figure 3-13). This characterized area appears to be highly conserved between these two species, with only three nucleotide mismatches. In addition, three separate consensus sequences for nuclear receptor binding half-sites were identified in a direct repeat conformation: AGGTCAattcAGTTCAataACTTCA (Figure 3-13). Similar cluster of consensus binding sites was not found within 1500 bp upstream of the coding area from two other lepidoptera species, *Manduca sexta* and *Bombyx mori*.

Figure 3-12 Local alignment between H. zea and S. frugiperda promoters
Box: conserved area from -289 to coding region.

Figure 3-13   Promoter analysis by ConSite and McPromoter 006

Genomic DNA sequence from -289 upstream of transcription start site to coding sequence. Red text region: three consensus nuclear receptor binding half-sites.

When using ConSite to search for any potential TATA box within the promoter sequence, the predicted site appeared at 115 bp upstream of the experimentally determined TSS. On the other hand, when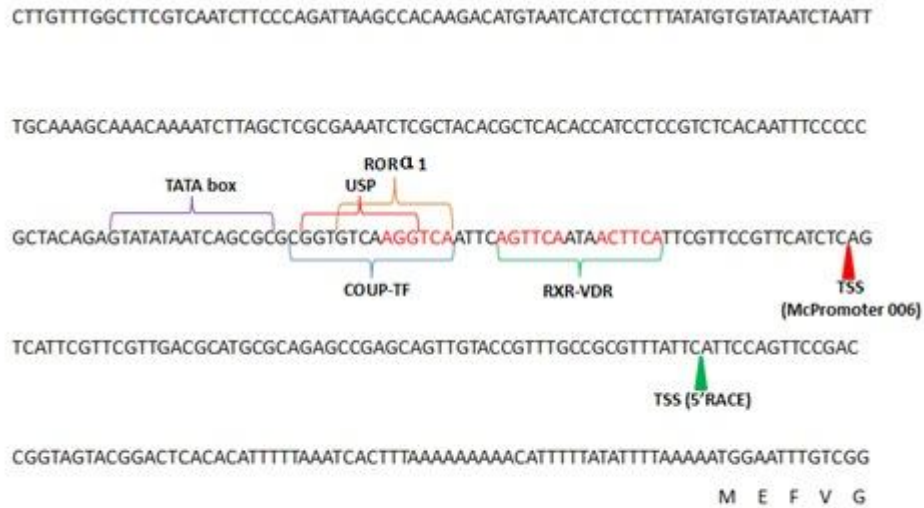 the promoter prediction server McPromoter 006 (http://www.cbs.dtu.dk/services/Promoter) was used to predict the location of the TSS, it appeared to locate at 62 bp upstream of the experimentally determined location (Figure 3-13). Nevertheless, the experimentally determined site region appears to be highly conserved in six different lepidoptera species; whereas less sequence identity was observed in the 62 bp upstream area (Figure 3-14).
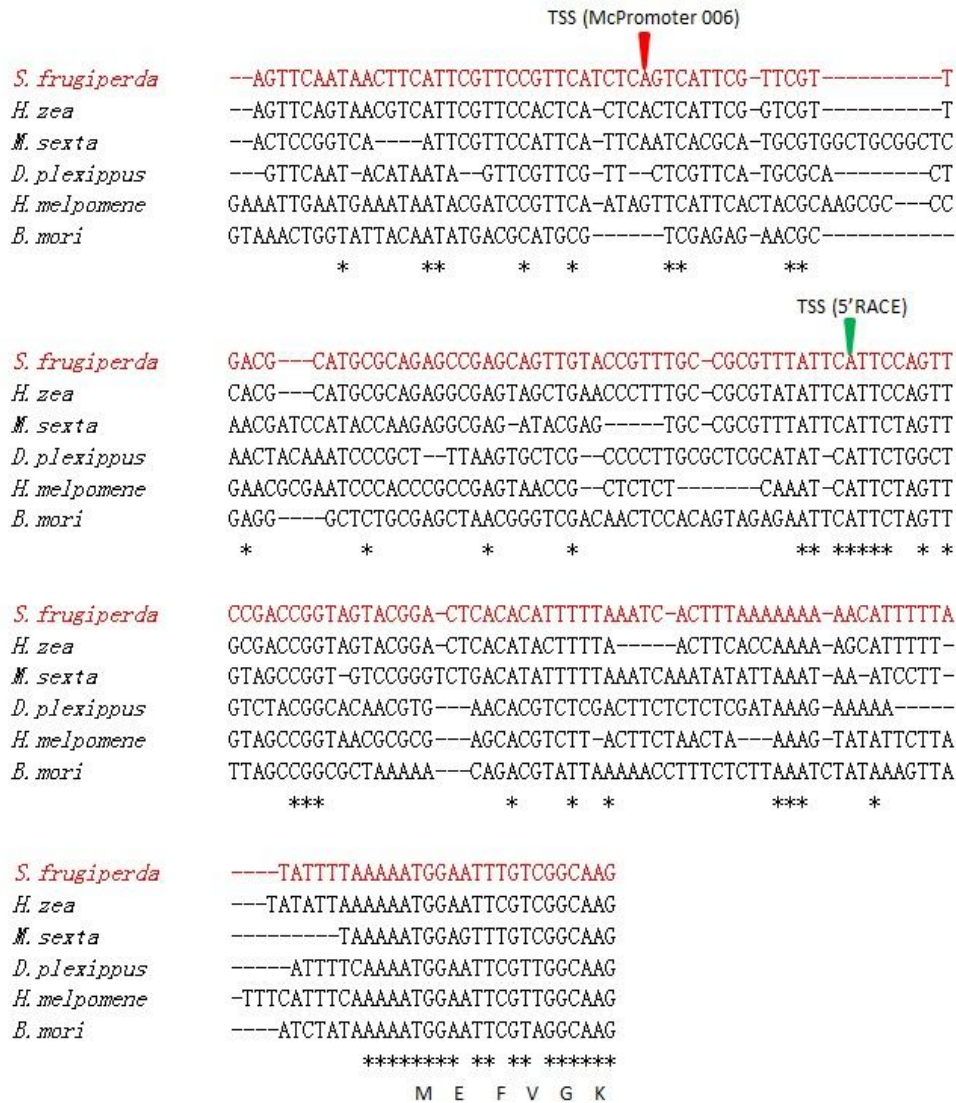
39

TSS (McPromoter 006)

```
S. frugiperda    --AGTTCAATAACTTCATTCGTTCCGTTCATCTCAGTCATTCG-TTCGT----------T
H. zea           --AGTTCAGTAACGTCATTCGTTCCACTCA-CTCACTCATTCG-GTCGT----------T
M. sexta         --ACTCCGGTCA----ATTCGTTCCATTCA-TTCAATCACGCA-TGCGTGGCTGCGGCTC
D. plexippus     ---GTTCAAT-ACATAATA--GTTCGTTCG-TT--CTCGTTCA-TGCGCA--------CT
H. melpomene     GAAATTGAATGAAATAATACGATCCGTTCA-ATAGTTCATTCACTACGCAAGCGC---CC
B. mori          GTAAACTGGTATTACAATATGACGCATGCG------TCGAGAG-AACGC-----------
                        *      **    *   *      **         **
```

TSS (5'RACE)

```
S. frugiperda    GACG---CATGCGCAGAGCCGAGCAGTTGTACCGTTTGC-CGCGTTTATTCATTCCAGTT
H. zea           CACG---CATGCGCAGAGGCGAGTAGCTGAACCCTTTGC-CGCGTATATTCATTCCAGTT
M. sexta         AACGATCCATACCAAGAGGCGAG-ATACGAG-----TGC-CGCGTTTATTCATTCTAGTT
D. plexippus     AACTACAAATCCCGCT--TTAAGTGCTCG--CCCCTTGCGCTCGCATAT-CATTCTGGCT
H. melpomene     GAACGCGAATCCCACCCGCCGAGTAACCG--CTCTCT-------CAAAT-CATTCTAGTT
B. mori          GAGG----GCTCTGCGAGCTAACGGGTCGACAACTCCACAGTAGAGAATTCATTCTAGTT
                     *         *         *        *           ** ***** * *
```

```
S. frugiperda    CCGACCGGTAGTACGGA-CTCACACATTTTTAAATC-ACTTTAAAAAAA-AACATTTTTA
H. zea           GCGACCGGTAGTACGGA-CTCACATACTTTTA-----ACTTCACCAAAA-AGCATTTTT-
M. sexta         GTAGCCGGT-GTCGGGTCTGACATATTTTTAAATCAAATATATTAAAT-AA-ATCCTT-
D. plexippus     GTCTACGGCACAACGTG---AACACGTCTCGACTTCTCTCTCGATAAAG-AAAAA-----
H. melpomene     GTAGCCGGTAACGCGCG---AGCACGTCTT-ACTTCTAACTA---AAAG-TATATTCTTA
B. mori          TTAGCCGGCGCTAAAAA---CAGACGTATTAAAAACCTTTCTCTTAAATCTATAAAGTTA
                     ***            *    *  *          ***      *
```

```
S. frugiperda    ----TATTTTAAAAATGGAATTTGTCGGCAAG
H. zea           ---TATATTAAAAAATGGAATTCGTCGGCAAG
M. sexta         ---------TAAAAATGGAGTTTGTCGGCAAG
D. plexippus     -----ATTTTCAAAATGGAATTCGTTGGCAAG
H. melpomene     -TTTCATTTCAAAAATGGAATTCGTTGGCAAG
B. mori          ----ATCTATAAAAATGGAATTCGTAGGCAAG
                         ******** ** ** ******
                           M  E   F  V  G  K
```

Figure 3-14   Multiple sequence alignment of crabp promoters

Promoter sequence of *H. zea* was obtained from previous unpublished work. *M. sexta* accession#: JH668378.1. *D. plexippus* scaffold#: DPSCF300001 (Zhan et al., 2011). *H. melpomene* accession#: HE667817.1. *B. mori* accession#: DF090534.1.

3.4. Discussion

3.4.1.   Transcription start site

Most of the sequence of the nested PCR product obtained from RACE matched with the EST sequence of *crabp*, indicating that the CRABP-1 and CRABP-2 primers

40

had targeted the correct gene (Figure 3-5). The sequence ends with an area that matched the sequence of the SMARTer II A Oligo, which annealed to the first cDNA strand during the 5'RACE reaction. If we assume that little mRNA degradation was present during the reaction, the sequence located next to the annealing site of the SMARTer II A Oligo would be the experimentally determined TSS (Figure 3-5). In *M. sexta*, the experimentally determined start site was 70 bp upstream of the start codon; its sequence is similar to that of *S. frugiperdia*. However, the sequence between the TSS and start codon share little conservation across the two species (Figure 3-14). The other four lepidoptera species, whose TSSs were not experimentally determined, contain the same conserved sequence at 69-76 bp upstream of the start codon (Figure 3-14). Based on these observations, there is strong likelihood that the experimentally determined start site is correct.

3.4.2. TATA box

The TATA box identified by ConSite has the sequence: TATATAA, which located at 127 bp upstream of the experimentally determined TSS (Figure 3-13). This distance seemed to be different from most eukaryotic genes, which varies between 25 to 35 bp upstream of the TSS. In *H. zea*, the predicted TATA box was at the same location, with the sequence TATAAAA. However, for more distant species such as *M. sexta*, no potential TATA box was found anywhere within the first few hundred base pairs upstream of the start site, which could suggest that such an element is not necessary for the transcription of this gene. Examples of genes lacking TATA box were found to be prominent throughout the animal kingdom, such as the rat insulin-like growth factor-binding protein-2 gene (Brown and Rechler, 1990) and various mammalian ribosomal-protein genes (Perry et al., 1992).

3.4.3. Nuclear receptor binding sites

The *crabp* promoter sequences from *M. sexta, D. plexippus, H. melpomene,* and *B. mori* were obtained from whole genome sequencing projects (Figure 3-14); whereas the sequence from *H. zea* was obtained from previous unpublished work. Among these five lepidoptera species, *H. zea* is the species closest to *S. frugiperda*; therefore, we focused on conserved regions between the two species in the search for potential

nuclear receptor binding sites which are evolutionarily conserved between different species. By using ConSite, potential human COUP-TF, RORα1, RXR-VDR heterodimer, and a *D. melanogaster* USP binding sites were predicted, in the conserved cluster between -112 and -79 (Figure 3-13).

Based on our hypothesis mentioned in section 1.5, nuclear receptors that regulate the expression of iLBP coding genes may become activated through binding to fatty acids or other hydrophobic molecules. Among the four nuclear receptors, COUP-TF is known to become activated by retinoic acid, and to initiate the expression of reporter construct containing corresponding regulatory elements. Although the concentration of retinoic acid needed in *in vitro* experiments was substantially higher than physiological levels, this evidence strongly suggests that this receptor can be activated through retinoic acid binding (Kruse et al., 2008). Similarly, RXR was also shown to become activated and dimerize with other protein components upon binding to 9-cis retinoic acid. Since CRABP is a protein thought to be involved in retinoic acid transport, regulation of its gene by retinoic acid is not unlikely, which could be mediated by retinoic acid activated receptor homologous to receptors such as COUP-TF or RXR. On the other hand, RORα1 is a receptor that does not bind all-trans retinoic acid, but rather, melatonin and cholesterol derivatives, making it less likely to be involved in the regulation of *crabp* (Wiesenberg et al., 1995). Lastly, USP is an invertebrate homolog of mammalian RXR, which can also dimerize with other receptors, such as the ecdysone receptor (EcR), and control developmental events. Natural high-affinity ligands for USP, a highly conserved protein with a large hydrophobic pocket, have not been identified, and thus USP remains an orphan receptor (Clayton et al., 2001). More than a decade ago (Jones and Sharp, 1997), it was demonstrated that USP can interact with juvenile hormone III (JH III) while forming homodimers, and activate reporter construct containing a direct repeat response element, so its role as juvenile hormone receptor remains highly controversial (Maki et al., 2004). Better characterized is the role of vertebrate RXR in the regulation of many genes, including the genes coding iLBPs (Hewson et al., 2002; Hwang et al., 2002; Poirier et al., 1997); as vertebrate and invertebrate iLBPs have highly conserved gene and protein structures, their genes could also be under similar control mechanisms, and hence it is possible that the insect RXR-analog USP

also plays a role in regulating the fatty acid-controlled expression of insect genes such as *crabp.*

In order to determine whether the sequence cluster between -112 and -79 indeed contains elements involved in regulation of the *crabp* gene, reporter gene studies could be carried out with constructs in which suspected regulatory elements in the promoter sequence have been altered. Such a study requires a suitable cell system that is easily transfected with the constructs, and is able to express genes under the control of the *crabp* promoter.

4.   Analysis of the SF9 cell line

4.1. Introduction

4.1.1.   GFP reporter gene

In molecular biology, the term reporter assay often refers to attaching the coding region of a gene (known as the reporter gene) to the promoter sequence of a different gene of interest. By transferring this artificial reporter construct into living bacteria, fungi, plants, animals, or cell cultures, one will be able to characterize and study the strength of the promoter by measuring the expression level of the reporter gene. In order to have the measurement achieved easily, the reporter gene must not be naturally expressed in the host organism, while possessing certain characteristics that allow easy detection and quantification of its expression. For instance, the most commonly used reporter genes often enable visual detection and quantification due to their enzymatic activity, which can result in the production of colored, fluorescent or luminescent molecules. One such example is the *lacZ* gene from the bacteria *Escherichia coli*, which encodes the polypeptide β-galactosidase, an enzyme that produces a blue dye by cleaving the substrate analog 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-gal). To quantify the activity of this enzyme, cells from the host organism need to be isolated and stained with X-gal. The disadvantage of this approach is that it is necessary to destroy the cells for the purpose of quantification, thus causing difficulties if the sample has to be kept for downstream experimental analysis.

In order to overcome this difficulty, the *gfp* gene discovered from the jellyfish *Aequorea victoria* can serve as an alternative. The *gfp* gene encodes the green fluorescence protein (GFP), a 26.9 kDa protein that exhibits bright green fluorescence when exposed to light at the blue to ultraviolet range (Prendergast and Mann, 1978). The properly folded form of this polypeptide possesses a major excitation peak at the wavelength of 395 nm and a minor one at 475 nm; whereas the major emission peak is

at 509 nm. Unlike β-galactosidase, the detection and quantification of GFP does not require the addition of substrates, but only needs the samples being exposed to light with the proper wavelength. As a result, samples can be kept away from potential toxicity of substrates and used for other experimental analysis after detection and quantification.

4.1.2.   Choosing a host system

Choosing a proper host system is a very important task if one decides to use the reporter assay to study the strength of the promoter of a target gene. Ideally, the host must be able to naturally express the target gene, so that when the reporter construct is transferred into the host, the reporter gene will be expressed in similar manner as the target gene. This is important when one wants to identify any regulatory elements within the promoter. In order to do so, a deletion analysis can be performed by deleting any potential sequence elements (such as the -112 to -79 sequences in section 3.4.3) within the promoter, followed by comparison of the reporter gene expression with the reporter construct that has the full length promoter (Figure 4-1). Using this strategy, key regulatory elements can be identified.
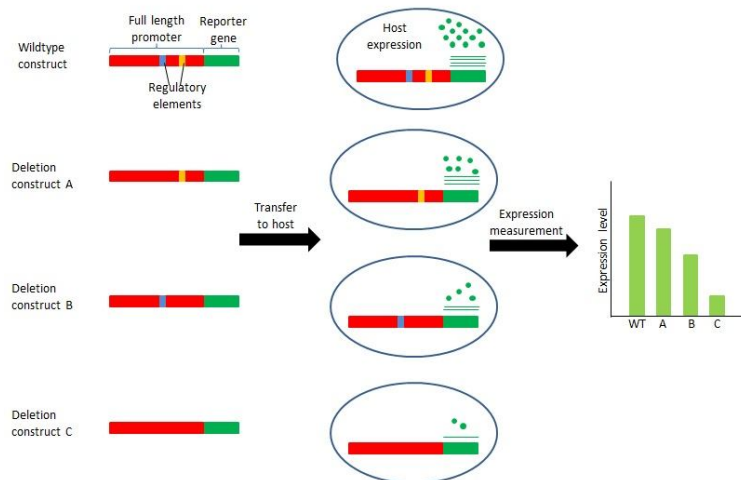

Figure 4-1   Overview of promoter analysis by reporter assay

Sometimes, after the reporter constructs have been transferred to the host, a treatment with hormone analogs is performed on the host. This treatment procedure may allow certain regulatory sequences that are hormone response elements (HREs) to

45

be identified (Beckers et al., 1997). As depicted in Figure 4-2, hormone treatment of host cells transfected with constructs containing the full length promoter increases the expression level of the reporter gene, indicating that at least one HRE is present in the promoter. When combined with deletion studies, it can be determined that element A is a HRE, while element B is not (Figure 4-2). This combined strategy has been widely used in finding and characterizing HREs (Montoliu et al, 1995). However, a major requirement for this type of study is to have a host system that not only expresses the target gene, but is also able to regulate its expression through the act of hormones or hormone analogs. One example related to the iLBP family is the discovery of a fatty acid response element within the promoter of rat H-FABP (Qu et al., 2007). This gene is naturally expressed in L6 cardiomyocytes, and up-regulated in response to fatty acid exposure.
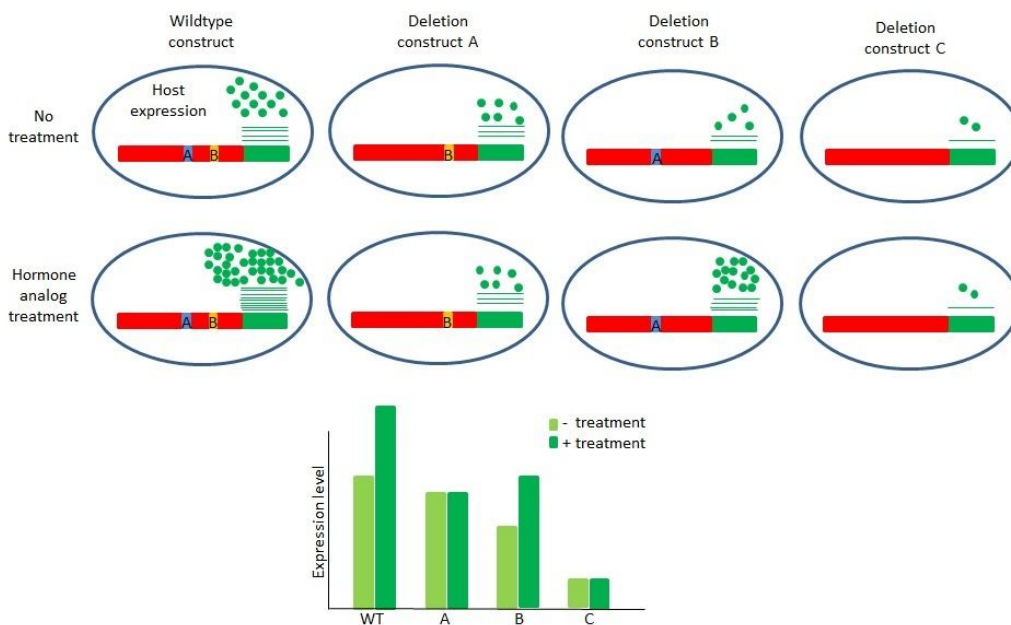

Figure 4-2   Strategy for finding HREs using reporter assay

While far fewer well-characterized insect cell lines exist, a commercial protein expression system is based on lepidoptera cells, which can result in very strong expression from a baculovirus promoter. The SF9 cells of this system, which is derived from an ovarian cell line of *S. frugiperda*, can be easily transfected and maintained in cell culture (Vaughn et al., 1977). In this study, we intended to test if the SF9 cells can be

used as a suitable system to identify any potential fatty acid response element within the promoter of the *crabp* gene.

4.1.3.   Cloning of the crabp promoter

The reporter construct was made by cloning the ~1.5 kb promoter sequence into pZsGreen1-DR vector, a promoterless vector that encodes ZsGreen1-DR, a variant of a *Zoanthus* species GFP, which has the maximum excitation and emission at 496 nm and 506 nm, respectively (Figure 4-3). In order to evaluate whether the construct was able to successfully produce GFP in *S. frugiperda*, it was used to transfect the SF9 cell line. The resulting GFP expression was then evaluated using a BioTek fluorescence microplate reader.
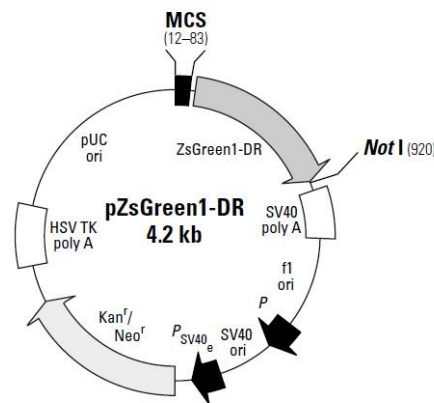

Figure 4-3   Map of pZsGreen1-DR vector
The GFP coding sequence is located downstream of the MCS. Clontech Laboratories, Inc. Used with permission.

It is commonly assumed most regulatory elements are contained within first kilobase upstream of the proximal promoter of a gene, and therefore we cloned 1.5 kb of sequence upstream of the start codon of the *crabp* gene into the reporter gene vector. To do so, we amplified the promoter sequence using a set of primers that contained adaptor sequences complementary to the sequence at the vector insertion site (Figure 4-4, A, B). Next, the vector was amplified with primers that also anneal at the same site (Figure 4-4, B). The two PCR products were then joined into a circular vector using the CloneEZ® PCR Cloning Kit from GenScript (Cat#: L00339) (Figure 4-4, C).   The

47

resultant vector, named pZsCRABP, had the complete 1.5 kb promoter sequence placed immediately upstream of the GFP start codon (Figure 4-4, D).

Figure 4-4   Cloning strategy for the crabp promoter

A: PCR amplification from genomic DNA. B: PCR amplification from pZsGreen1-DR vector. C: Recombination between PCR products. D: Circular product from recombination. Red and yellow highlights indicate homologous sequences.

4.1.4.   Transfection

Transfection is the process of deliberately introducing nucleic acids into cells. It usually involves opening of transient pores in the cell membrane in order for the uptake of foreign material to proceed. Some typical methods of transfection are electroporation, calcium phosphate-mediated delivery, and deposit by liposomes made from cationic lipids (Safari and Hosseinkhani, 2013; Schroeder et al, 2013; Sun et al, 2013). All these processes can successfully transport foreign nucleic acid through the hydrophobic cell membrane, but also result in varying degrees of cell damage. Once inside the cells, one cycle of cell division must be completed in order to let the foreign genetic material reach the transcription machinery inside the nucleus, so that its genes can be expressed.

In electroporation an electric field is applied to cells to increase the electric conductivity of the cell membrane, causing the lipid membrane to undergo structural

changes and creating hydrophilic pores that allow the entrance of foreign substances such as DNA plasmids, which could never passively diffuse across the hydrophobic lipid bilayer (Neumann et al., 1982). One problem this process causes is that the transient hydrophilic pores formed may not only contract and seal the cell membrane as intended, but could also expand and eventually rupture the membrane. The outcome often depends on the sizes of the pores, which are difficult to control (Joshi and Schoenbach, 2000). As a result, the degree of cell death often greatly reduces the efficiency of transfection.

The transfection method using calcium phosphate was originally discovered in the early 1970s (Graham, 1988). It works by forming precipitates of plasmid DNA and calcium ions, which attach to cell surface and are randomly taken into cells by endocytosis (Bose and Tarafder, 2012). It is well known for its simplicity, low toxicity to the cells, and low cost; however, a major disadvantage of the method is its low transfection efficiency.

Nowadays, one of most widely used methods for transfecting eukaryotic cells involves the usage of cationic lipids, which can physically attach to DNA molecules and form liposomes, small membrane enclosed vesicles that are structurally similar to cells. These vesicles can deliver DNA into the cells by easily fusing with the cell membranes because both contain lipid bilayers (Felgner et al., 1987). This gene delivery method is often advantageous over the previous two due to its higher transfection efficiency, while causing little cell death. In this experiment, transfection of SF9 cells was performed and optimized using a cationic lipid based transfection reagents: the Cellfectin® II Reagent from Invitrogen (Cat#: 10362-100).

4.1.5. *Optimization of transfection efficiency*

In order to accurately determine the transfection efficiency of a particular transfection experiment, the expression of the reporter gene from transfected cells must be distinguished from untransfected cells. To achieve this goal, a GFP reporter construct containing a strong promoter in SF9 cells and an effective detection method that provides visual identification of any GFP expressing cells is required.

In 1992, a second immediately early (IE-2) gene from the baculovirus *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus (OpMNPV) was identified and characterized. It was discovered that the constitutive expression of this gene relies primarily on the cis-acting regulatory elements within its promoter sequence (Theilmann and Stewart, 1992). Although the natural host of the virus is the Douglas fir tussock moth *Orgyia pseudotsugata*, this promoter was also found to be strongly active when used for protein expression in the SF9 cell line (Hegedus et al., 1998). This promoter, known as the OpIE2 promoter, was cloned into the MCS region of the pZsGreen1-DR vector, and used as a construct to determine transfection efficiency. Cells that constitutively express GFP are visually detectable, and thus can be used for estimating transfection efficiency (Figure 4-5).
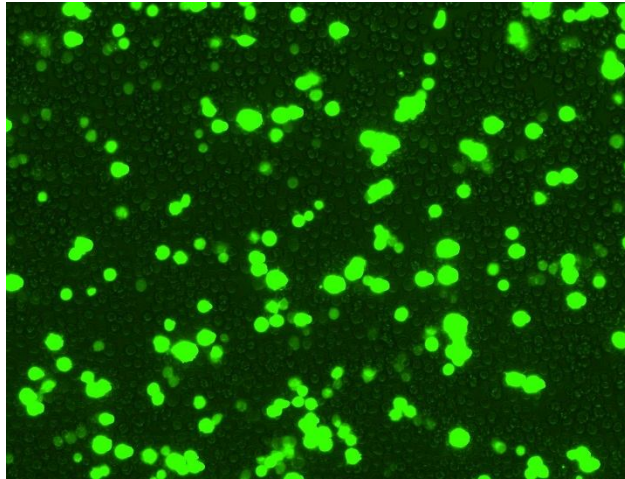

Figure 4-5   Transfection of SF9 cells

SF9 cells transfected with pZsOpIE2 construct using the Cellfectin® II Transfection Reagent.

With the aid of the pZsOpIE2 construct and the UV fluorescence microscope, the transfection efficiency from different experimental conditions can be accurately determined. Once the best experimental conditions are established, the cells can be transfected with the pZsCRABP construct, treated with fatty acids or other substrates, and subjected to GFP expression measurement.

4.2. Materials and methods

4.2.1. Fatty acid treatment

The following long chain fatty acids were obtained from Sigma Aldrich dissolved in DMSO to a concentration of 100 mM, and used to treat SF9 cells: palmitic acid (16:0, Cat#: P0500), oleic acid (18:1, O7501), linoleic acid (18:2, L1012), and arachidonic acid (20:4, A3555). Before treatment, 10^6 adherent SF9 cells were plated in each well of a 24-well tissue culture plate, and allowed to attach to the bottom. Once attached, growth medium in each well was aspirated and replaced with 500 µL of treatment solution containing 100 µM of fatty acid. The cells were then incubated at 27°C for four hours. Each treatment was repeated three times. When the incubation period was over, cells were subjected to total RNA extraction as outlined in section 2.2.1. cDNA synthesis and real-time PCR analysis were performed with the same procedure as outlined in section 2.2.2 and 2.2.5, using the same cycling parameters and primers. Each PCR reaction had four replicates.

4.2.2. PCR amplification of crabp promoter and vector

To add adaptor sequence homologous to the pZsGreen1-DR vector to the *crabp* promoter sequence, roughly 100 ng of *S. frugiperda* genomic DNA was mixed with 25 µmoles of CRABPpZs-F (AGCGCTACCGGACTCAGATCTAACATTCCGTTGTATCGTC C) and CRABPpZs-R (CCGTGCTTGGACTGGGCCATTTTTAAAATATAAAAATGTTTTT TTTTAAAGTGAT) primers, along with 1 µL of 10 mM dNTP mix. In addition, 10 µL of 5X iProof HF Buffer and 1 µL of iProof DNA Polymerase from BioRad's iProof™ High-Fidelity DNA Polymerase kit (Cat#: 172-5300) were added into the mixture, along with nuclease free water to bring the total reaction volume to 50 µL. Thermal cycling parameters were: initial denaturation at 98°C for 5 minutes, followed by 40 cycles of 98°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes, along with a final extension at 72°C for 10 minutes.

The pZsGreen1-DR vector was linearized using SmaI in the same procedure as outlined in section 3.2.3. Once digestion and purification were completed, roughly 50 ng of the linearized vector were mixed with 25 µmoles of pZsGFP-F (ATGGCCCAGTCCAA

GCACGG) and pZsMC-R (AGATCTGAGTCCGGTAGCGCT), along with the appropriate amount of PCR ingredient as described from the PCR above. All cycling parameters were maintained except that the annealing temperature and the time for extension has been changed to 62°C and 5 minutes, respectively. The resulting PCR products from both reactions were analyzed on a 1.0% agarose gel and purified.

4.2.3. Recombination cloning

Once purified and quantified using the Nanodrop spectrophotometer, ~220 ng of each of the two PCR products were mixed, along with 2 µL of 10X buffer and CloneEZ enzyme from the CloneEZ® PCR Cloning Kit from GenScript (Cat#: L00339), giving a total reaction volume of 20 µL. The mixture was incubated at 22°C for 30 minutes, then used for transformation directly.

Prior to transforming DH5α *E. coli* cells (Invitrogen, cat#: 18258-012) with the recombination reaction product, 50 µL of cells were thawed on ice. Once thawed, 8 µL of the reaction product were mixed with the cells. The mixture was incubated on ice for 30 minutes. Next, the cells were placed in 42°C water bath for 90 seconds, and placed on ice for 3 minutes. SOC medium (600 µL) was slowly added to the cells, followed by incubation at 37°C on a 225 rpm shaker for 1 hour. Cells were pelleted by centrifuging at 4000 rpm for 5 minutes. Next, an aliquot of 500 µL of the supernatant was then removed from the cells, and the cell pellet was re-suspended using the remaining supernatant. From this all suspension, 10 µL and 100 µL were plated separately on LB plates containing 100 µg/mL of kanamycin. The plates were incubated at 37°C overnight.

On the next day, a total of 24 colonies from both plates were selected and used as templates for colony PCR. For each reaction, 3.5 µL of Fermentas' 2X PCR Master Mix (Cat#: K0172) was mixed with 1.4 µmoles of pZs-1 (AGCGCTACCGGACTCAGATCT) and pZs-2 (CTCCACCACGCACAGGTTGATGGCCT) primers, along with nuclease free water to add up to a total reaction volume of 7 µL. To each of the PCR tube, a portion of a single colony was picked using a pipet tip and mixed with the reaction mixture. The remaining colony on the pipet tip was streaked over a LB plate (containing kanamycin) and allowed to grow at 37°C. PCR cycling parameters were: initial denaturation at 95°C for 5 minutes, followed by 35 cycles of 95°C for 1 minute, 57°C for

30 seconds, and 72°C for 2 minutes, along with a final extension at 72°C for 10 minutes. PCR products were analyzed on a 1.0% agarose gel in order to identify positive colonies.

When positive colonies were identified, each was used to inoculate 5 mL of liquid LB contained kanamycin, which was then placed in a 37°C shaker overnight. On the next day, plasmid DNA from the liquid bacteria culture was isolated using QIAGEN's QIAprep Spin Miniprep Kit (Cat#: 27104). The concentrations were measured using the Nanodrop spectrophotometer; and the plasmids were sent for sequencing using pZs-1 and pZs-2 primers. When sequencing results confirmed about the correct insertion orientation of the *crabp* promoter, the corresponding colony was used to make several liquid cultures and glycerol stocks.

4.2.4. Transfection

SF9 adherent cells were allowed to grow to ~80% confluence inside a 25 cm² Corning cell culture flask (Fisher Scientific, Cat#: 10-126-28). Upon reaching the desired cell density, the growth medium was replaced with 5 mL of transfection medium, which was composed of 75 µL of fetal bovine serum (Gibco® Sera, Cat#: 12483-020), 675 µL of Grace's supplemented insect medium (Invitrogen, Cat#: 11605-102), and 4250 µL of Grace's unsupplemented insect medium (Invitrogen, cat#: 11595-030). Next, 3 µg of pZsOpIE2 plasmid and 24 µL of Cellfectin® II Reagent were separately dissolved in 300 µL of unsupplemented medium, and incubated at room temperature for 30 minutes. The two solutions were mixed together, and incubated at room temperature for an additional 15 minutes. The reaction mixture was slowly added into the cell culture flask containing the transfection medium, dissolved by gentle swirling, and incubated at 27°C for 4 hours. This incubation period was later increased to 29 hours for maximizing transfection efficiency. After this period, the medium inside the flask was replaced with 5 mL of growth medium, followed by an additional 36 hours growth period at 27°C. Cells were then observed under the UV inverted microscope for estimating the transfection efficiency.

4.2.5. GFP quantification

Once the optimal experimental condition was established, cells were transfected with the pZsCRABP construct using the optimized transfection protocol. Transfected cells were allowed to grow to 100% confluence, and were plated onto 8 separate wells of a 24-well cell culture plate, with each well having 10^6 cells. The plate was placed inside a BioTek fluorescence microplate reader for fluorescence reading. Reading parameter settings were: excitation at 485 nm, emission at 528 nm, detection from top of the plate with 100% sensitivity. The average reading from the 8 wells were compared with those transfected with pZsOpIE2 and untransfected SF9 cells using unpaired student t-test.

4.3. Results

4.3.1. Fatty acid treatment

In order to determine if the expression of the housekeeping gene *rpl37a* was stable across different treatment groups, the ANOVA test was used to compare the Ct values (Appendix D). The results (F=1.172, P=0.38) indicated that no significant difference existed between test groups.

In order to normalize the initial quantity of *crabp* mRNA against *rpl37a* mRNA, the computation method mentioned in section 2.3.4 was applied, under the assumption that the efficiency of each reaction was similar to each other. The normalized expressions (Table 4-1) were compared between different treatment groups using the ANOVA test, which resulted in no significant differences (F=1.333, P=0.33). The results maintained the same when fatty acid concentration for each treatment was increased to 500 µM and 1000 µM, or when the treatment period was prolonged to 6, 8, 10, and 24 hours (data not shown).

Table 4-1   *crabp expression normalized against rpl37a*

Treatment	*crabp* expression	mean	variance
Control	0.07	0.06	0.0006
	0.09	0.06	0.0006

Treatment	*crabp* expression	mean	variance
	0.04	0.06	0.0006
Oleic acid	0.06	0.06	0.0004
	0.04	0.06	0.0004
	0.08	0.06	0.0004
Arachidonic acid	0.07	0.08	0.0007
	0.06	0.08	0.0007
	0.11	0.08	0.0007
Palmitic acid	0.08	0.09	0.0002
	0.09	0.09	0.0002
	0.11	0.09	0.0002
Linoleic acid	0.08	0.08	N/A

4.3.2.   Cloning of the crabp promoter into pZsGreen1-DR vector

When the primers CRABPpZs-F and CRABPpZs-R were used to amplify the *crabp* promoter sequence from genomic DNA template, a clear band appeared at the position of the 1.5kb band when the PCR products were analyzed by agarose gel electrophoresis (Figure 4-6). After purification, the DNA concentration was 41.1 ng/uL.
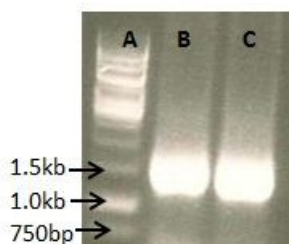

Figure 4-6   PCR amplification of the crabp promoter
A: 1.0 DNA ladder. B, C: PCR product.

When the pZsGreen1-DR vector digested with SmaI and undigested vector were analyzed by agarose gel electrophoresis, a shift in band position was observed, confirming the success of the digestion (Figure 4-7). When the digested vector was

used as template for PCR amplification using pZsGFP-F and pZsMC-R primers, the PCR product appeared to have a size between 4.0 and 5.0 kb, which matched with the expected size of 4194 (Figure 4-8). The PCR product was purified and had a concentration of 20.7 ng/uL.
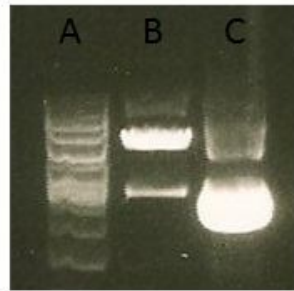

Figure 4-7 SmaI digestion of the pZsGreen1-DR vector
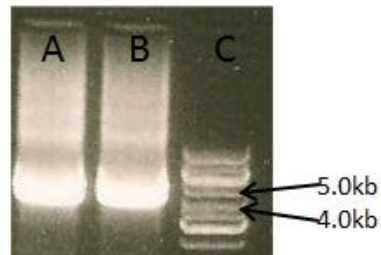A: 1.0 DNA ladder. B: SmaI digested vector. C: undigested vector.

Figure 4-8 PCR amplification of the pZsGreen1-DR vector
A, B: PCR product. C: 1.0 DNA ladder.

After transformation and colony PCR, gel electrophoresis analysis result indicated that four out of the twenty-four selected colonies appeared to have a strong band above the 1.5 kb band, which matched with the expected size of 1746 bp for successful insertion of the promoter sequence (Figure 4-9, lane B, H, O, U). These were colony #1, 7, 13, and 19. Nineteen colonies had a PCR band below the 250 bp band, consistent with the expected size of 171 bp for an insertion failure (Figure 4-9, lane C-G, I-K, M, P-T, V-Z). No PCR band was observed for colony #11 (Figure 4-9, lane L). The

plasmid isolated from colony #1 was sent for Sanger's sequencing using the primers pZs-1 and pZs-3, to confirm the correct insertion orientation (Figure 4-10).

Figure 4-9   Colony PCR results

A, N: 1.0 DNA ladder. B-M, O-Z: colony PCR products.

Figure 4-10   Cloning orientation confirmation

Sanger's sequencing results using pZs-1 and pZs-3 primers were aligned with hypothetical construct sequence near the multiple cloning region (A) and the GFP coding region (B). Yellow highlight: multiple cloning region sequence. Green highlight: GFP coding sequence. Red highlight: *crabp* promoter sequence.

4.3.3.   Transfection

When SF9 cells were transfected with the pZsOpIE2 construct, a transfection efficiency of about 8% was observed following an incubation period of 4 hours (Figure 4-11, A). This percentage was increased to about 30% when the incubation period was extended to 29 hours (Figure 4-11, B). Using the condition with the higher efficiency, SF9 cells were transfected with the pZsCRABP. The pZsOpIE2 construct and non-transfected SF9 cells were served as positive and negative control, respectively. After

three attempts, fluorescence produced by cells transfected with the pZsCRABP construct was similar to non-transfected SF9 cells (Figure 4-12). The cells transfected with pZsOpIE2 construct, however, were able to produce significantly more fluorescence, where P<0.0001 using the student t-test.
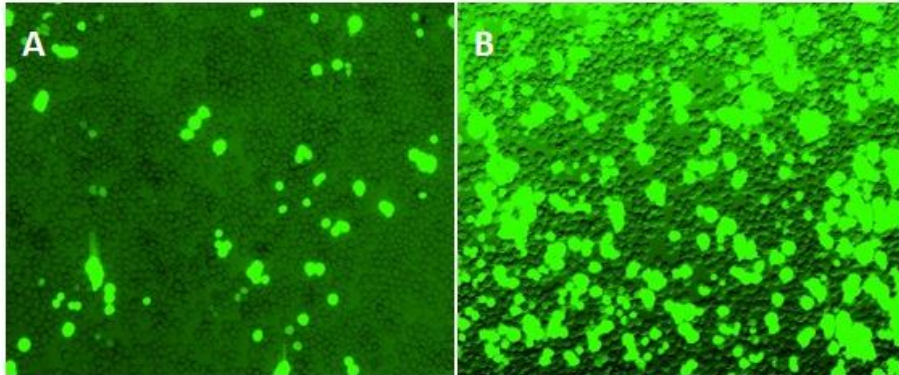

Figure 4-11   Optimization of transfection efficiency

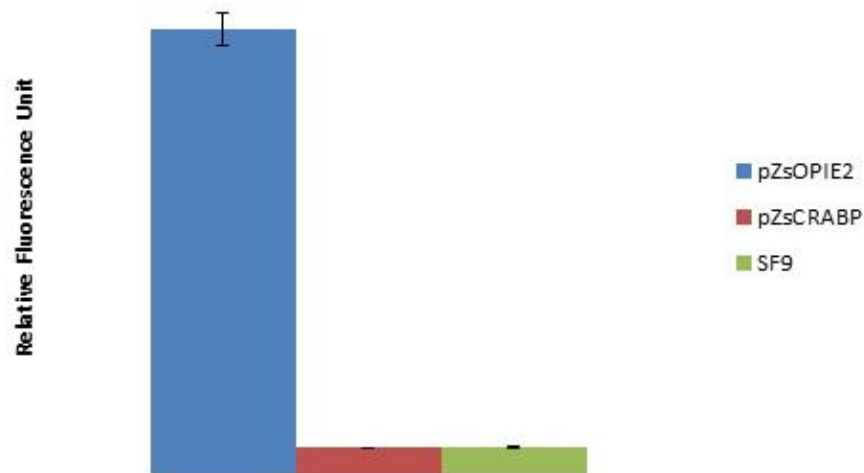A: 4 hours post-transfection incubation. B: 29 hours post-transfection incubation.

Figure 4-12   Fluorescence detection by microplate reader

Error bars represent standard deviations. n=8.

4.4. Discussion

4.4.1. Fatty acid treatment on SF9 cells

Based on the assumption that real-time PCR efficiency remained unchanged across all tested samples, the ANOVA test results suggested that none of the four fatty acid treatments affects the expression of *crabp* in SF9 cells. This is in contrast to previous observations in mammalian L6 myoblasts, where significant increases in H-FABP expression were observed following incubation with similar amounts of fatty acids, bound to bovine serum albumin (Qu et al., 2007). While this could indicate that the SF9 cells may require a carrier protein similar to bovine serum albumin to effectively deliver the fatty acid molecules to their surface, it is also possible that the expression of *crabp* cannot be up-regulated by fatty acids in SF9 cells, either because a necessary transcription factor is absent, or because a fatty acid response element does not exist in the gene. In any case, SF9 cells would not be a suitable system to study the fatty acid controlled regulation of *crabp*.

4.4.2. Testing of reporter construct

Although we do not have evidence to show that the expression of *crabp* can be up-regulated in response to fatty acid exposure, SF9 cells can still be used to test the ability of the pZsCRABP construct to express GFP. The pZsCRABP construct contains 1.5 kb sequence upstream of the start codon of *crabp*, which is generally believed to contain most of the elements crucial for transcription. However, when SF9 cells were transfected with the construct using the optimized condition, little differences in fluorescence was detected when compared with non-transfected cells, which suggests that the amount of GFP produced is too small to generate detectable fluorescence levels (Figure 4-12). This could be due to a very low level of natural expression of the *crabp* gene in SF9 cells, or due to limitations of the pZsCRABP reporter construct, which may not contain all the elements crucial for strong expression in the -1500 bp upstream of the transcription start site.

While it is still commonly believed that most regulatory sequences are contained in the vicinity to the promoter, many studies have demonstrated that important regulatory

sequences can be located in more distant locations, or within intron sequences. Chung and Perry (1989) have demonstrated that deletion of the first intron of the mouse ribosomal protein gene *rpl32* causes its transcription rate to drastically decrease to less than 20% of maximum level. In 2005, Logette and colleagues have described the presence of a sterol regulatory element binding protein response element located within the first intron of the human *CASP-2* gene (Logette et al., 2005). Moreover, recent evidence has demonstrated that non-intronic genes exhibit higher transcription rate when an intron has been inserted, whereas deletion of a natural intron can cause significant decrease in the transcription rate (Moabbi et al., 2012).

5. General discussion

5.1. Flight muscle iLBP

The initial goal of this study was to identify any FABP coding gene(s) in the flight muscle of *S. frugiperda* and study their up-regulation by fatty acid-dependent flight. Based on our analysis of the data from flight experiments, the expression of the *crabp* gene appears to be mildly affected by flight activity. However, the observed trend was significantly weaker than in 3 week old adult desert locusts, where the expression increases several-fold during flight (Chen and Haunerland, 1994). This may be because *S.* frugiperda do not usually migrate as far and as fast as locusts, and hence do not require the same high rate of fatty acid transport.

The insect order Lepidoptera is believed to have originated around 299 to 251 million years ago; whereas the more primitive insect order Orthoptera (which includes the desert locust) arose earlier (359 to 299 million years ago) (Peck, 1991). Even though considered "less primitive", *S. frugiperda* possess a 16 times smaller genome than the locust (407 Mb vs. ~6,500 Mb) (Negre et al., 2006). Thus, over the history of evolution, the insects may have lost certain genes, but there is no evidence that genes involved in fatty acid transport, metabolism, or signaling have been eliminated. In Lepidoptera as well as in Orthoptera, fatty acids are the primary energy source for extended flight activity, but *Lepidopterean* have a lower metabolic rate and hence express less FABP/CRABP. These small iLBP genes may act as a fatty acid transport protein, but to a lesser extent than in locusts, and may not be required in extremely high concentrations.

5.2. The SF9 cells

The SF9 cell line was originally established from ovarian tissues of *S. frugiperda* and is commonly used for recombinant protein production by means of a baculovirus expression system. In this study, we have chosen this easily maintained cell line as potential host cell for reporter genes to study the expression and regulation pattern of the *crabp* gene. However, treatment with linoleic acid, oleic acid, palmitic acid, or arachidonic acid all appeared to have no effect on the expression level of the *crabp* gene. Three of these four fatty acids were found in great abundance in total lipid extracts from the insect (Khani et al., 2007): linoleic acid (16%-30% of whole lipid composition) is necessary in the insect diet for the purpose of emergence and wing formation (Fraenkel and Blewett, 1947); oleic acid (35%-39%) and palmitic acid (23-33%) are known precursors of sex pheromones in lepidoptera species (Lofstedt et al., 1986; Roelofs and Wolf, 1988); and arachidonic acid, although not found in great abundance, has a positive effect on the growth of larvae (Fraenkel and Blewett, 1947). Since all four fatty acids play important roles in growth or development of the insect, one would expect the need for an intracellular binding protein, which in turn can be expected to be up-regulated by fatty acids. Our inability to detect any change in *crabp* expression could be caused by insufficient uptake, perhaps due to the lack of a lipoprotein in the growth medium, or simply because the SF9 cell line does not express the necessary nuclear receptors that can be activated by fatty acids. Although we found within the *crabp* promoter potential binding sites for the nuclear receptor USP, which is homologous to vertebrate RXR and is expressed in SF9 cells (Giraudo et al., 2011), it is not clear whether USP is involved in fatty acid signaling or in the regulation of *crabp*. In any case, since we could not induce the expression of *crabp* in SF9 cells, we conclude that the cell line is not a suitable reporter system to study its regulation.

5.3. Future work

The flight experiment designed in this thesis was a preliminary study which required improvement in various aspects in order to produce conclusive results. First, it is necessary to establish that the expression of the reference gene used for normalization in real-time PCR analysis remains constant between muscle tissues under

different experimental conditions. In this study, we used the *rpl37a* gene, one of the classic housekeeping genes used for real-time PCR experiments in mammals. Previously, no studies have been conducted to test the stability of this gene in insect muscle under various conditions, including flight. Thus, in order to minimize the errors produced from expression variations of the reference gene, other housekeeping genes could be tested and used alone or in combination with *rpl37a* for a more robust normalization. This can be achieved with the aid of statistic algorithms such as geNorm, NormFinder, and Bestkeeper (Andersen et al., 2004; Pfaffl et al., 2004; Vandesompele et al., 2002;). Secondly, once a stable reference gene has been chosen, reaction efficiency for both the target and reference gene must be accurately determined for individual samples. This step is crucial for the accurate calculation of the amount of template for each gene, which can be achieved by generating a standard curve based on a dilution series of each cDNA sample (Figure 2-3) (Ramakers et al., 2003). Lastly, in order to produce a conclusion based on statistically significant results, more subjects must be tested for each experimental group, which will minimize the errors induced by outliers. With these conditions optimized, one will able to provide more conclusive evidence about whether or not *crabp* and/or *fabp3* expression can be up-regulated through flight.

Since we have concluded that the SF9 cell line is not suitable for reporter gene study of *crabp*, alternative systems must be considered. However, the choice is limited due to the small number of lepidopteran cell lines, especially from tissues other than ovaries. In insects, the fat body is the major storage site for lipid molecules, equivalent to the mammalian adipose tissue. Hence, fatty acid transport should be prominent in this tissue, and thus fatty acid transport proteins should be expressed strongly. Very few cell lines from fat body have been established, and even fewer sufficiently characterized. Still, cell lines derived from the fat body, such as FB33, a cell line derived from *H. zea* fat body (Kariuki et al., 2000) or IPL-LdFB, derived from the fat body of the gypsy moth *Lymantria dispar* (Ferkovich et al., 1991), may be considered as a more suitable alternative.

5.4. Conclusion

In this study, preliminary flight experiments were performed to study the regulation of *crabp* and *fabp3* in the flight muscle of adult *Spodoptera frugiperda*. Although no solid evidence was found to support the hypothesis that flight activity can up-regulate the expression of either of the two genes, an expression trend similar to the muscle FABP of the desert migratory locust *Schistocerca gregaria* during flight was observed for *crabp*. To analyze the molecular mechanism that governs the regulation of *crabp*, we have sequenced its promoter and attempted to use the SF9 cell line to perform a reporter assay. Since the expression of *crabp* in the SF9 cell line does not appear to be up-regulated in response to fatty acid treatment, alternative strategies to analyze the regulation of this iLBP gene are necessary.

References

Andersen, C., Jensen, J., & Orntoft, T. (2004). Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Research, 64(15), 5245-5250.

Baillie, A., Coburn, C., & Abumrad, N. (1996). Reversible binding of long-chain fatty acids to purified FAT, the adipose CD36 homolog. *Journal of Membrane Biology, 153*(1), 75-81.

Banaszak, L., Winter, N., Xu, Z., Bernlohr, D. A., Cowan, S., & Jones, T. A. (1994). Lipid-binding proteins: A family of fatty acid and retinoid transport proteins. Advances in Protein Chemistry; Lipoproteins, Apolipoproteins, and Lipases, 45, 89-151.

Barber, R., Harmer, D., Coleman, R., & Clark, B. (2005). GAPDH as a housekeeping gene: Analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiological Genomics, 21*(3), 389-395.

Beckers, T., Reilander, H., & Hilgard, P. (1997). Characterization of gonadotropin-releasing hormone analogs based on a sensitive cellular luciferase reporter gene assay. *Analytical Biochemistry, 251*(1), 17-23.

Benning, M., Smith, A., Wells, M., & Holden, H. (1992). Crystallization, structure determination and least-squares refinement to 1.75 angstrom resolution of the fatty-acid-binding protein isolated from manduca-sexta L. *Journal of Molecular Biology, 228*(1), 208-219.

Bonen, A., Benton, C., Campbell, S., Chabowski, A., Clarke, D., Hang, X., Luiken, J. (2003). Plasmalemmal fatty acid transport is regulated in heart and skeletal muscle by contraction, insulin and leptin, and in obesity and diabetes. *Acta Physiologica Scandinavica, 178*(4), 347-356.

Bose, S., & Tarafder, S. (2012). Calcium phosphate ceramic systems in growth factor and drug delivery for bone tissue engineering: A review. *Acta Biomaterialia, 8*(4), 1401-1421.

Brelivet, Y., Kammerer, S., Rochel, N., Poch, O., & Moras, D. (2004). Signature of the oligomeric behaviour of nuclear receptors at the sequence and structural level. *EMBO Reports, 5*(4), 423-429.

Brown, A., & Rechler, M. (1990). Cloning of the rat insulin-like growth factor-binding protein-2 gene and identification of a functional promoter lacking a tata box. *Molecular Endocrinology, 4*(12), 2039-2051.

Cao, X., Luo, X., Liang, J., Zhang, C., Meng, X., & Guo, D. (2012). Critical selection of internal control genes for quantitative real-time RT-PCR studies in lipopolysaccharide-stimulated human THP-1 and K562 cells. *Biochemical and Biophysical Research Communications, 427*(2), 366-372.

Cavagnari, B., Scaraffia, P., Haller, J., De Burgos, N., & Santome, J. (2000). Presence of a fatty acid-binding protein and lipid stores in flight muscles of dipetalogaster maximus (hemiptera : Reduviidae). *Journal of Medical Entomology, 37*(6), 938-944.

Chen, X., & Haunerland, N. (1994). Fatty-acid-binding protein expression in locust flight-muscle - induction by flight, adipokinetic hormone, and low-density lipophorin. *Insect Biochemistry and Molecular Biology, 24*(6), 573-579.

Chung, S., & Perry, R. (1989). Importance of introns for expression of mouse ribosomal-protein gene Rpl32. *Molecular and Cellular Biology, 9*(5), 2075-2082.

Clavel, S., Farout, L., Briand, M., Briand, Y., & Jouanel, P. (2002). Effect of endurance training and/or fish oil supplemented diet on cytoplasmic fatty acid binding protein in rat skeletal muscles and heart. *European Journal of Applied Physiology, 87*(3), 193-201.

Clayton, G. M., Peak-Chew, S. Y., Evans, R. M., & Schwabe, J. W. R. (2001). The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation. *Proceedings of the National Academy of Sciences of the United States of America, 98*(4), 1549-1554.

Escriva, H., Bertrand, S., & Laudet, V. (2004). The evolution of the nuclear receptor superfamily. *Essays in Biochemistry: Nuclear Receptor Superfamily, 40*, 11-26.

Felgner, P., Gadek, T., Holm, M., Roman, R., Chan, H., Wenz, M., Danielsen, M. (1987). Lipofection - a highly efficient, lipid-mediated dna-transfection procedure. *Proceedings of the National Academy of Sciences of the United States of America, 84*(21), 7413-7417.

Ferkovich, S. M., Dillard, C., & Oberlander, H. (1991). Stimulation of embryonic-development in microplitis-croceipes (braconidae) in cell-culture media preconditioned with a fat-body cell-line derived from a nonpermissive host, gypsy-moth, lymantria-dispar. *Archives of Insect Biochemistry and Physiology, 18*(3), 169-175.

Folli, C., Ramazzina, I., Percudani, R., & Berni, R. (2005). Ligand-binding specificity of an invertebrate (manduca sexta) putative cellular retinoic acid binding protein. *Biochimica Et Biophysica Acta-Proteins and Proteomics, 1747*(2), 229-237.

Fraenkel, G., & Blewett, M. (1947). Linoleic acid and arachidonic acid in the metabolism of 2 insects, ephestia-kuehniella (lep) and tenebrio-molitor (col). *Biochemical Journal, 41*(3), 475-478.

Fruchart, J. (2001). Peroxisome proliferator-activated receptor-alpha activation and high-density lipoprotein metabolism. *American Journal of Cardiology, 88*(12), 24N-29N.

Giguere V. (1999). Orphan nuclear receptors: From gene to function. *Endocrine Reviews, 20*(5), 689-725.

Giraudo, M., Califano, J., Hilliou, F., Tran, T., Taquet, N., Feyereisen, R., & Le Goff, G. (2011). Effects of hormone agonists on Sf9 cells, proliferation and cell cycle arrest. *Plos One, 6*(10), e25708.

Glatz, J., Van Breda, E., & Van der Vusse, G. (1998). Intracellular transport of fatty acids in muscle - role of cytoplasmic fatty acid-binding protein. *Skeletal Muscle Metabolism in Exercise and Diabetes, 441*, 207-218.

Graham, F. (1988). Citation classic - a new technique for the assay of infectivity of human adenovirus-5 dna. *Current Contents/clinical Medicine,* (46), 16-16.

Gronemeyer, H., Gustafsson, J., & Laudet, V. (2004). Principles for modulation of the nuclear receptor superfamily. *Nature Reviews Drug Discovery, 3*(11), 950-964.

Harmon, M., Boehm, M., Heyman, R., & Mangelsdorf, D. (1995). Activation of mammalian retinoid-X receptors by the insect growth-regulator methoprene. *Proceedings of the National Academy of Sciences of the United States of America, 92*(13), 6157-6160.

Haunerland, N. (1994). Fatty-acid-binding protein in locust and mammalian muscle - comparison of structure, function and regulation. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 109*(2-3), 199-208.

Haunerland, N. (1997). Transport and utilization in insect flight muscles. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 117*(4), 475-482.

Haunerland, N., & Chisholm, J. (1990). Fatty-acid binding-protein in flight-muscle of the locust, schistocerca-gregaria. Biochimica Et Biophysica Acta, 1047(3), 233-238.

Haunerland, N., & Spencer, F. (2004). Fatty acid-binding proteins - insights from genetic manipulations. *Progress in Lipid Research, 43*(4), 328-349.

Hegedus, D., Pfeifer, T., Hendry, J., Theilmann, D., & Grigliatti, T. (1998). A series of broad host range shuttle vectors for constitutive and inducible expression of heterologous proteins in insect cell lines. *Gene, 207*(2), 241-249.

Hewson, Q., Lovat, P., Pearson, A., & Redfern, C. (2002). Retinoid signalling and gene expression in neuroblastoma cells: RXR agonist and antagonist effects on CRABP-II and RAR beta expression. *Journal of Cellular Biochemistry, 87*(3), 284-291.

Huang, Z., Zhou, D., Zheng, S., Feng, Q., & Liu, L. (2011). A midgut-specific fatty acid-binding protein from spodoptera litura: Cloning, expression and localization analysis. Unpublished manuscript.

Huang, J., Xiong, Y., Li, T., Zhang, L., Zhang, Z., Zuo, B., Ren, Z. (2012). Ectopic overexpression of swine PPAR gamma 2 upregulated adipocyte genes expression and triacylglycerol in skeletal muscle of mice. *Transgenic Research, 21*(6), 1311-1318.

Hwang, S., Urizar, N., Moore, D., & Henning, S. (2002). Bile acids regulate the ontogenic expression of ileal bile acid binding protein in the rat via the farnesoid X receptor. *Gastroenterology, 122*(5), 1483-1492.

Jia, Z., Pei, Z., Maiguel, D., Toomer, C. J., & Watkins, P. A. (2007). The fatty acid transport protein (FATP) family: Very long chain acyl-CoA synthetases or solute carriers? *Journal of Molecular Neuroscience, 33*(1), 25-31.

Johnson, S. (1987). Migration and the life-history strategy of the fall armyworm, spodoptera-frugiperda in the western hemisphere. *Insect Science and its Application, 8*(4-6), 543-549.

Jones, G., & Sharp, P. A. (1997). Ultraspiracle: An invertebrate nuclear receptor for juvenile hormones. *Proceedings of the National Academy of Sciences of the United States of America, 94*(25), 13499-13503.

Joshi, R., & Schoenbach, K. (2000). Electroporation dynamics in biological cells subjected to ultrafast electrical pulses: A numerical simulation study. *Physical Review E, 62*(1), 1025-1033.

Juge-Aubry, C., Hammar, E., Siegrist-Kaiser, C., Pernin, A., Takeshita, A., Chin, W., Meier, C. (1999). Regulation of the transcriptional activity of the peroxisome proliferator-activated receptor alpha by phosphorylation of a dependent trans-activating domain. *Journal of Biological Chemistry, 274*(15), 10505-10510.

Kammer, A. E., & Heinrich, B. (1978). Insect flight metabolism. *Advances in Insect Physiology, 13*, 133-228.

Kariuki, C., McIntosh, A., & Goodman, C. (2000). In vitro host range studies with a new baculovirus isolate from the diamondback moth plutella xylostella (L.) (plutellidae : Lepidoptera). *In Vitro Cellular & Developmental Biology-Animal, 36*(4), 271-276.

Kazemi, M., McDonald, C., Shigenaga, J., Grunfeld, C., & Feingold, K. (2005). Adipocyte fatty acid-binding protein expression and lipid accumulation are increased during activation of murine macrophages by toll-like receptor agonists. *Arteriosclerosis Thrombosis and Vascular Biology, 25*(6), 1220-1224.

Khani, A., Moharramipour, S., Barzegar, M., & Naderi-Manesh, H. (2007). Comparison of fatty acid composition in total lipid of diapause and non-diapause larvae of cydia pomonella (lepidoptera : Tortricidae). *Insect Science, 14*(2), 125-131.

Kruse, S. W., Suino-Powell, K., Zhou, X. E., Kretschman, J. E., Reynolds, R., Vonrhein, C., Xu, H. E. (2008). Identification of COUP-TFII orphan nuclear receptor as a retinoic acid-activated receptor. *Plos Biology, 6*(9), 2002-2015.

Kukalova-Peck, J. (1991). *Fossil history and the evolution of hexapod structures.* Melbourne University Press, Melbourne, Australia & UCL Press Limited, London, England.

Laudet V. (1997). Evolution of the nuclear receptor superfamily: Early diversification from an ancestral orphan receptor. *Journal of Molecular Endocrinology, 19*(3), 207-226.

Livak, K., & Schmittgen, T. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-delta delta C) method. *Methods, 25*(4), 402-408.

Lofstedt, C., Elmfors, A., Sjogren, M., & Wijk, E. (1986). Confirmation of sex-pheromone biosynthesis from (16-D3)palmitic acid in the turnip moth using capillary gas-chromatography. *Experientia, 42*(9), 1059-1061.

Logette, E., Solary, E., & Corcos, L. (2005). Identification of a functional DNA binding site for the SREBP-1c transcription factor in the first intron of the human caspase-2 gene. *Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids, 1738*(1-3), 1-5.

Maess, M. B., Sendelbach, S., & Lorkowski, S. (2010). Selection of reliable reference genes during THP-1 monocyte differentiation into macrophages. *Bmc Molecular Biology, 11*, 90.

Maki, A., Sawatsubashi, S., Ito, S., Shirode, Y., Suzuki, E., Zhao, Y., Kato, S. (2004). Juvenile hormones antagonize ecdysone actions through co-repressor recruitment to EcR/USP heterodimers. *Biochemical and Biophysical Research Communications, 320*(1), 262-267.

Mansfield, S., Cammer, S., Alexander, S., Muehleisen, D., Gray, R., Tropsha, A., & Bollenbacher, W. (1998). Molecular cloning and characterization of an invertebrate cellular retinoic acid binding protein. *Proceedings of the National Academy of Sciences of the United States of America, 95*(12), 6825-6830.

Maatman, R., Degano, M., Vanmoerkerk, H., Vanmarrewijk, W., Vanderhorst, D., Sacchettini, J., & Veerkamp, J. (1994). Primary structure and binding characteristics of locust and human muscle fatty-acid-binding proteins. *European Journal of Biochemistry, 221*(2), 801-810.

Moabbi, A. M., Agarwal, N., El Kaderi, B., & Ansari, A. (2012). Role for gene looping in intron-mediated enhancement of transcription. *Proceedings of the National Academy of Sciences of the United States of America, 109*(22), 8505-8510.

Montoliu, L., Blendy, J., Cole, T., & Schutz, G. (1995). Analysis of perinatal gene-expression - hormone response elements mediate activation of a lacz reporter gene in liver of transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America, 92*(10), 4244-4248.

Motojima, K., Passilly, P., Peters, J., Gonzalez, F., & Latruffe, N. (1998). Expression of putative fatty acid transporter genes are regulated by peroxisome proliferator-activated receptor alpha and gamma activators in a tissue- and inducer-specific manner. *Journal of Biological Chemistry, 273*(27), 16710-16714.

Murphy, E. (1998). L-FABP and I-FABP expression increase NBD-stearate uptake and cytoplasmic diffusion in L cells. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 275*(2), G244-G249.

Negre, V., Hotelier, T., Volkoff, A., Gimenez, S., Cousserans, F., Mita, K., Fournier, P. (2006). SPODOBASE : An EST database for the lepidopteran crop pest spodoptera. *Bmc Bioinformatics, 7*, 322.

Neumann, E., Schaeferridder, M., Wang, Y., & Hofschneider, P. (1982). Gene-transfer into mouse lyoma cells by electroporation in high electric-fields. *Embo Journal, 1*(7), 841-845.

Owen GI, Z. A. (2000). Origins and evolutionary diversification of the nuclear receptor superfamily. *Cellular and Molecular Life Sciences, 57*(5), 809-827.

Perry, R., Hariharan, N., Chung, S., Kelley, D., & Safrany, G. (1992). The importance of downstream and initiator elements for the tata box-lacking promoters of mammalian ribosomal-protein genes. *Faseb Journal, 6*(1), A274-A274.

Pfaffl, M., Tichopad, A., Prgomet, C., & Neuvians, T. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - excel-based tool using pair-wise correlations. *Biotechnology Letters, 26*(6), 509-515.

Pfister, C., Tatabiga, M., & Roser, F. (2011). Selection of suitable reference genes for quantitative real-time polymerase chain reaction in human meningiomas and arachnoidea. *BMC Research Notes, 4*, 275.

Poirier, H., Braissant, O., Niot, I., Wahli, W., & Besnard, P. (1997). 9-cis-retinoic acid enhances fatty acid-induced expression of the liver fatty acid-binding protein gene. *FEBS Letters, 412*(3), 480-484.

Portales-Casamar, E., Thongjuea, S., Kwon, A. T., Arenillas, D., Zhao, X., Valen, E., Sandelin, A. (2010). JASPAR 2010: The greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Research, 38*, D105-D110.

Prendergast, F., & Mann, K. (1978). Chemical and physical-properties of aequorin and green fluorescent protein isolated from aequorea-forskalea. *Biochemistry, 17*(17), 3448-3453.

Qu, H., Cui, L., Rickers-Haunerland, J., & Haunerland, N. H. (2007). Fatty acid-dependent expression of the muscle FABP gene - comparative analysis of gene control in functionally related, but evolutionary distant animal systems. *Molecular and Cellular Biochemistry, 299*(1-2), 45-53.

Ramakers, C., Ruijter, J., Deprez, R., & Moorman, A. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters, 339*(1), 62-66.

Renaud, J., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., & Moras, D. (1995). Crystal-structure of the rar-gamma ligand-binding domain bound to all-trans-retinoic acid. *Nature, 378*(6558), 681-689.

Roelofs, W., & Wolf, W. (1988). Pheromone biosynthesis in lepidoptera. *Journal of Chemical Ecology, 14*(11), 2019-2031.

RuizLozano, P., Merki, E., Doevendans, P., Zhou, M., & Chien, K. (1997). Rxr homodimer and Rxr/ppar heterodimer pathways negatively regulate an atrial marker, mlc-2a, at the transcriptional level in ventricular muscle cells. *Circulation, 96*(8), 1327-1327.

Safari, M. G., & Hosseinkhani, S. (2013). Lipid composition of cationic nanoliposomes implicate on transfection efficiency. *Journal of Liposome Research, 23*(3), 174-186.

Schleicher, C., Cordoba, O., Santome, J., & DellAngelica, E. (1995). Molecular evolution of the multigene family of intracellular lipid-binding proteins. *Biochemistry and Molecular Biology International, 36*(5), 1117-1125.

Schroeder, S., Bertram, B., & von Holst, A. (2013). High efficiency transfection and survival rates of oligodendrocyte precursor cells achieved by electroporation. *Glia, 61*, S149-S150.

Smith, A., Tsuchida, K., Hanneman, E., Suzuki, T., & Wells, M. (1992). Isolation, characterization, and cdna sequence of 2 fatty acid-binding proteins from the midgut of manduca-sexta larvae. *Journal of Biological Chemistry, 267*(1), 380-384.

Strube, C., Buschbaum, S., Wolken, S., & Schnieder, T. (2008). Evaluation of reference genes for quantitative real-time PCR to investigate protein disulfide isomerase transcription pattern in the bovine lungworm dictyocaulus viviparus. *Gene, 425*(1-2), 36-43.

Sun, Y., Li, X., Liang, X., Wan, Z., & Duan, Y. (2013). Calcium Phosphate/Octadecyl-quatemized carboxymethyl chitosan nanoparticles: An efficient and promising carrier for gene transfection. *Journal of Nanoscience and Nanotechnology, 13*(8), 5260-5266.

Tarter, M., Capaldi, S., Carrizo, M. E., Ambrosi, E., Perduca, M., & Monaco, H. L. (2008). Crystal structure of human cellular retinol-binding protein 11 to 1.2 angstrom resolution. *Proteins-Structure Function and Bioinformatics, 70*(4), 1626-1630.

Theilmann, D., & Stewart, S. (1992). Molecular analysis of the transactivating ie-2 gene of orgyia-pseudotsugata multicapsid nuclear polyhedrosis-virus. *Virology, 187*(1), 84-96.

Van der Horst, D., Van Marrewijk, W., & Diederen, J. (2001). Adipokinetic hormones of insect: Release, signal transduction, and responses. *International Review of Cytology - a Survey of Cell Biology, Vol 211, 211*, 179-240.

Van der Vusse, G., Glatz, J., Van Nieuwenhoven, F., Reneman, R., & Bassingthwaighte, J. (1998). Transport of long-chain fatty acids across the muscular endothelium. *Skeletal Muscle Metabolism in Exercise and Diabetes, 441*, 181-191.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology, 3*(7), RESEARH0034

Vaughn, J., Goodwin, R., Tompkins, G., & Mccawley, P. (1977). Establishment of 2 cell lines from insect spodoptera-frugiperda (lepidoptera-noctuidae). *In Vitro-Journal of the Tissue Culture Association, 13*(4), 213-217.

Vida, M., Serrano, A., Romero-Cuevas, M., Pavon, F. J., Gonzalez-Rodriguez, A., Gavito, A. L., Baixeras, E. (2013). IL-6 cooperates with peroxisome proliferator-activated receptor-alpha-ligands to induce liver fatty acid binding protein (LFABP) up-regulation. *Liver International, 33*(7), 1019-1028.

Weers, P., & Ryan, R. (2003). Apolipophorin III: A lipid-triggered molecular switch. *Insect Biochemistry and Molecular Biology, 33*(12), 1249-1260.

Weisiger, R. (1996). Cytoplasmic transport of lipids: Role of binding proteins. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 115*(3), 319-331.

Wiesenberg, I., Missbach, M., Kahlen, J., Schrader, M., & Carlberg, C. (1995). Transcriptional activation of the nuclear receptor rzr-alpha by the pineal-gland hormone melatonin and identification of cgp-52608 as a synthetic ligand. *Nucleic Acids Research, 23*(3), 327-333.

Wu, W., Niles, E. G., Hirai, H., & LoVerde, P. T. (2007). Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. *BMC Evolutionary Biology, 7*, 15. Retrieved from http://www.biomedcentral.com.proxy.lib.sfu.ca/1471-2148/; http://www.biomedcentral.com.proxy.lib.sfu.ca/content/pdf/1471-2148-7-27.pdf

Zheng, C., Xu, G., & Wang, D. (2012). Molecular cloning and tissue-specific expression of the heart-type fatty acid-binding protein (H-FABP). *Materials for Environmental Protection and Energy Application, Pts 1 and 2, 343-344*, 438-447.

Zheng, L., Roeder, R., & Luo, Y. (2003). S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell, 114*(2), 255-266.

Zhan, S., Merlin, C., Boore, J. L., & Reppert, S. M. (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell, 147*(5), 1171-1185.

Zhu, Y., Machleder, E., Chenchik, A., Li, R., & Siebert, P. (2001). Reverse transcriptase template switching: A SMART (TM) approach for full-length cDNA library construction. *Biotechniques, 30*(4), 892-897.

Appendices

Appendix A.

EST sequences of iLBP and housekeeping genes in *Spodoptera frugiperda*

fabp1 (Accession: Sf2M09275-5-1)

```
1     CGTTAAACGGAAGAAGAGTCAATAGGCCAAAATGTCTTTCCTCAACAAGAACTTCAAATTCGTGAAGCAGGAGAACTTCGATGGAT

                      FABP1-F
                   ─────────────────▶
87    TTCTCCGTGCTGTCGGTTTACCCGAGGAGAAGATTGCTCAGATCCTGAAGTACTCTCCTGACCAGAAGCTGGTGCAGGACGGTGA

                              FABP1-R
                   ◀─────────────────
172   CAGCTACACCTACATCACCAACGGCCCCGAGGGTGTCAAGGAGATCAAGTTCAAGTCCGGTGTGGAATTCGAAGATGTTATTGGC

257   ACTGACAAAGTTCCCATCAAGACCACGTACACAGTCGACGGCAACGTAGTGACACAGAACATCTCAGCTCCCCAAGGCACAGCCG

342   TCTTCAAGAGGGAATACAATGGCGACGAACTCACCGTGGTCATCACCGGTGACAAATTCGACGGTGTTGCCAAGAGGTACTACAA

427   GGCTGAATAAACTCATAGCTATTAGTAATTACGTTAATTTATTTTTTATAAATAATAAATATTATTATTAATTAAAAAAAAAAAAAAAAAA

517   AAAAAAAAAAAAAAAAAAAGGGG
```

fabp2 (Accession: Sf2M02196-5-1)

```
1     CTAAAAACTAACCGTGAAAAATGGCTTTCCTTAACAAGACCTACAAGTTTGTGAGCCAGGAGAACTTCGATGGTTTCCTCAAGGCA

86    TCTGGTGTCCCCGATGACAAGATAGAAAAGACTCAGAGCTACACACCAGACCAGAAGATCACTAAGGATGGTGACACCTACACTT

172   ACATCGTGTCTGGTGCTTTAGGCACAAAGGAAGTCAAATTCCAGTCTGGAGTTGAATTCGATGACAAACTTGGCGCTGAACAGACT

                          FABP2-F
                   ─────────────────▶
257   CCCGTCAAGAGCACAATCGTCGTTGACGGCAACACAGTTACTCAGACCGCTAAGGGAGCCACCGGAGTTGCTACCTTCAAGAGA

                              FABP2-R
                   ◀─────────────────
341   GAGTACAACGGAGACGATTTGGTTGTGACCATCACCATGGACAATTGGAACGGATCCGCCAAGAGGTACTACAAGGCTGCATGAA

426   CAAACTAACCAACAAACGAATAGTCCATAATTTAATTTATTTTTTGTAATAAAATCGAAATTAACACCTAAAAAAAAAAAAAAAAAAAAAAA

515   AAAAAAAAAAAA
```

fabp3 (Accession: Sf1F10466-5-1)

```
1    GCAGTTTGCTATTCGAAGTAACCTAACCTACCAAAATGTCTTTCCTAGGCAAGGAATACAAATTTGAGAGGCAGGAGAACTTTGAG

                                    FABP3-F →
87   GAATTCGTGAACTCTTTGGGTCTTCCCGCAGAGCAGACTCAAGGGTACCTCAACTACAACCCTGTACTGAAGTACACCAAGAATG

                                                    ← FABP3-R
172  GAGACATGTACACCCTGACTTCCGTCACTGCTCAGGGCACGAAGGAGGTCTCCTTCAAGTCTGGAGTCGCCTTCGATGAGACTGT

257  TGCTGGCAAGAAGGTCAACACCACTTACACCGTTGACGGTGACACCATCACACAGGTCCAGAAGTCTGATGATGGAGTCCTCACC

342  ATCACCAGGACTTTCTCTGGAAATGAGTTGGTTGTGACCCTGAAGACCAACAAATGGGACGGTGTTGCTACTAGATACTACAAAGT

428  TGTTTAATTTATTTTATTGTACCTCAAAAGTATTATTATATTTATTACTTGAATTTATTACTGTTATTTAAANNNTTTAATTTTACCC
```

crabp (Accession: Sf2L00595-5-1)

```
1    CCCACGCGTCCGCCCACGCGTCCGGACCGGTAGTACGGACTCACACATTTTTAAATCACTTTAAAAAAAAACATTTTTATATTTTAA

88   AAATGGAATTTGTCGGCAAGAAATACAAGATGGTCTCTTCGGAGAACTTCGACGAGTTCATGAAGGCTATCGGCGTTGGTCTCATC

                                    CRABP-F →
174  ACCCGCAAGGCTGCTAACGCCGTCACTCCGACGGTGGAACTTCGCAAGGAGGGAGACGTCTACAACCTAGTGACGTCATCCACC

                                                    ← CRABP-R
258  TTCAAGACCACTGAGGTCAAGTTCAAGCCTGGTGAGGAGTTCGATGAGGACCGAGCTGATGGTGCCAAGGTGAAGTCAGTGTGC

342  ACCTTCGAGGGTAACACCTTGAAGCAGATCCAGAAGGCTGCCGATGGCATGGAAGTCAGCTACGTCAGAGAATTCGGACCCGAA

426  GAGATGAAGGCTGTGATGACAGCTAAGGATGTGACCTGCACCAGGGTGTACAAGGTCCAATAATCTGCTTCTAAGACTGCGGGGC

511  AGAGCCCTACCTGTGTCTATACTGCACAACTACTATAACTCTATAAATTATGTCGTTAGGCTACTAGTATCTACGTTCGTTCGA
```

rpl37a (Accession: Sf9L07692)

```
1    GGTTTTTTCTTTTCTGTCAACCGGAGAGTTAGACAGGTTATCAAAAATGGCCAAACGCACGAAGAAGGTCGGAATTACTGGCAAAT

                            RPL37A-F →
87   ATGGCACACGTTACGGTGCCTCCCTCCGTAAAATGGTCAAAAAAATGGAAGTCACCCAGCACGCAAAATACACTTGCTCGTTCTGC

                                            ← RPL37A-R
173  GGTAAGGATGCCATGAAGCGTTCCTGTGTCGGCATCTGGTCATGCAAGCGCTGTAAGAGGACCGTAGCTGGTGGTGCGTGGGTA

257  TTCTCCACCACCGCTGCCTCCTCCTGCAGGTCCGCAGTCAGAAGATTGCGTGAGGTCAAGTAAACATCTGACGTTAGCTTAAGGA

342  AATAAACAAAAACAACC
```

gapdh (Accession: Sf9L06272)

```
1    GGAAAACTGTGGCGTGATGGCCGTGGTGCCCAGCAGAACATCATTCCCGCCGCCACCGGCGCTGCTAAAGCCGTCGGCAAAGT
```

GAPDH-F →

```
84   CATTCCTGCTCTGAATGGAAAGCTGACTGGTATGGCTTTCCGTGTTCCCGTTGCTAACGTCTCGGTCGTCGACCTGACTGTTCGTC
```

← GAPDH-R

```
170  TTGGCAAGCCCGCCAGCTACGATGCCATCAAGCAGAAGGTCAAGGAGGCCGCCCAAGGACCCCTGAAGGGCATTCTTGACTACA

254  CCGAGGAACAGGTCGTGTCATCCGACTTCATCGGTGACAACCACTCATCTATCTTCGATGCTGCTGCCGGTATCTCTCTGAACGA

349  CAACTTCGTCAAGCTCATCAGCTGGTATGACAACGAGTACGGTTACTCCAACCGCGTTATCGATCTCATCAAGTACATCCAGACCA

435  AGGATTAAATGTTAGATTGCATTGTAGAGATAAATGTTGTTATGAATGTTTATGAATATGATTTTGTCGACGATATCGACTTTTATAAT

524  TTAAGCTTAAGTATAGCGTTCAAGTCTAAATTATTATTTAACCACTTTTGAAAGTGATGTTATACGTGTAATAATG
```

Blue arrows indicate locations of primers.

Appendix B.

Optimization of Real-time PCR
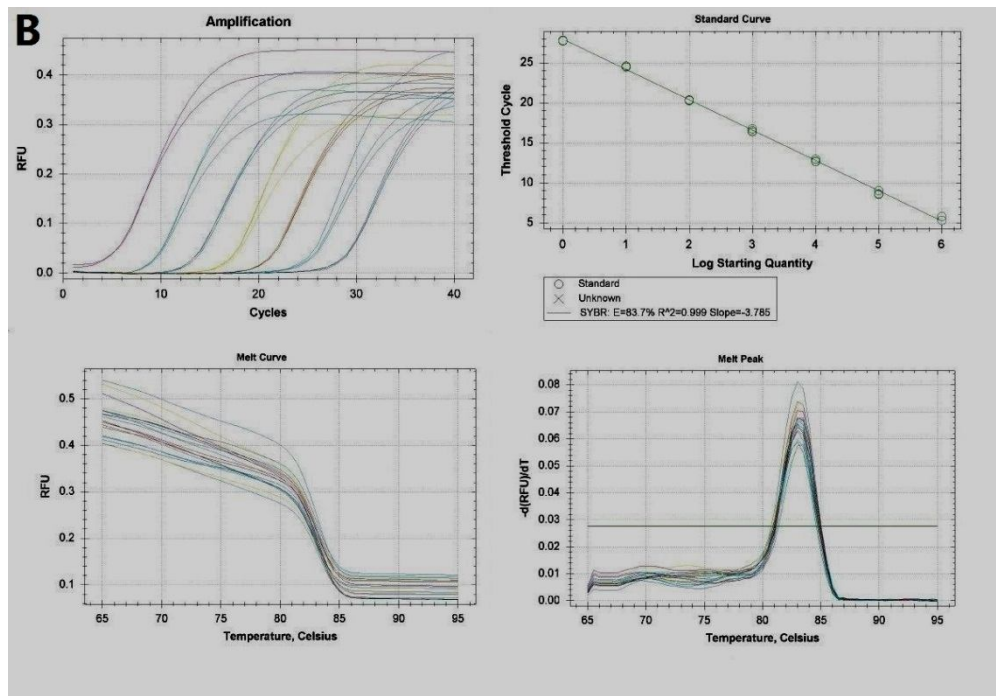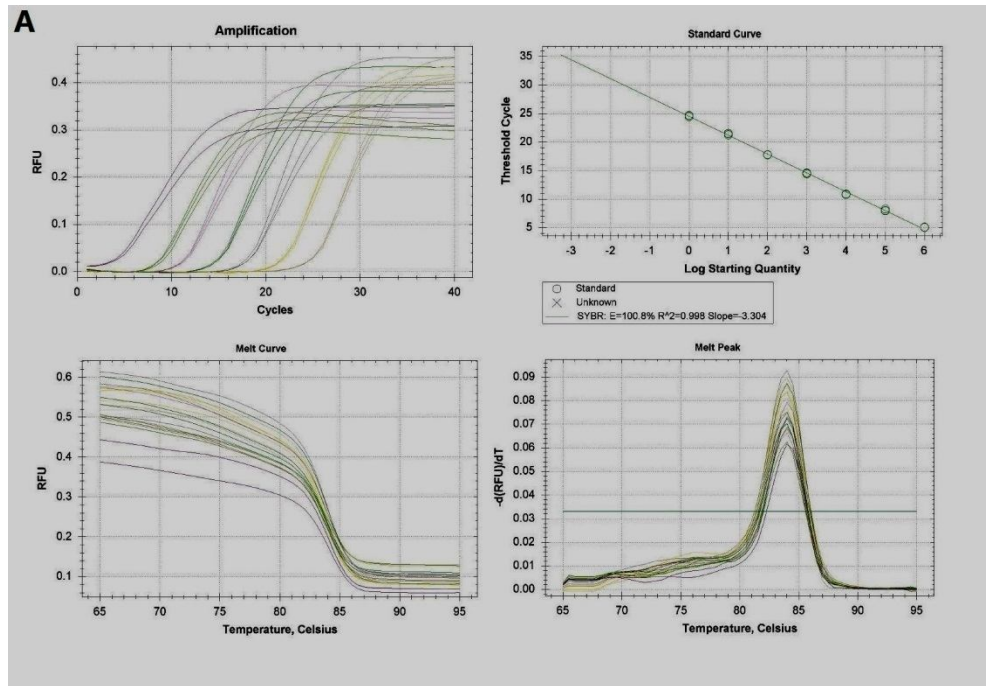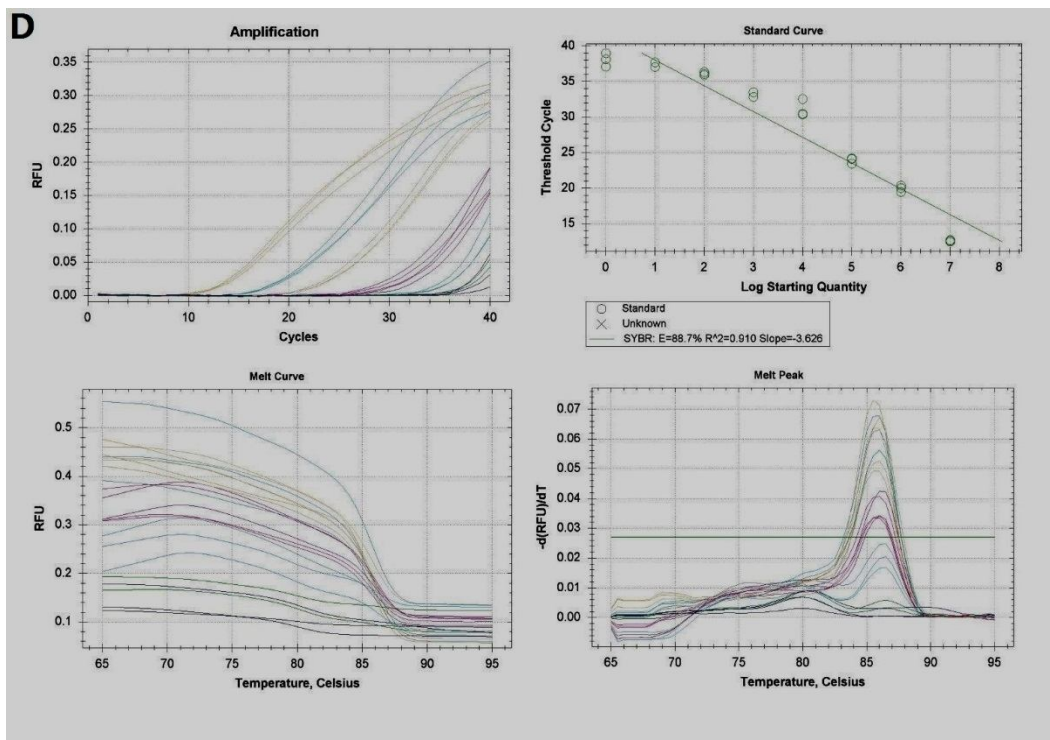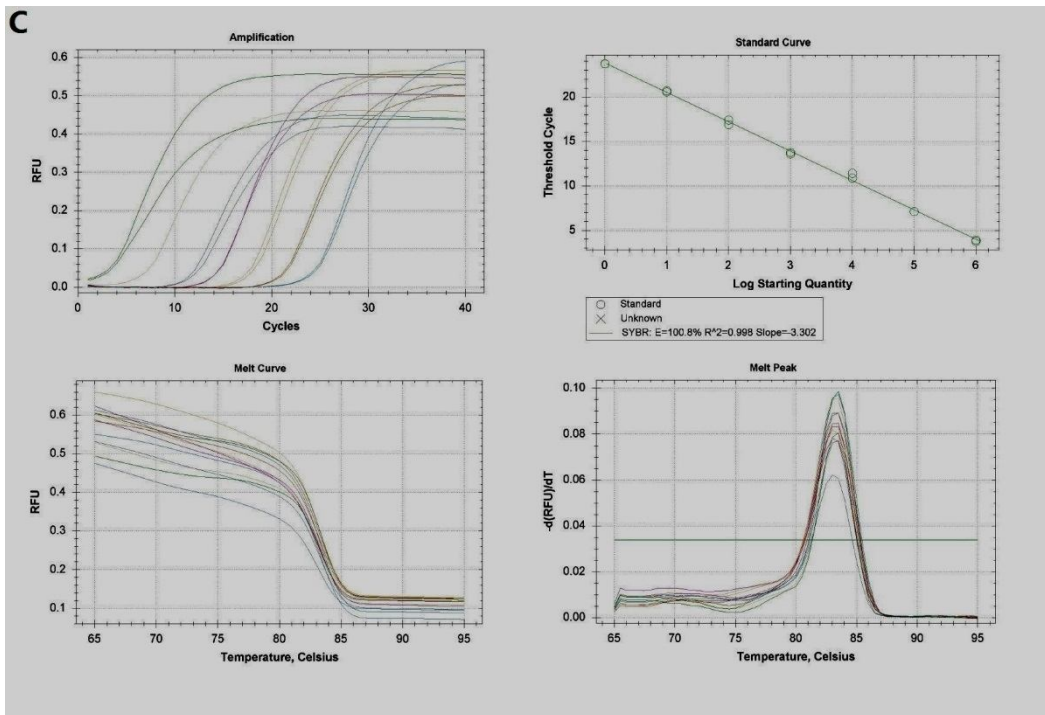
A: CRABP. B: FABP3. C: RPL37A. D: GAPDH.

Appendix C.

Real-time PCR data from flight experiment

Group	Subject ID	Ct_{CRABP}	Ct_{FABP3}	Ct_{RPL37A}
#1: Control	01	18.36	20.92	15.78
	02	18.89	21.77	16.06
	03	21.53	21.73	17.92
	04	16.56	19.64	15.65
#2: 4 hours flight 0 hour rest	05	16.44	20.19	15.81
	06	17.97	19.84	16.52
	07	17.43	19.82	15.91
	08	20.88	21.24	18.33
#3: 4 hours flight 2 hours rest	09	15.29	20.00	16.07
	10	16.00	22.19	15.26
	11	17.19	22.50	15.35
	12	15.81	20.47	15.31
#4: 4 hours flight 4 hours rest	13	16.74	20.26	15.48
	14	16.11	20.09	15.51
	15	15.69	20.25	16.08
	16	16.58	20.38	15.54

Appendix D.

Real-time PCR data from fatty acid treatment

Treatment	Ct_{CRABP}	Ct_{FABP3}	Ct_{RPL37A}
Control	17.29	13.38	13.38
	17.57	14.16	14.16
	18.33	13.53	13.53
Oleic acid	17.76	13.76	13.76
	17.74	13.17	13.17
	17.54	13.87	13.87
Arachidonic acid	17.77	14.01	14.01
	18.29	14.27	14.27
	17.00	13.86	13.86
Palmitic acid	17.18	13.62	13.62
	17.54	14.01	14.01
	17.17	14.00	14.00
Linoleic acid	17.17	13.56	13.56