Volume 327, number 3, 289–293 FEBS 12772 August 1993

A gene homologous to the *reg* gene is expressed in the human pancreas

Catherine Bartoli*, Bouchra Gharib, Dominique Giorgi, Alice Sansonetti, Jean-Charles Dagorn and Jean-Louis Bergé-Lefranc*

Unité 315 de l'INSERM, 46, Boulevard de la Gaye, 13258 Marseille Cedex 9, France

Received 12 June 1993

We have determined the nucleotide sequence of *reg1* a human genomic DNA fragment homologous to the *reg* gene which is expressed in the exocrine pancreas and regenerating islets. Sequence comparisons of *reg* and *reg1* suggested similar exon–intron organisation. Based on this assumption, specific oligonucleotides for *reg1* exons were used to demonstrate expression of the *reg1* gene in pancreas and liver. The proteins encoded by *reg* and *reg1* comprise 166 amino acids and differ by 22 amino acids only.

Pancreas; Lithostathine; *reg* gene

1. INTRODUCTION

Lithostatine, a secretory protein synthesized by the exocrine pancreas, controls the growth of calcium carbonate crystals in pancreatic juice [1]. Cloning lithostatine mRNA led to the demonstration that lithostatine gene expression was decreased in patients presenting with chronic calcifying pancreatitis [2]. Lithostatine was independently characterized in the regenerating endocrine pancreas and named accordingly the *reg* gene product [3]: the *reg* gene and a *reg* pseudogene were described [4]. Yet, when human genomic DNA was digested with restriction enzymes and hybridized with *reg* cDNA, we observed the presence of additional bands which could not reflect polymorphic variations. We recently demonstrated the presence in the human genome, besides the *reg* gene and the *reg* pseudo-gene, of another sequence, named *reg*1, highly related to the *reg* gene [5]. In the human genome, *reg* and *reg1* sequences have been located [5] at an identical position (2p12). We report here the nucleotide sequence of the *reg1* gene and present evidence that it codes for a transcript very similar to that of the *reg* gene. Expression of the *reg1* gene analyzed in several human tissues, was observed in pancreas and liver.

Correspondence address: J.-L. Bergé-Lefranc, Unité 315 de l'INSERM, 46, Boulevard de la Gaye, 13258 Marseille Cedex 9, France. Fax: (33) 91 80 43 19.

**Present address:* Unité 242 de l'INSERM, Faculté de Médecine, 27, boulevard Jean Moulin, 13385 Marseille cedex 5, France.

Genbank accession number: L08010.

2. MATERIALS AND METHODS

2.1. *Sequencing the* reg1 *gene*

Isolation of the *reg1* DNA fragment and determination of its restriction map have been previously described [5]. Briefly: a human genomic library (Stratagene Cloning System, La Jolla, CA) constructed in λ DASH phage vectors was screened with a human *reg* cDNA [3] probe. DNA of the positive clones was analysed using various restriction enzymes and a clone whose restriction map differed from those of *reg* and *reg* pseudo-gene [4] was further characterized. DNA from clone was digested with *Hin*dIII and *Pst*I and submitted to electrophoresis on a 0.8% agarose gel. The DNA fragments were transferred to nylon membranes and hybridized with the complete *reg* cDNA. Fragments containing sequences homologous to *reg* were subcloned in pUC18 vectors and sequenced using the dideoxy chain termination method of Sanger et al. [6] with standard modifications for sequencing double strand vectors. The nucleotide sequences were determined on both strands of the fragments.

2.2. *Detection of* reg1 *mRNA in various tissues*

Expression of the *reg1* gene was analyzed using RT-PCR tests. cDNAs were prepared from poly(A$^+$) RNA (Stratagene) using AMV reverse transcriptase and oligo dT primers. PCR primers were selected in order that their 3′ end did not cross-hybridize with *reg* sequences. The 3′ primer, CT5 (5′-CAAACAAAGGAGAACTTCTTC-3′) was localized at position 2,761–2,741 in the putative exon 6 of the *reg1* gene and the 5′ primer, CT4 (5′-CCAGACAGAGCTGCCTAATC-3′) was localized at position 1,073–1,092 in the putative exon 3. Amplification of the 457 nt fragment was carried out as follows for 30 cycles: denaturation at 94°C for 30 s, annealing at 55°C for 1 min and DNA synthesis at 72°C for 2 min. In the last cycle DNA synthesis lasted 8 min.

2.3. *Nucleotide sequence analysis of* reg1 *mRNA*

Amplification products were loaded on 6% acrylamide gels. The bands of interest were excised and the DNA was extracted and sequenced by asymmetric PCR methods as previously described [7]. RACE reactions [8] were performed in order to determine the nucleotide sequences of the ends of *reg1* mRNA. CT4 primer was the internal primer used for 3′ RACE and CT5 the internal primer for 5′ RACE. The nucleotide sequences of the amplified fragments were determined using a DNA cycle sequencing system (Stratagene). Sequences obtained were compared with those of the *reg1* genomic fragments and allowed determination of the *reg1* gene organization.

```
   aggaagggcaaagctcaacatcaacttggacagtttgccaacctgtttgtggtaagtt -181
 gatgtcatttgtgaccactcctaatgtgtgccaccaataagctattcctgatgccagaat -121
 ctcttactgtcagtgccctctgtaggccttctgatccttactccttgctccacccattgt  -61
 ttatatcatgtagttctctctcagaccctgatataaagctcctactctgtctgacctgac   -1
  AAGCCACCTCAAGTGGACAAGGCACTTACCAACAG gtaaaggggcattacaggagaag   58
 agcatgtctaacgtgggatttctcttttcattttgaggtagatacagggtgattttctg  118
 aataaaagatcccagtagtaatgaaacttaagcaagaccaaagctgatttcgggtaattt  178
 ggcctctgttatccccaaaccaaaagagaaatatctgggagtgtagctatctcagtggac  238
 ctttctgctcacaggaattcagagaggagaggatgttagaaagataacaggtgctctgct  298
 ctcttcttcaaaccctcttccctgtgttctcctacagAGATTGCTGATTTGCTCCTTAAG  358
CAAGAGATTCACTGCCGCTAAGC ATG GCT CAG ACC AAC TCG TTC TTC ATG    408
                        Met Ala Gln Thr Asn Ser Phe Phe Met      9
CTG ATC TCC TCC CTG ATG TTC CTG TCT CTG AGC CAA G gtgagatttt   455
Leu Ile Ser Ser Leu Met Phe Leu Ser Leu Ser Gln                21
 ccccacacttcccacaaccccaactctgaattcttcctccatcctcatgtataaggtt   515
 cacttgaaaaaaagcagagtcaacatcagggtttttttatgttgttcagtgatcattatg  575
 gctgattttatcccattcaaaaacaccctcaccttcattcatgggtttgagacagaattt  635
 aataggaccacttataggtgaccattgtggttgagtttatctgattgaatctatatgcga  695
 tggcagtttgggggatgtttttatgtagtcattgctaggatgagagctaaggcaaacgtg  755
 tgcagggaaaccgagagaaacttgagaaaggaggaagcctgggtctttaaaggcagaagc  815
 ctcagcctcagaattaaaggaaaacgagaactcatttatttagcctattcattgtgagct  875
 cttgtcttgagcagaggaaactagagagaaaagagataggatgcaggagggcagaagtga  935
 gcaatgccccagtattcactgtatccatatgttcttataaggacaccaagaagccccta   995
 ttcaccttccagccttttccttgcccctgagattctttcttagttatctccttttttttt 1055
ccccagGC CAG GAG TCC CAG ACA GAG CTG CCT AAT CCC CGA ATC AGC  1102
      Gly Gln Glu Ser Gln Thr Glu Leu Pro Asn Pro Arg Ile Ser    35
TGC CCA GAA GGC ACC AAT GCC TAT CGC TCC TAC TGC TAC TAC TTT   1147
Cys Pro Glu Gly Thr Asn Ala Tyr Arg Ser Tyr Cys Tyr Tyr Phe    50
AAT GAA GAC CCT GAG ACC TGG GTT GAT GCA GAT  gtgagtgaggagagc  1195
Asn Glu Asp Pro Glu Thr Trp Val Asp Ala Asp                    61
 agcaggggaagggaggcttatgaaggtagaggcagctgctaatttgcagtgtgttctgtg 1255
 gctgcaatgagataagattgatcccttccctattccaccactggtccaaaacttcccaat 1315
 ctactttatcccatcatttgacacattcccagcacagagatgctggtggtcagtgacagc 1375
 atcatcagggacatttctgtgctgtcctttttctgttacatcctctggaaggtctcagta 1435
 tatccctcacaccttccttctccactgagtgctccattttcttctccaacagCTC TAT  1493
                                                  Leu Tyr      63
TGC CAG AAC ATG AAT TCA GGC AAC CTG GTG TCT GTG CTC ACC CAG   1538
Cys Gln Asn Met Asn Ser Gly Asn Leu Val Ser Val Leu Thr Gln    78
GCG GAG GGT GCC TTC GTG GCC TCA CTG ATT AAG GAG AGT AGC ACT  1583
Ala Glu Gly Ala Phe Val Ala Ser Leu Ile Lys Glu Ser Ser Thr    93
GAT GAC AGC AAT GTC TGG ATT GGC CTC CAT GAC CCA AAA AAG  gtc 1628
Asp Asp Ser Asn Val Trp Ile Gly Leu His Asp Pro Lys Lys       107
 agtctgcagccacctctatctccttataaacatttttgagaggtaagagggacgtttaag 1688
 gtctggcaccgcaatcaccaacttttatctttttgtttgtttaaataaaagcaacctctt 1748
 tatagatcctataatgtatgagttgtgaagttcagtgtaggtagttagagacatgagctg 1808
 aaggctgaatttctgggctctgggaattcatgcacccactcattgtgtctacttctaga  1868
 aatgcatctttatgtacaacttttttccctattttgctattgtctgtcttggaagaggtcc 1928
 ctctgtagactatatagaaaatgagtagtggaggagaatctactgctggcatttgttata 1988
 cattttatacaagtgtataaaactgtacagtatattatttagtttaacactataaactaa 2048
 ataatatatcaacaactactctacagccaatgttatgctggatatgagagttctgagatt 2108
 caggaaaaaaatcagaaaccactctctgtaatgggcttttatgggtctctgtatcaaatt 2168
 ctgaacacttattatttgctagaagaggaggaggaattcggacattctagagaaggagaa 2228
 gcttagagcaaaagcagaggaaatgatatgatattcatggtgacaacaatgtttattctt 2288
 tctgctataacttggcctgtttctgagtgtgcacacaggcctggttattctattgatttt 2348
 tgagtgaccatggccctgttctggccttctccatctagAAC CGC CGC TGG CAC    2403
```

Fig. 1. Nucleotide sequence of the human *reg1* gene. Capital letters indicate exons. Introns are indicated by lower-case letters. Numbering begins at the 5′ end of the first exon of the gene. The polyadenylation recognition signal (position 2,942), the Goldberg-Hogness promoter sequence (position −29) and a CAAT-like sequence (position −147) are underlined. The predicted amino acids sequence is indicated in italics under the nucleotide sequence.

```
                                                  Asn Arg Arg Trp His   112
TGG AGT AGT GGG TCC CTG GTC TCC TAC AAG TCC TGG GAC ACT GGA  2448
Trp Ser Ser Gly Ser Leu Val Ser Tyr Lys Ser Trp Asp Thr Gly   127
TCC CCG AGC AGT GCT AAT GCT GGC TAC TGT GCA AGC CTG ACT TCA  2493
Ser Pro Ser Ser Ala Asn Ala Gly Tyr Cys Ala Ser Leu Thr Ser   142
 TGC TCA G gtgagaggcagacaatctatccacctgttgccatttccttcccacttatc 2550
Cys Ser                                                      144
 tctggggatgaacatggggactgggatagaggaaaggtaagctccttatctggaaaataa 2610
 agaagtatttcctctagtttttgttctgagtcctaggttgaggaggggctacactcctt 2670
 ctgatcctctatgtctgacacttctcattgtactatagGA TTC AAG AAA TGG AAG 2725
                                          Gly Phe Lys Lys Trp Lys   150
GAT GAA TCT TGT GAG AAG AAG TTC TCC TTT GTT TGC AAG TTC AAA  2770
Asp Glu Ser Cys Glu Lys Lys Phe Ser Phe Val Cys Lys Phe Lys   165
 AAC TAGAGGAAGCTGAAAAATGGATGTCTAGAACTGGTCCTGCAATTACTATGAAGTCA 2829
Asn                                                          166
 AAAATTAAACTAGACTATGTCTCCAACTCAGTTCAGACCATCTCCTCCCTAATGAGTTTG 2889
 CATCGCTGATCTTCAGTACCTTCACCTGTCTCAGTCTCTAGAGCCCTGAAAAATAAAAAC 2949
 AAACTTATTTTTATCCA gtgttctgtcttctgcatttgctctttctacagcccatgctt 3008
 gggtggttggggtgggaatgattgtcacactccagagcttgccatggcccatccacttgt 3068
 taaaaccccactcacattttatgtatgtcaggcttatgaacatgtggtggccttgtttat 3128
 gacaagataaaaagattaagatttcatccacaacacatgttagca
```

Fig. 1 (*continued*).

3. RESULTS

3.1. *Cloning and organization of the* reg *gene*

We have previously shown that a sequence homologous to that of the *reg* gene and the *reg* pseudogene, named *reg1*, was present in the human genome [5]. A DNA clone containing the *reg1* sequence was isolated from a human genomic library and its nucleotide sequence was established. Sequence similarities between *reg* and *reg1* were analyzed with the Bisance system (CITI 2, Centre Interuniversitaire d'Informatique, Paris) using the program of Goad and Kanehisa [9,10]. Alignment of the two sequences revealed about 75% identity.

Sequence of the *reg1* transcript had to be known in order to esablish the organization of the *reg1* gene. Regions of the *reg1* gene expected to be exonic, on the basis of similarities with the *reg* gene, were selected for *reg1*-specific oligonucleotide synthesis. These oligonucleotides (CT4 and CT5) were used in RT-PCR experiments with pancreatic cDNA. A fragment could be amplified and sequenced, providing part of *reg1* mRNA structure. Sequencing was completed on the 3′ and 5′ ends by performing RACE reactions [8]. Comparison of *reg1* mRNA and gene sequences showed that the gene was organized into six exons (Fig. 1). Consensus GT/AG sequences were present at splice sites. A TATA box and a CAAT sequence were present in positions −29 and −147, respectively. A 20 nt sequence specific of the enhancer regions of the pancreatic exocrine-specific gene [11] was observed between positions −91 and −111. The putative polyadenylation signal (AATAAA) was located in exon 6 in position 2,942.

Further comparison with the *reg* gene revealed identical organization, the sizes of the six exons being conserved. Sequence identity within exons 2 to 6 amounted to 92%. By contrast, exons 1 did not show significant homology.

3.2. *Structure and expression of the* reg1 *transcript*

Sequence of the *reg1* transcript in pancreas comprised 771 nucleotides (Fig. 1). A single open reading frame of 498 nucleotides encoded a protein of 166 amino acids. Comparison with the *reg*/lithostathine protein revealed that they had the same size and differed only by 22 amino acids out of 166 (Fig. 2). Position of cysteines and of the amino acids characteristic of calcium-dependent lectins [12] were conserved.

Besides being expressed in pancreas (Fig. 3, lane 1), *reg1* was also found expressed in liver (Fig. 3, lane 3) but not in kidney, lymphocytes, brain, lung, fibroblasts or placenta.

4. DISCUSSION

Lithostathine is an inhibitor of calcium carbonate crystal growth [1] synthesized and secreted by the acinar cells of the human pancreas. The *reg* product is a protein of unknown function expressed in islet cells, only during regeneration of the endocrine tissue. Yet, amino acid sequence comparison of lithostathine and the *reg* product pointed out that lithostathine was indeed en-

```
         10        20        30        40        50        60        70
MAQTNSFFMLISSLMFLSLSQGQESQTELPNPRISCPEGTNAYRSYCYYFNEDPETWVDADLYCQNMNSG
====  =-=====  =====  =====-=====-  ====================  ===============
MAQTSSYFMLISCLMFLSQSQGQEAQTELPQARISCPEGTNAYRSYCYYFNEDRETWVDADLYCQNMNSG
         10        20        30        40        50        60        70

         80        90       100       110       120       130       140
NLVSVLTQAEGAFVASLIKESSTDDSNVWIGLHDPKKNRRWHWSSGSLVSYKSWDTGSPSSANAGYCASL
====================-===  ============================  =-===  =  ===  ==
NLVSVLTQAEGAFVASLIKESGTDDFNVWIGLHDPKKNRRWHWSSGSLVSYKSWGIGAPSSVNPGYCVSL
         80        90       100       110       120       130       140

        150       160       170       180
TSCSGFKKWKDESCEKKFSFVCKFKN
==  -==  ====   ==  ==========
TSSTGFQKWKDVPCEDKFSFVCKFKN
        150       160
```

Fig. 2. Amino acid sequence comparison of human *reg1* (upper sequence) and *reg*/lithostathine (lower sequence) proteins. Amino acid identities (=) and conservative replacements (−) are indicated.

coded by the *reg* gene [13]. We report here the nucleotide sequence of *reg1*, a novel gene homologous to the *reg* gene. It is expressed in liver and pancreas but, in the latter, assignment to the exocrine or endocrine tissue could not be achieved with the methods used in this study.

Similarities between the *reg* and *reg1* transcripts suggest that the two transcripts are conjointly detected when the *reg* cDNA is used as probe. Hence, quantitation of *reg*/lithostathine transcript performed in the past [2] also comprised *reg1* mRNA. Specifics probes have now to be used to evaluate the relative expression of *reg* and *reg1* in the pancreas.

Similarities extend to several structural features of both proteins. A high degree of identity was observed between the signal peptide of the pre-lithostahine [14] and the putative signal peptide of the *reg1* product, indicating that the latter could also be a secretory protein. Structural characteristics which confer to the lithostathine homologies with the calcium-dependent lectins are present in the *reg1* product; the carbohydrate binding domain and position of the six half cysteines involved in the three disulfide bridges being conserved. The *reg1* product might therefore show, like lithostathine, the capacity of aggregating bacteria [15]. The other described activity of lithostathine is the inhibition of $CaCO_3$ crystal growth in pancreatic juice [1], that activity being borne by the N-terminal undecapeptide of the protein [16]. It should be stressed that the *reg1* product differs by three out of eleven amino acids in that portion of the molecule. Whether such changes have altered the inhibiting properties of the peptide remains to be demonstrated. The identical chromosomal localisation of the *reg* and *reg1* genes suggesting tandem organization, their high degree of sequence similarity and their coexpression in the human pancreas strongly support the hypothesis that the two genes result from the duplication of a common ancestral gene.

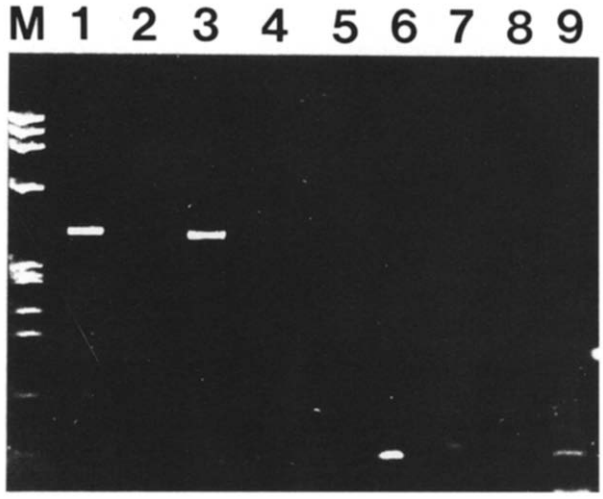

Fig. 3. Expression of the *reg1* gene in human tissues. cDNA was prepared from RNA extracted from various human tissues and *reg1* cDNA was further amplified using RT-PCR. Lane 1, pancreas; lane 2, control without RNA; lane 3, liver; lane 4, kidney; lane 5, lymphocytes; lane 6, brain; lane 7, lung; lane 8, fibroblasts; lane 9, placenta. M = markers ϕx174/*Hae*III.

Acknowledgements: The authors wish to thank M. Fontes for stimulating discussions and help in preparing the manuscript.

REFERENCES

[1] De Caro, A., Multigner, L., Dagorn, J.C. and Sarles, H. (1988) Biochimie 70, 1209–1214.
[2] Giorgi, D., Bernard, J.P., Rouquier, S., Iovanna, J., Sarles, H. and Dagorn, J.C. (1989) J. Clin. Invest. 84, 100–106.
[3] Terazono, K., Yamamoto, H., Takasawa, S., Shiga, K., Yonemura, Y., Tochino, Y. and Okamoto, H. (1988) J. Biol. Chem. 263, 2111–2114.

[4] Watanabe, T., Yonekura, H., Terazono, K., Yamamoto, H. and Okamoto, H. (1990) J. Biol. Chem. 265, 7432–7439.
[5] Gharib, B., Fox, M.F., Bartoli, C., Giorgi, D., Sansonetti, A., Swallow, D.M., Dagorn, J.C. and Bergé-Lefranc, J.L. (1993) Ann. Hum. Genet., in press.
[6] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[7] Lossi, A.M. and Bergé-Lefranc, J.L. (1989) FEBS Lett. 256, 163–166.
[8] Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA 85, 8998–9002.
[9] Goad, W.B. and Kanehisa, M.I. (1983) Nucleic Acids Res. 10, 243–263.
[10] Kanehisa, M.I. (1984) Nucleic Acids Res. 12, 203–213.
[11] Boulet, A.M., Ervin, C.R. and Rutter, W.J. (1986) Proc. Natl. Acad. Sci. USA 83, 3599–3603.
[12] Petersen, T.E. (1988) FEBS Lett. 231, 51–53.
[13] Stewart, T.A. (1989) Biochem. J. 260, 622–623.
[14] Itih, T., Tsuzuki, H., Katoh, T., Teraoka, H., Matsumoto, K., Yoshida, N., Terazono, K., Watanabe, T., Yonekura, H., Yamamoto, H. and Okamoto, H. (1990) FEBS Lett. 272, 85–88.
[15] Iovanna, J., Frigerio, J.M., Dusetti, N., Ramare, F., Raibaud, P. and Dagorn, J.C. (1993) Pancreas, in press.
[16] Bernard, J.P., Adrich, Z., Montalto, G., De Caro, A., De Reggi, M., Sarles, H. and Dagorn, J.C. (1992) Gastroenterology 103, 1277–1284.