Structural and functional analysis of a promoter of the human granulin/epithelin gene

Vijay BHANDARI*†, Rachael DANIEL*, Pheng Siew LIM‡§ and Andrew BATEMAN*||

*Endocrine Laboratory, Royal Victoria Hospital, 687 Pine Avenue West, Montreal, Quebec, Canada H3A 1A1, and ‡Institute of Molecular and Cell Biology, National University of Singapore, 10 Kent Ridge Crescent, Singapore 0511

Granulins (grns) or epithelins (epis) are peptides with molecular masses of approx. 6 kDa that modulate the growth of cells. The precursor for the grns/epis, which might itself be biologically active, is a secreted glycoprotein containing multiple repeats of the grn/epi motif. Grn/epi mRNA occurs widely *in vivo*, particularly in tissues rich in epithelial and haematopoietic cells. To understand better the role of the gene products for grn/epi it is important to determine the patterns of *grn/epi* gene expression

and how this is regulated. To assist in this we have obtained the 5' sequence of the human *grn/epi* gene, and using chimaeras of the *grn/epi* -5' sequence and the chloramphenicol acetyltransferase gene we have shown a strong promoter activity associated with the 5' sequence of the human *grn/epi* gene. We have further delineated regions of the 5' sequence that confer high-level expression on the chimaeric gene.

INTRODUCTION

The growth, development and maintenance of tissues depends on the concerted action of many secreted and membrane-bound regulatory factors. There is evidence that the products of the granulin/epithelin gene might be among the polypeptides involved in these processes. Granulins (grns) [1] or epithelins (epis) [2] are polypeptides of molecular mass approx. 6 kDa that promote or inhibit the growth of various cells in culture [2]. They are generated from a glycoprotein precursor containing 7½ repeats of the grn/epi motif [3–6]. The *grn/epi* gene is strongly expressed in cells of epithelial and haemic lineage *in vitro* and expression is also present in fibroblastic cell lines [3,5]. The mRNA for grn/epi occurs widely *in vivo*, particularly in the spleen, placenta, ovary, epididymis, adrenal, lung and kidney [4], with little or no expression in skeletal muscle, connective tissue or endothelia. The gene product for grn/epi is also abundant in spermatozoa, where it is localized to the acrosome as a glycoprotein called acrogranin [7]. The intact grn/epi precursor is secreted and might act as an autocrine growth factor for PC cells, a highly tumorigenic cell line derived from a mouse teratoma [8]. The N-terminal sequence of transforming growth factor ϵ (TGF- ϵ), a mitogen for epithelia and some mesenchymal cells, is indistinguishable from human grn/epi A, although it is not yet known whether TGF- ϵ is derived from the same gene as the grns/epis [9]. High-affinity receptors for grn/epi and TGF- ϵ have been reported on epithelial tumour cells [10,11].

To understand better the biology of the *grn/epi* gene products it is important to determine what regulates the spatial and temporal patterns of their expression. One approach to this is through structural and functional analysis of the 5' sequence of the *grn/epi* gene. Baba et al. [12] reported the structure and promoter activity of the mouse *grn/epi* gene (acrogranin). The exon/intron structure of the mouse protein-coding region is

organized in the same way as the human gene [13], with each 6 kDa *grn/epi* unit being encoded by two non-equivalent exons. The mouse *grn/epi* gene, which lacks a TATA box, showed promoter activity within a sequence of 121 bp upstream from the ATG initiator codon. The structure of the human *grn/epi* gene promoter is not known. Here we report the 5' sequence of the human *grn/epi* gene, compare it with the mouse gene and delineate regions likely to be functionally significant in the expression of the human *grn/epi* gene.

MATERIALS AND METHODS

Library screening and isolation of genomic clones

The genomic clone HGL1 was obtained as previously described [13]. To obtain clones containing more of the 5' sequence of the *grn/epi* gene, a human genomic library in EMBL-3 SP6/T7 (Clontech, Palo Alto, CA, U.S.A.) was screened with a 108 bp probe corresponding to nucleotides –35 to –142 of the 5' sequence of the precursor cDNA for human *grn/epi* gene. This probe was obtained by PCR amplification of a human bone marrow cDNA library in lambda gt11 (Clontech) with a *grn/epi*-specific reverse primer in combination with a flanking lambda gt11 primer (results not shown). Duplicate nitrocellulose filters (Schleicher and Schuell, Keene, NH, U.S.A.) were prehybridized in 5×SSC (where SSC is 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0), 5×Denhardt's reagent (where Denhardt's reagent is 0.02% polyvinylpyrrolidone, 0.02% Ficoll, 0.02% BSA), 0.2% SDS at 42 °C for 12 h. Hybridization was in 5×SSC, 2.5×Denhardt's reagent, 0.2% SDS, 50% (v/v) formamide, 10% (w/v) PEG-8000 at 42 °C for 15 h. Filters were then washed twice, for 60 min each time, in 2×SSC containing 0.1% SDS at 60 °C, and autoradiographs were obtained by exposure

Abbreviations used: CAT, chloramphenicol acetyltransferase; epi, epithelin; grn, granulin; TGF, transforming growth factor.

† Present address: Institute of Molecular and Cell Biology, National University of Singapore, 10 Kent Ridge Crescent, Singapore 0511.

§ Present address: Department of Molecular Medicine, School of Medicine, University of Auckland, Grafton Road, Auckland, New Zealand.

|| To whom correspondence should be addressed.

at -70°C with Kodak X-OMAT film (Eastman Kodak, Rochester, NY, U.S.A.) with an intensifying screen. DNA inserts from positively hybridizing clones were subcloned into Bluescript KSII and the nucleotide sequence was obtained by double-stranded dideoxy sequencing as modified for use with the Sequenase 2.0 enzyme (USB, Cleveland, OH, U.S.A.).

Primer extension analysis

The oligonucleotide primer used was a 17mer (5' AGCAGCA-ACCGGGTAGC 3') corresponding to nucleotides -35 to -51 of the precursor for human *grn/epi* (synthesized at the Sheldon Biotechnology Centre, McGill University). Primer (0.05–0.2 pmol), end-labelled with [^{32}P]ATP (7000 Ci/mmol; ICN Biomedicals, Costa Mesa, CA, U.S.A.), was hybridized to 20 μg of total RNA in a solution of 40 mM Pipes, pH 6.4, 1 mM EDTA, 400 mM NaCl, 80% formamide. The reaction was heated at 85°C for 10 min, slowly cooled to 50°C over 60 min and then incubated at 37°C for 12–16 h. The nucleic acids were precipitated with ethanol and resuspended in a solution of 50 mM Tris/HCl, pH 8.3, 75 mM KCl, 3 mM MgCl_2 , 500 μM of each of dATP, dGTP, dCTP and dTTP, 10 mM dithiothreitol, 0.4 unit/ μl human placental RNase inhibitor and 0.1 $\mu\text{g}/\mu\text{l}$ actinomycin D. The primer was extended with 20 units/ μl M-MLV reverse transcriptase at 37°C for 1 h. Unextended and template RNA were digested by RNase reaction mix (2.3 vol. of 100 $\mu\text{g}/\text{ml}$ salmon sperm DNA, 20 $\mu\text{g}/\text{ml}$ DNase-free RNase A in 100 mM NaCl, 10 mM Tris/HCl, pH 7.5, 1 mM EDTA) at 37°C for 15 min. The extended single-stranded DNA products were extracted with phenol/chloroform (1:1, by vol.), precipitated with ethanol, resuspended in sequencing gel loading buffer (0.3% Bromophenol Blue, 0.3% Xylene Cyanol FF, 10 mM EDTA, pH 7.5, 97.5% formamide; Pharmacia) and then analysed on an 8% (w/v) polyacrylamide/8 M urea sequencing gel.

Construction of human *grn/epi* promoter/chloramphenicol acetyltransferase (CAT) expression plasmids

The chimaeric *grn/epi* promoter/CAT plasmids pGRN-1662, pGRN-638, pGRN-492 and pGRN-342 were constructed by ligation of various 5' sequence fragments of the *grn/epi* gene into the promoterless CAT-gene-containing plasmid pBLCAT3. The 5' sequence restriction fragments of the human *grn/epi* gene were obtained by digesting a *SacI/XhoI* subclone of HGL1 with *SstII* and *PstI*, *SstII* and *SmaI*, and *SstII* and *XhoI* to generate pGRN-342, pGRN-492 and pGRN-638 respectively. A *SacI* subclone of HGLP26 was digested with *SstII* and *XhoI* to generate pGRN-1662. All restriction fragments were purified from low-melting-point agarose by phenol extractions followed by ethanol precipitation. The fragments were cloned into pBLCAT3 by blunt-end ligation into the *BamHI* cloning site of pBLCAT3, which was filled in with dNTPs by using T4 DNA polymerase. The orientation of the inserts with relation to the CAT gene coding sequence was verified by restriction enzyme analysis with *PstI*, *SacI* and *XhoI*, and by double-stranded dideoxynucleotide sequencing. pGRN-154, pGRN-116, pGRN-63 and pGRN-32 were generated by controlled digestions from the 5' end of pGRN-342 with exonuclease III as described elsewhere [14]. The 5' ends of these *grn/epi* promoter/CAT constructs were verified by double-stranded dideoxynucleotide sequencing. The CAT vectors pBLCAT2 and pBLCAT3 [15] were generously supplied by Dr. G. Hendy (Royal Victoria Hospital, Montreal, Quebec, Canada).

Cell culture and transfections

A549, CaSki, NIH-3T3 and COS-7 cells were purchased from American Type Tissue Culture. Cells were maintained in RPMI-1640 medium (for A549 and CaSki cells) or Dulbecco's minimal essential medium (for COS-7 and NIH-3T3 cells) supplemented with 10% fetal bovine serum and 10 mg/ml gentamycin in an atmosphere of 5% CO_2 in air at 37°C . Before transfection, cells growing as monolayers were trypsinized and replated in six-well (35 mm) tissue-culture plates or 60 mm tissue-culture dishes, and incubated until 50–70% confluency. Cells were transfected with 2 μg of chimaeric *grn/epi* promoter/CAT plasmid by liposome-mediated transfection with 10 μl of 1 mg/ml Lipofectin® reagent (Gibco-BRL) for 6 h as described by the manufacturer. Transfection efficiency with the liposome-mediated method is consistent [16–18] thus eliminating the need to normalize for transfection efficiency. Cell extracts were assayed for CAT activity 48–52 h after transfection. pBLCAT2, which contains the thymidine kinase promoter upstream of the CAT gene, was used as a positive control.

Assays for CAT activity

Transfected cells were harvested and resuspended in 100 μl of 0.25 M Tris/HCl, pH 7.8. The cells were disrupted by three cycles of freezing and thawing followed by centrifugation to remove cell debris. Equal volumes of cell extract were assayed for CAT activity by a modified procedure of Gorman et al. [19]. Briefly, the cell extracts were heated for 10 min at 65°C , centrifuged and 70 μl of the supernatant was added to a solution consisting of 66.5 μl of 0.25 M Tris/HCl, pH 7.8, 2 μl of [^{14}C]chloramphenicol (60 mCi/mmol; ICN), 10 μl of 40 mM acetyl-CoA and 1.5 μl of 0.5 M EDTA, pH 8. The samples were incubated for 2 h at 37°C , then extracted with 1 ml of ethyl acetate. A 900 μl portion of the organic phase was removed and dried under vacuum. The resulting pellet was resuspended in 25 μl of ethyl acetate and the unreacted and acetylated forms of the [^{14}C]chloramphenicol were separated by TLC (Silica Gel 1B; J. T. Baker, Phillipsburg, NJ, U.S.A.) with chloroform/methanol (19:1, v/v) as solvent, and, after autoradiography, areas corresponding to unreacted and acetylated forms were cut out from the TLC sheets and the β -particle emission was determined by scintillation counting. The CAT activity was then quantified as the percentage conversion of the [^{14}C]chloramphenicol to its acetylated form per μg of protein used in each assay. The protein content of each extract was determined by the Bradford protein analysis as described by the manufacturer with reagents supplied in kit form (Bio-Rad).

RESULTS

Isolation of genomic clones containing the 5' flanking region of the *grn/epi* gene

We have previously reported the isolation and nucleotide sequence of two genomic clones, HGL1 and HGL23, containing the protein-coding region of the *grn/epi* gene [13]. Partial characterization of HGL1 suggested that there was an intron in the 5' sequence of the human *grn/epi* gene seven nucleotides upstream of the translation initiation methionine [3, 13]. We have further characterized HGL1 by restriction-enzyme mapping, Southern hybridization with probes specific to the 5' sequence of the *grn/epi* gene, nucleotide sequencing and comparison with the cDNA sequence for human *grn/epi* (results not shown), and

Figure 1 Structure and nucleotide sequence of the 5' flanking and untranslated region of the human *grn/epi* gene

Upper panel, schematic and partial restriction map of HGLP26 containing the 5' sequence and part of the *grn/epi* protein coding region. The ATG codon is boxed. The hatched area indicates region sequenced. Lower panel, nucleotide sequence of the 5' sequence of the human *grn/epi* gene. Nucleotide +1 denotes the A residue of the translation initiation codon (ATG), and residues preceding it are indicated by negative numbers. The intron nucleotides are not numbered. The putative transcription initiation sites are indicated by a dot. The sequences of potential regulatory elements are doubly underlined, with r indicating reverse complement. The CCAAT boxes are underlined. An 11 bp direct repeat followed by eight 5 bp repeats in tandem similar to the *v-ras* oncogene repeats are doubly overlined. The *Alu* repeats are overlined.

confirm the presence of an intron in the 5' sequence of the *grn/epi* gene. The intron is approx. 3 kb, determined by restriction-enzyme mapping. An intron of 2.2 kb is also present at the same

position in the 5' untranslated region of the mouse *grn/epi* gene [12].

Analysis of HGL1 revealed that it contained part of the coding region of the *grn/epi* gene, the 3 kb intron and 675 bp of 5' sequence. To obtain genomic clones containing more of the 5' sequence, a human genomic library was probed with a cDNA fragment corresponding to nucleotides -35 to -142 of the 5' sequence of the *grn/epi* gene. Three clones, HGLP21, HGLP22 and HGLP26, were obtained. Restriction mapping of HGLP21, HGLP22 and HGLP26 followed by Southern hybridization with *grn/epi*-specific cDNA fragments indicated that they overlapped and that each clone contained a substantial portion of the 5' sequence of the *grn/epi* gene. We chose HGLP26, which has an insert of approx. 17 kb and overlaps with HGL1 by approx. 5.5 kb, for further analysis. Two overlapping restriction fragments from clone HGLP26, a 7.7 kb *SacI* fragment and a 4.1 kb *XhoI* fragment containing the 5' sequence of the *grn/epi* gene were subcloned into Bluescript KSII, characterized by restriction analysis and partly sequenced to obtain 2388 bp of the 5' sequence of the human *grn/epi* gene.

Structural analysis of the 5' flanking region of the human *grn/epi* gene

The nucleotide sequence of the 5' region of the human *grn/epi* gene is shown in Figure 1. In the human *grn/epi* gene transcription starts from at least two sites (see below), so we have numbered the 5' sequence of the *grn/epi* gene such that -1 refers to the nucleotide preceding the translation initiation codon (ATG). The 3 kb intron is not included in the numbering scheme. The classic consensus sequence for a TATA box, TATA(A/T)A [20] is missing, although a variant TATA sequence, TATATT, is present at nucleotides -148 to -143. This TATATT motif lies 35 bp upstream of one of the transcription initiation sites of the *grn/epi* gene and thus might serve as a functional TATA box for this start site [20]. The 5' sequence contains four CCAAT boxes, at nucleotides -103 to -99, -133 to -129, -251 to -247, and -398 to -394. An adjacent 11 bp direct repeat, GAATAGGGCAG, followed by eight repeats in tandem with the consensus sequence C/TAG/AGG/A occurs at nucleotides -928 to -871. Searches of the nucleic acid databases with the BLAST e-mail server offered by the National Library of Medicine, NIH, revealed that the 5' sequence of the rat sarcoma virus *v-ras* oncogene contains a similar stretch of nucleotides but in reverse complement, i.e. T/CCT/CTA/G, repeated $7\frac{1}{2}$ times consecutively [21]. The significance of this motif is unknown. Also found within the 5' sequence of the human *grn/epi* gene are 2 *Alu* repeats, several homopurine/homopyrimidine tracts and many potential binding sites for transcription factors. Multiple copies of palindromic sequences and direct or inverted repeats are also found, similar sequences being associated with enhancer and/or repressor activity [22].

Transcription initiation sites in the human *grn/epi* gene

Primer extension analysis was used to identify the probable transcription start site(s) of the human *grn/epi* gene. A 17-nucleotide primer specific to the 5' sequence -35 to -51 was annealed to total RNA from human fetal intestine and U937 and A549 cells and extended with reverse transcriptase. The extended products were separated by polyacrylamide gel electrophoresis and revealed by autoradiography. A representative primer extension experiment is shown in Figure 2. Two major primer

Figure 2 Transcription initiation sites of the human *grn/epi* gene

A 17 nucleotide primer corresponding to nucleotides —35 to —51 of the precursor to human *grn/epi* was annealed to 20 µg of total RNA from U937 cells (lane 1), A549 cells (lane 2), human fetal intestine (lane 3) and tRNA (lane 4) and extended with M-MLV reverse transcriptase. The primer-extended products were analysed on an 8% (w/v) polyacrylamide/8 M urea gel. A sequencing ladder generated with the same primer and the *grn/epi* clone HGLP26 was run in parallel as a size standard. Transcription initiation sites are indicated by the arrows at approx. 108 and 214 nucleotides.

extension products. of 73 and 179 bp were obtained. A control reaction with tRNA did not result in any extension products. These results indicate that the initiation of transcription of the *grn/epi* gene is heterogeneous, occurring approximately at nucleotides —108 and —214.

Activity of the promoter for human *grn/epi* gene

Chimaeric human *grn/epi* promoter/CAT constructs, each terminating at the same 3' site (nucleotide —21) but sequentially truncated at the 5' end (Figure 3) were tested for their ability to promote the expression of CAT activity in mammalian cell lines. The chimaeric constructs are named according to the length of the cloned 5' sequence.

To map the promoter of the human *grn/epi* gene, four mammalian cell lines, A549 (human lung carcinoma cells), CaSki (human cervical carcinoma cells), COS-7 (SV-40-transformed monkey kidney fibroblast-like cells) and NIH-3T3 (mouse fibroblasts), were used to analyse the promoter activity of the *grn/epi*/CAT chimaeras. A representative CAT analysis for each cell line is given in Figure 4, and the relative CAT activities of pGRN-32 to pGRN-1662 and the control plasmids, pBLCAT2 and pBLCAT3, are given in Table 1. The relative transcriptional activity of the *grn*/CAT chimaeras in these cell lines was consistent with the levels of expression of the *grn/epi* in these cells as determined by Northern blot analysis ([3], and V. Bhandari and A. Bateman, unpublished work).

In A549, CaSki, COS-7 and NIH-3T3 cell lines the transcriptional activities of pGRN-32 and pGRN-63 were not significantly different from the promoterless control plasmid, pBLCAT3 (Table 1), in agreement with the observation that pGRN-32 and pGRN-63 do not contain the predicted transcriptional start sites. As the length of the 5' sequence within the *grn/epi*/CAT construct is increased (pGRN-116, pGRN-154, pGRN-342) there is a corresponding increase in transcriptional activity. The strongest promoter activity is observed with pGRN-492 for A549 and CaSki cells, and with pGRN-638 for COS-7 and NIH-3T3 cells although the level of transcriptional activity is still

Figure 3 Schematic representation of the human *grn/epi*/CAT chimaeric plasmids

All constructs contain a common 3' end (nucleotide —21) but are sequentially truncated at the 5' end with each construct named according to the length (in bp) of the 5' sequence. The relative positions of the CCAAT boxes and putative response elements are shown. The letter r indicates reverse complement. The solid bar indicates the position of the tandem repeats similar to that found in the 5' sequence of the rat sarcoma virus *v-src* oncogene.

Figure 4 Representative experiment showing the activity of the promoter for the *grn/epi* gene in (A) A549 human lung carcinoma cells, (B) CaSki human cervical carcinoma cells, (C) COS-7 monkey kidney fibroblasts and (D) NIH-3T3 mouse fibroblasts

Cells were transfected with 2 μ g of the various *grn/epi*/CAT constructs by using Lipofectin® reagent, and CAT activity was determined 48–52 h after transfection as described in the Materials and methods section. pBLCAT2 is a CAT plasmid containing the thymidine kinase promoter; pBLCAT3 is a promoterless construct. pGRN-32, pGRN-64, pGRN116, pGRN-154, pGRN-342, pGRN-492, pGRN-638 and pGRN-1662 are CAT plasmids containing the 5' sequence of the human *grn/epi* gene with the names of the constructs reflecting the length (in bp) of the 5' sequence.

greater than that observed for the promoterless construct pBLCAT-3. This decrease in transcriptional activity is not observed in COS-7 cells.

Table 1 Promoter activity for the *grn/epi* gene in cell lysates of A549 (human lung carcinoma), CaSki (human cervical carcinoma), COS-7 (monkey kidney fibroblast-like) and NIH-3T3 (mouse fibroblast) cells

The levels of the unreacted and acetylated forms of [14 C]chloramphenicol were assayed for β -emission by scintillation counting as described in the Materials and methods section. The results are relative CAT activities expressed as percentages of [14 C]chloramphenicol conversion per μ g of protein, relative to the CAT activity of pBLCAT3, and are means $\pm$ S.D. for three or four separate experiments.

CAT construct	Relative CAT activity (% of pBLCAT3)			
	A549	CaSki	COS-7	NIH-3T3
pBLCAT2	6.02 $\pm$ 0.66	37.81 $\pm$ 18.86	3.74 $\pm$ 0.90	0.85 $\pm$ 0.10
pBLCAT3	1.00	1.00	1.00	1.00
pGRN-32	1.12 $\pm$ 0.30	3.50 $\pm$ 2.78	1.14 $\pm$ 0.22	0.94 $\pm$ 0.20
pGRN-63	1.42 $\pm$ 0.19	0.94 $\pm$ 0.14	1.05 $\pm$ 0.27	0.87 $\pm$ 0.20
pGRN-116	3.60 $\pm$ 0.61	2.86 $\pm$ 1.08	2.20 $\pm$ 0.54	0.96 $\pm$ 0.28
pGRN-154	3.24 $\pm$ 0.10	5.27 $\pm$ 1.51	11.09 $\pm$ 2.31	0.86 $\pm$ 0.10
pGRN-342	7.37 $\pm$ 1.88	7.17 $\pm$ 1.92	17.33 $\pm$ 4.44	3.07 $\pm$ 1.23
pGRN-492	14.32 $\pm$ 1.19	18.26 $\pm$ 3.29	15.46 $\pm$ 3.33	4.64 $\pm$ 2.84
pGRN-638	11.98 $\pm$ 0.39	7.52 $\pm$ 3.11	23.34 $\pm$ 5.78	4.80 $\pm$ 3.38
pGRN-1662	4.82 $\pm$ 1.03	8.10 $\pm$ 2.71	20.07 $\pm$ 7.96	2.08 $\pm$ 1.30

DISCUSSION

The aims of this study were to determine the 5' sequence of the human *grn/epi* and to delineate regions likely to be functionally important in maintaining the expression of the gene. We have shown previously [4] that the expression of the *grn/epi* gene *in vivo* occurs predominantly in epithelial cells, with little expression in mesenchymal cells, muscle or endothelium. In tissue culture, however, some cells of a mesenchymal lineage such as 3T3 fibroblasts do express the *grn/epi* gene [3]. We therefore chose to look at the promoter activity of the 5' sequence of the human *grn/epi* gene in cell lines of epithelial origin (A549 and CaSki), and also in non-transformed fibroblast-like cells (NIH-3T3) and a transformed fibroblastic cell line (COS 7) that expresses the *grn/epi* gene at a level equivalent to epithelial cell lines (V. Bhandari and A. Bateman, unpublished work). In future work it will be important to extend this to more physiological cell types and to cells growing under more physiological conditions. The 5' sequence of the human *grn/epi* gene is a strong functional promoter in the epithelial cell lines A549 and CaSki, and in COS-7-transformed fibroblasts. It is a weaker promoter in NIH-3T3 cells. Analysis of the transcriptional activity of the *grn/epi*/CAT chimaeras suggests that the 91-nucleotide sequence between –84 (pGRN-63) and –175 (pGRN-154) is sufficient to confer promoter activity for *grn/epi* in A549, CaSki and COS-7 but not in NIH-3T3 cells. However, further elements located between –175 (pGRN-154) and –513 (pGRN-492) are needed for high levels of expression of the *grn/epi* gene. Maximal promoter activity was seen with pGRN-492 in Caski and A549 cells, and with pGRN-638 in COS-7 and NIH-3T3 cells. Promoter activity in the larger construct, pGRN-1662, was diminished relative to pGRN-492 in Caski, A549 and NIH-3T3 cells but still remained elevated above the basal expression obtained with the promoterless pBLCAT3. This suggests that the 5' sequence of the *grn* contains both positive and negative response elements. Interestingly the expression of pGRN-1662 remained as high as pGRN-492 in COS-7 cells, suggesting that the transcriptional regulation of the *grn/epi* gene might vary between cells. Similar inhibitory regulation residing 5' to positive regulatory elements has been shown for the genes of other growth factors such as TGF β 1 [23] and basic fibroblast growth factor [24]. Negative regulation of the

gene for TGF β 1 is influenced by external factors such as the extracellular matrix [25]. It will be interesting to determine whether exogenous factors can have a negative effect on the expression of the *grn/epi* gene.

Features that are frequently found in the promoters of genes with growth-related functions include multiple transcriptional initiation sites, the lack of appropriately positioned TATA and CCAAT boxes, a high G+C content and one or more GC box elements that are binding sites for the transcription factor Sp1 [26]. Primer extension analysis of the human *grn/epi* gene indicates that there are at least two major transcriptional start sites, at positions -108 and -214. Consistent with this we previously reported two human *grn/epi*-precursor mRNAs differing in size by about 100 bp [3]. The *grn/epi* gene has four potential CCAAT boxes, although their functional status is not known. It has no typical TATA box but a potential variant of the TATA box is seen at positions -148 to -143. The mouse gene also does not contain a typical TATA sequence [12]. The 5' sequence of the human gene has a G+C content of 56% over the 500 bp upstream of the translation initiation site (omitting the intron sequences). This is relatively low compared with several other TATA-less genes; for example G+C contents of 88%, 80%, 79%, 77% and 70% have been reported for promoters of TGF β 1, the epidermal growth factor receptor, platelet-derived growth factor A, insulin receptor and TGF α genes [23,27-30].

Many potential regulatory elements that have been associated with functional activity in promoters of other eukaryotic genes are scattered throughout the 5' flanking region of the human *grn/epi* (Figure 1): 25 sequences show a 7 out of 8 bp match with the AP-1 consensus sequence TGA(C/G)TCAG [31,32]. Numerous potential AP2-binding sites [33,34] are present; nine sites show a 100% match. The promoter for the *grn/epi* contains four sequences that show a 7 out of 8 bp match with the CRE/ATF consensus sequence [35]. Several consensus sequences involved in cytokine and growth-factor-regulated gene expression are present in the 5' sequence of the human *grn/epi*. These include four copies in reverse complement of the IL6-RE hexanucleotide and three potential NFL-IL6-binding sites, both of which have been implicated in regulating gene expression in inflammation, the acute phase response and haematopoiesis [36,37]. Two sequences match the TGF β -1 inhibitory element, which is conserved in the promoter regions of TGF β -1-inhibited genes [38]. There were no steroid hormone-like regulatory elements, suggesting that the high levels of *grn* mRNA detected in reproductive tissues [4,6] are not the direct result of sex-steroid action on these tissues.

A comparison of the 5' sequences of the genes for human and mouse *grn/epi* gene by using the dot-plot identity matrix DOTPLOT program offered by DNASTAR revealed two major regions of conservation. There is 73% identity over the region of sequence 300 bp 5' of the translated start codon (not including the intron) (Figure 5). The first conserved region covers most of the -175 to -363 core that seems to be crucial in sustaining CAT activity in the various *grn/epi*/CAT chimaeras. Several putative binding sites for transcription factors are conserved between the two genes in this region, such as the nuclear factors SP1 (nucleotides -87 to -79, human sequence) [39], NFIL6 (nucleotides -255 to -247) [37] and CF1 (nucleotides -232 to -226) [40], as well as the tissue-specific transcription factors TCF1 (nucleotides -175 to -171, and -183 to -179) [41] and TEF2 (nucleotides -220 to -227) [42]. Also conserved within this region is a CCAAT box at nucleotides -133 to -129. The second conserved region occurs between nucleotides -950 and -1430 of the human *grn/epi*, corresponding to nucleotides -460 and -940 of the mouse gene (Figure 5). Most of this region is included in the pGRN-1662 CAT construct and appears, in some

Figure 5 Dot matrix analysis of the 5' flanking sequence of the genes for human (x-axis) and mouse (y-axis) *grn/epi* gene

The nucleotide sequence of the mouse 5' flanking region was obtained from Baba et al. [12]. The numbering of nucleotides is based on the initiator ATG codon. Each dot represents a 65% match (stringency) within a 20-nucleotide stretch (window size). Two highly conserved regions between the human and mouse sequence occur over a region of 300 bp 5' of the translation start codon and between nucleotides -950 and -1430 of the human *grn/epi* gene and nucleotides -460 and -940 of the mouse gene respectively. The dot matrix analysis was performed with the program DOTPLOT (DNASTAR, Madison, WI, U.S.A.).

cell lines, to exert a negative effect on the more 3' elements. It contains a conserved potential binding site for the ubiquitous transcription factor PEA3 [43] (nucleotides -1377 to -1372, human sequence) as well as two conserved short homopurine/homopyrimidine tracts. Functionally significant polypurine/polypyrimidine tracts are found in the promoters of several essential growth-control genes such as platelet-derived growth factor A, Ki-ras, epidermal growth factor receptor and *c-myc* [44-48].

In summary, the 5' sequence of the human *grn/epi* gene lacks a conventional TATA box but has several potential CCAAT boxes, contains several copies of GC box elements, and exhibits at least two transcription initiation sites. The *grn/epi*/CAT expression experiments delineate the sequence between -175 and -513 as markedly increasing the basal promoter activity. This region is strongly conserved between the genes for human *grn/epi* and mouse *grn/epi* (acroggranin). Sequences upstream from -650 might exert suppressive regulation over more 3' elements in some cells. Analysis of the *grn/epi* 5' sequence against known regulatory elements suggest that cAMP, agents stimulating protein kinase C such as phorbol esters, and several cytokines, especially those involved in the inflammatory response, might play a role in regulating the expression of the *grn/epi* gene. The large number and variety of putative regulatory elements implies considerable complexity and versatility in the regulation of expression of the *grn/epi* gene.

REFERENCES

- Bateman, A., Belcourt, D., Bennett, H. P. J., Lazure, C. and Solomon, S. (1990) *Biochem. Biophys. Res. Commun.* **173**, 1161-1168
- Shoyab, M., McDonald, V. L., Byles, C., Todaro, G. and Plowman, G. D. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7912-7916
- Bhandari, V., Palfree, R. G. E. and Bateman, A. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1715-1719
- Bhandari, V., Giald, A. and Bateman, A. (1993) *Endocrinology (Baltimore)* **133**, 2682-2689
- Plowman, G. D., Green, J. M., Neubauer, M. G., Buckley, S., McDonald, V. L., Todaro, G. J. and Shoyab, M. (1992) *J. Biol. Chem.* **267**, 13073-13078
- Baba, T., Hoff, H. B., Nemoto, H., Lee, H., Orth, J., Arai, Y., Anakwe, O. O. and Gerton, G. L. (1990) *Biol. Reprod.* **42**, 317-328
- Gerton, G. L. 1993 *Mol. Reprod. Develop.* **34**, 233-243
- Zhou, J., Gao, G., Crabb, J. W. and Serrero, G. (1993) *J. Biol. Chem.* **268**, 10683-10689

- 9 Parnell, P. G., Wunderlich, J., Carter, B. and Halper, J. (1992) *Growth Factors* **7**, 65–72
- 10 Culouscou, J.-M., Carlton, G. W. and Shoyab, M. (1993) *J. Biol. Chem.* **268**, 10458–10462
- 11 Parnell, P. G., Carter, B. J. and Halper, J. (1995) *J. Recept. Signal Transduct. Res.* **15**, 747–756
- 12 Baba, T., Nemoto, H., Watanabe, K., Arai, Y. and Gerton, G. L. (1993) *FEBS Lett.* **322**, 89–94
- 13 Bhandari, V. and Bateman, A. (1992) *Biochem. Biophys. Res. Commun.* **188**, 57–63
- 14 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1994) *Current Protocols in Molecular Biology*, John Wiley and Sons, New York
- 15 Luckow, B. and Schutz, G. (1987) *Nucleic Acids Res.* **15**, 5490
- 16 Behr, J. P., Demeneix, B. A., Loeffler, J. P. and Perez, J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6982–6986
- 17 Demeneix, B. A., Kley, N. and Loeffler, J. P. (1990) *DNA Cell Biol.* **9**, 335–345
- 18 Basthel, F., Boutiller, A. L., Giraud, P. and Demeneix, B. A., Behr, J. P. and Loeffler, J. P. (1992) *Neuroscience* **9**, 291–312
- 19 Gorman, C. M., Moffat, L. F. and Howard, B. H. (1982) *Mol. Cell. Biol.* **2**, 1044–1048
- 20 Lodish, H., Baltimore, D., Berk, A., Zipursky, S. L., Matsudaira, P. and Daenell, J. (eds.) (1995) *Molecular Cell Biology*, 3rd edn., Scientific American Books, New York
- 21 Rasheed, S., Norman, G. L. and Heidecker, G. (1983) *Science* **221**, 155–157
- 22 Isseman, I. and Green, S. (1990) *Nature (London) (London)* **347**, 645–650
- 23 Kim, S.-J., Glick, A., Sporn, M. B. and Roberts, A. B. (1989) *J. Biol. Chem.* **264**, 402–408
- 24 Florkiewicz, R. Z., Shibata, F., Barankiewicz, T., Baird, A., Gonzales, A.-M., Florkiewicz, E. and Shah, N. (1991) *Ann. N.Y. Acad. Sci.* **638**, 109–126
- 25 Streuli, C. H., Schmidhauser, C., Kobrin, M., Bissell, M. J. and Derynck, R. (1993) *J. Cell Biol.* **120**, 253–260
- 26 Mitchell, P. J. and Tjian, R. (1989) *Science* **245**, 371–378
- 27 Ishii, S., Xu, Y., Roe, B. A., Merlino, G. T. and Pastan, I. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4920–4924
- 28 Bonthron, D. T., Morton, C. C., Orkin, S. H. and Collins, T. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1492–1496
- 29 Araki, E., Shimida, F., Uzawa, H., Mori, M. and Ebina, Y. (1987) *J. Biol. Chem.* **262**, 16186–16191
- 30 Jakobovits, E. B., Schlokot, U., Vannice, J. L., Derynck, R. and Levinson, A. D. (1988) *Mol. Cell. Biol.* **8**, 5549–5554
- 31 Curran, T. and Franza, B. R. (1988) *Cell* **55**, 395–397
- 32 Angel, P., Hattori, K., Smeal, T. and Karin, M. (1988) *Cell* **55**, 875–885
- 33 Williams, T. and Tjian, R. (1991) *Genes Dev.* **5**, 670–682
- 34 Imigawa, M., Chiu, R. and Karin, M. (1987) *Cell* **51**, 251–260
- 35 Roesler, W. J., Vandenbark, G. R. and Hanson, R. W. (1988) *J. Biol. Chem.* **263**, 9063–9066
- 36 Hattori, M., Abraham, L. J., Northemann, W. and Fey, G. H. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2364–2368
- 37 Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T. and Kishimoto, T. (1990) *EMBO J.* **9**, 1897–1906
- 38 Kerr, L. D., Miller, D. B. and Matrisian, L. M. (1990) *Cell* **61**, 267–278
- 39 Saffer, J. D., Jackson, S. P. and Annarella, M. B. (1991) *Mol. Cell. Biol.* **11**, 2189–2199
- 40 Riggs, K. J., Merrell, K. T., Wilson, G. and Calame, K. (1991) *Mol. Cell. Biol.* **11**, 1765–1769
- 41 Van deWetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) *EMBO J.* **10**, 123–132
- 42 Davidson, I., Xiao, J. H., Rosales, R., Staub, A. and Chambon, P. (1988) *Cell* **54**, 931–942
- 43 Wasyluk, C., Flores, P., Gutman, A. and Wasyluk, B. (1989) *EMBO J.* **8**, 3371–3378
- 44 Wang, Z. Y. and Duell, T. K. (1992) *Biochem. Biophys. Res. Commun.* **188**, 433–439
- 45 O'Neill, D., Bornschlegel, K., Flamm, M., Castle, M. and Bank, A. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8953–8957
- 46 Hoffman, E. K., Trusko, S. P., Murphy, M. and George, D. L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2705–2709
- 47 Johnson, A. C., Jinno, Y. and Merlino, G. T. (1988) *Mol. Cell. Biol.* **8**, 4174–4184
- 48 Postel, E. H., Mango, S. E. and Flint, S. J. (1989) *Mol. Cell. Biol.* **9**, 5123–5133