Original Paper

Eur J Hum Genet 1995;3:344-350

Catherine Bartoli^a Jean Charles Dagorn^b Michel Fontes^a Jean Louis Bergé-Lefranc^a

a INSERM U 406 et

^b INSERM U 315, Marseille, France

A Limited Genomic Region Contains the Human REG and REG-IRelated Genes

Key Words

REG gene Physical mapping Chromosome 2

Abstract

A glycoprotein expressed in exocrine pancreas (where it has been called lithostathine) and endocrine pancreas (where it has been called the regeneration protein) is encoded by a gene (REG) which maps to 2p12. A REG-related sequence (REGL) is also located in 2p12 and expressed in the pancreas. Here we describe the physical mapping of these genes within a 100-kb genomic region. A YAC clone was converted into an ordered cosmid contig. We constructed a restriction map of the cosmid contig and localized the loci corresponding to the genes. A third REG-related sequence also maps to this region. Although this sequence was previously described as a pseudogene, we show here that it is also expressed in the pancreas.

Introduction

Lithostathine is a secreted glycoprotein of the pancreatic juice [1] produced by the acinar cells of the pancreas [2, 3]. The pancreatic juice is supersaturated in calcium carbonate [4] and lithostathine is believed to act as an inhibitor of calcium carbonate crystal growth [5]. In 1988, lithostathine mRNA was found in the islets of Langerhans during their regeneration [6]. The corresponding gene, REG (i.e. regenerating gene), has been cloned and

> Received: May 29, 1995 Revision received: August 7, 1995 Accepted: August 11, 1995

its structure has been reported [7]. It is now established that lithostathine and the REG protein are different names for a single protein encoded by a single-copy gene, the REG gene. Recently, Watanabe et al. [8] demonstrated that the REG protein stimulates β cell proliferation and is thus a growth factor which might be used in the treatment of diabetes mellitus.

Recently, we [9] and others [10] have characterized a gene (REGL, i.e. regenerating gene like) whose coding regions exhibited more

© 1995 S. Karger AG, Basel 1018-4813/95/ 0036-0344\$8.00/0 than 90% identity with those of the REG gene. The REGL gene is expressed in the human pancreas and liver.

We have determined the chromosomal localization of the REG and REGL genes using in situ hybridization [11], and both map to 2p12. Their identical localization, the strong homology between them and their concomitant expression in the pancreas suggested that they were organized in tandem.

Watanabe et al. [7] also described a third REG genomic sequence. The presence of a stop codon in the potential coding region allowed these authors to postulate that it was a pseudogene. We report here evidence that this sequence is specifically expressed in pancreatic tissue and also demonstrate that REG, REGL and the 'REG pseudogene' map very close together since they are included in a genomic region of about 100 kb.

Materials and Methods

YAC Library Screening

A YAC library obtained from CEPH [12] was screened by PCR using primers specific for REG and REGL sequences. REGL primers (5' primer: 5'GCT-CCACCCATTGTTTATATC3'; 3' primer: 5'GCATC-AACCCAGGTCTCAG3') were localized at positions 165–185 and 1415–1397, respectively, in the REGL gene [9]. REG primers (5' primer: 5'TTAGAAATATA-AATTTAACCTACCCCTTGAGG3'; 3' primer: 5'CAG-GCAGGAGATCAGGATGAAGTAT3') were localized at positions 710–741 and 1612–1588 in the REG gene [7].

Cosmid Library Construction and Screening

The YACs were partially digested with Sau3A and the resulting fragments were cloned into cosmid vectors (Supercos 1, Stratagene). A library of 2×10^3 recombinant clones was obtained and screened using total human DNA as probe. Positive clones were picked and further analyzed using the REG cDNA probe and REG- or REGL-specific probes that were constructed as described below. Cosmid DNAs containing REG-related sequences were prepared by alkaline lysis from about 3 out of 10 ml of overnight cultures [13] and digested with various restriction enzymes. To identify the extremities of the cloned fragments, cosmid DNA was digested with *Eco*RI, separated by electrophoresis and transferred onto nylon membranes. T3 and T7 oligonucleotides were 5' end labelled using γ^{32} P-ATP and hybridized to the membranes. End fragments were localized and purified.

Preparation of Probes and Hybridization Conditions

The REG cDNA probe was used to detect all the REG-related sequences [11]. The REG and REGL genes were separately detected using two different genomic DNA fragments located in intron IV of both genes: a *PstI-XbaI* fragment (706 nucleotides long) located at position 1644–2349 of the REG gene [7] and a *PstI-Hind*III fragment (596 nucleotides long) located at position 1632–2227 of the REGL gene [9]. DNA fragments were cloned in pUC vectors and used as probes. We have verified that these probes do not cross-hybridize.

Membranes were prehybridized and hybridized for 16 h at 65°C according to the manufacturer's recommendations (Pall Biodyne) and washed at 65°C for 20 min in $1 \times SSC$, 0.1% SDS, 20 min in 0.5 $\times SSC$, 0.1% SDS, and 60 min in 0.1 $\times SSC$, 0.1% SDS then exposed to X-ray films.

Tissue-Specific Expression and Sequencing of the REG Pseudogene Products

The expression of the REG pseudogene [7] was analyzed using RT-PCR tests: cDNA was prepared from polyA⁺ RNA (obtained from Stratagene) using M-MLV reverse transcriptase and oligo-dT primers. Primers of the REG pseudogene were designed in order that their 3' ends did not cross-hybridize with the and REGL sequences. The 5' primer REG (5'GGCTCAGACCAACTCATG3') was localized at position 156-173 and the 3' primer (5'TAAACC-CAGGTCTCATGC3') at position 953-936 in the sequence described by Watanabe et al. [7]. We have verified that these primers do not amplify REG- and REGL-containing clones and amplified a fragment of expected size (798 bp) when human genomic DNA was used as template. In control reactions, incubations were performed without reverse transcriptase. Amplification was performed as follows for 30 cycles: denaturation at 94°C for 1 min, annealing at 60°C for 1 min and DNA synthesis at 72°C for 1 min. The amplified fragments were cloned in a PCR II vector (TA cloning, Invitrogen) and the nucleotide sequences of the fragements were determined.



Fig. 1. Southern blot analysis of YAC DNA digested by *Hind*III and hybridized with a REG cDNA probe. Lane 1 = YAC 56A10; 2 = YAC 155E9; 3 = YAC 238E3; 4 = yeast genomic DNA; 5 = human genomic DNA.

Results

YAC Isolation and Analysis

The CEPH YAC library was screened using PCR primers corresponding to the REG and REGL genes (see Materials and Methods). Four positive YAC pools were detected by REG primers and five were obtained using REGL primers. Of the positive pools, three were detected by both REG and REGL, indicating the presence of YACs potentially containing both loci. The corresponding subpools were further screened using the same primers. Three individual YAC clones were isolated: 155E9, 56A10 and 238E3. DNA was prepared from each clone [14], digested with HindIII and hybridized with a REG cDNA probe to detect the sequences of REG, REGL and also the REG pseudogene [11]. The hybridization patterns obtained (fig. 1) were identical to those obtained with genomic DNA, exhibiting four hybridizing fragments (6.9, 3.6, 3.2 and 2.4 kb long). We have previously reported [9, 11] that the two HindIII fragments (3.2 and 2.4 kb) include the complete REGL gene, that the 6.9-kb fragment contains the REG pseudogene and that the 3.6-kb HindIII fragment contains the complete REG gene. Thus, a single YAC contained all three REG-related sequences. The YAC clone containing the smaller genomic insert (YAC 155E9) was analyzed further.

Cosmid Library Construction and Analysis

To establish a more precise restriction map of the cloned region, a cosmid library was constructed from the YAC 155E9. Partial Sau3A digestions of the YAC were performed and fragments with an average size of 40 kb were cloned into cosmid vectors. A library of 2 \times 10³ recombinants was obtained and screened with a probe for total human DNA. Forty positive clones were picked representing three times the coverage of the YAC, and were further analyzed with the REG cDNA probe. Six clones were obtained (named A, B, C, D, H and I). We used HindIII digestions of the cloned DNA and compared the sizes of the hybridizing fragments with those previously described [9, 11] for the genes. We observed (fig. 2) that clones B and H contained the two HindIII fragments of the REGL gene. Clone I contained the 3.6-kb HindIII fragment of the REGL gene. These first results were confirmed by using the REG- and REGL-specific probes (data not shown). Clones A and D con-



Fig. 2. Electrophoresis of cosmid DNAs digested by *Hin*dIII (a) and Southern blot analysis after hybridization with a REG cDNA probe (b). Lane 1 = clone B; 2 = clone H; 3 = clone I; 4 = clone A; 5 = clone C; 6 = clone D; M = DNA digested with *Hin*dIII.



Fig. 3. Physical map YAC 155E9 (upper line) and the cosmid contig. Cosmids were ordered using their *Eco*RI end fragments as walking probes. Bs = *Bss*HII; Nr = *Nar*I; Eg = *Eag*I; Sc = *Sac*II; Cl = *Cla*I; Xh = *Xho*I. The REGL, REG and REG pseudogene loci are indicated.

tained the *Hin*dIII fragment of the 'REG pseudogene', while clone C contained only a portion of this fragment. These data have been confirmed by PCR assays, using specific primers (data not shown). In that part of the study, we did not attempt to locate the exact position of the three REG-related genes in the

cloned sequences and focused our interest on the genomic organization of the region.

Contig Construction

A cosmid contig (fig. 3) was constructed after comparison of the sizes of restriction fragments and hybridization with the single-



Fig. 4. Expression of a sequence described as the REG pseudogene in various human tissues revealed by RT-PCR. Lane 1 = pancreas; 2 = kidney; 3 = fetal kidney; 4 = liver; 5 = fetal liver; 6 = brain; 7 = fetal brain; 8 = lung; 9 = fetal lung; 10 = placenta; 11 = fibroblast; 12 = HeLa cells; M = 100-bp ladder.

copy probes for REG, REGL or REG cDNA. Clones were oriented by hybridization with ³²P-labelled T3 and T7 oligonucleotides located on both sides of the cloning site of the cosmid vector. Clones A, B, C, D, H and I were digested with several rare-cutter restriction enzymes. A BssHII fragment (11 kb long) was detected in cosmids B and I. This fragment contained a NarI site. This observation allows the positioning of clones B and I. To expand this contig in both directions, the Eco-RI ends of clones B and I were used as probes for the other four clones. One extremity of clone I hybridized with clones A and D and the other end of clone B hybridized only with clone H. Other digestions performed by XhoI and ClaI restriction enzymes allowed us to locate the clones by alignment of the restiction sites. The total size of the cosmid contig was about 100 kb. Hybridization of the restriction fragments with a REG cDNA probe located the REGL gene in a ClaI-XhoI fragment (16 kb long) of clones B and H, the REG gene in the BssHII-XhoI fragment (16.5 kb long) of clone I and the REG pseudogene in the *XhoI-ClaI* fragment (15.5 kb long) of clone D.

Expression of REG Pseudogene

The expression of the REG and REGL genes has been reported previously. We have looked for the expression of the sequence named the REG pseudogene in several human tissues by RT-PCR tests. The primers used to amplify the cDNA were located in putative exons 2 and 3 of the REG pseudogene [7] encompassing a potential intron and and the stop codon described in this sequence. We obtained amplification (fig. 4) of two cDNA fragments (174 and 169 bp long) from pancreatic RNA. These two fragments hybridized to the REG cDNA probe and were observed only with pancreatic RNA. The quality of each mRNA preparation was verified by RT-PCR tests using primers for actin mRNA. The occurrence of two fragments was observed in at least three independent reactions starting from pancreatic RNA. The amplified fragments were cloned and sequenced. Sequence analysis of the 174-bp fragment confirmed the hypothesis of Watanabe, i.e. the exons of REG and the REG pseudogene are identical in size and aligned. Furthermore, we observed the in-frame stop codon. The sequence of the shorter fragment (169 bp) was identical to that of the REG pseudogene except for a deletion of 5 nucleotides located at the 5' junction of exon 3, suggesting that it had been generated by differential splicing. The sequence contained a complete open reading frame spanning over 56 amino acids.

Discussion

In the present paper we report the structural organization of a restricted genomic region containing three homologous genes, located on chromosome 2 in band p12. We have con-

structed a cosmid contig spanning about 100 kb including the three genes REG, REGL and another REG-related sequence, i.e. the REG pseudogene. The expression of the REG gene in the pancreas has been previously reported [7]. In prior work [9], we observed the expression of the REGL gene in the pancreas and liver. REG and REGL genes have the same exon-intron organization and their mRNAs have 91% nucleotide identity. The 5' regions of both genes contain the consensus sequence for the genes expressed in the exocrine or endocrine pancreas [15]. In the present work we observed the specific expression, in the human pancreas, of the genomic sequence previously named the REG pseudogene. Translation of its nucleotide sequence corresponding to the coding region of the REG gene indicated numerous amino acid replacements and the interruption of the translation by an in-frame stop codon [7]. For these reasons, the sequence was supposed to be a pseudogene. Although we do not have data concerning the in vivo translation of the mRNA that we detected by RT-PCR, our results indicate the presence, in the pancreas, of two mRNAs originating from this genomic sequence by alternative splicing. Analysis of the transcribed sequences indicated the presence of open reading frames. The putative reading frame of one of the mRNAs is interrupted, as expected, by a stop codon. However, this RNA might encode a polypeptide shorter than the products of REG or REGL. Analysis of the shorter mRNA indicated the presence of a complete open reading frame (56 amino acids) between the oligonucleotides used for RT-PCR. We do not know where the ATG initiating translation of the two mRNAs is located and the complete sequences of the transcripts have to be determined for further analysis. However, the probable presence of an alternative splicing site, the existence of a complete open reading frame in one transcript and the specific expression in the pancreas strongly suggest that this sequence is a genomic sequence encoding pancreas-specific protein(s).

Several REG-related genes and proteins have been reported in mammals: the rat PAP family [16-18], rat peptide 23 [19], bovine PTP [20] and human HIP/PAP [21, 22]. The human HIP/PAP gene and REG genes map to identical localizations (2p12) [23, 24]. However, HIP/PAP sequences were not detected in the YAC clone described here. Recently, two genes have been described in the mouse [25]. These genes are situated on two different chromosomes. The Reg I gene is localized on mouse chromosome 12, and the Reg2 gene is localized on chromosome 3. However, mouse Reg1 protein exhibits a higher degree of similarity to rat and human Reg proteins than mouse Reg2 and was thought to be the mouse homologue to rat and human Reg proteins. It has been suggested that the ancestral REG gene may have been duplicated prior to divergence of the mouse and rat [25]. In the human, the REG gene family members are tandemly organized in a locus that is not homologous to the loci of mouse Reg1 and Reg2 genes. It is now of interest to look for the presence of tissue-specific DNA domains that regulate the expression of these pancreas-specific genes. However, the patterns of tissue expression of the REG-related genes are different. The REG gene is expressed in the pancreas and at a low level in gastric mucosa and the kidney [7], REGL is expressed in the pancreas and liver [9], while expression of the REG pseudogene was only detected in the pancreas. A more precise structural analysis of the REG locus region has now to be performed. Determining the orientation of each gene will help towards identifying different tissue-specific promotors and enhancers; moreover, the presence of repetitive elements would argue strongly for sequential duplications of an ancestral gene that led to the genomic organization reported here.

References

- DeCaro A, Multigner L, Dagorn JC, Sarles H: The human pancreatic stone protein. Biochimie 1988;70: 1209-1214.
- 2 Giorgi D, Bernard JP, Rouquier S, Iovanna J, Sarles H, Dagorn JC: Secretory pancreatic stone protein messenger RNA: Nucleotide sequence and expression in chronic calcifying pancreatitis. J Clin Invest 1989;84:100-106.
- 3 Kimura N, Yonekura H, Okamoto H, Nagura H: Expression of human regenerating gene mRNA and its product in normal and neoplastic human pancreas. Cancer 1992;70: 1857–1863.
- 4 Moore EW, Verine H: Pathogenesis of pancreatic and biliary CaCO₃ lithiasis: The solubility product (K'sp) of calcite determined with the Ca²⁺ electrode. J Lab Clin Med 1985;106:611-618.
- 5 Multigner L, Sarles H, Lombardo D, De Caro A: Pancreatic stone protein. II: Implication in stone formation during the course of chronic calcifying pancreatitis. Gastroenterology 1985;89:387-391.
- 6 Terazono K, Yamamoto H, Takasawa S, Shiga K, Yonemura Y, Tochino Y, Okamoto H: A novel gene activated in regenerating islets. J Biol Chem 1988;263:2111-2144.
- 7 Watanabe T, Yonekura H, Terazono K, Yamamoto H, Okamoto H: Complete nucleotide sequence of the human REG gene and its expression in normal and tumoral tissues: The Reg protein, pancreatic stone protein, and pancreatic thread protein are one and the same product of the gene. J Biol Chem 1990;265: 7432-7439.
- 8 Watanabe T, Yonemura Y, Yonekura H, Suzuki Y, Miyashita H, Sugiyama K, Moriizumi S, Unno M, Tanaka O, Kondo H, Bone AJ, Takasawa S, Okamoto H: Pancreatic beta-cell replication and amelioration of surgical diabetes by Reg protein. Proc Natl Acad Sci USA 1994; 91:3589-3592.

- 9 Bartoli C, Gharib B, Giorgi D, Sansonetti A, Dagorn JC, Bergé-Lefranc JL: A gene homologous to the REG gene is expressed in the human pancreas. FEBS Lett 1993;327:289– 293.
- 10 Moriizumi S, Watanabe T, Unno M, Nakagawara K, Suzuki Y, Miyashita H, Yonekura H, Okamoto H: Isolation, structural determination and expression of a novel REG gene, human REG Iβ. Biochim Biophys Acta 1994;1217:199-202.
- 11 Gharib B, Fox MF, Bartoli C, Giorgi D, Sansonetti A, Swallow DM, Dagorn JC, Bergé-Lefranc JL: Human regeneration protein/lithostathine genes map to chromosome 2p12. Ann Hum Genet 1993;57:9-16.
- 12 Albertsen HM, Abderrahim H, Cann HM, Dausset J, Le Paslier D, Cohen D: Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. Proc Natl Acad Sci USA 1990;87: 4256-4260.
- 13 Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Laboratory Press, 1989, vol 1, pp 25-28.
- 14 Green ED, Olson MV: Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci USA 1990;87:1213-1217.
- 15 Boulet AM, Erwin CR, Rutter WJ: Cell-specific enhancers in the rat exocrine pancreas. Proc Natl Acad Sci USA 1986;83:3599–3603.
- 16 Iovanna J, Orelle B, Keim V, Dagorn JC: Messenger RNA sequence and expression of rat pancreatitisassociated protein, a lectin-related protein overexpressed during acute experimental pancreatitis. J Biol Chem 1991;266:24664-24669.
- 17 Frigerio JM, Dusetti N, Keim V, Dagorn JC, Iovanna J: Identification of a second rat pancreatitis associated protein: Messenger RNA cloning, gene structure, and expression during acute pancreatitis. Biochemistry 1993;32:9236-9241.

- 18 Frigerio JM, Dusetti N, Garrido P, Dagorn JC, Iovanna J: The pancreatitis associated protein III (PAP III), a new member of the PAP gene family. Biochem Biophys Acta 1993;1216:329-331.
- 19 Yamamoto T, Katsumata N, Tachibana K, Friesen HG, Nagy JI: Distribution of a novel peptide in the anterior pituitary, gastric pyloric gland and pancreatic islets of rat. J Histochem Cytochem 1992;40:221-229.
- 20 Gross J, Brauer AW, Bringhurst C, Corbett C, Margolies MN: An unusual bovine pancreatic protein exhibiting pH-dependent globule-fibril transformation and unique amino acid sequence. Proc Natl Acad Sci USA 1985;82:5627–5631.
- 21 Orelle B, Keim V, Masciotra L, Dagorn JC, Iovanna JL: Human pancreatitis-associated protein: Messenger RNA cloning and expression in pancreatic disease. J Clin Invest 1992;90:2284-2291.
- 22 Lasserre C, Christa L, Simon MT, Vernier P, Bréchot C: A novel gene (HIP) activated in human primary liver cancer. Cancer Res 1992;52: 5089-5095.
- 23 Dusetti N, Frigerio JM, Fox M, Swallow D, Dagorn JC, Iovanna JL: Molecular cloning, genomic organization, and chromosomal localization of the human pancreatitis-associated protein (PAP) gene. Genomics 1994;19:108–114.
- 24 Lasserre C, Simon MT, Ishikawa H, Diriong S, Nguyen VC, Christa L, Vernier P, Bréchot C: Structural organization and chromosomal localization of a human gene (HIP/PAP) encoding a C-type lectin overexpressed in primary liver cancer. Eur J Biochem 1994;224:29–38.
- 25 Unno M, Yonekura H, Nakagawara K, Watanabe T, Miyashita H, Moriizumi S, Okamoto H, Itoh T, Teraoka H: Structure chromosomal localization and expression of mouse *Reg* genes, *RegI* and *RegII*. J Biol Chem 1993;268:15974–15982.