

---

# Perspectives in Diabetes

## Comparative Analysis of Insulin Gene Promoters

### Implications for Diabetes Research

Colin W. Hay and Kevin Docherty

DNA sequences that regulate expression of the insulin gene are located within a region spanning ~400 bp that flank the transcription start site. This region, the insulin promoter, contains a number of *cis*-acting elements that bind transcription factors, some of which are expressed only in the  $\beta$ -cell and a few other endocrine or neural cell types, while others have a widespread tissue distribution. The sequencing of the genome of a number of species has allowed us to examine the manner in which the insulin promoter has evolved over a 450 million-year period. The major findings are that the A-box sites that bind PDX-1 are among the most highly conserved regulatory sequences, and that the conservation of the C1, E1, and CRE sequences emphasize the importance of MafA, E47/ $\beta$ 2, and cAMP-associated regulation. The review also reveals that of all the insulin gene promoters studied, the rodent insulin promoters are considerably dissimilar to the human, leading to the conclusion that extreme care should be taken when extrapolating rodent-based data on the insulin gene to humans. *Diabetes* 55:3201–3213, 2006

**T**he cloning and sequencing of the insulin gene in 1980 (1) was a landmark breakthrough that opened up a new field of research on the mechanisms controlling expression of the gene. This in turn led to the discovery of transcription factors that, in addition to regulating the insulin gene in a tissue-specific and temporal manner, participated in the development of the endocrine pancreas and in the maintenance of islet cell function (2). Some of these transcription factors have been identified as maturity-onset diabetes of the young (MODY) genes (3), and at least one has been associated with type 2 diabetes. Their use in the development of novel therapies for diabetes based on the differentiation of embryonic or adult stem cells toward a  $\beta$ -cell-like phenotype (4) and the forced expression of endogenous insulin genes in nonislet cells (5–7) has also been exploited.

The early work on characterizing the DNA sequences involved in regulating insulin gene expression focused on the rat insulin 1 gene (8). The reason for this was that at

the time there were no available human  $\beta$ -cell lines and it was felt important to correlate data from transfected promoter constructs with effects on the endogenous insulin gene. As it turned out, most of the studies involved transfecting the rat insulin gene constructs in the Syrian hamster HITm2.2.5 cell line, which transfected much more efficiently using available techniques than the rat RINm5F cell line. There was also a perception that human insulin promoter constructs would not function in transfected rodent cells. However, these worries proved to be unfounded after it was later shown that there is a very high degree of sequence and functional conservation within the transcription factors that regulate the gene (e.g., 89% identity between rat and human PDX-1) and the human insulin promoter exhibited the expected pattern of activity in transgenic mice (9,10). As a result of the decision to concentrate on a detailed analysis of the rat insulin promoter, most of the literature on the insulin promoter pertains to this promoter.

The structure and evolution of the insulin gene has been previously reviewed (11). In this article, we focus on the sequences that lie upstream of, or flank, the transcription start site and are known to affect transcription of the gene. One major conclusion is that the rodent promoters are markedly different from the human promoter, and we urge caution in extrapolating data from rodent promoter studies to the etiology and therapy of diabetes.

#### INSULIN GENE EXPRESSION

Humans, in keeping with the overwhelming majority of species, have a single copy of the insulin gene, which is located on chromosome 11 (p15.5) (12). Of the small number of species with two nonallelic insulin genes, the best known are *Xenopus laevis* (13) and the popular laboratory research rodents of rat (14) and mouse (15), with insulin two corresponding to the single copy in most animals.

In the adult insulin is expressed almost exclusively in the  $\beta$ -cells of the pancreatic islets of Langerhans (16), hence its name from Latin *insula* or “island.” Low levels of extrapancreatic insulin have been detected in a number of other tissues (17,18) including brain (19), thymus (20–22), lachrymal glands (23), and salivary glands (24). The role of insulin expression in non- $\beta$ -cells is unclear. In some tissues it may play a role in the complex hormonal communication required for the maintenance of overall energy balance (25,26) or in the establishment of immune tolerance (27). Very little is known about the regulatory sequences that control insulin gene expression in nonpancreatic tissue, although the sequence containing

---

From the School of Medical Sciences, University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen, U.K.

Address correspondence and reprint requests to Dr. K. Docherty, School of Medical Sciences, University of Aberdeen, Institute of Medical Sciences, Aberdeen, AB25 2ZD, U.K. E-mail: k.docherty@abdn.ac.uk.

Received for publication 8 June 2006 and accepted in revised form 24 August 2006.

ChIP, chromatin immunoprecipitation; COUP-TFII, chicken ovalbumin upstream promoter–transcription factor II; CRE, cyclic AMP response element; ECR, evolutionary conserved region; HNF, hepatocyte nuclear factor; ILPR, insulin-linked polymorphic region; PDX-1, pancreatic duodenum homeobox-1.

DOI: 10.2337/db06-0788

© 2006 by the American Diabetes Association.

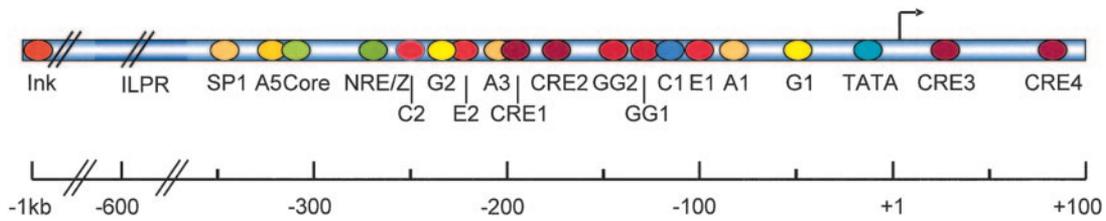


FIG. 1. Schematic representation of the human insulin promoter. The positions of the major *cis*-acting elements are indicated. The scale represents nucleotides relative to the transcription start site (+1 and forward arrow).

variable numbers of tandem repeats (see later) has been implicated in thymus expression of insulin.

In the  $\beta$ -cell, sophisticated mechanisms have evolved to control insulin expression at the correct time and place during embryonic development. In the adult related mechanisms and a variety of signaling pathways are involved in restricting insulin expression to  $\beta$ -cells (notwithstanding the low level extrapancreatic expression about which little is known) and in coordinating insulin expression in response to diverse afferent signals (16). Positive and negative crosstalk between the various signaling pathways, formation of homo- or heterodimers permitting individual transcription factors to act as activators, nonactivators or repressors, reversible phosphorylation of transcription factors, multiple isoforms of several transcription factors, and synergistic interactions between certain combinations of transcription factors extend the gamut of signals influencing the regulation of insulin gene expression.

Insulin transcriptional control is conferred by *cis*-acting regulatory sequences believed to be located within 300–400 bp from the transcription start site (28), which bind  $\beta$ -cell restricted and ubiquitous transcription factors (16). The principal regulatory elements within the human insulin promoter are outlined in Fig. 1. The compact nature of the insulin promoter results in the close proximity of regulatory elements that can bind an extensive range of factors thereby permitting a multiplicity of outcomes through additive and synergistic interactions between the bound proteins (29–31). In addition, regulatory elements can overlap in certain species e.g., the A3 and a cAMP response element (CRE) site in humans, introducing another layer of complexity through binding competition between alternative transcription factors.

**MULTIPLE-SPECIES COMPARISON OF INSULIN PROMOTERS**

There is no general approach to interpreting and predicting transcriptional evolution; however, the insulin promoter is one of the most extensively studied, and knowledge of the signals that bear upon insulin transcriptional regulation facilitates our understanding of possible functional consequences of insulin promoter evolutionary differences. By classic convention, the sequences that regulate basal promoters were divided into two classes. These are upstream regulatory elements (UREs) that are often located within 100–200 bp upstream of the site of initiation and display directional qualities, and enhancers that can function over distances of many kilobase pairs, regardless of orientation or whether they lie upstream or downstream of the start site. However, as more promoters and enhancers have been identified and studied, it has become apparent that there is a continuum between these two classes of regulatory elements with promoter and enhancer motifs sharing many physical and functional

traits. Therefore, in keeping with current opinion, we have reviewed the *cis*-regulatory elements within the compact insulin promoter without further categorization.

This review has drawn upon publicly available DNA sequences to compare the human insulin promoter sequence (–1,500 to +100) to the insulin promoters in an evolutionary and taxonomically divergent range of species. Definitive identification of insulin genes and their promoters lags well behind the isolation of the corresponding cDNA sequences; hence, care has been taken to ensure that only unambiguous insulin promoters have been included. These belong to human (*Homo sapiens*), great apes (chimpanzee [*Pan troglodytes*], orangutan [*Pongo pygmaeus*], and gorilla [*Gorilla gorilla*]), Old World monkeys (African green monkey [*Cercopithecus aethiops*] and rhesus macaque [*Macaca mulatta*]), New World monkey (owl monkey [*Aotus trivirgatus*]), rodents (rat [*Rattus norvegicus*] and mouse [*Mus musculus*]), mammals with diverse diets (carnivorous dog [*Canis familiaris*], herbivorous cow [*Bos taurus*], and omnivorous pig [*Sus scrofa*]), bird (chicken [*Gallus gallus*]), and fish (zebrafish [*Danio rerio*]). The promoter sequences of gorilla, orangutan, African green monkey, and owl monkey are currently incomplete extending upstream to positions –295, –290, –426, and –510, respectively. The phylogenetic relationships based on molecular analyses (32,33) between these species are outlined in Fig. 2.

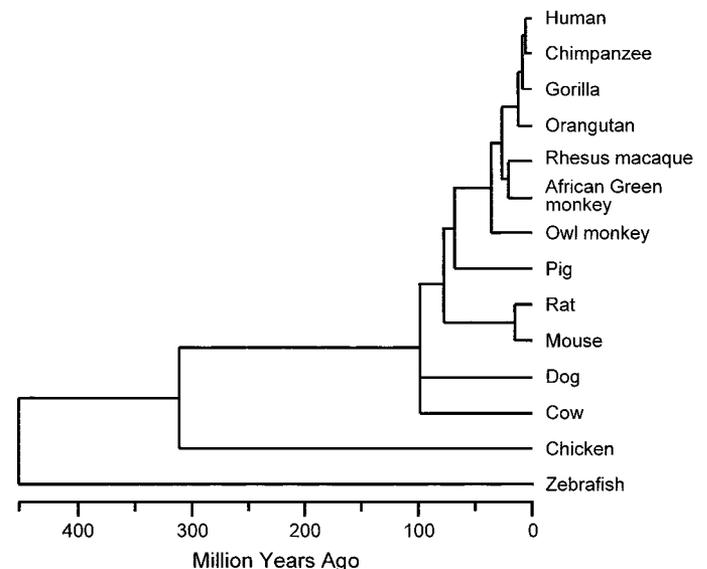


FIG. 2. Phylogenetic relationship between studied species. Evolutionary lineages of the studied species based on molecular analyses with branch lengths corresponding to time since divergence.

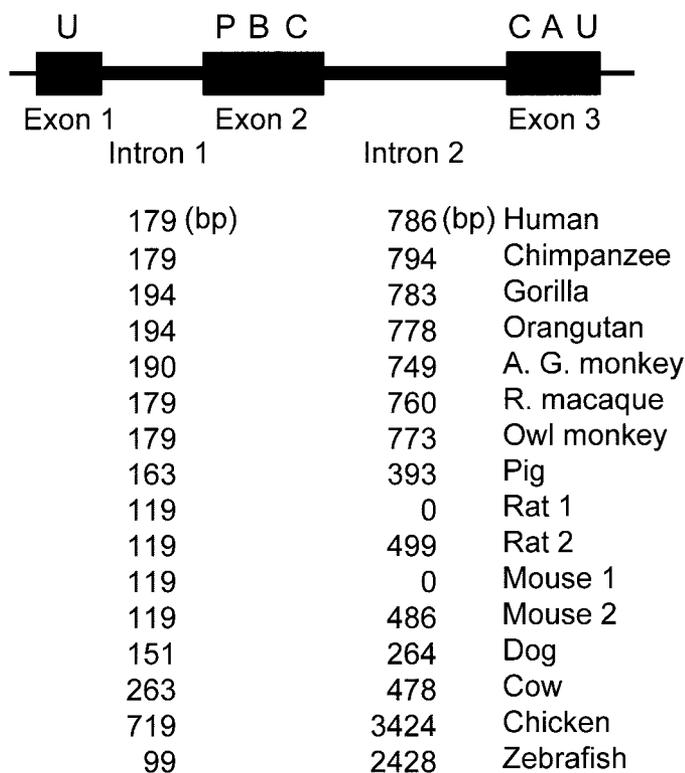


FIG. 3. Insulin gene intron sizes in studied species. The intron sizes (bp) are listed for all the reviewed insulin genes. The corresponding regions of the insulin protein are displayed above the exons. A, peptide A; B, peptide B; C, peptide C; P, preproinsulin leader peptide; U, untranslated. Figure is modified from ref. 34.

### INSULIN GENES

A preliminary evaluation of the relatedness of homologues can be generated from the number and relative position of introns, and these are shown in Fig. 3 (34). There are minor variations in the sizes of the introns among mammals while large dissimilarities are witnessed in the introns of chicken and zebrafish. The insulin 1 genes of rat and mouse have lost the second intron and also contain the remnant of a polydeoxyadenylate acid tract preceding the downstream direct repeat. Together, these structural features have led to the suggestion that the insulin 1 gene is a functional transposon (14) that was generated by an RNA-mediated duplication-transposition event involving a transcript of insulin 2 gene that was initiated upstream from the normal capping site. This duplication-transposition event clearly preceded separation of rat and mouse 15 million years ago. Along this evolutionary road, additional divergence has taken place resulting in rat having the two insulin genes residing about 55 Mbp apart on chromosome 1, whereas in the mouse they lie on different chromosomes, namely 6 and 7.

Synteny (i.e., the preserved order of genes between organisms) provides an expedient higher-level assessment of the association between homologues. The identification and annotation of genes in most genomes remains fragmentary; however, it is clear from currently available data that all of the studied insulin genes display remarkable synteny extending all the way back to zebrafish, which diverged from humans 450 million years ago. Not only are the immediate upstream and downstream flanking genes of tyrosine hydroxylase (TH) and insulin-like growth factor 2 (IGF-2) retained, but inspection of ~500 Mbp con-

firms extensive maintenance of synteny of many important genes including *syt8*, *lsp1*, *tnnt3*, *mrpl23*, *cd81*, and *tssc4*. While gene order and direction of transcription are preserved, the spacing between specific genes can vary. This is most dramatically illustrated with the insulin and TH genes, which are separated by 2–22 kbp in all species except mouse and rat, where the insulin 2 gene lies ~210 and 230 kbp distant from the TH gene, respectively. Despite evidence of different rates of insertion and deletion mutation within the insulin gene region, maintenance of synteny across vast evolutionary timescales points to a common and vital function for the insulin gene product, which is wholly consistent with the high degree of insulin protein conservation.

### HOMOLOGY BETWEEN INSULIN PROMOTERS

It has been estimated from large-scale studies that the number of conserved intergenic sequences is similar to that of coding sequences (35–37), and evolutionary changes in promoters together with their attendant alteration in transcriptional response to physiological and environmental demands have been documented (38,39). This is facilitated by the fact that promoters of protein-encoding genes are laid out into functional modules (40), allowing independent evolutionary selection of distinct characteristics of the overall transcription profile. Promoters are also considered to be more prone to genetic change than coding sequences (41,42) as the constraints typical of coding sequences are absent. In light of different regions of vertebrate genomes diverging at dissimilar rates (43) and this heterotachy being witnessed across different classes of mutation and lineages (44), this study utilized a variety of comparative alignment and transcription factor binding site search techniques with parameters that were appropriate for the evolutionary distances between species in order to detect meaningful evolutionary conserved regions (ECRs). The computational tools included CLUSTAL W (45), T-Coffee (46), GraphAlign (47), ECR Browser (48), Mulan (49), zPicture (50), TRES (51), and TRANSFAC (52).

Calculations of homology between the different insulin promoters and the human version were carried out across the region spanning –600 to +1. The downstream 100 bp, which contains two cyclic AMP response element (CRE) sites in human (see the section on CREs below), is comprised mostly of the extremely poorly conserved first intron that unduly influences the overall results. Percentage identity plots (PIPs) comparing the human insulin promoter to those of the other species reveal that, not surprisingly, the most closely related chimpanzee and other great apes share the greatest homology to human, making discernment of conserved regions impossible. Mammals that are more distantly diverged from human display several regions of conservation within the first 350 bp upstream, which correspond to the major regulatory elements. There is a clear fall off in homology beyond –350 or –400 bp upstream from the start of transcription, which is especially apparent in rhesus macaque. While PIPs are useful for identifying ECRs, a detailed breakdown of identity values for specific regions can expose the overall relatedness of different insulin promoters (Table 1). Interestingly, the degree of homology does not follow a simple direct correlation with time from divergence. For example, African green monkey and owl monkey diverged from humans 25 and 35 million years ago, but the main regulatory region of their promoters (–300 to +1) display

TABLE 1  
Homology between the human and studied insulin promoters

Species	Region of human insulin promoter (bp)						
	-600 to -550	-550 to -500	-500 to -450	-450 to -400	-400 to -350	-350 to -300	-300 to + 1
Chimpanzee	85	90	92	86	94	100	98
Gorilla	SNA	SNA	SNA	SNA	SNA	SNA	98
Orangutan	SNA	SNA	SNA	SNA	SNA	SNA	97
African green monkey	SNA	SNA	SNA	SNA	72	94	90
Rhesus macaque	38	42	39	46	75	94	94
Owl monkey	SNA	SNA	45	44	69	96	98
Pig	46	36	46	60	45	50	65
Rat 1	35	44	48	43	42	49	66
Rat 2	29	37	43	45	44	55	64
Mouse 1	38	49	37	43	46	51	65
Mouse 2	39	44	49	42	42	50	68
Dog	54	35	53	39	59	75	69
Cow	48	52	49	57	47	57	67
Chicken	24	41	34	32	32	30	37
Zebrafish	36	29	22	29	30	44	41

Percent identity values were calculated for the principal insulin promoter (-300 to + 1) and 50-bp windows further upstream to -600 bp from the start of transcription. SNA, sequence not available.

90 and 98% identity, respectively. Similarly, most nonprimate mammals have 65–69% identity in this region and 49–55% in the adjacent upstream 50 bp. Dog stands out in having much higher homology with 69 and 75% identity for these two regions, respectively.

Together, these results are in agreement with the opinion that vertebrate genes and immediate upstream flanks are highly constrained and, more important, confirm the accepted demarcation of the insulin promoter. There is no discernable significant homology between human and either chicken or zebrafish insulin promoters, which is in keeping with the view that most human DNA is not alignable to species separated by more than 200 million years. Likewise, there is no homology between chicken and zebrafish insulin promoters.

Computational analysis of the insulin promoters for novel evolutionary conserved sequences uncovered a single short region immediately upstream of the A3 box (see **A BOXES**); however, this region does not appear to contain any currently known transcription factor consensus sequences.

#### THE PROXIMAL PROMOTER REGION

Within a promoter, the fundamental component is the ~100-bp basal promoter that provides an assembly platform for the RNA polymerase II initiation complex. These modules vary among genes and can contain a TATA box 25–30 bp upstream of the transcription start site, an initiator element lacking the TATA sequence or a null basal promoter containing neither. All of the studied insulin promoters contain a TATA box. However, the chicken promoter is distinct from the others in that at least two isoforms can be transcribed from alternative initiation sites (53). In E1.5 chicken embryo pancreas, the single insulin gene is also transcribed from an upstream secondary promoter to yield an mRNA with an additional 32-bp leader sequence. Inspection of available chicken genome sequence reveals that this alternative start site must be the product of a secondary basal transcription complex, as the transcript includes the genomic sequence from immediately upstream of the TATA box (25 bp upstream from the

start of transcription) to the beginning of exon 1. The lack of another TATA box within the promoter and the presence of a C at -1 and an A at +1 of the longer transcript suggest that transcription is most likely established by an initiator element.

#### REGULATORY ELEMENTS WITHIN INSULIN PROMOTERS

Regulatory elements within promoters can originate at different times, and species comparisons indicate that promoters evolve through transcription factor binding site turnover and accretion (54,55). The relative numbers of the principal insulin promoter regulatory elements in the surveyed species are listed in Table 2.

**A boxes.** A-box sequences containing the TAAT motif bind homeodomain proteins (56), the most important of which is pancreatic duodenum homeobox-1 (PDX-1) (57–61), which has been shown to be a potent stimulator of transcription of rat, mouse, and human insulin genes (62). There are three principal A boxes in the human promoter: A1 (-82), A3 (-216), and A5 (-319) (Fig. 1). PDX-1 stimulates expression at A3 (58,63–65) and mutation of A3 has the most significant effect on transcription (61,65,66). Contrary to the opinion that A3 is not the most conserved (16), this survey has shown that A3 is the only A box present in all the mammals and, therefore, must be considered to be the most conserved and central to PDX-1 stimulation. PDX-1 bound to A1 has been shown to interact synergistically with E47/β2 in rat insulin 1 (30).

As the 4-bp TAAT motif can occur every 256 bp, the ability of PDX-1 to differentiate between potential regulatory elements must be influenced by adjacent sequences. The 3-bp flanking sequences have been shown to make an important contribution to the binding affinity of PDX-1 to TAAT core elements with a concomitant effect on activation. However, variations in these sequences are insufficient to completely explain differences in PDX-1 binding affinities (67). Therefore, the 8-bp flanking regions of all A boxes were assayed for homology (Table 3). The A3 box and 5' flanking region lie within a novel ECR, and this is reflected in the high degree of conservation. The lack of any other regulatory elements within this ECR based on

TABLE 2  
Major regulatory elements in insulin promoters

Species	Divergence (MYA)	Regulatory elements							
		A	GG	CRE	E	C1	SP1	G1	G2
Human	0	3	2	4	2:4	1	1	1	1
Chimpanzee	5	3	2	4	2:4	1	1	1	1
Gorilla	6.4	3*	2	4	2:3	1	SNA	1‡	1
Orangutan	11.3	3*	2	4	2:3	1	SNA	1	1§
Rhesus macaque	23	3	2	3	2:3	1	1	1‡	1§
African green monkey	25	3	2	4	1:2	1	0	1‡	1§
Owl monkey	35	3	1	2	1:4	1	1	1	1
Pig	65	1	1	1	1:3	1	0	0	1¶
Rat 1	75	2	1	1	2:4	1	0	1‡	0
Rat 2	75	2	0	1	1:3	1	0	1‡	0
Mouse 1	75	2	0	1	2:5	1	0	1‡	1¶
Mouse 2	75	2	0	1	1:4	1	0	1‡	1
Dog	92	2	2	1	1:4	1	0	0	1¶
Cow	94	1	1	1	1:2	1	0	0	1
Chicken	310	2	1	1	0:3	0	0	0	0
Zebrafish	450	3	0	1	0:2	0	0	0	0

The numbers of the principal regulatory elements present in insulin promoters are listed. Given the complexity and uncertainty of predicting the transcriptional outcome of mutations within regulatory elements, certain assumptions have had to be made and are described below. Since CRE sites can display high variability and remain active, those containing single base pair differences from human have been included. The chicken and zebrafish insulin promoters share so little homology with mammals that the values for A boxes simply reflect the presence of TAAT motifs. Similarly, it is unclear whether any of their putative nonconsensus CRE sequences are functional. E box values are given as two numbers, with the first referring to the generally recognized E1 and E2 boxes and the second number representing all CANNTG sequences up to position -400 or the 5' limit of available promoter sequence. \*The genomic sequence for the A5 box is unavailable; however, it has been included in the table because the immediate downstream region shares 98% identity with the human and all other primates contain an A5 box; †lacks the consensus motif but contains the adjacent guanine-rich region (GAGA box), which is known to also function; ‡1 bp substitution at position 2; §conservative transition at position 1 in the consensus motif; ¶the G2 site lies in a different location in comparison to human; ||no certainty that the regulatory element functions. "Divergence" refers to time since divergence from human; MYA, millions of years ago; SNA, sequence not available.

computational analysis raises the intriguing possibility that, while the TAAT motif is symmetrical, binding of PDX-1 to the promoter may be directional. Clear, though less well defined, asymmetrical homology of the other A box flanking regions to the human sequences is also apparent. Regulatory elements present in multiple copies often exist in both orientations (42), thereby increasing potential phenoplasty.

The A3 5' flanking region in rat insulin 1 has two additional TAAT sequences as a consequence of two single base pair changes. This creates the A4 site (29), which is

juxtaposed to A3 to generate an additional regulatory element that has been reported to bind other homeodomain transcription factors, some of which have been shown to affect transcription. One of the best studied is hepatocyte nuclear factor (HNF)-1 $\alpha$ , which has been reported to activate the rat insulin 1 gene in the HIT cell line (68). Similarly, Isl-1 has been found to bind to this site (69) and to interact with islet cell-specific transcription factor  $\beta$ 2 to stimulate rat 1 insulin expression (70). Other transcription factors reported to bind to the A3/A4 box include cdx-3 (29) and HMGI(Y) (71). Inspection of all other

TABLE 3  
Homology between the A box flanking regions in human and mammals.

Species	A5 box		A3 box		A1 box	
	5'	3'	5'	3'	5'	3'
Chimpanzee	100	100	100	100	100	100
Gorilla	SNA	SNA	100	100	100	100
Orangutan	SNA	SNA	100	100	100	100
Rhesus macaque	100	100	100	75	100	100
African green monkey	87	100	100	100	100	75
Owl monkey	100	100	100	62	75	100
Pig	—	—	100	50	—	—
Rat 1	—	—	75*	50*	50	75
Rat 2	—	—	100	50	37	62
Mouse 1	—	—	87	50	50	75
Mouse 2	—	—	100	50	50	62
Dog	62	87	100	62	—	—
Cow	—	—	100	50	—	—

The 5' and 3' 8-bp flanking regions of the insulin promoter A boxes (shown in their natural order along the promoter) were compared with human A boxes. Values are expressed as percent identity. \*The flanking regions of the two overlapping TAAT motifs at the adjacent A4 site have 87% identity with the human A3 flanking regions. [—], A box not present; SNA, sequence not available.

TABLE 4  
CRE sites within mammalian insulin promoters

Species	CRE1	CRE2	CRE3	CRE4
Consensus		tgacgtca		
Human	<u>tgacc<u>ccgc</u></u>	tgacg <u>acc</u>	<u>tgcatcag</u>	<u>tgcgtcag</u>
Chimpanzee	<u>tgacc<u>ccgc</u></u>	tgacg <u>acc</u>	<u>tgcatcag</u>	<u>tgcgtcag</u>
Gorilla	<u>tgacc<u>ccgc</u></u>	tgacg <u>acc</u>	<u>tgcatcag</u>	<u>tgcgtcag</u>
Orangutan	<u>tgacc<u>ccgc</u></u>	tga <u>Tgacc</u>	<u>tgcatcag</u>	<u>tgcgtcag</u>
Rhesus macaque	tgac <u>GTgc</u>	tgacg <u>acc</u>	<u>tgcatcag</u>	<u>cgcgtcag</u>
African green monkey	<u>tgacc<u>ccgc</u></u>	tgacg <u>acc</u>	<u>tgcatcag</u>	<u>cgcgtcag</u>
Owl monkey		tgacg <u>acc</u>	<u>tgatcaC</u>	
Pig		tga <u>Tgacc</u>		
Rat 1		tgacg <u>Tcc</u>		
Rat 2		tg <u>TTgTcc</u>		
Mouse 1		tgaTgTcc		
Mouse 2		tgacg <u>Tcc</u>		
Dog		tgacg <u>acc</u>		
Cow		tgacg <u>Gcc</u>		

The CRE site octamers present in the reviewed species are listed with deviations from the consensus CRE motif underlined and differences from the equivalent human CRE site shown in capital letters.

insulin promoters shows that this homeodomain-binding sequence is unique to rat insulin 1. It would, therefore, seem logical to conclude that these transcription factors play no role in other species. However, HNF-1 $\alpha$  provides an example of how the promiscuity of transcription factors creates obstacles in predicting insulin promoter effectors. Although the consensus binding sequence is not present in the human insulin promoter, the A3 region is sufficiently similar for the protein to bind, at least in vitro, and stimulate reporter assays (72). On the other hand, in vivo chromatin immunoprecipitation (ChIP) assays have shown that HNF-1 $\alpha$  is not necessary for either insulin 1 or 2 expression in mice, which lack A4 (73). Surprisingly, both the 5' and 3' flanking regions of each of the A4 TAAT sequences have higher homology to the human A3 region than rat insulin 1 A3, differing by only 1 bp. This evokes the interesting likelihood that, although the rat insulin 1 A3 box seems to be the main binding site (67), A4 could also bind and be regulated by PDX-1. Regardless of the regulatory capacity of the alternative A boxes, the binding kinetics of PDX-1 to the primary A3 regulatory element could be appreciably different in rat insulin 1 compared with humans and other mammals.

The greatly diverged chicken and zebrafish insulin promoters lack mammalian A boxes; however, several TAAT motifs are present. The chicken has two at -359 and -386, and zebrafish has three at -142, -347, and -359 plus two more further upstream at -473 and -510. The clustering of TAAT motifs is greater than would be expected from random nucleotide arrangements. While TAAT motifs are targets for a large number of homeodomain transcription factors, it is worthy to note that the 5' and 3' flanks of the zebrafish A boxes at -359 and -142 have 3-bp sequences associated with strong PDX-1 binding (67), suggesting a possible role for PDX-1 in regulating these insulin genes. The flanking regions share no homology with human. This is unlikely to reflect divergence of the PDX-1 proteins (rodent, chicken, and zebrafish PDX-1 proteins share 89, 26, and 49% amino acid sequence identity with the human protein, respectively) as the homeodomains are well conserved and there is no evidence of species specificity in DNA binding.

**GG boxes.** In addition to A boxes, the GGAAAT-containing GG2 motif (-145) is also activated by PDX-1 (74)

despite its deviation from the homeodomain consensus. The human insulin promoter contains a second GG motif 5 bp downstream of GG2 and commonly referred to as GG1 (75) or A2 (28). Mutation of these GG regulatory elements either singly or together has been shown to drastically reduce transcription (76), and the transcription factor binding to GG1 interacts with a transcription factor binding to the adjacent C1 site (77). Together, these findings suggest that both of the GG regulatory elements have a function in insulin expression. Of the two, GG2 is by far the more conserved being present in all mammals except the rodent insulin promoters. GG1, on the other hand, is absent from insulin promoters that diverged from human more than 25 million years ago, with the exception of the rat insulin 1 gene and dog. The presence of the highly conserved GG1 and C1 regulatory elements immediately downstream of GG2 and GG1, respectively, precludes useful comparison of flanking regions. The chicken insulin promoter has a GG motif at -130, which is in the same general region as GG1 in human (-133); however, there is no homology with the flanking regions of either human GG1 or GG2. The zebrafish insulin promoter does not contain any GG motifs.

**Cyclic AMP response element.** In the context of the insulin promoter, cAMP responsive elements bind the broadest array of transcription factors. These are generally closely related members of the bZIP CREB/ATF family, which can exist as multiple isoforms (78) that can interact with transcription factors activated by cAMP and diacylglycerol signaling pathways (79,80) to create activators, nonactivators, or repressors. The human insulin gene has four CRE sites: CRE1 at -210; CRE2 at -183; CRE3 at +18; and CRE4 at +61 (81). Although none of the CRE sites contains the consensus CRE sequence of TGACGTCA, mutagenesis experiments have shown that all are transcriptionally active (82).

Comparison of CRE sites between species (Table 4) reveals that only primates have multiple copies of CREs with other mammals containing a single CRE corresponding to CRE2. Of these, only the dog CRE is identical to the conserved human CRE2 site. The multiple CRE sites in primates could be due to several factors; the most likely being dietary. It should be noted that while gorillas are often considered to be predominantly folivorous, it has

become apparent that they also consume a significant amount of fruit (83). This is even truer of the Western gorilla (*Gorilla gorilla*), whose genome is being sequenced for assembly, than of Eastern gorillas (*Gorilla beringei*). Also, all the primates, especially the great apes, are partly omnivorous since they supplement their diets with birds, eggs, small reptiles, and insects. In comparison to the other mammals studied, only primates consume large quantities of fruit in their diet. However, the number of CRE sites is not in a simple direct correlation with the amount of fruit consumed, as all the studied primates eat large quantities. Another possible reason is that while primates are omnivorous to varying degrees, they often gorge themselves on a single food (e.g., ripe fruit when a tree is in season or meat when a whole carcass is consumed quickly), which would give rise to major alterations in metabolic demands. This would be particularly pertinent to early humans and necessitate an insulin promoter that could respond accordingly. The phenomenon of increased numbers of CREs in primates may be expedited by the fact that that primate promoters have an increased rate of evolution (44).

As with other regulatory elements, the chicken and zebrafish insulin promoters do not contain obvious CRE sites. The chicken insulin promoter contains four possible (three overlapping) nonconsensus sequences in the vicinity of the conserved mammalian CRE site, while the zebrafish has two potential nonconsensus octamers at -46 and -226.

It is impossible to draw conclusions on the effects of the numerous minor nucleotide changes on CRE site activity, as most regulatory elements can tolerate one or more substitutions without total loss of function (84,85). Therefore, it may be very significant that, even with the variability of the octamer in the conserved CRE site, sequences that include the CRE core along with at least 8 bp of both 5' and 3' flanking regions represent one of the most prominent ECRs in all mammalian insulin promoters. This strongly points to the importance of CRE sites in insulin gene regulation.

**C elements.** Initial expression studies on the C1 element at -128 (5'TGCAGCCTCAGCC) were carried out on the rat insulin 2 gene, showing that it binds the transcription factor RIPE3b1, which was subsequently identified as the basic leucine zipper (bZIP) protein MafA (86-90). Mutagenesis of the human C1 MafA binding site reduces promoter activity by 74% in INS-1  $\beta$ -cells (91) and blocks activation by glucose in MIN6  $\beta$ -cells (92). MafA can also interact with  $\beta$ 2 and PDX-1 (93). All the mammalian insulin genes show extremely high conservation of the C1 site, and all are identical to human, except dog and pig, which have 1- and 2-bp substitutions at the 3' and 5' regions of the consensus sequence, respectively. As the recognition site is 13 bp long, it is possible that mutations at its extremities would not necessarily eliminate MafA stimulation. Despite the clear conservation pressures on the C1 site, no comparable sequence was detected in the chicken and zebrafish insulin promoters.

The human insulin promoter has a bipartite C2 element (5'CAGGGACAGG) at -252 (94), and rat insulin 1 promoter has been reported to contain a dissimilar, though active, sequence between -329 and -307. The C2 site can bind PAX4 and PAX6, which repress (95) and stimulate (96), respectively. A search of insulin promoters showed that the human C2 site is present in all primates, although African green monkey has a single base pair substitution

between the two CAGG motifs. Among nonprimates, dog has two substitutions between the direct repeat and cow has three repeats with the intervening regions containing 1- and 2-bp deletions. It is not immediately apparent from DNA sequence alone whether these latter sites are functional.

**E boxes.** E boxes (5'CANNTG) bind proteins of the basic helix loop helix (bHLH) class of transcription factor with ubiquitous E47 forming a heterodimer with neuroendocrine cell specific NeuroD/ $\beta$ 2 (97). Two important E boxes were initially identified in the rat insulin 1 promoter between -104 and -112 (E1 or IEB1) and between -233 and -241 (E2 or IEB2) (98). The E1 box is the more conserved of the two and analysis showed that it is present in all mammal insulin promoters. Mutagenesis of this site in the rat insulin 1 and 2 promoters results in reduced transcription (98,99), and in the human insulin promoter drastically reduces basal transcription (91) and responsiveness to glucose (92). The E2 motif is less well conserved and the homologous sequence in the human insulin promoter at -239 (5'GCCACCGG) (75) contains a nonconsensus recognition site. The human E2 sequence can bind the ubiquitous transcription factor USF (100) but it does not appear to have a measurable effect on the overall activity of the promoter. In addition to the E1 and E2 boxes, a search of the insulin promoters revealed the presence of many other "CANNTG" consensus sequences (Table 2), including two in the negative regulatory element that lie just 23 and 33 bp upstream of the human E2 site. The presence of numerous potential E boxes suggests that regulation of the insulin promoter by bHLH transcription factors remains to be fully elucidated. Chicken and zebrafish insulin promoters contain neither E1 nor E2; however, they possess several consensus E box sites.

An unnamed sequence at -232 (5'GGGCC), which we have tentatively termed G2 in Fig. 1, overlaps the 5' end of the E2 box and binds a factor with limited tissue distribution (101). This sequence, which is known to induce DNA curvature, may serve to bring together proteins that bind at sites flanking this motif. Examination of the other insulin promoters reveals that within the primates, chimpanzee and gorilla contain the G2 sequence at the same location while orangutan, rhesus macaque, and African green monkey share a transition at the first nucleotide. The G2 site is absent from owl monkey; however, this primate has an alternative G2 motif at -453. Among the other mammalian insulin promoters, mouse insulin 2 and cow have a G2 site in the same region while dog, mouse insulin 1, and pig have alternative G2 sequences at -329, -400, and -16, respectively. Since a 6-bp motif would be expected to occur only once every 4,096 bp by random, the existence of alternative G2 motifs may indicate that G2-facilitated DNA bending abets interactions between proteins binding to the promoter. The G2 motif is absent from the rat insulin paralogues, chicken and zebrafish.

**Negative regulatory element.** The human insulin promoter contains an inhibitory sequence (-279 to -258) referred to as the negative regulatory element (NRE) (5'GAGACATTTGCCCCAGCTGT) (75,102) that lies within the glucose sensing Z element (-243 to -292) (103,104). It displays contrary properties acting as both a potent glucose-responsive transcriptional enhancer in primary cultured islet cells and as a transcriptional repressor in immortalized  $\beta$ - and non- $\beta$ -cells and in primary fibroblasts (103). Searches of the insulin promoters detected the NRE sequence in all primates; however, it is absent

from all other species, which is in agreement with reports that there is no evidence for a  $\beta$ -cell-specific NRE in rat insulin 1 (98,105).

**Insulin-linked polymorphic region.** A hypervariable region containing variable numbers of tandem 14-bp repeats (5'TCTGGGGAGAGGGG) (insulin-linked polymorphic region [ILRP] or variable number of tandem repeats) is located at approximately -360 in the human insulin promoter. The ILPR adopts an altered structure, which has been characterized as a quadriplex involving interactions between the G residues on the top strand (106). This sequence, which binds the transcription factor Pur-1/Maz (107) has a powerful effect on promoter activity in  $\beta$ -cells. Three classes of VNTR alleles have been identified based on the number of repeats of the 14-bp sequence: class I (20–63 repeats), class II (64–139 repeats), and class III (140–210 repeats). There is a correlation between the number of repeats in this region (IDDM2 locus) and susceptibility to type 1 diabetes with the highest risk conferred by class I (108), while class III has been linked to type 2 diabetes (109). On the other hand, studies involving large cohorts have shown that this region has no impact on early growth (110), insulin release, or diabetes (111). The class I allele is associated with higher levels of insulin mRNA in the pancreas, whereas class III alleles are associated with higher levels of insulin gene transcription in the thymus (20). The increased levels of insulin in the thymus may promote efficient deletion of autoreactive T-cells for proinsulin and immune tolerance to a key antigen implicated in type 1 diabetes. The ILPR sequence was found in only the chimpanzee promoter.

**G1 box.** The G1 box (5'GTAGGGGA) at -52 contains a sequence similar to that in the ILPR repeat sequence. The human insulin promoter G1 box binds the transcription factor Pur-1/MAZ (107,112). Although rat insulin 1 and 2 promoters lack the 5'GTAGGGGA motif, Pur-1/MAZ can bind to the adjacent guanine-rich region that often contains a GAGA box to stimulate transcription (113). A search of insulin promoters shows that chimpanzee, orangutan, and owl monkey have a G1 sequence identical to human. Gorilla and rhesus macaque share a single nucleotide change; however, like African green monkey, rat insulin 1, and both mouse paralogues, they retain the GAGA box. Therefore, it is likely that Pur-1/MAZ is active in regulation of these insulin promoters. Pig, cow, and dog all contain deletions in this region, and chicken and zebrafish lack homologous motifs.

**Enhancer core.** The core element (5'TGTGGAAAG) at -312 has a perfect match to the binding site for the CCAAT-enhancer binding protein (C/EBP) and probably other factors. There is very little known about this regulatory element, although it may act along with the adjacent A5 to mediate MafA-PDX-1 interactions (104). The enhancer core is present in all the primates for which sequence is available (not gorilla and orangutan). Rat insulin 2, mouse insulins 1 and 2, and dog share a single conservative transition at the most 3' position. Rat insulin 1 has an additional mutation at the most 5' position that may significantly reduce stimulatory potential, and the motif is absent from all other species.

**SP1 site.** The SP1 site (5'CCGCCC) at -345 was originally identified as a sequence that could bind a factor present in HIT T15  $\beta$ -cells. The SP1 site appeared to exhibit powerful transcriptional effects, but mutations that abolished protein binding had no effect on its transcriptional activity (A.R. Clark and K.D., unpublished findings), suggesting

possible interactions with adjacent sites. The SP1 site has also been identified as a potential binding site for the SP1-like factor KLF11, variants of which may contribute to the development of diabetes (114). Examination of primates for which sequence is available (not gorilla and orangutan) shows that all but African green monkey contain the identical SP1 site in the same position. African green monkey has a single nucleotide substitution of a C to T that reduces but does not eliminate KLF11 binding to oligonucleotide in electrophoretic mobility shift assay (EMSA) studies (114). It is absent from all other species.

**Ink box.** The Ink (for insulin kilobase upstream) sequence at -1,030 contains a cluster of potential binding sites comprising a palindromic element with zero spacing overlapping a direct-repeat element with 2 bp pairing (5'AGGTCCCCAGGTCATGCCCTC) and is responsive to both retinoic acid and thyroid hormone (115). Searches of available insulin promoters sequences upstream to -1,500 shows that the Ink box is absent from all nonprimates. Of the primates, distant upstream sequence is available for only chimpanzee and rhesus macaque. Both of these monkeys contain the Ink motif at -854 and -947, respectively. Although the positions are quite removed from the human, the immediate 30-bp regions display 95% identity with the human Ink region, suggesting that this regulatory element may be influential in insulin expression, perhaps playing a role in energy homeostasis.

#### SEQUENCE ELEMENTS THAT ARE ABSENT FROM THE HUMAN INSULIN PROMOTER

**CCAAT box.** Several of the descriptions of the effects of transcription factors on insulin expression are based on results from single species. For example, there is a transcriptionally active CCAAT regulatory element that overlaps the single CRE site in the insulin promoters of both rat and mouse. Expression studies using rat insulin 1 promoter have shown that the combined CRE/CCAAT site shows preferential binding for the nuclear transcription factor-Y (NF-Y), which leads to reduced influence of CRE-associated signaling (116). A search of the other insulin promoters revealed not only that no nonrodent species have a CCAAT site that overlaps with CRE, but that CCAAT sites are totally absent from all of the insulin promoters except zebrafish, which has three at -164, -130, and -85. Therefore, NF-Y signaling, which has an absolute requirement for all five bases in the CCAAT consensus sequence (117), is unique to rodents within mammals and does not typically play a role in insulin regulation.

**HNF-4 $\alpha$  regulatory element.** Rat and mouse insulin 1 and 2 promoters contain a consensus binding site for HNF-4 $\alpha$  (5'ACGGCAAAGTCC) located between -69 and -57. The rat insulin 1 promoter has been shown to be activated directly by HNF-4 $\alpha$ , which can interact synergistically with PDX-1 at the adjacent A1 site (118). In contrast, the HNF-4 $\alpha$  binding site does not exist in the human insulin promoter, and HNF-4 $\alpha$  fails to activate the gene in reporter assays (72). A search of all other insulin promoters found no evidence of any HNF-4 $\alpha$  binding sites. Therefore, HNF-4 $\alpha$  transactivation is unique to rodents and does not generally have a function in insulin regulation.

**STAT regulatory element.** Hormones involved in energy homeostasis and growth (e.g., leptin, prolactin, and growth hormone) have been reported to modify rat insulin

1 expression at  $-330$  to  $-322$  (5'TTCTGGGAA) through the transcription factors STAT3 (119) and STAT5 (120). Examination of all insulin promoters revealed that the STAT regulatory element is present only in the rat insulin 1 promoter, although the other rodent insulin promoters have only a single base pair substitution within the consensus sequence. The differences in the human insulin promoter were much greater, raising uncertainty about the relevance of direct influence by the STAT signaling pathways on insulin expression in humans.

**COUP-TFII binding element.** Chicken ovalbumin upstream promoter-transcription factor II (COUP-TFII), which is also known as NR2F and ARP-1, binds a direct repeat in the chicken ovalbumin promoter (121) and has been reported to bind an unrelated imperfect repeat in rat insulin 2 promoter between  $-55$  and  $-38$  (5'GGGT CAGGGGGGGGGTGC) (122) through a different molecular mechanism (123). COUP-TFII has recently been implicated in the control of blood glucose in heterozygous knockout transgenic mice that had increased insulin secretion in low glucose and decreased insulin secretion in high glucose (124). The corresponding regions in human and all primates (5'AGGTAGGGGAGATGGGCT) have several nucleotide differences that result in loss of essential guanine nucleotides. In rat insulin 1 and both mouse insulin promoters, there are transitions of essential guanines to adenines in positions where either purine can serve for recognition, if not intermolecular association. Given the irregular binding affinities of COUP-TFII, it is difficult to make unequivocal statements regarding its possible effects on these insulin promoters. The other mammals (cow, dog, and pig) have deletions in the COUP-TFII binding region, and no known form of consensus sequences are to be found in chicken and zebrafish insulin promoters. Within the context of this survey, the action of this COUP-TFII would seem to be limited to rodents.

#### REGULATORY ELEMENT SPACING

The spacing between the individual regulatory elements within the particularly well-conserved cassette of C1, E1, and A1 boxes has been shown to alter the relative stimulatory effects of the transcription factors that bind along with their synergistic interactions (91). Comparison of mammalian insulin promoters in this region showed that the relative spacing of the regulatory elements has been maintained for at least 35 million years, as there is no deviation in the primates. On the other hand, all the rodent insulin promoters contained insertions and deletions between all three sites. In mammals lacking A1, the C1-E1 spacing was maintained in both pig and dog while cow had a one base pair insertion between C1 and E1.

#### EFFECTS OF CHROMATIN STRUCTURE

Efficient transcription is the outcome of coordinated dynamic arrangements upon the promoter. ChIP assays using MIN6  $\beta$ -cells have shown that PDX-1, MafA, E47, and  $\beta$ 2 bind to the mouse insulin 2 promoter in a cyclical manner with a periodicity of  $\sim 10$ – $15$  min (125). Insulin gene regulation is also influenced by epigenetic factors that include DNA methylation and alterations in histone modifications, which affect the packaging of DNA within chromatin. There are a number of studies on the role of histone acetylation and methylation in the control of insulin gene expression. A key role for histone acetyltransferase (HAT) p300 in insulin promoter regulation has

been demonstrated by the observations that PDX-1 and  $\beta$ 2 mediate their effects on the rat insulin 2 gene through an interaction with p300 (31,126,127), while activation of a rat insulin 1 promoter construct in HeLa cells by PDX-1 requires interactions with p300 (128). It has also been shown that the effects of glucose on a rat insulin 1 promoter construct in the mouse MIN6  $\beta$ -cell line involved the recruitment by PDX-1 of HAT and histone deacetylase activities (HDAC) activities. Thus, under low-glucose conditions, PDX-1 associated with HDACs to repress transcription (129), whereas under high glucose conditions PDX-1 recruited the HAT p300 to activate transcription (130). PDX-1 has also been linked to the presence of methylated histone H3, i.e., H3K4me (nomenclature as per (131)), at the proximal promoter and coding regions of the insulin gene in rodent cells (132). More recently, the histone methyl transferase set9 has been localized to  $\beta$ -cells in association with the insulin gene (133).

Investigations into the role of chromatin accessibility in insulin expression have revealed that PDX-1 shows preferential binding to open chromatin (euchromatin) over condensed chromatin (heterochromatin). In particular, PDX-1 occupies the endogenous insulin promoter in mouse  $\beta$ TC3  $\beta$ -cells but not in mPAC ductal cells, which do not express insulin. Furthermore, the binding affinity of PDX-1 is strongly influenced by the position of nucleosomes relative to its regulatory element (134). Even within euchromatin, the degree of openness varies as the A3/A4 region ( $-126$  to  $-296$ ) to which PDX-1 can bind contained the most open chromatin structure based on micrococcal nuclease digestion, whereas the adjacent region ( $-297$  to  $-460$ ), which is not as crucial for  $\beta$ -cell-specific insulin transcription, was more condensed. Although it is likely that the insulin gene is embedded in euchromatin in  $\beta$ -cells and in more condensed heterochromatin in non- $\beta$ -cells, it may be of relevance that the synteny studies (see INSULIN GENES) show that the human insulin gene lies only 2 kbp from the transcriptionally active TH gene, whereas this distance is  $>100$ -fold greater in rodents. Thus, the diverse efforts to induce insulin expression in non- $\beta$ -cells may be less problematic in humans than in rodents.

#### CONCLUSIONS

The extraordinary synteny of insulin genes from zebrafish to human substantiates the key importance of the insulin hormone product. Comparison of insulin promoters spanning 450 million years of evolution has permitted identification of the central regulatory elements as well as several valuable observations.

1. The transcription factor PDX-1 emerges as one of the fundamental regulators of insulin expression for several reasons: all promoters have at least one A box with A3 being the most conserved, the weaker PDX-1-binding GG boxes also form ECRs with GG2 being the more conserved, and the transcription factor is known to interact with MafA and E47/ $\beta$ 2.
2. The other strongly conserved regulatory elements of C1, E1, and the conserved CRE site attest to the importance of MafA, E47/ $\beta$ 2, and cAMP-associated regulation. Of these regulatory elements, the CRE site is unusual due to the remarkable degree of variability and extensive array of associating transcription factors.
3. Regulatory element conservation is not limited to the consensus sequences as the flanking regions also con-

tribute to ECRs. This is true of regulatory elements with both short and long recognition sequences indicating that flanking regions are necessary for transcription factor specificity and binding. This may be of particular consequence for the capricious CRE sites while the asymmetrical nature of the conserved A3 box flanking regions may reflect directional binding of PDX-1.

4. Within mammals, dog stands out due to its much greater homology to humans. The similarities include higher percentage identity, possessing more PDX-1-binding A boxes and GG elements, and having a CRE site that is identical to human. It is interesting to speculate whether these likenesses between a carnivore and omnivore correlate to the increased contribution of meat in the human diet over evolutionary time compared with other primates.
5. The chicken and zebrafish insulin promoters bear no obvious homology with mammals and exhibit a dearth of readily discernible regulatory elements.

## IMPLICATIONS

Investigations based on rodents and their insulin genes have provided invaluable insights into diabetes and the workings of insulin promoters. However, the findings reported here illustrate that notable dissimilarities exist between the human and rodent promoters, which may reflect both divergence and the degree to which these promoters have been studied. The atypical characteristics of rodent insulin promoters are exemplified most manifestly with the rat insulin 1 promoter, whose unusual attributes include an active dominant CCAAT site overlapping the single CRE site, HNF-1 $\alpha$ , and HNF-4 $\alpha$  regulatory elements; a functional Isl-1 binding site at A3/A4; a STAT-3 binding site; a potential COUP-TFII binding site; a consensus-containing E2 site; loss of GG boxes; lower conservation of A3 flanking regions; and changed spacing between regulatory elements in the C1-E1-A1 module leading to alternative synergistic interactions. The most plausible basis for the complexity of rodent insulin promoters is the duplication of their associated genes. Gene duplication can lead to functional divergence of the *cis*-regulatory elements (135,136) that can be swift even in recently duplicated genes (137). In addition, the signaling pathways regulating an essential gene like insulin will undoubtedly incorporate redundancy to extend responses and to act as a buffer against the consequences of mutation of key components. The fundamental differences in regulatory elements should serve as a salutary warning to be cautious when extrapolating rodent-based data to humans.

A major obstacle in diabetes research has been the lack of a human pancreatic  $\beta$ -cell line that is functionally equivalent to primary  $\beta$ -cells. It is essential that new human  $\beta$ -cell lines be developed and widely distributed in order that physiologically and medically relevant studies on the human insulin promoter can be carried out. This is especially true of *in vivo* epigenetic and ChIP-based experiments that will accurately map the position and define the role of nucleosomes and undoubtedly help to unravel the precise mechanisms responsible for insulin gene regulation.

These are exciting times as genome sequencing progresses rapidly. The availability of insulin genes from a wider range of species will provide tools that will permit the relatively straightforward answering of points raised in this report and allow us to advance our comprehension

and appreciation of the subtle and sophisticated insulin promoter.

## REFERENCES

1. Bell GI, Pictet RL, Rutter WJ, Cordell B, Tischer E, Goodman HM: Sequence of the human insulin gene. *Nature* 284:26–32, 1980
2. Wilson ME, Scheel D, German MS: Gene expression cascades in pancreatic development. *Mech Dev* 120:65–80, 2003
3. Frayling TM, Evans JC, Bulman MP, Pearson E, Allen L, Owen K, Bingham C, Hannemann M, Shepherd M, Ellard S, Hattersley AT: Beta-cell genes and diabetes: Molecular and clinical characterization of mutations in transcription factors. *Diabetes* 50 (Suppl. 1):S94–S100, 2001
4. Stoffel M, Vallier L, Pedersen RA: Navigating the pathway from embryonic stem cells to beta cells. *Semin Cell Dev Biol* 15:327–336, 2004
5. Ferber S, Halkin A, Cohen H, Ber I, Einav Y, Goldberg I, Barshack I, Seiffers R, Kopolovic J, Kaiser N, Karasik A: Pancreatic and duodenal homeobox gene 1 induces expression of insulin genes in liver and ameliorates streptozotocin-induced hyperglycemia. *Nat Med* 6:568–572, 2000
6. Horb ME, Shen CN, Tosh D, Slack JM: Experimental conversion of liver to pancreas. *Curr Biol* 13:105–115, 2003
7. Choi KS, Shin JS, Lee JJ, Kim YS, Kim SB, Kim CW: *In vitro* trans-differentiation of rat mesenchymal cells into insulin-producing cells by rat pancreatic extract. *Biochem Biophys Res Commun* 330:1299–1305, 2005
8. Edlund T, Walker MD, Barr PJ, Rutter WJ: Cell-specific expression of the rat insulin gene: evidence for role of two distinct 5' flanking elements. *Science* 230:912–916, 1985
9. Fromont-Racine M, Bucchini D, Desbois P, Pictet R, Jami J: Human insulin gene in transgenic mouse lines. *Biomed Biochim Acta* 47:349–353, 1988
10. Fromont-Racine M, Bucchini D, Madsen O, Desbois P, Linde S, Nielsen JH, Saulnier C, Ripoche MA, Jami J, Pictet R: Effect of 5'-flanking sequence deletions on expression of the human insulin gene in transgenic mice. *Mol Endocrinol* 4:669–677, 1990
11. Steiner DF, Chan SJ, Welsh JM, Kwok SC: Structure and evolution of the insulin gene. *Annu Rev Genet* 19:463–484, 1985
12. Harper ME, Ullrich A, Saunders GF: Localization of the human insulin gene to the distal end of the short arm of chromosome 11. *Proc Natl Acad Sci U S A* 78:4458–4460, 1981
13. Shuldiner AR, Phillips S, Roberts CT, Jr, LeRoith D, Roth J: *Xenopus laevis* contains two nonallelic preproinsulin genes: cDNA cloning and evolutionary perspective. *J Biol Chem* 264:9428–9432, 1989
14. Soares MB, Schon E, Henderson A, Karathanasis SK, Cate R, Zeitlin S, Chirgwin J, Efstratiadis A: RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. *Mol Cell Biol* 5:2090–2103, 1985
15. Davies PO, Poirier C, Deltour L, Montagutelli X: Genetic reassignment of the insulin-1 (Ins1) gene to distal mouse chromosome 19. *Genomics* 21:665–667, 1994
16. Melloul D, Marshak S, Cerasi E: Regulation of insulin gene transcription. *Diabetologia* 45:309–326, 2002
17. Rosenzweig JL, Havrankova J, Lesniak MA, Brownstein M, Roth J: Insulin is ubiquitous in extrapancreatic tissues of rats and humans. *Proc Natl Acad Sci U S A* 77:572–576, 1980
18. Kojima H, Fujimiya M, Matsumura K, Nakahara T, Hara M, Chan L: Extrapancreatic insulin-producing cells in multiple organs in diabetes. *Proc Natl Acad Sci U S A* 101:2458–2463, 2004
19. Devaskar SU, Giddings SJ, Rajakumar PA, Carnaghi LR, Menon RK, Zahm DS: Insulin gene expression and insulin synthesis in mammalian neuronal cells. *J Biol Chem* 269:8445–8454, 1994
20. Pugliese A, Zeller M, Fernandez A, Jr, Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD: The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet* 15:293–297, 1997
21. Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C: Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* 15:289–292, 1997
22. Smith KM, Olson DC, Hirose R, Hanahan D: Pancreatic gene expression in rare cells of thymic medulla: evidence for functional contribution to T cell tolerance. *Int Immunol* 9:1355–1365, 1997
23. Cunha DA, Carneiro EM, Alves Mde C, Jorge AG, de Sousa SM, Boschero AC, Saad MJ, Velloso LA, Rocha EM: Insulin secretion by rat lacrimal glands: Effects of systemic and local variables. *Am J Physiol Endocrinol Metab* 289:E768–E775, 2005

24. Vallejo G, Mead PM, Gaynor DH, Devlin JT, Robbins DC: Characterization of immunoreactive insulin in human saliva: evidence against production in situ. *Diabetologia* 27:437–440, 1984
25. Schwartz MW, Porte D Jr: Diabetes, obesity, and the brain. *Science* 307:375–379, 2005
26. Porte D, Jr, Baskin DG, Schwartz MW: Insulin signaling in the central nervous system: a critical role in metabolic homeostasis and disease from *C. elegans* to humans. *Diabetes* 54:1264–1276, 2005
27. Pugliese A: Insulin expression in the thymus, tolerance, and type 1 diabetes. *Diabetes Metab Rev* 14:325–327, 1998
28. German M, Ashcroft S, Docherty K, Edlund H, Edlund T, Goodison S, Imura H, Kennedy G, Madsen O, Melloul D: The insulin gene promoter: a simplified nomenclature. *Diabetes* 44:1002–1004, 1995
29. German MS, Wang J, Chadwick RB, Rutter WJ: Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. *Genes Dev* 6:2165–2176, 1992
30. Glick E, Leshkowitz D, Walker MD: Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter. *J Biol Chem* 275:2199–2204, 2000
31. Qiu Y, Guo M, Huang S, Stein R: Insulin gene transcription is mediated by interactions between the p300 coactivator and PDX-1, BETA2, and E47. *Mol Cell Biol* 22:412–420, 2002
32. Kumar S, Hedges SB: A molecular timescale for vertebrate evolution. *Nature* 392:917–920, 1998
33. Hedges SB, Kumar S: Genomic clocks and evolutionary time scales. *Trends Genet* 19:200–206, 2003
34. Docherty K, Steiner DF: Molecular and cell biology of the beta cell. In *Ellenberg and Rifkin's Diabetes Mellitus*. 5th ed. Porte DEJ, Sherwin RS, Eds. Englewood Cliffs, NJ, Prentice Hall, 1997, p. 29–48
35. Onyango P, Miller W, Lehoczy J, Leung CT, Birren B, Wheelan S, Dewar K, Feinberg AP: Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res* 10:1697–1710, 2000
36. Shabalina SA, Ogurtsov AY, Kondrashov VA, Kondrashov AS: Selective constraint in intergenic regions of human and mouse genomes. *Trends Genet* 17:373–376, 2001
37. Bergman CM, Kreitman M: Analysis of conserved noncoding DNA in drosophila reveals similar constraints in intergenic and intronic sequences. *Genome Res* 11:1335–1345, 2001
38. Brakefield PM, Gates J, Keys D, Kesbeke F, Wijngaarden PJ, Monteiro A, French V, Carroll SB: Development, plasticity and evolution of butterfly eyespot patterns. *Nature* 384:236–242, 1996
39. Abouheif E, Wray GA: Evolution of the gene network underlying wing polyphenism in ants. *Science* 297:249–252, 2002
40. Arnone MI, Davidson EH: The hardwiring of development: organization and function of genomic regulatory systems. *Development* 124:1851–1864, 1997
41. Stern DL: Evolutionary developmental biology and the problem of variation. *Evolution Int J Org Evolution* 54:1079–1091, 2000
42. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA: The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* 20:1377–1419, 2003
43. Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L: Evolution and functional classification of vertebrate gene deserts. *Genome Res* 15:137–145, 2005
44. Taylor MS, Kai C, Kawai J, Carninci P, Hayashizaki Y, Sempere CAM: Heterotachy in mammalian promoter evolution. *PLoS Genet* 2:e30, 2006. In press
45. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Res* 22:4673–4680, 1994
46. Notredame C, Higgins DG, Heringa J: T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217, 2000
47. Spalding JB, Lammers PJ: BLAST filter and GraphAlign: rule-based formation and analysis of sets of related DNA and protein sequences. *Nucleic Acid Res* 32:W26–W32, 2004
48. Ovcharenko I, Nobrega MA, Loots GG, Stubbs L: ECR browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acid Res* 32:W280–W286, 2004
49. Ovcharenko I, Loots GG, Giardine BM, Hou M, Ma J, Hardison RC, Stubbs L, Miller W: Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res* 15:184–194, 2005
50. Ovcharenko I, Loots GG, Hardison RC, Miller W, Stubbs L: zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res* 14:472–477, 2004
51. Katti MV, Sakharkar MK, Ranjekar PK, Gupta VS: TRES: comparative promoter sequence analysis. *Bioinformatics* 16:739–740, 2000
52. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E: TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acid Res* 31:374–378, 2003
53. Hernandez-Sanchez C, Rubio E, Serna J, de la Rosa EJ, de Pablo F: Unprocessed proinsulin promotes cell survival during neurulation in the chick embryo. *Diabetes* 51:770–777, 2002
54. Ludwig MZ, Kreitman M: Evolutionary dynamics of the enhancer region of even-skipped in drosophila. *Mol Biol Evol* 12:1002–1011, 1995
55. Rockman MV, Wray GA: Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol* 19:1991–2004, 2002
56. Rudnick A, Ling TY, Odagiri H, Rutter WJ, German MS: Pancreatic beta cells express a diverse set of homeobox genes. *Proc Natl Acad Sci U S A* 91:12203–12207, 1994
57. Boam DS, Docherty K: A tissue-specific nuclear factor binds to multiple sites in the human insulin-gene enhancer. *Biochem J* 264:233–239, 1989
58. Ohlsson H, Karlsson K, Edlund T: IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J* 12:4251–4259, 1993
59. Miller CP, McGehee RE, Jr, Habener JF: IDX-1: A new homeodomain transcription factor expressed in rat pancreatic islets and duodenum that transactivates the somatostatin gene. *EMBO J* 13:1145–1156, 1994
60. Leonard J, Peers B, Johnson T, Ferreri K, Lee S, Montminy MR: Characterization of somatostatin transactivating factor-1, a novel homeobox factor that stimulates somatostatin expression in pancreatic islet cells. *Mol Endocrinol* 7:1275–1283, 1993
61. Marshak S, Totary H, Cerasi E, Melloul D: Purification of the beta-cell glucose-sensitive factor that transactivates the insulin gene differentially in normal and transformed islet cells. *Proc Natl Acad Sci U S A* 93:15057–15062, 1996
62. McKinnon CM, Docherty K: Pancreatic duodenal homeobox-1, PDX-1, a major regulator of beta cell identity and function. *Diabetologia* 44:1203–1214, 2001
63. Peers B, Leonard J, Sharma S, Teitelman G, Montminy MR: 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, 1994
64. Peshavaria M, Gamer L, Henderson E, Teitelman G, Wright CV, Stein R: XIHbox 8, an endoderm-specific xenopus homeodomain protein, is closely related to a mammalian insulin gene transcription factor. *Mol Endocrinol* 8:806–816, 1994
65. Petersen HV, Serup P, Leonard J, Michelsen BK, Madsen OD: Transcriptional regulation of the human insulin gene is dependent on the homeodomain protein STF1/IPF1 acting through the CT boxes. *Proc Natl Acad Sci U S A* 91:10465–10469, 1994
66. German MS, Wang J: The insulin gene contains multiple transcriptional elements that respond to glucose. *Mol Cell Biol* 14:4067–4075, 1994
67. Liberzon A, Ridner G, Walker MD: Role of intrinsic DNA binding specificity in defining target genes of the mammalian transcription factor PDX1. *Nucleic Acid Res* 32:54–64, 2004
68. Emens LA, Landers DW, Moss LG: Hepatocyte nuclear factor 1 alpha is expressed in a hamster insulinoma line and transactivates the rat insulin I gene. *Proc Natl Acad Sci U S A* 89:7300–7304, 1992
69. Karlsson O, Thor S, Norberg T, Ohlsson H, Edlund T: Insulin gene enhancer binding protein isl-1 is a member of a novel class of proteins containing both a homeo- and a cys-his domain. *Nature* 344:879–882, 1990
70. Peng SY, Wang WP, Meng J, Li T, Zhang H, Li YM, Chen P, Ma KT, Zhou CY: ISL1 physically interacts with BETA2 to promote insulin gene transcriptional synergy in non-beta cells. *Biochim Biophys Acta* 1731: 154–159, 2005
71. Ohneda K, Mirmira RG, Wang J, Johnson JD, German MS: 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, 2000
72. Okita K, Yang Q, Yamagata K, Hangenfeldt KA, Miyagawa J, Kajimoto Y, Nakajima H, Namba M, Wollheim CB, Hanafusa T, Matsuzawa Y: Human insulin gene is a target gene of hepatocyte nuclear factor-1alpha (HNF-1alpha) and HNF-1beta. *Biochem Biophys Res Commun* 263:566–569, 1999
73. Parrizas M, Maestro MA, Boj SF, Paniagua A, Casamitjana R, Gomis R, Rivera F, Ferrer J: Hepatic nuclear factor 1-alpha directs nucleosomal hyperacetylation to its tissue-specific transcriptional targets. *Mol Cell Biol* 21:3234–3243, 2001
74. Le Lay J, Matsuoka TA, Henderson E, Stein R: Identification of a novel

- PDX-1 binding site in the human insulin gene enhancer. *J Biol Chem* 279:22228–22235, 2004
75. Boam DS, Clark AR, Docherty K: Positive and negative regulation of the human insulin gene by multiple trans-acting factors. *J Biol Chem* 265: 8285–8296, 1990
  76. Tomonari A, Yoshimoto K, Tanaka M, Iwahana H, Miyazaki J, Itakura M: GGAAAT motifs play a major role in transcriptional activity of the human insulin gene in a pancreatic islet beta-cell line MIN6. *Diabetologia* 39:1462–1468, 1996
  77. Tomonari A, Yoshimoto K, Mizusawa N, Iwahana H, Itakura M: Differential regulation of the human insulin gene transcription by GG1 and GG2 elements with GG- and C1-binding factors. *Biochim Biophys Acta* 1446:233–242, 1999
  78. Foulkes NS, Sassone-Corsi P: Transcription factors coupled to the cAMP-signalling pathway. *Biochim Biophys Acta* 1288:F101–21, 1996
  79. Nishizuka Y: Studies and perspectives of protein kinase C. *Science* 233:305–312, 1986
  80. Masquillier D, Sassone-Corsi P: Transcriptional cross-talk: nuclear factors CREM and CREB bind to AP-1 sites and inhibit activation by jun. *J Biol Chem* 267:22460–22466, 1992
  81. Inagaki N, Maekawa T, Sudo T, Ishii S, Seino Y, Imura H: c-Jun represses the human insulin promoter activity that depends on multiple cAMP response elements. *Proc Natl Acad Sci U S A* 89:1045–1049, 1992
  82. Hay CW, Sinclair EM, Bermano G, Durward E, Tadayyon M, Docherty K: Glucagon-like peptide-1 stimulates human insulin promoter activity in part through cAMP-responsive elements that lie upstream and downstream of the transcription start site. *J Endocrinol* 186:353–365, 2005
  83. Doran DM, McNeillage A, Greer D, Bocian C, Mehlman P, Shah N: Western lowland gorilla diet and resource availability: new evidence, cross-site comparisons, and reflections on indirect sampling methods. *Am J Primatol* 58:91–116, 2002
  84. Latchman DS: *Eukaryotic Transcription Factors*. San Diego, CA, Academic Press, 1998
  85. Courey AJ: Regulatory transcription factors and cis-regulatory regions. In *Transcription Factors*. Locker J Ed. San Diego, CA., Academic Press, 2001, p. 17–34
  86. Olbrot M, Rud J, Moss LG, Sharma A: Identification of beta-cell-specific insulin gene transcription factor RIPE3b1 as mammalian MafA. *Proc Natl Acad Sci U S A* 99:6737–6742, 2002
  87. Kataoka K, Han SI, Shioda S, Hirai M, Nishizawa M, Handa H: MafA is a glucose-regulated and pancreatic beta-cell-specific transcriptional activator for the insulin gene. *J Biol Chem* 277:49903–49910, 2002
  88. Matsuoka TA, Zhao L, Artner I, Jarrett HW, Friedman D, Means A, Stein R: Members of the large maf transcription family regulate insulin gene transcription in islet beta cells. *Mol Cell Biol* 23:6049–6062, 2003
  89. Kajihara M, Sone H, Amemiya M, Katoh Y, Isogai M, Shimano H, Yamada N, Takahashi S: Mouse MafA, homologue of zebrafish somite maf 1, contributes to the specific transcriptional activity through the insulin promoter. *Biochem Biophys Res Commun* 312:831–842, 2003
  90. Kataoka K, Shioda S, Ando K, Sakagami K, Handa H, Yasuda K: Differentially expressed maf family transcription factors, c-maf and MafA, activate glucagon and insulin gene expression in pancreatic islet alpha- and beta-cells. *J Mol Endocrinol* 32:9–20, 2004
  91. Docherty HM, Hay CW, Ferguson LA, Barrow J, Durward E, Docherty K: Relative contribution of PDX-1, MafA and E47/beta2 to the regulation of the human insulin promoter. *Biochem J* 389:813–820, 2005
  92. da Silva Xavier G, Rutter J, Rutter GA: Involvement of per-arnt-sim (PAS) kinase in the stimulation of preproinsulin and pancreatic duodenum homeobox 1 gene expression by glucose. *Proc Natl Acad Sci U S A* 101:8319–8324, 2004
  93. Zhao L, Guo M, Matsuoka TA, Hagman DK, Parazzoli SD, Poutout V, Stein R: The islet beta cell-enriched MafA activator is a key regulator of insulin gene transcription. *J Biol Chem* 280:11887–11894, 2005
  94. Read ML, Masson MR, Docherty K: A RIPE3b1-like factor binds to a novel site in the human insulin promoter in a redox-dependent manner. *FEBS Lett* 418:68–72, 1997
  95. Campbell SC, Cragg H, Elrick LJ, Macfarlane WM, Shennan KI, Docherty K: Inhibitory effect of pax4 on the human insulin and islet amyloid polypeptide (IAPP) promoters. *FEBS Lett* 463:53–57, 1999
  96. Sander M, Neubuser A, Kalamaras J, Ee HC, Martin GR, German MS: Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. *Genes Dev* 11:1662–1673, 1997
  97. Naya FJ, Stellrecht CM, Tsai MJ: Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. *Genes Dev* 9:1009–1019, 1995
  98. Karlsson O, Edlund T, Moss JB, Rutter WJ, Walker MD: 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, 1987
  99. Whelan J, Poon D, Weil PA, Stein R: Pancreatic beta-cell-type-specific expression of the rat insulin II gene is controlled by positive and negative cellular transcriptional elements. *Mol Cell Biol* 9:3253–3259, 1989
  100. Read ML, Clark AR, Docherty K: The helix-loop-helix transcription factor USF (upstream stimulating factor) binds to a regulatory sequence of the human insulin gene enhancer. *Biochem J* 295:233–237, 1993
  101. Reibel L, Besnard C, Lores P, Jami J, Gacon G: An insulinoma nuclear factor binding to GGGCCC motifs in human insulin gene. *Nucleic Acid Res* 21:1595–1600, 1993
  102. Clark AR, Wilson ME, Leibiger I, Scott V, Docherty K: A silencer and an adjacent positive element interact to modulate the activity of the human insulin promoter. *Eur J Biochem* 232:627–632, 1995
  103. Sander M, Griffen SC, Huang J, German MS: A novel glucose-responsive element in the human insulin gene functions uniquely in primary cultured islets. *Proc Natl Acad Sci U S A* 95:11572–11577, 1998
  104. Pino MF, Ye DZ, Linning KD, Green CD, Wicksteed B, Poitout V, Olson LK: Elevated glucose attenuates human insulin gene promoter activity in INS-1 pancreatic beta-cells via reduced nuclear factor binding to the A5/core and Z element. *Mol Endocrinol* 19:1343–1360, 2005
  105. Walker MD, Edlund T, Boulet AM, Rutter WJ: Cell-specific expression controlled by the 5'-flanking region of insulin and chymotrypsin genes. *Nature* 306:557–561, 1983
  106. Hammond-Kosack MC, Dobrinski B, Lurz R, Docherty K, Kilpatrick MW: The human insulin gene linked polymorphic region exhibits an altered DNA structure. *Nucleic Acid Res* 20:231–236, 1992
  107. Lew A, Rutter WJ, Kennedy GC: Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor pur-1/MAZ. *Proc Natl Acad Sci U S A* 97:12508–12512, 2000
  108. Bennett ST, Wilson AJ, Cucca F, Nerup J, Pociot F, McKinney PA, Barnett AH, Bain SC, Todd JA: IDDM2-VNTR-encoded susceptibility to type 1 diabetes: Dominant protection and parental transmission of alleles of the insulin gene-linked minisatellite locus. *J Autoimmun* 9:415–421, 1996
  109. Ong KK, Phillips DI, Fall C, Poulton J, Bennett ST, Golding J, Todd JA, Dunger DB: The insulin gene VNTR, type 2 diabetes and birth weight. *Nat Genet* 21:262–263, 1999
  110. Bennett AJ, Sovio U, Ruokonen A, Martikainen H, Pouta A, Taponen S, Hartikainen AL, King VJ, Elliott P, Jarvelin MR, McCarthy MI: Variation at the insulin gene VNTR (variable number tandem repeat) polymorphism and early growth: Studies in a large finnish birth cohort. *Diabetes* 53:2126–2131, 2004
  111. Hansen SK, Gjesing AP, Rasmussen SK, Glumer C, Urhammer SA, Andersen G, Rose CS, Drivsholm T, Torekov SK, Jensen DP, Ekstrom CT, Borch-Johnsen K, Jorgensen T, McCarthy MI, Hansen T, Pedersen O: Large-scale studies of the HphI insulin gene variable-number-of-tandem-repeats polymorphism in relation to type 2 diabetes mellitus and insulin release. *Diabetologia* 47:1079–1087, 2004
  112. Kennedy GC, German MS, Rutter WJ: The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat Genet* 9:293–298, 1995
  113. Kennedy GC, Rutter WJ: Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates an insulin promoter in heterologous cells. *Proc Natl Acad Sci U S A* 89:11498–11502, 1992
  114. Neve B, Fernandez-Zapico ME, Ashkenazi-Katalan V, Dina C, Hamid YH, Joly E, Vaillant E, Benmezroua Y, Durand E, Bakaher N, Delannoy V, Vaxillaire M, Cook T, Dallinga-Thie GM, Jansen H, Charles MA, Clement K, Galan P, Herberg S, Helbecque N, Charpentier G, Prentki M, Hansen T, Pedersen O, Urrutia R, Melloul D, Froguel P: Role of transcription factor KLF11 and its diabetes-associated gene variants in pancreatic beta cell function. *Proc Natl Acad Sci U S A* 102:4807–4812, 2005
  115. Clark AR, Wilson ME, London NJ, James RF, Docherty K: Identification and characterization of a functional retinoic acid/thyroid hormone-response element upstream of the human insulin gene enhancer. *Biochem J* 309:863–870, 1995
  116. Eggers A, Siemann G, Blume R, Knepel W: Gene-specific transcriptional activity of the insulin cAMP-responsive element is conferred by NF-Y in combination with cAMP response element-binding protein. *J Biol Chem* 273:18499–18508, 1998
  117. Mantovani R: A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acid Res* 26:1135–1143, 1998
  118. Bartoov-Shifman R, Hertz R, Wang H, Wollheim CB, Bar-Tana J, Walker MD: Activation of the insulin gene promoter through a direct effect of hepatocyte nuclear factor 4 alpha. *J Biol Chem* 277:25914–25919, 2002
  119. Morton NM, Emilsson V, de Groot P, Pallett AL, Cawthorne MA: Leptin

- signalling in pancreatic islets and clonal insulin-secreting cells. *J Mol Endocrinol* 22:173–184, 1999
120. Seufert J, Kieffer TJ, Habener JF: Leptin inhibits insulin gene transcription and reverses hyperinsulinemia in leptin-deficient ob/ob mice. *Proc Natl Acad Sci U S A* 96:674–679, 1999
  121. Tsai SY, Tsai MJ: Chick ovalbumin upstream promoter-transcription factors (COUP-TFs): Coming of age. *Endocr Rev* 18:229–240, 1997
  122. Hwung YP, Crowe DT, Wang LH, Tsai SY, Tsai MJ: The COUP transcription factor binds to an upstream promoter element of the rat insulin II gene. *Mol Cell Biol* 8:2070–2077, 1988
  123. Hwung YP, Wang LH, Tsai SY, Tsai MJ: Differential binding of the chicken ovalbumin upstream promoter (COUP) transcription factor to two different promoters. *J Biol Chem* 263:13470–13474, 1988
  124. Bardoux P, Zhang P, Flamez D, Perilhou A, Lavin TA, Tanti JF, Hellemans K, Gomas E, Godard C, Andreelli F, Buccheri MA, Kahn A, Le Marchand-Brustel Y, Burcelin R, Schuit F, Vasseur-Cognet M: Essential role of chicken ovalbumin upstream promoter-transcription factor II in insulin secretion and insulin sensitivity revealed by conditional gene knockout. *Diabetes* 54:1357–1363, 2005
  125. Barrow J, Hay CW, Ferguson LA, Docherty HM, Docherty K: Transcription factor cycling on the insulin promoter. *FEBS Lett* 580:711–715, 2006
  126. Qiu Y, Sharma A, Stein R: p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. *Mol Cell Biol* 18:2957–2964, 1998
  127. Sharma A, Moore M, Marcora E, Lee JE, Qiu Y, Samaras S, Stein R: The NeuroD1/BETA2 sequences essential for insulin gene transcription colocalize with those necessary for neurogenesis and p300/CREB binding protein binding. *Mol Cell Biol* 19:704–713, 1999
  128. Stanojevic V, Habener JF, Thomas MK: Pancreas duodenum homeobox-1 transcriptional activation requires interactions with p300. *Endocrinology* 145:2918–2928, 2004
  129. Mosley AL, Ozcan S: The pancreatic duodenal homeobox-1 protein (pdx-1) interacts with histone deacetylases hdac-1 and hdac-2 on low levels of glucose. *J Biol Chem* 279:54241–54247, 2004
  130. Mosley AL, Corbett JA, Ozcan S: Glucose regulation of insulin gene expression requires the recruitment of p300 by the beta-cell-specific transcription factor pdx-1. *Mol Endocrinol* 18:2279–2290, 2004
  131. Nightingale KP, O'Neill LP, Turner BM: Histone modifications: Signalling receptors and potential elements of a heritable epigenetic code. *Curr Opin Genet Dev* 16:125–136, 2006
  132. Chakrabarti SK, Francis J, Ziesmann SM, Garmey JC, Mirmira RG: Covalent histone modifications underlie the developmental regulation of insulin gene transcription in pancreatic beta cells. *J Biol Chem* 278:23617–23623, 2003
  133. Francis J, Chakrabarti SK, Garmey JC, Mirmira RG: Pdx-1 links histone H3-lys-4 methylation to RNA polymerase II elongation during activation of insulin transcription. *J Biol Chem* 280:36244–36253, 2005
  134. Francis J, Babu DA, Deering TG, Chakrabarti SK, Garmey JC, Evans-Molina C, Taylor DG, Mirmira RG: Role of chromatin accessibility in the occupancy and transcription of the insulin gene by the pancreatic transcription factor PDX-1. *Mol Endocrinol*. In press
  135. Ferris SD, Whitt GS: Evolution of the differential regulation of duplicate genes after polyploidization. *J Mol Evol* 12:267–317, 1979
  136. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545, 1999
  137. Gu Z, Nicolae D, Lu HH, Li WH: Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet* 18:609–613, 2002