Regulatory elements in the first intron of the rat fatty acid synthase gene

Babak OSKOUIAN, Vangipuram S. RANGAN and Stuart SMITH*

Children's Hospital Oakland Research Institute, 747 52nd Street, Oakland, CA 94609, U.S.A.

Sequence elements have been identified within the 1.2 kb-long first intron of the fatty acid synthase (FAS) gene that mediate both positive and negative effects on transcription. The negative regulatory element, when positioned downstream of either the FAS or simian virus 40 promoter, down-regulates the expression of a coupled reporter gene in an orientation-dependent manner. Sequences mediating this effect have been mapped, by deletion mutagenesis, to two regions approximately within nucleotides +405 to +768 and +924 to +1083. Both regions contain sequence elements that are strongly protected from DNase I digestion by nuclear extracts prepared from liver, but not by those prepared from spleen. The results of run-on assays performed with nuclei derived from tissues that express FAS at either high or low levels indicate that the different rates of

transcription of the endogenous FAS gene result from differences in the extent of initiation, so it is unlikely that the negative effect is caused by transcriptional pausing in the first intron. The positive element maps to nt +292 to +297 and corresponds to an authentic binding site for upstream stimulatory factor (USF). This USF-binding element can up-regulate transcription from a heterologous promoter in a position- and orientation-independent manner. However, in the context of the entire FAS first intron, the effect of the USF-binding site is masked unless the effect of the negative elements is ablated by mutagenesis. These results suggest that the dominant negative element of the first intron may play a role in determining the tissue-specific expression of the FAS gene.

INTRODUCTION

In animals, the conversion of malonyl-CoA into fatty acid is catalysed by a single multifunctional protein, fatty acid synthase (FAS). Expression of the FAS gene is regulated in a tissue-specific manner in response to various developmental, nutritional and hormonal signals. For example, the concentrations of FAS mRNA and protein are elevated in mammary gland epithelia during the lactogenic phase of development [1], in pre-adipocytes as they differentiate into mature adipocytes [2] and in fetal lung type II cells at the onset of surfactant production [3]. In tissues such as liver and adipose, which play a key role in the maintenance of energy balance, FAS mRNA and protein levels are depressed on fasting or on ingestion of diets high in polyunsaturated fat [4–6]. In other tissues, such as spleen, FAS mRNA and protein are barely detectable, regardless of developmental or nutritional status [7].

In recent years, evidence has accumulated indicating that many, if not all, of these tissue-specific changes in the concentrations of FAS mRNA and protein result, at least in part, from changes in the level of transcription of the FAS gene, and several hormones, including insulin, glucagon, thyroid hormone and glucocorticoids, have been implicated as agents capable of influencing FAS expression [8–11]. It is likely that the key to understanding these regulatory processes will be found in the DNA sequences flanking the transcribed portion of the FAS gene, and these sequences have been identified for both the rat [7] and the goose [12]. The 5'-proximal promoter region of the rat FAS gene is characterized by the presence of a TATA box between nucleotides –33 and –26, a positively acting inverted CCAAT box between nucleotides –98 and –92, and a functional Sp1-binding site at positions –241 to –236 [13]. Transcriptional analysis has also identified the location of sequence elements required for mediating the positive action of insulin as being between nt –68 and –52 [14], and of those required for

mediating the negative action of cAMP between nt –98 and –92 [15]. The objective of the present study was to ascertain whether sequences present in the first intron might play a role in modulating transcription of the FAS gene.

MATERIALS AND METHODS

Cell cultures

Rat H4IIE hepatoma cells were cultured in minimum essential medium containing Earl's balanced salt solution and non-essential amino acids supplemented with 10% (v/v) fetal bovine serum and 10% (v/v) bovine calf serum. The murine pre-adipocyte cell line 3T3-30A5 was obtained from Dr. Ki-Han Kim (Purdue University, IN, U.S.A.) and cultured in Dulbecco's modified Eagle's medium (DME-H21) containing 10% (v/v) fetal bovine serum. Adipocyte differentiation was induced by 'nutrient deprivation' followed by insulin treatment [16]. All cells were grown at 37 °C under 5% CO₂.

Transient transfections

Cell lines were grown to 60–70% confluency on 60 mm² dishes and were transfected with 10 µg of plasmid DNA per plate by the calcium phosphate precipitation method [17]. Normalization of transfection efficiency was facilitated by co-transfection with a β-galactosidase expression plasmid.

Stable transfections

3T3-30A5 cell lines with stably integrated –249/+1272 and –249/+68 FAS-CAT genes (where CAT is chloramphenicol acetyltransferase) were obtained by separately co-transfecting the cell line with plasmids containing these genes together with the linearized pSV-Neo plasmid. Stable clones were selected by their resistance to the antibiotic G418, and approx. 50 colonies were pooled [18].

Abbreviations used: FAS, fatty acid synthase; CAT, chloramphenicol acetyltransferase; TK, thymidine kinase; USF, upstream stimulatory factor; RT-PCR, reverse transcription-PCR; SV40, simian virus 40.

* To whom correspondence should be addressed.

Enzyme assays

CAT activity was assayed radiochemically [19], and product formation was quantified by analysis on a Molecular Dynamics Phosphorimager. The activity of β -galactosidase was determined spectrophotometrically [20].

Preparation of nuclear extracts

Rats were either fasted throughout or fasted and then re-fed on a fat-free, high-carbohydrate diet to produce, respectively, very low or very high levels of hepatic FAS expression. Nuclear extracts were prepared from both liver and spleen [21]. Nuclear extracts were prepared from 3T3-30A5 and H4IIE cells essentially as described by Dignam et al. [22].

Electrophoretic mobility shift assays

DNA fragments were 5'-end-labelled with [γ - 32 P]ATP using T4 polynucleotide kinase and purified using Sep-Pak C18 [20]. Nuclear extracts were first incubated at either 20 °C or 0 °C for 20 min in 20 μ l of reaction mixture containing 20 mM Hepes (pH 7.9), 1 mM dithiothreitol, 60 mM KCl, 0.5 mM EDTA, 5 % glycerol and 20–60 μ g/ml double-stranded poly(dI-dC). The radiolabelled DNA probes [(1–5) $\times$ 10⁴ c.p.m./pmol] were then added and incubation continued for a further 30 min at 20 or 0 °C. The DNA–protein complexes were separated by electrophoresis through 5 % polyacrylamide gels (acrylamide/bisacrylamide ratio of 38:2) in 22.5 mM Tris borate/0.5 mM EDTA. The gels were dried and exposed to X-ray film at –70 °C.

DNase I footprinting assays

Various amounts of nuclear extract, or BSA, were incubated with 20–60 μ g/ml double-stranded poly(dI-dC) at 0 °C for 20 min in 20 mM Hepes (pH 7.9), 60 mM KCl, 1 mM dithiothreitol, 0.5 mM EDTA and 5 % glycerol contained in a final volume of 50 μ l. The end-labelled probes were then added and the mixture was incubated for a further 40 min at 0 °C. An equal volume of a solution containing 10 mM MgCl₂, 10 mM CaCl₂ and 0.2 unit of DNase I (Promega, Madison, WI, U.S.A.) was then added, the digestion was allowed to proceed for 3 min at 20 °C and the products were analysed [23].

Reverse transcription–PCR (RT-PCR)

3T3-30A5 preadipocytes were transiently transfected with different CAT reporter gene constructs. Poly(A)⁺ mRNA was purified using Oligotex beads (Qiagen, Chatsworth, CA, U.S.A.) as specified by the manufacturer. The mRNA was used to programme a coupled RT-PCR using avian myoblastosis virus reverse transcriptase and the thermostable *Tfi* DNA polymerase from *Thermus flavus* (Promega) according to the manufacturer's instructions. The cDNA primer oligonucleotide bore the sequence 5'-CCAGCTCACCGTCTTTCATTGCCATACGG-3', corresponding to the region comprising positions 2533–2561 of the pCAT-Basic plasmid, and the FAS specific oligonucleotide had the sequence 5'-TCCAGTTTCGCTGCCGCGCTCCTCG-CTTGTCG-3', corresponding to positions +24 to +55 of the FAS sequence.

Run-on transcription assay

Intact nuclei were isolated from livers and spleens of fasted/re-fed rats and from livers of fasted rats essentially as described by Gorski et al. [21]. The transcripts were labelled by incubation of approx. 5 $\times$ 10⁶ nuclei with 200 μ Ci of [α - 32 P]rGTP, as described by Greenberg [24]. The RNA was then purified and 5 $\times$ 10⁶ c.p.m.

of labelled RNA was hybridized to a nitrocellulose filter containing 10 μ g of each denatured DNA probe per slot. Hybridization and washing conditions were as described [20].

Construction of plasmids containing a series of chimaeric FAS–reporter genes

Plasmids containing various FAS–CAT fusion genes were generated by engineering a series of nested deletions from the FAS portion of the DNA using exonuclease III and mung bean nuclease by the method developed by Henikoff [25]. Two different plasmids served as substrates for the deletion constructions. The first plasmid contained FAS sequence from –608 to +1272 (where +1 is the start of transcription) cloned upstream of promoterless plasmid pCAT-Basic (Promega). The second plasmid contained FAS sequence from –450 to +68 cloned in pCAT-Basic as a *Sau*3AI fragment. The resulting deletion constructs were sized by agarose-gel electrophoresis, and the end points of the deletions were determined by nucleotide sequencing of the fusion junctions.

Construction of thymidine kinase (TK)-CAT derivatives

An oligonucleotide with the sequence 5'-GATCCGCTGTCACGTGGGCGCCGCG-3', corresponding to positions +283 to +305 within the FAS intronic sequences, preceded by GATC was synthesized and inserted into the *Bam*HI site of the vector TK-CAT in both orientations as well as in tandem.

Insertion of the FAS first intron into a heterologous gene

An oligonucleotide with the sequence 5'-CTCCATTTTAGCTTCCTGAGCTCCTGAAAATCTCGCC-3' was used to introduce a *Sac*I site (underlined) 14 bp upstream from the initiation codon of the CAT gene in plasmid pCAT-Control using the protocol developed by Kunkel et al. [26]. This unique *Sac*I site is placed within the transcriptional unit of the CAT gene driven by simian virus 40 (SV40) promoter and enhancer sequences. PCR was employed to amplify the FAS first intron along with its splice signals and simultaneously to introduce flanking *Sac*I sites (underlined sequences); the primers used were: (i) 5'-GCCTC-CAGAGCTCAGACAGGTACGCC-3' and (ii) 5'-AAATTT-GAGCTCTTCTCTGTGGGGGAAGGT-3'. The *Sac*I sites were then used to clone the FAS first intron into the pCAT-Control plasmid in a position identical to that in which these sequences lie with respect to the transcription and translational signals in the FAS gene.

Engineering of deletions within the first intron

The removal of specific intronic sequences was achieved either by exploiting the available restriction sites or by using the exonuclease/mung bean nuclease method [25]. All other DNA manipulations were performed using standard methods [20].

RESULTS

The presence of intronic sequence down-regulates the FAS promoter

The possibility that the first intron of the FAS gene may harbour regulatory elements capable of affecting transcription was investigated initially by comparing the levels of reporter gene expression produced by two series of FAS–CAT fusion plasmids, one containing the FAS intron and the other devoid of intronic sequence. The intron-containing plasmids all have the same FAS 3' end at the FAS–CAT fusion junction corresponding to FAS nt +1272, whereas the intronless plasmids all have nt +68 of the

Table 1 Promoter activity of the FAS gene in the presence or absence of the intronic sequences

A series of three 5'-deletions (—450, —249 and —149) was engineered in the upstream region of the FAS gene, in the absence or presence of the FAS first intron. The constructs were cloned in front of a promoterless CAT gene and the resulting plasmids were transiently transfected into either 3T3-30A5 pre-adipocytes or H4IIE cells. CAT activity was assayed in cell extracts. Values are means $\pm$ S.D. of four independent experiments.

Construct used	Chloramphenicol acetylated (%/h)	
	3T3-30A5	H4IIE
—450/+68	13.8 $\pm$ 1.0	3.2 $\pm$ 0.5
—450/+1272	4.1 $\pm$ 0.8	1.0 $\pm$ 0.3
—249/+68	16.7 $\pm$ 1.8	5.1 $\pm$ 0.8
—249/+1272	4.1 $\pm$ 0.9	1.3 $\pm$ 0.7
—149/+68	7.6 $\pm$ 0.7	2.9 $\pm$ 0.5
—149/+1272	3.4 $\pm$ 1.2	1.4 $\pm$ 0.7

Table 2 The negative effect of the FAS first intron is maintained throughout differentiation of 3T3-30A5 cells into adipocytes

3T3-30A5 pre-adipocytes, stably transfected with the —249 FAS-CAT reporter gene constructs either lacking or containing the FAS first intron, were induced to differentiate for 0 (pre-adipocyte), 4 (intermediate) or 10 (adipocyte) days prior to assay of CAT activity. Values are means $\pm$ S.D. of four independent experiments.

Construct used	Chloramphenicol acetylated (%/h)		
	Preadipocyte	Intermediate	Adipocyte
—249/+68	5.4 $\pm$ 0.9	13.3 $\pm$ 2.8	27.0 $\pm$ 2.6
—249/+1272	0.8 $\pm$ 0.02	1.1 $\pm$ 0.05	2.5 $\pm$ 0.3

FAS gene at the fusion junction. Members of each series differ in the length of the upstream, non-transcribed FAS sequences contained at their 5'-ends. Without exception, plasmids of the intron-containing series, when transfected into H4IIE cells and 3T3-30A5 preadipocytes, produced levels of reporter gene expression lower than those of corresponding plasmids of the intronless series (Table 1). Furthermore the level of reporter gene expression was essentially the same in cells transfected with the intron-containing plasmids, regardless of the length of FAS 5'-sequences included in the constructs. This was despite the presence of a functional Sp1 binding site between nucleotides —241 and —236, which has been shown to direct increased transcription of the —249/+68 intronless construct compared with the —149/+68 intronless construct (compare lines 3 and 5 in Table 1), and a negative element between —319 and —301 that down-regulates transcription, particularly in H4IIE cells (compare lines 1 and 3 in Table 1). Evidence for the presence of these positive- and negative-acting elements in the proximal promoter region was presented previously [13]. The new data suggested that the presence of FAS intronic sequence has a negative effect on transcription that completely masks the effects of the regulatory elements present in the non-transcribed 5'-flanking region.

Negative regulation of chimaeric FAS-reporter-gene constructs persists during differentiation of stably transfected 3T3 cells

A series of stable 3T3-30A5 cell lines was constructed harbouring chimaeric FAS-CAT genes with and without the first intron.

Figure 1 Analysis of FAS-CAT mRNA by RT-PCR

3T3-30A5 pre-adipocytes were transiently transfected with FAS-CAT fusion plasmids and the resultant mRNAs were reverse transcribed and amplified (RT-PCR) as described in the Material and methods section. The size of the amplified products was ascertained by electrophoresis on a 2% agarose gel. Lane M, 1 kb DNA ladder (Gibco BRL, Gaithersburg, MD, U.S.A.); lane 1, untransfected cells; lane 2, cells transfected with —249/+68 construct; lane 3, cells transfected with —121/+1272 (reverse +405/+1083) plasmid; lane 4, cells transfected with —249/+1272 construct; lane 5, product of standard PCR using —249/+68 DNA as a control. The sizes of the amplified products were as predicted: lanes 2 and 5, 334 nt; lanes 3 and 4, 341 nt.

These cells were then induced to differentiate so that expression of the endogenous FAS gene was increased [2]. Although CAT expression also increased during differentiation in cells harbouring either the intron-containing or the intronless reporter gene, the level of expression was always 7–12-fold higher in cells carrying the intronless constructs than in those carrying the corresponding intron-containing constructs (Table 2). Thus the negative effect of the FAS intronic sequences on reporter gene expression can be demonstrated in stably transfected as well as transiently transfected cells, and persists even when expression of the endogenous FAS gene is activated during adipocyte development.

The effect of intronic sequence on gene expression is not an artifact caused by anomalous splicing

In order to evaluate the possibility that inclusion of intronic sequence in the chimaeric reporter gene might affect CAT expression indirectly, perhaps through improper splicing of the transcribed mRNA, we examined the effect of engineering a large deletion, Δ +295/+1083, from the middle of the intron in the context of the —249/+1272 construct. CAT expression in cells transfected with the deletion constructs (13.5 $\pm$ 1.6% chloramphenicol acetylated/h) was approx. 3-fold higher than in cells transfected with the parental construct (4.1 $\pm$ 0.9%) and was of similar magnitude to that observed in cells transfected with the intronless —249/+68 construct (16.7 $\pm$ 1.8%). Since both the splice donor and acceptor sequences are identical in the constructs containing partial and intact intron sequences, the difference in reporter gene expression could not be attributed to a defective splicing mechanism, but rather must have resulted from the presence of certain internal intronic sequences between nt +295 and +1083. Furthermore, when the mRNAs resulting from transfection of 3T3-30A5 cells with various intron-containing or intronless constructs were examined by coupled RT-PCR, we detected only an mRNA species with a size exactly matching that expected from a proper splicing event (Figure 1).

Figure 2 Nuclear run-on transcription assay

(A) Labelled transcripts synthesized using nuclei isolated from livers of fasted/re-fed (lane 1) and fasted (lane 2) rats and spleens of fasted/re-fed rats (lane 3) were hybridized to various DNA probes immobilized on nitrocellulose filters (10 μ g of denatured DNA/slot). The probes used were: A, genomic clone, -120/+1272 (includes exon 1 and intron 1); B-G, cDNA clones; B, 66-693 (from exons 1-4); C, 642-2544 (from exons 4-15); D, 2544-4858 (from exons 15-27); E, 2544-7615 (from exons 15-43); F, 6281-7615 (from exons 36-43); G, 66-7615 (full-length cDNA). The genomic and cDNA numbering is according to Amy et al. [7,27]. (B) Quantitative analysis of the hybridization signal in relation to the size of the probe. Densitometric analysis of the autoradiographs was performed using NIH Image software. The signal obtained from a full-length FAS cDNA probe (probe G) was taken as 100%. The signals from different tissues were normalized using an actin probe.

Run-on transcription assays

To determine whether the intronic sequences exert their negative effect on transcription at the level of initiation or elongation, we performed a series of nuclear run-on assays. Intact nuclei were isolated from liver, in which FAS expression is regulated according to nutritional status, and from spleen, in which FAS is barely detectable at all times. The nascent RNAs were labelled using [32 P]rGTP and hybridized either to a full-length FAS cDNA probe or to shorter probes that collectively spanned the entire coding region of the rat FAS gene as well as the first exon and first intron (Figure 2). In this assay, the hybridization signal associated with each probe reflects the distribution of RNA polymerase at various points on the FAS gene defined by the specific probes. The strongest hybridization signals were obtained using liver nuclei from fasted/re-fed animals; in this case, each of the FAS probes gave a hybridization signal that was approximately proportional to the probe size and guanine nucleotide content, indicating that, at the time the nuclei were isolated, the

Figure 3 DNase I footprint analysis of the FAS intronic region containing the putative E-box

A radiolabelled probe corresponding to FAS nucleotides -23 to +405 was incubated with nuclear extracts from H4IIE cells or 3T3-30A5 pre-adipocytes, then digested with DNase I as described in the Materials and methods section. Lanes 1, 5 and 9 contained no extract, lanes 2, 3 and 4 contained 15 μ g, 30 μ g and 45 μ g respectively of H4IIE nuclear extract; lanes 6, 7 and 8 contained 15 μ g, 30 μ g and 45 μ g respectively of 3T3-30A5 pre-adipocyte nuclear extract.

RNA polymerase complexes were randomly distributed along the gene. However, in the case of nuclei isolated from spleen and from the livers of fasted rats, only the full-length and long FAS probes gave clearly detectable hybridization signals. Based on the hybridization signal generated by the full-length probe, we calculated that the rate of FAS gene transcription in liver from fasted/re-fed rats is 13-fold higher than in liver from fasted animals and 24-fold higher than in spleen. Importantly, the probes representing exon 1 and intron 1 (probe A) and exons 1-4 (probe B) did not detect any radiolabelled RNA. Thus it is unlikely that low transcription of the gene results from an elongation block in the first intron.

Upstream stimulatory factor (USF) binds to a site within the first intron

Analysis of the nucleotide sequence of the deleted intronic fragment (nt +295 to +1083) for possible regulatory elements revealed an E-box sequence motif (CACGTG) at nt +292 to +297. To ascertain whether this sequence element might play a role in the regulation of transcription, we first determined whether proteins present in nuclear extracts were able to bind to this region of the gene. A DNA probe encompassing FAS nt sequences -23/+405, end-labelled at the +405 position, was found to be protected from DNase I digestion in the presence of nuclear extracts prepared from H4IIE cells or 3T3-30A5 pre-adipocytes. A distinct footprint was observed encompassing nucleotides +285 to +303, coinciding exactly with the location of the E-box (Figure 3).

Confirmation of the ability of nuclear proteins to recognize the intronic E-box was sought using electrophoretic mobility shift assays. The mobility of a radiolabelled probe encompassing FAS sequences +283/+305 and carrying the E-box core sequence at

Figure 4 Detection of proteins binding to the FAS intronic region +285/+305 by the electrophoretic mobility shift assay

A radiolabelled DNA oligonucleotide corresponding to the FAS promoter region +285 to +305 was used as probe. Lane 1, free probe alone; lane 2, 10 μ g of 3T3-30A5 pre-adipocyte nuclear extract; lane 3, as lane 2 but with a 100-fold molar excess of unlabelled +285/+305 oligonucleotide; lane 4, as lane 2 but with a 100-fold molar excess of an unlabelled oligonucleotide representing the authentic adenovirus USF-binding site; lane 5, as lane 2 but with 2 μ g of anti-USF antibody added before addition of the probe; lane 6, as lane 5 except that the antibody was added 20 min after the addition of the probe; lane 7, 10 ng of purified USF; lane 8, as lane 2 but with a 100-fold molar excess of an unlabelled oligonucleotide containing the consensus sequence for Sp1; lane 9, 10 μ g of heat-treated H4IIE cell nuclear extract.

its centre was retarded to a similar extent in the presence of nuclear extracts prepared from either H4IIE or 3T3-30A5 cells; the formation of the protein–DNA complexes was unaffected by heating the nuclear extracts at 65 °C for 20 min (Figure 4, lane 9). The heat-stable nature of these DNA-binding factor(s), together with the fact that their recognition sequence perfectly matched that of USF, prompted us to examine the possibility that the protein binding to this region is indeed USF. We verified this hypothesis in three different ways. First, an unlabelled double-stranded oligonucleotide bearing the USF-binding site of adenovirus major late promoter competed effectively with the radiolabelled probe for binding to the protein(s) present in the pre-adipocyte nuclear extract (Figure 4, lane 4). Secondly, when rabbit polyclonal anti-USF antibodies (Santa Cruz Biotechnology, Santa Cruz, CA, U.S.A.) were incubated with the nuclear extracts, either prior to or after the addition of the radiolabelled probe (Figure 4, lanes 5 and 6 respectively), the mobility of the protein–DNA complex was retarded. Finally, when USF was used as the protein source, a radiolabelled complex was formed with the same electrophoretic mobility as the complexes formed using the nuclear extracts (Figure 4, lane 7).

These experiments identified USF as the factor present in the nuclear extracts of H4IIE and 3T3-30A5 cells that binds to the E-box in the first intron of the FAS gene. To assess the functional significance of the interaction between the FAS E-box and USF, an oligonucleotide identical to the FAS intronic E-box and its flanking sequences was cloned, in both orientations, upstream of the TK basal promoter in the TK-CAT plasmid. Plasmid TK-CAT was chosen as the vector for this experiment because the TK basal promoter directs low-to-moderate CAT expression in 3T3-30A5 cells that conveniently permits detection of either positive or negative modulation of transcription. Insertion of a single copy of the FAS intronic sequence in either orientation conferred a markedly higher activity on the TK basal promoter,

Table 3 Effect of the FAS intronic USF-binding site on promoter activity in 3T3-30A5 pre-adipocytes

(1) Using a heterologous promoter. An oligonucleotide corresponding to the USF-binding site of the FAS first intron, FAS nt +285 to +305, was cloned in front of the TK promoter in the TK-CAT plasmid, giving rise to three different constructs: pUSF(+1) and pUSF(+2) containing one and two copies respectively of the insert in the same orientation as that of FAS, and pUSF(rev) with one copy of the insert in the reverse orientation. Equal amounts of these three constructs, as well as the TK-CAT vector, were then separately transfected and the resulting CAT activity was determined 48 h later. (2) Using the FAS promoter. The USF-binding site was either inactivated by insertion of an 8 bp linker between nt +294 and +295 (USF mutant) or removed by making a 80 bp deletion spanning that region (Δ +257/+337). Values are means $\pm$ S.D. of four independent experiments.

Promoter used	Construct used	Chloramphenicol acetylated (%/h)
TK	TK-CAT	5.3 $\pm$ 0.7
	pUSF(+1)	13.2 $\pm$ 1.5
	pUSF(+2)	21.3 $\pm$ 2.8
	pUSF(rev)	12.9 $\pm$ 1.5
FAS	–450/+68	13.8 $\pm$ 1.0
	–450/+1272	3.3 $\pm$ 0.4
	–450/+1272 (USF mutant)	3.7 $\pm$ 0.2
	–450/+1272 (Δ +257/+337)	4.8 $\pm$ 0.5

and the presence of two tandem copies of the insert resulted in an even greater stimulatory effect (Table 3). However, when the USF-binding site was either deleted or insertionally disrupted in the context of the FAS–CAT construct, no effect on FAS promoter activity was observed (Table 3). These experiments revealed that, although the intronic USF-binding element within the first intron is capable of acting as an enhancer when translocated to a heterologous promoter, in the context of the FAS promoter this effect appears to be masked.

Sequences between positions +405 and +1083 mediate the negative regulation

None of the results of studies on the intronic USF-binding site were consistent with a role for this element in the negative effect on transcription ascribed to the first intron. Indeed, when nt +405 to +1083 were deleted from the intron-containing FAS–CAT construct, leaving the putative E-box sequence element intact, the negative effect of the intronic sequence was abolished, as had been observed previously with the +295/+1083 deletion (results not shown). Clearly, then, the intronic region responsible for the negative effect on transcription is independent of the E-box and is located between nt +405 and +1083.

FAS first intron can down-regulate a heterologous promoter

The possibility that the intronic negative element could function out of its normal context was assessed by cloning FAS nt +405 to +1083 both upstream and downstream of heterologous promoters. Insertion of this intronic sequence upstream of the TK and SV40 promoters had no effect on the expression of the associated reporter gene (results not shown). Indeed, neither did the intronic sequence act as a negative element when cloned upstream of the FAS promoter in the –249/+68 FAS–CAT construct (results not shown). These results suggested that these sequences cannot affect transcription when placed upstream of transcriptional units and prompted us to examine the effect of placing the sequence element downstream of a heterologous promoter. The strategy for construction of an appropriate vector involved engineering a unique restriction site downstream of the

Table 4 Orientation-dependence of the negative regulatory element of the FAS intron

3T3-30A5 pre-adipocytes or H4IIE cells were transiently transfected with the plasmids indicated. Construct +124/+68 lacked intronic sequences; construct -121/+1272 contained the intact first intron. Two intron-containing constructs also had an internal region of the intron (+405 to +1083) flipped in the reverse orientation. In one construct the USF-binding site was disrupted by insertion of an 8 bp linker between nt +294 and +295. Values are means $\pm$ S.D. of four independent experiments.

Construct used	Chloramphenicol acetylated (%/h)	
	3T3-30A5	H4IIE
-124/+68	10.5 $\pm$ 1.8	2.9 $\pm$ 0.3
-124/+1272	2.3 $\pm$ 0.9	1.8 $\pm$ 0.2
-121/+1272 (+405 to +1083 in reverse orientation)	7.8 $\pm$ 0.3	2.8 $\pm$ 0.5
-121/+1272 (+405 to +1083 in reverse orientation, USF site disrupted)	5.1 $\pm$ 0.4	1.0 $\pm$ 0.1

transcription start site but upstream of the translation initiation codon of a CAT reporter gene driven by the SV40 promoter and enhancer sequences, pCAT-Control (Promega). This site was then used to clone a PCR-amplified fragment containing the intact intron and its splicing signals. The resulting reporter gene construct contained two new exons, one carrying the 5'-untranslated sequences and the other the protein coding region. The new reporter gene, when transiently transfected into 3T3-30A5 cells, resulted in drastically lower CAT expression than did the intronless control reporter gene (14.7 $\pm$ 1.0 and 57.8 $\pm$ 5.1 % chloramphenicol acetylated/h respectively). To ensure that the effect was not caused by aberrant splicing of the FAS intron from the reporter gene, FAS nt +405 to +1083 were deleted from the FAS-intron-containing pCAT-Control plasmid and transcriptional activity was reassessed. The level of reporter gene expression produced from this vector containing the partial intron (56.9 $\pm$ 3.0 % chloramphenicol acetylated/h) was restored to the level exhibited for the intronless control, as we had observed when the same deletion was made from the intron in the context of the FAS promoter. These results established that sequences within the +405 to +1083 region of the FAS first intron, when placed downstream of the promoter, can shut down transcription from the strong SV40 promoter as well as from the FAS promoter.

The regulatory element is orientation-dependent

To evaluate the possible orientation-dependence of the negative regulatory element, a CAT reporter gene driven by the FAS promoter and containing FAS nt -121/+1272 was constructed with nt +405 to +1083 inverted. The levels of CAT expression obtained with the original and the deletion plasmids were compared in transient transfection experiments using 3T3-30A5 and H4IIE cells. The results showed that the negative effect of the intronic sequences was largely abolished when their orientation with respect to the promoter was reversed (Table 4).

Disruption of the negative effect of the first intron allows transcriptional activation via the intronic E-box

Earlier attempts to define a role for the E-box located at nt +292 to +297 had concluded that any positive effects of this sequence element were probably masked by the presence of the negative element further downstream in the first intron. This

Figure 5 Effect of various deletions in the FAS first intron on reporter gene expression

3T3-30A5 pre-adipocytes were transiently transfected with the deletion mutants shown. The deletion end-points are denoted by numbers within the boxes, and the location of the FAS-CAT fusion junction is denoted by the numbers outside the boxes.

possibility was re-investigated utilizing the FAS-CAT chimaeric gene in which the negative effect on transcription has been eliminated by flipping the orientation of FAS nt +405/+1083. Disruption of the E-box in this construct by an 8 bp mutational insertion resulted in a decrease in reporter gene expression, using either 3T3-30A5 preadipocytes or H4IIE cells (Table 4). Furthermore, if sequences between nt +405 and +1083 were deleted in the context of a -121/+1272 construct with the wild type E-box, the resulting CAT activity was about 40 % higher than that obtained with the same plasmid bearing the insertionally mutated E-box (6.68 $\pm$ 1.06 compared with 4.12 $\pm$ 0.73 % acetylated chloramphenicol/h respectively). These results indicate that the E-box can contribute to enhanced transcription in the context of the FAS promoter, provided that the negative effect of the FAS first intron is abolished.

Deletion analysis of the negative regulatory region

In an attempt to pinpoint the DNA sequences responsible for the negative effect of the first intron on FAS transcription, the effects of deleting several different regions of the first intron were evaluated in 3T3-30A5 cells (Figure 5). Deletion of nt +405 to +819, nt +662 to +860 or nt +819 to +1083 eliminated much of the negative effect of the intronic sequence, suggesting that each of these three deleted regions contain sequence elements that contribute to the negative effect on transcription. On the other hand, deletion of either nt +768 to +924 or nt +787 to +840 had no effect on transcription, suggesting that nucleotides within these deleted regions do not contribute to the negative effect. Taken together, the data suggest that two separate sequence elements contribute to the negative effect on transcription, one lying approximately between nt +405 and +768 and the other approximately between nt +924 and +1083.

A number of tissue-specific nuclear proteins bind to the FAS first intronic sequences

Intronic sequence elements between nt +405 and +1083 that might be capable of binding nuclear proteins were mapped by

Figure 6 DNase I footprint analysis of the FAS first intron

Two end-labelled probes were used to search for nuclear-protein-binding sites in the first intron of the FAS gene. (A) Probe +405/+819: lane 1, G+A reaction; lanes 2 and 6, free probe; lane 3, 50 μ g of nuclear extract from livers of fasted rats; lane 4, 50 μ g of nuclear extract from livers of fasted/re-fed rats; lane 5, 50 μ g of nuclear extract from spleens of fasted/re-fed rats. (B) Probe +819/+1083: lanes 1 and 5, free probe; lane 2, 50 μ g of nuclear extract from livers of fasted rats; lane 3, 50 μ g of nuclear extract from livers of fasted/re-fed rats; lane 4, 50 μ g of nuclear extract from spleens of fasted/re-fed rats.

DNase I footprinting. Liver nuclear extracts prepared from rats that had been either fasted or fasted and then re-fed a fat-free diet protected four distinct regions, nt +696 to +745, nt +757 to +789 (Figure 6A), nt +870 to +888 and nt +994 to +1023 (Figure 6B). However, nuclear extracts prepared from spleen offered little or no protection to any of these regions. This result indicated that several tissue-specific nuclear proteins are capable of binding to the first intron of the FAS gene. We also detected binding of nuclear proteins present in extracts prepared from 3T3-30A5 cells, both pre- and post-differentiation, to probes representing the footprinted regions by gel shift analysis (results not shown).

DISCUSSION

The negative effect of the FAS intron on reporter gene expression conceivably could be mediated at the level of RNA processing rather than at the transcriptional level. We therefore evaluated several alternative explanations for the effects observed as a result of the inclusion of intronic sequences in the reporter gene constructs. One possibility considered was that the intronic sequences might not be properly spliced out or that alternative splicing of reporter gene transcripts might generate mRNA species that were translated with different efficiencies. Alternatively, the introduction of a long intron into the reporter gene might hinder release of the mature mRNA from the nucleus by introducing a bottleneck at the RNA processing stage. However, each of these possible explanations can be discounted for the following reasons. First of all, when 3T3 cells were transfected with FAS-CAT reporter genes containing or lacking the FAS first intron, only one mRNA species, of the appropriate size in both cases, could be detected by RT-PCR. Secondly, flipping the

orientation of specific internal intronic sequences, leaving the splice donor and acceptor sites intact, eliminated the negative effect on reporter gene expression; again, only one mRNA species could be detected by RT-PCR, which had the same size as the mRNA produced from the construct containing the intact intron. These results indicate that the effect of the presence of the first intron on reporter gene expression cannot be attributed to the problem of splicing out of a long intron. Indeed, results of the internal deletion experiments implicate, as mediators of the negative effect on reporter gene expression, specific regions of the first intron located approximately between nt +405 and +768 and between nt +924 and +1083.

Although, in eukaryotes, regulation of transcription is commonly achieved at the level of initiation, a number of genes have been identified in which transcription is regulated at the elongation stage. In these genes the presence of specific 'attenuator' sequences, usually located near the exon 1/intron 1 boundary, appears to induce a pausing of the polymerase that can lead to the premature termination of transcription and the production of short, unstable RNA transcripts [28]. For many of the genes exhibiting transcriptional pausing, elongation beyond the first exon is governed by metabolic, tissue-specific and/or developmental signals [28,29]. Read-through of the pause sites in these genes may be influenced by recruitment of specific transactivator proteins to the promoter through enhancer elements or by specific elongation factors that directly modify the properties of RNA polymerase II at sites within a gene [30]. No clear consensus exists as to the nature of these attenuator sequences, possibly because different attenuation sites may promote pausing by different mechanisms. Among the various mechanisms that have been suggested are steric hindrance to polymerase progress, caused either by the association of proteins at specific sites or by local perturbances in DNA structure, and the formation of stable secondary structural elements in the RNA transcripts. Because of their very nature, these attenuator sequences operate only when located downstream of a promoter, and typically they function in an orientation-dependent, promoter-independent fashion [31]. The properties of the FAS intronic negative element(s) are consistent with these characteristics: that is, it operates in a promoter-independent manner, but only when positioned downstream of the promoter in the correct orientation. Furthermore, within the region of intron 1 implicated in the down-regulation of the FAS gene, between nt +449 and +465, is a GC-rich sequence (GAGGCCCCAGGGCCTC) capable of forming a stable RNA hairpin. However, deletion of this sequence from intron-containing FAS-CAT reporter gene constructs had no effect on transcription (B. Oskouian, unpublished work). It is unlikely, therefore, that this sequence mediates a transcriptional arrest that results in lower expression of an associated reporter gene.

As we completed the present study, Hsu et al. [32] reported the discovery of a negative regulator of transcription within the first intron of the human FAS gene. The human gene exhibits 90% sequence identity with the rat gene in the promoter region, marked by the presence of a TATA box, an inverted CCAAT box and a functional Sp1 binding site. Hsu et al. [32] found evidence for the presence of a second, very weak, promoter in the first intron of the human gene and suggested that this intronic promoter may exert a roadblock for the polymerase that has initiated transcription at the strong promoter site in exon 1. However, the presence of a second intronic promoter in the rat gene seems unlikely, based on the results of primer extension experiments with mRNA isolated from H4IIE cells. Thus a primer complementary to the rat DNA sequence downstream of the translational start site in exon 2 was extended to give a single

Figure 7 Map of regulatory elements in the promoter and first intron regions of the rat FAS gene

Ex1, exon 1; Ex2, exon 2; NRE, negative regulatory element; I-NRE, intronic negative regulatory element (approximate location); Sp1, binding site for the transcription factor Sp1; CCAAT, inverted CCAAT box; EBOX, insulin response element; TATA, TATA box. The transcription (nt +1) and translation (ATG) start sites are shown.

product corresponding exactly in length to that predicted from reverse transcription of an mRNA that included sequences representative of the entire first exon (results not shown). In practice, the second-promoter hypothesis proposed by Hsu et al. [32] predicts the same accumulation of short transcripts that one would expect from the pausing mechanism discussed above. However, using the nuclear run-on assay, we concluded that the low level of FAS expression observed in spleen and in the livers of fasted rats could be accounted for by a low level of initiation of FAS transcription, and we found no evidence for the accumulation of the short transcripts predicted by a pausing mechanism (Figure 4). With the caveat that short nascent transcripts might be too unstable for detection by the assay, our results do not support the operation of a transcriptional pausing mechanism within the first intron. Nevertheless, it is possible that a pausing mechanism may operate under specific metabolic conditions that we have not reproduced in the nuclear run-on experiments. Conceivably, the human FAS gene may be regulated in a different manner than that of the rat. Indeed, we can find no evidence for the operation of an intronic promoter in the rat gene, and it may be pertinent that the region of the rat gene that we have implicated in the negative effect on transcription (nt +405 to +1083) is relatively poorly conserved in the human gene, exhibiting only 34% sequence identity, compared with 90% identity in the proximal promoter region.

An alternative explanation for the observed phenomenon is that the first intron contains *cis*-regulatory elements that directly down-regulate initiation of transcription at the promoter site. Although relatively uncommon, such a situation would not be unique, since both positive and negative regulators of transcription initiation, as well as those conferring hormonal responsiveness and tissue specificity, have been identified within intronic sequences of other genes. In some instances a number of elements with opposing effects have been identified within the same intron. For example, Telliez et al. [33] have recently reported the presence of five footprinted regions in the first intron of mouse *Ha-ras* gene; two of these sites appear to be associated with enhancer-like elements and one with a negative-acting element. In another study, Tourmente et al. [34] found that 360 bp of the first intron of the *Drosophila* $\beta 3$ tubulin gene are required for the regulation of the gene by ecdysone. This intronic region too was shown to contain both positive- and negative-acting regulatory sequences. Another *Drosophila* gene, tropomyosin I, is under the control of at least two regulatory regions within the first intron of the gene [35]. These regions contain multiple muscle-specific positive and negative *cis*-elements which together contribute to both temporal and tissue-specific expression of the tropomyosin gene. Although our finding that the negative regulatory element of the FAS intron is effective only when positioned downstream of a promoter could be

explained by a transcriptional pausing mechanism, it is also not inconsistent with an effect at the level of transcription initiation. For example, Wang et al. [36] have reported the presence of a negative regulatory element in the first intron of the A-chain gene for human platelet-derived growth factor that is functional when located downstream, but not upstream, of a reporter gene driven by either the homologous promoter or a heterologous promoter. In another report, Hanke et al. [37] showed the presence of a both position- and orientation-dependent negative element in the last intron of the human CD8 α gene. Our results obtained with the run-on assays appear consistent with a mechanism whereby the sequences within the FAS first intron exert their negative effect at the level of initiation of transcription.

Our findings indicate that the FAS first intron contains, in addition to the negatively acting element, a second discrete element capable of mediating positive control on transcription. This positive element, located at nt +292 to +297, appears to mediate transcriptional activation through interaction with USF. During the course of our work the presence of this USF-binding site in the FAS first intron was also reported by Foufelle et al. [38]. These authors found that the location of the USF-binding site coincided with the presence of a site of hypersensitivity to DNase I present in liver, but not spleen, and suggested that the sequence confers glucose-responsiveness on the FAS gene; however, they did not assess the functional activity of the USF-binding site, either in or out of the context of the first intron. Our results indicate that this USF-binding site can mediate activation of basal transcription in the context of the FAS, or a heterologous promoter only when the effect of the downstream negative intronic element(s) is negated, and suggest that regulation ultimately depends on a complex interplay between elements in the first intron as well as the upstream promoter region.

The physiological role of the negative element of the first intron remains largely a matter of speculation at present. Neither our results nor those of Hsu et al. [32] offer an explanation as to the physiological significance of the effect of the intronic sequences on transcription. However, the observation that spleen, a tissue that does not actively transcribe the FAS gene, apparently is deficient in proteins that bind to the putative intronic regulatory sequences suggests that perhaps these sequence elements could play a role in the tissue-specific expression of the gene. Studies designed to test this hypothesis are presently under way.

Several sequence elements in the proximal promoter region and first intron have now been implicated in the basal and regulated transcription of the rat FAS gene, and these are summarized in Figure 7. Initiation of transcription from a single start site is enhanced by the presence of an adjacent TATA motif, an inverted CCAAT box and an upstream binding site for the transcription factor Sp1; further modulation of transcription may be achieved through interactions between these positive-

acting elements and negative-acting elements located at positions -319 to -301 and in the first intron [13]. The increased transcription of hepatic FAS that follows ingestion of a high-carbohydrate diet is regulated by insulin, and the insulin response element has been identified as an E-box located at positions -68 to -60 [14,15]. The insulin effect appears to be mediated by USF, which can bind to the -68/-60 E-box [39]. The E-box located in the first intron can also bind USF and may be responsible in part for the glucose-responsiveness of the gene (the present paper; [38]). Antagonism of the insulin effect by cAMP appears to involve the inverted CCAAT box located at -98/-92 [15]. The repression of FAS transcription by sterols is thought to require both a sterol response element, located at -150/-141, and the E-box located at -68/-60 [40]. Polyunsaturated fatty acids can also repress FAS transcription, possibly by antagonism of the insulin response mediated through the E-box at -68/-60; however, it is not yet clear whether this sequence constitutes a polyunsaturated fatty acid response element or whether these fatty acids inhibit the signalling pathway from the insulin receptor to the FAS gene [41]. Clearly, further work will be required to unravel details of the complex interplay of factors that regulate expression of this gene.

This work was supported by grant DK42636 from the National Institutes of Health. We thank Dr. Fattah Kashanchi (National Institutes of Health) for generously donating recombinant USF, and Dr. Jürgen Naggert and Dr. Gordon Watson for their critical reading of the manuscript.

REFERENCES

- Smith, S. and Ryan, P. (1979) *J. Biol. Chem.* **254**, 8932-8936
- Moustaid, N. and Sul, H. S. (1991) *J. Biol. Chem.* **266**, 18550-18554
- Pope, T. S., Smart, D. A. and Rooney, S. A. (1988) *Biochim. Biophys. Acta* **959**, 169-177
- Alberts, A. W. and Greenspan, M. D. (1984) in *Fatty Acid Metabolism and its Regulation* (Noma, S., ed.), pp. 29-58, Elsevier, Amsterdam
- Leveille, G. A. (1967) *J. Nutr.* **91**, 25-34
- Schwartz, R. S. and Abraham, S. (1982) *Biochim. Biophys. Acta* **711**, 316-326
- Amy, C. M., Williams-Ahlf, B., Naggert, J. and Smith, S. (1990) *Biochem. J.* **271**, 675-679
- Paulauskis, J. D. and Sul, H. S. (1989) *J. Biol. Chem.* **264**, 574-577
- Swierczynski, J., Mitchell, D. A., Reinhold, D. S., Salati, L. M., Stapleton, S. R., Klautsky, S. A., Struve, A. E. and Goodridge, A. G. (1991) *J. Biol. Chem.* **266**, 17459-17466
- Iritani, N. (1992) *Eur. J. Biochem.* **205**, 433-442
- Moustaid, N., Sakamoto, K., Clarke, S., Beyer, R. S. and Sul, H. S. (1993) *Biochem. J.* **292**, 767-772
- Kameda, K. and Goodridge, A. G. (1991) *J. Biol. Chem.* **266**, 419-426
- Oskouian, B., Rangan, V. S. and Smith, S. (1996) *Biochem. J.* **317**, 257-265
- Moustaid, N., Beyer, R. S. and Sul, H. S. (1994) *J. Biol. Chem.* **269**, 5629-5634
- Rangan, V. S., Oskouian, B. and Smith, S. (1996) *J. Biol. Chem.* **271**, 2307-2312
- Park, K. and Kim, K.-H. (1991) *J. Biol. Chem.* **266**, 12249-12256
- Kingston, R. E., Chen, C. A. and Okayama, H. (1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 9.1.1-9.1.3, John Wiley & Sons, New York
- Southern, P. J. and Berg, P. (1982) *J. Appl. Gen.* **1**, 327-341
- Kingston, R. E. and Sheen, J. (1995) in *Current Protocols in Molecular Biology*, vol. 1 (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 9.7.1-9.7.11, J. Wiley & Sons, New York
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, pp. 16.66-16.67, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Gorski, K., Carneiro, M. and Schibler, U. (1986) *Cell* **47**, 767-776
- Dignam, J. D., Lebovitz, R. M. and Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475-1489
- Appel, B. and Sakonju, S. (1993) *EMBO J.* **12**, 1099-1109
- Greenberg, M. E. (1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 4.10.1-4.10.8, John Wiley & Sons, New York
- Henikoff, S. (1984) *Gene* **28**, 351-359
- Kunkel, T. A., Roberts, J. D. and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367-382
- Amy, C., Witkowski, A., Naggert, J., Williams, B., Randhawa, Z. and Smith, S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3114-3118
- Wright, S. (1993) *Mol. Biol. Cell* **4**, 661-668
- Spencer, C. A. and Groudine, M. (1990) *Oncogene* **5**, 777-785
- Krumm, A. and Groudine, M. (1995) *Science* **269**, 1400-1401
- Bentley, D. L. and Groudine, M. (1988) *Cell* **53**, 245-256
- Hsu, M. H., Chirala, S. S. and Wakil, S. J. (1996) *J. Biol. Chem.* **271**, 13584-13592
- Telliez, J. B., Plumb, M., Balmain, A. and Bailleul, B. (1995) *Mol. Carcinogen.* **12**, 137-145
- Tourmente, S., Chapel, S., Dreau, D., Drake, M. E., Bruhat, A., Couderc, J. L. and Dastugue, B. (1993) *Insect Biochem. Mol. Biol.* **23**, 137-143
- Gremke, L., Lord, P. C., Sabacan, L., Lin, S. C., Wohlwill, A. and Storti, R. V. (1993) *Dev. Biol.* **159**, 513-527
- Wang, Z. Y., Masaharu, N., Qiu, Q. Q., Takimoto, Y. and Deuel, T. F. (1994) *Nucleic Acids Res.* **22**, 457-464
- Hanke, J. H., Hambor, J. E. and Kavathas, P. (1995) *J. Mol. Biol.* **246**, 63-73
- Foufelle, F., Lepetit, N., Bosc, D., Delzenne, N., Morin, J. and Raymondjean, M. (1995) *Biochem. J.* **308**, 521-527
- Wang, D. and Sul, H. S. (1995) *J. Biol. Chem.* **270**, 28716-28722
- Bennett, M. K., Lopez, J. M., Sanchez, H. B. and Osborne, T. F. (1995) *J. Biol. Chem.* **270**, 25578-25583
- Fukuda, H., Iritani, N., Katsurada, A. and Noguchi, T. (1996) *Biochem. Mol. Biol. Int.* **38**, 987-996