Upstream stimulatory factor (USF) and neurogenic differentiation/β-cell E box transactivator 2 (NeuroD/BETA2) contribute to islet-specific glucose-6-phosphatase catalytic-subunit-related protein (IGRP) gene expression

Cyrus C. MARTIN, Christina A. SVITEK, James K. OESER, Eva HENDERSON, Roland STEIN and Richard M. O'BRIEN[1]

Department of Molecular Physiology and Biophysics, 761 PRB, Vanderbilt University Medical School, Nashville, TN 37232-0615, U.S.A.

Islet-specific glucose-6-phosphatase (G6Pase) catalytic-subunit-related protein (IGRP) is a homologue of the catalytic subunit of G6Pase, the enzyme that catalyses the final step of the gluconeogenic pathway. The analysis of IGRP-chloramphenicol acetyltransferase (CAT) fusion-gene expression through transient transfection of islet-derived βTC-3 cells revealed that multiple promoter regions, located between −306 and −97, are required for maximal IGRP-CAT fusion-gene expression. These regions correlated with *trans*-acting factor-binding sites in the IGRP promoter that were identified in βTC-3 cells *in situ* using the ligation-mediated PCR (LMPCR) footprinting technique. However, the LMPCR data also revealed additional *trans*-acting factor-binding sites located between −97 and +1 that overlap two E-box motifs, even though this region by itself conferred minimal fusion-gene expression. The data presented here show that these E-box motifs are important for IGRP promoter activity, but that their action is only manifest in the presence of distal promoter elements. Thus mutation of either E-box motif in the context of the −306 to +3 IGRP promoter region reduces fusion-gene expression. These two E-box motifs have distinct sequences and preferentially bind NeuroD/BETA2 (neurogenic differentiation/β-cell E box transactivator 2) and upstream stimulatory factor (USF) *in vitro*, consistent with the binding of both factors to the IGRP promoter *in situ*, as determined using the chromatin-immunoprecipitation (ChIP) assay. Based on experiments using mutated IGRP promoter constructs, we propose a model to explain how the ubiquitously expressed USF could contribute to islet-specific IGRP gene expression.

Key words: diabetes, pancreas, promoter, transcription.

INTRODUCTION

The transcriptional mechanisms that direct the specific or enriched expression of genes such as insulin and glucagon to the islets of Langerhans in the pancreas have been intensely investigated. An effort has been made both to identify the transcription factors that bind to the promoters of these genes and to understand how they interact to drive gene expression specifically in islet cells. The results of these studies suggest that islet-specific transcription factors do not exist. Rather, islet-specific expression is thought to be conferred by the unique combination of transcription factors bound to a given promoter, with islet-enriched transcription factors playing a major role [1,2]. In addition or alternatively, for some genes it is possible that initiation of transcription in non-islet tissues is prevented by repressors found in these tissues [1].

It is apparent from gene-knockout studies in mice that many islet-enriched transcription factors are not only important for conferring islet-specific gene expression in the adult, but are also critical for pancreatic/islet development [3]. The observation that mice lacking PDX-1 (pancreatic duodenal homeobox-1) fail to develop a pancreas is a striking example of this [4]. Other islet-enriched transcription factors, such as Pax-6 and Nkx2.2, appear to be required for later stages in islet/pancreatic development [5,6]. This suggests that islet-enriched transcription factors may be arranged in a hierarchy such that each is required for a unique developmental stage [2]. In addition, a reduction in activity of at least some of these same factors can result in abnormal islet physiology. This is best illustrated by the fact that four of the five subtypes of maturity-onset diabetes of the young ('MODY'), rare monogenic forms of type II diabetes, are caused by mutations in genes encoding islet-enriched transcription factors [7]. Together these data suggest that the study of islet-specific gene transcription can not only yield information about the regulation of individual genes, but can contribute to an understanding of islet physiology, pancreatic/islet development and the mechanisms of tissue-specific gene expression in general.

In an effort to supplement the information gained from studying the promoters of previously characterized islet-specific/enriched genes we have begun to characterize the promoter of a gene that encodes an islet-specific glucose-6-phosphatase (G6Pase) catalytic-subunit-related protein (IGRP). G6Pase catalyses the final step of the gluconeogenic pathway, the hydrolysis of glucose 6-phosphate to glucose and inorganic phosphate, and is predominantly expressed in liver, intestine and kidney [8,9]. An IGRP cDNA was isolated through a subtractive hybridization methodology used to isolate β-cell-enriched transcripts and it encodes a protein that has approx. 50 % homology at the amino acid level with the catalytic subunit of G6Pase [10]. Cloning of

Abbreviations used: G6Pase, glucose-6-phosphatase; IGRP, islet-specific G6Pase catalytic-subunit-related protein; LMPCR, ligation-mediated PCR; CAT, chloramphenicol acetyltransferase; bHLH, basic helix-loop-helix; USF, upstream stimulatory factor; ChIP, chromatin immunoprecipitation; HIT, hamster insulinoma tumour; DMEM, Dulbecco's modified Eagle's medium; SV40, simian virus 40; PDX-1, pancreatic duodenal homeobox-1; NeuroD, neurogenic differentiation; BETA2, β-cell E box transactivator 2.

[1] To whom correspondence should be addressed (e-mail richard.obrien@mcmail.vanderbilt.edu).

© 2003 Biochemical Society

the mouse and human genomic sequences revealed that IGRP is highly conserved across these species, arguing that it plays a functional role in the islet [11,12]. However, IGRP does not catalyse glucose 6-phosphate hydrolysis and the rat IGRP gene is not expressed, partially as a result of a mutated TATA box [12]. These studies raise the possibility that in mice and humans either the enzyme is non-functional or it acts on an unidentified substrate. Despite these uncertainties as to the function of the protein, RNA and immunohistochemical analyses indicated that IGRP gene expression is restricted to islets [12].

Our studies have focused on understanding the underlying mechanisms for islet-specific IGRP gene expression. We recently determined by fusion gene analysis that multiple regions of the mouse IGRP promoter are required for full activity in β-cell-derived cell lines [13]. These functionally important regions correlated with *trans*-acting factor-binding sites that were identified using the ligation-mediated PCR (LMPCR) *in situ* footprinting technique [13]. One exception, however, were *trans*-acting factor-binding sites, identified by LMPCR, that coincided with two E-box motifs located between -82 and -63 in the proximal region of the promoter [13]. The importance of these motifs was not apparent from the fusion-gene analysis, since a truncated IGRP-chloramphenicol acetyltransferase (CAT) fusion gene, with a 5′ end point of -97 that contains these elements, conferred minimal reporter-gene expression [13].

E-boxes have the consensus sequence CANNTG and bind to transcription factors of the basic helix-loop-helix (bHLH) family [14]. The centre two nucleotides in this consensus sequence are of primary importance with respect to specifying which bHLH factor will bind, whereas the flanking sequence plays a more secondary role [15,16]. Massari and Murre [14] have defined seven classes of HLH factors based on tissue distribution, dimerization characteristics and DNA-binding activity. Much of what is known about bHLH proteins has been derived from studies in muscle that demonstrated the ability of these proteins both to activate muscle-specific genes and to stimulate differentiation [17,18]. With respect to the pancreas, the class II HLH protein neurogenin 3 plays a critical role in islet development [19], whereas in the adult E-boxes are thought to be critical for the expression of the glucagon and insulin genes by binding to heterodimers of the class II HLH protein NeuroD [neurogenic differentiation; also known as BETA2 (β-cell E box transactivator 2)] and either of the ubiquitous class I HLH proteins E12/47 or HEB (HeLa E-box-binding factor) [20–22].

The present study demonstrates that both proximal E-boxes in the IGRP promoter are important for high basal IGRP gene expression, but only in the context of the full-length promoter. Thus the activity of these E-boxes is only manifest in the presence of distal promoter elements. In addition, we show using gel-retardation assays that IGRP E-box 1 preferentially binds the islet-enriched class II factor NeuroD, whereas IGRP E-box 2 preferentially binds the ubiquitously expressed class III HLH leucine-zipper members, upstream stimulatory factor (USF)-1 and -2. These observations are supported by the demonstration that both factors bind to the IGRP promoter *in situ*, as determined using the chromatin-immunoprecipitation (ChIP) assay. Based on experiments using mutated IGRP promoter constructs, we propose a model to explain how the ubiquitously expressed USF might contribute to the tissue-restricted pattern of IGRP gene expression. This model suggests that the low-affinity binding of USF to the IGRP promoter serves as a safeguard by preventing transactivation of the IGRP promoter by USF in non-islet tissues, and hence ectopic expression of the IGRP gene, but that islet-enriched accessory factors stabilize binding of USF to the IGRP promoter in β-cells.

Table 1 Mutation of the E-box motifs in the mouse IGRP promoter reduces basal fusion-gene expression

βTC-3 cells were transiently co-transfected, as described in the Materials and methods section, using LIPOFECTAMINE™ solution containing various IGRP-CAT fusion genes (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The IGRP-CAT fusion genes represented either the wild-type promoter sequence located between -306 and $+3$ (-306 WT) or site-directed mutations of the E-box 1 motif (-306 E-box 1 MUT), the E-box 2 motif (-306 E-box 2 MUT) or both (-306 E-box 1 + 2 MUT), all generated within the context of the -306 and $+3$ promoter fragment. HIT cells were transiently co-transfected, as described in the Materials and methods section, by addition of a calcium phosphate–DNA co-precipitate containing various IGRP-CAT fusion genes (15 μg), as described above, and an expression vector encoding β-galactosidase (2.5 μg). Following transfection, HIT and βTC-3 cells were incubated for 18–20 h in serum-free or serum-containing medium, respectively. The cells were then harvested and CAT, β-galactosidase and luciferase activities assayed as decribed previously [11,27]. Results are presented as the ratio of CAT/luciferase (βTC-3 cells) (a) or CAT/β-galactosidase (HIT cells) (b) activities, expressed as percentages relative to the values obtained with the -306 WT fusion gene, and represent means ± S.E.M. from three experiments, each using an independent preparation of all fusion-gene plasmids.

(a)

Plasmid	CAT/luciferase (% relative to -306 WT)
βTC-3 cells	
-306 WT	100
-306 E-box 1 MUT	75.3 ± 0.9
-306 E-box 2 MUT	47.2 ± 1.7
-306 E-box 1 + 2 MUT	25.4 ± 1.3

(b)

Plasmid	CAT/β-galactosidase (% relative to -306 WT)
HIT cells	
-306 WT	100
-306 E-box 1 MUT	84.2 ± 8.4
-306 E-box 2 MUT	43.4 ± 2.5
-306 E-box 1 + 2 MUT	36.8 ± 0.6

MATERIALS AND METHODS

Materials

[α-^{32}P]dATP (> 3000 Ci · mmol^{-1}) and [^{3}H]acetic acid, sodium salt (> 10 Ci · mmol^{-1}), were obtained from Amersham Biosciences and ICN, respectively. Specific antisera to USF-1 (sc-229 and sc-8983), USF-2 (sc-861 and sc-862), NeuroD (sc-1084), E12/47 (sc-349) and rabbit IgG (sc-2027) were all obtained from Santa Cruz Biotechnology.

Fusion-gene plasmid construction

The USF-2-VP16 plasmid (where VP16 indicates herpes simplex viral protein 16) was a generous gift from Howard Towle (Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN, U.S.A.) [23]. The construction of mouse IGRP-CAT fusion genes, containing promoter sequence from -306 to $+3$ and -97 to $+3$, in the pCAT(An) expression vector [24], has been described previously [11,13]. A three-step PCR strategy [25] was used to create site-directed mutants of the two E-box motifs. The resulting constructs, designated -306 E-box 1 MUT and -306 E-box 2 MUT (Table 1), were generated within the context of the -306 to $+3$ IGRP promoter fragment. Briefly, for each construct two complementary PCR primers were designed to mutate nucleotides within the E-box motifs [25]. The sequences of the sense-strand oligonucleotides were as

© 2003 Biochemical Society

follows (mutated nucleotides in lower case): for −306 E-box 1 MUT, 5′-ATCGTGCTTGCTCCAacGAgtGTCAGCATCACAT-GT-3′; and for −306 E-box 2 MUT, 5′-ACAGATGGTCAG-CATacCAgtTCACGTAATGGCTCA-3′. With the −306 IGRP-CAT plasmid as the template, these sense-strand oligonucleotides were used in conjunction with a 3′ PCR primer to generate the 3′ half of the IGRP promoter, whereas the complementary antisense-strand oligonucleotides were used in conjunction with a 5′ PCR primer to generate the 5′ half of the IGRP promoter. The 3′ primer (5′-CCGCTCGAGATCCAGATCCTC-3′; *Xho*I cloning site underlined) and 5′ PCR primer (5′-CGGGATCCAAGCTCTAGCCAAGC-3′; *Bam*HI cloning site underlined) were designed to conserve the junctions between the IGRP promoter and pCAT(An) vector to be the same as those in the wild-type −306 to +3 IGRP-CAT fusion-gene plasmid. The PCR products from each reaction pair were then combined and used themselves as both primer and template in a second PCR reaction step to generate small amounts of the full-length, mutated IGRP promoter fragments. Finally, the 5′ and 3′ PCR primers were then used to amplify these fragments. A construct designated −306 E-box 1 + 2 MUT in which both E-box motifs were mutated was generated using the same three-step PCR strategy, but with the −306 E-box 2 MUT plasmid as the template and 5′-ATCGTGCTTGCTCCAacGAgtGTCAGCATACCAGTT-3′ as the sense-strand primer. A construct designated −306 E-box 2 to Ins E-box, in which the core and flanking sequence of IGRP E-box 2 were replaced with the rat insulin I E1 motif, was also generated using the same PCR strategy but with the −306 IGRP-CAT plasmid as the template and 5′-CCACAGATGGTCAGCgcCAtcTGcCACGTAATGGCTCAG-3′ as the sense-strand primer (mutated nucleotides are in lower case letters).

Constructs designated −97 E-box 1 Opt and −97 E-box 2 Opt (Table 2), in which the IGRP E-box 1 and 2 sequences, respectively, were replaced with the core and flanking sequence of an optimal (Opt) USF-binding site, were generated using PCR and the following 5′ primers (*Xba*I cloning sites are underlined and mutated nucleotides are in lower-case letters): 5′-GCTCTAGATCGTGCTTGCgCatCAcgTGacacGCATCACATG-3′ and 5′-GCTCTAGATCGTGCTTGCTCCACAGATGGTCAG-CATCACgTGaCACGTAATGGCT-3′, respectively, in conjunction with the same 3′ PCR primer described above and the −97 IGRP-CAT plasmid as the template. Note that the E-box 1 flanking sequence is switched to that of E-box 2, with the exception of a single base pair. Promoter fragments generated by PCR were completely sequenced to ensure the absence of polymerase errors. Plasmid constructs were purified by centrifugation through CsCl gradients [26].

For some of the experiments described in this paper, CAT was replaced with the more sensitive firefly luciferase reporter. This was achieved by re-isolating the various IGRP promoter fragments, as *Hin*dIII/*Xho*I or *Bam*HI/*Xho*I fragments, respectively, from the plasmids described above and ligation into the pGL3-Mod vector. pGL3-Mod was generated by replacing the polylinker in the pGL3-Basic vector (Promega) with a polylinker containing the following restriction-endonuclease-recognition sites: *Kpn*I, *Bam*HI, *Hin*dIII, *Xba*I, *Xho*I and *Bgl*II.

Cell culture and transient transfection

Hamster insulinoma tumour (HIT) cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 2.5 % (v/v) fetal bovine serum and 15 % (v/v) horse serum and were transfected by the addition of a calcium phosphate–DNA co-precipitate containing 15 μg of a CAT construct and 2.5 μg of Rous sarcoma virus–β-galactosidase as described previously [11].

Mouse islet β-cell-derived βTC-3 cells were grown in the same media, but were transfected with 0.5 μg of an expression vector encoding simian virus 40 (SV40)–*Renilla* luciferase (Promega) and either 2 μg of a CAT plasmid or 2 μg of a firefly luciferase plasmid using the LIPOFECTAMINE™ reagent (Gibco BRL) as described previously [13]. In some experiments an expression vector encoding a USF-2-VP16 fusion protein (40 ng [23]) was also included.

Human HeLa cervical carcinoma cells were grown in DMEM containing 10 % (v/v) calf serum and were co-transfected as described previously [27] using a calcium phosphate precipitate containing 15 μg of a firefly luciferase fusion gene construct and 1.2 ng of an expression vector encoding SV40–*Renilla* luciferase (Promega) respectively.

CAT, luciferase and β-galactosidase assays

Transfected HIT cells were harvested by trypsin digestion and then sonicated in 300 μl of 250 mM Tris (pH 7.8) containing 2 mM PMSF. The lysate was assayed for β-galactosidase activity as described previously [11]. Transfected βTC-3 and HeLa cells were harvested by trypsin digestion and then solubilized in passive lysis buffer (Promega). After two cycles of freeze/thawing, firefly and *Renilla* luciferase activities were assayed as described previously [27]. The remaining HIT and, when applicable, βTC-3 lysate was heated for 10 min at 65 °C and cellular debris was removed by centrifugation. CAT assays were then performed on the supernatant as described previously [11]. To correct for variations in transfection efficiency, the results are expressed as the ratio of CAT/β-galactosidase activity in HIT cell transfections, firefly/*Renilla* luciferase activity in HeLa cell transfections and either CAT/*Renilla* luciferase or firefly/*Renilla* luciferase activities in βTC-3 cell transfections. In addition, three independent preparations of each IGRP-CAT or IGRP-luciferase plasmid construct were analysed to obtain the data shown in each Table.

Gel-retardation assay

Labelled probes

Sense and antisense oligonucleotides representing wild-type or mutant IGRP E-box 1, E-box 2 and the rat insulin I E1 element (see Figure 1) were synthesized with *Bam*HI-compatible ends and subsequently gel-purified, annealed and labelled with [α-^{32}P]dATP using the Klenow fragment of *Escherichia coli* DNA polymerase I to a specific radioactivity of approx. 2.5 μCi/pmol [26].

High-salt nuclear-extract preparation

βTC-3 nuclear extract was prepared as described in [28] except that nuclei were lysed by resuspension in a buffer containing 800 mM NaCl, 20 mM Hepes, pH 7.9, 0.75 mM spermidine, 0.15 mM spermine, 0.2 mM EDTA, 2 mM EGTA, 2 mM dithiothreitol, 25 % glycerol and 1 mM PMSF. After incubation for 30 min at 4 °C to ensure complete lysis, samples were centrifuged at 541 000 ***g*** in a Beckman TLA 100.3 rotor for 40 min at 4 °C. The supernatant was dialysed against buffer containing 100 mM KCl, 20 mM Hepes, pH 7.9, 0.2 mM EDTA, 0.2 mM EGTA, 2 mM dithiothreitol, 20% glycerol and 1 mM PMSF. Protein concentrations were determined by the Bio-Rad protein assay reagent and the extract was divided into portions for storage at −80 °C.

© 2003 Biochemical Society

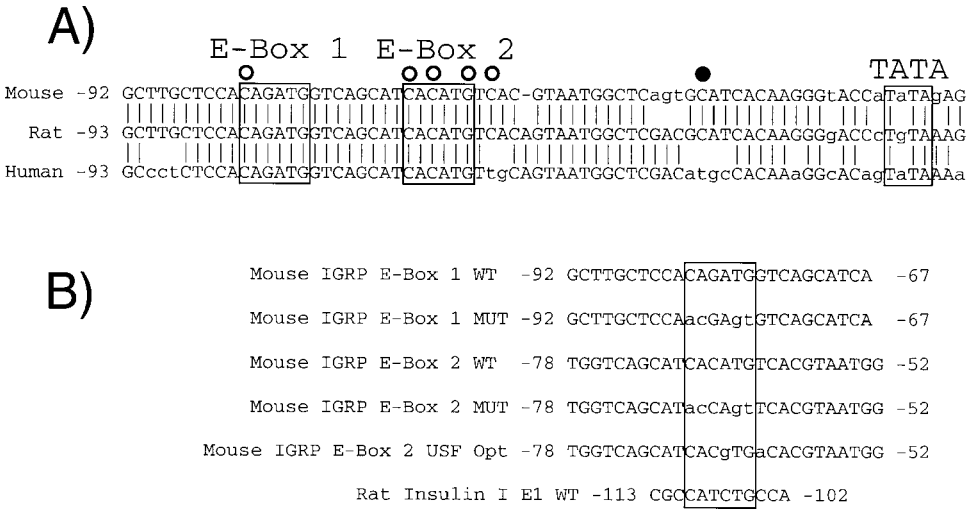

Figure 1 Conservation of E-box motifs in the mouse, rat and human IGRP promoters

(**A**) An alignment of the proximal mouse, rat and human IGRP promoters is shown. Two E-box motifs and the TATA motif are boxed. Increases (●) or decreases (○) in DMS methylation between *in situ*- versus *in vitro*-methylated mouse βTC-3 cell DNA are shown; this information was taken from [13]. (**B**) Oligonucleotides used in these studies. The wild-type (WT), mutated (MUT) and optimal (Opt) E-box motifs are boxed. All sequences are labelled relative to the transcription start site at + 1, and mutated nucleotides are in lower-case letters.

IGRP E-box 1 and rat insulin I E1 binding assays

Approx. 7 fmol of radiolabelled probe (approx. 30 000 c.p.m.) was incubated with 3 μg of βTC-3 nuclear extract in a 20 μl reaction volume containing 20 mM Hepes, pH 7.9, 0.1 mM EDTA, 1 mM EGTA, 12.5% glycerol (v/v), 1 mM dithiothreitol, 1 μg of poly(dI - dC) · poly(dI-dC), 50 mM KCl and 50 mM NaCl. After incubation at room temperature for 10 min the reactions were placed on ice for another 10 min before loading on to a 6% polyacrylamide gel containing 0.5 × TBE (45 mM Tris base/45 mM boric acid/1 mM EDTA) and 2.5% (v/v) glycerol. Samples were electrophoresed for 3 h at 150 V in 0.5 × TBE buffer at 4 °C before the gel was dried and exposed to Kodak XB film with intensifying screens.

IGRP E-box 2 binding assays

Approx. 7 fmol of radiolabelled E-box 2 probe (approx. 30 000 c.p.m.) was incubated with 1 μg of βTC-3 nuclear extract in a 20 μl reaction containing 20 mM Hepes, pH 7.9, 0.1 mM EDTA, 1 mM EGTA, 12.5% glycerol (v/v), 1 mM dithiothreitol, 2 μg poly(dI-dC) · poly(dI-dC), 50 mM KCl and 50 mM NaCl. After incubation at room temperature for 20 min samples were loaded on to a 6 % polyacrylamide gel containing 1 × TGE (25 mM Tris base/190 mM glycine/1 mM EDTA) and 2.5 % (v/v) glycerol. Samples were electrophoresed for 1.5 h at 150 V in 1 × TGE buffer at room temperature before the gel was dried and exposed to Kodak XB film with intensifying screens.

Competition experiments and gel supershifts

For competition experiments, unlabelled competitor DNA was mixed with the radiolabelled oligomer at the indicated molar excess prior to addition of nuclear extract. For supershift experiments, specific antisera (1 μl) were pre-incubated with βTC-3 nuclear extract for 10 min on ice prior to the addition of the labelled IGRP wild-type oligonucleotide probe. All subsequent steps were carried out as described above.

ChIP assays

ChIP assays were performed using a modification of published procedures [29]. Briefly, βTC-3 cells were grown to approx. 80 % confluence in sets of three 100 mm dishes. The cells were washed briefly in PBS at 37 °C and then cross-linked with 5 ml of serum-free DMEM containing 1 % formaldehyde for 10 min at room temperature. The cross-linking reaction was quenched by adding 250 μl of 2.5 M glycine to give a final concentration of 125 mM. The cells were then washed in PBS at 4 °C and subsequently scraped into 1.5 ml tubes. After pelleting by centrifugation, the cells were lysed by incubation for 10 min on ice in 200 μl of a buffer containing 1 % SDS, 10 mM EDTA, 50 mM Tris, pH 8, and 1 mM PMSF. Lysates from three dishes of cells were combined, added to a tube containing 250 mg of glass beads and subjected to 10 cycles of sonication with a Virsonic 60 cycler (Virtis Company) at 6 W. The vast majority of chromatin fragments generated were under 500 bp (results not shown).

The sonicated chromatin was then either purified as described in [29] or subjected to immunoprecipitation as follows. Aliquots (100 μl) of the sonicated chromatin were mixed with a buffer (900 μl) containing 0.01 % SDS, 1.1 % Triton X-100, 1.2 mM EDTA, 16.7 mM Tris, pH 8, 167 mM NaCl and 1 μl/ml protease inhibitor cocktail (Sigma catalogue no. P-8340) and then samples were pre-cleared by adding 25 μl of Protein A/G–agarose (Santa Cruz Biotechnology) with rotation at 4 °C for 1 h. After centrifugation, 10 μg of USF-1 (sc-8983), NeuroD (sc-1084) or rabbit IgG (sc-2027) antisera was added along with 50 μg of BSA to the supernatant and immune complexes were allowed to form at 4 °C for 18 h, again with rotation. Protein A/G–agarose (80 μl) was then added and the incubation continued for an additional 4 h at 4 °C. The immune complexes were then precipitated by centrifugation and the resulting pellets were washed for 5 min in a low-salt buffer (0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris, pH 8, and 150 mM NaCl), a high-salt buffer (0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris, pH 8, and 500 mM NaCl), a LiCl buffer (250 mM LiCl, 1 % Nonidet P40, 1 % sodium deoxycholate, 1 mM EDTA,

© 2003 Biochemical Society

and 10 mM Tris, pH 8), and finally twice in TE (10 mM Tris, pH 8/1 mM EDTA).

Protein–DNA complexes were eluted and cross-links reversed by incubating at 65 °C overnight in 500 μl of elution buffer (0.2 % SDS/0.1 M $NaHCO_3$). The following day, proteinase K (50 μg) and RNase A (20 μg) were added and samples were subjected to digestion at 55 °C for 2 h prior to extraction twice with phenol/chloroform and once with chloroform. DNA fragments were subsequently ethanol-precipitated with the aid of a co-precipitant (Novagen Pellet Paint) at room temperature and immediately pelleted by microcentrifugation for 30 min. After dissolving the pellets in nuclease-free water, the samples and a 1:100 dilution of the purified immunoprecipitation input chromatin were subjected to 28 PCR cycles (at 95 °C for denaturation, 61 °C for annealing and 72 °C for extension) using the Qiagen Master Mix. The products were visualized by electrophoresis on 2 % agarose gels containing ethidium bromide. PCR primers were designed to amplify the IGRP promoter (5′-$^{-215}$CAGAGAGGGTGCCGAAGAAATTCAAGTTC^{-187}–3′ and 5′ - $^{+86}$TCCTGCAGATGATGAATAATAAGCACTCCACTC^{+54}–3′) and exon 5 (5′-$^{+593}$GGATGCTAGTAGCCGAGGCCTTTG-AAC^{+619}–3′ and 5′-$^{+932}$GTCAGAGCACAGAGCAAGCGGAA-GC^{+908}–3′).

RESULTS AND DISCUSSION

Two E-boxes contribute to IGRP promoter activity

An analysis of IGRP-CAT fusion-gene expression through the transient transfection of the HIT cell line revealed that the region of the IGRP promoter between − 306 and + 3 was sufficient to drive maximal IGRP-CAT fusion-gene expression [11]. This same promoter region is sufficient to direct transgene expression to islets *in vivo* (C. Frigeri and R.M. O'Brien, unpublished work). Additional analyses of truncated IGRP-CAT fusion-gene expression through the transient transfection of HIT cells and the mouse islet-derived βTC-3 cell line revealed that multiple promoter regions, located between − 306 and − 97, are required for maximal IGRP-CAT fusion-gene expression [13]. These regions correlated with *trans*-acting factor-binding sites in the IGRP promoter that were identified in βTC-3 cells *in situ* using the LMPCR footprinting technique [13]. However, the LMPCR data also revealed additional *trans*-acting factor-binding sites located between − 97 and + 1 in the IGRP promoter, even though this region by itself conferred minimal fusion-gene expression [13]. An alignment of this − 97 to + 1 region of the IGRP promoter from mouse, rat and human reveals that these protein–DNA interaction sites that were identified in the mouse IGRP promoter *in situ* overlap two E-box motifs, designated E-box 1 and E-box 2, that are strongly conserved between species (Figure 1A).

To investigate the functional significance of these observations, the two E-box motifs were mutated (Figure 1B), either individually or together, in the context of the − 306 to + 3 IGRP promoter region. The level of reporter-gene expression directed by fusion genes containing these mutations was then analysed by transient transfection of βTC-3 and HIT cells. Table 1 shows that, in βTC-3 cells, the mutation of E-box 1 resulted in an approx. 25 % reduction in promoter activity, while mutation of E-box 2 resulted in an approx. 50 % reduction in activity. Elimination of both elements together leads to a further reduction in promoter activity to a level that is approx. 25 % that of the activity exhibited by the wild-type promoter. Similar results were obtained in HIT cells, except that the relative contribution of E-box 2 as compared with E-box 1 with respect to basal fusion-gene expression was increased (Table 1). These results demonstrate that these E-boxes are critical for IGRP promoter activity (Table 1), but that their activity is only manifest in the presence of distal promoter elements [13]. Thus the functional relevance of these sites was not apparent from the analysis of 5′-truncated IGRP-CAT fusion genes since the − 97 to + 1 promoter region, which includes both E-box motifs, lacks these distal elements [13]. Interestingly, a similar conclusion was reached in studies on the insulin I and II gene promoters, in which the insulin E elements were shown to be dependent on other elements for optimal function in the context of both the native promoter and heterologous promoters [30,31].

IGRP E-box 1 and 2 bind NeuroD and USF with distinct affinities

Since both the *in situ* footprinting (Figure 1 [13]) and transient transfection (Table 1) data suggest that functionally important transcription factors occupy both E-box motifs in the IGRP promoter, we next sought to identify the factors binding these elements. To address this question protein binding to these elements was analysed using the gel-retardation assay.

When a labelled oligonucleotide representing the wild-type IGRP promoter sequence from − 92 to − 67 (Figure 1B) that encompasses E-box 1 was incubated with nuclear extract prepared from βTC-3 cells, several protein–DNA complexes were detected (Figure 2). Competition experiments, in which a varying molar excess of unlabelled DNA was included with the labelled probe, were used to correlate protein binding with basal IGRP gene expression. The wild-type E-box 1 oligonucleotide competed effectively for the formation of all of these protein–DNA complexes (Figure 2A, see arrows). By contrast, an oligonucleotide, designated E-box 1 MUT (Figure 1B), that contains a mutation identical with that described in the − 306 E-box 1 MUT construct (Table 1), failed to compete with the labelled probe for formation of two complexes designated X and Y (Figure 2A). This indicates that these complexes represent specific protein–DNA interactions and that their formation correlates with basal gene expression conferred by E-box 1. In contrast, the other bands detected in the assay must represent either non-specific protein–DNA interactions or protein binding to sequences outside the E-box motif.

To identify the factors present in complexes X and Y, gel-retardation assays were performed in which βTC-3 cell nuclear extract was pre-incubated with antisera specific for members of the bHLH class of transcription factors, which are known to bind E-box motifs (Figure 2B [14]). Because E12/47 and NeuroD have been shown to bind as heterodimers to the E-box motifs in the insulin and glucagon promoters, these proteins were primary candidates [20,21]. As can be seen in Figure 2(B), addition of antibodies recognizing NeuroD resulted in a selective disruption in the formation of complex X, with no effect on the formation of complex Y. Concomitant with the disruption of complex X, a clear supershift was apparent upon addition of the NeuroD antiserum. Similarly, the E12/47 antiserum also disrupted the formation of complex X, though it only generated a weak supershift. These results suggests that complex X represents a heterodimer of these factors.

Since Whelan et al. [32] have previously shown that both NeuroD/E12/47 and USF bind the rat insulin II E1 element *in vitro*, the effect of antibodies to USF on the formation of complex Y was investigated. As can be seen in Figure 2(B), addition of antibodies recognizing either USF-1 or USF-2 both resulted in a selective disruption in the formation of complex Y, with no effect on the formation of complex X. Two separate USF-1 and USF-2 antisera that recognize distinct epitopes in

© 2003 Biochemical Society

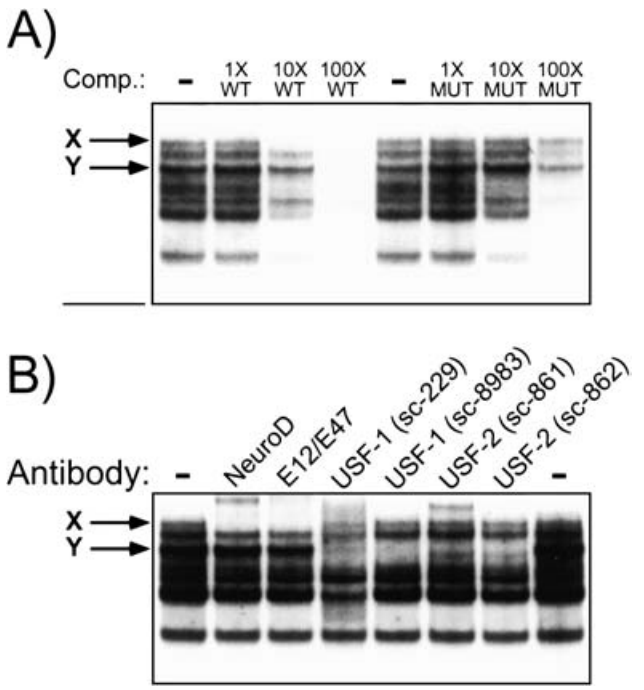

Figure 2 The IGRP E-box 1 motif binds NeuroD, E12/47 and USF

(**A**) The labelled IGRP E-box 1 WT oligonucleotide probe (Figure 1B) was incubated in the absence (–) or presence of the indicated molar excesses of the unlabelled IGRP E-box 1 WT or IGRP E-box 1 MUT oligonucleotide competitors prior to the addition of βTC-3 cell nuclear extract. Protein binding was then analysed as described in the Materials and methods section. (**B**) βTC-3 cell nuclear extract was incubated in the absence (–) or presence of the indicated antisera for 10 min on ice prior to the addition of the labelled IGRP E-box 1 WT oligonucleotide probe and incubation for an additional 10 min at room temperature and then 10 min on ice. Protein binding was then analysed as described in the Materials and methods section. For both panels, in the representative autoradiographs shown only the retarded complexes are visible, but not the free probes, which were present in excess. The arrows point to two complexes, designated X and Y, that bind the WT but not the MUT oligonucleotide and contain NeuroD/E12/47 and USF, respectively.

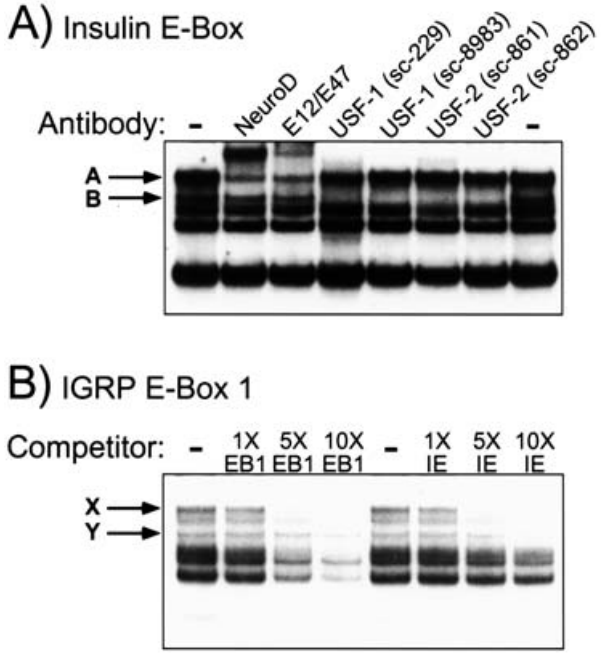

Figure 3 The rat insulin I E1 motif binds NeuroD, E12/47 and USF

(**A**) βTC-3 cell nuclear extract was incubated in the absence (–) or presence of the indicated antisera for 10 min on ice prior to the addition of the labelled rat insulin I E1 oligonucleotide probe (Figure 1B) and incubation for an additional 10 min at room temperature and then 10 min on ice. Protein binding was then analysed as described in the Materials and methods section. The arrows point to two complexes, designated A and B, that bind NeuroD–E12/47 and USF respectively. (**B**) The labelled IGRP E-box 1 WT oligonucleotide probe (Figure 1B) was incubated in the absence (–) or presence of the indicated molar excesses of the unlabelled IGRP E-box 1 WT (EB1) or rat insulin I E1 (IE) oligonucleotide competitors prior to the addition of βTC-3 cell nuclear extract. Protein binding was then analysed as described in the Materials and methods section. For both panels, in the representative autoradiographs shown only the retarded complexes are visible, but not the free probes, which were present in excess. The arrows point to two complexes, X and Y, that selectively bind the wild-type but not the mutated IGRP E-box 1 oligonucleotide (see Figure 2A). It should be noted that the exposure time for the autoradiograph shown in (**A**) was longer than that for (**B**), consistent with the competition analysis (**B**) that indicates that NeuroD binds both the rat insulin I E1 motif and IGRP E-box 1 with similar affinities.

these proteins were analysed with slightly different results. While USF-1 antiserum sc-8983 selectively disrupted the formation of complex Y, USF-1 antiserum sc-229 caused a general reduction of all protein–DNA binding, although this was clearly most specific for complex Y (Figure 2B). Neither USF-1 antiserum generated a supershifted complex. In contrast, USF-2 antiserum sc-861, but not USF-2 antiserum sc-862, did generate a supershifted complex (Figure 2B). These results suggest that complex Y contains a mixture of USF-1 and USF-2 that likely represents a USF-1/2 heterodimer, since this combination is believed to be favoured in most tissues [33]. It should be noted that the complexes which form on E-box elements in gel-retardation assays are critically dependent on the binding conditions [20,32]. Thus under different buffer conditions we only detect NeuroD/E12/47 and not USF binding (results not shown).

Since NeuroD and E12/47 bind to E-box 1 in the IGRP promoter we decided to directly compare the binding properties of IGRP E-box 1 with an established binding site for these factors, the E1 element from the rat insulin I promoter [34]. When βTC-3 nuclear extract was incubated with a labelled oligonucleotide representing the insulin I E1 element (Figure 1B), under exactly the same gel-retardation conditions that were used with IGRP E-box 1, multiple complexes were observed (Figure 3A). Performing experiments identical with those described above in which βTC-3 cell nuclear extract was pre-incubated with antisera specific for various bHLH proteins revealed that one of these complexes, designated A, represents a heterodimer of NeuroD and E12/47, whereas another complex, designated B, represents a mixture of USF-1 and USF-2 (Figure 3A).

To directly compare the affinity of the IGRP E-box 1 and rat insulin I E1 motifs for these factors competition experiments were performed in which the labelled IGRP E-box 1 WT oligonucleotide (Figure 1B) was pre-incubated with various concentrations of the unlabelled IGRP E-box 1 WT and rat insulin I E1 oligonucleotides before addition of βTC-3 nuclear extract and analysis of protein binding using the gel-retardation assay. Both elements competed equally effectively for formation of the NeuroD–E12/47 complex, indicating that they bind these proteins with similar affinity (Figure 3B; complex X). At a 10-fold molar excess neither element competed effectively for formation of the USF-1–USF-2 complex (Figure 3B; complex Y). This is consistent with the competition experiment shown in Figure 2(A), in which the IGRP E-box 1 WT oligonucleotide only competed for formation of this complex at a 100-fold excess but not at a 10-fold excess. This observation can be explained kinetically by the fact that USF is abundant in βTC-3 nuclear extract and binds the IGRP E-box 1 motif with very low affinity (see below). The fact that IGRP E-box 1 and rat insulin I E1 elements bind to the same

© 2003 Biochemical Society

transcription factors *in vitro* with similar affinities is consistent with the observation that the critical core sequence [15,16] of the IGRP E-box 1 element, CAGATG, and the rat insulin I E1 element, CATCTG, are identical, but inverted relative to each other (Figure 1B).

In contrast with the complex pattern of protein binding detected with IGRP E-box 1, when a labelled oligonucleotide representing IGRP E-box 2 (Figure 1B) was incubated with βTC-3 cell nuclear extract a single protein–DNA complex was detected (Figure 4). Competition experiments, in which a varying molar excess of unlabelled DNA was included with the labelled probe, were used to correlate protein binding with basal IGRP gene expression. The wild-type E-box 2 oligonucleotide competed effectively for the formation of this protein–DNA complex (Figure 4A). By contrast, an oligonucleotide, designated E-box 2 MUT (Figure 1B), that contains a mutation identical with that described in the -306 E-box 2 MUT construct (Table 1), failed to compete with the labelled probe for complex formation (Figure 4A). This indicates that this complex represents a specific protein–DNA interaction and that its formation correlates with basal gene expression conferred by E-box 2. As described above for IGRP E-box 1, to identify the factor(s) present in the complex formed with IGRP E-box 2, gel-retardation assays were performed in which βTC-3 cell nuclear extract was pre-incubated with antisera specific for members of the bHLH class of transcription factors. As can be seen in Figure 4(B), addition of antibodies recognizing USF-1 and USF-2 resulted in the selective disruption of this complex, whereas antibodies to NeuroD and E12/47 had no effect (results not shown). This result suggests that the specific complex formed with the IGRP E-box 2 probe represents a mixture of USF-1 and USF-2. As stated above, this mixture is probably a heterodimer of USF-1 and -2, based on the known binding characteristics of these factors [33]. As with the E-box 1 experiments (Figure 2B), two separate USF-1 and USF-2 antisera that recognize distinct epitopes in these proteins were analysed. In contrast with the IGRP E-box 1 (Figure 2B) and rat insulin I E1 (Figure 3A) gel-retardation experiments, where only the USF-2 antiserum sc-861 generated a supershifted complex, all four antisera generated a supershifted complex with the E-box 2 probe (Figure 4B). This observation may indicate that USF-1 and -2 bind the E-box 2 motif in a conformation distinct from that which forms upon IGRP E-box 1 and rat insulin I E1 binding.

The gel-retardation supershift analyses described above suggest that IGRP E-box 1, but not E-box 2, can bind a NeuroD/E12/47 heterodimer. To further characterize this difference cross-competition analyses were performed, using nuclear extract prepared from βTC-3 cells, in which a varying molar excess of unlabelled DNA was included with the labelled probe. When the oligonucleotide representing the wild-type IGRP E-box 1 sequence (Figure 1B) was used as the labelled probe, the unlabelled wild type E-box 1 oligonucleotide competed effectively for the formation of the NeuroD E12/47 heterodimer complex at a 10-fold molar excess (Figures 5A and 2A, complex X). In contrast, at a 10-fold molar excess the unlabelled wild type E-box 2 oligonucleotide did not compete for formation of this complex; partial competition was only seen at a 100-fold molar excess (Figure 5A, complex X). This result indicates that the NeuroD–E12/47 heterodimer binds the IGRP E-box 1 motif with a much higher affinity than it binds IGRP E-box 2.

This same experiment revealed that USF binds the IGRP E-box 2 motif with a higher affinity than it binds E-box 1. Thus when the wild-type IGRP E-box 1 sequence was used as the labelled probe, the unlabelled wild-type E-box 1 oligonucleotide only competed effectively for the formation of the USF-1 and -2 complex at a 100-fold molar excess (Figures 5A and 2A, complex Y). In contrast, only a 10-fold molar excess of the unlabelled wild-type E-box 2 oligonucleotide was required to prevent formation of this complex (Figure 5A, complex Y). This conclusion was supported by the results of the inverse experiment. Thus when the wild-type IGRP E-box 2 sequence was used as the labelled probe, the unlabelled wild-type E-box 2 oligonucleotide competed effectively for the formation of the USF-1 and -2 complex (Figures 4A and 5B). In contrast, the unlabelled wild-type E-box 1 oligonucleotide failed to compete for formation of this complex even at a 100-fold molar excess (Figure 5B).

To support the conclusion that USF binds to E-box 2 with a much higher affinity than it binds E-box 1, the ability of a USF-2-VP16 fusion protein [23] to transactivate various IGRP-CAT fusion genes was analysed through transient co-transfection of βTC-3 cells. As expected, overexpression of USF-2-VP16, with its potent *trans*-activation domain, stimulated expression of the full-length wild-type IGRP-CAT fusion gene containing promoter sequence between -306 and $+3$ more than 3-fold (results not shown). A similar induction of reporter-gene expression was seen when this experiment was repeated using the -306 IGRP-CAT fusion gene that contains a mutation in E-box 1 (results not shown). In contrast, however, expression of the -306 IGRP-CAT fusion gene that contains a mutation in E-box 2 was almost completely refractory to induction by USF-2-VP16 (results not shown). These results suggest that the USF-2-VP16 fusion protein exhibits a much higher affinity for E-box 2 than E-box 1, consistent with the gel-retardation data, and thus it can serve as a more potent activator through the E-box 2 motif.

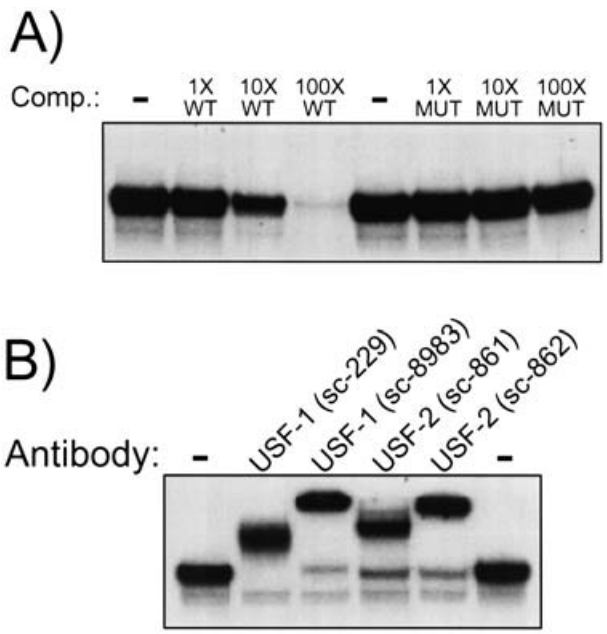

Figure 4 The IGRP E-box 2 motif binds USF

(**A**) The labelled IGRP E-box 2 WT oligonucleotide probe (Figure 1B) was incubated in the absence (−) or presence of the indicated molar excesses of the unlabelled IGRP E-box 2 WT or IGRP E-box 2 MUT oligonucleotide competitors prior to the addition of βTC-3 cell nuclear extract. (**B**) βTC-3 cell nuclear extract was incubated in the absence (−) or presence of the indicated antisera for 10 min on ice prior to the addition of the labelled IGRP E-box 2 WT oligonucleotide probe and incubation for an additional 20 min at room temperature. For both panels, protein binding was analysed as described in the Materials and methods section. In the representative autoradiographs shown only the retarded complexes are visible, but not the free probes, which were present in excess.

© 2003 Biochemical Society

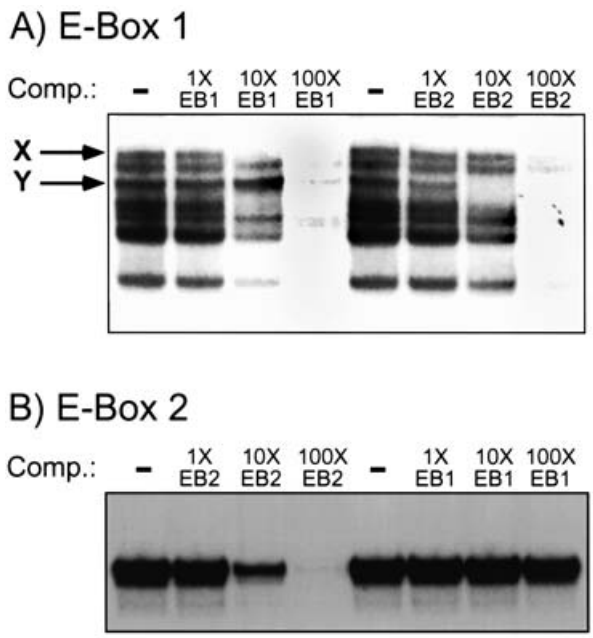

Figure 5 The IGRP E-box 1 and 2 motifs bind NeuroD–E12/47 and USF with distinct affinities

(**A**) The labelled IGRP E-box 1 WT oligonucleotide probe (Figure 1B) was incubated in the absence (−) or presence of the indicated molar excesses of the unlabelled IGRP E-box 1 WT (EB1) or IGRP E-box 2 WT (EB2) oligonucleotide competitors prior to the addition of βTC-3 cell nuclear extract. The arrows point to the two complexes, designated X and Y, that selectively bind the wild-type but not the mutated IGRP E-box 1 oligonucleotides (see Figure 2A). (**B**) The labelled IGRP E-box 2 WT oligonucleotide probe (Figure 1B) was incubated in the absence (−) or presence of the indicated molar excesses of the unlabelled IGRP E-box 2 WT (EB2) or IGRP E-box 1 WT (EB1) oligonucleotide competitors prior to the addition of βTC-3 cell nuclear extract. For (**A**) and (**B**), protein binding was then analysed as described in the Materials and methods section. In the representative autoradiographs shown only the retarded complexes are visible, but not the free probes, which were present in excess.

USF and NeuroD bind to the IGRP promoter *in situ*

To complement the results of the *in vitro* gel-retardation analyses, ChIP assays [35] were performed to assess NeuroD and USF binding to the IGRP promoter within intact cells. Whereas ChIP analyses are generally important for confirming that interactions that occur *in vitro* also occur *in situ*, this is particularly true for bHLH proteins, which comprise a large family of transcription factors that all recognize the core E-box motif [14].

Fragmented chromatin from formaldehyde-cross-linked βTC-3 cells was subjected to immunoprecipitation with antibodies to either USF or NeuroD. The presence of the IGRP promoter in the immunoprecipitates was then analysed by PCR using primers that recognize the IGRP gene sequence between -215 and $+54$. As can be seen from Figure 6(A), the IGRP promoter appears to be enriched in the USF and NeuroD immunoprecipitates compared with the IgG control. To test the specificity of the antibody–protein interactions, these immunoprecipitates were also analysed for the presence of exon 5 of the IGRP gene [11] using PCR primers that recognize the IGRP coding sequence between $+593$ and $+932$. Approx. 6 kbp of genomic DNA separates exon 5 and the IGRP promoter, and neither USF nor NeuroD would be predicted to associate with exon 5. As expected, no enrichment of IGRP exon 5 was detected in the USF or NeuroD immunoprecipitates compared with the IgG control (Figure 6A). The low signal in the experimental lanes cannot be explained by the lack of exon 5 promoter in the starting material, as a signal of the expected size can be seen in the chromatin input prior to immunoprecipitation.

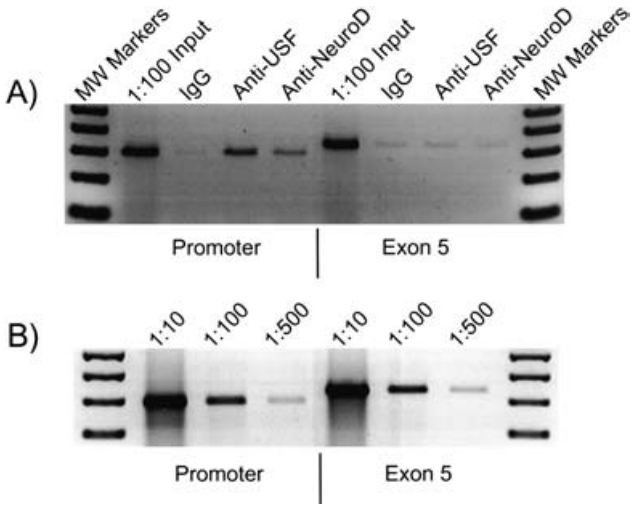

Figure 6 The IGRP promoter binds NeuroD and USF *in situ*

NeuroD and USF binding to the IGRP promoter were analysed *in situ* using the ChIP assay. Chromatin from formaldehyde-treated βTC-3 cells was immunoprecipitated using anti-USF or anti-NeuroD antibodies or, as a control, using IgG. The presence of the IGRP promoter and exon 5 in the chromatin preparation prior to immunoprecipitation (1:100 input) and in the immunoprecipitates was then assayed using PCR as described in the Materials and methods section (**A**). Both the IGRP promoter and exon 5 are amplified with similar efficiencies in the chromatin preparation (**B**). MW Markers, molecular-mass markers.

In addition, both the IGRP promoter and exon 5 are amplified with similar efficiencies (Figure 6B). These results demonstrate that both USF and NeuroD bind to the IGRP promoter inside intact cells.

Significance of selective protein binding to IGRP E-box 1 and E-box 2

The fact that the core and flanking sequences of E-box 1 and 2 are highly conserved between species (Figure 1) suggests that there may be a specific reason for this sequence conservation and the preferential binding of NeuroD to E-box 1 and USF to E-box 2. To explore this possibility, experiments were performed in which this binding specificity was reversed by switching the sequences of E-box 1 and 2 to those of USF- and NeuroD-binding sites, respectively.

The sequence of IGRP E-box 2 was changed to that of the rat insulin I E1 element, in the context of the -306 to $+3$ IGRP promoter fragment, by mutating the core E-box sequence and 2 bp either side of the core (see the Materials and methods section). The level of reporter-gene expression directed by a fusion gene containing this mutation was analysed by transient transfection of βTC-3 cells. This mutation reduced IGRP promoter activity approx. 35 % in comparison with the wild-type promoter (results not shown). This was surprising because an approx. 50 % reduction in promoter activity was seen with the IGRP E-box 2 block mutant, which should not bind to any bHLH factor (Table 1). This suggests that even though NeuroD/E12/47 is presumably binding to the rat insulin I E1 sequence in the context of this mutated IGRP promoter fragment, it cannot function effectively at the E-box 2 site to enhance basal gene expression *in situ*. Instead, this result suggests that there is a specific requirement for USF binding to IGRP E-box 2 for maximal IGRP gene transcription.

The reverse experiment, switching the sequence of E-box 1 to that of a high-affinity USF-binding site, was incorporated into a more wide-ranging study that examined the significance

© 2003 Biochemical Society

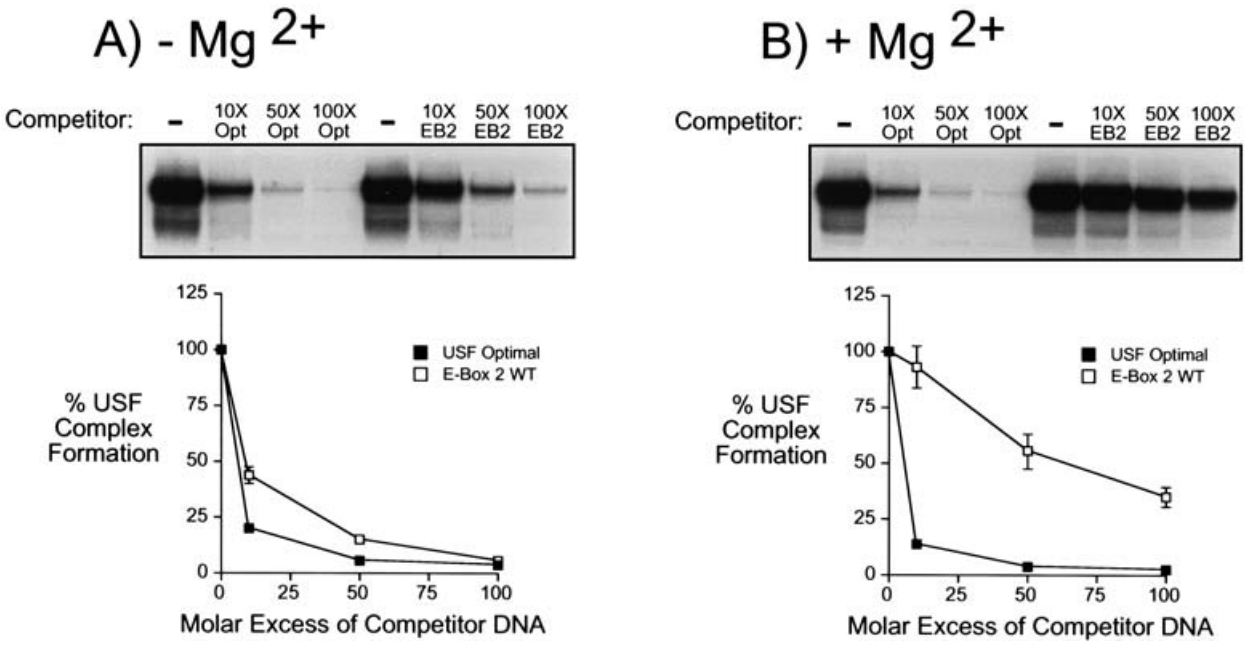

Figure 7 Magnesium affects the relative affinity of USF binding to the IGRP E-box 2 motif and an optimal USF-binding site

A labelled oligonucleotide probe representing a modified IGRP E-box 2 motif, designated E-box 2 USF Opt (Figure 1B), the sequence of which had been changed to that of an optimal (Opt) USF-binding site, was incubated in the absence (−) or presence of the indicated molar excesses of the unlabelled IGRP E-box 2 Opt or IGRP E-box 2 wild-type (EB2) oligonucleotide competitors prior to the addition of βTC-3 cell nuclear extract. Protein binding was then analysed in the absence (**A**) or presence (**B**) of 5 mM $MgCl_2$ as described in the Materials and methods section. In the representative autoradiographs shown only the retarded complexes are visible, but not the free probes, which were present in excess. Protein binding was quantified by using a Packard Instant Imager to count ^{32}P label associated with the retarded complex. Data represent the means ± S.D. from three experiments.

of low-affinity USF binding to the IGRP promoter. Thus even though USF binds to IGRP E-box 2 with much higher affinity than it binds to E-box 1 (Figure 5B), even E-box 2 is not an optimal USF-binding site; in the IGRP promoter the sequence of both E-boxes 1 and 2 differ from the optimal USF-binding sequence defined by Bendall and Molloy [16]. To test whether IGRP E-box 2 is indeed a sub-optimal USF-binding site, gel-retardation competition analyses were performed using a labelled oligonucleotide probe representing a modified IGRP E-box 2 motif, designated E-box 2 USF Opt (Figure 1B), the sequence of which had been changed to that of an optimal USF-binding site as determined by Bendall and Molloy [16]. As can be seen in Figure 7(A), the unlabelled USF optimal site (Opt) competed approximately twice as effectively for USF binding as the unlabelled oligonucleotide representing the wild-type E-box 2 sequence. The competition experiments shown in Figure 7(A) were performed in the absence of magnesium. However, Bendall and Molloy [16] showed that the presence of magnesium further restricted the ability of USF to bind to suboptimal sites. Consistent with their observations, when these competition experiments were repeated in the presence of 5 mM Mg^{2+} a dramatic difference was seen in the affinity of USF binding to these oligonucleotides (Figure 7B). Thus as can be seen in Figure 7(B), the unlabelled USF optimal site now competed at least 10-fold more effectively for USF binding than the unlabelled oligonucleotide representing the wild-type E-box 2 sequence. Since these are *in vitro* assays, the binding conditions do not precisely simulate the *in vivo* environment; however, it is interesting to note that the intracellular concentration of Mg^{2+} is estimated to be 10 mM, with 0.5 mM of that being free Mg^{2+} ions [36].

The results of these competition experiments demonstrate that although IGRP E-box 2 binds USF with higher affinity than it binds IGRP E-box 1 (Figure 5B), it is not an optimal USF-binding site (Figure 7). Given that the core and 5′-flanking sequence of E-box 2 is highly conserved between species, this suggests that there may be a specific reason for the suboptimal binding of USF to E-box 2. As USF is expressed ubiquitously, we hypothesized that the IGRP promoter has evolved a weak USF-binding site as part of the mechanism to restrict IGRP gene expression to islets. If so, a model can be envisaged in which (i) islet-enriched accessory factors stabilize USF binding to the IGRP promoter in β-cells, explaining the functional importance of E-box 2 to IGRP fusion-gene expression (Table 1) and consistent with the ChIP analysis (Figure 6A), but (ii) in non-islet cells the absence of these accessory factors coupled with weak USF binding prevents transactivation of the IGRP promoter by USF and hence ectopic expression of the IGRP gene. We predicted that, if this proposed model is correct, changing the sequence of the IGRP E-boxes to those of optimal USF-binding sites would (i) allow a truncated IGRP promoter lacking these accessory-factor-binding sites to confer fusion-gene expression in βTC-3 cells and (ii) allow misexpression of the mutated fusion gene in non-islet cell lines. To investigate these predictions, IGRP-luciferase fusion gene constructs were generated in which the sequences of IGRP E-boxes 1 and 2 were switched to that of an optimal USF-binding site in the context of a truncated -97 to $+3$ IGRP promoter fragment (Table 2). The level of reporter-gene expression directed by these constructs was then analysed by transient transfection of βTC-3 and HeLa cells (Table 2).

In βTC-3 cells the full-length IGRP promoter was highly active, as expected, whereas the truncated -97 to $+3$ IGRP promoter fragment conferred a level of reporter-gene expression that was indistinguishable from the promoterless luciferase vector control (Table 2). However, when either E-box 1 or E-box 2 was mutated

© 2003 Biochemical Society

Table 2 Converting E-boxes 1 and 2 to optimal USF-binding sites enhances IGRP fusion-gene expression in βTC-3 cells and imparts expression in HeLa cells

βTC-3 cells were transiently co-transfected, as described in the Materials and methods section, using LipofectAMINE solution containing various IGRP-luciferase fusion genes (2 μg) and an expression vector encoding *Renilla* luciferase (0.5 μg). The IGRP-luciferase fusion genes represented either the wild-type promoter sequence located between − 306 and + 3 (− 306 WT) or between − 97 and + 3 (− 97 WT) or site-directed mutations of the E-box 1 motif (− 97 E-box 1 Opt) or the E-box 2 motif (− 97 E-box 2 Opt), both generated within the context of the − 97 and + 3 promoter fragment, in which the core and flanking sequences of the respective E-box motifs were switched to those of an optimal (Opt) USF-binding site. HeLa cells were transiently co-transfected, as described in the Materials and methods section, by addition of a calcium phosphate–DNA co-precipitate containing various IGRP-luciferase fusion genes (15 μg), as described above, and an expression vector encoding *Renilla* luciferase (1.2 ng). Following transfection, HeLa and βTC-3 cells were incubated for 18–20 h in serum-free or serum-containing medium, respectively. The cells were then harvested and both firefly and *Renilla* luciferase activities assayed as decribed previously [27]. Results are presented as the ratio of firefly/*Renilla* luciferase activities expressed as percentages relative to the values obtained with the pGL3-MOD vector, and represent means ± S.E.M. from three experiments, each using an independent preparation of all fusion-gene plasmids.

Plasmid	Firefly/*Renilla* luciferase (% relative to pGL3-MOD)
βTC-3 cells	
pGL3-MOD	100
− 306 WT	11274 ± 1006
− 97 WT	103 ± 10
− 97 E-box 1 Opt	560 ± 67
− 97 E-box 2 Opt	329 ± 59
HeLa cells	
pGL3-MOD	100
− 306 WT	97 ± 5
− 97 WT	89 ± 2
− 97 E-box 1 Opt	627 ± 48
− 97 E-box 2 Opt	190 ± 14

to the optimal USF-binding site in the context of this truncated − 97 to + 3 IGRP promoter fragment, a level of reporter gene activity was detected that was clearly greater than that conferred by the promoterless luciferase vector control (Table 2). This result implies that increasing the affinity of USF for the IGRP promoter obviates the requirement for distal accessory elements, by allowing expression of truncated fusion genes that lack these accessory elements that are normally required for E-box function in the context of the full-length promoter (Table 2 [13]).

To determine whether optimal USF sites allow IGRP fusion-gene expression in ectopic tissues, these same constructs were transiently transfected into HeLa cells, a non-islet cell line. As expected for an islet-specific promoter, in HeLa cells both the full-length − 306 to + 3 and truncated − 97 to + 3 IGRP promoter fragment conferred a level of reporter-gene expression that was indistinguishable from the promoterless luciferase vector control (Table 2). However, when either E-box 1 or E-box 2 was mutated to the optimal USF-binding site in the context of the truncated − 97 to + 3 IGRP promoter fragment, a level of reporter gene activity was detected that was clearly greater than that conferred by the promoterless luciferase vector control (Table 2). Similar results were obtained in HeLa cells when the same mutations were generated in the context of the − 306 to + 3 IGRP promoter region (results not shown).

If a ubiquitously expressed factor such as USF transactivates a gene whose expression is tissue specific, one would predict there must be mechanisms in place that prevent the factor from inducing ectopic transcription of the gene. The data in Table 2 suggest that, in the case of IGRP, it is possible that the suboptimal binding sites serve this role by making it necessary for USF to co-operate with islet-enriched accessory factors to circumvent this obstacle. Certainly, it is apparent that low-affinity transcription-factor binding can play an important role in the regulation of gene transcription. For example, the multi-hormonal modulation of glucocorticoid-stimulated phosphoenolpyruvate carboxykinase gene transcription is dependent in large part on the low-affinity binding of the glucocorticoid receptor to the phosphoenolpyruvate carboxykinase gene promoter [37]. In addition, a recent developmental study in *Caenorhabditis elegans* demonstrated the utility of low-affinity transcription-factor-binding sites in the temporal control of gene expression. Thus the time of onset of expression of different pharyngeal genes in *C. elegans* is controlled, at least in part, by the relative affinity of the PHA-4 forkhead transcription factor for their promoters [38].

The observation that E-box 1 and E-box 2 are important for IGRP gene expression is consistent with the critical role played by these motifs in other islet-specific promoters. Elimination of the E1 or E2 elements in the rat insulin I promoter, for example, leads to a greater than 90 % reduction in promoter activity in transformed β-cell lines [39]. Likewise, large reductions in transcriptional activity are seen upon elimination of the E3 site in the glucagon promoter [40]. Similar to our results with IGRP E-box 1, these E-boxes in the insulin and glucagon promoters bind to a heterodimer of NeuroD and E12/47, a complex that is both enriched in islets and able to transactivate these promoters [20,21,40]. The majority of studies examining the mechanism by which NeuroD–E12/47 activates transcription have used the insulin promoter as a model and have shown that this factor synergistically activates transcription by co-operating with PDX-1, a β-cell-enriched homeodomain protein that binds to adjacent A/T-rich elements [41]. This synergy between NeuroD, E12/47 and PDX-1 is thought to result from both co-operative DNA binding and the formation of a strong interaction surface for the co-activators p300 and CBP [34,42,43]. Although this evidence collectively points to NeuroD–E12/47 as probably being the important activator operating through the insulin E elements, USF also binds *in vitro* to the rat insulin I E1 element (Figure 3A) and the rat insulin II E1 element [32]. Similarly, USF also binds to IGRP E-box 1 (Figure 2B). Although it is theoretically possible that either factor may operate through the rat insulin E elements and the IGRP E-box 1 motif, ChIP experiments show that, *in situ*, NeuroD binds the insulin E elements (E. Henderson and R. Stein, unpublished work). Similarly, although ChIP assays demonstrate that NeuroD and USF are both bound to the IGRP promoter *in situ* (Figure 6A), given the inability of a NeuroD-binding site to effectively replace E-box 2 (results not shown), the low affinity of E-box 1 for USF (Figure 5), as well as the inability of USF-2-VP16 to transactivate the IGRP fusion gene containing an E-box 2 mutation (results not shown), it is most likely that, *in situ*, USF binds IGRP E-box 2 and NeuroD binds IGRP E-box 1.

The importance of USF for islet-specific gene expression has previously been suggested by studies showing that USF binds to a functionally important element in the proximal region of the PDX-1 promoter [44]. Interestingly, the core sequence of this E-box in the PDX-1 promoter is optimal for USF binding [16,44], but, in contrast with the model proposed above for IGRP gene expression, this optimal binding of USF does not lead to expression of PDX-1 in multiple tissues. One possible reason for this is that the PDX-1 promoter is organized in a fundamentally different fashion. Indeed, the PDX-1 promoter has no TATA box [44] and therefore USF may act as part of the general transcriptional apparatus, as it appears to in the human transcobalamin II promoter, which also

© 2003 Biochemical Society

lacks a TATA-box motif [45]. PDX-1 plays a key role in pancreatic development, the maintenance of normal islet physiology in the adult and the regulation of insulin gene transcription [4,43,46]. The latter may explain why reducing USF levels in islet β-cell lines is sufficient to lower insulin mRNA levels, given that this manipulation may result in impaired PDX-1 expression [47].

Collectively the results of this study confirm the importance of E-box motifs to the regulation of islet gene expression. Future studies on the IGRP promoter will focus on characterizing the accessory factors that bind the IGRP promoter and either functionally or physically interact with USF and NeuroD. Additional experiments will also address whether IGRP gene expression is regulated by metabolites such as glucose *in vivo*. If so, it will be interesting to see if E-box 1 and/or E-box 2 play a role in these responses, given the importance of these motifs to the induction of insulin transcription by glucose [48,49].

We thank Howard Towle for providing the pCAT(An) and USF-2-VP16 expression vectors, Shimon Efrat for providing the βTC-3 cell line and Tim Osborne for advice on the performance of ChIP assays. Research in the laboratory of R.M.O'B. was supported by a grant from the Juvenile Diabetes Foundation International (JDFI). Research in the laboratory of R.S. was supported by National Institutes of Health (NIH) grant DK55091. Research in both laboratories was supported by NIH grant P60 DK20593, which funds the Vanderbilt Diabetes Center Core Laboratory. C.C.M. was supported by the Vanderbilt Viruses, Nucleic Acids and Cancer Training Program (5T32 CA09385-17).

REFERENCES

1 Ohneda, K., Ee, H. and German, M. (2000) Regulation of insulin gene transcription. Semin. Cell. Dev. Biol. **11**, 227–233
2 Bramblett, D. E., Huang, H. P. and Tsai, M. J. (2000) Pancreatic islet development. Adv. Pharmacol. **47**, 255–315
3 Habener, J. F. and Stoffers, D. A. (1998) A newly discovered role of transcription factors involved in pancreas development and the pathogenesis of diabetes mellitus. Proc. Assoc. Am. Physicians **110**, 12–21
4 Offield, M. F., Jetton, T. L., Labosky, P. A., Ray, M., Stein, R. W., Magnuson, M. A., Hogan, B. L. and Wright, C. V. (1996) PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development **122**, 983–995
5 St-Onge, L., Sosa-Pineda, B., Chowdhury, K., Mansouri, A. and Gruss, P. (1997) Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. Nature (London) **387**, 406–409
6 Sussel, L., Kalamaras, J., Hartigan-O'Connor, D. J., Meneses, J. J., Pedersen, R. A., Rubenstein, J. L. and German, M. S. (1998) Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. Development **125**, 2213–2221
7 Winter, W. E. (2000) Molecular and biochemical analysis of the MODY syndromes. Pediatric Diabetes **1**, 88–117
8 van de Werve, G., Lange, A., Newgard, C., Mechin, M. C., Li, Y. and Berteloot, A. (2000) New lessons in the regulation of glucose metabolism taught by the glucose 6-phosphatase system. Eur. J. Biochem. **267**, 1533–1549
9 Van Schaftingen, E. and Gerin, I. (2002) The glucose-6-phosphatase system. Biochem. J. **362**, 513–532
10 Arden, S. D., Zahn, T., Steegers, S., Webb, S., Bergman, B., O'Brien, R. M. and Hutton, J. C. (1999) Molecular cloning of a pancreatic islet-specific glucose-6-phosphatase catalytic subunit-related protein. Diabetes **48**, 531–542
11 Ebert, D. H., Bischof, L. J., Streeper, R. S., Chapman, S. C., Svitek, C. A., Goldman, J. K., Mathews, C. E., Leiter, E. H., Hutton, J. C. and O'Brien, R. M. (1999) Structure and promoter activity of an islet-specific glucose-6-phosphatase catalytic subunit-related gene. Diabetes **48**, 543–551
12 Martin, C. C., Bischof, L. J., Bergman, B., Hornbuckle, L. A., Hilliker, C., Frigeri, C., Wahl, D., Svitek, C. A., Wong, R., Goldman, J. K. et al. (2001) Cloning and characterization of the human and rat islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP) genes. J. Biol. Chem. **276**, 25197–25207
13 Bischof, L. J., Martin, C. C., Svitek, C. A., Stadelmaier, B. T., Hornbuckle, L. A., Goldman, J. K., Oeser, J. K., Hutton, J. C. and O'Brien, R. M. (2001) Characterization of the mouse islet-specific glucose-6-phosphatase catalytic subunit-related protein gene promoter by *in situ* footprinting. Correlation with fusion gene expression in the islet derived bTC-3 and hamster insulinoma tumor cell lines. Diabetes **50**, 502–514
14 Massari, M. E. and Murre, C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol. Cell. Biol. **20**, 429–440
15 Dang, C. V., Dolde, C., Gillison, M. L. and Kato, G. J. (1992) Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. Proc. Natl. Acad. Sci. U.S.A. **89**, 599–602
16 Bendall, A. J. and Molloy, P. L. (1994) Base preferences for DNA binding by the bHLH-Zip protein USF: effects of $MgCl_2$ on specificity and comparison with binding of Myc family members. Nucleic Acids Res. **22**, 2801–2810
17 Puri, P. L. and Sartorelli, V. (2000) Regulation of muscle regulatory factors by DNA-binding, interacting proteins, and post-transcriptional modifications. J. Cell. Physiol. **185**, 155–173
18 Arnold, H. H. and Winter, B. (1998) Muscle differentiation: more complexity to the network of myogenic regulators. Curr. Opin. Genet. Dev. **8**, 539–544
19 Gradwohl, G., Dierich, A., LeMeur, M. and Guillemot, F. (2000) Neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc. Natl. Acad. Sci. U.S.A. **97**, 1607–1611
20 Naya, F. J., Stellrecht, C. M. and Tsai, M. J. (1995) Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. Genes Dev. **9**, 1009–1019
21 Dumonteil, E., Laser, B., Constant, I. and Philippe, J. (1998) Differential regulation of the glucagon and insulin I gene promoters by the basic helix-loop-helix transcription factors E47 and BETA2. J. Biol. Chem. **273**, 19945–19954
22 Peyton, M., Moss, L. G. and Tsai, M. J. (1994) Two distinct class A helix-loop-helix transcription factors, E2A and BETA1, form separate DNA binding complexes on the insulin gene E box. J. Biol. Chem. **269**, 25936–25941
23 Kaytor, E. N., Shih, H. and Towle, H. C. (1997) Carbohydrate regulation of hepatic gene expression. Evidence against a role for the upstream stimulatory factor. J. Biol. Chem. **272**, 7525–7531
24 Jacoby, D. B., Zilz, N. D. and Towle, H. C. (1989) Sequences within the 5′-flanking region of the S14 gene confer responsiveness to glucose in primary hepatocytes. J. Biol. Chem. **264**, 17623–17626
25 Higuchi, R., Krummel, B. and Saiki, R. K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res. **16**, 7351–7367
26 Sambrook, J., Fritsch, E. F. and Maniatis, E. F. (1989) Molecular Cloning: a Laboratory Manual, 2nd edn, Cold Spring Harbor Press, Plainview, NY
27 Chapman, S. C., Ayala, J. E., Streeper, R. S., Culbert, A. A., Eaton, E. M., Svitek, C. A., Goldman, J. K., Tavare, J. M. and O'Brien, R. M. (1999) Multiple promoter elements are required for the stimulatory effect of insulin on human collagenase-1 gene transcription. Selective effects on activator protein-1 expression may explain the quantitative difference in insulin and phorbol ester action. J. Biol. Chem. **274**, 18625–18634
28 O'Brien, R. M., Noisin, E. L., Suwanichkul, A., Yamasaki, T., Lucas, P. C., Wang, J. C., Powell, D. R. and Granner, D. K. (1995) Hepatic nuclear factor 3- and hormone-regulated expression of the phosphoenolpyruvate carboxykinase and insulin-like growth factor-binding protein 1 genes. Mol. Cell. Biol. **15**, 1747–1758
29 Gerrish, K., Cissell, M. A. and Stein, R. (2001) The role of hepatic nuclear factor 1 alpha and PDX-1 in transcriptional regulation of the pdx-1 gene. J. Biol. Chem. **276**, 47775–47784
30 German, M. S., Moss, L. G., Wang, J. and Rutter, W. J. (1992) The insulin and islet amyloid polypeptide genes contain similar cell-specific promoter elements that bind identical beta-cell nuclear complexes. Mol. Cell. Biol. **12**, 1777–1788
31 Karlsson, O., Walker, M. D., Rutter, W. J. and Edlund, T. (1989) Individual protein-binding domains of the insulin gene enhancer positively activate beta-cell-specific transcription. Mol. Cell. Biol. **9**, 823–827
32 Whelan, J., Cordle, S. R., Henderson, E., Weil, P. A. and Stein, R. (1990) Identification of a pancreatic beta-cell insulin gene transcription factor that binds to and appears to activate cell-type-specific expression: its possible relationship to other cellular factors that bind to a common insulin gene sequence. Mol. Cell. Biol. **10**, 1564–1572
33 Viollet, B., Lefrancois-Martinez, A. M., Henrion, A., Kahn, A., Raymondjean, M. and Martinez, A. (1996) Immunochemical characterization and transacting properties of upstream stimulatory factor isoforms. J. Biol. Chem. **271**, 1405–1415
34 Glick, E., Leshkowitz, D. and Walker, M. D. (2000) Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter. J. Biol. Chem. **275**, 2199–2204
35 Orlando, V. (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked chromatin immunoprecipitation. Trends Biochem. Sci. **25**, 99–104
36 Gupta, R. K. and Moore, R. D. (1980) ^{31}P NMR studies of intracellular free Mg^{2+} in intact frog skeletal muscle. J. Biol. Chem. **255**, 3987–3993
37 Stafford, J. M., Wilkinson, J. C., Beechem, J. M. and Granner, D. K. (2001) Accessory factors facilitate the binding of glucocorticoid receptor to the phosphoenolpyruvate carboxykinase gene promoter. J. Biol. Chem. **276**, 39885–39891

© 2003 Biochemical Society

38 Gaudet, J. and Mango, S. E. (2002) Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. Science **295**, 821–825
39 Karlsson, O., Edlund, T., Moss, J. B., Rutter, W. J. and Walker, M. D. (1987) A mutational analysis of the insulin gene transcription control region: expression in beta cells is dependent on two related sequences within the enhancer. Proc. Natl. Acad. Sci. U.S.A. **84**, 8819–8823
40 Cordier-Bussat, M., Morel, C. and Philippe, J. (1995) Homologous DNA sequences and cellular factors are implicated in the control of glucagon and insulin gene expression. Mol. Cell. Biol. **15**, 3904–3916
41 Peers, B., Leonard, J., Sharma, S., Teitelman, G. and Montminy, M. R. (1994) Insulin expression in pancreatic islet cells relies on cooperative interactions between the helix loop helix factor E47 and the homeobox factor STF-1. Mol. Endocrinol. **8**, 1798–1806
42 Ohneda, K., Mirmira, R. G., Wang, J., Johnson, J. D. and German, M. S. (2000) The homeodomain of PDX-1 mediates multiple protein-protein interactions in the formation of a transcriptional activation complex on the insulin promoter. Mol. Cell. Biol. **20**, 900–911
43 Qiu, Y., Guo, M., Huang, S. and Stein, R. (2002) Insulin gene transcription is mediated by interactions between the p300 coactivator and PDX-1, BETA2, and E47. Mol. Cell. Biol. **22**, 412–420
44 Sharma, S., Leonard, J., Lee, S., Chapman, H. D., Leiter, E. H. and Montminy, M. R. (1996) Pancreatic islet expression of the homeobox factor STF-1 relies on an E- box motif that binds USF. J. Biol. Chem. **271**, 2294–2299
45 Li, N. and Seetharam, B. (1998) A 69-base pair fragment derived from human transcobalamin II promoter is sufficient for high bidirectional activity in the absence of a TATA box and an initiator element in transfected cells. Role of an E box in transcriptional activity. J. Biol. Chem. **273**, 28170–28177
46 Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K. and Edlund, H. (1998) Beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. Genes Dev. **12**, 1763–1768
47 Qian, J., Kaytor, E. N., Towle, H. C. and Olson, L. K. (1999) Upstream stimulatory factor regulates Pdx-1 gene expression in differentiated pancreatic beta-cells. Biochem. J. **341**, 315–322
48 Sharma, A. and Stein, R. (1994) Glucose-induced transcription of the insulin gene is mediated by factors required for beta-cell-type-specific expression. Mol. Cell. Biol. **14**, 871–879
49 German, M. S. and Wang, J. (1994) The insulin gene contains multiple transcriptional elements that respond to glucose. Mol. Cell. Biol. **14**, 4067–4075

Received 9 October 2002/3 January 2003; accepted 23 January 2003
Published as BJ Immediate Publication 23 January 2003, DOI 10.1042/BJ20021585

© 2003 Biochemical Society