Intron 1 sequences are required for pancreatic expression of the human proglucagon gene

Li Zhou, Min Nian, Jun Gu, and David M. Irwin

 Departments of Laboratory Medicine and Pathobiology and the Banting and Best Diabetes Centre, University of Toronto, Toronto, Ontario, Canada

Submitted 16 August 2005; accepted in final form 11 October 2005

Zhou, Li, Min Nian, Jun Gu, and David M. Irwin. Intron 1 sequences are required for pancreatic expression of the human proglucagon gene. Am J Physiol Regul Integr Comp Physiol 290: R634–R641, 2006. First published October 13, 2005; doi:10.1152/ajpregu.00596.2005.—The mammalian proglucagon gene is expressed in pancreatic islet A-cells, intestinal L-cells, and select neurons of the brain, where posttranslational processing results in the liberation of a unique profile of peptides. Despite the importance of proglucagon-derived peptides in human biology, little is known about the regulation of the human gene, as the rat gene has been the preferred model for understanding the regulation of proglucagon gene expression. Previously, we have shown that although the immediate promoter region of the rat proglucagon gene is sufficient for expression in pancreatic islet cells, the homologous human proglucagon promoter sequences are not sufficient. We have now used a comparative genomic approach to identify noncoding sequences near the human proglucagon gene that are conserved among mammals, and thus potentially are regulatory sequences. Our alignments identified three evolutionarily conserved noncoding regions (ECR), one is the immediate promoter region (ECR1), the second is about 5 kb 5’ to the mRNA start site (ECR2), and the third is near the 3’ end of the first intron (ECR3). Our in vitro transient transfection assays with reporter gene constructs that include the human ECR3 support expression in rodent islet cell lines. Complementary studies with transgenic mice possessing a reporter gene regulated by a human proglucagon gene promoter-intron 1 (including ECR3) sequences express the reporter gene in the pancreas, as well as the intestine and selected neurons. These studies suggest that conserved sequences within intron 1 of the human proglucagon gene are important for expression in the pancreas.

The human proglucagon gene is expressed in A-cells of the pancreatic islets, L-cells of the small and large intestine, and selected neurons of the brain (4, 15, 16). Tissue-specific posttranslational processing of proglucagon results in the production of glucagon in the pancreas and two glucagon-like peptides, GLP-1 and GLP-2, in the intestine (15, 27). The proglucagon-derived peptide hormones have essential and diverse roles in human physiology. Glucagon regulates carbohydrate, lipid, and amino acid metabolism and acts as a counterregulatory hormone to insulin in regulating blood glucose levels (14). GLP-1 is an incretin hormone that potentiates insulin secretion from the pancreas and has additional roles in gastric emptying and feeding behavior (1, 2, 3). GLP-1 is currently under development as a potential therapy for the treatment of diabetes (7, 20). GLP-2 is a key regulator of intestinal epithelial growth and function (1, 2, 3).

Extensive knowledge has been gained on the sequences and DNA-binding proteins that are required for expression of the rat proglucagon gene (4, 15, 16). Approximately 300 bases of rat proglucagon 5’ flanking sequence have been found to be sufficient to support expression of reporter genes in glucagon gene-expressing rodent islet cell lines (5). Within the immediate promoter of the rat proglucagon gene, four enhancer elements (designated G2 through G5) and an A-cell-specific promoter element (G1), as well as a large number of DNA-binding proteins have been identified and characterized (4, 15, 16). In contrast, much less is known about the regulation of the human proglucagon gene. Despite extensive sequence similarity of the immediate promoter region, short human proglucagon promoter-reporter gene constructs fail to generate significant reporter gene activity in rodent islet cell lines (21). It was concluded that divergent mechanisms are used for tissuespecific regulation of the human and rat proglucagon genes, as more than 3,300 bases of human proglucagon 5’-flanking sequences were required for expression in rodent islet cell lines (21). Similar conclusions were also drawn from observations with transgenic mice. Although 1,300 bases of rat proglucagon gene 5’ flanking sequences allowed expression of a transgene in islets of mice (6), 1,600 bases of human proglucagon 5’ flanking sequence did not generate expression in islets of mice, yet this human proglucagon promoter construct did support expression in two other proglucagon gene expressing tissues, the intestine, and selected neurons of the brain (21).

To help identify which lineage, either the primate lineage leading to human, or rodent lineage leading to rat, produces the pancreatic islet cell transcriptional activity change, proglucagon gene promoters from the chicken and diverse mammals have been characterized (Tsai B, Yue S, and Irwin DM, unpublished observations, Ref. 30). The chicken proglucagon gene promoter was found to have extremely limited sequence similarity to mammalian promoter, yet it could support reporter gene expression in rodent islet cell lines (30). This work suggested that significant changes in the sequence of the proglucagon promoter could occur yet retain promoter activity. A survey of proglucagon promoters from diverse mammals that included two primates (human and rhesus monkey), three rodents (rat, mouse, and hamster), a carnivore (dog), and an artiodactyl (cow) found that only the primate proglucagon promoters failed to support reporter gene expression in tran-
sient-transfection studies with rodent islet cell lines (Tsai B, Yue S, and Irwin DM, unpublished observations). These two studies strongly suggest that promoter sequence changes have occurred on the lineage leading to primates that have yielded an immediate promoter that fails to support expression in rodent islet cell lines. It was concluded that additional sequences, either further 5' or 3' to the mRNA start site, are necessary for the regulation of expression of the human proglucagon gene (Tsai B, Yue S, and Irwin DM, unpublished observations) (21, 30).

In this study, we have compared proglucagon gene sequences from diverse mammals and the chicken to identify evolutionarily conserved noncoding regions (ECRs) that potentially regulate expression of the human proglucagon gene. ECRs are sequences that do not code for protein yet have a regulatory function. Our genomic comparisons identified three ECRs: the immediate promoter region, a potential regulatory function, and a 5-kb region immediately downstream of the proglucagon promoter, both in vitro and in vivo.

MATERIALS AND METHODS

Reagents. Reagents and chemicals were purchased from BioShop Canada (Burlington, ON), Caledon (Georgetown, ON), Difco (Detroit, MI), AP BioTech (Baie d’Urfe, PQ), Sigma (St. Louis, MO), and Invitrogen (Burlington, ON). Restriction endonucleases and DNA-modifying enzymes were purchased from New England Biolabs (Mississauga, ON). Taq DNA polymerase, and deoxyuridinyl triphosphates were obtained from Roche (Laval, PQ). Turbo Pfu was purchased from Stratagene (La Jolla, CA). Oligonucleotide primers and DNA sequencing were purchased from ACGT (Toronto, ON), while pGL3 expression vectors were purchased from Promega (Madison, WI).

Bioinformatic analysis. Genomic sequences that encoded proglucagon genes were downloaded from the ensemble Web site (http://www.ensemble.org). Genome databases for human (NCBI35), mouse (NCBI33), dog (BROAD1), cow (Btau1.0), opossum (NCBI33), rat (RGSC3.4), and chicken (30) proglucagon cDNAs were downloaded for BLAST searches. The cow proglucagon genomic sequence was identified by searching the cow proglucagon cDNA sequence. The opossum proglucagon genomic sequence was identified from the opossum genome using the opossum BLAST search. The cow proglucagon genomic sequences were downloaded from the ensemble Web site (http://www.ensemble.org). The BLAST search identified genomic sequences similar to the mammalian proglucagon gene.

Generation of transgenic mice. The 5.7-kb genomic fragment that included 5775 bp of the 5' flanking region and 96 bp of exon 1 was ligated upstream of the human growth hormone (hGH) reporter gene. The 3.7-kb genomic fragment containing 602 bases of the 5' flanking region, all of exon 1 and intron 1, and 8 bases of untranslated region of exon 2 of the human proglucagon gene were ligated upstream of the β-galactosidase (β-gal) reporter gene. The University of Toronto Faculty of Medicine Transgene and Knockout Facility used the transgene constructs to generate transgenic founders in the FVB strain. Transgenic mice were characterized by Southern blot analysis or PCR analysis of tail DNA, as previously described (21). Primers for detection of transgene expression were designed in Table 1. Germ line transmission was identified from two founders for each transgene, with transgene expression characterized in mice from the second generation of each line. The University Health Network Animal Care Committee approved the experimental protocol for the generation and characterization of transgenic mice.

Reporter gene plasmids. Genomic fragments containing putative promoter and intron 1 regions of the human proglucagon gene were amplified by PCR using primers −602f and 3071r (see Table 1) from a P1 genomic clone that has previously been characterized (21) and cloned into pCR2.1 TA cloning vector (Invitrogen). Smaller promoter and intron fragments were amplified from this promoter-intron clone using the primers listed in Table 1, were released from the TA vector by digestion with appropriate restriction endonucleases (see Table 1), and cloned into pGL3 basic, pGL3 promoter, or modified pGL3 luciferase reporter gene vectors (Promega). Large genomic fragments (>1 kb) were amplified using Turbo Pfu (Stratagene), while smaller fragments were amplified with Taq (Roche, Laval PQ). The 602-base and minimal 82-base human promoters were amplified using the primers −602f and 58r and −82f and 58r, respectively. Intron 1 sequences were amplified using the primers 98f and 3063r. A series of 500-base intron fragments were amplified using the following pairs of primers: 98f and 552r, 553f and 1018r, and 1454f and 1455f and 1920r, 1921f and 2378r, 2379r and 2862r, and 2862f and 3063r. The DNA sequence of all promoter-reporter constructs was confirmed.

Cell culture and transfections. The hamster islet cell line InR1-G9, mouse islet cell line αTC-1, and baby hamster fibroblasts (BH7 cells) were grown in DMEM (4.5 g glucose/l) supplemented with 5% fetal calf serum and 1% penicillin and streptomycin. Luciferase reporter gene constructs were transfected into the cell lines using lipofectamine (Invitrogen). Controls for transfections included the promoterless pGL3 basic, the positive transfection pGL3 control vector (containing the SV40 promoter enhancer; Promega), and a control containing 312 bases of promoter and 58 bases of exon 1 of the rat proglucagon gene cloned into pGL3 Basic. Reporter gene construct activity was measured by detecting luciferase activity using a Lumat LB 9501 luminometer (EG&G Berthold, Wellesley, MA).

Data analysis. Results are presented as means ± SE from at least three independent experiments. Luciferase assays were performed in triplicate. Data were analyzed by ANOVA followed by Tukey’s test or paired Student’s t-tests using the computer software InStat 2.0 (GraphPad Software, San Diego, CA). Differences were considered statistically significant at P < 0.05.

RESULTS

Identification of evolutionarily conserved regions near the proglucagon gene promoter. Human (31), mouse (24), rat (10), dog (11), cow (18), and chicken (30) proglucagon cDNAs and/or genes have previously been characterized and were used to identify genomic sequences that included proglucagon genes and flanking sequences. To identify the opossum proglucagon gene, we searched the draft opossum genomic sequence using TBLASTN with mammalian proglucagon amino acid sequences. Our searches of the opossum genome identified a...
proglucagon gene that is similar in structure to previously characterized mammalian proglucagon genes, as it consisted of six exons and encoded a predicted proglucagon polypeptide that potentially could be proteolytically processed to release the proglucagon polypeptide (Fig. 1). We were unable to identify any sequence with greater than 50% identity to exon 3; bases that appear in italics are linkers added for cloning. "Position of 5' base (excluding added linker) relative to the mRNA start site. "Strand primer to which orientation is designed. Sense primers are +; antisense are −. "Sequence relative to the mRNA start site of the human growth hormone 1 mRNA variant 1 mRNA (accession no. NM_000515). "Sequence relative to the mRNA start site of the mouse GAPDH mRNA (accession no. BC096440). "Located within the luciferase coding sequence in the pGL3 vector sequence. "Sequences relative to the initiation codon of β-galactosidase. "Sequence relative to the mRNA start site of the mouse proglucagon mRNA (accession no. NM_008100). 

Table 1. Primers for construction and characterization of reporter genes

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Location</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hGH5′</td>
<td>GAAGAAGCCTATATCCCAAAG</td>
<td>234</td>
<td>+</td>
</tr>
<tr>
<td>hGH3′</td>
<td>GTAGTGCTCTAGTGTTG</td>
<td>609</td>
<td>−</td>
</tr>
<tr>
<td>GAPDH5′</td>
<td>CATTGACATGCTAGTCAGG</td>
<td>166</td>
<td>+</td>
</tr>
<tr>
<td>GAPDH3′</td>
<td>AGATCCGCATGCGAACAGG</td>
<td>795</td>
<td>−</td>
</tr>
<tr>
<td>7f</td>
<td>ACAAGCATTGAGACAGAGACAG</td>
<td>21</td>
<td>+</td>
</tr>
<tr>
<td>luc-r</td>
<td>CGTTGCTGACGCTGCGACACGAC</td>
<td>293</td>
<td>−</td>
</tr>
<tr>
<td>β-galr</td>
<td>TTAGAGACCTCGATGACTTCGCTC</td>
<td>88</td>
<td>+</td>
</tr>
<tr>
<td>β-galr</td>
<td>GCCGGTAACTCGGCTTCCATCG</td>
<td>688</td>
<td>−</td>
</tr>
<tr>
<td>mouse PG5′</td>
<td>TAGAACCCAATTATTCGCTG</td>
<td>103</td>
<td>+</td>
</tr>
<tr>
<td>mouse PG3′</td>
<td>CCTGTGCGCAAGATTGTGCCAGAAT</td>
<td>596</td>
<td>+</td>
</tr>
</tbody>
</table>

Rat proglucagon primers

r – 312f | GCGGACTTCCTAGCTGCTATTGATGTCAA | −312 | + |
|r + 58r | GCAACGTCAGGTTATAGCTGTTGGAAA | 58 | − |

plot of the mammalian and chicken proglucagon genes; similar PIP plots were generated using other mammalian sequences as the reference sequence (results not shown). Identification of ECRs requires comparisons of genomic sequences that have diverged enough to allow nonfunctional sequences to accumulate differences, whereas sequences under selection have accumulated few if any changes. Since sequences under selection can have differing levels of selective constraints, genome comparisons representing different amounts of evolutionary divergence can differentially detect conserved sequences with differing levels of constraints. The most conserved sequences will be detected by comparisons covering the greatest evolutionary divergence, whereas lower levels of conservation can be detected using more closely related genome sequences (28). When the chicken proglucagon genomic sequences are compared with mammalian proglucagon genomic sequences, the only sequences near the proglucagon gene that showed greater than 50% sequence identity are exons 2 through 6 of the human proglucagon gene, exons that encode the proglucagon polypeptide (Fig. 1). We were unable to identify any sequence with greater than 50% identity to exon 1.
show very high sequence conservation. ECR1 is the 300-base immediate immediate promoter region and is known to be important for expression of the rat proglucagon gene (5). ECR2 is ~5 kb upstream from the mRNA start site and contains a region and of more than 200 bases with greater than 90% identity between human and the opossum (and more than 300 bases showing greater than 90% identity within placental mammals). ECR3, which is found near the 3' end of intron 1 is even better conserved than the other two ECRs, composed of ~300 bases, showing greater than 92% identity between opossum and human (and ~300 bases with greater than 97% identity within placental mammals).

Expression of human proglucagon transgene with 5.7 kb of 5' flanking sequence. We have previously reported that 5.7 kb of human proglucagon gene 5' flanking sequence could drive expression of luciferase reporter gene in rodent islet cell lines (21). ECR2 is included within the 5.7 kb of 5' flanking sequence, thus potentially contributing to islet-specific expression of the proglucagon gene. To determine whether the 5.7 kb of 5' flanking sequence was sufficient to drive expression in glucagon-gene expressing cells in vivo, we generated the 5.7 GLU-GH transgene. Tissue-specific expression of the reporter gene was characterized in two founder lines of transgenic mice by RT-PCR (Fig. 2). hGH reporter transgene mRNA was detected in the stomach, duodenum, jejunum, ileum, colon, and brain, but not pancreas or liver of mice descending from founder no. 10 (Fig. 2A). In contrast, low levels of growth hormone transgene mRNA were detected by RT-PCR in all tested tissues of mice descendent from founder no. 6 (Fig. 2B). To confirm expression of the transgene product, immunocytochemistry was used to assay for expression of the transgene protein product hGH. hGH-immunoreactive cells were detected in the stomach, duodenum, jejunum, ileum, colon, and brain, but not in the pancreas or liver of mice from both line no. 10 and 6 (data not shown). Despite widespread expression of the transgene as detected by RT-PCR (Fig. 2B), immunocytochemistry could only detect the hGH reporter in the stomach, duodenum, jejunum, ileum, colon, and brain in mice from line no. 6, suggesting that only some tissues properly express the transgene to generate mRNA that can be translated to yield hGH.

Enhancer activity of human proglucagon intron 1. In addition to the ECRs identified in the 5' flanking sequence of the proglucagon gene sequence, an ECR was identified within intron 1 (Fig. 1). To determine whether intron 1 sequences could promote expression on the proglucagon gene in glucagon gene-expressing pancreatic islet cells, we constructed a reporter gene plasmid containing 300 bases of intron 1 (Fig. 2A). The levels of reporter gene activity for the 3.7-kb Glu-Luc construct, was found to drive significantly higher levels of luciferase reporter gene activity than the promoterless construct or one that contained only 602 bases of intron 1 (21). ECR1 is the 300-base immediate immediate promoter region and is known to be important for expression of the rat proglucagon gene (5). ECR2 is ~5 kb upstream from the mRNA start site and contains a region and of more than 200 bases with greater than 90% identity between human and the opossum (and more than 300 bases showing greater than 90% identity within placental mammals). ECR3, which is found near the 3' end of intron 1 is even better conserved than the other two ECRs, composed of ~300 bases, showing greater than 92% identity between opossum and human (and ~300 bases with greater than 97% identity within placental mammals).

Expression of human proglucagon transgene with 5.7 kb of 5' flanking sequence. We have previously reported that 5.7 kb of human proglucagon gene 5' flanking sequence could drive expression of luciferase reporter gene in rodent islet cell lines (21). ECR2 is included within the 5.7 kb of 5' flanking sequence, thus potentially contributing to islet-specific expression of the proglucagon gene. To determine whether the 5.7 kb of 5' flanking sequence was sufficient to drive expression in glucagon-gene expressing cells in vivo, we generated the 5.7 GLU-GH transgene. Tissue-specific expression of the reporter gene was characterized in two founder lines of transgenic mice by RT-PCR (Fig. 2). hGH reporter transgene mRNA was detected in the stomach, duodenum, jejunum, ileum, colon, and brain, but not pancreas or liver of mice descending from founder no. 10 (Fig. 2A). In contrast, low levels of growth hormone transgene mRNA were detected by RT-PCR in all tested tissues of mice descendent from founder no. 6 (Fig. 2B). To confirm expression of the transgene product, immunocytochemistry was used to assay for expression of the transgene protein product hGH. hGH-immunoreactive cells were detected in the stomach, duodenum, jejunum, ileum, colon, and brain, but not in the pancreas or liver of mice from both line no. 10 and 6 (data not shown). Despite widespread expression of the transgene as detected by RT-PCR (Fig. 2B), immunocytochemistry could only detect the hGH reporter in the stomach, duodenum, jejunum, ileum, colon, and brain in mice from line no. 6, suggesting that only some tissues properly express the transgene to generate mRNA that can be translated to yield hGH.

Enhancer activity of human proglucagon intron 1. In addition to the ECRs identified in the 5' flanking sequence of the proglucagon gene sequence, an ECR was identified within intron 1 (Fig. 1). To determine whether intron 1 sequences could promote expression on the proglucagon gene in glucagon gene-expressing pancreatic islet cells, we constructed a reporter gene plasmid containing 300 bases of 5' flanking sequence, all 96 bp of exon 1, all 2,970 bp of intron 1, and the first 12 (untranslated) bases of exon 2 of the human proglucagon gene ligated upstream of a luciferase reporter gene (3.7 kb Glu-Luc). The 3.7-kb Glu-Luc, containing the intron 1 construct, was found to drive significantly higher levels of luciferase reporter gene activity than the promoterless construct or one that contained only 602 bases of human proglucagon promoter (Fig. 3A). The levels of reporter gene activity for the 3.7-kb Glu-Luc construct were 50–60% of the levels generated by a 312-base rat proglucagon promoter construct (Fig. 3A).

To confirm that transcription initiated at the human proglucagon mRNA initiation site in our reporter gene construct, human proglucagon reporter gene cDNA was amplified by RT-PCR using a sense primers (7f) in exon 1 for the human
proglucagon gene and an antisense primers (luc-r) for the luciferase sequence (see Table 1 for primers). The cDNA product generated by RT-PCR was the predicted size, and its sequence showed that the mRNA generated by this reporter gene construct initiated with the proglucagon exon 1 sequence and had properly spliced intron 1 to join human proglucagon exon 1 and exon 2 sequences (results not shown). As an alternative approach, to exclude the possibility that intron 1 was acting as a promoter element, we cloned intron 1 into luciferase reporter plasmids with (pGL3P) and without (pGL3B) promoters. Significant reporter gene activity was only detected when intron 1 was cloned into a promoter-containing vector and not with a vector that lacks a promoter (Fig. 3B).

Enhancer activity of the evolutionarily conserved region in intron 1. To determine which portions of intron 1 (bases 97 to 3,063 of the proglucagon gene) contribute to the enhancer activity, a series of smaller fragments of intron 1 were tested for enhancer activity (Fig. 4A). A total of seven fragments, each of ~500 base pairs, spanning the intron were cloned upstream of the SV40 promoter in pGL3P. Only one fragment, spanning bases 2,379 to 2,862, significantly enhanced luciferase activity in rodent islet cell lines (Fig. 4A). This fragment from bases 2,379 to 2,862, which enhances reporter gene activity, corresponds to the ECR3 that was found within intron 1 of the proglucagon gene (Fig. 1). The short ECR3 portion of intron 1 was also found to enhance reporter gene activity when placed upstream of 332-base or 82-base human proglucagon gene promoters or a minimal 82-base rat proglucagon promoter, although to a lower extent than that generated from the 312-base rat promoter (Fig. 4B). No other fragment of intron 1 significantly enhanced reporter gene activity, but surprisingly, one portion of the intron, from bases 1,018 to 1,454, essentially silenced reporter activity (Fig. 4A). The sequence between bases 1,018 and 1,454 of the proglucagon gene does not appear to correspond to any highly conserved portion of intron 1, although most of intron 1 is better conserved than other noncoding sequences near the proglucagon gene (Fig. 1).

Fig. 3. Expression of human proglucagon promoter constructs that contain intron 1 sequences. To determine whether intron 1 sequences could direct tissue-specific expression of a reporter gene in vivo, a transgenic mouse was generated that contained 602 bases of human proglucagon 5' flanking sequence, all of exon 1 and intron 1, and the first 12 untranslated bases of exon 2.

Expression of human proglucagon transgene containing intron 1 sequences. To determine whether intron 1 sequences could direct tissue-specific expression of a reporter gene in vivo, a transgenic mouse was generated that contained 602 bases of human proglucagon 5' flanking sequence, all of exon 1 and intron 1, and the first 12 untranslated bases of exon 2.
ligated upstream of a β-gal reporter gene. This construct was identical to the construct in Fig. 3A, except for the reporter gene. A total of three founders were generated, two of which produced germ line transmission. To assess reporter gene expression, RT-PCR was conducted using primers for the β-gal sequence. As shown in Fig. 5A, reporter gene expression was detected in both lines showing germ line transmission in brain and intestinal tissues, as previously detected in 1.6 kb (21) and 5.7 kb (Fig. 2), 5′ flanking transgenic mice. In addition, both of our intron 1-containing transgenic mice also expressed the transgene in the pancreas and did not express the transgene in tissues, such as liver and muscle (Fig. 5A), suggesting that this intron 1-containing transgene construct expresses the reporter gene in all major sites of proglucagon expression. An attempt to amplify a portion of human proglucagon intron 1 (Fig. 5B) in these same cDNA samples failed to generate the expected band, indicating that the mRNA was not contaminated with genomic DNA or partially processed mRNA, yet GAPDH could be amplified from all samples (Fig. 5D), suggesting that the mRNA was intact and converted to cDNA. The pattern of expression of the transgene (Fig. 5A) was identical to that of the endogenous mouse proglucagon gene (Fig. 5C).

**DISCUSSION**

The availability of a number of mammalian genomes (e.g., human, mouse, rat, dog, cow, opossum, chicken, see www.ensemble.org) has allowed the use of comparative genomics to identify sequences that are evolving at a slower rate than expected for DNA sequences that are not functional (8, 19, 23, 28). Comparative genomics readily identifies protein coding regions of genes, but, in addition, it has found an equally large amount of the noncoding sequence that has been constrained by selective pressures, presumably because they have a biological function, such as regulatory sequences (19). We have Fig. 4. ECR3 in intron 1 is an enhancer. A: enhancer activities of fragments of human proglucagon gene intron 1. Schematic structure (not to scale) of luciferase reporter genes are shown to the left with luciferase activity in glucagon gene-expressing hamster islet cell line InR1-G9 and mouse islet cell line αTC-1 on the right. Reporter gene activity is expressed relative to the activity of the pGL3P plasmid. *Activity is significantly greater than that of the promoterless plasmid. B: intron 1 enhancer work on minimal proglucagon promoters. Schematic structure (not to scale) of luciferase reporter genes are shown to the left with luciferase activity in glucagon gene-expressing hamster islet cell line InR1-G9 and mouse islet cell line αTC-1 on the right. Reporter gene activity is expressed relative to the activity of the human or rat proglucagon promoter plasmids without the intron fragment. *Activity is significantly greater than that of the human or rat proglucagon promoter plasmids without the intron fragment. The rat 312-base promoter generated about 15–20 times more reporter gene activity compared with the 82-base human or rat promoters or the 332-base human promoter.

Fig. 5. Tissue-specific expression of β-galactosidase (β-gal) in an intron 1-containing human proglucagon promoter transgene. A–D: RT-PCR for β-gal, human proglucagon intron 1, mouse, proglucagon, and GAPDH, respectively, in various mouse tissues from a transgenic mouse that has β-gal downstream of 602 bases of 5′ flanking, all of exon 1 and intron 1, and 12 untranslated bases of exon 2 of the human proglucagon gene sequence. Tissues used for isolation of RNA for RT-PCR are indicated above the lanes, with negative referring to a mock reaction without RNA, and positive being an RT-PCR with a cloned DNA product. The positive control for A and B is the transgene used to make the mouse, for C, a cloned mouse proglucagon cDNA, and for D, a cloned mouse GAPDH cDNA.
applied these techniques to identify sequences near or within the human proglucagon gene that have been conserved within mammals (Fig. 1).

We first compared the genomic sequences encoding the proglucagon gene of the chicken to those sequences encoding proglucagon genes from mammals, thus representing sequence divergence of ~310 million years (9, 17), and found that the only conserved sequences are the exons of the gene (Fig. 1). The protein coding exons 2 through 5 that encode most of the proglucagon precursor show about 75% sequence conservation, while exon 6, which encodes the C-terminus of proglucagon and the 3' untranslated region, had more varied conservation (Fig. 1). Although our chicken proglucagon genomic sequence did contain the first exon of the gene (30), this exon was not conserved with the mammalian sequences. The mammalian proglucagon gene is flanked by the FAP gene, about 16 kb upstream, and the D1IV gene, about 65 kb downstream (12), with a similar structure observed in the chicken genome (30). The last exon of the upstream FAP gene is conserved between chicken and mammalian proglucagon genomic sequences (Fig. 1), whereas downstream, nothing similar was identified between the proglucagon and D1IV genes (data not shown). Similar results were observed with comparisons of proglucagon genomic sequences from fish or frogs to mammals (data not shown). Therefore, comparisons of chicken (or fish or frogs) and mammals fail to detect any ECRs, which potentially could be regulatory sequences.

Our next comparisons involved genomic sequences from the placental mammals (human, mouse, rat, dog, and cow), which diverged from each other over the past ~100 million years (9, 17). In comparisons within placental mammals, most of the nonrepetitive sequence was found to align between species (Fig. 1), with most of the “gaps” in the alignment due to deletion or insertion of sequence rather than low conservation. A large number of noncoding sequences are highly conserved between proglucagon genes of placental mammals (see Fig. 1). Comparisons that included the proglucagon gene sequence from the marsupial opossum, which diverged from placental mammals about 170 million years ago (9, 17), result in the identification of fewer ECRs (Fig. 2). In comparisons between the opossum and placental mammals, the noncoding sequences immediately upstream of exon 1 (ECR1), a sequence about 5 kb 5' to exon 1 (ECR2), and intron 1, especially near the 3' end of intron 1 (ECR3) (Fig. 2). The sequences of the three ECRs are better conserved than other flanking or intronic sequences, suggesting that they are being maintained by selection because they have a function, possibly as regulatory elements. In the opossum-placental mammal comparisons, the best conserved genomic sequence is ECR3 located in intron 1, which shows higher sequence identity than any of the protein coding exons, implying that it is undergoing a very slow rate of sequence evolution. Yet despite a slow rate of sequence evolution in mammals for ECR3, no sequence similar to ECR3 was found within the chicken proglucagon gene (see Fig. 1). This suggests that the selective constraints, and hence function of the sequence, have changed between chicken and mammals. A similar change appears to have also occurred for part of ECR1 in the immediate promoter region between these species (Tsai B, Yue S, and Irwin DM, unpublished observations).

ECR1 has a clear role in the regulation of rat proglucagon gene expression. ECR1 corresponds to the well-characterized immediate promoter region of the rat proglucagon gene (4, 15, 16), which in many mammals is sufficient to support expression in rodent islet cell lines (Tsai B, Yue S, and Irwin DM, unpublished observations). Although this sequence is well conserved in the human proglucagon gene, it is not sufficient to support reporter gene expression in vitro or in vivo, suggesting that a few changes may prevent transcriptional activity (21). Additional sequences have been hypothesized to be necessary for proper expression of the human proglucagon gene (21). In vitro transfection assays suggested that human proglucagon gene sequences 3.3 to 5.7 kb upstream of the mRNA start site could support expression of a reporter gene in rodent islet cell lines (21). ECR2 at about 5 kb upstream of the human proglucagon mRNA start site falls within the sequence defined by transient transfection assays as having enhancer activity. Transgenic mice, which include the 5.7 kb of 5' flanking sequence, including ECR2, failed to express the reporter gene in the pancreas, yet did express the reporter in the other major sites of endogenous proglucagon gene expression, the intestine and brain (Fig. 3). Thus the presence of both ECR1 and ECR2 was not sufficient to drive expression of a reporter gene in all sites of expression of the endogenous proglucagon gene.

Genomic comparisons indicated that the most conserved portion of mammalian proglucagon genes is not a protein coding exon but rather is a 450-base segment (ECR3) near the 3' end of intron 1 (Fig. 1) and potentially confers a regulatory function. Testing the enhancer activity of the proglucagon gene intron 1 is greatly facilitated by the fact that the initiator methionine for human gene (and other mammalian) proglucagon gene is encoded by exon 2 (31). Thus all of the exon 1 and intron 1, together with the 5' end of exon 2, could be included in our constructs, as they would only contribute 5' untranslated sequence and would not interfere with the expression, or function, of downstream reporter genes. Both in vitro and in vivo, reporter gene constructs that contained 602 bases of 5' flanking sequence (containing ECR1), all of exon 1 and intron 1 (containing ECR3), and the first 12 bases of exon 2, allowed expression of reporter genes in rodent islet cell lines (Figs. 3 and 5). Because the 602 bases of 5' flanking sequence, as well as exon 1 sequences, did not support reporter gene expression (Fig. 3), these results suggest that sequences within intron 1 were acting as an enhancer. Analysis of fragments of intron 1 suggests that most of the enhancer activity is located in a fragment between bases 2,379 and 2,860 of intron 1 (Fig. 5), the sequence that corresponds to ECR3 (Fig. 2). Searches of the ECR3 sequence identified potential binding sites for transcription factors such as HNF3β, Cdx-2/3, GATA factors, homeobox-containing factors (e.g., Isl-1, Brm4), and Pax factors. The potential transcription factor binding sites included many factors that previously had been identified as having a role in the expression of the rat proglucagon gene (4, 15, 16).

Most of intron 1 is better conserved than most of the noncoding sequences flanking or within mammalian proglucagon genes (Fig. 1), suggesting that selective pressure is conserving the sequence, possibly for additional functions. The observation that a sequence between bases 1,018 and 1,454 may repress expression supports this suggestion. Sequences distributed through intron 1 may act to modulate expression of proglucagon rather than directly act as an enhancer. The disassociation of sequences necessary for expression and those necessary for regulation has previously been demonstrated for.
intestinal expression of the human proglucagon gene. While 1.6 kb of 5′ flanking sequence is sufficient for expression in intestinal cell (21), these sequences were not sufficient for regulated expression via nutrient sensing (22).

ACKNOWLEDGMENTS

This work was supported in part by an operating grant from the Canadian Institutes of Health Research (to D. M. Irwin) and Studentships from the Banting and Best Diabetes Centre (to L. Zhou and M. Nian). We thank Drs. D. Drucker, H. Elsholtz, T. Jin, and anonymous reviewers for suggestions.

REFERENCES