Roles of an Ets motif and a novel CACGAC direct repeat in transcription of the murine dihydrolipoamide dehydrogenase (*Dld*) gene

Hsin-Sheng YANG*, Mark JOHNSON† and Mulchand S. PATEL*¹

*Department of Biochemistry, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY 14214, U.S.A., and †Departments of Biochemistry and Genetics, Case Western Reserve University, School of Medicine, Cleveland, OH 44106, U.S.A.

The 5'-flanking region of the murine dihydrolipoamide dehydrogenase (*Dld*) gene was characterized for its promoter activity. DNase I footprinting analysis of the promoter region (–545 bp to +41 bp) revealed six major protein-binding domains (termed P1 to P6) that were protected by NIH3T3 fibroblast nuclear extracts. Transient transfection assays, using a series of nested deletions of the 2.5 kb 5'-flanking region ligated to the chloramphenicol acetyltransferase reporter gene, identified that the –42-bp to +41-bp region, which harbours the P1, P2, and P3 domains, had minimal transcriptional activity. When the 5'-flanking region was extended from –42 bp to –82 bp, there was an approx. 5-fold increase in promoter activity. To identify further the *cis* elements involved in transcription of the *Dld* gene (–82 bp to +41 bp), a series of mutations were introduced into this region and evaluated for functional effects using transient transfection and electrophoretic mobility shift assays. Mutation

or deletion of the CACGAC direct repeat, located from –61 bp to –46 bp, resulted in minimal promoter activity. Mutation of the Ets motif, located from –37 bp to –32 bp, reduced the minimal promoter activity by approx. 50%, whereas the deletion of this motif almost abolished the promoter activity. These results indicate that: (i) the Ets motif is required for the minimal promoter activity and (ii) the CACGAC direct repeat enhances promoter activity. Database searches failed to identify the direct repeat with the CACGAC motif and hence the CACGAC sequence may represent a novel motif. The requirement of both the Ets motif and the direct repeat element for optimal promoter activity represents a unique combination for gene transcription.

Key words: gene regulation, mitochondrial enzymes, promoter analysis.

INTRODUCTION

Dihydrolipoamide dehydrogenase (E3; EC 1.8.1.4) is a common component of the three 2-oxo acid dehydrogenase complexes, namely the pyruvate dehydrogenase complex, 2-oxoglutarate dehydrogenase complex and branched-chain 2-oxo acid dehydrogenase complex, as well as glycine synthase. These E3-dependent complexes participate in oxidative pathways of carbohydrate, lipid and amino acid metabolism. The catalytic mechanism of E3 involves the transfer of two electrons from reduced lipoyl moieties covalently linked to specific lysine residues of the dihydrolipoamide acyltransferases and H protein to NAD⁺ via a flavin intermediate [1–3].

During mammalian development, the activity of the pyruvate dehydrogenase complex in rat liver [4] and muscle [5] increases several-fold from birth to adulthood. This increase in pyruvate dehydrogenase complex activity is paralleled by increases in the activity of all three catalytic subunits, namely, pyruvate dehydrogenase, dihydrolipoamide acetyltransferase and E3. During *in vitro* differentiation of murine 3T3-L1 preadipocytes into adipocytes, the activity of the pyruvate dehydrogenase complex increases approx. 7-fold after the cells are treated with insulin, dexamethasone and 1-methyl-3-isobutylxanthine. The induction of pyruvate dehydrogenase and E3 activities in 3T3-L1 adipocytes is attributed to an increase in the rates of synthesis of these enzymes [6,7], which are regulated at the pretranslational level.

Most TATA-less promoters contain an initiator element which directs transcription initiation. One group of initiators with the consensus sequence 5'-(A/C)GGAA(G/A)-3', which was found

in the promoters of the murine cytochrome *c* oxidase subunit IV and Vb (*COXIV* and *COXVb*) genes, binds with GA binding protein (GABP), a member of the Ets family [8,9]. The Ets family contains more than 30 members, which may function as transcription activators, repressors or initiators, and which play important roles during embryogenesis, haematopoiesis and tumour metastasis [10].

Recently, the nucleotide sequence of the proximal promoter region (300 bp) of the murine *Dld* gene was determined [11]. Similarly to the human *DLD* gene [12,13], the promoter region of the murine *Dld* gene has multiple transcription initiation sites and lacks a TATA motif. The sequence of approx. 2 kb of the human *DLD* 5'-flanking region has been determined and shown to contain several consensus sequences for DNA binding proteins, including a glucocorticoid-receptor element, a cAMP-response element, a CAAT box, a fat-specific element and a nuclear-respiratory factor. The presence of these regulatory motifs suggests some regulatory features of this gene. Progressive deletions of the human *DLD* 5'-flanking region fused with the chloramphenicol acetyltransferase (CAT) reporter gene in transiently transfected HepG2 cells indicated a negative regulatory element in a segment from –1223 bp to –769 bp and a minimal promoter region (–19 bp to +47 bp) that supported significant CAT expression [12]. This short promoter sequence shows similarity (73%) to the murine genomic sequence [11]. However, the transcriptional regulation of human *DLD* is not well understood. As a stage to understanding *Dld* gene expression, we have cloned and characterized the murine *Dld* 5'-flanking region. The functional domains and nuclear protein binding sites in the

Abbreviations used: E3, dihydrolipoamide dehydrogenase; EMSA, electrophoretic mobility shift assay; CAT, chloramphenicol acetyltransferase; TBP, TATA-binding protein; GABP, GA-binding protein.

¹ To whom correspondence should be addressed (e-mail mspatel@acsu.buffalo.edu).

The nucleotide sequence reported in this article has been deposited in the GenBank[™] Data Library with accession number U73468.

proximal promoter region were evaluated. The requirement of two *cis*-elements for optimal promoter activity, an Ets motif (CGGAAA) and a direct repeat element with a novel CACGAC motif, has been identified.

MATERIALS AND METHODS

Genomic library screening and sequencing

A 9 kb *SalI*–*BamHI* murine genomic DNA fragment containing the 5'-flanking and -coding regions of the murine *Dld* gene was cloned from a λ EMBL3 mouse genomic library, as described previously [11]. A 3 kb fragment upstream of the coding region extending from a *SalI* site to a *BsaI* site within the 5' untranslated region (+41) was subcloned into a pUC19 vector (Life Technologies) and termed pPE37–3SB. Approx. 1 kb of the nucleotide sequence upstream of the transcription initiation site was determined by the dideoxy-chain termination method, using the Sequenase II kit (Amersham) with synthesized primers hybridizing to internal or flanking vector sequences.

Construction of plasmids

Plasmid pPE37–3SB was digested with *EcoRI* and the resulting 2.5 kb fragment containing the murine *Dld* 5'-flanking region was ligated into the *EcoRI* site of the promoterless CAT reporter plasmid, pBSKCAT [14], to produce the construct D1 (–2.5 kb to +41 bp) (see Figure 3). Restriction enzyme or exonuclease III (Erase-a-Base System; Promega) digestion produced a series of 5'-deletions of the *Dld* promoter upstream of the transcription initiation site, which was determined by murine liver mRNA [11]. The sizes of the various 5'-deletions were –2.0 kb (*BglII*), –1.4 kb (*EcoRV*), –740 bp, –573 bp, –483 bp, –362 bp, –242 bp (*HindIII*), –82 bp (*PvuII*) and –17 bp (*BglII*). The constructs D10 (–42 bp to +41 bp) and D13 (–25 bp to +41 bp) were generated by PCR with *pfu* polymerase (Stratagene), using primer pairs (mE3–42, T3) and (mE3–25, T3) respectively (mE3–42, 5'-CTCGAGGCGACCGGAAAGCCT-CTGCC-3'; mE3–25, CTCGAGCCTTTCGG GCCAATCGCG CTGC-3') using the following conditions: 95 °C for 30 s, 55 °C for 1 min and 68 °C for 2 min 30 s. The PCR products were digested by *XhoI* and *PstI* and subsequently cloned into the *XhoI* and *PstI* sites of pBSKCAT [14]. All constructs were verified by either restriction-enzyme analysis or dideoxyribonucleotide sequencing.

DNase I footprinting

Nuclear proteins were prepared from cultured NIH3T3 cells by a modification of the method described previously [15]. Briefly, cells were washed with ice-cold PBS, scraped from a 100-mm plate into 0.5 ml of Dignam's buffer A [16] and lysed in a glass Dounce homogenizer with 10–15 strokes. The nuclei were collected by centrifugation at 14000 *g* for 5 min, resuspended in 200 μ l of Dignam's buffer C [16] and shaken for 1 h at 4 °C. After re-centrifugation at 14000 *g* for 5 min, the supernatant was dialysed against 50 vol. of Dignam's buffer D [16] for 2 h at 4 °C, and stored at –80 °C.

DNase I footprinting reactions (50 μ l final volume) were carried out with nuclear proteins isolated from NIH3T3 cells (50–100 μ g), end-labelled DNA probe (20000 c.p.m.), 1 μ g of poly(dI-dC)(dC-dI) (Sigma) in 10 mM Hepes, pH 7.9/10% (v/v) glycerol/50 mM KCl/2% (v/v) poly(vinyl alcohol)/0.1 mM EDTA/0.5 mM PMSF/0.5 mM dithiothreitol. After incubation at room temperature for 20 min, 5 μ l of cofactor solution (100 mM $MgCl_2$ /30 mM $CaCl_2$) were added to the reaction mixture, followed by 5 μ l of RNase-free DNase I (0.01–1 unit;

Promega) and the incubation was continued for a further 2 min. The reaction was stopped by the addition of 150 μ l of 150 mM NaCl/0.1% (v/v) SDS/15 mM EDTA containing 30 mg/ml of yeast tRNA and DNA was extracted with phenol/chloroform (1:1, v/v) [17]. The sample was resolved on a denaturing 6% polyacrylamide/urea gel alongside a G + A ladder prepared by the Maxim and Gilbert method [18].

Electrophoretic mobility shift assay (EMSA)

The end-labelled DNA fragments were made either with [γ - 32 P]ATP (7000 Ci/mmol; ICN) using T4 polynucleotide kinase or [α - 32 P]ATP (3000 Ci/mmol; ICN) using the large fragment of DNA polymerase I. DNA–protein reactions for EMSA were carried out in volumes of 15 μ l containing 3–5 μ g of nuclear proteins isolated from NIH3T3 cells, 3–5 μ g poly(dI-dC)(dC-dI), 10 mM Hepes (pH 7.9), 10% (v/v) glycerol, 50 mM KCl, 0.1 mM EDTA and 1–2 fmol (10000 c.p.m.) of labelled probe. After incubation of this mixture at room temperature for 20 min, the products of the binding reactions were resolved on 4% polyacrylamide non-denaturing gels in 0.5 $\times$ 45 mM Tris/borate, pH 8.0/1 mM EDTA buffer. For competition EMSA, the reaction mixtures were preincubated with various concentrations of unlabelled competitor oligonucleotide before the addition of nuclear extracts.

Cell culture, transient transfection and CAT assay

The NIH3T3 cells were maintained in complete medium [90% minimal essential medium, 10% (w/v) fetal-bovine serum, 100 units/ml penicillin, 100 μ g/ml streptomycin and 2 mM glutamine] (Life Technologies) at 37 °C in a 5% CO_2 atmosphere, and cells at 50–60% confluence were used in transient transfection assays. Transient transfections were performed using the calcium phosphate co-precipitation method [19] with 15 μ g of test DNA and 5 μ g of β -galactosidase-expressive plasmid (pRSV- β -gal) [14]. After incubation for 6 h at 37 °C, the medium was replaced with complete medium containing 15% DMSO for 3 min, the cell monolayer was washed twice with PBS and then fresh complete medium was added. Cells were harvested after 48 h by washing twice with ice-cold PBS and then scraping the cells into 200–400 μ l of 20 mM potassium phosphate buffer (pH 7.5). Cells were lysed by sonication for 5–10 s (30% duty cycle, output control 3; Branson Sonifier 450), centrifuged at 14000 *g* for 10 min and an aliquot of the supernatant was assayed for β -galactosidase activity [20]. The remaining supernatant was heated at 65 °C for 10 min to inactivate the deacetylase and then assayed for CAT activity by the phase-extraction method [19] using [3 H]chloramphenicol as the substrate. The background level of CAT activity was subtracted. The background level of CAT expression is defined as CAT expression of pBSKCAT [14] in transiently transfected NIH3T3 cells, which is approx. 10% of CAT activity of the D1 construct.

Site-directed mutagenesis

A series of oligonucleotides with specific mutations were synthesized. Site-specific mutations were introduced using a Quick-Change™ site-directed mutagenesis kit (Stratagene) using D9 (–82 bp to +41 bp) or D10 (–25 bp to +41 bp) constructs as templates. After mutagenesis, the promoter fragments were cut with *XhoI* and *PstI* and cloned into the pBSKCAT plasmid [14]. All of the mutations were confirmed by nucleotide sequencing.

RESULTS

Cloning and sequencing the 5'-flanking region of murine *Dld* gene

In order to study the elements involved in *Dld* gene expression, the 5'-flanking region of the murine *Dld* gene was cloned. Approx. 300 bp of the 5'-flanking region of the murine *Dld* gene region have been reported previously [11]. The nucleotide sequence has now been extended to more than 1.0 kb (Figure 1). The sequence identity between the 1.0 kb murine *Dld* 5'-flanking region and the 1.0 kb human *DLD* 5'-flanking sequence [12] is approx. 52% using the GAP program from the Genetic Computer Group [21].

Nuclear protein binding domains of the murine *Dld* promoter

To identify protein binding domains, the murine *Dld* 5'-flanking region from -545 bp to +41 bp was analysed by DNase I footprinting with nuclear extracts isolated from NIH3T3 cells. The fragments of the coding strand of the murine *Dld* 5'-flanking region were end-labelled and partially digested by DNase I in the presence or absence of nuclear extracts (Figure 2). Six distinct footprinting domains on the coding strand were identified and termed P1 (+1 bp to +22 bp), P2 (-22 bp to -4 bp), P3 (-38 bp to -27 bp), P4 (-237 bp to -203 bp), P5 (-348 bp to -324 bp) and P6 (-392 bp to -375 bp). The boundaries of the protected domains differed slightly between coding and non-coding strands, with the exception of P3 and P5 domains (non-coding strand: P1, -1 bp to +24 bp; P2, -27 bp to -5 bp; P3,

		<u>GATA (R)</u>				
-1002	CTGCATACAT	AAATCATTAT	CTATGCTTTC	TCTAATGGCA	TAGATGGTAA	GAAGTAATAT
-942	TCATCTAATC	CCTAACTCTA	TGAGTTCAGT	ACAGTGCCTG	GGCTGTAAAC	TTTGTAGAAG
-882	TATAGAGTTA	AGGTTTAAAC	TAGAGACTGT	<u>TCF</u> TATTACAAAG	<u>GATA (R)</u> TGSTATCAGA	CCTTAATCTC
-822	CTGTGCTCAA	TTTTTTTTTT	TTTTTGTTTT	<u>GATA</u> TGTTTTGTGT	TTTCGAGATA	GGGAGCTCTG
-762	GCTGTCTTGG	AATCTACCTC	<u>AP2</u> GTAGACCAGG	CCTGGCCTGG	CCTCGAAGTG	AGAAATCTGC
-702	CTGCCTCTGC	CTCCCGAGTG	CTGGGATTAA	AGGCGTAGCC	CACCACGCCA	GGCTCTGTGC
-642	TCATCTATAG	<u>c-Myb</u> ACAACAGTTA	CGAATCTTTT	<u>AP1 (R)</u> AGGTCGCTCG	TCCTCAATGAA	AACCTAACAC
-582	CTCACCCACC	<u>AP2</u> ATCCCAACCC	TTCTAGTTTC	<u>NR-1L6</u> TTCTACAAGG	GAGGCGTGCT	TTCAAAATAC
-552	AAAAATAGAA	<u>p53</u> GGGAGCTTGT	TTTCTATACT	<u>GR (R)</u> TTTTGTAGTC	GCAGTTTTCA	<u>TR (R)</u> ELA-F GAACGAGATG
-462	TTGGTGAGAG	AAAGGTTCAA	GCCTATCCTT	<u>c-Myb</u> TCCATTGCCT	<u>TCF (R)</u> TCGAGAACAG	AGATTGTAAA
-402	<u>c-Myb</u> CCAGCCGTTA	AAATCAAAAG	CTTTGTTCGT	<u>Oct-R</u> AGAGCTGCAG	<u>TCF (R)</u> TTACCTCTCC	TTTGCCTTCT
-342	TTAACTCTCT	AGGGCTCCAT	CCAGAAAATG	CAACAGTTC	ATACGGTTTT	TGTAAGCTA
-282	GATAAAGCCT	ACTGCCACAG	AATATCTAGC	ACAACCATAA	GCTTGTCAAG	TTGTCAAAT
-222	<u>Ets (R)</u> TTTCTTCTCT	TCCTATTATA	ACTTTACTTT	<u>TCF (R)</u> CTGTGGGCTT	TACTGAAGAA	ACAACATTTT
-162	<u>GATA (R)</u> TATTATATCTA	GTTCGTCTTT	CAAGGATAAT	TTGCAACTCT	ACTCCAGTCA	TTTGAGCAAC
-102	AGTTCTCCAA	TTCCCAAGCT	GCCACAAAGC	GCAGCGCCTT	<u>DR II</u> TCACGACAGC	<u>DR III</u> GCACGACGAA
-42	<u>Ets</u> GGCAGCGGAA	<u>NF-1</u> AGCCTCTGCT	<u>HIP</u> TTTCGGGGCA	<u>CAAT</u> ATCGCGCTGC	<u>AP2</u> TCCGCGGCGA	<u>CRE</u> TGACGTATAT
+19	<u>NR-1</u> GTTCGCGCAG	CGCAGGGTGG	GGAGACCTTG	GATAAGTCTT	ACCAGGCGTT	CACGCGAGGT
+79	GAAGAACCCCT	GTTTGGAGGG	ACGGGACCGA	GGGACGATGC	GAGCTGGAGT	CGTGTGTACC
+139	GCTCCTTGGC					

Figure 1 Nucleotide sequence of the 5'-flanking region of the murine *Dld* gene

Nucleotides are numbered relative to the transcription initiation site (+1). Potential consensus sequences for transcriptional factor binding sites are overlined. Arrows indicate the upstream direct repeat element. The translation initiation site is underlined. The asterisks indicate the regions protected by NIH3T3 nuclear extracts.

Figure 2 DNase I footprinting analysis of protein binding domains of the *Dld* promoter

A 545-bp fragment of the murine *Dld* proximal promoter was end-labelled and subjected to digestion by DNase I in the absence (-) or presence (+) of 50–100 µg NIH3T3 nuclear proteins. The protected regions in the coding strand are identified by nucleotide numbers and denoted as domains P1 to P6. NE, nuclear extract.

-56 bp to -29 bp; P4, -235 bp to -211 bp; P5, -329 bp to -313 bp; P6, -385 bp to -370 bp).

Functional analysis of the murine *Dld* promoter

To identify the functional region of the murine *Dld* promoter, a series of progressive deletions of the *Dld* 5'-flanking region (this site was determined previously by primer extension using murine liver mRNA) [11] were constructed and ligated to the upstream of the CAT reporter gene. The constructs were transiently transfected into NIH3T3 cells and CAT activity was determined after 48 h. The CAT activity was normalized to β -galactosidase activity for transfection efficiency and the results were expressed as fold change relative to construct D1 (-2.5 kb to +41 bp) (Figure 3).

The deletions from -2.5 kb to -740 bp (Figure 3, constructs D2 to D4) led to a 3- to 4-fold increase in CAT activity relative to the -2.5 kb construct (D1). Further deletions to -573 bp (D5) and to -483 bp (D6) resulted in approx. 7.5- to 9-fold

Figure 3 Functional analysis of the murine *Dld* promoter in NIH3T3 cells

Names of various *Dld* promoter–CAT constructs used and the deletion start points relative to the transcription initiation site are listed on the left. The lines represent the murine *Dld* promoter and the open boxes represent the CAT structural gene. The shaded circles represent the nuclear protein binding sites identified from DNase I footprinting analysis. The level of CAT expression in transiently transfected NIH3T3 cells is shown on the right. Transfection efficiency of CAT constructs was normalized to β -galactosidase activity and the relative CAT activity is expressed as fold change over that of the D1 construct (—2.5 kb to +41 bp). The results are the means $\pm$ S.D. of six independent transfections.

Figure 4 Site-specific mutations of the upstream direct repeat in transient transfection assays

A schematic representation of the *Dld* promoter fused with the CAT structural gene (construct D9; —82 bp to +41 bp) is presented at the top. The mutations (M1–M10) are listed on the left. The sequence and the position of each mutation is shown. M3 + M10 represents a double mutation. The level of CAT expression in transiently transfected NIH3T3 cells is shown on the right. Transfection efficiency of CAT constructs was normalized to β -galactosidase activity. The wild-type D9 construct was assigned a value of 100%. The results are the means $\pm$ S.D. of six independent transfections.

Figure 5 EMSA analysis of the upstream direct repeat with the CACGAC motif binding activity in NIH3T3 nuclear proteins

(A) Sequences of oligonucleotides used in the present study are shown. The CACGAC motif is boxed. The positions of mutations in the CACGAC motif are marked with asterisks. (B) Competition EMSA with various mutant upstream-direct-repeat fragments. The radiolabelled probe A was incubated with 5 µg NIH3T3 nuclear proteins in the binding reaction in the presence of a 0-, 50- and 200-fold molar excess of the wild-type probe (probe A, lanes 1–3) and three mutant probes (probe B, lanes 4–6; probe C, lanes 7–9 and probe D, lanes 10–12).

increases in CAT expression. Deletion to -362 bp (D7), which removed the P6 domain, showed a small reduction in the level of CAT expression (approx. 6-fold). CAT activity was dramatically reduced to the level of approx. 1.5-fold by the removal of the P5 domain, with a deletion to -242 bp (D8), suggesting that a positive regulatory element may be located within the P5 domain. Extending the deletion to -82 bp (D9) removed the P4 domain and resulted in an increase in CAT expression to approx. 5-fold. This finding suggests that the P4 domain may function as a negative regulatory-factor(s) binding site. Deletion to -42 bp (D10) resulted in reduction in the level of CAT expression to approx. 1-fold, suggesting that positive regulatory factor(s) binding site(s) may be located within the region from -82 bp to -43 bp, although DNase I footprinting did not identify any protein binding sites (see below). The minimal promoter activity seen with the D10 construct was abolished by deletion to -17 bp (D11), which removed the P3 domain and a part of the P2 domain.

Requirement of direct repeat elements for optimal *Dld* promoter activity

Progressive deletions indicated that the region from -82 bp to -43 bp is important for optimal promoter activity. In this region, three direct repeats were identified (Figure 4). Direct repeat I had an AGCGCA motif separated by 14 bp. Direct repeat II had an AGCGC motif with no nucleotides between half-sites, and direct repeat III had a CACGAC motif separated by 4 bp. In an attempt to identify the possible role of these direct repeats in transcription of the *Dld* gene, a series of mutations were introduced in each direct repeat. Mutant M1, which interrupted direct repeats I and II, did not significantly change

the level of CAT expression in transient transfection studies. Substitution of the first three nucleotides of the upstream CACGAC motif in direct repeat III (M3) dramatically reduced CAT expression to approx. 20% of the wild-type promoter, whereas replacement of the last three nucleotides of the upstream CACGAC motif (M4) had approx. 60% of wild-type activity. Similar results were observed when the downstream CACGAC motif in the direct repeat III was mutated (M6 and M7). Two additional mutations (M2 and M5), which replaced two nucleotides immediately upstream of the direct repeat III or between the two CACGAC motifs had no significant effect on the level of CAT expression. The results of site-specific mutations indicate that the direct repeat III with CACGAC motif is required for the optimal promoter activity in the region from -82 bp to +41 bp. A double mutation (M3+M10), affecting both the direct repeat III and Ets motifs, had approx. 10% of CAT activity of the wild-type promoter, whereas mutation of direct repeat III (M3) alone reduced CAT activity to 20% and mutation of the Ets motif (M10) diminished CAT activity to approx. 50%.

Site-specific mutagenesis analysis indicated that the direct repeat III is important for transcription of the *Dld* gene. However, DNase I footprinting did not identify any protection in this region (Figure 2). To investigate further whether any nuclear factors bind to the direct repeat III, a 22-bp oligonucleotide segment (Figure 5A, probe A) including the sequence of direct repeat III was end-labelled and used for EMSA. As shown in Figure 5B, one prominent DNA-protein complex was observed when end-labelled probe A was incubated with NIH3T3 nuclear extracts. The presence of this complex was completely blocked by the addition of an approx. 200-fold excess of unlabelled probe A (Figure 5B, lanes 1–3), demonstrating a specific interaction between nuclear factors and direct repeat III. Unlabelled probes B, C and D, in which the upstream or downstream CACGAC motif was mutated, did not effectively compete for binding with the wild-type probe, despite an approx. 200-fold excess of unlabelled probe (Figure 5B, lanes 4–12), suggesting that nuclear factors have low affinity for the mutant sequences. In contrast to the results of DNase I footprinting, the EMSA shown in Figure 5(B) indicates that the nuclear factors do bind to the direct-repeat element.

A search of the Transcription Factor Database (GBF Bio-informatics) using the Pattern Search program [22] identified no consensus binding sequences of known transcription factors that had more than 80% identity with the CACGAC motif, suggesting that the CACGAC motif may represent a novel motif for binding a nuclear factor(s).

Requirement of an Ets motif for minimal promoter activity

Progressive deletions indicated that the D10 construct (-42 bp to +41 bp) was sufficient for minimal promoter activity (Figure 3). To identify the element responsible for transcription within the -42 bp to +41 bp region, a series of mutations based on the D9 (-82 bp to +41 bp) construct were made, fused with the CAT reporter gene and transiently transfected into NIH3T3 cells (Figure 6).

In the region from -42 bp to -8 bp, there are several consensus sequences for transcription factors, such as Ets (-37 bp to -32 bp) [23], housekeeping protein 1 (-22 bp to -13 bp) [24], activator protein 2 (-15 bp to -8 bp) [25], nuclear factor 1 (-26 bp to -11 bp) [26] and a CAAT box (-16 bp to -11 bp). To determine the potential function of the proximal Ets motif (CGGAAA) in the murine *Dld* promoter, two mutations were made, each having three nucleotide substitutions in the CGGAAA motif (M10, ATTAATA and M11, CGGCCC).

Figure 6 Site-specific mutations of the murine *Dld* basal transcriptional region in transient transfection assays

A schematic representation of the *Dld* promoter fused with the CAT structural gene is represented at the top. The mutations (M8–M19) are listed on the left. The sequences and positions of the mutations are shown. (A) Site-specific mutations based on the construct D9 (–82 bp to +41 bp). The level of CAT activity of the wild-type D9 construct is assigned a value of 100%. (B) Site-specific mutations based on the construct D10 (–42 bp to +41 bp). The transfection efficiency of CAT constructs was normalized to β -galactosidase activity. The level of CAT activity of the wild-type D10 construct was assigned a value of 100%. A deletion construct (D13) is shown at the bottom with CAT activity relative to that of the D10 construct. The results are the means $\pm$ S.D. of six independent transfections. The asterisks indicate $P < 0.01$.

The promoter activity of each mutation (M10 or M11) had approx. 40–50% of the activity of the wild-type promoter, suggesting that this Ets motif was functionally important for expression of the murine *Dld* gene. Whereas mutations of housekeeping protein 1, activator protein 2, nuclear factor 1 and CAAT consensus sequences (M13 to M17) did not significantly change the level of CAT expression, indicating that these *cis*-elements were not required for transcription of the *Dld* promoter. A mutation of the transcription initiation site (M19) retained approx. 75% of the CAT activity of the wild-type promoter. Wiley et al. [27] have shown that the TATA-binding protein (TBP) can functionally bind to the –30 bp region in a TATA-

less promoter. To test whether the TBP binds to the CGGAAA motif, the –82 bp to +41 bp fragment of murine *Dld* promoter containing the CGGAAA motif was end-labelled by T4 kinase. The results of the EMSA indicated that purified TBP did not bind to the –82-bp to +41-bp fragment and an approx. 500-fold excess of TBP consensus oligonucleotide could not compete with the binding of the –82-bp to +41-bp fragment to the NIH3T3 nuclear extracts (results not shown). In addition, the M10 mutation produced a sequence close to the TATA consensus (ATTAAA), which preserved only approx. 40% of CAT activity of the wild-type promoter. These results indicate that the CGGAAA motif is not a TBP binding site.

Figure 7 EMSA analysis of the Ets motif in the *Dld* promoter

(A) Sequences of oligonucleotides used in this study are shown. The Ets motif is boxed. The positions of mutations in the Ets motifs are designated by asterisks. (B) Competition EMSA with various mutant Ets motifs. The radiolabelled probe E was incubated with 5 μ g of NIH3T3 nuclear proteins in the binding reaction in the presence of a 0-, 50-, 100- or 200-fold molar excess of the wild-type (probe E, lanes 1–4) and mutant Ets probes (probe F, lanes 5–9 and probe G, lanes 10–12). (C) EMSA analysis of the binding activity of GABP with the wild-type or mutant Ets motif. The radiolabelled probe E (wild-type) or probe F (mutant) was incubated with different concentrations of recombinant GABP as indicated in the binding reactions.

To identify further the role of the Ets motif in the transcriptional activity without direct repeat III, several mutations based on construct D10 (–42 bp to +41 bp) were made, fused with the CAT reporter gene and transiently transfected into NIH3T3 cells. As shown in Figure 6(B), mutation of the Ets motif (M20) resulted in an approx. 50% reduction in CAT expression of the wild-type promoter. Mutation of the Ets motif in the construct without direct repeat III (M20) caused a similar percentage reduction in CAT expression as did mutation of the construct with the direct repeat III (M10 in Figure 6A). Mutations of nuclear factor 1 consensus sequence (M21 and M22), HIP-1 consensus sequence (M22) and transcription initiation site (M23) did not change the level of CAT expression. Deletion to –25 bp

(D13), which removed the Ets motif, led to a background level of CAT expression.

To investigate the binding ability of Ets and its mutant motifs with nuclear proteins, competition EMSA was performed using a 20-bp oligonucleotide segment of the promoter sequence containing the wild-type Ets motif as a probe (Figure 7A, probe E). As shown in Figure 7(B), one prominent band was observed when end-labelled probe E was incubated with NIH3T3 nuclear proteins. This prominent band was blocked by the addition of excess unlabelled probe E (Figure 7B, lanes 1–4), indicating a specific interaction between the Ets motif and nuclear proteins. The oligonucleotide segments mutated in the Ets motif, probes F (Figure 7B, lanes 5–8) and G (Figure 7B, lanes 9–12), competed poorly against the wild-type Ets motif with an approx. 200-fold excess of unlabelled probe.

The results of mutation and deletion indicated that the Ets motif is required for the promoter activity and may function as an initiator element. In the Ets family, GA-binding protein (GABP) has been reported to function as initiator in several mitochondrial genes [7,8]. To investigate whether GABP binds to the Ets motif in the *Dld* promoter, end-labelled probe E (wild-type Ets motif) was incubated with purified recombinant GABP. The recombinant GABP used here was a truncated protein containing both GABP α and GABP β DNA-binding domains (gifts from Dr. A. H. Batchelor and Dr. C. Wolberger, Johns Hopkins University, Baltimore, MD, U.S.A.) [28]. A DNA–protein complex was observed when probe E was incubated with recombinant GABP (Figure 7C). The amount of DNA–protein complex correlated directly with the amount of recombinant GABP. However, no DNA–protein complex was observed by incubation of probe F, an Ets mutant probe, with recombinant GABP.

DISCUSSION

This study presents evidence for a unique combination of two *cis*-elements, one of them a hitherto unrecognized CACGAC motif, in transcription of the murine *Dld* housekeeping gene. Housekeeping genes are expressed in all cells and usually lack both TATA and CCAAT boxes. They are also rich in G+C and contain one or more Sp1 binding sites [29]. The murine *Dld* promoter has several characteristics of a housekeeping gene promoter [11], i.e. no TATA motif and high G+C content within the proximal promoter region. However, no Sp1 binding site was found within an approx. 1 kb (–1002 bp to +41 bp) murine *Dld* promoter region. EMSA, using purified Sp1, indicated that Sp1 did not bind to a –82-bp to +41-bp fragment (results not shown), demonstrating that Sp1 does not regulate transcription of this proximal 5'-flanking region. Although a CCAAT box was found in the proximal promoter region (–16 bp to –11 bp), mutation analysis indicated also that this CCAAT box did not modulate transcription (Figure 4).

DNase I footprinting analysis of the –545 bp to +41 bp region (Figure 2) and progressive deletions extending from the 5' end of the –2.5 kb region (Figure 3) revealed several important features. A negative regulatory-element binding site was identified in the region between –740 bp and –574 bp. The region spanning from –483 bp to –242 bp, which includes the P6 and P5 domains, had a positive effect on transcription, whereas the region from –242 bp to –82 bp, which harbours the P4 domain, had a suppressive effect on transcription. The region from –42 bp to +41 bp had minimal promoter activity, indicating that the P1, P2 and P3 domains were required for minimal transcription activity. However, in the human *DLD* promoter, only one protein binding domain (–23 bp to –3 bp) near the

transcription initiation site was observed (J. Morris, G. Johanning and M. S. Patel, unpublished work). These observations suggest that the transcriptional initiation of the *DLD* gene in human and mouse may vary in some regard. It is not surprising that the same gene from different species has different transcriptional initiation mechanisms. Such differences have been found also between the human and mouse terminal deoxynucleotidyl transferase promoters [30–33].

Deletion and mutation analyses combined with transient transfection assays identified two elements involved in the expression of the murine *Dld* gene, i.e. the Ets motif located from –37 bp to –32 bp and the direct repeat with a novel CACGAC motif located at 8 bp upstream of the Ets motif. Deletion analysis showed that construct D10 (–42 bp to +41 bp) had minimal levels of *Dld* promoter activity (Figure 3). Since the Ets motif may function as an initiator, removal of the Ets motif (construct D13 in Figure 6B) almost abolished the promoter activity. In addition, mutations in the Ets motif resulted in reduction (approx. 50 %) of CAT activity in transiently transfected NIH3T3 cells (Figures 6A and 6B). This reduction may be due to weak binding of nuclear protein to the mutated sequences. A weak interaction between the mutated Ets sequence and NIH 3T3 nuclear extracts was also observed in the competition EMSA, when a > 500-fold excess of unlabelled mutant probe (probe F) was used to compete with the labelled wild-type probe (results not shown). However, mutations at consensus sequences, such as the CCAAT box (–16 bp to –11 bp), activator protein 2 (–15 bp to –8 bp) and housekeeping protein 1 (–22 bp to –13 bp), within the proximal promoter region produced no significant change in levels of CAT expression, indicating that these sequences were not required for the expression of the *Dld* gene. The results of mutation and deletion assays suggest that the Ets motif may function as an initiator element in the murine *Dld* promoter. The Ets family has a highly conserved C-terminal DNA-binding domain, which recognizes a GGA motif. It has been noted that different Ets family members have different DNA binding specificities [34–36]. For example, an Ets motif in the nicotinic acetylcholine receptor promoter binds with GABP but not with either Ets-2 or Elf-1 in EMSAs [37]. The Ets motif (AAAGGAAA; the underlined type indicates the Ets core sequence) of the human pyruvate dehydrogenase β subunit promoter failed to bind recombinant GABP (results not shown), whereas recombinant GABP bound the Ets motif (CGGAAA) of the murine *Dld* promoter, implying that this motif is capable of functioning as a GABP binding site (Figure 7C). In a previous study, primer extension of the 5' end of *Dld* mRNA showed multiple transcription initiation sites positioned from –43 bp to +14 bp [11]. In other studies of promoters utilizing the GABP initiator, initiation of transcription from multiple sites extending over more than 20 nucleotides has been observed [8,9,38,39]. Although the Ets motif is distal to the predominant transcription initiation site of murine liver mRNA, this motif is close to several upstream minor transcription initiation sites (–43 bp, –29 bp and –26 bp) [11]. The Ets motif has been found in a number of nuclear genes encoding mitochondrial respiratory chain proteins, such as cytochrome *c* oxidase subunits IV and Vb (*COXIV* and *COXVb*) [8,9,40,41]. In murine *COXIV* and *COXVb* promoters, Ets motifs have been suggested to be GABP binding sites and function as initiators [8,9]. For maximal level of transcription activity in *COXIV* and *COXVb* promoters, two GABP binding sites that are separated by 26 nucleotides are required [41–43]. However, a single GABP binding site has also been shown to be sufficient for minimal promoter activity in transfected NIH3T3 and COS cells [8,9], which was also seen for the *Dld* promoter in the present study (Figure 3, construct D10). EMSA, using purified recombinant

GABP, also demonstrated that GABP binds with a single GABP binding motif (Figure 7C).

Similarly to other GABP initiator elements [8,41,44], the Ets motif in the murine *Dld* gene is activated by upstream elements. A direct repeat element with the CACGAC motif located 8 bp upstream of the Ets motif activates transcription. Construct D9 (with direct repeat III) had approx. 5-fold increased CAT activity of that of construct D10 (without the direct repeat III), indicating that the direct repeat with the CACGAC motif greatly enhances transcriptional activity of the *Dld* promoter. Site-specific mutation or deletion (Figures 3 and 4) of the direct repeat with the CACGAC motif demonstrated the abolition of the activation effect of the direct repeat element on the *Dld* promoter. A double mutation affecting the direct repeat element and Ets motif had an almost additive effect (Figure 4).

Although DNase I footprinting analysis did not show the presence of the direct repeat with the CACGAC motif within the protected sequences (Figure 2), competition EMSA using oligonucleotide segments with the wild-type or mutant CACGAC motif (Figure 7) did identify interactions of the direct repeat with the CACGAC motif and NIH3T3 nuclear extracts. It is possible that the DNase I footprinting assay is not sensitive enough to detect the trace amount of DNA–protein complex formed [19], or that the interaction between the direct repeat III element and nuclear protein(s) was weakened by the conditions of the DNase I footprinting assay. Oligonucleotide segments with mutations at either the CACGAC motif did not compete against the wild-type probe, indicating that the DNA–protein interaction was specific for CACGAC motifs and that the mutant probes had little binding affinity with the nuclear extracts. Mutation of nucleotide sequences between two CACGAC motifs (M5) did not alter the level of CAT expression, suggesting that the sequence between CACGAC motifs may not be important for activating transcription. The results of EMSAs using different nuclear proteins from rat liver, murine MIN6, human HepG2 and human HeLa cells (results not shown) indicated similar mobility shift effects as those seen with NIH3T3 nuclear extracts, suggesting that the CACGAC direct repeat binding factor is ubiquitous. In addition to the murine *Dld* promoter, the direct repeat with CACGAC or CACGAC-like sequences was also present in other promoters of nuclear genes encoding mitochondrial proteins, such as the human *DLD* gene [12] and human pyruvate dehydrogenase α subunit [45] and β subunit [46] genes, suggesting that the direct repeat with CACGAC motif may play a role in regulating these genes. The CACGAC motif is a novel transcription-factor binding motif but additional studies are required in order to identify the nuclear factor(s) that interacts with this novel *cis* element.

We thank Dr. Adrian H. Batchelor and Dr. Cynthia Wolberger (Johns Hopkins University, Baltimore, MD, U.S.A.) for supplying the recombinant GABP, and Dr. Daniel Kosman (State University of New York, Buffalo, NY, U.S.A.) and Dr. David Samols (Case Western Reserve University, Cleveland, OH, U.S.A.) for critical reading of this manuscript. This work was supported by U.S. Public Health Service Grant DK 42885 to M.S.P., and M.T.J., a predoctoral trainee, was supported by Metabolism Training Grant AM 07319.

REFERENCES

- Patel, M. S. and Harris, R. A. (1995) *FASEB J.* **9**, 1164–1172
- Patel, M. S. and Roche, T. E. (1990) *FASEB J.* **4**, 3224–3233
- Reed, L. J. (1974) *Acc. Chem. Res.* **7**, 40–46
- Knowles, S. E. and Ballard, F. J. (1974) *Biol. Neonate* **24**, 41–48
- Sperl, W., Sengers, R. C. A., Trijbels, J. M. F., Ruitenbeek, W., De Graaf, R., Ter Laak, H., Van Lith, T., Kerkhoff, C. and Janssen, A. (1992) *Biol. Neonate* **61**, 188–200
- Hu, C.-W. C., Utter, M. F. and Patel, M. S. (1983) *J. Biol. Chem.* **258**, 2315–2320

- 7 Carothers, D. J., Pons, G. and Patel, M. S. (1988) *Biochem. J.* **214**, 177–181
- 8 Carter, R. S., Bhat, N. K., Basu, A. and Avadhani, N. G. (1992) *J. Biol. Chem.* **267**, 23418–23426
- 9 Carter, R. S. and Avadhani, N. G. (1994) *J. Biol. Chem.* **269**, 4381–4387
- 10 Sharrocks, A. D., Brown, A. L., Ling, Y. and Yates, P. R. (1997) *Int. J. Biochem. Cell Biol.* **29**, 1371–1387
- 11 Johnson, M., Yang, H.-S., Johanning, G. L. and Patel, M. S. (1997) *Genomics* **41**, 320–326
- 12 Johanning, G. L., Morris, J. I., Madhusudhan, K. T., Samols, D. and Patel, M. S. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10964–10968
- 13 Feigenbaum, A. S. and Robinson, B. H. (1993) *Genomics* **17**, 376–381
- 14 Gorman, C. M., Moffat, L. F. and Howard, B. H. (1982) *Mol. Cell. Biol.* **2**, 1044–1055
- 15 Pollock, R. and Treisman, R. (1990) *Nucleic Acids Res.* **18**, 6197–6204
- 16 Dignam, J. D., Liebowitz, R. M. and Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475–1489
- 17 Roesler, W. J., Vandenbark, G. R. and Hanson, R. W. (1989) *J. Biol. Chem.* **264**, 9657–9664
- 18 Maxim, A. M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560
- 19 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1992) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York
- 20 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- 21 Devereux, J., Haeblerli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- 22 Wingender, E., Kel, A. E., Kel, O. V., Karas, H., Heinemeyer, T., Dietze, P., Knueppel, R., Romaschenko, A. G. and Kolchanov, N. A. (1997) *Nucleic Acids Res.* **25**, 265–268
- 23 Brown, T. A. and McKnight, S. L. (1992) *Genes Dev.* **6**, 2502–2512
- 24 Means, A. L. and Farnham, P. J. (1990) *Mol. Cell. Biol.* **10**, 653–661
- 25 Williams, T. and Tjian, R. (1991) *Genes Dev.* **5**, 670–682
- 26 Goyal, N., Knox, J. and Gronostajski, R. M. (1990) *Mol. Cell. Biol.* **10**, 1041–1048
- 27 Wiley, S. R., Kraus, R. J. and Mertz, J. E. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5814–5818
- 28 Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McNight, S. T. and Wolberger, C. (1998) *Science* **279**, 1037–1041
- 29 Black, M. C., Jambou, R. C., Swock, A. G., Kahn, J. W. and Azizkhan, J. C. (1990) *Mol. Cell. Biol.* **10**, 6632–6641
- 30 Bhaumik, D., Yang, B., Trangas, T., Bartlett, J. S., Coleman, M. S. and Sorscher, D. H. (1994) *J. Biol. Chem.* **269**, 15861–15867
- 31 Sorscher, D. H., Yang, B., Bhaumik, D., Trangas, T., Philips, A. V., Chancellor, K. E. and Coleman, M. S. (1994) *Biochemistry* **33**, 11025–11032
- 32 Smale, S. T. and Baltimore, D. (1989) *Cell* **57**, 103–113
- 33 Smale, S. T., Schmidt, M. C., Berk, A. J. and Baltimore, D. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4509–4513
- 34 Urness, L. D. and Thummel, C. S. (1990) *Cell* **63**, 47–61
- 35 Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D. and Graves, B. J. (1992) *Genes Dev.* **6**, 975–990
- 36 Woods, D. B., Ghysadal, J. and Owen, M. J. (1992) *Nucleic Acids Res.* **20**, 699–704
- 37 Schaeffer, L., Duclert, N., Huchet-Dymanus, M. and Changeux, J.-P. (1998) *EMBO J.* **11**, 3078–3090
- 38 Bottinger, E. P., Shelley, C. S., Farokhzad, O. C. and Arnaout, M. A. (1994) *Mol. Cell. Biol.* **14**, 2604–2615
- 39 Salmon, P., Giovane, A., Wasyluk, B. and Klatzmann, D. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7739–7743
- 40 Nye, J. A., Peterson, J. M., Gunther, C. V., Jonsen, M. D. and Graves, B. J. (1992) *Genes Dev.* **6**, 975–990
- 41 Basu, A., Park, K., Atchinson, M. L., Carter, R. S. and Avadhani, N. G. (1994) *J. Biol. Chem.* **268**, 4188–4196
- 42 Virbasius, J. V. and Scarpulla, R. C. (1991) *Mol. Cell. Biol.* **11**, 5631–5638
- 43 Yu, M., Yang, X.-Y., Schmidt, T., Chinenov, Y., Wang, R. and Martin, M. E. (1997) *J. Biol. Chem.* **272**, 29060–29067
- 44 Virbasius, J. V., Virbasius, C. A. and Scarpulla, R. C. (1993) *Genes Dev.* **7**, 380–392
- 45 Maragos, C., Hutchison, W. M., Hayasaka, K., Brow, G. K. and Dahl, H.-H. M. (1989) *J. Biol. Chem.* **264**, 12294–12298
- 46 Madhusudhan, K. T., Naik, S. S. and Patel, M. S. (1995) *Biochemistry* **34**, 1288–1294

Received 22 September 1998/9 December 1998; accepted 8 January 1999