Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues

Jun Takeda, Susumu Seino and Graeme I. Bell*

Howard Hughes Medical Institute and Departments of Biochemistry and Molecular Biology and Medicine, The University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL 60637, USA

Received April 13, 1992; Revised and Accepted July 31, 1992

EMBL accession nos Z11898–Z11901 (incl.)

ABSTRACT

Transcription factors containing the POU-domain have been shown to be important regulators of tissue-specific gene expression in the pituitary and lymphoid cells. Using a polymerase chain reaction (PCR)-based strategy, we have searched for similar factors that may be expressed in adult human pancreatic islets. This approach resulted in the amplification of sequences encoding the octamer binding proteins Oct1 and Oct3 (also called Oct4). The isolation of cDNAs encoding Oct3 revealed the expression of two isoforms of this transcription factor termed Oct3A and Oct3B that are generated by alternative splicing. Human Oct3A and Oct3B are composed of 360 and 265 amino acids, respectively, of which the 225 amino acids at the COOH-termini are identical. The sequence of human Oct3A shows 87% amino acid identity with mouse Oct3. Reverse-transcriptase PCR showed low levels of expression of both Oct3A and Oct3B mRNA in all adult human tissues examined. We also isolated and characterized the human Oct3 gene (*OTF3*) and a related gene, *OTF3C*. The human Oct3 gene, localized to human chromosome 6 in the region of the MHC complex, spans about 7 kb and consists of five exons. The Oct3-related gene, *OTF3C*, is a retroposon and has been localized to human chromosome 8. Southern blotting and PCR amplification of human DNA indicated the presence of other *OTF3*-related genes as has been previously noted in the mouse. Two polymorphisms which can be typed using PCR were identified in *OTF3* which will facilitate genetic studies of this gene.

INTRODUCTION

Transcription factors containing the POU-homeo domain (1–3) have been shown to be important regulators of tissue-specific gene expression in lymphoid (4–6) and pituitary differentiation (3,7–9) and in early mammalian development (10–14). To identify members of the POU family of transcription factors that

may be involved in the tissue-specific regulation of expression of genes in insulin-secreting cells of the pancreas, we used the polymerase chain reaction and degenerate oligonucleotide primers specific for the homeo-domain to amplify POU-related sequences in human pancreatic islet mRNA. The sequences of two of the PCR products were identical to human Oct1 and two others were homologous to mouse Oct3.

Since the sequence of human Oct3 had not been previously reported, we characterized the human Oct3 cDNAs and showed that two forms of Oct3 mRNA are expressed in adult tissues as a result of alternative splicing. We also isolated and characterized the human Oct3 gene. The human Oct3 gene (designated *OTF3*) and a pseudogene-like sequence (*OTF3C*) were localized to chromosomes 6 and 8, respectively. Linkage studies showed that *OTF3* is tightly linked to the genes for *HLA-A*, *B*, *C* and *DR*.

MATERIALS AND METHODS

General methods

Standard procedures were carried out as described in Sambrook et al. (15) and as described previously (16). Human pancreatic islets were provided by D.W. Scharp and P.E. Lacy (Washington University School of Medicine). RNA was isolated using the guanidinium thiocyanate/CsCl method. DNA sequencing was done by the dideoxynucleotide chain-termination procedure (17) after subcloning appropriate DNA fragments into M13mp18 and M13mp19. Both strands were sequenced.

Cloning of cDNAs encoding POU-domain transcription factors expressed in human islets

Five micrograms of DNase I-treated total human pancreatic islet RNA were reverse-transcribed using AMV reverse transcriptase (GIBCO BRL, Gaithersburg, MD) and 10 pmol of primer POU-1, 5'-CGGGATCCGTT(A/G)CA(A/G)AACCAIA-C(A/G)C(G/T)IAT(C/T)AC(C/T)TC-3'. POU-domain related sequences were amplified (18) using primers POU-1 and POU-2, 5'-AGIAGIAA(A/G)AA(A/G)CGIACIA(C/G)IAT-3'. Amplification was carried out using GeneAmp reagents and a

* To whom correspondence should be addressed

Perkin Elmer Cetus DNA Thermal Cycler for 30 cycles using the following conditions: cycles 1–2, denaturation at 95°C for 1 min, annealing at 37°C for 2 min, and extension at 72°C for 3 min; cycles 3–4, denaturation at 95°C for 1 min, annealing at 45°C for 2 min, and extension at 72°C for 3 min; and cycles 5–30, denaturation at 95°C for 1 min, annealing at 55°C for 2 min and extension at 72°C for 3 min (the final extension was for 10 min). The spacing of these degenerate primers is such that they would amplify a 122 bp segment encoding part of the POU-homeo domain. The PCR products were separated on a 3% NuSieve GTG agarose (FMC BioProducts, Rockland, ME) gel and fragments in the region of 122 bp were cloned into the *Hinc* II site of M13mp18 and sequenced.

Isolation of human Oct3 cDNA clones

Human Oct3 cDNA clones were isolated using the rapid amplification of cDNA ends (RACE) procedure (19). For 5'-RACE, one microgram of human pancreatic islet RNA was reverse-transcribed using 1 pmol of the antisense primer, 5'-TCGGACCACATCCTTCTCGA-3'. Second-strand cDNA was prepared using 10 pmol of (dT)₁₇-adapter primer and amplified with the adapter primer and the internal Oct3-specific-antisense primer, 5'-TGCAGGAACAAATTCTCCAGGTGCTCTCACT-3'. For 3'-RACE, the cDNA was prepared using the oligo (dT)₁₇-adapter primer and amplified with the adapter-primer and the Oct3-specific-sense primer, 5'-AGAGGCAACCTGGAGAATTT-3'. The PCR products were separated on a 1% low-melting point agarose gel, cloned into the *Hinc* II site of M13mp18 and sequenced. At least two PCR products were sequenced in order to identify nucleotide substitutions arising as a result of misincorporation by *Taq* DNA polymerase. Human Oct3 cDNA clones were also isolated from an adult human kidney cDNA library in λ gt10 by hybridization with ³²P-labeled Oct3B PCR fragment [corresponding to nucleotides 16–999 in Fig. 1B]. The hybridization and washing conditions were: hybridization—50% formamide, 5×SSC, 2×Denhardt's solution, 20 mM sodium phosphate buffer (pH 6.5), 0.1% sodium dodecyl sulfate, 100 μ g/ml of sonicated and denatured salmon testes DNA, 10% dextran sulfate, and 1×10⁶ cpm/ml of ³²P-labeled probe at 42°C for 16–20 h; and washing—0.1×SSC and 0.1% sodium dodecyl sulfate at room temperature for 1 h and then at 60°C for 30 min.

Amplification of Oct3A and Oct3B mRNA in adult tissues

cDNA was prepared from one microgram of DNase I-treated total RNA from adult human heart, kidney, liver, spleen, and pancreatic islets and from term placenta using reverse transcriptase (RT) and 200 ng of random hexadeoxynucleotide primers (Pharmacia LKB Biotechnology). The cDNA sample was divided into two tubes and Oct3A and Oct3B sequences were amplified in the presence of 1 μ Ci of [α -³²P]-dCTP (Amersham, 3000 Ci/mmol) and a common downstream antisense primer (from exon 1b), 5'-CCACATCGGCCTGTGTATAT-3', from the POU-specific domain and upstream sense primers specific for Oct3A, 5'-CTCCTGGAGGGCCAGGAATC-3' (from exon 1a), and Oct3B, 5'-ATGCATGAGTCAGTGAACAG-3' (from exon 1b), mRNAs. The PCR was for 25 cycles of denaturation at 95°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1 min. The products were separated on a 5% polyacrylamide gel. The gel was dried and exposed to X-ray film for 3 h. The expected sizes of the Oct3A and Oct3B-specific products are 381 and 303 bp, respectively.

Isolation of human Oct3 gene and related sequence

The partial *Alu* I/*Hae* III fetal liver genomic library of Lawn et al. (20) was screened by hybridization as described above using kidney cDNA clones as probes.

Gene mapping

The chromosomal locations of the human Oct3 gene (*OTF3*) and a processed pseudogene-like sequence (*OTF3C*) were determined using Bios PCRable™ DNA from Somatic Cell Hybrids (Bios, New Haven, CT) and primer-pairs specific for *OTF3* [sense (exon 3) primer 5'-AGAGGCAACCTGGAGAATTT-3' and antisense (exon 4) primer 5'-CGTTGTGCATAGTCGCTGCT-3'] and *OTF3C* [sense (5'-flanking region) primer 5'-TGAAGAATCTAGCATCATGGTATG-3' and antisense (POU-specific domain region) primer 5'-AACACCTTCCCAAATAGAAC-3']. The localization of *OTF3* on chromosome 6 was confirmed by linkage by typing a *Hind* III RFLP at this locus in nine CEPH (Centre d'Etude Polymorphisme Humain) families (family numbers 66, 1341, 1362, 1377, 1408, 1423, 1424, 1444 and 1463) (21). Two-point lod scores between this RFLP and others in the CEPH public database, version 4, were determined using the LINKAGE program CLODScore (22). The *Hind* III RFLP was typed using a PCR-based procedure. Briefly, 50 ng of *Eco*RI-digested genomic DNA was amplified using the sense (intron 4) primer 5'-AGCTCATTGTCTAATGTCAT-3' and the antisense (3'-flanking region) primer 5'-CAGCTACATGGTGACTGAGT-3' for 30 cycles: denaturation at 95°C for 1 min, annealing at 63°C for 1 min, and extension at 72°C for 1 min. The 802 bp PCR product was digested with *Hind* III and the fragments were separated on a 1.2% agarose gel; the presence of the *Hind* III site gave fragments of 546 and 256 bp on digestion.

Allele-specific amplification of *OTF3*

The T→G replacement in the second position of the codon for the initiating methionine of Oct3B mRNA was characterized by allele-specific amplification using the sense primers 5'-GGGCAGCTTGGGAAGGCAGAT-3' and 5'-GGGCAGCTTGGGAAGGCAGAG-3' and the antisense primer (intron 2) 5'-ATGCGGTCAGAATCTGCAGA-3'. Briefly, 50 ng of *Eco*RI-digested DNA was amplified for 40 cycles: denaturation at 95°C for 1 min, and annealing and extension at 60°C for 4 min; and the MgCl₂ concentration was 0.5 mM. The PCR product is 290 bp.

RESULTS

Identification of POU-domain transcription factors in pancreatic islets

Messenger RNAs encoding members of the POU-family of transcription factors expressed in human pancreatic islets were amplified in vitro using the primers POU-1 and -2 which correspond to conserved regions of the homeo domain (1–3). Since the expected size of POU-homeo-domain-containing PCR products obtained using these primers was 122 bp, fragments of this size were cloned into M13mp18 and sequenced. 100 clones were sequenced and of these four were homologous to the homeo-domain of the POU family. The sequences of two clones, A1 and A33, were identical to that of human Oct1 cDNA (23). The sequences of the other two clones, K27 and A11, were related

Human	1	MAGHLASDFAFSPPPGGGDPGGPPEPGWVDPRTWLSFQGGPPGGPGIGPG	50
Mouse		-----SA-L-----	49
		VGPGEVWVGIPPCPPPYEFCGGMAYCGPQVGVGLVPQGGLETSQPEGEAGVGVESNSD	108
		:-:-:-L-S-:-:-A-----L-----V-V-L-----Q--AR-----E	103
		GASPEPCTVTPGAVKLEKEKLEQNPEESQDIKALQKELEQFAKLLKQKRITLGYTQAD	166
		-T-S--ADR-N-----:V-PT-----M-----	159
		VGLTLGVLFGRVFSQTTICRFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAE	224
		-----L-----E-----	217
		TLVQARKRKRTSIENVRGNLENLFLQCPKPTLQQITHIAQQLGLEKDVVRVWFCNRR	282
		-----WS--TM--K-----S-----D-----	275
		QKGKRSSSDYAQREDFAAGSPFSGGPFVSFPLAPGPHFGTPGYGSPHFTALYSVSPFP	340
		-----IE-S-:-EY--T-T-P-A-----P-----T-----	332
		EGEAFPPVSVTLGSPMHSN	360
		-----S-P--A-----	352

Figure 2. Comparison of amino acid sequences of human Oct3A and mouse Oct3. Only differences between the two sequences are noted. Gaps introduced to generate this alignment are represented by colons.

a region of 1136 bp and encode a protein of 265 amino acids ($M_r = 30.0$ kDa) (Fig. 1B); there is an inframe translational stop codon upstream of the putative initiating methionine. There is 89% amino acid identity and 92% similarity between residues 41–265 of the human protein and amino acids 129–352 of mouse Oct3 (11–13) consistent with the human protein being Oct3. However, the sequence of amino acids 1–40 of the human protein were unrelated to corresponding region of mouse Oct3. Consequently, we designated this sequence as human Oct3B. Of the three 5'-RACE-generated clones that we analyzed, all encoded Oct3B and none encoded a protein whose NH_2 -terminal sequence was homologous to those of mouse Oct3.

In order to identify other Oct3-related sequences expressed in human tissues, we screened an adult human kidney cDNA library by hybridization. Of 1.2×10^6 phage screened, four clones were identified. One of them λ KHO-4 contained a 1065 bp insert whose sequence was identical to the corresponding region of the composite pancreatic islet Oct3 sequence. The sequence of a second clone, λ KHO-1, having an insert of 1161 bp, showed 99% nucleotide sequence identity with λ KHO-4 and encoded a protein whose sequence was 99.3% identical with that of human Oct3B (Fig. 1B). The third clone, λ KHO-9, had an insert of 1328 bp, excluding polyA tract, and encoded a protein of 359 amino acids with 87% identity with the sequence of mouse Oct3 including a region homologous to amino acids 2–128 of mouse Oct3; this sequence was designated as human Oct3A. Since the sequence of the insert in λ KHO-9 lacked the ATG of the initiating methionine, this region of the cDNA was obtained using the 5'-RACE procedure and human adult kidney RNA; a 45 bp fragment was isolated in this manner. The composite sequence of the human Oct3A cDNA is 1373 bp and encodes a protein of 360 amino acids ($M_r = 38.6$ kDa) (Fig. 1A). There is 87% overall identity between the sequences of human Oct3A and mouse Oct3 and eight of the 56 differences are insertion/deletions (Fig. 2). The fourth kidney cDNA clone, λ KHO-14, had a 532 bp insert corresponding to nucleotides 624 to 1155 of the human Oct3B cDNA, a region that is common to both Oct3A and Oct3B sequences. The nucleotide sequence of the insert in λ KHO-14 was identical to the corresponding region of λ KHO-4 and -9 except for A and G substitutions at nucleotides 1087 and 1136 of the Oct3B cDNA sequence, respectively. The latter nucleotide change results in the loss of a *Hind* III site (Fig. 1B); as described below this sequence difference represents an RFLP.

Figure 3. Expression of Oct3A and Oct3B mRNAs in human tissues by RT-PCR. RT-PCR was carried out as described in Material and Methods. The PCR fragments derived from Oct3A and Oct3B mRNAs are shown.

Expression of Oct3 mRNA

Since our cDNA probes did not show hybridization to RNA blots prepared using total RNA from adult heart, kidney, liver and spleen and term placenta, we used RT-PCR to screen for the presence of Oct3A and Oct3B mRNA as described in Materials and Methods. Since the primer pairs used span an intron in the Oct3 gene, PCR products derived from Oct3A and Oct3B mRNA can be readily distinguished from those generated by amplification of any contaminating DNA. PCR-products of the expected size of 381 and 303 bp derived from human Oct3A and Oct3B mRNA, respectively, were observed in all the adult tissues that we examined as well as pancreatic islets (Fig. 3). In pancreatic islets, the relative abundance of Oct3B mRNA was higher than Oct3A, whereas in the other tissues that we tested, the levels were similar.

Isolation and characterization of the human Oct3 gene

About 1×10^6 phage were screened by hybridization with ^{32}P -labeled Oct3B cDNA clone. Twenty-one clones were identified and of these five hybridized with probes derived from the 5'- and 3'-ends of the Oct3A and Oct3B cDNA clones. These latter clones were characterized further.

Clones λ HG-13, -14, -15 and -16 formed a contig of 18 kb that included the human *OTF3* (Fig. 4A). The sequence of a region of 6.2 kb indicated that *OTF3* was comprised of five exons (Fig. 4A, Table 1) that span ~ 7 kb. The exon-intron organization of the human Oct3 gene is very similar to that previously determined for the mouse gene (24); the four introns located in identical positions in both genes. Of interest in this gene is the presence of an atypical splice donor site at the beginning of intron 1a. The human Oct3 gene has a GC here rather than the usual GT. As reviewed by Jackson (25), there are 26 other examples where the 5'-splice site is GC which is the only variant of the 5'-splice site that has been identified. The biological consequences of the presence of a GC with respect to expression of Oct3A and Oct3B mRNAs are unknown.

A comparison of the sequences of the gene and of the Oct3A and Oct3B cDNA clones indicated that Oct3A and Oct3B mRNA are derived from the same gene by alternative splicing (Fig. 4B). Oct3A mRNA is generated from exon 1a, the 3'-half of exon 1b, and exons 2–4 through the use of an internal splice acceptor site in exon 1b. Oct3B mRNA includes exons 1b–4.

The sequence of the exon 1b of clones λ HG-13 to -16 showed one interesting difference. Clones λ HG-13, -14 and -15, whose sequences appear to be identical, have AGG in the position of the initiating methionine of Oct3B mRNA while λ HG-16 has an ATG at this position. Since the nucleotide sequences of the exons,

Figure 4. Exon-intron organization of the human Oct3 gene and alternative splicing. A. Map of clones comprising the human Oct3 gene. The five exons are shown as boxes. The locations of the *Hind* III (H) and ATG/AGG polymorphisms described in the text are shown. The locations of sites for several restriction endonuclease are noted: S, *Sal* I; R, *Eco*RI; and X, *Xba* I. B. Alternative splicing of Oct3 mRNA. The arrow shows the location of the internal splice acceptor site in exon 1b.

Table 1. Exon-intron organization of *OTF3*

Sequence of exon-intron junction					
Exon no.	Exon size	5'-Splice donor	3'-Splice acceptor	Intron size	Amino acid interrupted
1a	447 bp	GAGGAG gcaagtgaagcttc.....ttttaaatccag	TCCCAG	5 kb	Glu-135
1b	344 bp	TATTTG gtgggtcccttc.....tgccctcccccag	GGAAGG	227 bp	Gly-176
2	132 bp	GAGGAG gtaagggtggag.....cccactgcacag	ATATGC	285 bp	Glu-219
3	159 bp	AAGGAT gtgagtgccatgt.....cattcacttcag	GTGGTC	260 bp	Asp-272
4	513 bp				

The sites at which introns interrupt the mRNA and protein sequence are indicated. An atypical splice donor sequence occurs in intron 1a, GC. Exon and intron sequences are shown in capital and lower-case letters, respectively. Note that part of Oct3B mRNA is encoded by the 'intron' upstream of exon 1b. The partial sequence of human *OTF3* has been deposited in the EMBL/GenBank database with accession number Z11900.

introns and flanking regions of these two types of clones are 99.6% identical, we believe that they are the products of alleles of *OTF3*. Moreover, this sequence difference implies that some individuals may not express Oct3B. Allele-specific amplification of both types of alleles suggests that this is indeed the case (Fig. 5); the frequencies of the ATG and AGG alleles in a group of 51 unrelated Caucasians are 0.67 and 0.33, respectively.

Whereas the sequences of clones λHG-13 to -16 indicate that *OTF3* is interrupted by introns, the sequence of the region of clone λHG-17 homologous to Oct3 mRNA showed 97.5% nucleotide sequence identity with the sequence of Oct3A cDNA and lacked introns (Fig. 6). The presence of a polyA tract at the 3'-end of the Oct3-like sequence in λHG-17 and of flanking direct repeats are consistent with this sequence, designated as *OTF3C*, being a retroposon (26). The sequence of the Oct3A-like protein

Figure 5. DNA polymorphisms in the human Oct3 gene. A. and B. Allele-specific amplification of the ATG and AGG alleles, respectively. The genotypes of the five individuals shown here are from left to right: ATG/AGG, ATG/ATG, ATG/ATG, ATG/AGG and AGG/AGG. C. Characterization of *Hind* III polymorphism. The 802 bp PCR product was digested with *Hind* III. The genotypes of the individuals shown are from left to right: +/–, –/–, +/+, +/+ and +/–. The molecular weight markers at the left are a *Hae* III digest of ΦX174 DNA.

encoded by λHG-17 differs from Oct3A at 14 residues and by the deletion of a triplet corresponding to Gln²⁵⁸; a similar deletion was noted in one of the PCR products, A11, obtained using DNase-I-treated human pancreatic islet RNA.

Hybridization of Southern blots of *Eco*RI and *Hind* III digests of human DNA with an Oct3B cDNA probe revealed multiple hybridizing fragments (Fig. 7) suggesting that there may be a family of Oct3-related sequences in humans as there are in mice (27). By contrast, the relatively simple patterns noted with

```

1  CGTTCACCAATCTGCAGGAATAAGAATCAGAAAGAAGGACACACCATTCATAGCACAGAACACAAGTAGATGGCTCCCATGAAGAATCTAGCATCATGGTATGGCAGAAGCAGAAT
120 GGACTCAGGAATCATGGTGCCTACACTGGAATTTTCAGATCGGCACCTGACACAATGGACACATTATTCACATTTCCAAATCTTGGCATTCTTATCCACAAGTAGAATAAATTTGT
240 CAATTCACAGGTGATGATTATGATTAAAGAGATTACTTTTGAAGAGTTCTCAACACATTCAGTCAACATTTAATGATGCTTCAGGCACAGAGTTTCATTGTAGTGAGCGTATGACACACAC
1  Met Ala Gly His Leu Ala Ser Asp
360 AGCCATACGGTCACAGAGCTTTCAATGAAAAGTAACATAATTG [CTCATTTTACCAGGCCCCGGCTTGGGGCGCTTCCTTCCCC ATG GCG GGA CAC CTG GCT TCG GAT
10 20 30
Phe Ala Phe Ser Pro Pro Gly Gly Gly Gly Asp Gly Pro Trp Gly Ala Glu Pro Gly Trp Val Asp Pro Leu Thr Trp Leu Ser Phe
470 TTC GCC TTC TCG CCC CCT CCA GGC GGT GGG GGT GAT GGG CCA TGG GGG GCG GAG CCG GGC TGG GTT GAT CCT CTG ACC TGG CTA AGC TTC
40 50 60
Gln Gly Pro Pro Gly Gly Pro Gly Ile Gly Pro Gly Val Gly Pro Gly Ser Glu Val Trp Gly Ile Pro Cys Pro Pro Pro Tyr Glu
560 CAA GGC CCT CCT GGA GGG CCA GGA ATC GGG CCG GGG GTT GGG CCA GGC TCT GAG GTG TGG GGG ATT CCC CTG TGC CCC CCG CCG TAT GAG
70 80 90
Leu Cys Gly Gly Met Ala Tyr Cys Gly Pro Gln Val Gly Val Gly Leu Val Pro Gln Gly Gly Leu Glu Thr Ser Gln Pro Glu Ser Glu
650 TTA TGT GGG GGG ATG GCG TAC TGT GGG CCT CAG GTT GGA GTG GGG CTA GTG CCC CAA GGC GGC TTG GAG ACC TCT CAG CCT GAG AGC GAA
100 110 120
Ala Gly Val Gly Val Glu Ser Asn Ser Asn Gly Ala Ser Pro Glu Pro Cys Thr Val Pro Pro Gly Ala Val Lys Leu Glu Lys Glu Lys
740 GCA GGC GTC GGG GTG GAG AGC AAC TCC AAT GGG GCC TCC CCG GAA CCC TGC ACC GTC CCC CCT GGT GCC GTG AAG CTG GAG AAG GAG AAG
130 140 150
Leu Glu Gln Asn Pro Glu Lys Ser Gln Asp Ile Lys Ala Leu Gln Lys Glu Leu Glu Gln Phe Ala Lys Leu Leu Lys Gln Lys Arg Ile
830 CTA GAG CAA AAC CCG GAG AAG TCC CAG GAC ATC AAA GCT CTG CAG AAA GAA CTC GAG CAA TTT GCC AAG CTC CTG AAG CAG AAG AGG ATC
160 170 180
Thr Leu Gly Tyr Thr Gln Ala Asp Val Gly Leu Ile Leu Gly Val Leu Phe Gly Lys Val Phe Ser Gln Lys Thr Ile Cys Arg Phe Glu
920 ACC CTG GTC AAT ACA CAG GCC GAT ATC GGG CTC ATC AAG AGA AAG CTA TTT GGG GAT Leu Leu Phe Gly Lys Val Phe Ser Gln Lys Thr Ile Cys Arg Phe Glu
190 200 210
Ala Leu Gln Leu Ser Phe Lys Asn Ile Cys Lys Leu Arg Pro Leu Leu Gln Lys Trp Val Glu Glu Ala Asp Asn Asn Glu Asn Leu Gln
1010 GCT CTG CAG CTT AGC TTC AAG AAC ATC TGT AAG CTG CCG CCC TTG CTG CAG AAG TGG GTG GAG GAA GCT GAC AAC AAT GAA AAT CTT CAG
220 230 240
Glu Ile Cys Lys Ala Glu Thr Leu Met Gln Ala Arg Lys Arg Lys Arg Thr Ser Ile Glu Asn Arg Val Arg Gly Asn Leu Glu Asn Leu
1100 GAG ATA TGC AAA GCA GAA ACC CTC GAG CAG GCC CGA AAG AGA AAG CTA TTT GGG GAT Leu Leu Phe Gly Lys Val Phe Ser Gln Lys Thr Ile Cys Arg Phe Glu
250 260 270
Phe Leu Gln Cys Pro Lys Pro Thr Leu Gln Ile Ser His Ile Ala Gln Gln Leu Gly Leu Glu Lys Asp Val Val Arg Val Trp Phe Cys
1190 TTC CTG CAG TGC CCG AAA CCC ACA CTG CAG ATC AGC CAC ATC GCC CAG CAG CTT GGG CTC GAG AAG GAT GTG GTC CGA GTG TGG TTC TGT
280 290 300
Asn Arg Arg Gln Lys Gly Lys Arg Ser Ser Ser Asp Tyr Ala Gln Arg Glu Asp Phe Glu Ala Ala Gly Ser Pro Phe Ser Gly Gly Pro
1280 AAC CGG CGC CAG AAG GGC AAG CGA TCA AGC AGC GAC TAT GCA CAA CGA GAG GAT TTT GAG GCT GCT GGG TCT CCT TTC TCA GGG GGA CCA
310 320 330
Val Ser Phe Pro Pro Ala Pro Gly Pro His Phe Gly Thr Pro Gly Tyr Gly Ser Pro His Phe Thr Ala Leu Tyr Ser Ser Val Pro Phe
1370 GTG TCC TTT CCT CCG GCC CCA GGG CCC CAT TTT GGT ACC CCA GGC TAT GGG AGC CCT CAC TTC ACT GCA CTG TAC TCC TCG GTC CTT TTC
340 350 359
Pro Glu Gly Glu Val Phe Pro Val Ser Val Ile Thr Leu Gly Ser Pro Met His Ser Asn OP
1460 CCT GAG GGG GAA GTC TTT CCC CCA TCC GTC ATC ACT CTG GGC TCT CCC ATG CAT TCA AAC TGA GGTGCTGCCCCCTTAGGAATGGGGAACAGG
1558 GGAGGGGAGGAGCTAGGGAAGAGAACCTGGAGTTTGGCCAGGGCTTTTGGGATTAAAGTTCTTCACTTCAAGGAAGGAATTGGGAACACAAAGGGTGGGGCAGGGGAGTTTGGGG
1678 CAACTGGTTGGAGGGAAGGTGAAGTTCAATGATGCTCTTATTATTTATCCACATCATGATCACTTTTCTTAAATAAAGAAGCCTGGGACACAGTAAAAAAGAAAAAGAAAG
1798 AAAAGAAAA] GAAAAGTAACATAATTGAGTAATAATTTTAAAGTGTGTAATAATATGTCGCACATAGTAGGTATTCAATAAATACTTGCTTTTTCTCTCTTTTACCTTTTCTCTAG
1916 AGTCCTACTTGATGGTAGCTGAGTTTCTTCAAGACATCTATTGTCTGTATACAAAACCTCTTTTGACCGCTAACCGCATAGTCCG

```

Figure 6. Nucleotide and predicted amino acid sequence of the Oct3 retroposon-like sequence, *OTF3C*. The boundaries of the Oct3A-like sequence are noted by square brackets. The direct repeats at the ends of this sequence are underlined. The accession number in the EMBL/GenBank database for the sequence of *OTF3C* is Z11901.

monkey (*Macaca nemestrina*), rat and guinea pig DNAs suggest that they may also be species differences in the size of this gene family. The Oct3B probe hybridized to human *EcoRI* fragments of 17, 12, 8.2, 6.1, 4.8 and 2.4 kb of which the 6.1 and 17 kb fragments correspond to the sizes of the *EcoRI* fragments containing *OTF3* and *OTF3C*, respectively.

In order to estimate the size of *OTF3* family, we amplified human DNA with the primers used for mapping *OTF3* (described in Material and Methods). Agarose gel electrophoresis of the PCR products showed fragments of 441 and 181/178 bp corresponding to the products from genes with and without introns in this regions, respectively (data not shown). The sequences of five clones showed that one was identical to intron-containing *OTF3* sequence. The other four clones lacked the intron between exons 4 and 5 (Fig. 8). The comparison of the sequences suggest that there may be one *OTF3* and at least four related sequences in the human genome, at least two of which are likely to be pseudogenes since they contain in-frame translational stop codons.

Chromosomal location of the human Oct3 gene

Using the pairs of sequence-specific primers described above and DNA from a panel of somatic-cell hybrids, the human *OTF3* and *OTF3C*, were assigned to chromosomes 6 and 8, respectively (Table 2). The localization of *OTF3* to chromosome 6 was confirmed by linkage by typing a *Hind III* polymorphism (AAGCTTAAGCGT at nucleotide 1136 of the Oct3B cDNA sequence (Fig. 1B)) in nine CEPH families. The two-point lod scores indicate that *OTF3* is tightly linked to *HLA-A*, *-B*, *-C* and *-DR* (Table 2).

Figure 7. Southern blot of Oct3 gene and related sequences. Ten micrograms of human, monkey (*Macaca nemestrina*), cat, rat, mouse and guinea pig DNA were digested with *EcoRI* or *Hind III*, blotted and hybridized with ³²P-labeled human Oct3B cDNA probe.

Figure 8. Comparison of sequences of PCR products after amplification of *OTF3* and related sequences. The sequences of the PCR products are shown. Only differences between the sequences are noted. The arrow indicates the location of intron 4; this intron is not present in the products derived from the Oct3-related genes shown here. The colons denote a gap, and the in-frame termination codon (TAA) is underlined. The sequence of the PCR product denoted as Oct3-related gene-4 is identical to that of *OTF3C*.

DISCUSSION

Using a PCR-based strategy, two members of the POU-family of transcription factors were identified in human pancreatic islets. One of these, Oct1 (23), is expressed in many tissues and has been implicated in transcription activation of a number of genes including histone H2B (28). Oct1 can also stimulate initiation of adenovirus DNA replication suggesting that it may be involved in regulation of the cell cycle (29,30). The other member of the POU-family, Oct3, had only been shown previously to be expressed in embryonic tissues where it is involved in the regulation of differentiation of embryonic stem cells (10–14). Our data indicate that Oct3 mRNA is present at low but detectable levels in adult human tissues. In addition, whereas only a single form of Oct3 mRNA has been identified in embryonic mouse tissues, two forms generated by alternative splicing were identified in adult human tissues. Oct3A and Oct3B mRNAs encode proteins that share a common DNA binding POU-homeo-domain but differ in sequence at their NH₂-termini. This region of Oct3A is rich in proline and glycine residues (24 and 28 of 135 amino acids, respectively) whereas the NH₂-terminus of Oct3B is not. Since proline-rich regions form part of the transcriptional activation domain of other transcription factors such as Oct2 and CTF/NF1 (31,32), alternative splicing of Oct3 mRNA may generate proteins with common DNA-binding domains but different activation domains thereby endowing Oct3A and Oct3B with different functional properties. Both Oct1 and Oct3 bind to the octamer motif ATGCAAAT to activate transcription (1,2). However, none of the cell-type specific genes expressed in pancreatic islets including insulin (33), islet amyloid polypeptide (34), GLUT2/liver glucose transporter (unpublished data) or glucokinase (35) contain a canonical octamer motif in the region of the promoter although there are related AT-rich elements. Thus, the target(s) for Oct3A and Oct3B in pancreatic islets is unknown.

There are multiple Oct3-like genes in both mice and humans. The human Oct3 gene, *OTF3*, has been mapped to chromosome 6 and is tightly linked to the HLA loci. The cognate mouse gene, *Otf-3*, has been localized to chromosome 17 (14,27) in a region which is syntenic with the MHC region of human chromosome 6. The *OTF3C* maps to chromosome 8 and thus may be the human homolog of the locus *Otf-3c* on mouse chromosome 3 (27).

Although the nucleotide sequences of λ KHO-1, -4, and -14 are similar, they are not identical. Since we expected to isolate

Table 2. Chromosomal assignment of human Oct3 (*OTF3*) and a related gene (*OTF3C*)

A. Segregation of *OTF3* and *OTF3C* with human chromosomes in a human-hamster cell hybrid panel

	Human chromosome																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
<i>OTF3</i>																								
Discordant	7	5	6	5	18	0	4	7	6	5	7	4	8	8	8	6	6	8	7	5	5	8	5	4
Informative	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
% Discordant	28	20	24	20	72	0	16	28	24	20	28	16	32	32	32	24	24	32	28	20	20	32	20	16
<i>OTF3C</i>																								
Discordant	8	6	9	3	19	7	5	0	8	8	6	8	9	10	7	6	4	7	10	8	10	7	4	9
Informative	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
% Discordant	32	24	36	12	76	28	20	0	32	32	24	32	36	40	28	24	16	28	40	32	40	28	16	36

B. Pairwise linkage of *OTF3* with chromosome 6 markers

Marker	Recombination fraction	Z_{max}
	$(\theta_a = \theta_f)$	
HLA-A	0.0000	5.72
HLA-B	0.0000	10.84
HLA-C	0.0000	6.62
HLA-DR	0.0000	7.83

only two different sequences corresponding to the products of both alleles of *OTF3*, these results imply that one or more of the Oct3-related genes is transcribed. Moreover, the isolation of a RT-PCR product from pancreatic islet RNA, clone A11, whose sequence was identical to that of *OTF3C* suggests that this locus is transcribed. SCIP/Oct6/Tst-1, another member of the POU-domain family of transcription factors that is expressed in proliferative glial progenitor cells, also has features of an expressed retroposon (36). If *OTF3C* mRNA encodes a protein, this protein may have altered DNA binding properties because of a deletion of one amino acid in a region involved in DNA binding as well as in specific interactions with other proteins including other members of the POU-family (37–40).

In our analysis of *OTF3*, we identified two polymorphisms one of which was in the codon that we believe represents the initiating methionine of Oct3B, ATG→AGG. This polymorphism has important functional consequences since individuals having AGG should not be able to express Oct3B. We have confirmed this polymorphism using a PCR-based strategy and the frequencies of the three genotypes, ATG/ATG, ATG/AGG and AGG/AGG, among a groups of 51 healthy unrelated Caucasians was 0.41, 0.51 and 0.08, respectively. These results indicate that a

significant fraction of the population has no or reduced levels of Oct3B protein. Since these individuals are otherwise normal, the absence of Oct3B seems to have few if any physiological consequences. Moreover, we have examined the sequence of the mouse *Otf3* gene (24) for a region homologous to that encoding the unique NH₂-terminus of human Oct3B. The sequences of the human and mouse genes in the region upstream of exon 2 are quite divergent suggesting that the mouse gene does not encode a protein similar to human Oct3B. This result implies that Oct3B does not serve a function that is crucial for mammalian development.

Our analysis of human pancreatic islet RNA for the presence of mRNAs encoding members of the POU-family resulted in the identification of Oct1 and Oct3. Further studies will be necessary to determine whether the endocrine pancreas, like the pituitary, expresses a tissue-specific POU domain protein.

ACKNOWLEDGEMENTS

We thank Drs. H. Singh and E.B. Crenshaw III for their critical reading of this manuscript. These studies were supported by the Howard Hughes Medical Institute and NIH grants DK-20595 and 42086.

REFERENCES

- Ruvkun, G. and Finney, M. (1991) *Cell*, **64**, 475–478.
- Rosenfeld, M.G. (1991) *Genes & Dev.*, **5**, 897–907.
- Ingraham, H.A., Albert, V.R., Chen, R., Crenshaw, E.B. III, Elsholtz, H.P., He, X., Kapiloff, M.S., Mangalam, H.J., Swanson, L.W., Treacy, M.N. and Rosenfeld, M.G. (1990) *Annu. Rev. Physiol.*, **52**, 773–791.
- Müller, M.M., Ruppert, S., Schaffner, W. and Matthias, P. (1988) *Nature*, **336**, 544–551.
- Scheidereit, C., Cromlish, J.A., Gerster, T., Kawakami, K., Balmaceda, C.-G., Currie, R.A. and Roeder, R.G. (1988) *Nature*, **336**, 551–557.
- Clerc, R.G., Corcoran, L.M., LeBowits, J.H., Baltimore, D. and Scharp, P.A. (1988) *Genes Dev.*, **2**, 1570–1581.
- Bodner, M., Castrillo, J.-L., Theill, L.E., Deerinck, T., Ellisman, M. and Karin, M. (1988) *Cell*, **55**, 505–518.
- Ingraham, H.A., Chen, R., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988) *Cell*, **55**, 519–529.
- Simmons, D.M., Voss, J.W., Ingraham, H.A., Holloway, J.M., Broide, R.S., Rosenfeld, M.G. and Swanson, L.W. (1990) *Genes Dev.*, **4**, 695–711.
- Schöller, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N. and Gruss, P. (1989) *EMBO J.*, **8**, 2543–2550.
- Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H. (1990) *Cell*, **60**, 461–472.
- Schöller, H.R., Ruppert, S., Suzuki, N., Chowdhury, K. and Gruss, P. (1990) *Nature*, **344**, 4354–438.
- Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W.J. and Staudt, L.M. (1990) *Nature*, **345**, 686–692.
- Schöller, H.R., Dressler, G.R., Balling, R., Rohdewohld, H. and Gruss, P. (1990) *EMBO J.*, **9**, 2185–2195.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Fukumoto, H., Seino, S., Imura, H., Seino, Y., Eddy, R.L., Fukushima, Y., Byers, M.G., Shows, T.B. and Bell, G.I. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 5434–5438.
- Sanger, F., Coulson, A.R., Barrel, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.*, **143**, 161–178.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) *Science*, **239**, 487–491.
- Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 8998–9002.
- Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) *Cell*, **15**, 1157–1174.
- Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J.M. and White, R. (1990) *Genomics*, **6**, 575–577.
- Lathrop, G.M., Lalouel, J.M., Julier, C. and Ott, J. (1985) *Am. J. Hum. Genet.*, **37**, 482–498.
- Sturm, R.A., Das, G. and Herr, W. (1988) *Genes Dev.*, **2**, 1513–1516.
- Okazawa, H., Okamoto, K., Ishino, F., Ishino-Kaneko, T., Takeda, S., Toyoda, Y., Muramatsu, M. and Hamada, H. (1991) *EMBO J.*, **10**, 2997–3005.
- Jackson, I.J. (1991) *Nucleic Acids Res.*, **19**, 3795–3798.
- Vanin, E.F. (1985) *Annu. Rev. Genet.*, **19**, 253–272.
- Siracusa, L.D., Rosner, M.H., Vigano, M.A., Gilbert, D.J., Staudt, L.M., Copeland, N.G. and Jenkins, N.A. (1991) *Genomics*, **10**, 313–326.
- Sive, H.L. and Roeder, R.G. (1986) *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 6382–6386.
- Verrijzer, C.P., Kal, A.J. and van der Vliet, P.C. (1990) *EMBO J.*, **6**, 1883–1888.
- Robert, S.B., Segil, N. and Heintz, N. (1991) *Science*, **253**, 1022–1025.
- Mermod, N., O'Neil, E.A., Kelly, T.J. and Tjian, R. (1989) *Cell*, **58**, 741–753.
- Gerster, T., Balmaceda, C.-G. and Roeder, R.G. (1990) *EMBO J.*, **9**, 1635–1643.
- Bell, G.I., Pictet, R.L., Rutter, W.J., Cordell, B., Tischer, E. and Goodman, H.M. (1980) *Nature*, **284**, 26–32.
- Nishi, M., Sanke, T., Seino, S., Eddy, R.L., Fan, Y.-S., Byers, M.G., Shows, T.B., Bell, G.I. and Steiner, D.F. (1989) *Mol. Endocrinol.*, **3**, 1775–1781.
- Magnuson, M.A. and Shelton, K.D. (1989) *J. Biol. Chem.*, **264**, 15936–15942.
- Stern, H.L., Tanaka, M. and Herr, W. (1989) *Nature*, **341**, 624–630.
- Ingraham, H.A., Flynn, S.E., Voss, J.W., Albert, V.R., Kapiloff, M.S., Wilson, L. and Rosenfeld, M.G. (1990) *Cell*, **61**, 1021–1033.
- Voss, J.W., Wilson, L. and Rosenfeld, M.G. (1991) *Genes Dev.*, **5**, 1309–1320.
- Verrijzer, C.P., van Oosterhout, J.A.W.M. and van der Vliet, P.C. (1992) *Mol. Cell. Biol.*, **12**, 542–551.