Pergamon

Insect Biochemistry and Molecular Biology 31 (2001) 553–562

Insect Biochemistry and Molecular Biology

www.elsevier.com/locate/ibmb

Cloning and sequence of the gene encoding the muscle fatty acid binding protein from the desert locust, *Schistocerca gregaria*

Qiwei Wu, Peter Andolfatto [1], Norbert H. Haunerland [*]

Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

Received 25 May 2000; received in revised form 18 September 2000; accepted 18 September 2000

Abstract

Muscle fatty acid binding protein (FABP) is a major cytosolic protein in flight muscle of the desert locust, *Schistocerca gregaria*. FABP expression varies greatly during development and periods of increased fatty acid utilization, but the molecular mechanisms that regulate its expression are not known. In this study, the gene coding for locust muscle FABP was amplified by PCR and cloned, together with 1.2 kb of upstream sequence. The sequence coding for the 607 bp cDNA is interrupted by two introns of 12.7 and 2.9 kb, inserted in analogous positions as the first and third intron of the mammalian homologues. Both introns contain repetitive sequences also found in other locust genes, and the second intron contains a GT-microsatellite. The promoter sequence includes a canonical TATA box 24 bp upstream of the transcription start site. The upstream sequence contains various potential myocyte enhancer sequences and a 160 bp segment that is repeated three times. In database searches in the genome database of *Drosophila melanogaster*, a gene with the same gene organization and promoter structure was identified, likely the dipteran homologue of muscle FABP. Upstream of both insect genes, a conserved 19 bp inverted repeat sequence was detected. A similar but reverse palindrome is also present upstream of all mammalian heart FABP genes, possibly representing a novel element involved in muscle FABP expression.

Keywords: FABP; Muscle; Locust; *Drosophila melanogaster*; *Schistocerca gregaria*; Regulatory elements; Repetitive elements; Microsatellite

1. Introduction

Fatty acid binding proteins (FABPs) are ubiquitous 15 kDa proteins believed to be involved in intracellular fatty acid transport and metabolism (Veerkamp et al., 1991). These proteins belong to a conserved multi-gene superfamily of binding proteins for fatty acids and related hydrophobic ligands, such as retinoic acid or retinol. FABPs form a typical beta-barrel structure that encloses the binding site for their ligands (Banaszak et al., 1994), which are bound by both hydrophobic and polar interactions. The expression of these proteins is tissue-specific and appears to be subject to different regulatory influences. Based on their tissue distribution, at least eight different FABPs have been identified in mammals, each encoded by a separate gene (Veerkamp and Maatman, 1995). Despite their different regulation, the general organization of all mammalian FABP genes is identical. The coding sequence is always interrupted by three introns of varying sizes, inserted in analogous positions, suggesting that these introns are of ancient evolutionary origin (Matarese et al., 1989).

Related FABPs have been found in lower vertebrates and in invertebrates. Homologous muscle FABPs (abbreviated H-FABP or M-FABP) have been identified in cardiac and skeletal muscles not only of various mammalian species, but also in birds (Guglielmo et al., 1998), fish (Ando et al., 1997), and insects (Haunerland, 1994). In spite of the evolutionary distance between these species, muscle FABP is highly conserved in its amino acid sequence and its three-dimensional structure. Moreover, a growing body of evidence suggests that the expression

Abbreviations: FABP, fatty acid binding protein; H-FABP, cardiac fatty acid binding protein; M-FABP, muscle fatty acid binding protein; PPAR, peroxisome proliferator activated receptor; RACE, rapid amplification of cDNA ends. The sequences have been deposited at Genbank (Accession numbers: AF244980, AF244981).

* Corresponding author. Tel.: +1-604-291-3540; fax: +1-604-291-4312.

E-mail address: haunerla@sfu.ca (N.H. Haunerland).

[1] Current address: Institute of Cell, Animal, Population Biology, University of Edinburgh, Edinburgh, UK.

0965-1748/01/$ - see front matter
PII: S0965-1748(00)00158-2

of these genes is controlled by similar mechanisms as well. This includes the observation that muscle FABP increases markedly during muscle differentiation in vertebrates and invertebrates (van Nieuwenhoven et al., 1995; Haunerland et al., 1992). Generally, the concentrations of FABP appear to be proportional to the rate of fatty acid oxidation encountered in various differentiated muscles (Haunerland, 1994): vertebrate skeletal muscles that are fueled mostly by carbohydrates contain only small amounts of the protein, while FABP is a major cytosolic component in fatty acid-dependent muscles, such as cardiac muscle. This trend extends to other classes and phyla: the flight muscle of a migratory shorebird that encounters metabolic rates twice as high as the mammalian heart also possesses approximately twice as much FABP (Guglielmo et al., 1998), and three-fold higher metabolic rates and FABP concentrations were measured in locust flight muscle (Haunerland et al., 1992). FABP expression can be further stimulated by increased lipid-dependent metabolism (endurance training (van Breda et al., 1992), extended flight (Chen and Haunerland, 1994) or fatty acid exposure in vivo and in vitro (Chen and Haunerland, 1994; van der Lee et al., 2000).

These similarities in FABP expression suggest gene control mechanisms that are conserved between the various species. To date, however, the gene for muscle FABP has been cloned only from four mammalian species, namely human (Phelan et al., 1996), mouse (Treuner et al., 1994), rat (Zhang et al., 1999), and pig (Gerbens et al., 1997). Here we report the cloning, sequencing, and analysis of the muscle FABP gene from the desert locust, *Schistocerca gregaria*.

2. Materials and methods

2.1. Genomic digestion and Southern blotting

S. gregaria were reared in crowded conditions at 32°C under continuous light. For the isolation of genomic DNA, freshly excised flight muscle tissue from adult locusts was minced and powdered under liquid nitrogen. Genomic DNA was isolated by a method modified from Jowett (1986). For Southern blots, approximately 100 μg of high molecular weight genomic DNA was digested with 150 units of *Eco*RI, *Sst*I, *Xho*I, *Xba*I or *Bam*HI (Amersham Pharmacia Biotech) at 37°C for 4–6 h. The digested DNA was ethanol-precipitated and separated on a 0.6% agarose gel. DNA was nicked, denatured and transferred to a Genescreen Nylon membrane (NEN-Dupont) by unidirectional capillary blotting. The membranes were UV-crosslinked (UV Stratalinker, Stratagene) and prehybridized in 5×SSPE (1×SSPE=125 mM $NaCl_2$, 10 mM NaH_2PO_4, 1 mM EDTA, pH7.4), 5×Denhart's solution (1×Denhard's solution=0.1% Ficoll-400, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin) and 0.3% SDS for at least 4 h prior to the addition of the labeled probe. Hybridization cocktails typically contained $1.0–2.0\times10^6$ cpm/ml. Hybridizations were carried for 12 h at 65–68°C. After hybridization, the membrane was washed once with 1×SSPE/1% SDS and twice with 0.1×SSPE/0.3% SDS at 65 or 68°C for 30 min. The membrane was then exposed to Kodak XK-1 film with intensifying screens at −80°C for up to 1 week.

2.2. Probes

For probe preparations, approximately 3 μg of locust FABP plasmid (Price et al., 1992) was double digested with *Eco*RI and *Sst*I restriction endonucleases (Amersham Pharmacia Biotech) yielding a 600 bp fragment which contained all of the FABP cDNA sequence with minimal flanking plasmid sequence. Approximately 50 ng of gel purified probe was labeled by Nick Translation (BRL) with 60 μCi of [α-^{32}P]ATP (3000 mCi/mmol, Amersham Pharmacia Biotech). The labeled probe was purified on a small Sephadex G-50 gel filtration spin-column (Amersham Pharmacia Biotech). Specific activities of the probe were typically $0.5–1.0\times10^9$ cpm/μg. For non-radioactive library screening, the probes were labeled with Digoxigenin-11-dUTP (Boehringer Mannheim).

2.3. λZAP library construction and screening

Genomic DNA (150 μg) was digested with *Eco*RI and separated on a 0.5% agarose gel. Southern blotting revealed a single 10 kb band after *Eco*RI digestion. Fragments in the 9.5–10.5 kb range were eluted from the gel and used to construct a sub-genomic library in λZAP II, (Gigapack cloning system, Stratagene). The library was screened with a digoxigenin labeled FABP cDNA probe. The presence of the FABP gene in λ plate lysates was confirmed by PCR with primers that anneal in the second exon of the FABP gene. A 10 kb clone containing exon 2 and 3 but not the 5′-end of the gene (FABP 6.1) was placed in Bluescript KS, restriction mapped and sequenced.

2.4. Long template PCR amplification of intron I

PCR was performed by using the Expand Long Template PCR Kit (Boehringer Mannheim) in the GeneAmp 2400 PCR System (Perkin Elmer). The reaction consisted of 250 ng of genomic DNA, 1×PCR buffer with 2.25 mM $MgCl_2$, 500 μM dNTP, 300 nM of each primer, and 2.5 units of polymerase mix. The PCR conditions were: 2 min at 94°C, 30 cycles of 10 s at 94°C, 30 s at 63°C, and 8–15 min extension at 68°C, followed by a final 10 min period at 68°C. The extension time was kept

at 8 min for the first 10 cycles, and then increased by 20 s per cycle. The 31 bp upper primer U2, selected from the 5′-end of the cDNA, was modified to carry an *Xba*I restriction site at its 5′-end; the lower primer L2 was chosen to anneal within the first intron, 1.7 kb upstream of the second exon (see Fig. 1).

2.5. Construction and screening of adapter-ligated genomic DNA libraries

For the cloning of the 5′-flanking region of locust FABP gene, five adapter-ligated locust genomic DNA libraries were constructed (Universal GenomeWalker Kit, Clontech). Aliquots of genomic DNA were digested with *Dra*I, *Eco*RV, *Pvu*II, *Sca*I or *Stu*I (New England Biolabs). The five digested DNA preparations were ligated to oligonucleotide adapter primers containing compatible ends for the above five enzymes. Since all the digested genomic DNA fragments were blunt ended, a common double-stranded adapter was formed by annealing the long oligonucleotide 5′-GTA ATA CGA CTC ACT ATA TAG GGC ACG CGT GGT CGA CGG CCC GGG CTG GT-3′ with the short oligonucleotide 5′-PO_4-ACC AGC CC-NH_2-3′. The 3′-NH_2 group was added to block the nonspecific extension of the adapter strand.

The 5′-flanking region of the locust FABP gene was obtained using the Universal GenomeWalker Kit, which involved a primary and a secondary PCR reaction. The upper, adapter-specific PCR primers were AP1: 5′-GTA ATA CGA CTC ACT ATA GGG C-3′ (primary) and AP2: 5′-ACT ATA GGG CAC GCG TGG T-3′ (secondary). For the first walk (*Dra*I library), the lower, gene specific PCR primers L4 and the nested primer L5 were used; for the second walk (*Stu*I library) L6 and L7 (primer sequences see Fig. 1). The PCR reaction mixtures in the primary amplification contained 10 ng of adapter-ligated genomic DNA, 10 pmol each of forward and reverse primers, 2.5 units of Advantage genomic polymerase mix (Clontech), 10 mM of dNTP and 5% DMSO. A 50-fold dilution of the primary PCR products was used as the DNA template for the secondary PCR. PCR was carried out on a GeneAmp PCR system 2400 (Perkin Elmer) with a touch-down protocol (seven cycles of: 2 s at 94°C, 2 min annealing and extension from 72°C down to 67°C, followed by 32 cycles of: 2 s at 94°C, 2 min annealing and extension at 67°C, and a final 4 min extension period at 67°C. The final PCR products were purified by agarose gel electrophoresis, and subcloned into the pCR2.1 vector (Invitrogen). To confirm that the sequence obtained from the genomic walking reactions are located immediately upstream of the FABP gene, a PCR product spanning from 1.1 kb upstream of the promoter into the first intron was generated from genomic DNA and sequenced (primers U9 and L9, see Fig. 1).

2.6. 5′-RACE

RACE-PCR was performed using the SMART RACE cDNA Amplification Kit (Clontech). Total RNA from

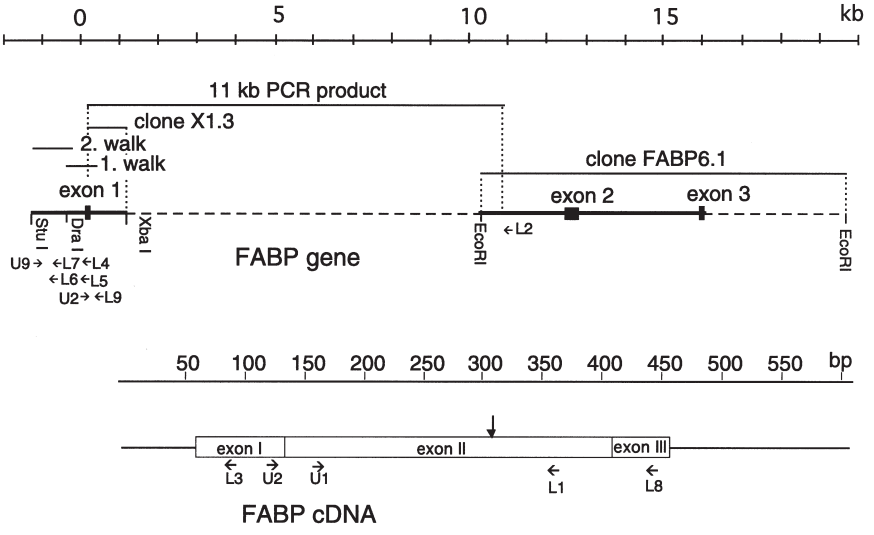

Fig. 1. Schematic representation of the locust muscle FABP gene and cDNA. Major clones and PCR products used in this study are shown above the depiction of the gene. Sequenced areas are shown as a the solid bold line; the areas marked with a dashed line were not sequenced. The approximate annealing sites for upper (→) and lower (←) PCR primers are indicated. The arrow pointing to the second exon of the cDNA indicates the location where an intron is present in most other FABP genes. Primer sequences: U1, 5′-TCG AGC GGA AGG CAG GTC-3′; L1, 5′-GAT GAT GGT GGG GTG GTC-3′; U2, 5′-AGC TCG ACT CGC AGA CCA ATT TTG AGG AAT A-3′; L2, 5′-CAAA TTA CAC TAG CAT CTC AG-3′; L3, 5′-TTG ATG CCT GCG AAT TCC TTC A-3′; L4, 5′-ATT GGT CTG CGA GTC GAG CTT GTA CTT GAT GCC TGC G-3′; L5, 5′-GCT GGC GGT GGG CAG TGG TCG GAG A-3′; L6, 5′-CCA TTG AAT GGG TTT TTA GTG CAG-3′; L7, 5′-GAC ACC ACT TGC ATT TTA CTG AAT-3′; L8, 5′-TAT TCT CGT TGC CAC CAG GTC G-3′; U9, 5′-TTC CCA TCC AAC AAA ATC ACA C-3′; L9, 5′-AGC AGA ACT AGG CAA GGA AAT A-3′. For further details on the cloning and sequencing strategies, see Section 2.

the flight muscle of 7-day-old adult locusts was isolated by a one step guanidinium isothiocyanate/phenol/chloroform extraction method adapted from Chomczynski and Sacchi (1987). First strand cDNAs were produced by reverse transcription using an oligo(dT) primer. This was followed by tailing at the 5′-end with dCTP and terminal transferase. The tailed cDNA product was amplified using an anchored forward primer 5′-AAG CAG TGG TAA CAA CGC AGA GTA CGC GGG-3′ (smart II) and a reverse primer L8 which anneals 15 bp upstream of the stop codon of the locust FABP cDNA (see Fig. 1). The amplified product was subcloned into the pCR2.1 cloning vector (Invitrogen) and sequenced.

2.7. Primer extension

Primer extension analysis was carried out using a primer Extension System (Promega). A reverse oligonucleotide L3 corresponding to nucleotides +26 to +4 of the coding sequence was 5′ end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. The radiolabeled primer was added to 20 μg of total RNA, incubated for 30 min at 60°C and then extended with AMV reverse transcriptase for 1 h at 42°C. The extension products were analyzed on 8% polyacrylamide gels containing 7 M urea.

2.8. Sequencing

Sequencing was carried out on an ABI sequencer at the Molecular Biotechniques Center at Simon Fraser University or with ABI's AmpliTaq DyeDeoxy terminator cycle sequencing chemistry (PE Biosystems) on the automatic ABI Model 373 Stretch DNA sequencer at Biotechnology Laboratory/Nucleic Acid Service Unit of the University of British Columbia. Clones or PCR products were sequenced on both strands, and ambiguous areas re-sequenced with alternative primers. Sequences were compared against Genbank.

3. Results

3.1. Cloning of the FABP gene

The large size of the locust genome (9.3 gigabases; Li, 1997, p. 383) posed a challenge for the cloning of the FABP gene. Highly specific probes were difficult to obtain from the 450 bp cDNA clone that was expected to represent several exons. The size of the introns was determined with PCR primers specific for sequences flanking the introns identified in all other sequenced members of the FABP gene family (Table 1). These experiments indicated that the first and third intron found in the other members of the FABP gene family are also present in the locust muscle FABP gene; the second intron Fig. 5, however, is missing. Following restriction digestion and Southern hybridization, a single band of approximately 10 kb was detected in *Eco*RI-digested genomic DNA, while the application of other enzymes resulted in several smaller fragments. A λZAP library was constructed from the 9–11 kb fraction of *Eco*RI-digested genomic DNA. To further reduce the screening, lysates of individual plates (50,000 pfu/plate) were analyzed by PCR for the presence of the FABP gene. Positive lysates were replated at 50,000 pfu/plate and screened with the FABP cDNA probe. Positive clones were re-plated and screened at 5000 pfu/plate and finally at 2p.383.00 pfu/plate. The insert of the resulting FABP6.1 was excised with *Eco*RI and cloned into Bluescript KS plasmid (Stratagene).

PCR analysis and sequencing revealed that the 10 kb clone (Fig. 1) contained exons 2 and 3 of the FABP gene, as well as 2 kb of intron I, the entire intron II (2.9 kb) and 3 kb of sequence downstream of the gene. The 3′ end of the gene was sequenced as far as the polyadenylation signal. Numerous attempts to obtain a clone upstream of FABP6.1 that contained the first exon failed; positive clones identified in library screening always turned out to be unstable in various host strains. Therefore, PCR primers suitable for long-range PCR were selected in the first exon as well as upstream of the second exon (Fig. 2). Long-template PCR resulted in an approx. 11 kb amplification product, which also could not be cloned into a plasmid vector. Restriction mapping of the PCR product revealed a unique *Xba*I site approx. 1.3 kb downstream of the first exon. The 1.3 kb fragment was cloned into Bluescript and sequenced (clone X1.3). This clone, and fragments thereof, were unsuccessfully used to screen genomic libraries, likely because of the highly repetitive nature of the sequence (see below). In order to obtain upstream sequence, a genome walking strategy was employed. As the previously published cDNA sequence (Price et al., 1992) was not complete, 5′-RACE was carried out to obtain the 5′ end of the coding sequence, as well as 57 bp of nontranslated cDNA. This sequence allowed the design of nested lower primers annealing upstream of the first intron, which were used together with adapter-specific upper primers for the PCR amplification of the promoter from *Dra*I-digested, adapter-ligated genomic DNA (Fig. 3). A 550 bp PCR product was cloned into pCR2.1 vector (Invitrogen) and sequenced. Nested primers designed from the 5′-end of this fragment were used to amplify an additional 700 bp of upstream sequence, this time from *Stu*I-digested, adapter-ligated genomic DNA. Thus, a total of 1.2 kb of upstream sequence was obtained. To prove that the clones obtained by the genomic walking approach represent indeed the sequence upstream of the first intron of the FABP gene, genomic DNA was used as a template for a PCR reaction with a primer combination

Table 1
Gene organization of the cloned muscle fatty acid binding protein genes

Gene	Intron I	Intron II	Intron III	Accession
Rat H-FABP	3.5	1.5	1.1	AF144090
Mouse H-FABP	3.4	1.5	1.1	U02884
Human H-FABP	3.4	1.9	1.4	U17081
Locust M-MABP	12.7	–	2.9	AF244980, AF244980
Fruit fly FABP	2.2	–	0.1	AAF54655

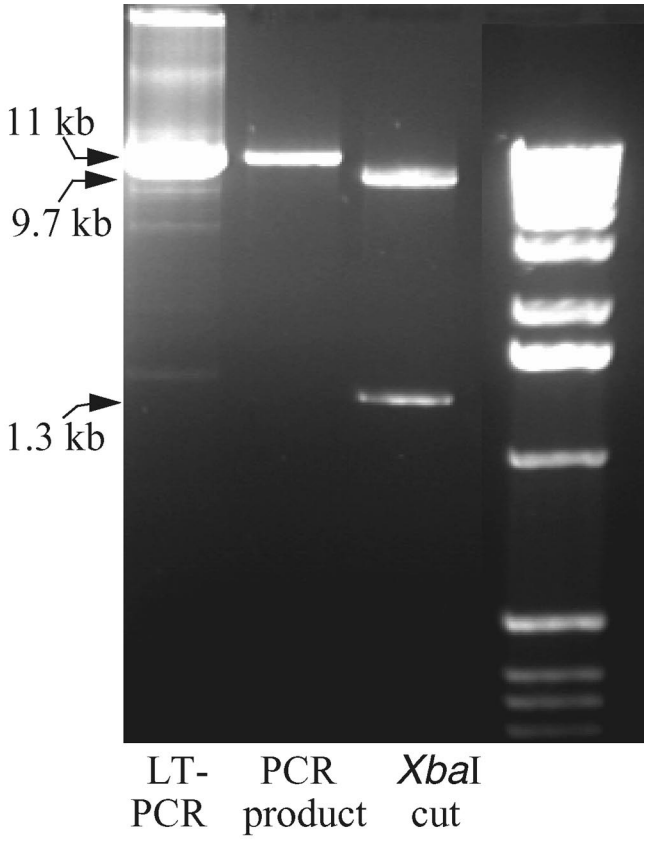

Fig. 2. Long template PCR of the first intron. PCR with primers U2 and L2 was carried out with the Expand Long Template (LT) PCR system as described in Section 2. The initial 11 kb LT-amplification product (first lane) was gel-purified (second lane). Alter restriction digestion with *Xba*I (third lane) the 1.3 kb fragment was eluted and cloned into Bluescript.

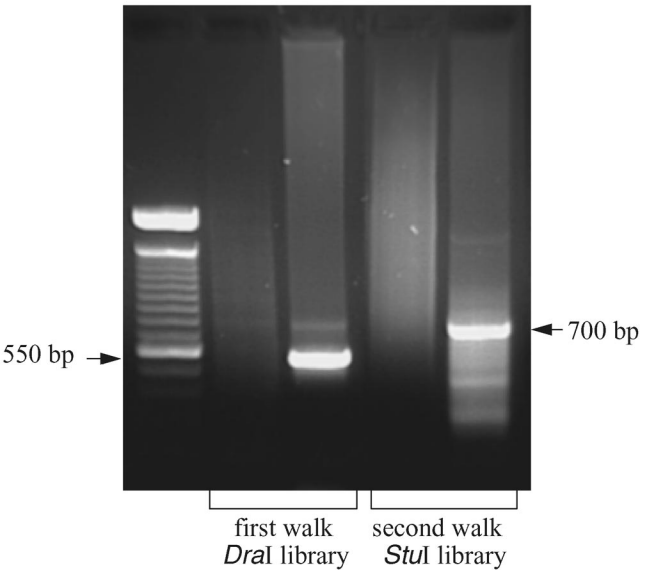

Fig. 3. Genome walking upstream of the FABP gene. Two separate genome walking reactions were carried out to obtain the sequence upstream of the FABP gene. In the first walk, nested lower primers specific for the 5′ end of exon I (L4, L5) and adapter specific upper primers were used to amplify a 550 bp product from *Dra*I digested, adapter ligated genomic DNA. Primers annealing near the 5′ end of the amplification product (L6, L7) were used for the second walk, with *Stu*I digested, adapter ligated DNA. For details, see Section 2.

annealing 800 bp upstream of the promoter and in the first intron, 500 bp downstream of the first exon. The sequence of the 1.3 kb PCR product obtained was identical to the previously determined sequence of exon 1 and its flanking regions.

3.2. Primer extension

Primer extension experiments from primer L3 resulted in a ~75–80 bp product (Fig. 4), representing ~50–55 bp of untranslated sequence upstream of the translation start codon. The transcription start site was determined from the 5′-RACE products, which contained 57 bp upstream of the translation start codon. The transcription start determined by RACE (24 bp downstream of the TATA box) is identical to the start site predicted by computer analysis.

4. Discussion

The gene for locust muscle FABP follows the same gene organization as the other members of the FABP superfamily, with the important exception that one of the introns is missing. Generally, three introns interrupt the coding sequence of this family of proteins (Veerkamp and Maatman, 1995): the first intron is inserted in the sequence coding for glycine 26, which is located in the turn of the helix-turn-helix motif that shields the binding cavity; the second following lysine 84, and the third before threonine 117 (numbering according to the struc-

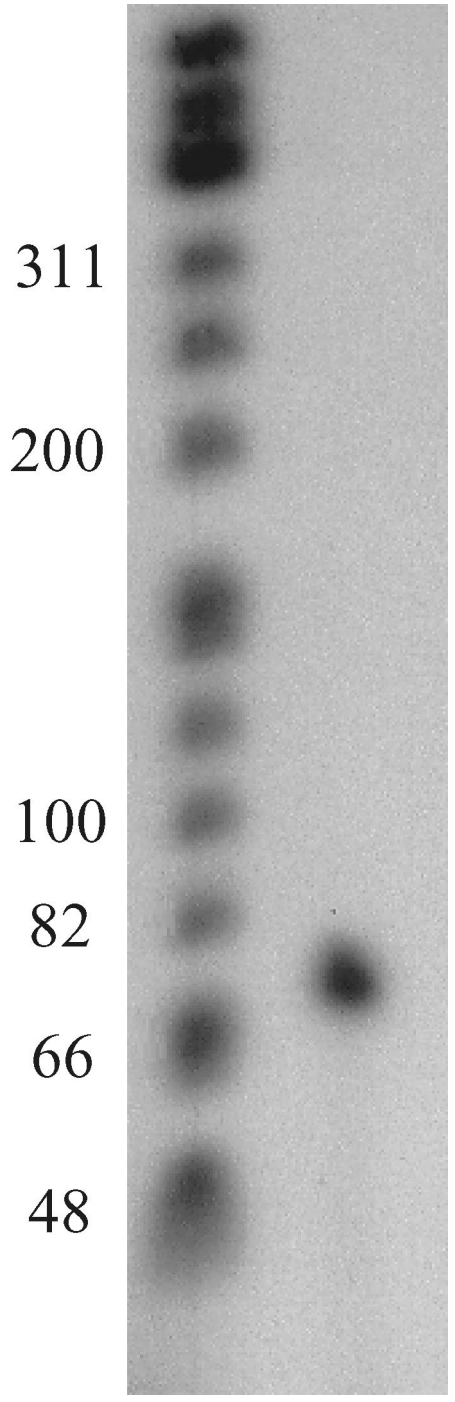

Fig. 4. Primer extension. Freshly isolated total RNA was reverse transcribed from ^{32}P-labeled lower primer L3, which anneals within the first exon. The extension product was separated on an 8% polyacrylamined gel, containing 7 M urea. Size markers were ^{32}P-labeled dephosphorylated ΦX174 *Hinf*I fragments. For details, see Section 2.

ture of locust FABP; Haunerland et al., 1994). These three introns are present in all other known FABP genes. The locust gene also has introns at glycine 26 and threonine 117, but lacks the second intron (Fig. 5). While the location of the introns is apparently conserved between the various species, their size varies widely (Table 1). The locust gene contains rather large introns. Coding DNA comprises only 4% of the 16 kb locust muscle FABP gene. While large introns are not unusual in organisms with very large genomes, they are less common for very strongly transcribed genes, like the locust FABP.

A 190 bp repeat is present in intron I (864–1052) and intron II (15419–15612); this repetitive sequence has been found before in the genome of *Schistocerca* species. The repeat has been identified in an intron of an antennapedia-class homeodomain gene of *S. gregaria* (Dawes et al., 1994) (88% identical over 120 bases, Genbank accession number 396776) and in the second intron of the *S. nitens* adipokinetic hormone II gene (Noyes and Schaffer, 1993) (91% over 183 bases, Genbank accession number L08775). This element is similar to a repetitive element found in multiple copies in the migratory locust, *Locusta migratoria* (Bradfield et al., 1985) (Genbank accession number M12077). In addition, a 250 bp sequence string further upstream in the first intron (296–545) is very similar to part of the first intron of the *S. gregaria* adipokinetic hormone I gene (Noyes and Schaffer, 1993) (71%, Genbank accession number L08771), indicating that this also represents a repetitive sequence occurring in multiple copies in the locust genome. The presence of this sequence near the 5′-end of the first intron explains why screening with hybridization probes from this area was not successful and resulted in numerous false positives.

Few other noteworthy features were detected in the first intron, but it can be expected that some unusual sequence prevented the cloning of this intron. The second intron contains an extended GT-dinucleotide repeat (15678–15745), reminiscent of a microsatellite. Interestingly, a GT-microsatellite is also present in the homologous intron of the human heart FABP gene (Arlt et al., 1996). However, these microsatellites are probably not of ancient origin and may have evolved independently, since similar sequences are not present in other FABP genes, including the H-FABP gene from either mouse or rat. It remains to be seen whether this repeat will be useful for distinguishing between individual populations of *S. gregaria*.

Few insect FABPs have been identified, and muscle FABP is known only from two closely related locust species (*L. migratoria*; Van der Horst et al., 1993; *S. gregaria*, Price et al., 1992). The recent release of the complete *Drosophila melanogaster* genome (Adams et al., 2000) offered the opportunity to search for an analogous gene in a Dipteran species. Only one homologous gene could be found (Genbank accession AAF54655), which codes for an EST from a *D. melanogaster* adult brain library (Genbank accession number AF083313) previously identified as similar to a putative mosquito FABP (della Torre et al., 1996) (*Anopheles gambiae*, Genbank accession number U50472). The deduced

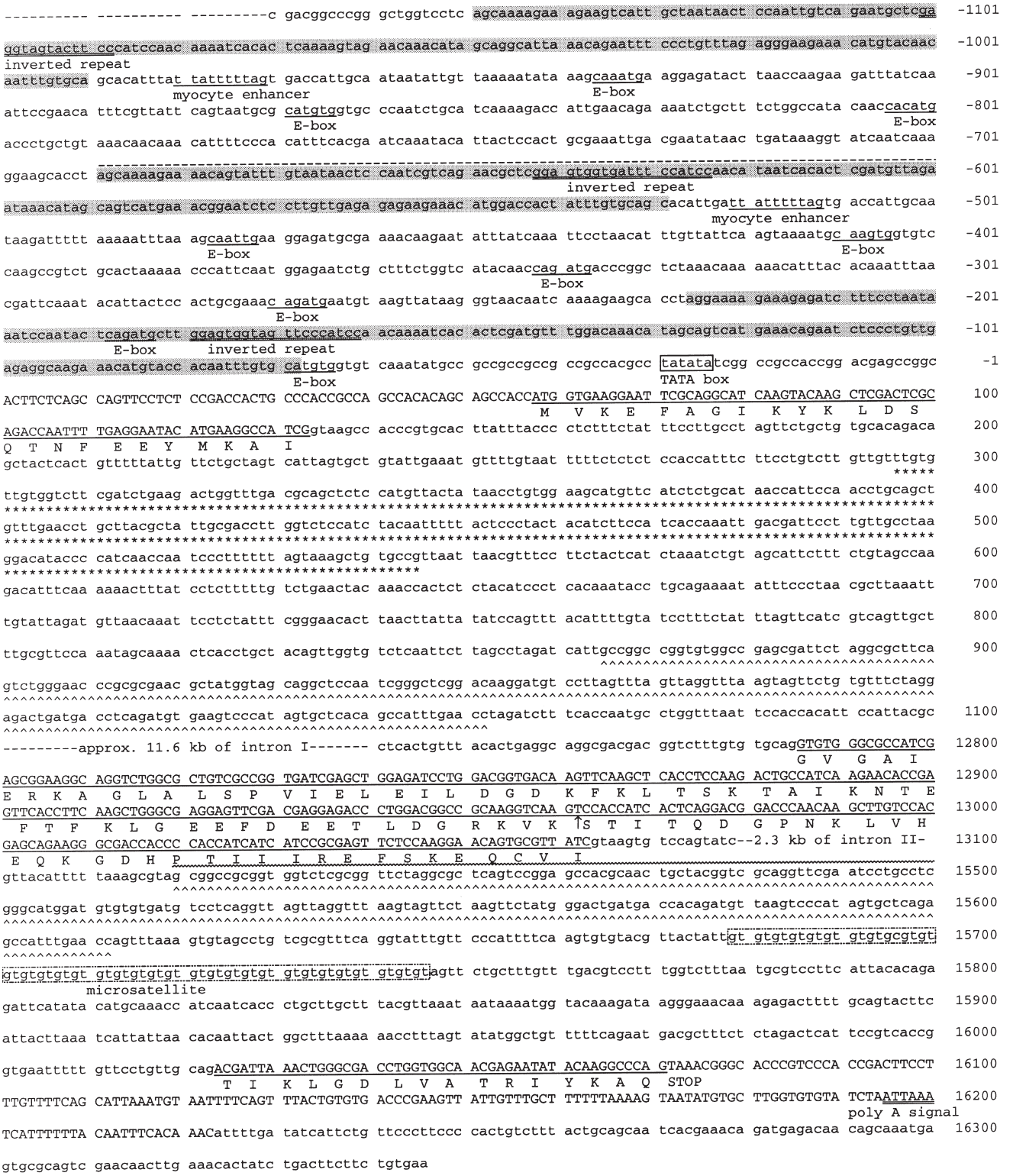

Fig. 5. Sequence of the locust FABP gene. Exons are shown as capital letters, while the introns and the 5′- and 3′-untranscribed regions are shown as lower case letters. The TATA box is located at −24. Potential regulatory elements are underlined. Also identified are a microsatellite in intron 2, and the polyadenylation signal. Various repeat regions are marked as follows: shading indicates the 180 bp sequence repeats in the upstream region. The circumflex sign below the sequence shows segments in intron I and II that are homologous to a repetitive element found in other locust species, and stars below the sequence mark the region that is homologous to a sequence segment found in the intron of the AKH gene.

amino acid sequence of the *Drosophila* protein is highly homologous to the locust muscle FABP (52% identity, and an additional 17% conservative substitutions), a higher degree of sequence homology than found with the *A. gambiae* clone. Vertebrate heart FABPs are also highly homologous to the *Drosophila* cDNA, as are the FABPs found in adipocytes, brain, the retina, and nerve cells (myelin P2 protein), which together with muscle FABP form a subfamily of fatty acid binding proteins that branched out less than 300 million years ago, long after the vertebrate–invertebrate divergence that occurred more than 600 million years ago (Matarese et al., 1989). Another member of this subfamily appears to be a protein found in the brain of the moth, *Manduca sexta*, which has been described as a cellular retinoic acid-binding protein (Mansfield et al., 1998). It is hence likely that the gene identified in *D. melanogaster* represents the dipteran homologue of the locust muscle FABP. The putative FABP gene from fruit fly and the locust FABP gene are similarly organized: the promoter in *D. melanogaster* is located 157 bp upstream of the first exon (17190), uses the same TATA box (TATATA) and contains GC-rich sequences as well (Fig. 6). The introns of the *D. melanogaster* gene are located in identical positions as in the *S. gregaria* gene, although they are much smaller (Table 1). The first intron spans 2.2 kb (17032–14833), and the second exon found in vertebrate FABP genes is absent, just as reported here for the locust FABP gene. The final intron in *D. melanogaster* is only 180 bp (14555–14832).

We have previously demonstrated that the locust FABP gene described here is indeed expressed in locust flight muscles (Zhang and Haunerland, 1998). Using a primer combination specific for a 597 bp sequence located in the first intron (1722 bp upstream of exon II), we carried out RT-PCR from total RNA of locust flight muscle, amplifying only unprocessed primary transcript. Because of the rapid splicing and degradation of intron sequences the amount of primary transcript is a good indicator of the rate of gene expression. Quantitative studies revealed that FABP is not expressed in the mesothorax muscles prior to metamorphosis. In fully differentiated flight muscles, however, FABP comprises almost 20% of all cytosolic proteins (Haunerland et al., 1992). These levels are a consequence of a high expression rate of the FABP gene as well as the long half-life of FABP and its mRNA. FABP gene expression commences immediately after metamorphosis and reaches extremely high levels within one day. Up to 800,000 copies of primary transcript/ng RNA can be found at this time, equivalent to ~0.5% of the total RNA. Several days later, when FABP approaches its maximal value, the primary transcript levels decrease 10,000-fold (Zhang and Haunerland, 1998). These findings are suggestive of a very strong promoter that can be tightly controlled.

Inspection of the promoter region of the locust FABP gene reveals several features indicative of a strong promoter. The TATA-box is located 24 bp upstream of the predicted transcription start site, which was verified experimentally by 5′-RACE and confirmed by the primer extension experiment. Immediately upstream and downstream of the TATA box, GC-rich areas can be found that are frequently associated with strong promoters. Comparison of the promoter with that of the vitellogenin gene from *L. migratoria*, a gene that is strongly expressed in fat body during oogenesis (Locke et al., 1987) (Genbank accession numbers M17333 and M17334), shows similar core elements (Fig. 6). In both promoters, identical TATA-boxes (TATATA) are located 24 bp upstream of the transcription start site. Both genes contain GC-rich areas flanking the TATA-box, which resemble the activator sequences found in many strong promoters (GC-box). As mentioned before, a similar core promoter structure is also seen in the putative FABP gene from *D. melanogaster* (Fig. 6).

The identification of other potential elements that control tissue specific expression is more difficult, as few consensus sequences are known for insect muscle genes. From locust or other orthopteran species, no nuclear gene coding for a muscle-specific protein has ever before been cloned. Several genes that are expressed in *D. melanogaster* muscles, however, have been shown to contain E-boxes, just like most vertebrate muscle genes, which can act as recognition sites for the transcription factor MyoD (Edmondson and Olson, 1993). Nine potential muscle E-boxes (consensus sequence CANNTG) are located within 1 kb upstream of the promoter of the locust FABP gene, some of which may direct muscle specific gene expression (Fig. 5). Upstream of the putative *Drosophila* FABP gene several E-boxes are present as well. Also noteworthy are the sequence motifs at −513 to −522, and at −972 to −981, potential binding sites for the myocyte enhancer MEF2 (consensus sequence YTAWWWWTAR) (Black and Olson, 1998) (Fig. 5).

The 1.2 kb of upstream sequence cloned here show considerable repetitiveness; a stretch of ~160 bp is repeated three times (−1150 to −991; −690 to −532;

Fig. 6. Sequence comparison of the core promoter. The core promoter of *L. migratoria* vitellogenin (VG b L.m.) and of the putative *D. melanogaster* FABP (FABP D.m.) are aligned with the locust FABP promoter (FABP S.g.). Bases identical with the locust FABP promoter are shaded.

−228 to −69), but no similar sequences have so far been detected elsewhere in the FABP gene or in any other orthopteran genes. This raises the possibility that the repeats contain regulatory elements which could act more efficiently in multiple copies. A comprehensive sequence analysis did not reveal known transcription factor binding sites within these repeats. Noteworthy, however, is the presence of a 19 bp inverted repeat sequence, 5′-GGAGTGGTA N TTCCCATCC-3′. A similar, partially palindromic sequence is also found upstream of the putative *Drosophila* FABP promoter (Fig. 7). This sequence does not resemble any known regulatory consensus sequences. A strikingly similar, but reversed palindromic sequence is found within 600 bp upstream of the promoter of all mammalian heart FABP genes (Fig. 7). It is tempting to speculate that these elements serve similar purposes, perhaps in the metabolic regulation of muscle FABP genes, which to date is only poorly understood.

In earlier studies we demonstrated that the expression of FABP gene in locust muscle is stimulated by increased fatty acid supply (Chen and Haunerland, 1994), and fatty acids have also been shown to increase FABP expression in mammalian muscle cells (van der Lee et al., 2000). There is substantial evidence that fatty acids or their metabolites can modulate gene expression at the level of transcription initiation (Van Bilsen et al., 1998), by mechanisms similar to lipophilic hormones such as steroids, retinoids, and thyroxins. The best characterized class of transcription factors that can be activated by fatty acids are the peroxisome proliferator activated receptors (PPARs), so called because of their activation by fibrate drugs known to stimulate the formation of peroxisomes (Issemann and Green, 1990). The involvement of these receptors in gene control has been established for a number of proteins related to lipid-metabolisms that are expressed in adipose and hepatic tissue, including the adipose fatty acid binding protein (Frohnert et al., 1999). However, peroxisome proliferator-response elements, direct repeat elements with the consensus sequence AGGTCA N AGGTCA, have not been identified upstream of the mammalian muscle FABP genes, although potential candidate sequences are present in the rodent genes (Treuner et al., 1994; Zhang et al., 1999). No direct repeat elements could be found

Fig. 7. Alignment of a potential FABP-specific upstream element. Sequences of inverted repeats found upstream of various muscle FABP genes are aligned. The sequences of the mammalian elements are shown in reverse orientation (3′→5′).

within the upstream sequence of the locust FABP gene described here. It may well be possible that other, not yet discovered factors, whether novel forms of PPAR or entirely different proteins, are responsible for the recognition of free fatty acid accumulation in the muscle cells. Reporter gene studies with both mammalian and locust FABP promoters will be required to further identify and characterize such regulatory elements.

Acknowledgements

This work was funded by grants from the Natural Science and Engineering Research Council of Canada and the Heart and Stroke Foundation of British Columbia and Yukon. We thank Rick Martel for technical assistance.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G. et al., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Ando, S., Xue, X.-H., Tibbits, G.F., Haunerland, N.H., 1997. Cloning and sequencing of complementary DNA for fatty acid binding protein from rainbow trout heart. Comp. Biochem. Physiol. 119B, 213–217.

Arlt, M.F., Goodfellow, P.J., Rottman, J.N., 1996. Dinucleotide repeat in the third intron of the FABP3/MDGI putative tumor suppressor gene. Dis. Markers 13, 57–59.

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. Adv. Protein Chem. 45, 89–151.

Black, B. L, Olson, E.N., 1998. Transcriptional control of muscle development by myocyte enhancer factor-2, MEF2 proteins. Annu. Rev. Cell Dev. Biol. 14, 167–196.

Bradfield, J.Y., Locke, J., Wyatt, G.R., 1985. An ubiquitous interspersed DNA sequence family in an insect. DNA 4, 357–363.

Chen, X., Haunerland, N.H., 1994. Fatty acid binding protein expression in locust flight muscle. Induction by flight, adipokinetic hormone, and low density lipophorin. Insect Biochem. 24, 573–579.

Chomczynski, P., Sacchi, N., 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.

Dawes, R., Dawson, I., Falciani, F., Tear, G., Akam, M., 1994. Dax, a locust Hox gene related to fushi-tarazu but showing no pair-rule expression. Development 120, 1561–1572.

Della Torre, A., Favia, G., Mariotti, G., Coluzzi, M., Mathiopoulos, K.D., 1996. Physical map of the malaria vector *Anopheles gambiae* and possible non-random distribution of coding regions. Genetics 143, 1307–1311.

Edmondson, D.G., Olson, E.N., 1993. Helix-loop-helix proteins as regulators of muscle-specific transcription. J. Biol. Chem. 268, 755–758.

Frohnert, B.I., Hui, T.Y., Bernlohr, D.A., 1999. Identification of a functional peroxisome proliferator-responsive element in the murine fatty acid transport protein gene. J. Biol. Chem. 274, 3970–3977.

Gerbens, F., Rettenberger, G., Lenstra, J.A., Veerkamp, J.H., Te Pas, M.F., 1997. Characterization, chromosomal localization, and gen-

etic variation of the porcine heart fatty acid-binding protein gene. Mamm. Genome 8, 328–332.

Guglielmo, C.G., Haunerland, N.H., Williams, T.D., 1998. Fatty acid binding protein, a major protein in the flight muscle migrating Western Sandpipers. Comp. Biochem. Physiol. 119B, 549–555.

Haunerland, N.H., 1994. Fatty acid binding proteins in locust and mammalian muscle. Comparison of structure, function, and regulation. Comp. Biochem. Physiol. 109B, 199–208.

Haunerland, N.H., Andolfatto, P., Chisholm, J.M., Wang, Z., Chen, X., 1992. Fatty-acid-binding protein in locust flight muscle: developmental changes of expression, concentration and intracellular distribution. Eur. J. Biochem. 210, 1045–1051.

Haunerland, N.H., Jacobson, B.L., Wesenberg, G., Rayment, I., Holden, H.M., 1994. Three-dimensional structure of the muscle fatty-acid-binding protein isolated from the desert locust *Schistocerca gregaria*. Biochemistry 33, 12378–12385.

Issemann, I., Green, S., 1990. Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. Nature 347, 645–650.

Jowett, T., 1986. Preparation of nucleic acids. In: Roberts, D.B. (Ed.), *Drosophila*: A Practical Approach. IRL Press, Oxford, pp. 275–279.

Li, W.-H., 1997. Molecular Evolution. Sinauer, Sunderland, MA.

Locke, J., White, B.N., Wyatt, G.R., 1987. Cloning and 5′ end nucleotide sequences of two juvenile hormone-inducible vitellogenin genes of the African migratory locust. DNA 6, 331–342.

Mansfield, S.G., Cammer, S., Alexander, S.C., Muehleisen, D.P., Gray, R.S., Tropsha, A., Bollenbacher, W.E., 1998. Molecular cloning and characterization of an invertebrate cellular retinoic acid binding protein. Proc. Natl. Acad. Sci. USA 95, 6825–6830.

Matarese, V., Stone, R.L., Waggoner, D.W., Bernlohr, D.A., 1989. Intracellular fatty acid trafficking and the role of cytosolic lipid binding proteins. Prog. Lipid Res. 28, 245–272.

Noyes, B.E., Schaffer, M.H., 1993. The closely related neuropeptide genes encoding adipokinetic hormones I and II have very different 5′-flanking regions. DNA Cell Biol. 12, 509–516.

Phelan, C., Morgan, K., Baird, S., Korneluk, K., Narod, S., Pollak, M., 1996. The human mammary-derived growth inhibitor, MDGI. gene: genomic structure and mutation analysis in human breast tumors. Genomes 34, 63–68.

Price, H.M., Ryan, R.O., Haunerland, N.H., 1992. Primary structure of locust flight muscle fatty acid binding protein. Arch. Biochem. Biophys. 297, 285–290.

Treuner, M., Kozak, C.A., Gallahan, D., Grosse, R., Müller, T., 1994. Cloning and characterization of the mouse gene encoding mammary-derived growth inhibitor/heart-fatty acid-binding protein. Gene 147, 237–242.

Van Bilsen, M., Van der Vusse, G.J., Reneman, R.S., 1998. Transcriptional regulation of metabolic processes: implications for cardiac metabolism. Pflügers Archs. 437, 2–14.

van Breda, E., Keizer, H.A., Vork, M.M., Surtel, D.A.M., de Jong, Y.F., van der Vusse, G.J.K., Glatz, J.F.C., 1992. Modulation of fatty-acid-binding protein content of rat heart and skeletal muscle by endurance training and testosterone treatment. Pflügers Archs. 421, 274–279.

Van der Horst, D.J., Van Doorn, J.M., Passier, P.C.C.M., Vork, M.M., Glatz, J.F.C., 1993. Role of fatty acid binding protein in lipid metabolism of insect flight muscle. Mol. Cell Biochem. 123, 145–152.

van der Lee, K.A., Vork, M.M., De Vries, J.E., Willemsen, P.H., Glatz, J.F., Reneman, R.S., Van der Vusse, G.J., Van Bilsen, M., 2000. Long-chain fatty acid-induced changes in gene expression in neonatal cardiac myocytes. J. Lipid Res. 41, 41–47.

van Nieuwenhoven, F.A., Verstijnen, C.P.H.J., Abumrad, N.A., Willemsen, P.H.M., van Eys, G.J.J.M., van der Vusse, G.J., Glatz, J.F.C., 1995. Putative membrane fatty acid translocase and cytoplasmic fatty acid-binding protein are co-expressed in rat heart and skeletal muscles. Biochem. Biophys. Res. Commun. 207, 747–752.

Veerkamp, J.H., Maatman, R.G.H.J., 1995. Cytoplasmic fatty acid-binding proteins: their structure and genes. Prog. Lipid Res. 34, 17–52.

Veerkamp, J.H., Peeters, R.A., Maatman, R.G.H.J., 1991. Structural and functional features of different types of cytoplasmic fatty acid-binding proteins. Biochim. Biophys. Acta 1081, 1–24.

Zhang, J., Haunerland, N.H., 1998. Transcriptional regulation of FABP expression in flight muscle of the desert locust, *Schistocerca gregaria*. Insect Biochem. Mol. Biol. 28, 683–691.

Zhang, J., Rickers-Haunerland, J., Dawe, I., Haunerland, N.H., 1999. Structure and chromosomal location of the rat gene encoding the heart fatty acid-binding protein. Eur. J. Biochem. 266, 347–351.