Pancreatic Islet Expression of the Homeobox Factor STF-1 Relies on an E-box Motif That Binds USF*

Seema Sharma‡§, James Leonard¶, Soon Lee‡, Harold D. Chapman**, Edward H. Leiter***‡, and Marc R. Montminy$$

From The Clayton Foundation Laboratories for Peptide Biology, The Salk Institute, La Jolla, California 92037, the §School of Medicine, Department of Biology, University of California, San Diego, La Jolla, California 92037, the ¶Strang-Cornell Cancer Research Laboratory, New York, New York 11203, and the **Jaxton Laboratories, Bar Harbor, Maine 04609

The vertebrate pancreas consists of endocrine and exocrine components, which arise from a common progenitor cell in the duodenal anlage (1). Within the endocrine component of the pancreas, a pluripotent precursor cell, which initially expresses multiple islet hormones, undergoes progressive restriction to form the four subpopulations of cells comprising the adult islets of Langerhans: insulin, somatostatin, glucagon, and pancreatic polypeptide-producing cells (2, 3). The mechanism by which these developmental pathways are activated is unclear, but current evidence implicates the homeobox factor STF-1 (IPF-1) as an important determinant in this process. Indeed, the requirement for STF-1 in development is supported by homologous recombination studies in which targeted disruption of the STF-1/IPF-1 gene leads to congenital absence of the pancreas (4).

STF-1 (also referred to as Pdx1) expression is first detectable at embryonic day 8.5 in cells of the pancreatic anlage and in pluripotent precursor cells. Transiently expressed in both endocrine and exocrine components of the developing pancreas, STF-1 production is progressively restricted to insulin- and somatostatin-producing islet cells (5, 6). In these cells, STF-1 appears to regulate both insulin and somatostatin genes by binding to functional elements within each promoter (5, 7–16).

Although STF-1 appears to be an important regulator of pancreatic genes, the mechanism by which STF-1 expression is itself targeted to pancreatic cells remains uncharacterized. Here, we show that a 6.5-kb fragment of the STF-1 promoter is sufficient to direct islet-specific expression of a β-galactosidase reporter gene in transgenic mice. Within this 6.5-kb fragment, an E-box element located at -104 is recognized primarily by the helix loop helix/leucine zipper nuclear factor USF. As point mutations within the -104 E-box that disrupt USF binding correspondingly impair STF-1 promoter activity, our results demonstrate that USF is an important component of the regulatory apparatus which directs STF-1 expression to pancreatic islet cells.

The commitment of cells to specific lineages during development is determined in large part by the relative expression of various homeodomain (HOX) selector proteins, which mediate the activation of distinct genetic programs. But the mechanisms by which individual HOX genes are themselves targeted for expression in different cell types remain largely uncharacterized. Here, we demonstrate that STF-1, a homeodomain protein that functions in pancreatic morphogenesis and in glucose homeostasis is encoded by an “orphan” homeobox gene on mouse chromosome 5. When fused to a β-galactosidase reporter gene, a 6.5-kibase genomic fragment of 5′-flanking sequence from the STF-1 gene shows pancreatic islet specific activity in transgenic mice. Two distinct elements within the STF-1 promoter are required for islet-restricted expression: a distal enhancer sequence located between -3 and -6.5 kilobases and a proximal E-box sequence located at -104, which is recognized primarily by the helix loop helix/leucine zipper nuclear factor USF. As point mutations within the -104 E-box that disrupt USF binding correspondingly impair STF-1 promoter activity, our results demonstrate that USF is an important component of the regulatory apparatus which directs STF-1 expression to pancreatic islet cells.

MATERIALS AND METHODS

Chromosome Mapping—Chromosome mapping of the STF-1 gene was performed using a (B6 × SPRET)F1 × SPRET backcross panel of DNAs from The Jackson Laboratory Backcross DNA Panel Map Service using a 22P-labeled STF-1 cDNA fragment as a probe.

Generation of Transgenic Mice and β-Galactosidase Staining—A fusion gene containing 6500 bp of upstream STF-1 sequence in front of the β-galactosidase reporter gene was constructed using standard cloning techniques and injected into male pronuclei of fertilized oocytes. Founder mice were identified using Southern blotting and polymerase chain reaction amplification techniques. Expression of the STF-1/β-galactosidase gene in transgenic tissues was evaluated on 20 μm sections of paraformaldehyde-fixed tissues using X-Gal as chromogenic substrate.

Antibodies—Nonselecting USF-1, USF-2 antibodies used in supershift experiments were purchased from Santa Cruz Biotechnology. TFE3 antibodies were generous gift of K. J ones. Antibodies that specifically recognize USF-1 or USF-2 were provided by M. Sawadogo (17, 18).

Isolation of STF-1 Genomic Clones—The STF-1 gene was isolated from an EMBL 3 rat genomic library using a 22P-labeled STF-1 cDNA fragment as hybridization probe. STF-1 positive genomic fragments were cloned into the EcoRI sites of the SK II plasmid (Stratagene).

RNase Protection and Primer Extension—Poly(A) RNA was prepared using an Oligo(dT)30 system (Quiagen). Oligonucleotide primers for primer extension analysis were 5′-end labeled using γ-22P-ATP and T4 polynucleotide kinase. 5 μg of Poly(A) RNA was incubated with end-labeled antisense primer at 80 °C for 5 min followed by 16 h of incubation at 42 °C. Primer extension reactions were performed using

*This work was supported by the Foundation for Medical Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported in part by National Institutes of Health-MSTP Training Grant GM07198.

§ Supported by National Institutes of Health Grants DK36175 and DK27722.

¶ Foundation for Medical Research Investigator. To whom correspondence should be addressed: The Clayton Foundation Laboratories for Peptide Biology, The Salk Institute, 10010 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 619-453-4100 (ext. 1502); Fax: 619-552-1546.

**Supported by the Foundation for Medical Research. The Salk Institute, La Jolla, California 92037, the School of Medicine, Department of Biology, University of California, San Diego, La Jolla, California 92037, and the Jaxton Laboratories, Bar Harbor, Maine 04609.

†† The abbreviations used are: kb, kilobase(s); bp, base pair(s); X-Gal, 5-bromo-4-chloro-3-indolyl β-D-galactosidase; Luc, luciferase; CAT, chloramphenicol acetyltransferase.
USF Promotes Islet-specific STF-1 Expression

- Chromosomal location of the STF-1 gene. The STF-1 gene is encoded by an orphan homeobox gene located in the distal region of chromosome 5 (mouse). Top, schematic diagram showing position of STF-1 (referred to as Pdx1) relative to other markers on mouse chromosome 5. Centimorgan scale is indicated on right. Bottom, table showing recombination frequency between STF-1 and various markers on chromosome 5. Left column indicates markers employed for chromosomal assignment. STF-1 referred to here as Pdx1.

- avian myeloblastosis virus reverse transcriptase at 37 °C for 1 h, and products were analyzed on 5% denaturing polyacrylamide. RNase protection analysis was performed using 25 μg of total RNA extracted from Tu6 cells (13). Antisense STF-1 RNA probe was generated using a STF-1 genomic fragment extending from −185 to +93 relative to the translation start site. 32P-Labeled antisense STF-1 RNA was synthesized in vitro using T7 RNA polymerase and [32P]UTP. STF-1 antisense RNA probe and mRNAs were annealed at 80 °C for 5 min followed by 16 h of incubation at 65 °C. Annealing reactions were subsequently treated with 40 μg/ml RNase at room temperature for 1 h, and the digestion products were analyzed on 5% denaturing polyacrylamide electrophoresis.

- Reporter Clones and Transient Transfections—All promoter fragments were cloned into the pA0 luciferase backbone, provided by D. Helinski (20), using standard cloning methods. 5'-Flanking sequences from the STF-1 promoter were fused to the luciferase gene at +68 (−600 STF Luc, −3500 STF Luc, −540 STF Luc, −410 STF Luc, −225 STF Luc, −140 STF Luc) or at +78 (−190 STF Luc, −130 STF Luc, −120 STF Luc, −95 STF Luc, −35 STF Luc) relative to the major transcriptional initiation site. Plasmids were transfected into HIT-T15 cells (ATCC), βTC 3 (kindly provided by D. Hanahan (21)), PC12, COS, and HeLa cells by calcium phosphate precipitation. Luciferase values were quantitated on a Monolight luminometer and normalized to transfection efficiency.

- Gel Shift Assays and DNase Protection Assays—For electrophoretic mobility shift assays, oligonucleotide probes were labeled with [α-32P]dCTP by fill-in reaction using Klenow fragment. 4 μg of nuclear extract was incubated with 0.5 ng of 32P-labeled, double-stranded oligonucleotide and subjected to nondenaturing polyacrylamide electrophoresis as described previously (9). For supershift analysis, proteins were preincubated with the antibody followed by incubation with radiolabeled double-stranded oligonucleotide and electrophoresis. DNase protection assays were performed as described previously (22).

RESULTS

Chromosomal Location and Genomic Organization of the STF-1 Gene—Using the STF-1 cDNA as a hybridization probe on a backcross panel of DNAs from Jackson Laboratories, we mapped the single copy STF-1 gene to the distal region of mouse chromosome 5 (Fig. 1, top). No recombinants were found with the distal markers Pmv12 or Iapls3-9 while six recombinants were observed with the more distal Actb locus (Fig. 1, bottom). These results predict that the STF-1 gene would cor-
respondingly be found on rat chromosome 14 and human chromosome 7q, loci which do not correspond to any of the four homeotic HOX clusters. These results indicate that STF-1 should be classified as an "orphan" homeobox gene.

To isolate the gene encoding STF-1, we screened 106 bacteriophage clones from a rat EMBL 3 genomic library with a 32P-labeled STF-1 cDNA probe and obtained two positive clones, each containing a genomic insert of 15 kb. In addition to 6.5 kb of 5' flanking and 3.5 kb of 3' flanking sequence, the 15-kb STF-1 genomic fragment contained the entire STF-1 coding region, which was interrupted by a single 4-kb intron inserted immediately upstream (A1a-135) of the homeobox coding sequence (amino acids 140–215) (Fig. 2A).

The absence of consensus TATA box or initiator sequences in the 5' flanking region of the STF-1 genomic clone (Fig. 2A) prompted us to map the transcriptional initiation sites for this gene. Using RNase protection and primer extension analysis on mRNAs from the insulin-producing cell lines RIN and Tu6 (Fig. 2, A and B), we identified three principle initiation sites, termed S1, S2, and S3, which were located 91, 107, and 120/125 nucleotides upstream of the translational start site, respectively. A fourth minor transcriptional initiation site 137 nucleotides upstream of the translational start site was also observed. Like other TATA-less promoters, the STF-1 promoter contains G/A and G/C-rich sequences 30 bp upstream of the S1 and S2 start sites (23–25).

STF-1 Promoter Activity in Pancreatic Islet Cells—To determine whether sequences within the 5' flanking region of the STF-1 gene were sufficient to target expression of STF-1 to pancreatic islet cells, we fused 6500 bp of 5' flanking STF-1 sequence to the β-galactosidase gene and examined the expression of this STF-1-lacZ reporter in transgenic mice. Using X-Gal as chromogenic substrate, we detected β-galactosidase activity in pancreatic islets of transgenic but not control littermates from three independent founder lines (Fig. 3). In keeping with the previously described expression pattern for endogenous STF-1 protein, no significant β-galactosidase activity was detected in exocrine acinar cells (Fig. 3) or in non-pancreatic tissues such as liver and spleen of transgenic mice (not shown). In keeping with the reported expression of the endogenous STF-1 gene in the duodenum (8, 13), in situ hybridization studies with antisense β-galactosidase RNA probe also revealed transgene expression in epithelial cells of the duodenum from transgenic animals (not shown). These results indicate that 6500 bp of the STF-1 promoter are indeed sufficient to target expression of STF-1 to pancreatic islet and duodenal cells.
To define functional elements that direct STF-1 expression to pancreatic islet cells, we examined the activity of the −6500 STF-1 Luc reporter in two distinct pancreatic islet cell lines (βTC 3, HIT). As predicted from results in transgenic mice, the STF-1 reporter showed 20–100-fold more activity in these islet cells compared to non-islet lines such as HeLa, PC12, and COS (Fig. 4A). By contrast, the 4-kb intron and 3-kb 3′-flanking region of the STF-1 gene showed no such activity when inserted into a minimal SV40 chloramphenicol acetyltransferase promoter plasmid (not shown), suggesting that the 6.5-kb STF-1 promoter fragment is specifically responsible for targeted expression of STF-1 in islet cells.

To delineate sequences within the STF-1 promoter that confer islet cell expression, we generated a series of 5′-deletion constructs and analyzed these reporters by transfection into HIT cells (Fig. 4B). Deletion of sequences from −6500 to −3500 bp from the −6500 STF reporter construct reduced STF-1 reporter activity 4-fold, suggesting the presence of a distal activating sequence within that region. Further truncation of the STF-1 promoter from −3500 to −190 bp did not affect reporter activity in HIT cells significantly (Fig. 4B), but deletion of STF-1 promoter sequences from −190 to −95 bp severely attenuated reporter activity in HIT cells, indicating that a proximal element was also required for STF-1 promoter function. Inspection of the sequence in the −190 to −95 region of the STF-1 promoter revealed three consensus E-box motifs (Fig. 2A). Although removal of two tandem E-boxes at −177 did not reduce promoter activity, deletion of the proximal E-box sequence at −104 (−95 STF Luc) completely abolished STF-1 expression in HIT cells.

A Proximal E-box in the STF-1 Promoter Recognizes a USF-Containing Complex—To characterize upstream factors that bind to functional elements in the STF-1 promoter, we performed DNase I protection assays using nuclear extracts from HIT and HeLa cells (Fig. 5A). In both extracts, we observed a predominant footprinting activity whose boundaries coincided with the functionally important proximal E-box motif (−118/−95). To further characterize the proteins that bind to the critical −104 E-box motif in HIT versus HeLa extracts, we performed gel mobility shift assays. Using a double-stranded STF-1 oligonucleotide extending from −118 to −95, we observed three complexes, termed C1, C2, and C3 (Fig. 5B). Formation of C1, C2, and C3 complexes was inhibited by a 50-fold excess of unlabeled STF 1E-box competitor oligonucleotide in binding reactions. Mutant E-box oligonucleotide or non-specific competitor DNAs had no effect on these binding activities, however, indicating that C1, C2, and C3 are indeed specific for the STF-1 E-box sequence (Fig. 5B). No qualitative difference in the pattern of these complexes was detected between HeLa and HIT extracts (not shown), suggesting that the −104 E-box motif may recognize factors that are comparably expressed in both cell types.

Previous reports demonstrating that E-boxes like the −118/−95 STF-1 motif (CACGTG) preferentially bind bHLH-ZIP proteins such as Myc, Max, TFE-3, TFE-B, and USF prompted us to examine whether any of these candidate proteins was contained within the C1, C2, or C3 complexes (26, 27). The ability of the −118/−95 E-box binding protein to withstand
heat denaturation (not shown) led us to first examine whether USF, a heat-stable upstream factor, was a component of C1, C2, or C3 (28). Remarkably, addition of anti-USF antiserum to gel mobility shift reactions inhibited formation of all three complexes (Fig. 5, left panel), but anti-TFE-3 antiserum had no effect on complexes C1, C2, or C3, suggesting that these complexes were most likely formed by USF proteins. In gel shift assays, recombinant USF-1 gave rise to a protein DNA complex, which migrated at the same relative position as complex C2 (not shown), and in DNase I protection studies, recombinant USF-1 footprinting activity coincided with that observed in HIT extracts (Fig. 5A).

Two forms of USF, termed USF-1 and USF-2, appear to be expressed in most cell types (18). To distinguish which of these USF proteins was contained within the C1, C2, and C3 complexes, we incubated HIT or HeLa extract with either anti-USF-1 specific or anti-USF-2 specific antiserum (Fig. 5C, right panel). Although USF-1 antiserum could “supershift” all three complexes, the USF-2 specific antiserum only inhibited formation of the C1 and C3 complexes. These results suggest that complexes C1 and C3 correspond to USF-1-USF-2 heterodimers, whereas C2 may contain a USF-1 homodimer.

To verify whether the CACGTG E-box sequence was essential for STF-1 promoter activity, we constructed a mutant STF-1 oligonucleotide that contains two base pair substitutions in the E-box (−118/−95). In gel mobility shift assays with HIT nuclear extracts, this mutant E-box motif (AACGCG) could not form C1, C2, and C3 complexes and could not compete for binding of USF-1 to wild-type E-box oligonucleotide (Fig. 6A). Correspondingly, full-length (6.5 kb) STF-1 and truncated (−190 STF) reporter plasmids containing the mutant STF E motif were nearly 10-fold less active than their wild-type counterparts in pancreatic islet cells (Fig. 6B). These results indicate that the proximal E-box, which binds USF, is indeed critical for STF-1 promoter activity.

**DISCUSSION**

The majority of vertebrate homeobox genes are confined to four chromosomal clusters, termed HOX A–D (29, 30). Within each cluster, individual homeobox genes are ordered in a 5′ to 3′ pattern, which is co-linear with each antero-posterior expression pattern during development. It is not entirely clear whether this colinear organization is critical for proper expression of hox genes, but current evidence suggests that such clusters may contain upstream enhancers that coordinately regulate hox gene expression (29, 31). Remarkably, the STF-1 gene does not map to any hox cluster but rather belongs to a group of so-called orphan hox genes. Although the regulatory implications of this distinct chromosomal location for STF-1 remain to be shown, our results suggest that orphan homeobox genes like STF-1 may be regulated by signals that are distinct from those employed for the HOX clusters. In this regard, the STF-1 promoter displays pancreatic islet cell-specific activity both in transgenic animals as well as in transient transfection assays, and the lineage specific activity of this transgene contrasts with the segmental expression pattern of most hox genes.

Two elements within the first 6500 bp of STF-1 5′-sequence appear to be important for islet-specific expression: a distal element located between −6500 and −3500 and a proximal element located at −104. Although the identity of the distal element remains to be elucidated, the proximal −104 element consists of a consensus E-box motif that predominantly recognizes the upstream activator USF. Multiple lines of evidence suggest that USF is important for STF-1 promoter activity. First, both non-discriminating USF-1 and USF-2 antibodies as
well as USF-1- and USF-2-specific antibodies recognize the complexes specific for the STF-1 E-box. Second, the STF-1 E-box binding activity in HIT nuclear extracts has characteristics reminiscent of USF: the complexes are heat stable and demonstrate half-lives similar to recombinant USF-1. Finally, point mutations that inhibit formation of USF complexes on the STF E-box correspondingly attenuate STF-1 reporter activity. These results suggest that USF complexes are indeed important for STF-1 promoter activity and consequently for pancreatic organogenesis.

Other nuclear factors in addition to USF, most notably Myc and Max, can also bind with high affinity to the STF-1 E-box (CACGTG) motif. Myc has been shown to stimulate target gene transcription by binding as a heterodimer with Max to E-box motifs (32, 33). As myc gene expression is typically undetectable in post-mitotic cells such as those in pancreatic islets, Myc-Max complexes may not be involved in STF-1 promoter regulation there. During development, however, STF-1 expression appears to be concentrated in proliferating ductal cells (6), and myc may consequently stimulate STF-1 expression under those conditions. In this regard, it is tempting to speculate that the profound changes in STF-1 expression, which are observed during pancreatic development, may in part reflect changes in E-box binding activities that ultimately restrict STF-1 production to pancreatic islet cells.

Acknowledgments—We thank M. Sawadogo for USF-1 and USF-2 specific antibodies, M. Voz for providing recombinant USF-1 protein, D. Hanahan for βTC 3 cells, and members of the Montminy laboratory for helpful discussions. We also thank Lucy Rowe for providing the resources of the Jackson Laboratory Backcross DNA Panel Map Service.

REFERENCES

1. Pictet, R. L., Clark, W. R., Williams, R. H., and Rutter, W. J. (1972) Dev. Biol. 29, 436–467
2. Gittes, G. K., and Rutter, W. J. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 606–609

2 M. R. Montminy and S. Sharma, unpublished observations.
Pancreatic Islet Expression of the Homeobox Factor STF-1 Relies on an E-box Motif That Binds USF
Seema Sharma, James Leonard, Soon Lee, Harold D. Chapman, Edward H. Leiter and Marc R. Montminy

J. Biol. Chem. 1996, 271:2294-2299. doi: 10.1074/jbc.271.4.2294

Access the most updated version of this article at http://www.jbc.org/content/271/4/2294

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 33 references, 14 of which can be accessed free at http://www.jbc.org/content/271/4/2294.full.html#ref-list-1